学位論文

Epidemiological studies on effective reproductive number and asymptomatic infections of COVID-19

（COVID-19 の実効再生産数及び不顕性感染についての疫学的研究）

2022 年 3 月

北海道大学

中條　航

学位論文

Epidemiological studies on effective reproductive number and asymptomatic infections of COVID-19

（COVID-19 の実効再生産数及び不顕性感染についての疫学的研究）

2022 年 3 月

北海道大学

中條　航

Table of Contents

List of manuscripts

Part of this study was presented in manuscripts below:

1. Ko Nakajo, Hiroshi Nishiura. Assessing Interventions against Coronavirus Disease 2019 (COVID-19) in Osaka, Japan: A Modeling Study. J Clin Med. 2021 Mar 18;10(6):1256.

2. Ko Nakajo, Hiroshi Nishiura. Transmissibility of asymptomatic COVID-19: Data from Japanese clusters. Int J Infect Dis. 2021 Apr;105:236-238.

3. Ko Nakajo, Hiroshi Nishiura. Exploring secondary SARS-CoV-2 transmission from asymptomatic cases using contact tracing data. Theor Biol Med Model. 2021 Jul 16;18(1):12.

Abstract

【Background and Objectives】

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that reached pandemic levels in 2020. It remains unclear whether the pandemic will come to an end soon with the potential emergence of new variants of concerns. The key parameters governing infectious dynamics of ongoing epidemics of COVID-19 are 1) the effective reproduction number, R(t), and 2) the relative transmissibility of asymptomatic cases as compared with symptomatic cases. To accurately estimate these critical epidemiological parameters, we must address the issue of "observability". COVID-19 infection events are generally not directly observable, and full datasets are rarely available. As a result, most studies have leveraged information on illness onset or the serial interval, or both, to estimate R(t). While the estimation of R(t) based on illness onset is conventional, accumulating evidence suggests that pre-symptomatic transmission contributes to the secondary transmission of COVID-19. Estimates of R(t) based on illness onset data seldom consider pre-symptomatic transmission. Similarly, individuals with asymptomatic SARS-CoV-2 infection can propagate the virus without being "noticed" and thus understanding viral transmissibility among asymptomatic individuals is critical for designing strategy controlling COVID-19. However, we seldom have opportunities to investigate on this because asymptomatic transmissions are usually hard to be traced (= observed). To address the issue of "observability" in estimation of each parameter, the objectives of this thesis are 1) to propose an alternative method of estimating R(t) as a function of time of infection using observable data while explicitly incorporating the pre-symptomatic transmission into the model and 2) to estimate the relative reproduction number of asymptomatic cases compared with symptomatic cases.

【Methods】

This thesis consists of two parts: In Section 1, we explored an alternative method of estimating R(t) and proposed a modified renewal equation using observable data while explicitly incorporating the pre-symptomatic transmission into the model. This method was applied to the early epidemic in Osaka prefecture to see the impact of public health and social measures (PHSM) on R(t). The negative log likelihood values of all possible combinations of "event-based" models were compared to account for the loss of single or combination of events. A Joinpoint segmented regression model was also used to assess whether a significant change in R(t) during each wave was associated with any of the start dates of key interventions. In Section 2, the relative reproduction number of

asymptomatic cases compared with symptomatic cases was estimated. The data of two early clusters in Japan was used, in which the information of symptomatic status and time of illness onset were meticulously collected. Assuming that the number of secondary cases resulting from either primary symptomatic or asymptomatic cases independently followed negative binomial distributions, the relative reproduction numbers of an asymptomatic case compared with a symptomatic case as well as dispersion parameter was estimated with 95% CI in each cluster. Whether symptomatology was associated with transmission of symptomatic vs. asymptomatic infections was also assessed. We explored the impact of isolation on the transmissibility of asymptomatic cases, using the probability distribution function of the serial interval shortened by isolation.

【Results】

The “simple” alternative method was successful in estimating the R(t) for COVID-19 over the course of the epidemic in Osaka. Based on estimated R(t), the epidemic was found to come under control around 2 April 2020 during the first wave, and 26 July 2020 during the second wave. Consistent patterns were revealed when our estimates were compared with the R(t) estimates based on the back-projected incidence of infection. R(t) did not decline drastically following any single intervention. However, when multiple interventions were combined, the relative reductions in R(t) during the first and second waves were 70% and 51%, respectively. During the second wave, a significant impact on the observed secondary transmission patterns was not produced by the combined effect of interventions that focus on high-risk groups. The reproduction number of symptomatic cases in Tokyo/Kanagawa cluster and Kyoto cluster was estimated at 1.2 (95% CI: 0.5–2.9) and 1.14 (95% CI: 0.61–2.09), respectively. The relative reproduction number for asymptomatic cases for each cluster was estimated at 0.27 (95% CI: 0.03–0.81) and 0.19 (95% CI: 0.03–0.66), respectively. Any apparent increased tendency for symptomatic primary case to produce symptomatic secondary cases was not found. We also assessed the relative transmissibility in the model using the probability density function of the generation interval adjusted for the isolation period in Kyoto cluster because movement of all identified close contacts was restricted for 14 days. However, the effectiveness of case isolation could not be estimated jointly with other parameters.

【Discussion】

We proposed the first method to estimate R(t) as a function of time of infection using observable data (= incident cases as a function of time of illness onset). The limitations in our method included 1) it could not estimate R(t) for the most recent 12 days and 2) the heterogeneities of host population were not incorporated into the model. The results

of PHSM impact on R(t) in second wave might suggest infection control focusing on high-risk groups is not substantial. The relative reproduction number for asymptomatic cases was estimated to be 20 – 30 % of symptomatic cases, which was broadly consistent with the previous report. Contact tracing focusing on symptomatic index cases may be justified when there is limited testing capacity. The sample size was small for each cluster, involving a broad uncertainty bound and a wide 95% CI.

【Conclusion】

Two key epidemiological parameters governing transmission dynamics of COVID-19 were investigated using mathematical modeling approach while addressing the issue of "observability" carefully. The alternative method of estimating R(t) based on observable illness onset data was firstly devised and successful in estimating the R(t) for COVID-19. The transmissibility of asymptomatic cases of SARS-CoV-2 infection was shown to be small relative to symptomatic cases. Our findings suggested that concerted efforts would be required to curb the COVID-19 epidemic and contact tracing focusing on symptomatic index cases may be justified when there is limited testing capacity. This study would assist planning the strategy for the efficient and feasible public health actions in future pandemics.

Abbreviations

CI	Confidence interval
COVID-19	Coronavirus disease 2019
PHSM	Public health and social measures
R(t)	The effective reproduction number
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2

Overall introduction

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that reached pandemic levels in 2020. Clinical manifestations range from non-specific upper or lower respiratory symptoms to severe pneumonia and even death. Mass vaccination including booster dose has been implemented globally, but it remains unclear whether the pandemic will come to an end soon with the potential emergence of new variants of concerns.

The key parameters governing infectious dynamics of ongoing epidemics of COVID-19 include 1) the effective reproduction number, R(t), and 2) the relative transmissibility of asymptomatic cases as compared with symptomatic cases. To accurately estimate these critical epidemiological parameters, we must address the issue of "observability": in estimating R(t), most previous studies have relied on information on observable illness onset, then to back-calculate incidence of infection that is usually unobservable. Generation interval is also usually unobservable, so serial interval would be substituted into a renewal equation at the expense of ignoring pre-symptomatic transmission, which is the unique feature of COVID-19. Similarly, individuals with asymptomatic SARS-CoV-2 infection can propagate the virus without being "noticed" and thus understanding viral transmissibility among asymptomatic individuals is critical for designing strategy controlling COVID-19. However, we seldom have opportunities to investigate on this because asymptomatic transmissions are usually hard to be traced (= observed).

To address the issue of "observability" in estimation of each key parameter governing COVID-19 dynamics, this thesis consists of two parts: Section 1 will deal with estimation of R(t) and propose an alternative method of estimating R(t) as a function of time of infection using observable data while explicitly incorporating the pre-symptomatic transmission into the model. This method will be applied to the early epidemic in Osaka prefecture to see the impact of public health and social measures on R(t). In following Section 2, the relative reproduction number of asymptomatic cases compared with symptomatic cases will be estimated. We will leverage the data of two early clusters in Japan in which the information of symptomatic status and time of illness onset were meticulously collected. Whether symptomatology was associated with transmission of symptomatic vs. asymptomatic infections will be assessed exploratory.

Section 1 – Study on effective reproductive number of COVID-19

Introduction

The effective reproduction number (=R(t)) is the actual average number of secondary cases per primary case at calendar time t. Because the frequency of transmission over time is reflected in R(t), it has been recognized as an epidemiologically useful tool in assessing transmission dynamics in populations with varying susceptibility levels and experiencing various public health interventions. It has been employed to evaluate effects of PHSMs implemented during COVID-19 and other recent pandemic (i.e., SARS) [Abbott et al., 2020; Caicedo-Ochoa et al., 2020; Cori et al., 2013; Flaxman et al., 2020; Kucharski et al., 2020; Kuniya et al., 2020; Pan et al., 2020; Ryu et al., 2020; Scire et al., 2020; Tariq et al., 2020; Wallinga and Teunis, 2004].

There are, however, a few caveats in utilizing R(t); first, a threshold for vaccination cannot be provided by R(t) but R0, i.e., basic reproductive number. Second, there exist technical challenges in accurately estimating R(t) in real-time. Renewal equation, from which R(t) can be theoretically delivered, needs both the incidence of infection and the generation time (i.e., the time from infection in the primary case to infection in the secondary case) [Gostic et al., 2020; Nishiura and Chowell, 2009; Wallinga and Lipsitch, 2007]. We cannot, however, observe infection events in most cases and have full datasets available and this is also case for COVID-19. Consequently, information on illness onset or the serial interval (i.e., the time from illness onset in the primary case to illness onset in the secondary case), or both have been conventionally leveraged to estimate R(t) in most studies. Third, the substantial contribution of pre-symptomatic transmission to the secondary transmission of COVID-19 has been suggested by accumulating evidences [He et al., 2020; Nishiura, Linton and Akhmetzhanov, 2020; Liu et al., 2020; Yang et al., 2020]. Therefore, estimates of R(t) of COVID-19 should take into account of pre-symptomatic transmission.

A novel method to estimate R(t) as a function of the date of infection using the observable data of illness onset is proposed to address the abovementioned challenges. This method is applied to epidemiological data from the Osaka prefecture, Japan. The relationship between trends in R(t) and the implementation of PHSM in Osaka was also exploratory assessed.

Methods

1. Epidemiological Data

The date of illness onset of COVID-19 cases in the Osaka prefecture was analyzed.

COVID-19 cases are confirmed by reverse transcriptase polymerase chain reaction and all confirmed cases are notified to the government. The index COVID-19 case was reported on 29 January 2020 and a total of 11,249 cases had been reported as of 13 October 2020. 8818 cases for which information of illness onset by 28 September 2020 was available were analyzed. 17 February as day 0 was used in subsequent analyses because no secondary cases were likely to be produced by the index case given the time window of illness onset from 29 January to 17 February (i.e., the day of illness onset for the second case identified).

The governor of the Osaka prefecture implemented several PHSMs throughout the period of this investigation (Table 1). In late March, Osaka residents had been requested to voluntarily abstain from weekend social activities and to refrain from going to bars and nightclubs before the declaration of a state of emergency by the Japanese government on 7 April. The criteria for its alert system, referred to as the “Osaka model,” to confront future epidemics were announced by the governor when the first wave was close to its end. They issued the first alert on 12 July and sequentially requested the voluntary restriction of social behaviors and the closing of shops/bars in specified areas of Osaka city.

Table 1. Chronology of the coronavirus disease 2019 (COVID-19) epidemic in Osaka.

