**Is Japan succeeding in suppressing COVID-19?
Estimation of infection rate and the size of the population potentially exposed
to SARS-CoV-2**

Motohiko Naito

Affiliations: Ronin (Not affiliated with any institute)

Correspondence:

Motohiko Naito

Email: litmanen0316@yahoo.co.jp

Tel: +81 090-7766-5250

Conflicts of Interest:

The author declares no competing financial interests.

Abstract

Background: The novel emerging virus SARS-CoV-2 has affected all human-kind during the first half of 2020. The aim of the study was to survey the actual circumstances from January until May. Methods: The data are collected and released systematically, by law, from the National Epidemiological Surveillance of Infectious Disease (NESID). Findings: Analysis of these data revealed that the infection spread in Japan from late March to early April 2020. The SARS-CoV-2 infection rate at its peak was estimated to be 10%. Thus, the size of the population who may have been exposed to the novel virus in Japan is estimated at 0.2 million, which is relatively small. The number of related deaths is likely to converge on 1,000 people. Interpretation: Applying the law of large numbers allows estimation of the infection rate as well as of the size of the affected population by statistical analysis. How to collect the data must be defined before the data analysis is suggested to be important to reflect the actual circumstances about COVID-19.

Introduction

The novel emerging respiratory disease, COVID-19,¹ has inspired fear in human-kind during the first half of 2020. COVID-19 is caused by the novel emerging coronavirus SARS-CoV-2, named by the International Committee on Taxonomy.² The first outbreak was confirmed in China, in December 2019.³ This outbreak led to 4,642 deaths among 84,149 confirmed cases overall in China, as of April 9, 2020.⁴ SARS-CoV-2 has spread exponentially in China and from China throughout the world. The World Health Organization officially described COVID-19 as a pandemic on March 11, 2020.⁵ SARS-CoV-2-positive cases were reported in 114 countries globally, on March 11, 2020.⁵ Based on the data released by the Johns Hopkins University, approximately 0.19 million deaths occurred among 2.7 million confirmed cases as of April 27 and around 0.38 million deaths among 6.5 million confirmed cases as of May 27, 2020.⁶

In Japan, the first SARS-CoV-2-positive confirmed cases were reported in January 16, 2020,⁷ and the number of cases increased almost daily. On April 7, 2020, the Japanese government declared state of emergency in seven prefectures: Tokyo, Osaka, Saitama, Chiba, Kanagawa, Hyogo, and Fukuoka; this declaration was expanded to other 40 prefectures on April 16, 2020.⁸ The declaration was lifted in Japan overall until May 25, 2020.⁹ The number of SARS-CoV-2-positive cases is small relative to the population of Japan (137 million),¹⁰ which is considered unusual. It has been suggested at April, that this may be due to the few PCR tests conducted.^{11,12} In May, it appeared that the Japanese government had succeeded in suppressing the novel virus infection.^{11,13}

We aimed to investigate the actual circumstances about SARS-CoV-2 infection in Japan since the first positive case reported. The data used in the study were officially announced by the Ministry of Health, Labour and Welfare (MHLW) in Japan, daily since January 22, 2020. MHLW data were thoroughly collected and released systematically to facilitate the survey of the actual circumstances. Data analysis and use of the mathematical theorem lead to the infection rate estimation in Japan as well as to the population size, potentially exposed in SARS-CoV-2, estimation.

Materials and Methods

Data used in this study

The data analyzed here are cited from the official announcements of the MHLW in Japan.¹⁴ We did not distinguish the data between the MHLW and Japanese government; we simply referred to “government data” henceforth.

