Real time monitoring of COVID-19 outbreak using mobility trend data for Tokyo and Osaka, Japan

Junko Kurita[1)], Tamie Sugawara[2)], Yasushi Ohkusa[2)]

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

2) National Institute of Infectious Diseases, Tokyo, Japan

Corresponding author: Junko Kurita, kuritaj@tokiwa.ac.jp

Keywords: Apple, COVID-19, mobility trend data, peak, real time monitoring ,effective reproduction number

Abstract

Background: In Japan, as a countermeasure against the COVID-19 outbreak, voluntary restrictions against going out (VRG) were applied.

Object: We examined some associations among trip data provided by Apple Inc. and estimated an effective reproduction number R(t) at the city level.

Method: We regressed R(t) on a polynomial function of daily data from Apple Inc. using the whole period and the sub-period after March 10, until the end of April. We predicted R(t) prospectively using the estimation results obtained in May.

Results: Estimation results indicate R(t) as 5.22 in Tokyo and 1.62 in Osaka if VRG ceases and mobility reverts to a normal level. The critical level of R(t)<1 was found to be 70.0% in Tokyo and 91.2% in Osaka compared with the normal level.

Discussion and Conclusion: Results demonstrate that mobility data from Apple Inc. are useful for prediction of R(t). Moreover, a 9–30% point reduction in trips from home was sufficient to maintain R(t)<1.

Introduction

In preference to lockdowns such as those instituted in European and North American countries, national and local governments in Japan set voluntary restrictions against going out (VRG) to begin from the end of March as a countermeasure against the COVID-19 outbreak [1]. However, the intensity of VRG requirements, and popular compliance and cooperation with them, have changed over time. Although lockdowns such as those instituted in Europe and North America might have their cessation determined by governments, people in Japan can independently adjust their degree of personal effort when supporting Japan's VRG voluntary measures. Therefore, the government must monitor the actual circumstances and policy outcomes to moderate their requirements for VRG. Moreover, some delay occurs with reporting of the number of newly infected patients. One delay is attributable to the incubation period from infection to symptom onset. Another important delay occurs from the time of onset to the time of reporting it. Because of those delays, for approximately two weeks, one is unable to observe the precise number of newly infected people for a given day. A government working blindly for two weeks when conducting decision-making can not act in a timely or effective manner.

Particularly, although the emergency declaration was applied throughout Japan on April 8, the Osaka prefectural government loosened VRG starting May 16. By contrast,

VRG were retained in Tokyo. Therefore, careful and timely monitoring of circumstances in Osaka must be done to detect any signs of reemerging outbreak. Even for Tokyo, policy recommendations to loosen VRG must be made based on scientific evidence. Therefore, a system must be constructed to monitor the current outbreak situation and to evaluate how relaxed the VRG should become.

To date, susceptible–infected–recovery (SIR) models for COVID-19 incorporating countermeasures have emphasized the date of initiation of countermeasures [2–4]. However, VRG have expanded gradually, at least in the case of Japan. Therefore, an all-or-nothing approach such as that implied by a SIR model might not be appropriate. A more continuous variable is expected to be necessary to represent compliance with VRG over time.

Data for those variables have been reported at several sites including Apple Inc. and Alphabet Inc. (hereinafter Apple and Google, respectively) worldwide, and at Nippon Telegraph and Telephone (NTT) and East and West Japan Railway companies (JR) in Japan. Of those, Apple, as the front-runner of this reporting service, started providing related data from January 13. The daily information included aggregated data for the number of trips a telephone user makes from home by mode of transport, such as driving and walking [5].

Using mobility information provided by Apple, we examined associations and estimated the effective reproduction number R(t) in Tokyo and Osaka.

Methods

When applying a simple susceptible–infected–recovery (SIR) model [2–4] to the epidemic curves of Tokyo and Osaka, with their respective populations of 13 and 8.8 million, we assume an incubation period that conforms to the empirical distribution in Japan. The number of symptomatic patients reported by the Ministry of Labour, Health and Welfare (MLHW) for February 10 – May 15 published [6] on May 16 was used. Some patients were excluded from data: 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 the duration extending from onset of symptoms to the report date among patients for whom the onset date had been reported.