Calendar date	Analysis date	Description
First wave		
17 February	Day 0	The index case developed illness.
1 March	Day 13	The governor requested the government to dispatch an emergency operations center team ("cluster busters team").
19 March	Day 31	Voluntary restrictions on crossing the border of Osaka (especially, between Osaka and Hyogo) were requested.
27 March	Day 39	Voluntary restrictions on weekend outings were requested.
31 March	Day 43	Voluntary restrictions on eating at restaurants operating at nighttime were requested.
3 April	Day 46	Voluntary restrictions on weekend outings were requested for the second time.
7 April	Day 50	A state of emergency was declared by the Prime Minister.
5 May	Day 78	The original classification of the epidemiological situation (the "Osaka model") was announced.
21 May	Day 94	Osaka was released from the state of emergency.
Second wave		
12 July	Day 146	A first alert was issued according to the Osaka model.
28 July	Day 162	Voluntary restrictions on social events involving the consumption of alcohol by more than four persons were requested.
31 July	Day 165	Restaurants and bars in specific districts of Osaka city were requested to voluntarily close.
21 August	Day 186	Restaurants and bars were allowed to open.
31 August	Day 196	Restrictions on social events involving the consumption of alcohol by more than four persons were lifted.

2. Estimation of R(t)

A renewal process model is used for estimation of R(t). If j(t) represents the incidence of infection at calendar time t, the commonly used renewal equation is:

$$j(t) = R(t) \int_0^\infty j(t-s)g(s)ds, \quad (1)$$

where g(s) is the probability density function of the generation time s. The fact that j(t) must be known is a fundamental practical issue for estimating R(t) from Equation (1). We cannot, however, directly observe j(t) in infectious disease epidemiology. This is reason why most analyses of COVID-19 epidemiology prior to my work had to back-calculate j(t) based on the incidence of illness onset, c(t), which is readily observable. With the assumption that the density function of the incubation period is known and the incubation period is independently and identically distributed, we could non-parametrically back-project the daily incidence of infection. Two steps of inference (i.e., back-calculation and solving the renewal equation) are required in this estimation process for R(t), rendering it not statistically rigorous method.

Here, the improved renewal equation is proposed to avoid this issue by adhering to the date of illness onset while estimating R(t) based on the date of infection. Let us assume that the relative frequency of secondary transmission with respect to disease-age (i.e., the time since illness onset) u, λ(u), is known. j(t), the daily incidence of infection at calendar time t, can then be written as follows, using the daily number of new illness onsets, c(t):

$$j(t) = R(t) \int_{-x}^\infty c(t-u)\lambda(u)du, \quad (2)$$

where R(t) is the instantaneous measure as a function of the date of infection, and pre-symptomatic secondary transmission is assumed to take place from x days prior to illness onset. Assuming that the probability density function of the incubation period is f(τ), the relationship between the daily number of new illness onsets and the daily incidence of infection can be described as:

$$c(t) = \int_0^\infty j(t-\tau)f(\tau)d\tau. \quad (3)$$

Instead of using this equation for back-calculating j(t) from c(t), j(t) in the right-hand side of Equation (3) is substituted by the right-hand side of Equation (2). We can then describe the relationship between the daily number of new illness onsets at time t and the effective reproduction number R(t) by:

$$c(t) = \int_0^\infty R(t-\tau) \int_{-x}^\infty c(t-\tau-u)\lambda(u)f(\tau)du\, d\tau. \quad (4)$$

Of note, the epidemic dynamics can be described by Equation (4) using observable data by producing c(t) (i.e., the epidemic curve drawn by date of illness onset). With this modified renewal equation, we can estimate R(t) as a function of the date of infection, t, using information on observable illness onset.

A gamma distribution with the peak at symptom onset and 12.3 days as the starting point for infectiousness was used for the relative frequency of secondary transmission at disease-age u, λ(u); these values were estimated from 77 transmission pairs [He et al., 2020]. A lognormal distribution with a mean of 5.2 days was used for the probability density of the incubation period, $f(\tau)$, which was estimated from 425 patients in Wuhan [Li et al., 2020], consistent with the results of Linton et al. [Linton et al. 2020]. The probability density function of the incubation period and the frequency of secondary transmission relative to disease-age were assumed to be independently and identically distributed and we ignored the heterogeneity of transmission, including age dependence and spatial dependence. It was also assumed that the extent of underreporting (i.e., ascertainment bias) remained unchanged over time.

By discretizing Equation (4) and changing the lower limit of the integration range in the second convolution to zero, we can model the expected value of the daily incidence (i.e., the number of new illness onsets) as follows:

$$E(c_t) = \sum_{\tau=0}^{t} R_{t-\tau} f_\tau \sum_{v=0}^{t-\tau+x} c_{t-\tau+x-v} \lambda_{v-x}. \quad (5)$$

where f_τ is now discrete and referred to as the probability mass function of the incubation period. Let us assume that the daily number of reported cases will follow a Poisson distribution. The likelihood function to estimate R_t is then:

$$\prod_t \frac{\left(\sum_{\tau=0}^{t} R_{t-\tau} f_\tau \sum_{v=0}^{t-\tau+x} c_{t-\tau+x-v} \lambda_{v-x}\right)^{c_t} \exp\left[-\sum_{\tau=0}^{t} R_{t-\tau} f_\tau \sum_{v=0}^{t-\tau+x} c_{t-\tau+x-v} \lambda_{v-x}\right]}{c_t!} \quad (6)$$

A piecewise constant model for R(t) that changes its value every 5 days was used [Nishiura and Chowell, 2014]. We choose this 5-day period specifically because it is in the range of published estimates of the serial interval for COVID-19 [Ganyani et al., 2020]. As a part of the sensitivity analyses, we modeled R(t) using 3-day and 4-day periods instead of the 5-day period. Using these different models, we compared the dates on which the R(t) took values below 1 for the first time during each wave.

Maximum likelihood estimates of R(t) were obtained by minimizing the negative logarithm of Equation (5). Parametric bootstrapping using the Hessian matrix *H* was implemented to quantify the confidence intervals (CIs) of R(t). 1000 resamples of

parameters were obtained from the normal distribution with mean $\boldsymbol{\theta}_0$ and standard deviation $\boldsymbol{\sigma}$, equal to the square root of diagonal elements of the inverse Hessian matrix ($\boldsymbol{\sigma}^2 = \mathrm{diag}(H^{-1}(\boldsymbol{\theta}_0))$). For each identical set of parameters, the potential variation in estimated parameter values was assessed. 95% CIs for R(t) were obtained by taking the 2.5th and 97.5th percentiles of the simulated distributions. The 95% CI for incidence was also computed using parametric bootstrapping.

In order to validate our method, our estimates of R(t) were compared with ones from a conventional renewal equation model, in which the distributions of generation intervals and daily incidence obtained by back-calculation are convoluted. The incidence of infection was originally back-calculated using the non-parametric back-projection method [Becker, Watson and Carlin, 1991]. Specifically, the expected value of j(t) was given by the following renewal process model:

$$E(j(t)) = R(t)\int_0^\infty j(t-\tau)g(\tau)\frac{F(T-t)}{F(T-t+\tau)}dt, \qquad (7)$$

where $j(t)$ is the back-calculated incidence of infection, $g(\tau)$ is the probability density function of the generation interval, $F(.)$ is the cumulative distribution function of the time delay from infection to reporting, and T is the latest calendar time of observation. Although adjustment for reporting delay was added to the renewal equation (7), be noted that the structure is in principle comparable to the commonly used renewal equation, as well as Equation (2). Using Equations (5) and (7), we could obtain two different sets of *R(t)* estimates and overlaid with epidemic curves to see their responsiveness to the implemented PHSMs. The coverage of the proposed Equation (5) in terms of maximum likelihood estimates from Equation (7) was calculated to assess the similarity between these equations throughout the epidemic.

3. Impact of PHSMs on R(t)

To understand the importance of interventions, three different types of analyses were carried out. First, the negative log likelihood values of all possible combinations of "event-based" models using step functions to approximate R(t) were compared to account for the loss of single or combination of events; the negative log likelihood value of "full" model that accounted for all event dates affecting R(t) on days 31, 39, 43, 46, 94, 162, 165, 186 and 196 was computed. Among these dates, PHSMs were implemented on days 31, 39, 43 and 46 (first wave) and days 162 and 165 (second wave). In the first wave, 15 possible combinations of four dates (days 31, 39, 43 and 46) were considered as alternative models. In the second wave, three combinations (i.e., on–off, off–on, and off–off models) of two dates (days 162 and 165) were considered as

alternative models. We performed likelihood ratio tests and Bonferroni correction of the global alpha level to account for multiple comparisons. Second, we used a joinpoint segmented regression model to assess whether a significant change in R(t) during each wave was associated with any of the start dates of key interventions [Joinpoint Regression Program, 2020]. By employing permutation tests for model selection, inflection points ("joinpoints") within trends over a specific period can be identified. Third, the relative reduction in the effective reproduction number associated with these dates was estimated. We estimated the relative reproduction number following each event date.

All statistical data were analyzed using R version 4.0.3 [R Foundation for Statistical Computing, 2020]. The R code used for this analysis is in Appendix.

4. Ethical consideration

The present study used publicly available data, and thus, did not require ethical approval.

Results

1. Epidemic curve and model prediction

The epidemic curve as a function of the date of illness onset is shown in Figure 1. Two distinct epidemic waves, the first occurring from February to May, 2020 and the other from June to October, 2020 and onwards, were observed. The first wave had the highest daily incidence of 69 cases on 1 and 3 April, and the second wave had the highest daily incidence of 186 cases on 29 July. The date of reporting was delayed by 5–7 days on average from the actual real-time epidemic data due to delays in diagnosis and reporting. The second wave has not smoothly declined over time: the rate of decline stagnated from late August, and a second hump with peak daily incidence of 86 cases was observed on 6 September 2020.

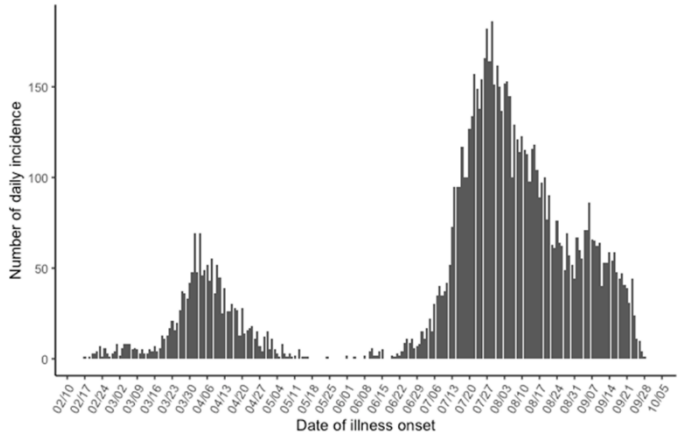

Figure 1. Daily numbers of new coronavirus disease 2019 (COVID-19) cases in Osaka prefecture from 17 February to 28 September 2020. Cases were counted as a function of the date of illness onset.

Our model predictions and associated 95% CIs using the bootstrap method against the observed epidemic curve are shown in Figure 2. The overall observed pattern of the epidemic was qualitatively effectively reproduced by our model. The cumulative number of cases of 8520 by the end of the investigation period was estimated by our model, whereas 8818 cases were actually observed. Bootstrap method showed that the 95% CI for the total epidemic size ranged from 6367 to 11,066 cases.

