

## Asymptomatic infection and herd immunity of COVID-19 in Wuhan and Japan

Junko Kurita<sup>1)</sup>, Tamie Sugawara<sup>2)</sup>, Yasushi Ohkusa<sup>2)</sup>

1) Department of Nursing, Tokiwa University, Ibaraki, Japan

2) National Institute of Infectious Diseases, Tokyo, Japan

Corresponding author: Junko Kurita, [kuritaj@tokiwa.ac.jp](mailto:kuritaj@tokiwa.ac.jp)

Keywords: peak, COVID-19, SIR model, proportion of asymptomatic cases , Wuhan, Japan, herd immunity

## Abstract

**Background:** The COVID-19 outbreak has shown two inconsistent phenomena: its reproduction number is almost two; and it shows earlier and lower peaks for new cases and the total number of patients.

**Object:** To resolve this inconsistency, we constructed a mathematical model to explain these phenomena.

**Method:** To outbreak data from Wuhan, China and Japan, we applied a susceptible–infected–recovery model with the proportion of asymptomatic patients among infected people ( $q$ ) as a key parameter for estimation, along with the basic reproduction number ( $R_0$ ).

**Results:** The first outbreak peak was recorded in Japan on April 3 for those infected on March 29. Their  $R_0$  and  $q$  were estimated respectively as 3.19 and 99.32% in Wuhan. In Japan, these were estimated around the peak as 2.96 and 99.99%.

**Discussion and Conclusion:** By introducing a very high proportion of asymptomatic cases, the two inconsistent phenomena might be resolved. Especially in Japan, the asymptomatic cases were 60 times higher than those of China.

## Introduction

The COVID-19 outbreak emerged in Wuhan, China on December 1 [1]. Approximately 50,000 cases in all were reported by March 6 [2]. The initial case of COVID-19 in Japan was a patient who showed symptoms when returning from Wuhan, China on January 3, 2020. As of April 26, 2020, the Ministry of Labour, Health and Welfare (MLHW) in Japan announced that there were 7,741 cases in Japan, including asymptomatic but infected people, but excluding those infected on a large cruise ship: the Diamond Princess [3]. In fact, the peak in Wuhan occurred on February 12. In Japan, the first peak was observed at the end of March. Subsequently, despite a declining epidemic curve, the government declared a state of emergency in Japan on April 7.

The COVID-19 outbreaks have two associated and inconsistent phenomena: their respective reproduction numbers are almost two; and outbreaks show earlier and lower peaks of new cases and the total number of patients. That reproduction number of two indicates that the peak will be reached when half of the population is infected. However, the total number of patients in Wuhan up to the peak was around 48,000. In Japan, the peak was reached after approximately 7000 cases. A similar phenomenon was confirmed in South Korea. How can one reconcile these two inconsistent characteristics

of the COVID-19 outbreak? To explain the two phenomena and resolve the difficulty, we constructed a mathematical model.

## Methods

We applied a simple susceptible–infected–recovery (SIR) model [4–6] to the epidemic curve of Wuhan, with 11 million population, and Japan, with 120 million populations. We assume an incubation period that conforms to the empirical distribution in Japan.

The proportion of asymptomatic people among those infected ( $q$ ) is a crucially important parameter for estimation in the model, as is the basic reproduction number ( $R_0$ ). For simplicity, we assumed equal infectiousness among asymptomatic cases as symptomatic cases [7].

Data used for Wuhan were the numbers of symptomatic patients from January 20 through March 5 published by Hubin [2]. For Japan, we used the number of symptomatic patients reported by the Ministry of Labour, Health and Welfare (MLHW) for January 14 – April 24, published [3] on April 26. We excluded some patients from data in Japan: those presumed to be persons infected abroad or infected as passengers on the Diamond Princess. Those patients were presumed not to represent

community-acquired infection in Japan. For onset dates of some symptomatic patients that were unknown, we estimated their onset date from an empirical distribution with duration extending from the onset to the report date among patients for whom the onset date had been reported.

The Wuhan data included no reported onset date for any patient. Therefore, we applied the empirical distribution of duration between onset to report in Japan to them.