The official government data are reported according to the Japanese domestic law, known as “infectious disease control law” enforce since 1998.¹⁵ In brief, a patient visits the doctor in a routine examination. If the patient is suspected of an infectious disease such as COVID-19, then the doctor must report this case immediately to the Public Health Center, through which this information is shared with the MHLW and related institutions. The data used in our study were collected systematically using the National Epidemiological Surveillance of Infectious Disease (NESID) system defined by this law. For example, data on hospitalized COVID-19 patients who underwent PCR test to confirm recovery and being discharged were not included.^{14,15} The cases who underwent PCR tests (simply referred to as “tests” in this study) and SARS-CoV-2-positive patients (simply referred to as “patients” henceforth) were reported from January 22, 2020, and the deaths from February 13, 2020, the day of the first death due to COVID-19. To our knowledge, the data on tests and patients for each prefecture were officially released by the Japanese government since March 14, 2020. Therefore, we recorded the overall Japanese data from January 22 until May 31, 2020, and from each prefecture from March 22 until May 31, 2020.

As seven prefectures were first declared in an emergency state by the Japanese government,⁸ the data of these prefectures were analyzed. If necessary, additional data officially released by each prefecture were added to the analysis. According to the note of the official MHLW announcements, some samples were duplicated in Chiba and Osaka until March 21, as well as in Kanagawa until March 23. In addition, public health center in Kanagawa rechecked the samples taken before April 27, 2020, from April 28, 2020 until May 16, 2020. Thus, the Kanagawa data from April 28, 2020, until May 16, 2020, are not the data from January 2020, but rather specifically reflect these 20 days. The data of March 23 and May 4, 2020, do not exist officially, and thus data from these two dates are missing.

Primary methods

The positive rate was defined as (the number of patients) / (the number of tests) and the death rate was defined as (the number of deaths) / (the number of patients). A

shift toward the positive rate was observed in the former and a shift toward the death rate in the latter.

As all data were presented in Japanese, we have established a data archive in English, and these data are included in the supplemental file.

R software (<https://www.r-project.org/>) was used for data analysis and for drawing graphs. The sample codes for drawing graph are available on the supplemental file.

Results

Primary data analysis of overall Japan

The number of tests conducted and the number of patients has increased exponentially from January 15, 2020, the date on which the first patients were confirmed test-positive (Fig. 1a). Figure 1b shows the positive rate of SARS-CoV-2 in Japan. First, the rate was drastically dropped at March 4. This happened because the number of test-positive patients markedly increased from 1,855 cases on March 3 to 5,690 cases on March 4, 2020, reflecting an increment of 3,835 cases in a single day. Although the reason for this drastic increase is unclear, this increment is not worrisome, as such gross increments or decrements do sometimes appear in the data. Next, overall, there are clearly many spikes, reflecting rapid fluctuations. These spikes coincided with Saturdays and holidays (Fig. S1a). Upon careful analysis of raw data, it was clear that tests were always less on weekends and holidays than on weekdays, while the number of patients increased on weekend or holidays. We concluded that the spikes could be ignored because occur naturally in the data. We introduced a mathematical theorem that was an especially useful tool for this study.

Theorem: The (strong) law of large numbers

Given that the data set is followed by the probability distributions with probability p , the parameter is not usually known. Picking any subset from the set with probability p : the set with p is called a population, and the subset is called a sample. The sample is chosen to allow calculation of the probability q . If the size of the sample is sufficiently large, then probability q converges on probability p in the population.

This theorem is proven in standard textbooks.¹⁶ Applying the law of large numbers is necessary to ensure that the sample size is sufficiently large to apply the theorem.

This analysis showed that the positive rate stabilized later in March after the marked increment recorded on March 4, 2020. In fact, the number of tests exceeded 5,000 on March 4 and the number of tests increased over time thereafter. With the increase in the number of patients, the positive rate remained stable at approximately 6% after the first week of March. Thus, we concluded that the number of tests was sufficiently large at March 4, 2020, to apply the law of large numbers and the positive rate at March converged to about 6%. Moreover, the positive rate varied widely before March 3 because the sample size was too small to apply the law of large numbers. Similarly, from April 11 until the end of April 2020, the positive rate converged to about 10%. The time between these two converging periods was worth investigating closely. This data analysis demonstrated that the positive rate increased from 6% to 10% during the first 10 days of April. It is logical to consider that SARS-CoV-2 spread markedly during these 10 days, a period named the “infection spreading period.” If we focus on the data of May, we note that the positive rate gradually decreases. The positive rate on May 1 is 9.3%, and decreases to 6.8% by May 31, 2020, demonstrating a gradual decrease in the positive rate. In short, the number of tests conducted was sufficiently large in March to estimate the infection rate, which reached ca. 10% at the peak of the infection.