We estimated the onset date of patients for whom 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 for whom 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 $\Sigma_{k=1}f(k)N_s+k$ for the long duration extending from s.

Moreover, the reporting delay for published data from MHLW might be considerable. In other words, if $s+k$ is larger than that in the current period t, then $s+k$ represents the future for period t. For that reason, $Ns+k$ is not observable. Such a reporting delay engenders underestimation bias of the number of patients. For that reason, it must be adjusted as $\Sigma_{k=1}^{t-s}f(k)N_s+k\,/\Sigma_{k=1}^{t-s}f(k)$. Similarly, patients for whom the onset dates were available are expected to be affected by the reporting delay. Therefore, we have $M_s|_t/\Sigma_{k=1}^{t-s}f(k)$, 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 defined R(t) as the number of the infected patients on day t divided by the number of patients who were presumed to be infectious. The number of infected patients was calculated from the epidemic curve by the onset date using a distribution of incubation period. The distribution of infectiousness in symptomatic and asymptomatic

cases was assumed to be 30% on the onset day, 20% on the following day, and 10% for the subsequent five days [7].

To clarify associations among R(t) and Apple data, we regressed R(t) on a polynomial function of daily Apple data in Tokyo and Osaka. The order of the polynomial function was determined stepwise from the model including only a linear term to the model including higher order if all estimated coefficients were significant. In addition to estimation using the whole period, the sub-periods split by March 10 were analyzed. When we use data in the whole sample or after March 10, we use data until the end of April for estimation and predict and evaluate it using data in May.

We inferred significance for any result under the 5% level.

Results

As of May 17, using data for February 10 – May 16, 3,742 in Tokyo and 1,824 in Osaka, community-acquired cases were identified, excluding asymptomatic cases. Figure 1 presents the empirical distribution of the duration of onset to report in Japan. The maximum delay was 30 days. Figure 2 depicts the empirical distribution of incubation periods among 125 cases for which the exposed date and onset date were published by MHLW in Japan. The mode was six days. The average was 6.6 days.

Figure 3 depicts Apple data obtained for Tokyo and Osaka. These almost coincide. The peak of Apple data occurred in February, but a second and smaller peak occurred around the middle of March. Subsequently it declined until the end of the data period.

Figure 4 depicts an epidemic curve, newly infected people by the infection date, estimated R(t),and prediction based on the result shown in the table for Tokyo and Osaka. The peak of the epidemic curve occurred at around April 3 in both Tokyo and Osaka. The peak of R(t) in Tokyo occurred around March 13, but it was slightly later in Osaka: March 17.

Table 1 presents the estimation result. For Tokyo, only the linear term was selected in the whole period and after the March 10 period. Conversely, up to cubic terms were selected in the whole period in Osaka, but the quadratic terms were selected after the March 10 period. The adjusted determinant coefficients in Tokyo were worse than in Osaka. Especially, they were higher than 0.9 in Osaka on the subperiod after March 10.

Figure 5 depicts R(t) and predicted line from Apple using data after March 10 until the end of April. In May, the prediction from Apple data was prospective. The goodness of fit in Osaka was much better than in Tokyo, even for May.

The estimation result indicates that R(t) will be 5.22 in Tokyo and 1.62 in Osaka

if VRG ceases and mobility reverts to a normal level, which was 100 for the Apple data of January 13. Therefore, complete cessation of VRG will re-initiate the outbreak. The critical level of $R(t)<1$ was found to be 70.0 in Tokyo and 91.2 in Osaka.

Discussion

Estimation results and their depiction in Figure 5 suggest that prediction in Osaka was better than that in Tokyo. Especially, the goodness of fit achieved in Osaka using data after March 10 indicated it as having the best fit. It might be consistent with an apparent change in the dominant virus among SARS-CoV-2 occurring in March [8]. Estimations seem to indicate that the Wuhan-originated strain was dominant before the end of March. However, after the end of March, a mutated virus became dominant in European countries and the USA. These two virus subtypes might have different infectiousness and therefore might have different associations with the mobility data.