We estimated the onset date of patients whose onset dates were not reported as follows: Letting  $f(k)$  represent this empirical distribution and letting  $N_t$  denote the number of patients for whom onset dates were not available published at date  $t$ , then the number of patients for whom the onset date was known is  $t-1$ . The number of patients for whom onset dates were not available was estimated as  $f(1)N_t$ . Similarly, the number of patients with onset date  $t-2$  and whose onset dates were not available were estimated as  $f(2)N_t$ . Therefore, the total number of patients for whom the onset date was not available, given an onset date of  $s$ , was estimated as  $\sum_{k=1} f(k)N_{s+k}$  for the long duration passing from  $s$ .

Moreover, the reporting delay for published data from MHLW might be considerable. In other words, if  $s+k$  was larger than that in the current period  $t$ , then  $s+k$  represents the future for period  $t$ . Therefore,  $N_{s+k}$  was not observable. Such a reporting

delay leads to underestimation bias of the number of patients. For that reason, it must be adjusted as  $\sum_{k=1}^{t-s} f(k)N_s+k / \sum_{k=1}^{t-s} f(k)$ . Similarly, patients for whom the onset dates were available are expected to be affected by the reporting delay. Therefore, we have  $M_s|t / \sum_{k=1}^{t-s} f(k)$ , where  $M_s|t$  represents the reported number of patients for whom onset dates were within period  $s$ , extending until the current period  $t$ .

We sought  $R_0$  and  $q$  to fit the data to minimize the sum of the absolute values of discrepancies among the bootstrapped epidemic curve and the fitted values. The estimated distributions of the three reproduction numbers were calculated using 10,000 fully replicated iterations of bootstrapping for the empirical epidemic curve of symptomatic patients.

Because the number of patients in Wuhan was not reported daily until January 20, we presumed only one initial case and ran the model. After the model prediction reached the number of patients on January 20, we compared the data and the model prediction.

Moreover, the epidemic curve in Japan has three periods marked by activation of school closure and voluntary events cancellation (VECSC) during February 27 – March 19. Before VECSC, the curve was increasing gradually, but it became almost entirely flat during the VECSC period. Subsequently, it began increasing exponentially. Then it

reached a peak and began a monotonic decreasing trend. Therefore, we inferred different reproduction numbers for the three periods. We assign  $R_0$  to the period before VECSC,  $R_v$  to the VECSC period, and  $R_a$  to the period after VECSC.

## Results

During January 20 through March 5 in Wuhan, 68,289 patients were reported.

During January 14 – April 26 in Japan, 12,936 community-acquired cases were identified, excluding asymptomatic cases.

Figure 1 depicts an empirical distribution of the duration of onset to report in Japan. The maximum delay was 30 days. Figures 2 and 3 show an epidemic curve by onset date in Wuhan and Japan adjusted to represent the empirical distribution of incubation period depicted in Figure 1. Figure 3 implies that the first peak of the outbreak began in Japan on April 3, comprising people who had been infected on March 29.

Figure 4 depicts the empirical distribution of incubation periods among 91 cases whose exposed date and onset date were published by MHLW in Japan. Its mode was six days. The average was 6.6 days.

The estimated  $R_0$ ,  $R_v$ , and  $R_a$  were 3.19 with range of [3.08, 3.31], 147.7 [145.3, 157.0], and 2.048 [2.048, 2.024] in Wuhan. Also,  $q$  was estimated as 99.32 [99.31,

99.36]%. In Japan,  $R_0$ ,  $R_v$ , and  $R_a$  were estimated respectively as 2.16 [1.97, 2.20], 1.13 [1.00, 1.44], and 2.96 [2.81, 3.08]. Moreover,  $q$  in Japan was estimated as 99.9888% [99.9885, 99.9890]%. Figures 2 and 3 also include the fitted line based on estimated parameters in Wuhan and Japan.

## Discussion

We observed the peak in Wuhan as February 7. The epidemic curve was apparently unaffected by blocking of traffic, which started on January 23. Moreover, the peak was two weeks later than the introduction of lockdown. Therefore, most patients whose onset contributed to the peak were infected after blocking of traffic. The peak is probably not the result of the blocking of traffic.