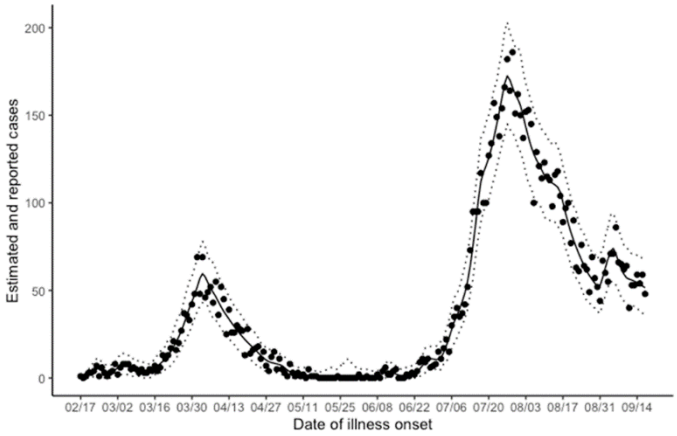

Figure 2. Comparison between the observed and model-predicted incidence of coronavirus disease 2019 (COVID-19) in Osaka prefecture. Comparisons were made as a function of the date of illness onset. Solid circles represent observed cases, while the continuous black line shows the maximum likelihood estimate of predicted incidence. Dotted lines represent the lower and upper boundaries of the 95% confidence intervals based on the bootstrap method.

2. Estimated R(t)

Figure 3A shows the maximum likelihood estimates of R(t) by our method. Our estimates showed that during the first wave, the R(t) reached <1 (0.8; 95% CI: 0.6–1.0) for the first time on 2 April 2020 and remained <1 until 26 May, indicating that the epidemic was under control during this period. It was subsequently shown that the R(t) had an explosive oscillating pattern and increased until 26 July, when it again dropped to <1 (0.9; 95% CI: 0.8–1.0). Very consistent patterns were revealed when our estimates were compared with the R(t) estimates based on the back-projected incidence of infection. The daily R(t) based on the estimated incidence of infection was contained within the 95% CI of our estimate for 141 days (68.1% of 207 days as the observation period).

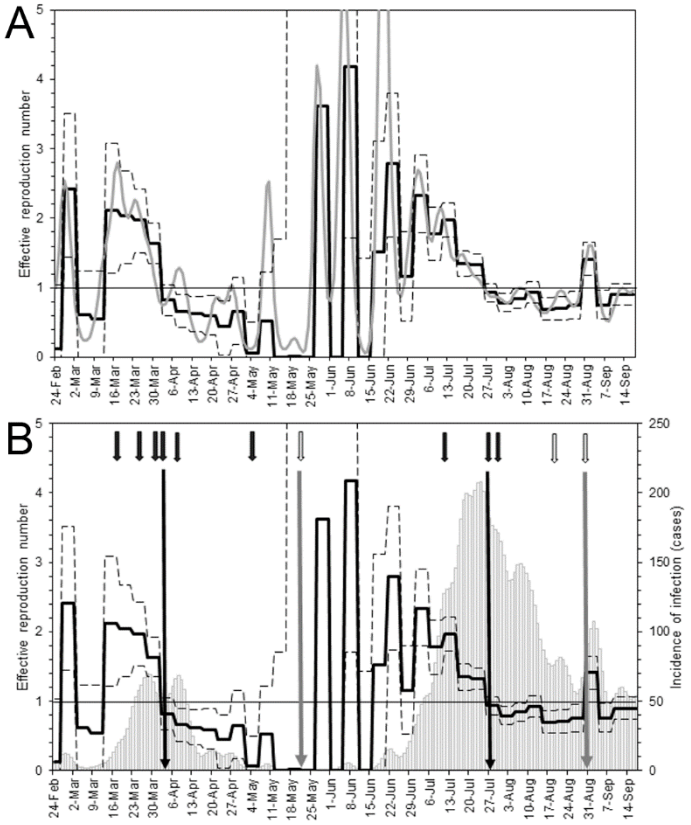

Figure 3. Estimation of the effective reproduction number, R(t), using two methods. (A) Comparison of R(t) estimates calculated using a novel method based on the date of illness onset (continuous black line) and using the existing method based on non-parametrically back-projected incidence of infection. Dashed lines show the 95% CIs of R(t) based on the bootstrap method. The horizontal gray line indicates R(t) = 1; below this value incidence declines. (B) Chronological relationship between announcements in Osaka and R(t). Black arrows represent announcements of requests to reduce contacts or any other announcements associated with infection control. White arrows represent announcements of the cessation of specific countermeasures. The overlaid bar chart shows incidence by the estimated date of infection.

The relationship between R(t) dynamics and the start and end dates of key PHSMs implemented throughout the epidemic is shown in Figure 3B. R(t) was shown to respond well to the implementation of each PHSM. The governor in Osaka prefecture requested successive voluntary restrictions starting on 19 March 2020 during the first wave. The R(t) was then shown to decrease step-by-step in response to these interventions. He requested voluntary restrictions on weekend outings for the second time on 3 April, and subsequently the R(t) declined to <1. The R(t) showed multiple humps 6 days after Osaka was released from the state of emergency on 21 May, indicating the beginning of the second wave. The R(t) was shown to decline to <1 again from 26 July, during which the governor of Osaka requested time voluntary restrictions on social events involving the consumption of alcohol by five or more persons. The R(t) sporadically reached >1 soon after this request was ended on 31 August.

As a sensitivity analysis, when employing 4-day and 3-day period models, the R(t) reached <1 for the first time on 1 and 2 April 2020, respectively, in the first wave. For the second wave, using the 4-day and 3-day period models, the R(t) declined to <1 from 26 and 28 July, respectively.

3. Impact of PHSMs on R(t)

Figure 4 compares models using 5-day period estimates of R(t) against event-based models of R(t). Estimates of R(t) could be qualitatively captured by the full model, except on dates with limited numbers of cases and broad uncertainty bounds of R(t). Table 2 and 3 show the results of multiple model comparisons. For the first wave, the full model was compared with fifteen different models. For nine models to which full model fitted better, two or more event dates were removed, along with either or both of day 43 and day 46. The model fit was significantly poorer (likelihood ratio test p-value = 0.0006) when four important events during the first wave were ignored. It was shown that models that removed only a single event date did not differ significantly from the full model. Our conclusions were not altered when using Holm's method rather than Bonferroni correction. Days 43 and 46 were identified as inflection points marking changes in the trend of R(t) by Joinpoint analysis (Table 2 and Figure 5). The relative reproduction numbers following days 31, 39, 43 and 46 compared with values prior to day 31 were 1.56 (95% CI: 0.83, 2.81), 0.86 (95% CI: 0.33, 2.01), 0.79 (95% CI: 0.21, 2.63) and 0.32 (95% CI: 0.12, 1.09), respectively. The upper boundaries of the 95% CIs of all single-day estimates were found to exceed 1 though the estimates of the single-day effect for days 39, 43 and 46 were all below 1. We estimated the relative reproduction number at 0.34 (95% CI: 0.19, 0.59) when combining the effects of all four events.

In the second wave, two events dates (day 162 and 165) were involved. Significant differences compared with the full model were not produced when removing either or both events (Table 3). We could not identify neither of these two event dates as a joinpoint (Table 3 and Figure 6). The relative reproduction numbers following days 162 and 165 compared with values prior to day 162 were 0.57 (95% CI: 0.13, 1.54) and 0.87 (95% CI: 0.28, not calculable), respectively. For the second wave, it was shown that the estimates of the single-day effect were below 1, but the upper boundaries of the 95% CIs exceeded 1 or were not calculable. The relative reproduction number was 0.49 (95% CI: 0.28, 0.79) when combining the effects of these two dates.

Table 2. Multiple comparisons of the effective reproduction number accounting for the absence of dates of specific event of interventions and identification of joinpoints during the first wave of coronavirus disease 2019 in Osaka. *Statistically significant following Bonferroni correction ($p < 0.00333$). 0 indicates that the corresponding model ignored the corresponding date, whereas 1 indicates that the date was taken into account. P-values show the results from likelihood ratio tests compared against the full model. Y/N indicates whether a joinpoint was identified or not within the 5-day period including the corresponding day.

Model	Day 31	Day 39	Day 43	Day 46	p-value	
1	0	1	1	1	0.24059	
2	1	0	1	1	0.76617	
3	1	1	0	1	0.67850	
4	1	1	1	0	0.15528	
5	0	0	1	1	0.36975	
6	0	1	0	1	0.40291	
7	0	1	1	0	0.00000	*
8	1	0	0	1	0.00000	*
9	1	0	1	0	0.00000	*
10	1	1	0	0	0.00000	*
11	0	0	0	1	0.00000	*
12	0	0	1	0	0.00000	*
13	0	1	0	0	0.00000	*
14	1	0	0	0	0.00000	*
15	0	0	0	0	0.00059	*
joinpoint	N	N	Y	Y		

Table 3. Multiple comparisons of the effective reproduction number accounting for the absence of dates of specific event of interventions and identification of joinpoints during the second wave of coronavirus disease 2019 in Osaka.

Model	Day 162	Day 165	p-value
1	0	1	0.45741
2	1	0	0.66792
3	0	0	0.86868
joinpoint	N	N	

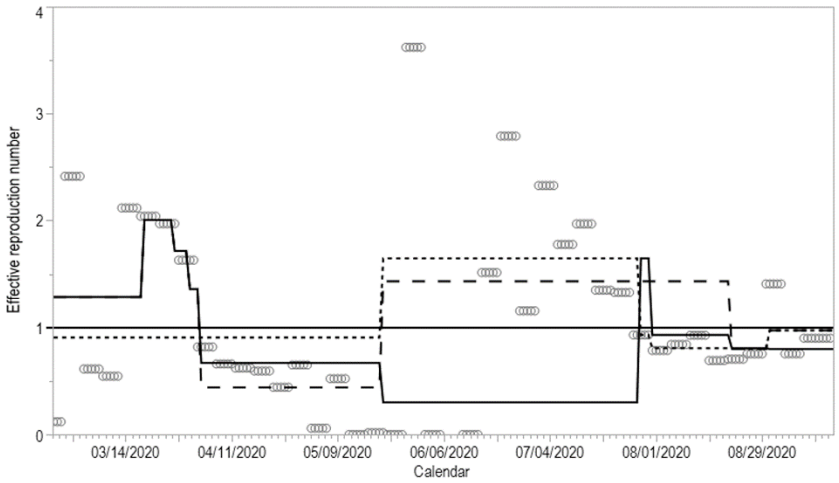

Figure 4. Comparison of event-based models of the effective reproduction number. Unfilled circles represent estimates based on the 5-day piecewise constant model. Solid line shows the full event-based model accounting for all changes in $R(t)$ on Day 31, 39, 43, 46, 94, 162, 165, 186 and 196. The dotted line represents an alternative model ignoring any intervention dates (i.e., Day 31, 39, 43 and 46) during the first wave. The dashed line represents another alternative model ignoring intervention dates (Day 162 and 165) during the second wave.

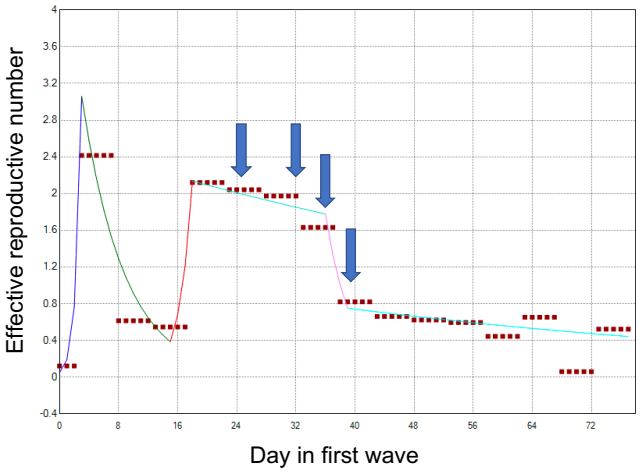

Figure 5. Joinpoint analysis during the first wave of coronavirus disease 2019 in Osaka. Closed squares indicate effective the reproductive number. The solid line indicates the regression line linking identified joinpoints. Black arrows represent starting days of the four key interventions during the first wave. Day 0 corresponds to 24 February 2020.