Next, we analyzed the data on COVID-19-related deaths (Fig. 2). The first report of death was on February 13, 2020. Thereafter, the number of deaths increased exponentially, along with the increase in the number of patients (Fig. 2a, b). First, there was a rapid increase in the number of deaths on May 22, 2020 (Fig. 2b), which appeared unusual. Inspecting the government data more closely, we found that the number of deaths were limited to those that the government had officially confirmed. The method of official announcements changed from April 22, 2020; it simply reflected a change in format. Hence, this increment on that date can be ignored. Considering this, the death data can be analyzed. Figure 2c shows a graph reflecting the changes in death rate. The death rate increased from the first report, except for the period from middle of March until middle of April, 2020, and then reached approximately 5% by May 14. The decrease in death rate in this period appeared strange, and we named it the “strange period”. To investigate the reason for the decrease during the strange period, we generated a graph of the logarithm of the number of deaths (Fig. 2d). The graph is almost linear except for the rapid increase

on April 22, 2020; the slope of the logarithm of death changed significantly on that date. The proof is described in supplemental results. However, the reason for the decrease during the strange period was unclear from the analysis of the death data. The decrease in death rate during the strange period coincided with the “infection spreading period” (Fig. 2c). This suggested that the novel virus spread faster than the increase in deaths during the first 10 days of April, strongly indicating that the “infection spreading period” should be re-defined for the period from late March until April 11, 2020.

Estimating the overall number of people who will infect and die of COVID-19

As previously discussed, the peak positive rate was calculated to approximately 10%. According to Fig. 2a, the rate of patient’s infection reduced gradually in May, plateauing in the last third of May 2020. Thus, the spread of SARS-CoV-2 seemed to slow down in May. The number of patients were considered below 20,000 during the spring of 2020. Therefore, we speculated that the number of patients under surveillance in Japan were at the most 20,000. Applying the law of large numbers, the infection rate was estimated, especially around 10% at the “infection spreading period”, and the population of individuals living in Japan who are likely to have been exposed to the virus can be calculated to be 200,000 (0.2 million) at most.

The death rate reached almost 5% in the final third of May 2020 (Fig. 2c). According to the data from China, the death rate in that country was 5.5% by April 9, 2020. In Germany, one of the first countries in the EU to ease the lockdown in the cities,¹⁷ recorded 8,727 deaths among 186,233 confirmed cases, with a death rate of 4.7% by June 9, 2020. Globally, the death rate was 7.0% by April 27 and 5.9% by May 27, 2020. Consequently, the death rate in Japan is near the global rate and that of other countries. Thus, the number of deaths in Japan is likely to amount to approximately 1,000.

Comparing the change in the positive rate in seven prefectures of Japan and overall

The positive rate discussed until now was the overall positive rate for Japan. Now, we will focus on the changes in the positive rate in the seven prefectures in which the state of emergency was initially declared.⁸ Data for each prefecture were available

from March 14, 2020. A detailed comparison in the positive rate in these seven prefectures and of Japan overall during the 2 months are discussed in detail in the supplemental results. In brief, qualified graphs of the positive rate in these seven prefectures as well as in Japan overall are shown in Fig. 3a. The “infection spreading period” could be observed in the data of four of the seven prefectures: Tokyo, Saitama, Hyogo, and Fukuoka, as well as in the data of Japan overall. The positive rate in Saitama, Hyogo, and Fukuoka showed similar tendencies as in the overall data: the positive rates increased during the “infection spreading period”, then stabilized during the rest of April, and gradually decreased during May 2020 (Fig S2). In Kanagawa and Chiba, the positive rate was stable after the “infection spreading period”, and in May 2020 either remained stable or decreased. Because of the quality check of Osaka data, it is concluded that Osaka data for May, 2020 should be treated as missing data. The Tokyo data are described in detail below.