We observed the first peak of the outbreak in Japan as occurring with the April 3 onset date and the inferred March 29 infection date. Because Figure 1 depicts almost all cases reported up to 30 days, the first peak might not change over time. Therefore, we conclude that the first peak of outbreak in Japan has already passed.

It is noteworthy that no countermeasure was implemented when the peak was reached, which occurred after VECSC period. Moreover, a state of emergency was declared on April 7. During March 21 through April 6, the only measure was a



recommendation that events with large audiences or numerous participants be cancelled.

Therefore, the peak infection date of the end of March can not be regarded as a result of strong countermeasures in Japan. Nevertheless, VECSC has apparently been effective because the epidemic curve is below the fitted line, as shown by the model in Figure 4.

Moreover, the end of March might be an almost identical climate to that of the beginning of March. Results therefore suggest that temperature and humidity might not affect the virus infectiousness.

We applied a simple SIR model including the proportion of asymptomatic cases that had not been incorporated into the model to date. Therefore, this is the first model which can show a peak and declining phase without a change in the reproduction number for COVID-19 outbreak. Figures 2 and 3 prove that the model fits well in Wuhan and Japan.

Earlier studies [8–10] have estimated  $R_0$  for COVID-19 as 2.24–3.58 in Wuhan. Our obtained  $R_0$  was similar but slightly smaller. By contrast, an earlier study [11] estimated  $R_0$  in Japan as 0.6. That figure might foster misguided policies for countermeasures in Japan. If results of the present study are correct, then they would necessitate adherence to contact tracing to detect clusters when more than 60 million people would become infected.

We estimated the proportion of asymptomatic cases among infected people as 99.32 and 99.98%, respectively, in Wuhan and Japan, which means that the respective ratios of asymptomatic cases to symptomatic cases were 147.8:1 and 8941:1. In other words, for every symptomatic patient confirmed, approximately 150 or 9000 asymptomatic cases are presumed to exist.

A report of an earlier study [7] described the proportion of asymptomatic cases among infected people as 3/23. That proportion represents a huge difference from our results presented herein. A possible reason might be laboratory test procedures. Earlier studies used PCR tests, which detected infection at the time of specimen collection. Therefore, a negative PCR test result does not contra-indicate a past infection. To confirm our results through laboratory testing, complete laboratory-based surveillance in the community using an IgG antibody test, not PCR, is expected to be necessary.

Such a trial was launched in New York City. It revealed from antibody testing that about 15% of residents were positive [12]. At that time, the prevalence was reported as 0.88% in New York state. Therefore,  $q$  was 94% in New York state. That result might be similar to our results obtained for  $q$ .

Recently, Keio University Hospital reported that about 6% of newly administrated and non-COVID-19 patients were infected asymptotically during April 13–19, 2020

[13]. During that period, Figure 3 implies that 2095 patients were reported, representing 11% ( $= 2095 \times 6985/120$  million) of the total population. The figure was slightly higher than 6%. However, they were not healthy people. Therefore, their protection against infection might be more of a concern; their usual activity might be less than that of healthy people. Therefore, incidence among those patients might be lower.

Why were the proportions of asymptomatic cases much different in Wuhan and Japan? The primary reason might be health conditions affecting some or all of the population such as smoking, diabetes, and air pollution. Residents in Japan might be healthier than people in Wuhan. Therefore, a person in Wuhan might be more likely to show symptoms than in Japan, even if their respective conditions of exposure were equivalent. The secondary reason might be the test strategy. If PCR tests in Japan were limited to use only with more severe patients than those in Wuhan [14], then more mild cases were probably classified to asymptomatic cases in the present study in Japan. Thirdly, the virus which has circulated in Japan since the middle of March might have mutated to show different characteristics from those of the virus which circulated in Wuhan.

The present study has some limitations. First, these results, which were obtained with a very high proportion of asymptomatic cases, are merely hypothetical. This

hypothesis should be verified through additional study.

Secondly, the peak in Japan on April 2 might be only the first peak: a second and third wave might occur. Moreover, a second or third wave peak might be higher than the first peak. In fact, the peak of the entire outbreak might eventually be such a second or third peak. One must particularly consider that April is the first month of the school year and the fiscal year in Japan. For that reason, the population in Japan reshuffles many of its activities at this time. Many new students and new employees move to Tokyo from outside Tokyo. In addition, Tokyo residents move away from the city. Therefore, the outbreak can be expected to increase again in middle or late April. Evaluation of the outbreak of COVID-19 in Japan in its entirety must be postponed until the outbreak ends.