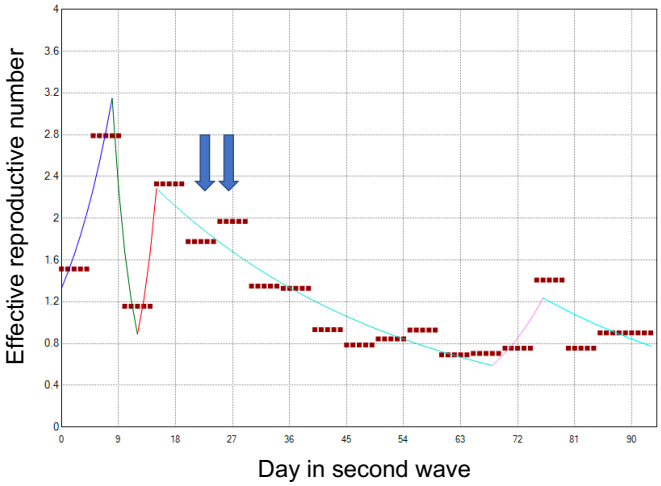

Figure 6. Joinpoint analysis during the second wave of coronavirus disease 2019 in Osaka**.** Black arrows represent starting days of two key interventions during the second wave. Day 0 corresponds to 16 June 2020.

Discussion

Here we proposed an alternative method for computing the R(t) of COVID-19. This method was applied to the epidemic in Osaka prefecture from February to September 2020. A modified renewal equation based on the date of illness onset was employed, leveraging the frequency of secondary transmission to capture the pre-symptomatic transmission of COVID-19. It was shown that using a piecewise constant model with a 5-day time interval to estimate R(t), the epidemic came under control around 2 April during the first wave and around 26 July during the second wave. We could also show that the estimates from the alternative method agreed well with those from an established method using the back-calculated incidence of infection. R(t), described by the date of infection, was demonstrated not to decline drastically following any single PSHM event. When multiple interventions were combined, the relative reductions in R(t) during the first and second waves were 70 and 51%, respectively. During the second wave, a significant impact on the observed secondary transmission patterns was not produced by the combined effect of interventions that focus on high-risk groups. This exercise using epidemiological data in Osaka prefecture suggests that estimates of R(t) calculated using our modified renewal equation can be used to assess the effectiveness of PHSMs in a retrospective manner.

R(t) as a function of time of infection is essential to evaluate intervention programs [Flaxman et al., 2020; Li et al., 2020]. Non-parametric back-projection of the time of infection using the observed time of illness onset has so far been required in such evaluations [Li et al., 2020]. There are, however, some caveats in using non-parametric

back-projection method because it is not a simple procedure, sometimes involving expectation-maximization techniques and smoothing [Becker, Watson and Carlin, 1991], and using two different inferential steps (estimation of the epidemic curve by the time of infection and subsequent estimation of R(t)) is not undesirable. The advantage of our method, compared with the conventional method, is the simplicity of the estimation procedure while accounting for the pre-symptomatic transmission of COVID-19. We decomposed the data generation process into two components: (i) the relative frequency of secondary transmission with respect to disease-age, and (ii) the probability density function of the incubation period. In other words, the generation time is decomposed into two components (i.e., (i) and (ii)). Then, by taking the double integral of the modified renewal equation, R(t) as a function of the time of infection can be estimated only by using the time of illness onset data. It must be noted that the R(t) estimates were consistent with those from conventional methods. To avoid complex smoothing procedures, we employed a step function for R(t) and varied the period of time-step as a sensitivity analysis. This function can be improved more flexibly; by employing a spline function, far smoother estimates of R(t) than ours can be obtained. Other infectious disease datasets can be applicable to our method if the two functions, i.e., the probability density function of the incubation period and the relative frequency of secondary transmission with respect to disease-age, are both known and quantified. Smallpox is among such examples [Nakajo and Nishiura, submitted].

The objective monitoring of epidemiological dynamics in real-time and the assessment of the effectiveness of PHSMs are the pragmatic values that R(t) can provide. There has been a technical debate over preferred methods for the estimation of R(t) depending on practical public health needs, including real-time or short-term monitoring of epidemics, forecasting of future trends, and assessing the impacts of interventions [Gostic et al. 2020; Lipsitch et al., 2020; Petermann and Wyler, 2020]. Our results showed that the R(t) did not abruptly decline following single PHSM events and the combined effects of multiple interventions reduced the R(t) to <1. During the first wave, the combined effect by various forms of self-restraint in reducing contacts was to reduce the R(t) by 70%; when individuals were asked to abstain from night life and reduce contacts in high-risk settings (e.g., eating and drinking behaviors), the R(t) declined to <1. A new trend of reduced R(t) following the implementation of these two interventions was also shown by joinpoint analysis. During the second wave, is was shown that the R(t) declined to <1 and the combined effect of interventions was to reduce R(t) by 50%. The model fitting was, however, not significantly altered by removing intervention dates during the second wave in our model comparison. Joinpoint analysis also showed that none of the

starting dates of interventions were identified as inflection points of trends in the R(t), which was in contrast with the findings for the first wave. These findings imply that the impact of interventions, including requests for self-restraint in avoiding eating and drinking with five or more people, on the observed transmission dynamics was limited. Humped patterns of R(t) with significant variance soon after the ending of major restrictions on 21 May 2020 were revealed by our model, indicating that R(t) estimates could reflect the cessation of interventions including requests for self-restraint behaviors. It was shown that infections did not continue to increase on and after the hump, possibly suggesting that even after the end of restrictions in restaurants, self-restraint in reducing contact behaviors persisted until the end of our study period.

We can use several different statistical techniques to assess the relationships between intervention timing and R(t). To name a few, these include (i) permutation methods (e.g., synthetic control methods that have recently been applied to COVID-19 [Kendall et al., 2020]), (ii) time series methods such as vector autoregression models, and (iii) model comparisons using penalized log likelihood values. The option (iii) was chosen because 1) we have the limited number of important events to be examined and 2) we find it not feasible to select appropriate controls in other prefectures where single PHSMs or combinations of PHSMs were implemented during the period of our analysis. There are several advantages of model comparison, including the ability to account for the degree of freedom (i.e., the number of parameters to be estimated) and the capacity to examine if events are essential to describe epidemic dynamics. On the other hand, if there are too many candidate dates on which trends changed, model comparison cannot be applied, and permutation would be preferred. In addition, causal links between time events and transmission cannot be examined by model comparison and vector autoregression may be preferred for this objective. Model comparison enabled us to demonstrate that combinations of interventions, not single interventions, during the first wave were critical in describing transmission dynamics. On the other hand, results from model comparison showed that intervention dates during the second wave were not significant in describing transmission dynamics, perhaps because of the reduced level of R(t) during the second wave. We estimated the combined effect in reducing R(t) during the second wave as the product of the effects of each PHSM event, which was as large as a 50% reduction. On the other hand, following correction for multiple testing, the effect of combined interventions during the second wave was not found to be significant. Significant changes in reduced R(t) corresponding to the beginnings of interventions during the second wave were not identified by joinpoint analysis. Taken together, infection control policy relying on single interventions that

focus on high-risk groups may not be efficient.

Several limitations should be noted. First, we ignored the heterogeneity of transmission in our model. Age has been suggested to be a modifying variable in terms of secondary transmission [Ayoub et al., 2020; Davies et al., 2020]. To characterize the COVID-19 epidemic more accurately, we should explore the impact of age structure or occupation, or both, on R(t). Second, R(t) for recent calendar time-points could not be estimated, which is the case for other methods. Both reporting delays and the natural history of the disease are related to this limitation. Our method assumed that pre-symptomatic transmission could occur 12 days prior to illness onset at the earliest [He et al., 2020]. So, R(t) for the most recent 12 days could not be estimated and this may be an issue for assessing the epidemic in real-time. Third, stochasticity during the transmission process was not considered in our model. Forth, we had to assume that transmission potential relative to disease-age was independent of calendar time, which may be a simple assumption. Ali et al. has shown that the serial interval of COVID-19 decreased throughout the epidemic [Ali et al., 2020], indicating that the assumption of a stable distribution for the serial interval may not be valid. The new variant could modify our original assumption as well. The proportion of symptomatic transmission could also be modified by interventions targeting manifestations of symptoms (e.g., case isolation and contact tracing) over the course of the epidemic. This will be investigated in an exploratory analysis, Section 2. Lastly but not the least, our model based on data of illness onset could not take into account for asymptomatic infection (i.e., transmission from an individual who never had symptoms). The magnitude of their transmission relative to symptomatic infection will be extensively studied in Section 2.

Section 2 – Study on asymptomatic infection of COVID-19

Introduction

Since the early stages of the pandemic of COVID-19, public health has been paying attention to individuals with asymptomatic SARS-CoV-2 infection (i.e., individuals who never develop symptoms throughout the course of infection) [Bai et al., 2020; Nishiura et al., 2020; Oran et al., 2020]. Nishiura et al. reported that the early estimate of asymptomatic ratio has been as ranging from 8 to 54% [Nishiura et al., 2020]. In addition, more than half of secondary transmissions with coronavirus disease 2019 (COVID-19) are reported to occur from asymptomatically infected people [Johansson et al., 2021]. It is critical to elucidate the transmissibility of asymptomatic infections for successful control of COVID-19 because these individuals can propagate the virus unknowingly. If transmissibility is substantial and asymptomatic infections are frequent, we might find controlling the epidemic via screening of symptomatic cases not be an effective strategy. On the other hand, if the transmissibility of asymptomatic cases is limited, limited resources of health authorities can be allocated to tracing primary symptomatic cases to bring the epidemic under control [Johansson et al., 2021; Koo et al., 2020; Kucharski et al., 2020].

There still exists uncertainty in the transmissibility of asymptomatic infections [Johansson et al., 2021; McEvoy et al., 2020; The Royal Society, 2020]. Contradictory findings have been gained from previous studies [Chaw et al., 2020; He et al., 2020; Kimball et al., 2020; Lee et al., 2020; McEvoy et al., 2020; Sayampanathan et al., 2021; Zou et al., 2020]. Retrospective study of 303 symptomatic and asymptomatic patients in a community treatment center in Korea showed that the viral loads of asymptomatic patients were similar to those of symptomatic patients [Lee et al., 2020]. This notion was supported by other studies leveraging viral load as a surrogate of transmissibility [Kimball et al., 2020; Zou et al., 2020]. In two epidemiological studies conducted in Singapore and Brunei, the incidence rate ratio (asymptomatic vs. symptomatic cases) and attack rate ratio (asymptomatic and pre-symptomatic vs. symptomatic cases) were both shown to be below 1 (0.24 and 0.78, respectively) and it was suggested that asymptomatic infections may be less transmissible than symptomatic infections [Chaw et al., 2020; Sayampanathan et al., 2021]. Analyzing transmission data in Ningbo from January 21 to March 6, 2020 [Chen et al., 2020], He et al. estimated the reproduction numbers of asymptomatic and symptomatic cases as 0.20 and 0.78, respectively, indicating that the relative transmissibility of asymptomatic cases was below 1 [He et al., 2020].

We should strengthen these findings and extend to other populations and age groups (e.g., young adults). In addition, we find it unclear whether transmission from asymptomatic cases is more likely to lead to asymptomatic infections. In Japan, cluster-based approach to COVID-19 Japan has been adopted [Oshitani et al., 2020], focusing on contact tracing in high-risk setting that satisfies three C's condition (i.e., close contact in confined and crowded space), granting us an opportunity to analyze the well-observed cluster data. Here, we aimed to analyze two distinct clusters of SARS-CoV-2 infections in Japan. The primary objective in Section 2 is to understand the transmissibility of asymptomatic SARS-CoV-2 infections, which is the key parameter governing the transmission dynamics of COVID-19 but Section 1 discussing on R(t) could not investigate at all. We also try to examine the potential role of symptomatology in giving rise to secondary symptomatic vs. asymptomatic infections.