Results also show critical levels of $R(t)<1$ as 83.8 in Tokyo and 91.2 in Osaka from Apple data. Consequently, a 10% point in Osaka and 16% point in Tokyo reduction in trips from home will be necessary, probably for more than one year, until herd immunity is achieved. Therefore, complete cessation of VRG might not be necessary to avoid re-emergence of an outbreak. Current VRG data indicate a 40% point reduction from Apple data. Therefore, restrictions against going out can be relaxed by

10% points in Tokyo and 30% points in Osaka. Such relaxation might include restarting of schools or private events and sports that involve small numbers of players, little player contact, and large outside spaces. Large entertainment events or professional sports events with numerous participants might continue to be risky. In Tokyo, more stringent restriction than in Osaka might be necessary. In other words, partial relaxation of VRG consistent with a controlled outbreak and economic activity can be monitored for R(t) to maintain it as less than one, but not as an extremely small number. Especially, we showed its predictive power prospectively in May as very high. Therefore, credibility might be high, even in Tokyo. Apple Inc. data might support fine control of the outbreak.

An advisory council for control of the COVID-19 outbreak in Japan has required an 80% point reduction since the emergency declaration issued on April 8 [9]. Actually, trips outside the home decreased 40% from the normal level, but the outbreak has been waning, as shown in Figure 5. It has become apparent that the requirement was greater than necessary. The experience has at least produced some evidence of what a request of 80% reduction elicits from people. Our study results indicate that a mere 30% point level in Tokyo or 9% point reduction might be sufficient to control an outbreak. For the early stage of the outbreak, R_0 of 0.6 was found [10], suggesting conditions under which

an outbreak would never occur in Japan. Such estimates might also mislead policies for countermeasures in Japan, which have so far necessitated adherence to contact tracing to detect clusters.

Earlier studies [11] that have used Apple and Google data worldwide have examined only linear terms. Nevertheless, we found that higher terms might be important to predict R(t). Higher predictive power from Apple data might imply that prediction of R(t) using Apple data is more reliable. Especially, the estimated R(t) for a few days was less precise because the incubation period is about six days, on average. Increasing R(t) beyond time passing until six days was a widely observed phenomenon. Conversely, Apple data are available the next day. It might be possible to predict the latest R(t) from Apple data.

We used Apple Inc. data only instead of data from Google or other sites. Apple started to provide data the earliest and has provided it continuously on a timely basis. Therefore, it has remained the most appropriate to evaluate the outbreak dynamics and to predict them prospectively.

The present study has some limitations. First, R(t) was not determined for the number of contacts only. It depends on other circumstances related to outbreaks including the proportion of susceptible persons or the infectiousness of asymptomatic

people. These were not incorporated into the analysis of R(t). To take such data into consideration, one must use a SIR model that includes asymptomatic cases. That is anticipated as a challenge for future research.

Secondly, although we showed Apple data have predictive power, they were examined only in May when the outbreaks were almost contained. In other words, Apple data can facilitate precise prediction at least if R(t) were small and stable. If the outbreak reemerged as it first did in March, the predictive power might be worse in Tokyo because Apple data were insufficiently able to explain the peak of R(t) in March, as shown in Figure 5. Conversely, in Osaka, one can expect sufficient predictive power even if an outbreak were to be re-initiated.

Thirdly, one must be reminded that Apple data reflect the proportions of users leaving their residence. The data do not indicate a number or rate of contact directly. In other words, Apple data show no intensity of the respective contacts. However, measurements of contact intensity are extremely difficult. Such measurement methods therefore represent a future research objective.

Fourthly, although Apple data were better than others, the users of Apple products and services might be more likely to include young and healthy persons. However, information from NTT or JR might not be limited to users. Therefore, combinations of

these data with Apple data might yield better results than those obtained from the present study. Exploration of those possibilities remains as a challenge for our future research.