Thirdly, as described above, VECSC is apparently effective. Therefore, its effects must be incorporated as effects influencing the model. Assessment of those effects constitutes our next challenge for future research efforts.

Fourthly, although we obtained a very high proportion of asymptomatic cases, they might include some effects of under-ascertainment [15]. Under-ascertainment was estimated as 9% in Wuhan. In other words, about 10 times in the estimated about 150 times of the reported cases were probably mild cases. In that sense, 140 times of the

reported cases were asymptomatic cases in Wuhan. What about Japan?

Under-ascertainment might be related to the strategy of PCR testing. Investigation of that possibility is anticipated as a future challenge for study.

Fifthly, although the distribution of incubation periods is probably almost identical in Wuhan and Japan, the distribution of the delay from onset to report might be much different. It might be affected by testing procedures and capacity, reporting systems, and human resources to apply them. We do not know the distribution in Wuhan. However, we must evaluate the inference that the distributions were the same.

## Conclusion

Results indicate that that the first peak of COVID-19 outbreak was April 2. The central government of Japan declared an emergency on April 7. However, as shown in Figure 1, the number of symptomatic patients had already declined. Probably, the declaration served to mitigate the outbreak.

By introducing a very high proportion of asymptomatic cases, two inconsistent phenomena might be resolved as a result of this study: the high reproduction number and low peak. Nevertheless, it is currently only a hypothesis. Its validity must be

verified using data of outbreaks of prefectures in Japan, or from other countries including the United States.

The present study is based on the authors' opinions, but does not reflect any stance or policy of their professionally affiliated bodies.

#### Acknowledgments

We acknowledge the great efforts of all staff at public health centers, medical institutions, and other facilities who are fighting the spread and destruction associated with COVID-19.

#### Reference

1. 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..  
Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020, 395, 497-506.
2. Health commission of Hubei province. Situation of novel corona virus pneumonia in Hubei province. <http://wjw.hubei.gov.cn/bmdt/ztzl/fkxxgzbdgrfyyq/> (in Chinese)  
[access on 30 April]

3. Japan Ministry of Health, Labour and Welfare. Press Releases.  
[https://www.mhlw.go.jp/stf/newpage\\_10723.html](https://www.mhlw.go.jp/stf/newpage_10723.html) (in Japanese) [accessed on 26 April, 2020]
4. Kurita J, Sugawara T Ohkusa Y. Preliminary evaluation of voluntary event cancellation as a countermeasure against the COVID-19 outbreak in Japan as of 11 March, 2020. <https://www.medrxiv.org/content/10.1101/2020.03.12.20035220v1>
5. Kurita J, Sugawara T Ohkusa Y. Forecast of the COVID-19 outbreak, collapse of medical facilities, and lockdown effects in Tokyo, Japan.  
<https://medrxiv.org/cgi/content/short/2020.04.02.20051490v1>
6. Ohkusa Y, Sugawara T, Taniguchi K, Okabe N. Real-time estimation and prediction for pandemic A/H1N1(2009) in Japan. *J Infect Chemother.* 2011;17:468-72.
7. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, Bardossy AC, Oakley LP, Tanwar S, Chisty Z, Bell JM, Methner M, Harney J, Jacobs JR, Carlson CM, McLaughlin HP, Stone N, Clark S, Brostrom-Smith C, Page LC, Kay M, Lewis J, Russell D, Hiatt B, Gant J, Duchin JS, Clark TA, Honein MA, Reddy SC, Jernigan JA; Public Health ? Seattle & King County; CDC COVID-19 Investigation Team. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in

