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

Supplemental file

Data used in this study

Data_archive.zip contains following data set without password.

Primary_data_set.csv

These data are used in the first chapter of the Results and Fig.1.

Bar_plot column’s number indicate the Saturdays or holidays to create Fig. S1.

deaths_data.csv

these data are focus on death data in the previous data, and used to create Fig. 2

seven_pref_data_withoutQualifying.csv

these data are primary data of third chapter of the Results as well as supplemental results.

X_conducted means counting cases PCR tests conducted along the NECID system in X prefecture.

X_poisitve means the cases confirmed positive the novel virus in X prefecture.

X_positive_rate means the positive rate in X prefecture, calculated by definition.

qualified_data_with_osaka.csv

qualified data of the seven_pref_data_withoutQualifying.csv

qualified_data_exp_osaka.csv

these are qualified_data_with_osaka.csv deleting Osaka data for May

Tokyo_in_May.csv

These data are a special data for Tokyo set in which the number of tests and patients were limited to May, resetting April 30, 2020, as day 0.

Tokyo_conducted, Tokyo_positive, and Tokyo_positive_rate columns are same as the one in qualified_data_exp_osaka.csv limiting in May, 2020.

fix_tokyo_() are revised the May data for Tokyo by recording the government data with April 30 set as day 0.

Tokyoto_() are cited from official release from Tokyo (1).

Sample code for drawing graph

```
#we use the package “tidyverse”, if you don’t have it, install it
> install.package(“tidyverse”)
# using the tidyverse package,
> library(tidyverse)

#Creating the bar plot
>   ggplot(data= Primary_data_set   mapping = aes(x=date) )+
geom_bar(mapping = aes(y=overall_PCR_test_conducted), stat = "identity", position
= "identity", fill="springgreen2" )+
geom_bar(mapping = aes(y=overall_positive_confirmed), stat = "identity", position =
"identity", fill="red" )+
xlab("Date")+ylab("Numbers")

#Creating graphs of the seven prefectures with overall
> ggplot(data= qualified_data_exp_osaka, mapping = aes(x=date) )+
```

showing the national holiday as black background

Note that ylim must have been decided before making bar plot, otherwise it'll be deleted.

geom_bar(mapping = aes(y= bar_plot_you), stat = "identity", position = "identity", fill="black", alpha=0.3)+

Main Body

geom_line(mapping = aes(y=overall_positive_rate), color="red")+

geom_point(mapping = aes(y=overall_positive_rate), color="red")+

geom_line(mapping = aes(y=hyogo_positive_rate), color="dodgerblue1")+

geom_point(mapping = aes(y=hyogo_positive_rate), color="dodgerblue1", shape=2)+

geom_line(mapping = aes(y=tokyo_positive_rate), color="purple1")+

geom_point(mapping = aes(y=tokyo_positive_rate), color="purple1")+

geom_line(mapping = aes(y=osaka_positive_rate), color="blue1")+

geom_point(mapping = aes(y=osaka_positive_rate), color="blue1")+

geom_line(mapping = aes(y=kanagawa_positive_rate), color="darkorange1")+

geom_point(mapping = aes(y=kanagawa_positive_rate), color="darkorange1")+

geom_line(mapping = aes(y=saitama_positive_rate), color="forestgreen")+

geom_point(mapping = aes(y=saitama_positive_rate), color="forestgreen")+

geom_line(mapping = aes(y=fukuoka_positive_rate), color="chocolate1")+

geom_point(mapping = aes(y=fukuoka_positive_rate), color="chocolate1")+

geom_line(mapping = aes(y=chiba_positive_rate), color="darkgoldenrod1")+

geom_point(mapping = aes(y=chiba_positive_rate), color="darkgoldenrod1", shape=2)+

```
xlab("Date") +ylab("Positive Rate")+
#exchanging y-axis as percent scale
#limits indicate the range to show
scale_y_continuous(labels = scales::percent, limits = c(0,0.20))
```

Supplemental Results

Demonstration of the slope of death logarithm significantly changed on April 20, 2020

If the number of deaths increases exponentially, one can describe this event as an exponential function. If *x* denotes the number of deaths, *t* denotes the time (day), *k* denotes the parameter that is characteristic of this function and the function is:

$$x = e^{kt} = \exp\,(kt)$$

or

$$log_e\, x = kt$$

The last equation indicates that the logarithm of the deaths is linearly related to time (the date), and k is the slope of the graph. The slope for the 30 days until April 20, 2020 was calculated and is indicated as K1, and that from April 26, 2020, as K2: K1 = 0.0574, and K2 = 0.0346. To demonstrate that the difference of K1 and K2 is significant, we calculated the number of expected deaths using both parameters and resetting the number of dead people to 35, which is the number at March 21, at day 0 we compared the results of the calculation with both Ks at day 30. Using K1, we calculated 195.8 deaths, while using K2, we calculated 98.8 deaths, which are significantly different. Thus, K1 is significantly greater than K2. The rate of increase in deaths seemed to reduce around April 26, 2020.