Conclusion

Results demonstrated that mobility data from Apple are useful for prediction of R(*t*). Moreover, a 9–30% point reduction in trips from home was found to be sufficient to maintain R(*t*)<1. An emergency declaration in Osaka was released on May 21. Careful monitoring of Apple data for Osaka is necessary.

Emergency declaration in Japan on April 7 had been ceased on May 25. However, the observed R(t) reached about two on May 16 as of the end of May shown in Figure 6, theough the end of May was beyond study period in the present study. We have to monitor carefully Apple data which is real time index as well as the observed R(t) which will delay for some days to stable.

This report of our study is based on the authors' opinions: it does not reflect any stance or policy of their respective professionally affiliated bodies.

Acknowledgments

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

References

1. Japan Times. Tokyo governor urges people to stay indoors over weekend as virus cases spike https://www.japantimes.co.jp/news/2020/03/25/national/science-health/tokyo-logs-40-coronavirus-cases/#.Xr4TV2eP604 [accessed on May 14, 2020]
2. 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
3. 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

4. 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.

5. Apple. Mobility trend Data.https://www.apple.com/covid19/mobility (in Japanese) [accessed on May 13]

6. Japan Ministry of Health, Labour and Welfare. Press Releases. https://www.mhlw.go.jp/stf/newpage_10723.html (in Japanese) [accessed on May 14, 2020]

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. Morb Mortal Wkly Rep. 2020 Apr 3;69(13):377-381. doi: 10.15585/mmwr.mm6913e1.

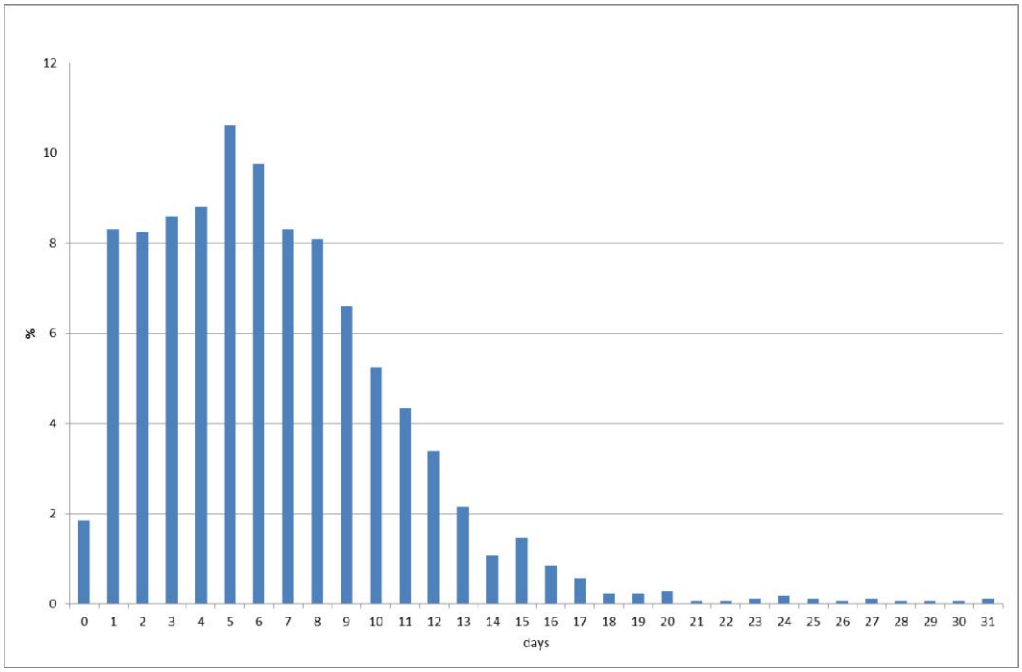
8. National Institute of Infectious Diseases. Survey of genomic molecular epidemiology survey for SARS-CoV-2. https://www.niid.go.jp/niid/ja/basic-science/467-genome/9586-genome-2020-1.html (in Japanese) [accessed on May 14, 2020]