Methods

1. Epidemiological Data

The first-identified two clusters of SARS-CoV-2 infections were analyzed for this investigation. Indoor environments were identified as focal areas of transmission by cluster-based approaches in Japan. All close contacts of confirmed cases that could be identified retrospectively were brought under observation and subjected to laboratory testing [Furuse et al., 2020; Oshitani et al., 2020]. That means that clusters from January to March 2020 were extremely well traced, and it should be also noted that incidence did not exceed contact tracing capacity. Both restriction of movement for 14 days and laboratory testing by polymerase chain reaction to confirm SARS-CoV-2 infection were requested for all identified close contacts. Collected information included transmission networks, age, symptomatic status (i.e., manifested any symptoms by the end of isolation vs. never manifested symptoms) and time of illness onset (for symptomatic cases only).

1-1: Tokyo and Kanagawa cluster

The cluster occurred in Tokyo in January 2020 and involved a total of 36 confirmed cases. The transmission event started with a night party on January 18, 2020 in Tokyo that stemmed from an exposure to tourist from Wuhan. Transmission events, subsequently, took place either at facilities or households in Tokyo and Kanagawa prefectures through end of January to early March 2020. The epidemic curve is shown in Figure 7.

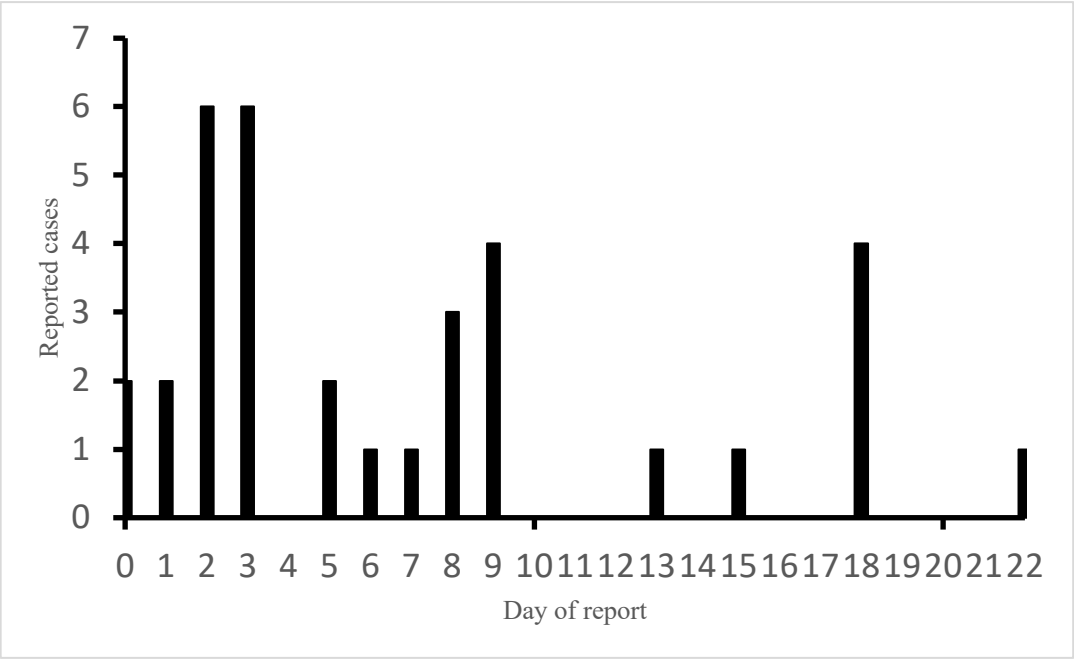

Figure 7. Epidemic curve in a cluster of SARS-CoV-2 infections in Tokyo and Kanagawa, Japan. Daily counts of confirmed cases are shown as a function of the day of report.

1-2: Kyoto cluster

The cluster occurred in Kyoto prefecture, Japan, in March 2020 and involved a total of 74 confirmed cases, majority of which were university students. It was revealed that the three index cases traveled to Europe in early March to celebrate their graduation from university, before returning to Japan on 14 March 2020. Transmission events took place during three parties on the nights of 19, 21, and 22 March in Kyoto and these parties were attended by the index cases independently. One of the secondary cases, having been infected at one of the parties, contributed to subsequent transmission events during another party on 23 March. The first confirmed case, which later turned out to be one of the index cases, was reported on 26 March in Ehime prefecture, southwest of Kyoto. Additional cases were notified in other prefectures as well as in Kyoto, forcing local public health centers to start contact tracing on 29 March. The epidemic curve is shown in Figure 8.

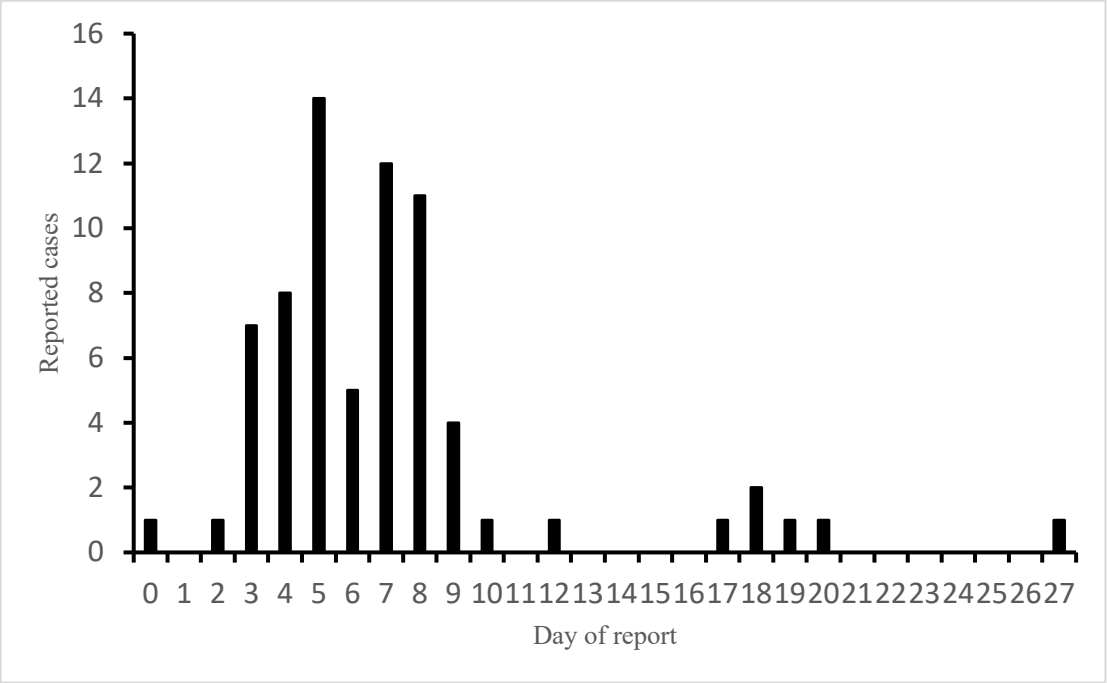

Figure 8. Epidemic curve in a cluster of SARS-CoV-2 infections among university students in Kyoto, Japan. Daily counts of confirmed cases are shown as a function of the day of report. The recognition of the cluster was notified on day 3.

2. Estimation of relative reproduction number of asymptomatic cases

SARS-CoV-2 transmissibility and its dependence on symptomatology were explored using discrete and two-type branching process models. Conventionally, offspring distributions of SARS-CoV-2 infections, characterized by superspreading events, have been modeled using the negative binomial distribution [Adam et al., 2020; Endo et al., 2020; Lau et al., 2020]. Adhering to this custom, we assumed that the number of secondary cases arising from either primary symptomatic or asymptomatic cases independently followed negative binomial distributions with means R_s or R_a, respectively, and common dispersions k. Let us set v as the relative reproduction number of asymptomatic cases compared with symptomatic cases such that $R_a = vR_s$. Let D represent observed data (with total sample size n in each cluster), then the likelihood of observing the number of secondary cases can be described as:

$$L(v, R_s, k \mid D) = \prod_i^n p(r_i | v, R_s, k), \quad (1)$$

where r_i denotes the observed number of secondary cases arising from primary case i and $p(.)$ represents the probability mass function of the negative binomial distribution.

The negative log-likelihood of Eq. (1) was minimized. The 95% confidence intervals (CIs) of these parameters were obtained from profile likelihood. A Poisson or geometric distribution was also employed as an alternative likelihood function. For Kyoto cluster, an alternative scenario was assumed in which an exponential decrease in the reproduction number as a function of calendar time and $R_{t=0}$, the reproduction number of symptomatic cases at calendar time of zero, was estimated [Chan and Nishiura, 2020]. In this scenario, it was assumed that the time of transmission for primary symptomatic and asymptomatic cases was the time of illness onset and the time of exposure, respectively.

3. Estimation of proportions of secondary symptomatic cases produced by primary symptomatic and asymptomatic cases

The potential role of symptomatic transmission in producing symptomatic secondary infections was explored. It was assumed that the number of symptomatic secondary cases among all secondary cases followed a binomial distribution with parameters p and q, representing the proportions of secondary symptomatic cases produced by primary symptomatic and asymptomatic cases, respectively. These parameters were estimated jointly with R_s , v and k

4. Estimation of parameters for serial interval

In analysis on Kyoto cluster, we estimated parameters governing the probability distribution function of transmissibility relative to illness onset. It was assumed that the probability followed a gamma distribution according to He et al. [He et al., 2020]. We can model the probability distribution function of the serial interval, $s(\tau)$, by convolution as:

$$s(\tau) = \int_{-m}^{\tau} h(\sigma) f(\tau - \sigma) d\sigma, \tag{2}$$

where $h(\tau)$ and $f(\tau)$ are probability density functions of the relative frequencies of secondary transmission with respect to time since illness onset and the incubation period, respectively, and m represents the start of infectiousness relative to illness onset. For probability density function of the incubation period, a lognormal distribution with a mean of 5.2 days, estimated using data from 425 patients in Wuhan, China [Li et al., 2020], was used. We assumed the value of m as 6 days based on the shortest observed serial interval in the transmission network in the cluster. Given the total of w observations of serial intervals for secondary cases j, we can describe the likelihood of observing the serial intervals τ_j as:

$$L(\theta|D) = \prod_{j}^{w} s(\tau_j|\theta), \quad (3)$$

where θ is the vector of the parameters (e.g., the shape and rate parameters) of the gamma distribution. We included total 18 pairs of symptomatic primary cases and secondary cases with dates of illness onset available in the analysis.

5. Estimation of the relative risk of secondary transmission among isolated individuals

For Kyoto cluster, we complemented our investigation of the transmissibility of asymptomatic cases, exploring the impact of isolation on these estimates using the probability distribution function of the serial interval shortened by isolation [Chan and Nishiura, 2020]. A parameter, ε, denoting the relative risk of secondary transmission among isolated individuals, was formulated similarly to a previous study [Chan and Nishiura, 2020]. ε was estimated jointly with R_s, v and k.