Comparison of the changes in the positive rate in seven prefectures and in Japan overall from March 26 until May 31, 2020

Graphs of the changes in positive rates in the seven prefectures and in Japan overall are shown in Fig. S2a. The positive rates in three prefectures, i.e., Chiba, Kanagawa, and Osaka, are beyond or close to 100% around the middle of March. According to the notes in the government data, some of the data counts had been duplicated in these three prefectures up to March 23, 2020. This explain this unusual statistical data. To qualify the data, we deleted data yielding a positive rate exceeding 80% for March. Then, the first day with a rate under 80% is set as day 1 in this study. Based on this qualification, we re-recorded the data from March 26, 2020 for analysis. Consequently, data of March consist of only 6 days, and hence we added these data to April data. Thus, the data of April range from March 26 until April 30, 2020.

Qualified graphs are shown in Fig. 3a and Fig. S1b. The positive rate in Chiba, Kanagawa, and Osaka varied widely in early April 2020. The sample size of the data of these three prefectures at that time was so small that the law of large numbers could not be applied. As the number of tests and patients increased in these three prefectures, we observed that the positive rate started to converge during April. For Tokyo, the graph showed an exponential increase during April, to converge at approximately 40%. Such an increase in Tokyo is characteristic compared with other prefectures. This graph shows that SARS-CoV-2 spread significantly more in Tokyo than in the whole of Japan, even after the “infection spreading period.” The remaining three prefectures, Saitama, Hyogo, and Fukuoka showed positive rates that coincided with the overall rate. The graphs show that the positive rate for these prefectures increased rapidly during the Infection Spreading Period, as it did for Japan overall. Taken together, SARS-CoV-2 spread widely in most of Japan during April, particularly during the Infection Spreading Period.

Quality check of all data for May

In Japan, there are a number of consecutive national holidays, on April 29 and from May 1 to May 6, which are termed the “Golden Week.” The official announcements of the data for the Golden Week stated that the data would be cited in each prefecture’s homepage because the general office was closed. Thus, the data for Golden Week were not guaranteed to be collected according to the National Epidemiological Surveillance of Infectious Disease (NESID) system. Hence, the quality checks for May data should be more stringent.

For the Osaka data, according to the notes described in the margin of data lists released officially by the Japanese government, all data of May for Osaka are cited from Osaka prefecture’s homepage. Examining the data from the Osaka prefecture’s homepage,[2] we found that the PCR tests were counted for all the cases conducted in Osaka. There are no Osaka data available in the NESID system.

As shown in the analysis of Tokyo data, when using the data counting all PCR tests, some of which may be outside of the NESID system, the positive rate can be underestimated. Therefore, we deleted all Osaka data of May; eventually, the data for Japan overall consisted of only 46 prefectures for May. To investigate the effects of this deletion, we compared the overall positive rate with and without the Osaka data (Fig. S2b); the observed difference was smaller than the increment or decrement of spikes. Thus, deletion of the Osaka data would not affect data analysis. Corrected Osaka data and the overall graph are shown in Fig. 3a.

Comparison of positive rates for May

Keeping data quality issues in mind, we now focus on analysis of May data. Through the quality check, the overall positive rate (with omission of the Osaka data for May) was 8.8% on May 1, 2020, and 6.9% on May 31, 2020. Compared to the data that converged to approximately 6% before the Infection Spreading Period and to 10% after the Infection Spreading Period (Fig 3a in main text), it appears that the overall positive rate gradually deceased in May 2020. The graphs of the five prefectures (excluding Tokyo and Osaka), are shown in Fig S2b along with the overall positive rate graph. As mentioned in the Materials and Methods, the Kanagawa data during the middle of May are used only for reference, as the variation itself is not meaningful. Ignoring such variation, the Kanagawa positive rate remained higher than that in the other four prefectures as well as that of Japan overall, although it showed a tendency to decrease. Four other prefecture rates (Saitama, Chiba, Hyogo, and Fukuoka), as well as the overall rate, showed a similar decreasing trend. In particular, the positive rate of Saitama, Chiba, Hyogo, and Fukuoka for these 2 months converged into that of Japan overall.
Tokyo data for May are described in main text.