- Residents of a Long-Term Care Skilled Nursing Facility - King County, Washington, March 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69:377-81. doi: 10.15585/mmwr.mm6913e1.
8. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Lou Y, Gao D, Yang L, He D, Wang M. 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* 2020[Online ahead of print]
  9. Liu Y, Gayle AA, Wilder-Smith A, Rockly J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med.* 2020.DOI: 10.1093/jtm/taaa021
  10. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Coronavirus disease-2019 (COVID-19): The Epidemic and the Challenges. *Int J Antimicrob Agents.* DOI:10.1016/j.ijantimicag.2020.105924
  11. H Nishiura, H Oshitani, T Kobayashi, T Saito, T Sunagawa, Takaji Wakita, MHLW COVID-19 Response Team, Mi Suzuki. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). doi: <https://doi.org/10.1101/2020.02.28.20029272>



12. Reuter. New York survey suggests nearly 14% in state may have coronavirus antibodies.

<https://www.reuters.com/article/us-health-coronavirus-usa-new-york/new-york-survey-suggests-nearly-14-in-state-may-have-coronavirus-antibodies-idUSKCN2252W>

N [accessed on April 26, 2020]

13. The Asahi Shimbun. Number of new patients testing positive for virus alarms

hospital. <http://www.asahi.com/ajw/articles/13321354> [accessed on April 26, 2020]

14. OECD. Testing for COVID-19: A way to lift confinement restrictions.

<https://www.oecd.org/coronavirus/policy-responses/testing-for-covid-19-a-way-to-lift-confinement-restrictions/> [accessed May 1, 2020]

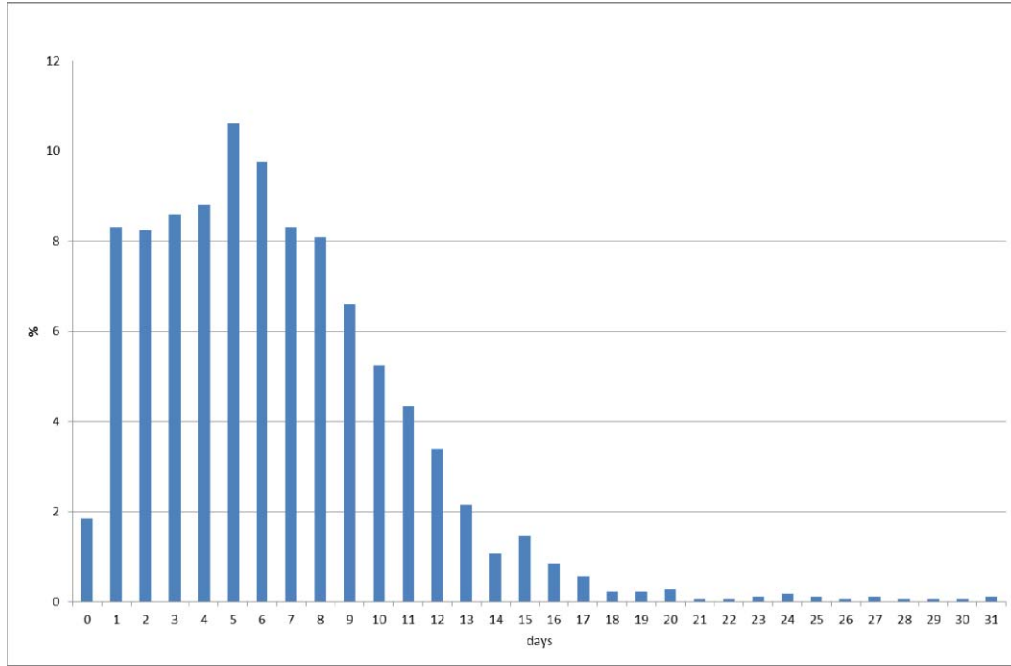
15. Omori R, Mizumoto K, Ascertainment rate of novel coronavirus disease

(COVID-19) in Japan. medRxiv 2020.03.09.20033183; doi:

<https://doi.org/10.1101/2020.03.09.20033183>

Figure 1: Empirical distribution of duration from onset to report by MLHW, Japan.