Analysis of the Tokyo data

The Tokyo positive rate increased exponentially during April 2020, converging on about 40% (Fig. 3a). Such an increase in the positive rate in Tokyo was compared with that in the other six prefectures and in Japan overall. Fig. 3a shows that the positive rate in Tokyo showed a markedly greater spread than that in Japan overall, even after the “infection spreading period.”

Next, we focus on the May 2020 data. As mentioned in the supplemental results, in Japan, there are several consecutive national holidays, on April 29 and from May 1 to May 6. Thus, the data for this period were not guaranteed that have been collected according to the NESID system. Hence, the quality checks for May data should be more stringent.

The Tokyo positive rate remained markedly higher than that in other prefectures during that month (Fig. 3a, b). Considering that the positive rate for April was significantly greater in Tokyo than in the other prefectures, it may be inappropriate to include the April data when comparing the Tokyo data for May with that of other prefectures. Thus, we revised the May data for Tokyo by recording the government data, with April 30 set as day 0. The positive rate using these data is shown as a dashed line in Fig. 3b; the graph shows that even when the revised May data are

used for Tokyo, the positive rate for this prefecture remains markedly higher than that of other prefectures.

During the data analysis, we noted that the number of tests performed in Tokyo in May 2020 was markedly lower than that performed in April. To verify the analysis, we examined the raw data for Tokyo more carefully. The number of tests per day for most of May was below 100 (Fig. 3c). Considering unlike the limitation of the tests in Tokyo prefecture for any reason, we collected the released data by the Tokyo prefecture;¹⁸ these data are referred to as Tokyoto data (because the Japanese use the term Tokyo-to, not Tokyo prefecture).

We compared the number of tests between the Tokyoto and government data (Fig. 3c). The number of tests of the Tokyoto data was significantly larger than that of the government data. According to the Tokyoto data, over 1,000 tests were performed on most of the weekdays (Fig. 3c). Therefore, we concluded that in Tokyo the number of tests performed was not limited. The Tokyoto data included all the PCR tests conducted in Tokyo, but these tests were never recorded in the NESID system, and therefore the Tokyoto and Japanese government data differed.

Although both definitions were suitable for statistical analysis, the difference in data would have led to opposite conclusions. Figure 3d shows the graph using the two different test datasets as the denominator for the rate. The tests from the government data were termed Tests-in-law, while the tests from the Tokyoto data were named All-Tests. Using the Tests-in-law, the positive rate remained stable at ca. 25% or decreased subtly after variation. Conversely, in the All-Tests data, the positive rate decreased exponentially early in May and converged to approximately 14% for most of May.

When the Tests-in-law data were used, the positive rate remained markedly higher in Tokyo than in Japan overall. In contrast, when the All-Tests data were used, the positive rate for Tokyo converged to a value less than the April, 2020, leading to conclude Tokyo succeeded in suppressing the spread of the novel virus.

Discussion

It is important to survey the real events about SARS-CoV-2 for both doctors and governors. Surveillance is the basis of the policies about COVID-19. In the cases of doctors, doctors will accept all patients as usual or inform them about the specialized

COVID-19 hospitals in advance to prevent spread of infection in hospital/clinic. The governor must decide the policies about COVID-19 to suppress or minimize the effect of SARS-CoV-2 infection at the current time. Therefore, governors of countries with high infection rate decided to limit their citizens' lifestyle. Theoretically, data analysis aims to state actual circumstances or reveal new perspectives. However, it is really difficult to achieve this scope. To overcome this difficulty, Japanese government enforces the Japan domestic law as well as establishes the NECID system.¹⁵ Throughout this study, we always validated that the data reflected the actual circumstances. Some data do not need to be subjected in changes: only changing the format (Fig. 2b), re-check the sample data (Fig. 3b), etc.