First step is to derive the probability density function of the serial interval shortened by isolation and expected number of secondary cases. The renewal equation was employed as a basis of this approach [Nishiura and Chowell, 2009]:

$$j(t) = \int_0^\infty A(\tau) j(t-\tau) d\tau, \quad (4)$$

where $j(t)$ is the incidence at calendar time t and $A(\tau)$ is the rate of secondary transmission per case at infection age τ. It should be noted that $g(\tau)$, the probability density function of the generation interval (i.e., the time from infection of a primary case to the infection of a secondary case by the primary case), is given by:

$$g(\tau) = \frac{A(\tau)}{\int_0^\infty A(x)dx}. \quad (5)$$

Differences in the transmissibility profiles of symptomatic cases and asymptomatic cases are interesting for our investigation. For a symptomatic individual, $\lambda(\tau)$, the probability density function of the interval from his/her illness onset to transmission to a secondary case was used to capture information on transmissibility relative to the time since illness onset, including before the time of illness onset. In Kyoto cluster, all cases were immediately isolated after laboratory confirmation, and thus it was assumed that the transmissibility profile for each symptomatic case was altered following isolation. Let us denote ε as the relative reduction in the rate of secondary transmission in an isolated individual, similarly to a previous study [Chan and Nishiura, 2020]. ε=0 indicates isolation has no effect on transmission profile. We can then describe the probability density function of infectiousness relative to illness onset with isolation,

$h(\tau)$, as:

$$h(\tau) = \begin{cases} \dfrac{\lambda(\tau)}{\int_{-m}^{k} \lambda(s)ds + (1-\varepsilon)\int_{k}^{\infty} \lambda(s)ds} & \text{for } \tau \leq k \\ \dfrac{(1-\varepsilon)\lambda(\tau)}{\int_{-m}^{k} \lambda(s)ds + (1-\varepsilon)\int_{k}^{\infty} \lambda(s)ds} & \text{for } \tau > k \end{cases}, \quad (6)$$

where k and m represent the time of isolation and the beginning of infectiousness relative to illness onset, respectively. Then, the probability density function of the observed serial interval will be given as a convolution of two probability density functions, $h(\tau)$ and $f(\tau)$; the latter is the probability density function of the incubation period,

$$s(\tau) = \int_{-m}^{\tau} h(\sigma) f(\tau - \sigma) d\sigma. \quad (7)$$

We can then describe R_s, the expected number of secondary cases generated by a primary symptomatic case under isolation, as:

$$R_s = R_{0,s}\left(\int_{-m}^{x} \lambda(\tau)d\tau + (1-\varepsilon)\int_{x}^{\infty} \lambda(\tau)d\tau\right), \quad (8)$$

where $R_{0,s}$ indicates the basic reproduction number for a primary symptomatic case and x is the disease age (i.e., time since illness onset) when primary cases are isolated. For a primary asymptomatic case (i.e., a case who never manifested symptoms during the observation period), the probability density function of the generation interval, $g(\tau)$, was substituted for $\lambda(\tau)$ in Eq. (8). $g(\tau)$ is modeled as the convolution of two probability density functions, $\lambda(\tau)$ and $f(\tau)$, as:

$$g(\tau) = \int_{0}^{\tau} f(\sigma)\lambda(\tau - \sigma) d\sigma. \quad (9)$$

Assuming the reduction of transmissibility resulting from isolation is the same as for symptomatic cases, R_a, the expected number of secondary cases generated by a primary asymptomatic case under isolation can be described as:

$$R_a = R_{0,a}\left(\int_{0}^{x} g(\tau)d\tau + (1-\varepsilon)\int_{x}^{\infty} g(\tau)d\tau\right), \quad (10)$$

where $R_{0,a}$ and x indicate the basic reproduction number for a primary asymptomatic case and the time of isolation, respectively. Note that time zero in the integral range indicates the time of exposure. Let us denote v as the relative reduction in the basic reproduction number for an asymptomatic case with respective to a symptomatic case. Then, equation (10) can be rewritten as:

$$R_a = vR_{0,s}\left(\int_0^l g(\tau)d\tau + (1-\varepsilon)\int_l^\infty g(\tau)d\tau\right), \tag{11}$$

where l indicates the time of isolation. The datasets used here were reported in a discrete time interval (days) so we discretized Eqs. (7), (8) and (11) as:

$$s_\tau = \sum_{s=1}^{m+\tau+1} h_{s+m+1} f_{\tau-s-m-1}, \tag{12}$$

$$R_s = R_{0,s}\left(\sum_{\tau=1}^{k+m+1} \lambda_\tau + (1-\varepsilon)\sum_{\tau=k+m+2}^{T} \lambda_\tau\right), \tag{13}$$

and

$$R_a = vR_{0,s}\left(\sum_{\tau=1}^{l+m+1} g_\tau + (1-\varepsilon)\sum_{\tau=l+m+2}^{T} g_\tau\right), \tag{14}$$

where h_τ, f_τ, λ_τ and g_τ are the probability mass functions corresponding to $h(\tau), f(\tau), \lambda(\tau)$ and $g(\tau)$, respectively, and $g_\tau = \sum_{s=1}^{\tau-1} h_{\tau-s}\lambda_s$.

A maximum-likelihood method to estimate parameters was used. The likelihood function, $L(\theta|D)$, consists of two parts, the observed length of the serial interval, $L_1(\theta|D)$, and the observed number of secondary cases, $L_2(\theta|D)$:

$$L(\theta|D) = L_1(\theta|D)L_2(\theta|D). \tag{15}$$

Let D_1 be the number of datasets of observed serial intervals. Then, the first likelihood is written as:

$$L_1(\varepsilon, \boldsymbol{\alpha}|D) = \prod_i^{D_1} s(t_i|\varepsilon, \boldsymbol{\alpha}), \tag{16}$$

where $\boldsymbol{\alpha}$ is the vector of parameters for $\lambda(\tau)$ and t_i is the observed serial interval. It was assumed that $\lambda(\tau)$ follows a gamma distribution [He et al., 2020]. For $f(\tau)$, we used a lognormal distribution with a mean of 5.2 days estimated based on data from 425 patients in Wuhan [Li et al., 2020]. The value of m was set as 6 days based on the observed smallest serial interval in our transmission network. It was assumed that the observed number of secondary cases derived from either a primary symptomatic or asymptomatic case will follow a negative binomial distribution with R_s or R_a as the mean, respectively. Let D_2 be the number of datasets for observed secondary cases. Then, the second likelihood can be described as:

$$L_2(\varepsilon, v, R_{0,s}, \boldsymbol{\alpha}, k|D) = \prod_i^{D_2} p(r_i|\varepsilon, v, R_{0,s}, \boldsymbol{\alpha}, k) = \prod_i^{D_2} NB\left(r_i;\ k, \frac{k}{R_- + k}\right), \quad (17)$$

where R_- is R_s or R_a, k is the dispersion parameter for the offspring distribution and r_i is the observed number of secondary cases arising from primary case i. We adopted two assumptions for $R_{0,s}$ (i.e., taking a constant value throughout the course of the epidemic vs. the value exponential decreases as a function of calendar time). In the latter scenario where $R(t) = R_{0,s}e^{-\delta t}$, δ, the exponential rate, was jointly estimated with the initial reproduction number $R_{0,s}$. The negative log-likelihood of Eq. (15) was minimized. The 95% CIs for the parameters were obtained by profile likelihood.

All statistical data were analyzed using R version 4.0.3 [R Foundation for Statistical Computing, 2020]. The R code used for this analysis is in Appendix.

6. Ethical consideration

The present study used publicly available data, and thus, did not require ethical approval.

Results

1. Reconstructed chains of transmission in the clusters

Figure 9 and 10 show reconstructed chains of transmission in the cluster in Tokyo and Kanagawa and the one in Kyoto, respectively. In Tokyo and Kanagawa cluster, it was found that except for household transmission events, the majority of cases from this cluster did not contribute to secondary transmission. Chains of clusters in healthcare and welfare facilities and a household in remote area in a neighboring prefecture Kanagawa were produced by one of the tertiary cases contributed to subsequent, causing the first fatal case in Japan. It was found that the second case in Kanagawa was unlinked to clusters of cases in Tokyo while there were no other cases observed in the relatively remote area in Kanagawa. So, it was highly probable that the case was a part of the very first cluster in Japan. 12 asymptomatic and 24 symptomatic cases were confirmed. The average age of confirmed cases was 60.3 years.

In Kyoto cluster involving 74 cases, we could reconstruct the transmission networks of 64 cases. For the remaining 10 cases, transmission events could not be identified, so we excluded them from subsequent analyses. The network comprised a total of five

generations, with two symptomatic cases acting as “super-spreaders” and giving rise to more than 10 secondary infections (Case 1 and Case 4 in Figure 10). 51 symptomatic and 13 asymptomatic cases were confirmed. Most infections (55%) occurred in individuals aged 20–29 years.

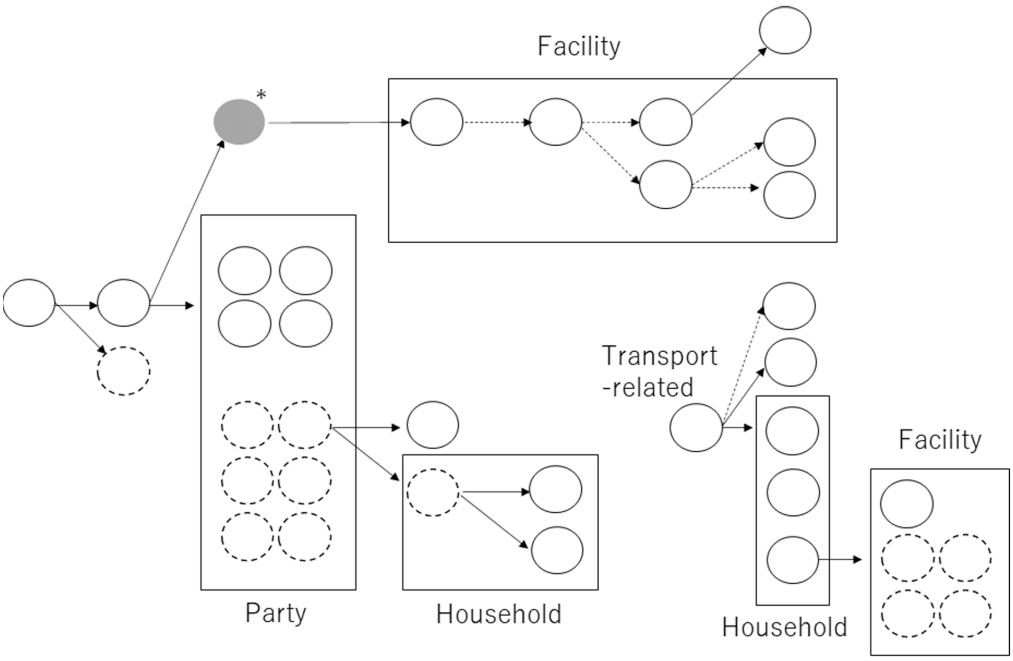

Figure 9. Transmission network of coronavirus disease 2019 (COVID-19) with cluster information in Tokyo and Kanagawa, Japan. Solid and dotted circles indicate symptomatic and asymptomatic cases, respectively. A large frame indicates common indoor environment/facility of transmission. Arrows indicate established link and dotted arrows represent the most plausible link, both based on contact tracing. A grey circle with asterisk-mark (*) indicates a case with unknown symptomatic status, and the transmission event from this case was thus omitted from our analysis. The chains of transmission on the lower right part of figure start with an unlinked case associated with transportation company, and the case was considered to have been associated with undiagnosed case associated with this cluster.