9. Japan Times. Japan state of emergency to cover Tokyo, Osaka and five other prefectures https://www.japantimes.co.jp/news/2020/04/06/national/japan-state-of-emergency-covid-19/#.XpPANv37RaQ [accessed on May 15,2020]

10. Nishiura H, Oshitani H, Kobayashi T, Saito T, Sunagawa T, Wakita T, Suzuki M, MHLW COVID-19 Response Team. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). doi: https://doi.org/10.1101/2020.02.28.20029272

11. Bergman N, Fishman R. Mobility Reduction and Covid-19 Transmission Rates. doi: https://doi.org/10.1101/2020.05.06.20093039

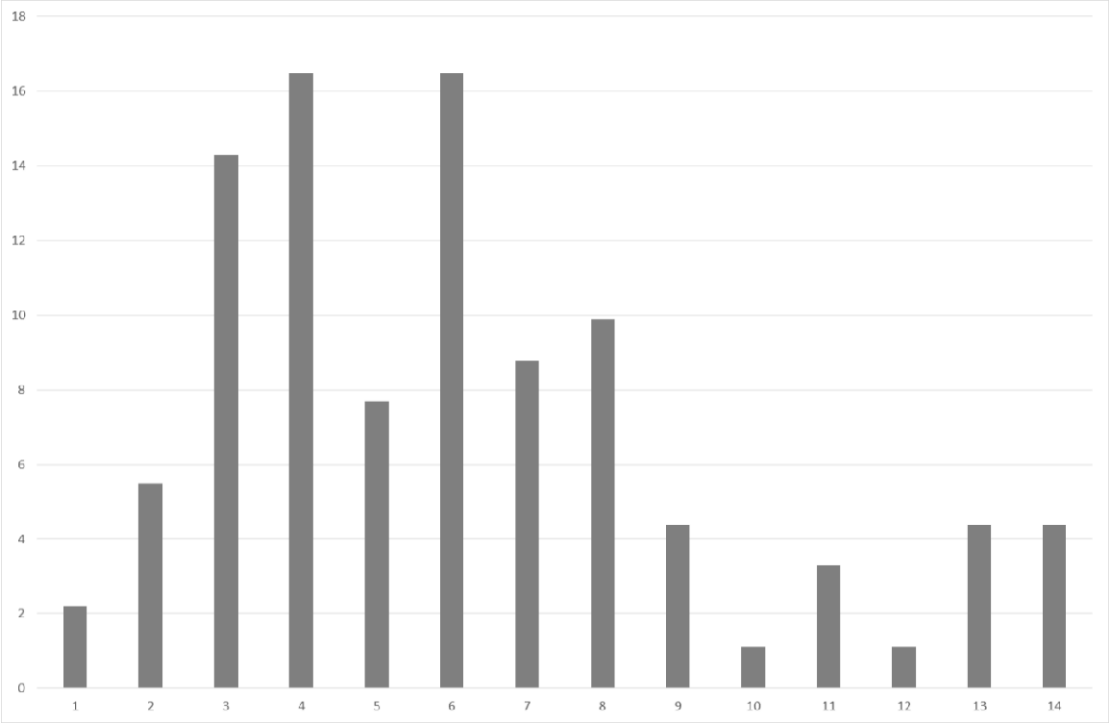
Figure 1: Empirical distribution of durations from onset to reporting (MLHW, Japan). (%)



Note: Bars represent the probability of duration from onset to reporting based on 657 patients in Japan for whom onset dates were available.

Figure 2: Empirical distribution of the incubation period (MLHW, Japan).