In this study, we sought to investigate the actual COVID-19 situation in Japan. During the preliminary collection of data and primary analysis, we noted that Japan already has an advanced surveillance system, which is systematically implemented in line with Japanese domestic laws. The NESID system facilitates data analysis due to its unique order. Using data from the NESID system, and because sufficient numbers of tests were performed to allow application of the law of large numbers, we estimate that the infection rate in Japan was about 10% at its peak. The number of positive confirmed patients increased exponentially in April and arrived to plateau in May (Fig. 2a). We speculate the number of positive confirmed cases to be 20,000 cases at most, during the spring of 2020. Applying the law of large numbers, we concluded that the size of the population exposed to SARS-CoV-2 was estimated as 0.2 million. Considering that Japan's population is 137 million,¹⁰ this exposed population size is relatively small validating that Japanese government minimized the infection spread. An expert meeting on February 24, 2020 declared that the primary goal related to COVID-19 is to slow down the spread of infection and minimize the number of severe cases and deaths.¹⁹ This goal seemed to have been achieved from Jan 22 until May 31. After the end of COVID-19 pandemic, many studies will focus on to explain why Japan has minimized SARS-CoV-2 infection from many points of view.

The data analysis also highlighted the importance of the definition of counting the number of PCR tests conducted. The data released for Tokyo by the Japanese government and by Tokyo prefecture differed markedly, and the difference could lead to opposite conclusions. The governor of Tokyo, Koike, has emphasized in a recent press conference that we should "prepare the policies for the secondary wave of

novel virus infection". However, the data suggest that Tokyo maintains a higher positive rate of SARS-CoV-2, and is thus currently still in the "first wave" of infection. For this reason, it is regrettable that Osaka data for May were not recorded in the NESID system. We urge the governor of Osaka, Yoshimura, to align the data with the NESID system as soon as possible.

Concluding, NESID system facilitates data analysis to survey the real event about SARS-CoV-2 infection. Analysis of these data revealed the "infection spreading period" and estimated that the population exposed to SARS-CoV-2 is approximately 0.2 million and the related deaths 1,000 people. In addition, the variation to define counting the number of PCR tests conducted would have led to opposite conclusion. How to collect the cases conducted PCR tests must be defined before the data analysis is suggested to be important to reflect the actual circumstances about COVID-19.

Acknowledgements

The author thanks all those who supported this study. The author would also like to thank Editage (www.editage.jp) for English language editing. This report is dedicated to all victims of COVID-19 worldwide.

References

- 1 WHO director-general's remarks at the media briefing on 2019-nCoV on 11 February 2020.
<https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020> [accessed June 10, 2020].
- 2 Gorbalenya AE, Baker SC, Baric RS, et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020; **5**: 536–44.
- 3 Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395**: 497–506.
- 4 WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int> [accessed June 10, 2020].
- 5 WHO director-general's remarks at the media briefing on 2019-nCoV on 11 March 2020.
<https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> [accessed June 10, 2020].
- 6 COVID-19 Map - Johns Hopkins Coronavirus Resource Center.
<https://coronavirus.jhu.edu/map.html> [accessed April 27 and May 27, 2020].
- 7 Ministry of Health, Labor, and Welfare. Press release (written in Japanese)
https://www.mhlw.go.jp/stf/newpage_08906.html [accessed June 10, 2020].
- 8 Coronavirus: Japan declares nationwide state of emergency. *BBC News*.
<https://www.bbc.com/news/world-asia-52313807> [accessed June 24, 2020]
- 9 Prime Minister of Japan and His Cabinet [COVID-19] The declaration of the lifting of the state of emergency in response to the novel coronavirus disease. May 25, 2020. http://japan.kantei.go.jp/ongoingtopics/_00027.html [accessed June 17 2020]
- 10 Searching “How many people living in Japan?” on wolfram alpha.
<https://www.wolframalpha.com> [accessed June 17, 2020]
- 11 Sposato W. Japan's halfhearted coronavirus measures are working anyway. *Foreign Policy* 2020. Online May 14, 2020.