(%)



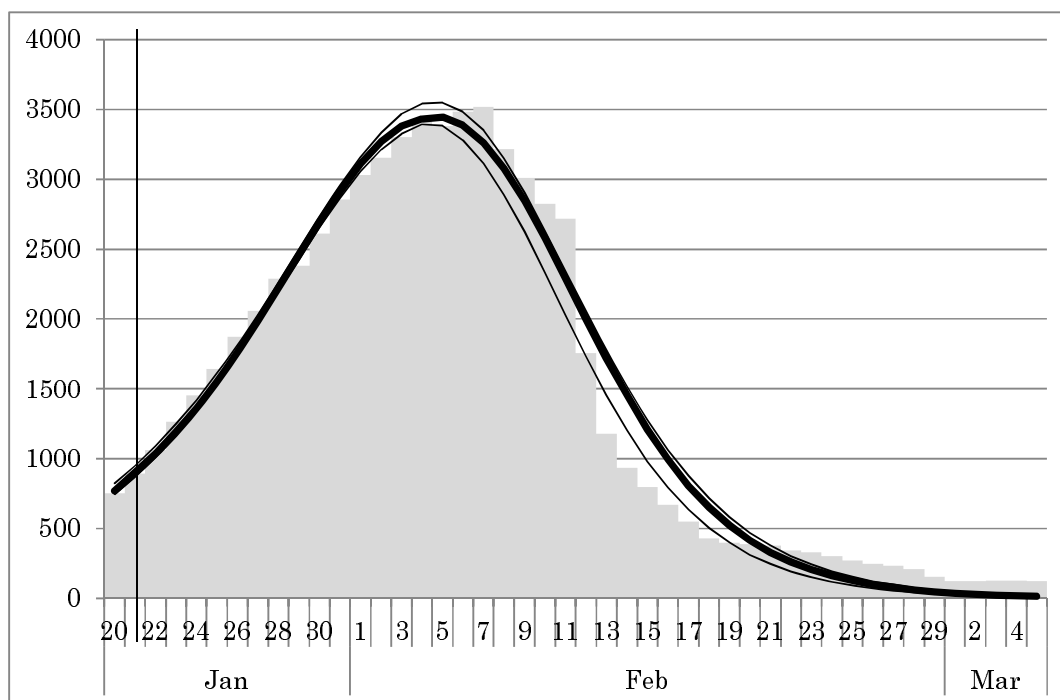
Note: Bars represent the probability of duration from onset to report based on 657

patients for whom the onset date was available in Japan. Data were obtained from

MLHW, Japan.

Figure 2: Epidemic curve based on the published by the Health Commission of Hubei province, China and the fitted lines.

(number of patients)