References

1 Official release from Tokyo. Latest report about infection of novel virus. https://stopcovid19.metro.tokyo.lg.jp/en [accessed June 18, 2020]

2 Official release from Osaka. Special site about the Covid-19 related information (written in Japanese). http://www.pref.osaka.lg.jp/default.html

Supplemental Figure Legends

Fig. S1 The positive rate graph highlights Saturdays and holidays

(a) The positive rate across Japan. Saturdays and holidays are highlighted with black shadow. Rapid increases and decreases essentially coincide with Saturdays and holidays.
(b) The positive rate of seven prefectures added to the graph of Japan overall. A similar coincidence is observed as in (a).

Fig. S2 Graphs for checking the quality of May data

(a) The shift of data of overall Japan and seven prefectures, Tokyo, Osaka, Saitama, Chiba, Kanagawa, Hyogo, and Fukuoka. Osaka and Kanagawa data are beyond 1 and Chiba data are close to 1 in the middle of March.
(b) Positive rate graphs for overall Japan compared with that based on Osaka data (blue), and the positive rate calculated from the data deleted, the Osaka data in May (red). The difference is smaller than the increments or decrements of spikes, thus, within the range of error.
(c) Comparing the graphs of the data of five prefectures, excepted Tokyo and Osaka, with overall Japan graphs. The large variations in Kanagawa graph in the middle of May make no sense mathematically. Kanagawa graph is relatively higher than the others. The other graphs almost coincident with the overall graph: there are a gradual decrease.