- <https://foreignpolicy.com/2020/05/14/japan-coronavirus-pandemic-lockdown-testing/> [accessed June 24, 2020]
- 12 Health Alert-U.S. Embassy Tokyo (April 3, 2020).
<https://jp.usembassy.gov/health-alert-us-embassy-tokyo-april3-2020/>
[accessed June 16, 2020]
 - 13 Iwasaki A, Grubaugh ND. Why does Japan have so few cases of COVID-19? *EMBO Mol Med* 2020; **12**: e12481.
 - 14 Ministry of Health, Labor, and Welfare. Official release archive about novel coronavirus infection (written in Japanese).
https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00086.html
[accessed almost everyday since April 16 until Jun 2, 2020].
 - 15 National Institute of Infectious Disease. The guidelines of the NESID system (written in Japanese).
<https://www.niid.go.jp/niid/images/epi/PDF/gijisyo-gildeline-200110.pdf>
 - 16 Kolmogorov AN. Foundations of the Theory of Probability. 1st edition, 1956. New York; AMS publishing;
 - 17 Official release: Easing coronavirus restrictions. Germany hands responsibility for easing virus restriction to states. German Government.
<https://www.deutschland.de/en/news/easing-coronavirus-restrictions>
[accessed June 16, 2020]
 - 18 Official release from Tokyo. Latest report about infection of novel virus.
<https://stopcovid19.metro.tokyo.lg.jp/en> [accessed June 22, 2020]
 - 19 Novel Coronavirus Expert meeting. The opinion for realized the basic policy about the novel coronavirus infectious diseases. Ministry of Health, Labor, and Welfare. Feb 24, 2020.
https://www.mhlw.go.jp/stf/seisakunitsuite/newpage_00006.html [accessed June 16, 2020]

Figure legends

Fig. 1 Graphs of the PCR test conducted (Tests) and the COVID-19-positive cases confirmed by the Tests (Patients)

- (a) The numbers of Tests (green) and Patients (red) since the first day the first SARS-CoV-2-positive case was confirmed. Both numbers have been growing exponentially.
- (b) The positive rate in Japan from January 16, 2020, until May 31, 2020. The positive rate varied widely during January and February, which is typical and is also observed in subsequent graphs. The black circle suggested that SARS-CoV-2 has spread widely in Japan.

Fig. 2 COVID-19-related deaths

- (a) The number COVID-19-related deaths (black), and patients (Red) since the day on which the first COVID-19 death was confirmed
- (b) The number of deaths is shown as a line
- (c) Comparison of the positive rate (red) and the death rate (black)
- (d) Logarithm of the number of deaths. The x-axis of this graph starts from March 14, 2020, because the death rate seemed to decrease around that date. The slope decreases in May 2020.

Fig. 3 Graphs of positive rates in seven prefectures as well as Japan overall

- (a) Qualified graphs of the positive rates in seven prefectures: Tokyo, Osaka, Saitama, Chiba, Kanagawa, Hyogo, and Fukuoka, as well as Japan overall. Osaka data in May are treated as missing data. The data for May 4 do not exist officially for individual prefectures, only overall data are available.
- (b) Comparison between the positive rates in six prefectures (excluding Osaka) in May with that of Japan overall. The Tokyo rates remain markedly higher than those of the other prefectures. Given the higher rates of April in Tokyo, we considered that it would be inappropriate to compare the rates of May. Consequently, we created a special data set in which the number of tests and patients were limited to May, resetting April 30, 2020, as day 0. even the Tokyo data are modified (dashed line), Tokyo graph still remains markedly higher than the graphs of the other five prefectures and Japan overall The Kanagawa graph is

relatively higher than that of the other prefectures, except that of Tokyo. The graphs of the other four prefectures coincide with the overall graph: the positive rate appears to decrease gradually.

- (c) Comparison of the number of PCR tests conducted per day based on data released by Tokyo (green) and by the Japanese government (purple)
- (d) Comparison of the positive rate of the data released by Tokyo (green) and those released by the government (purple). The government data are the same as in Fig. 3b. red line indicates the overall Japan's positive rate.

Fig 1(a)

Fig1(b)

Fig 2a

Fig 2b

Fig 2c

Fig 2d

Fig 3a

Fig 3b

Fig 3c

Fig 3d