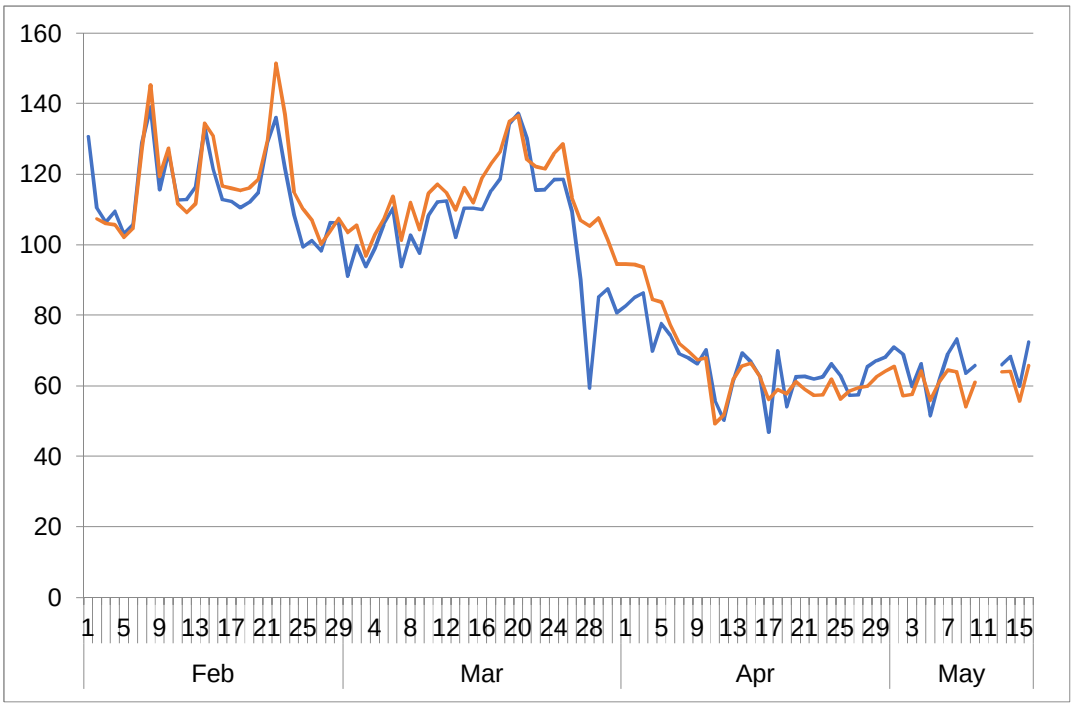
(%)



(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. Patients for whom incubation was longer than 14 days are included in the bar shown for day 14.

Figure 3: Apple data for Tokyo and Osaka.