Supplemental Figures

Fig S1a

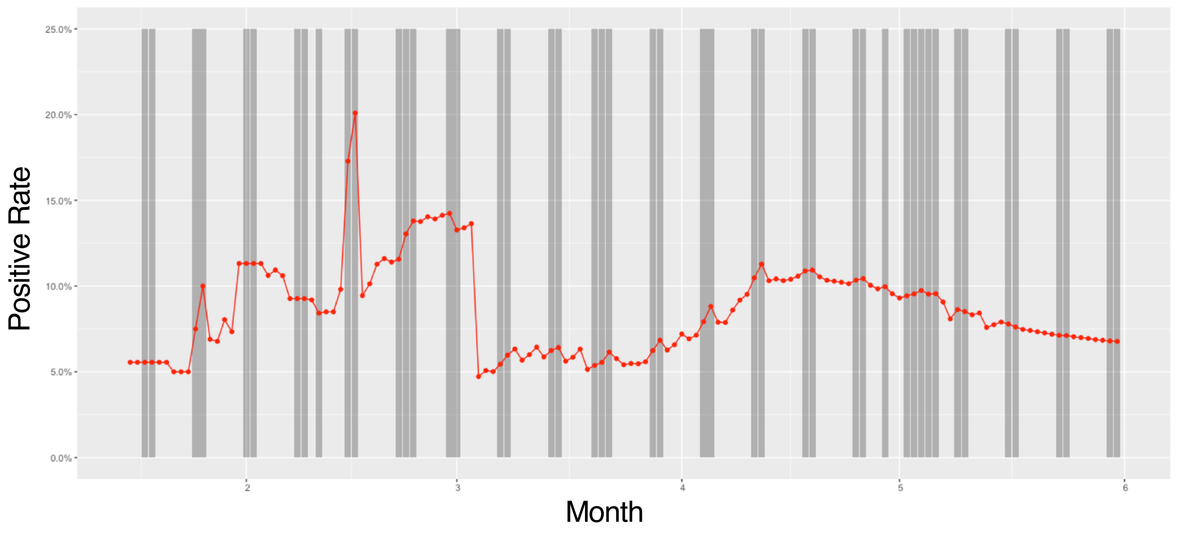

Fig S1b

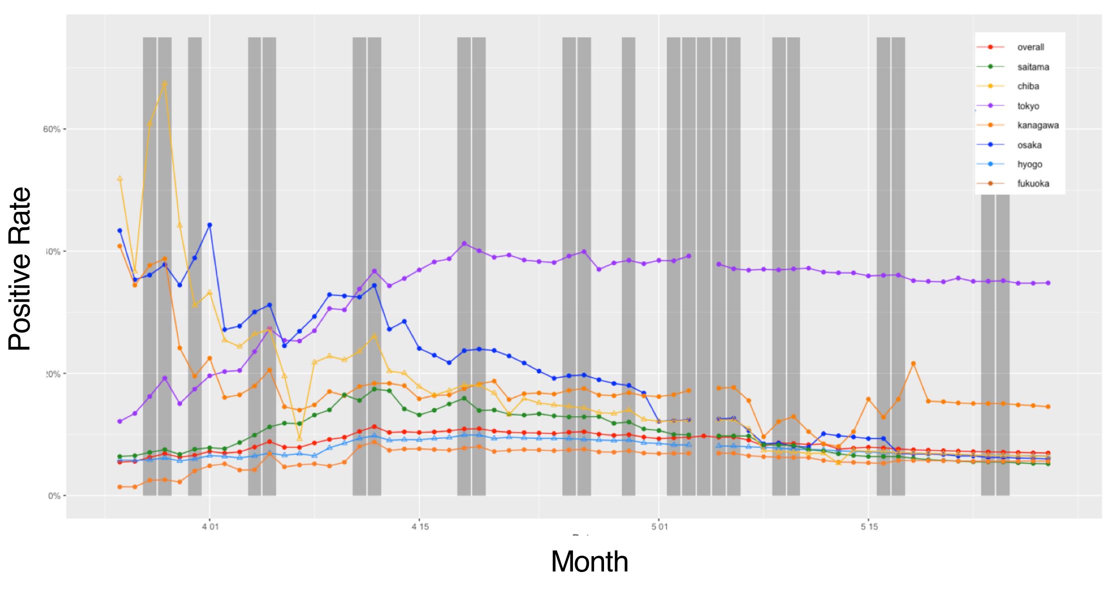

Fig S2a