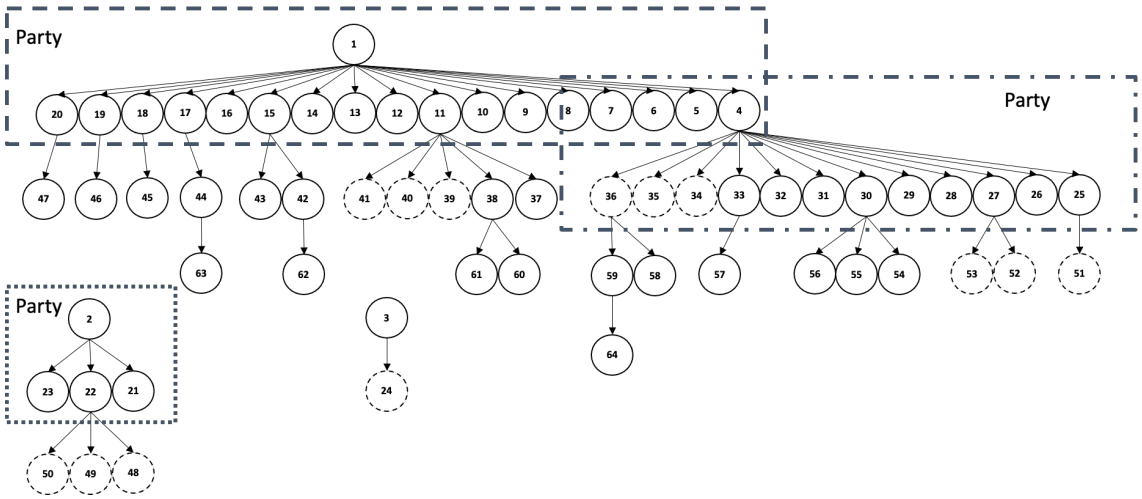

Figure 10. Chains of transmission in a cluster of SARS-CoV-2 infections among university students in Kyoto, Japan. Solid and dotted circles indicate symptomatic and asymptomatic cases, respectively. large frames indicate common indoor environment.

2. Relative reproduction number of asymptomatic cases

Table 4 and 5 show our estimates for the reproduction number and the dispersion parameter for symptomatic cases in Tokyo and Kanagawa cluster and Kyoto cluster, respectively. For Tokyo cluster, R_s , the reproduction number and the dispersion parameter for symptomatic cases were estimated to be 1.2 (95% confidence interval (CI): 0.5–2.9) and 0.23 (95% CI: 0.09–0.62), respectively. Relative infectiousness of asymptomatically infected cases was estimated to be 0.27 (95% CI: 0.03–0.81) of symptomatic cases. It indicates that symptomatic cases appeared to be 3.7 times more infectious than asymptomatic COVID-19. In Kyoto cluster (see Base case model in Table 5), R_s was estimated at 1.14 (95% CI: 0.61–2.09). v, the relative reproduction number for asymptomatic cases, was estimated at 0.19 (95% CI: 0.03–0.66), indicating that symptomatic cases appeared to be 5.3 times more infectious than asymptomatic cases. The dispersion parameter was estimated at 0.24 (95% CI: 0.13–0.47). In either cluster, a geometric or Poisson distribution analysis yielded similar estimate while AIC was minimum for a negative binomial distribution analysis.

To assess the uncertainty of v in Kyoto cluster, R_s and v were estimated given fixed dispersion parameters of 0.05, 0.1, 0.2, 0.3, 1, or 10. The range of 0.05–0.2 was consistent with the reported 95% credible interval of the dispersion parameter based on the World Health Organization situation report [Endo et al., 2020], while dispersion parameters of 1 and 10 implied that the numbers of secondary cases per primary symptomatic case approximately follow a geometric or Poisson distribution,

respectively. We found that the 95% CI of v narrowed as dispersion parameters increased, with a 95% CI of 0.06–0.5 at k=10 (Table 6).

Table 4. Epidemiological parameters of SARS-CoV-2 transmission in Tokyo and Kanagawa cluster. CI: confidence interval. Parameters included the reproduction number of symptomatic cases (Rs), the reproduction number for an asymptomatic case with respective to a symptomatic case (v) and the dispersion parameter (k).

Parameters	Maximum likelihood estimate (95% CI)
R_s	1.2 (0.5–2.9)
v	0.27 (0.03–0.81)
k	0.23 (0.09–0.62)

Table 5. Epidemiological parameters of SARS-CoV-2 transmission in Kyoto cluster. CI: confidence interval. Parameters included the reproduction number of symptomatic cases (Rs), the reproduction number of symptomatic cases at calendar time zero (Rt=0), the reproduction number for an asymptomatic case with respective to a symptomatic case (v), the dispersion parameter (k), and the exponentially decreasing rate of secondary transmission (δ)

	Base case model	Exponential decrease model
Parameters	Maximum likelihood estimate (95% CI)	Maximum likelihood estimate (95% CI)
R_s or $R_{t=0}$	1.14 (0.61–2.09)	12.6 (0.69–37.0)
v	0.19 (0.03–0.66)	0.07 (0.01–0.79)
k	0.24 (0.13–0.47)	0.26 (0.11–0.63)
δ	-	0.29 (0.10–0.61)

Table 6. Sensitivity of the reproduction number (R_s) and relative transmissibility among asymptomatic individuals (v) in Kyoto cluster to assumed values of the dispersion parameter, k. CI: confidence interval.

	R_s	v
Dispersion parameter, k	95% CI	95% CI
0.05	0.31–3.78	0.02–0.90
0.10	0.45–2.76	0.02–0.79
0.20	0.58–2.20	0.03–0.68
0.30	0.64–1.99	0.04–0.63
1.00	0.78–1.66	0.05–0.54
10.00	0.87–1.50	0.06–0.50

We estimated $R_{t=0}$ and v at 12.6 (95% CI: 0.69–37.0) and 0.07 (95% CI: 0.01–0.79), respectively, under a scenario where the reproduction number exponentially decreases as a function of calendar date (see Exponential decrease model in Table 5). We could not calculate the value of v precisely when jointly estimated with the proportion of symptomatic cases; when v was fixed as 0.2, we could estimate the proportion of symptomatic cases arising from each primary symptomatic case and asymptomatic case at 0.82 (95% CI: 0.49–0.95) and 0.78 (95% CI: 0.64–0.87), respectively.

3. The relative risk of secondary transmission among isolated individuals

We assummed that infectiousness started 6 days prior to illness onset and that the probability density function of secondary transmission relative to illness onset followed a gamma distribution. The shape and scale parameters were estimated at 5.49 (95% CI: 1.99–14.74) and 1.00 (95% CI: 0.35–2.57), respectively. We found the model using the probability density function of the serial interval shortened by case isolation not converge to allow an explicit estimation of ε, the isolation effect, although the estimate was found to be close to the lower bound. R_s, v and k were estimated at 1.58 (95%

CI: 0.68–3.47), 0.32 (95% CI: 0.02–0.95) and 0.26 (95% CI: 0.12–0.57), respectively (Table 7). We could not obtain a stable estimate of ε successfully even when assuming that the relative frequency of secondary transmission with respect to time since illness onset was known and fixed using the estimates of He et al. [He et al., 2020] (Table 8).

Table 7. Parameters estimated in the two models. The included parameters were the reproduction number of symptomatic cases (R_s), the reproduction number of symptomatic cases at calendar time of zero ($R_{t=0}$), the relative reduction in the rate of secondary transmission in an isolated individual (ε), the relative reduction in the reproduction number for an asymptomatic case with respect to a symptomatic case (v), the dispersion parameter (k), the shape (α_1) and rate (α_2) parameter of the gamma distribution, and the exponentially decreasing rate of secondary transmission (δ). CI: confidence interval; NaN, not a number

	Base case	Exponential decrease
Parameters	Mean (95% CI)	Mean (95% CI)
R_s or $R_{t=0}$	1.58 (0.68–3.47)	8.05 (1.11–17.7)
ε	0.001 (0.001–NaN)	0.001(0.001–0.999)
v	0.32 (0.02–0.95)	0.15 (0.02–0.75)
k	0.26 (0.12–0.57)	0.43 (0.17–1.06)
α_1	5.49 (1.17–9.19)	5.49 (1.96–14.4)
α_2	1.00 (0.35–2.57)	1.00 (0.35–2.58)
δ	-	0.23 (0.10–0.46)

Table 8. Parameters estimated in the two models. Assumed that the relative frequency of secondary transmission with respect to time since illness onset was known and fixed using the estimates of He et al. [He et al., 2020]. m was set as 12. The included parameters were the reproduction number of symptomatic cases (R_s), the reproduction number of symptomatic cases at calendar time of zero ($R_{t=0}$), the relative reduction in the rate of secondary transmission in an isolated individual (ε), the relative reduction in the reproduction number for an asymptomatic case with respect to a symptomatic case (v), the dispersion parameter (k) and the exponentially decreasing rate of secondary transmission (δ). CI: confidence interval; NaN, not a number

	Base case	Exponential decrease
Parameters	Mean (95% CI)	Mean (95% CI)
R_s or $R_{t=0}$	1.58 (0.69–3.53)	8.04 (1.1–17.7)
ε	0.001 (NaN–NaN)	0.001(0.001–NaN)
v	0.32 (0.02–0.95)	0.15 (0.02–0.74)
k	0.26 (0.12–0.57)	0.43 (0.17–1.06)
δ	-	0.23 (0.10–0.45)

Discussion

In Section 2, we investigated the remaining key parameter of COVID-19 that we could not incorporate into our model explicitly in Section 1, i.e., the relative transmissibility of asymptomatic SARS-CoV-2 infections. It was assessed in terms of the reproduction number and the serial dependence of symptomatic infection, using two cluster data and chains of transmission during the early stages of the epidemic in Japan. Assuming that the distribution of secondary cases followed a negative binominal distribution, the reproduction number of symptomatic cases in Tokyo/Kanagawa cluster and Kyoto cluster was estimated at 1.2 (95% CI: 0.5–2.9) and 1.14 (95% CI: 0.61–2.09), respectively. The relative reproduction number for asymptomatic cases for each cluster was estimated at 0.27 (95% CI: 0.03–0.81) and 0.19 (95% CI: 0.03–0.66), respectively, suggesting that the transmissibility of asymptomatic cases may not be highly sensitive to setting. In Kyoto cluster, $R_{t=0}$ (=12.6) was found to be much larger than R_s This can be explained because $R_{t=0}$ is defined as the reproduction number at the biggening

of the epidemic calendar time, with assumption of an exponential decrease over the course of epidemic, while R_s was assumed to be a constant throughout the course of epidemic. We could not find any apparent increased tendency for symptomatic primary case to produce symptomatic secondary cases. We also assessed the relative transmissibility in the model using the probability density function of the generation interval adjusted for the isolation period in Kyoto cluster because movement of all identified close contacts was restricted for 14 days. We could not, however, successfully estimate the effectiveness of case isolation jointly with other parameters.

Reproduction number estimates for asymptomatic cases of SARS-CoV-2 infection was reported in one study [He et al., 2020] (Table 9). Analyzing data on contact tracing from a first generation of 191 cases (161 symptomatic and 30 asymptomatic cases) during the very early stages of the epidemic in Ningbo, China, He et al. reported that the reproduction numbers of symptomatic and asymptomatic cases were 0.78 and 0.20, respectively, indicating a risk ratio for transmission of 0.26 in asymptomatic cases [He et al., 2020]. Our point estimates of the relative reproduction number for asymptomatic cases from both clusters were consistent with this finding. The transmissibility of asymptomatic infections were assessed using epidemiological measurements other than the reproduction number in two other contact tracing studies. The relative transmissibility of asymptomatic cases was shown to be around one-third (0.26) that of symptomatic cases from a recent report using the incidence rate ratio adjusted for age, sex and serological status [Sayampanathan et al., 2021], which agrees with our results. It should be noted that this result was based on regular screening of workers in specific industries, not intensive investigations triggered by notification of clusters such as in our studies. By investigating an outbreak followed by a cluster at a religious event in Brunei, however, the secondary attack rate ratio for asymptomatic vs. symptomatic cases (including pre-symptomatic cases) was found to be close to parity (1.1; calculated from data in Supplementary Information) [Chaw et al., 2020]. When the analysis was restricted to cases in households, the attack rate ratio was 0.37. In a household transmission study in Japan, it was indicated that the secondary attack risk of asymptomatic primary cases was 11.8% while overall secondary attack risk was 19.0% [Miyahara et al., 2021]. Taken together, these results suggest that the transmissibility of asymptomatic cases is less than half that of symptomatic cases.