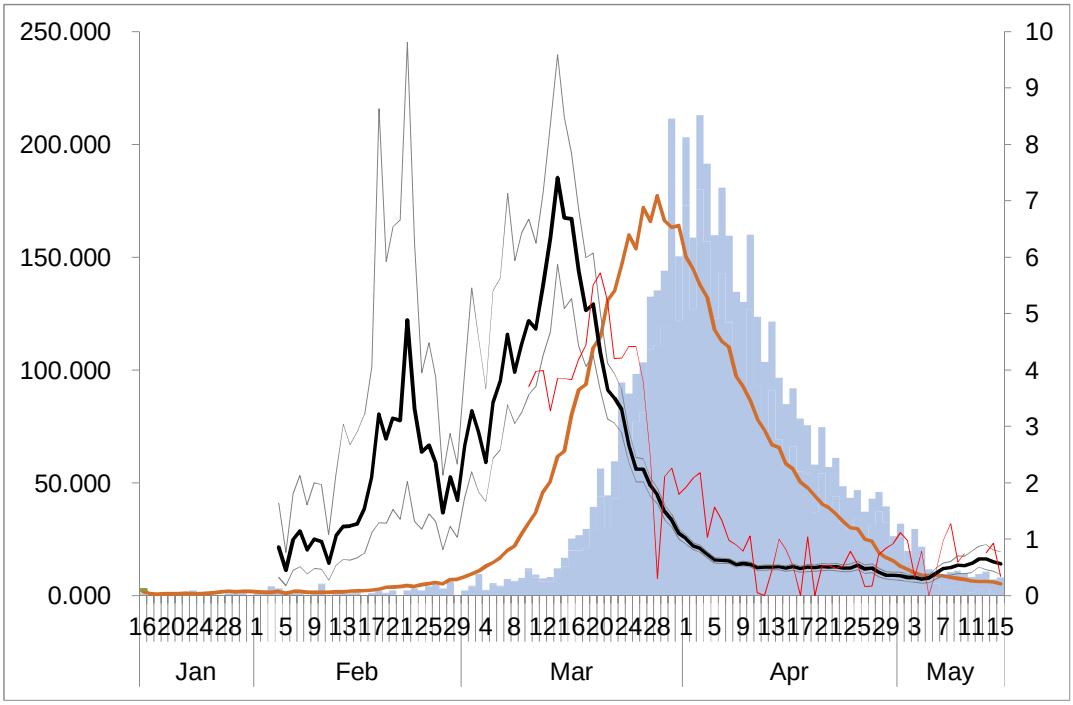


Note: The blue line represents Apple data for Tokyo. The red line represents data for Osaka. These were normalized to 100 on January 23, 2020.

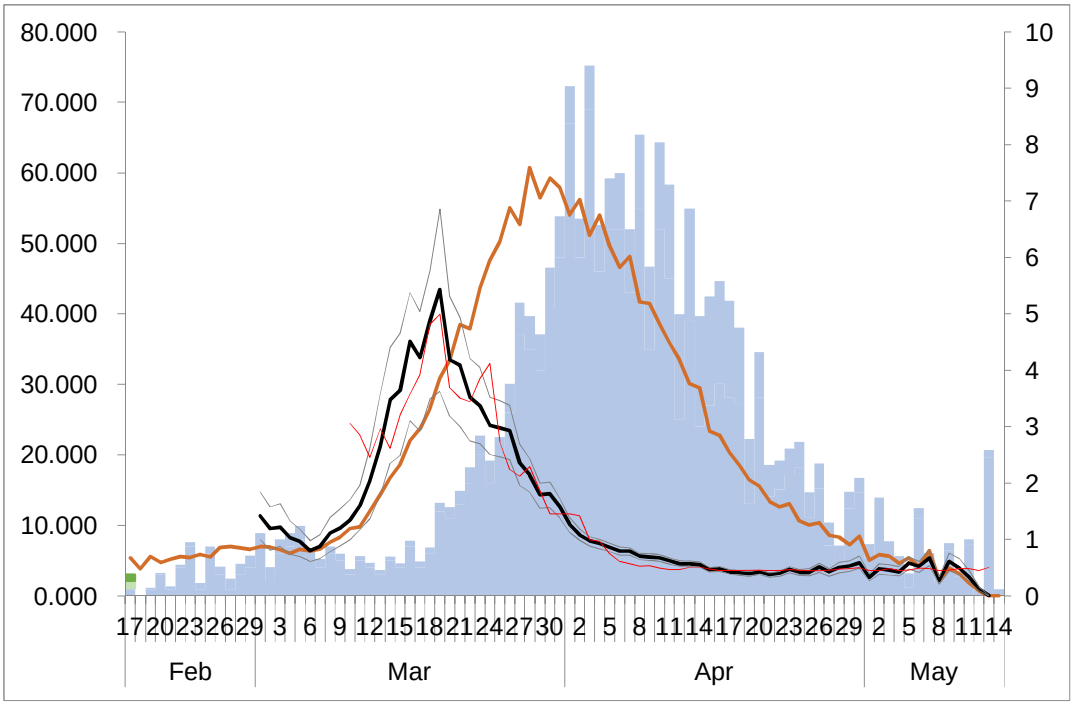
Figure 4: Epidemic curve, R(t) and prediction from the Apple data in Tokyo and Osaka

A: Tokyo

(Number of patients)