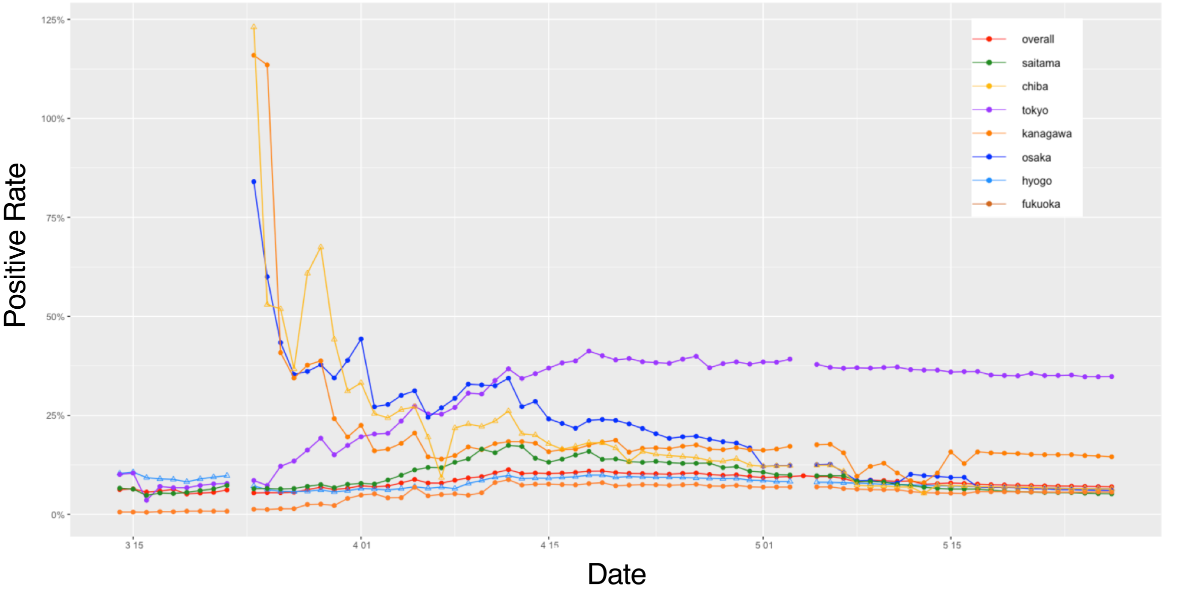

Fig S2b

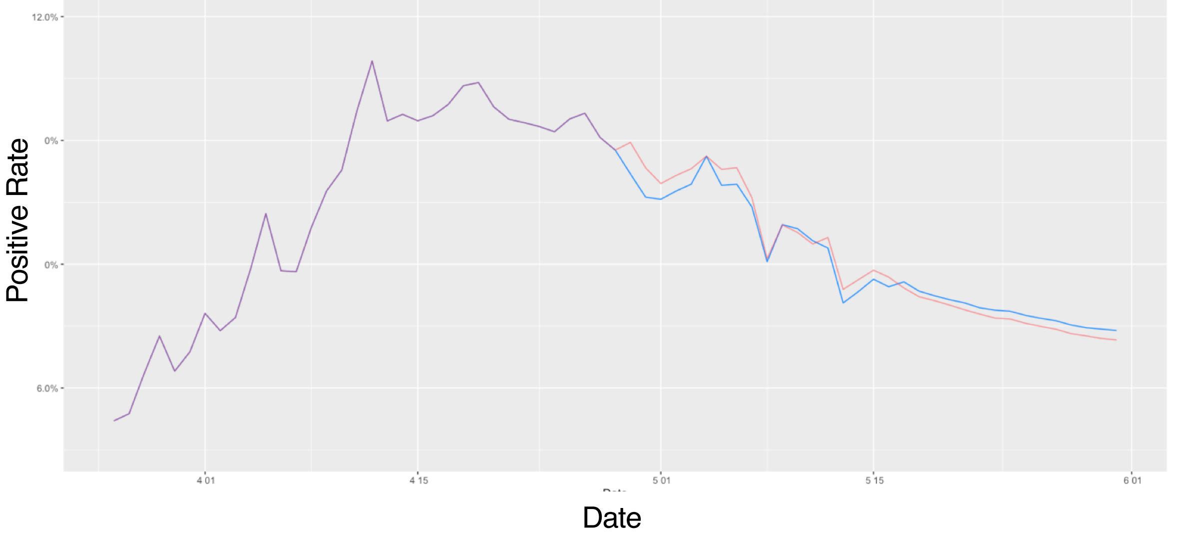

Fig S2c

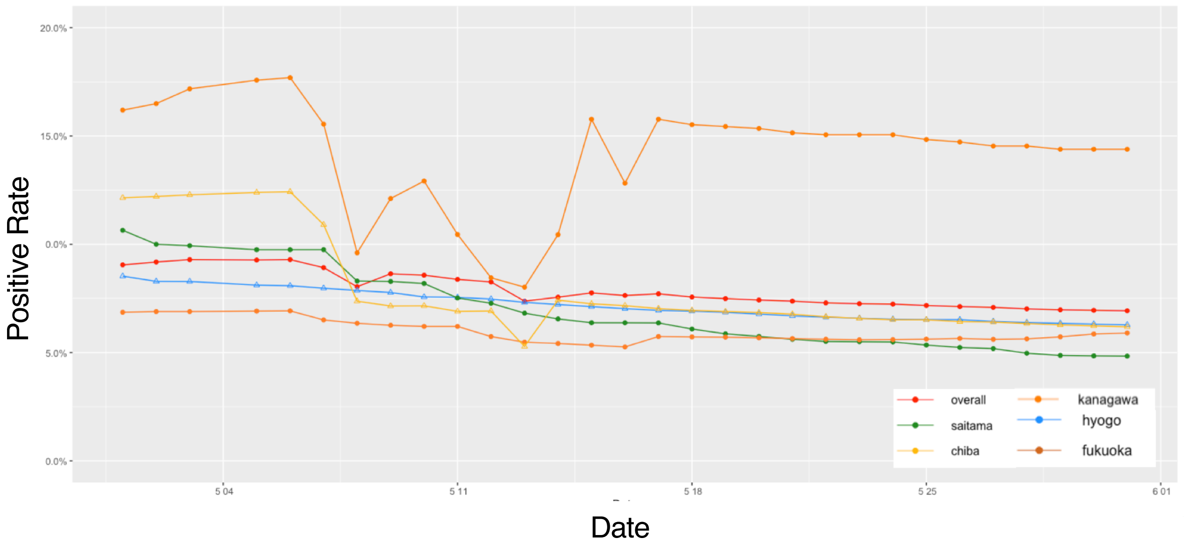

Fig S2d