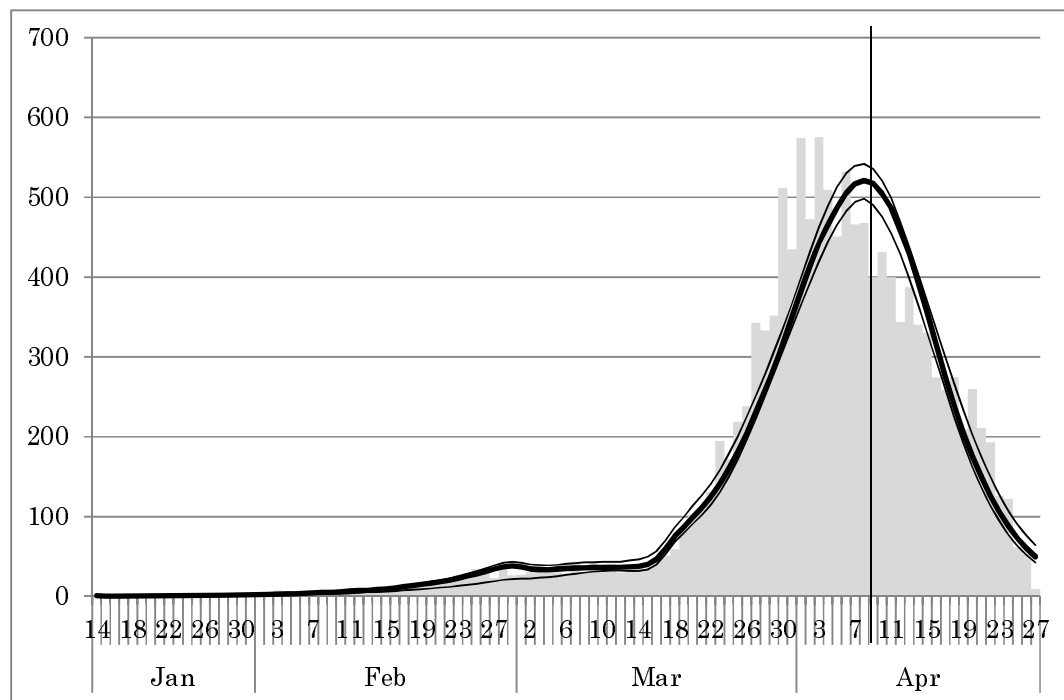


Notes: Bars represent the number of patients by onset date. The onset dates of patients were inferred from the empirical distribution of duration between onset to report depicted in Figure 1. The curve represents the fitted line by the model. The thin lines show its range. The vertical line represents January 23, when Wuhan was locked down.

Figure 3: Epidemic curve published by the Ministry of Labour, Health and Welfare

(MLHW), Japan and fitted lines

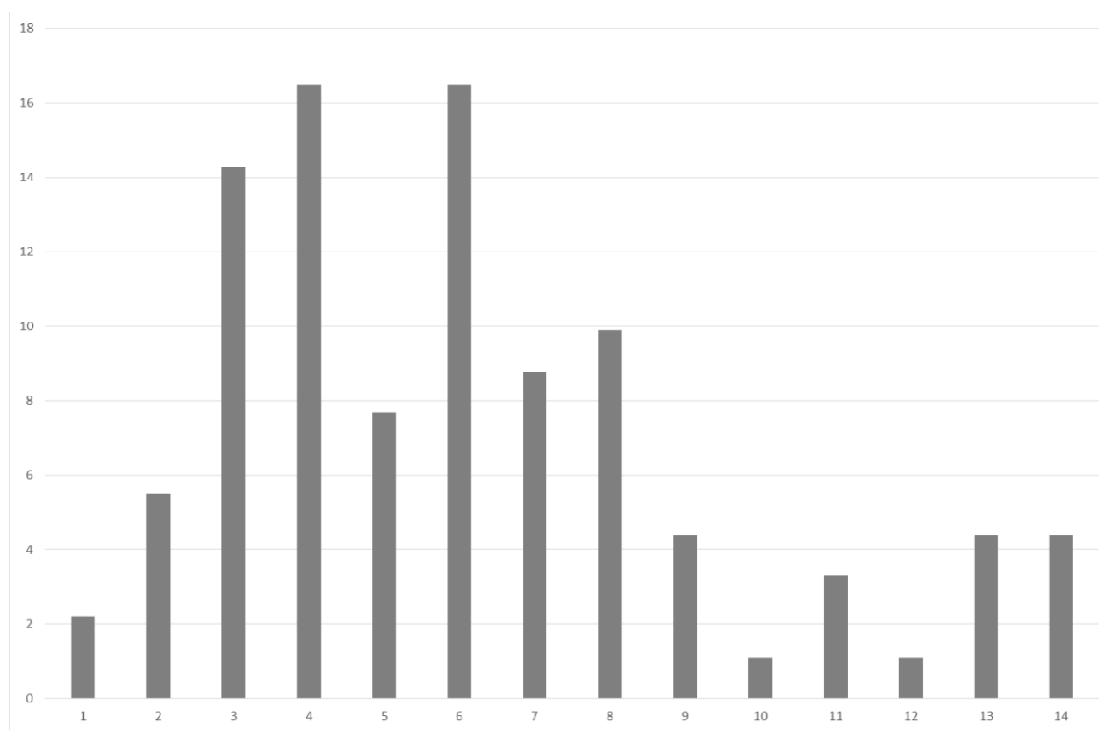
(Number of patients)



Notes: Bars represent the number of patients by onset date. The onset date of patients for whom the onset date was not reported were inferred from the empirical distribution of the duration between onset to report depicted in Figure 1. The curve represents the fitted line by the model. The thin line represents its range. The vertical line represents April 8, when the state of emergency was activated.

Figure 4: Empirical distribution of the incubation period published by MLHW, Japan.

(%)



(days)

Notes: Bars show the distribution of incubation periods for 91 cases for which the exposure date and onset date were published by MLHW, Japan. The patients for whom incubation was longer than 14 days are included in the bar shown for day 14.