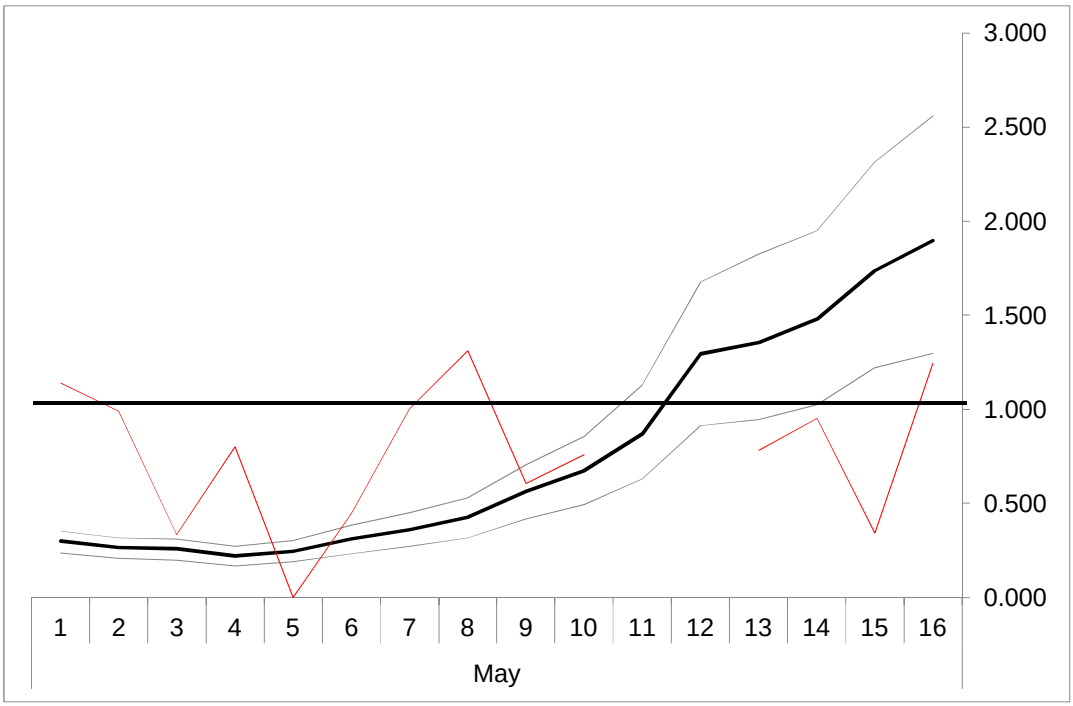


B: Osaka



Note: Bars represent the epidemic curve. The bold red line represents newly infected persons, measured according to the scale on the left-hand-side. The bold black line represents the estimated R(*t*). The two thin black lines show its 95% confidence interval. The thin red line represents R(*t*) predicted from Apple data. R(*t*) and the related prediction were measured according to the scale on the right-hand-side.

Figure 5: R(*t*) and prediction from the Apple data in Tokyo in May as of the end of May



Note: The bold black line represents the estimated R(*t*) as of the end of May. The two thin black lines show its 95% confidence interval as of the end of May.. The thin red line represents R(*t*) predicted from Apple data as of the end of May.. R(*t*) and the related prediction were measured according to the scale on the right-hand-side.

Table 1: Results of estimated R(*t*) inferred from Apple data

	Tokyo				Osaka			
	Whole period		After March 10		Whole period		After March 10	
	Est. coeff.	*p*	Est. coeff.	*p*	Est. coeff.	*p*	Est. coeff.	*p*
Linear term	0.0500	0.000	0.069	0.000	-0.5412	0.034	-0.08117	0.000
Quadratic term					0.5698×10^{-2}	0.032	0.7147×10^{-3}	0.000
Cubic term					-0.1740×10^{-4}	0.049		
Constant	-2.448	0.000	-3.813	0.000	16.18	0.035	2.752	0.002
Adjusted R^2	0.4368		0.2258		0.5177		0.9104	
No. observ,	81		52		74		64	

Note: The dependent variable was R(*t*); explanatory variables were polynomial functions of Apple data. Order was selected stepwisely where all coefficients were significant.