Table 5. Summary of transmission profiles of asymptomatic cases in contact tracing studies. *pre-symptomatic cases were counted as symptomatic cases; **reciprocal of reported values; ***calculated manually from Supplementary Table 1.

Study	Setting	Sample size	Measurement of relative infectivity of asymptomatic cases
Tokyo and Kanagawa cluster	Tokyo and Kanagawa, Japan	12 asymptomatic and 24 symptomatic cases	Ratio of reproduction numbers: 0.27 (95% CI: 0.03–0.81)
Kyoto cluster	Kyoto, Japan	13 asymptomatic and 51 symptomatic cases	Ratio of reproduction numbers: 0.19 (95% CI: 0.03–0.66)
He et al., 2020.	Ningbo city, China	52 asymptomatic and 271 symptomatic cases	Ratio of reproduction numbers: 0.26**
Sayampanathan et al., 2021.	Singapore	3035 contacts of asymptomatic cases and 755 contacts of symptomatic cases	Incidence rate ratio: 0.26**
Chaw et al., 2020.	Brunei	106 contacts of asymptomatic cases and 1595 contacts of symptomatic cases*	Attack rate ratio: 1.12***

The value of contact tracing for COVID-19, including the cluster-based approach [Oshitani et al., 2020], in preventing major epidemics can be critically assessed by findings from our investigation. Our findings would assure the validity of the cluster-based approach, and observing clusters starting with symptomatic transmission might be sufficient for the control. First, even if we missed and untraced an initially asymptomatic index case, the reproduction number was estimated as R_a=0.32 or 0.21, substantially below parity, and it is very likely that the resulting outbreak would decline to extinction. Second, let us suppose that only a proportion x of contacts are traced, which may be correlated with 1- the asymptomatic ratio. The reproduction number with contact tracing, then, would be $(1-x)((1-z)R_s+ zR_a) +ux(yR_s+(1-y)R_a)$, where z is the asymptomatic ratio, y is the proportion of traced asymptomatic contacts that would eventually develop symptoms, and u is the reduction factor resulting from contact tracing. Assuming that $u\approx0$ and $x=kz$, where k is a constant, we can simplify the

reproduction number with contact tracing to $(1-kz)((1-z)R_s+ zR_a)$. Assuming that k=1, z is 0.30 or 0.50 [Nishiura et al., 2020] and R_s =1.14, we can calculate the resulting reproduction number at 0.60 and 0.34, respectively. Thus, when tracing capacity is substantial, we could justify implementing contact tracing beginning with symptomatic cases, especially if testing capacity is limited.

A symptomatic case was found not to have a higher tendency to produce additional symptomatic infections than asymptomatic cases. In addition, there was no suggestion that asymptomatic secondary cases were more likely to be generated from an asymptomatic primary case. Our approach of using a classical branching process model of the generation-dependent transmission process in an independent manner might be supported by these findings.

Our study on the relative transmissibility of asymptomatic infections has several limitations. First, the sample size was small in each cluster, involving a broad uncertainty bound and a wide 95% CI. For instance, the number of asymptomatic cases was 13 in Kyoto cluster, accounting for approximately 20% out of all cases, and we found that the number of primary asymptomatic case was only one in this cluster. Thus, to obtain more precise estimates of the relative transmissibility and effectiveness of case isolation, a larger number of asymptomatic cases would be needed. Second, the tracing results did not perfectly capture all chains of transmission in both clusters: there was the missing link of second Kanagawa case in Tokyo/Kanagawa cluster, which might even be pre-symptomatic or asymptomatic transmission. In Kyoto cluster, 10 symptomatic cases had to be excluded because specific transmission events could not be identified. It was possible that three cases among these 10 were infected by asymptomatic individuals in the transmission network. Under a hypothetical scenario that a single asymptomatic case infected all three cases, it was estimated that the relative transmissibility for asymptomatic cases was 0.35, which is compatible with our baseline estimate. Third, the effect of the first 14 days of quarantine period on transmission profile could not be successfully estimated in Kyoto cluster, because we could not have the date of identification as being close contact for each subject available.

Conclusion

In this thesis, two key epidemiological parameters governing transmission dynamics of COVID-19, i.e., R(t) and the relative transmissibility of asymptomatic cases, were investigated using mathematical modeling approach while addressing the issue of "observability" carefully. Findings through Section 1 and 2 in this thesis were as follows:

- The alternative method of estimating R(t) based on observable illness onset data was firstly devised and successful in estimating the R(t) for COVID-19.
- A modified renewal equation based on the date of illness onset was employed, leveraging the frequency of secondary transmission to capture the unique characteristics of COVID-19 (i.e., pre-symptomatic transmission).
- Using a piecewise constant model, the epidemic in Osaka prefecture came under control around 2 April 2020 during the first wave and around 26 July 2020 during the second wave.
- Using transmission networks of early clusters in Japan, the transmissibility of asymptomatic cases of SARS-CoV-2 infection was shown to be small relative to symptomatic cases.
- There was no apparent increased tendency for symptomatic primary cases to produce symptomatic secondary cases.

Our findings suggested that concerted efforts would be required to curb the COVID-19 epidemic and contact tracing focusing on symptomatic index cases may be justified when there is limited testing capacity. The outcome of the COVID-19 or future pandemics would continue to rely on political leadership to swiftly design and implement combined interventions to reduce contacts broadly and appropriately. This study would assist planning the strategy for the efficient and feasible public health actions in future pandemics.

Acknowledgements

The manuscript 1 was supported by the German Federal Ministry of Health (BMG) COVID-19 Research and Development funding from the WHO and by funding from Health and Labor Sciences Research Grants (19HA1003, 20CA2024, and 20HA2007); the Japan Agency for Medical Research and Development (AMED; JP19fk0108104 and JP20fk0108140); the Japan Society for the Promotion of Science (JSPS) KAKENHI (17H04701), the Inamori Foundation, GAP fund program by Kyoto University, and the Japan Science and Technology Agency (JST) CREST program (JPMJCR1413) and SICORP (e-ASIA) program (JPMJSC20U3).

The manuscript 2 was supported by funding from a Health and Labor Sciences Research Grant (19HA1003, 20CA2024, and 20HA2007), the Japan Agency for Medical Research and Development (AMED; JP19fk0108104 and JP20fk0108140), the Japan Society for the Promotion of Science (JSPS) KAKENHI (17H04701), the Inamori Foundation, and the Japan Science and Technology Agency (JST) CREST program (JPMJCR1413).

The manuscript 3 was supported by German Federal Ministry of Health (BMG) COVID-19 Research and Development funding to the World Health Organization and by funding from Health and Labor Sciences Research Grants (19HA1003, 20CA2024, and 20HA2007); the Japan Agency for Medical Research and Development (AMED; JP19fk0108104, JP20fk0108140 and JP20fk0108535s0101); the Japan Society for the Promotion of Science (JSPS) KAKENHI (17H04701 and 21H03198); the GAP fund program of Kyoto University; the Japan Science and Technology Agency (JST) CREST program (JPMJCR1413); and the SICORP (e-ASIA) program (JPMJSC20U3). The funders were not involved in the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit the work for publication.

I would like to thank:
My Ph.D. advisor, Hiroshi Nishiura, for introducing me to mathematical epidemiology and fun of it.
Kei, Sari and Yuko for their love, much patience and support.

Conflict of interests

The author declares no conflict of interest.

References

Abbott, S., Hellewell, J., Thompson, R.N., Sherratt, K., Gibbs, H.P., Bosse, N.I., Munday, J.D., Meakin, S., Doughty, E.L., Chun, J.Y., et al. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. Wellcome Open Res. 5, 112.

Adam, D.C., Wu, P., Wong, J.Y., Lau, E.H.Y., Tsang, T.K., Cauchemez, S., Leung, G.M., Cowling, B.J. (2020). Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat Med. 26, 1714–9.

Ali, S.T., Wang, L., Lau, E.H.Y., Xu, X.K., Du, Z., Wu, Y., Leung, G.M., Cowling, B.J. (2020). Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 369, 1106–1109.

Ayoub, H.H., Chemaitelly, H., Seedat, S., Mumtaz, G.R., Makhoul, M., Abu-Raddad, L.J. (2020). Age could be driving variable SARS-CoV-2 epidemic trajectories worldwide. PLoS ONE. 15, e0237959.

Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D.Y., Chen, L., Wang, M. (2020). Presumed asymptomatic carrier transmission of COVID-19. JAMA. 323, 1406–7.

Becker, N.G., Watson, L.F., Carlin, J.B. (1991). A method of non-parametric back-projection and its application to aids data. Stat. Med. 10, 1527–1542.

Caicedo-Ochoa, Y., Rebellón-Sánchez, D.E., Peñaloza-Rallón, M., Cortés-Motta, H.F., Méndez-Fandiño, Y.R. (2020). Effective Reproductive Number estimation for initial stage of COVID-19 pandemic in Latin American Countries. Int. J. Infect. Dis. 95, 316–318.

Chan, Y.H., Nishiura, H. (2020). Estimating the protective effect of case isolation with transmission tree reconstruction during the Ebola outbreak in Nigeria, 2014. J. R. Soc. Interface. 17, 20200498.

Chaw, L., Koh, W.C., Jamaludin, S.A., Naing, L., Alikhan, M.F., Wong, J. (2020). Analysis of SARS-CoV-2 transmission in different settings, Brunei. Emerg Infect Dis.

26, 2598–606.

Chen, Y., Wang, A.H., Yi, B., Ding, K.Q., Wang, H.B., Wang, J.M., Shi, H.B., Wang, S.J., Xu, G.Z., et al. (2020). Epidemiological characteristics of infection in COVID-19 close contacts in Ningbo city. Chin J Epidemiol. 41, 667–71.

Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. Am. J. Epidemiol. 178, 1505–1512.

Davies, N.G., Klepac, P., Liu, Y., Prem, K., Jit, M., Pearson, C.A.B., Quilty, B.J., Kucharski, A.J., Gibbs, H., Clifford, S., et al. (2020). Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat. Med. 26, 1205–1211.

Endo, A.; Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group., Abbott, S., Kucharski, A.J., Funk, S. (2020). Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res. 5:67.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H.J.T., Mellan, T.A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J.W., et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 584, 257–261.

Furuse, Y., Sando, E., Tsuchiya, N., Miyahara, R., Yasuda, I., Ko YK., Saito, M., Morimoto, K., Imamura, T., Shobugawa, Y., et al. (2020). Clusters of coronavirus disease in communities, Japan, January–April 2020. Emerg Infect Dis. 26, 2176–9.

Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., Hens, N. (2020). Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Euro Surveill. 25, 1-8.

Gostic, K.M., Mcgough, L., Baskerville, E.B., Abbott, S., Joshi, K., Tedijanto, C., Kahn, R., Niehus, R., Hay, J., De Salazar, P.M., et al. (2020). Practical considerations for measuring the effective reproductive number, Rt. medRxiv 2020.

He, D., Zhao, S., Lin, Q., Zhuang, Z., Cao, P., Wang, M.H., Yang, L., et al. (2020). The

relative transmissibility of asymptomatic COVID-19 infections among close contacts. Int J Infect Dis. 94, 145–7.

He, X., Lau, E.H.Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., Tan, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat. Med. 26, 672–675.

Lau, M.S.Y., Grenfell, B., Thomas, M., Bryan, M., Nelson, K., Lopman, B. (2020). Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. Proc Natl Acad Sci U S A. 117, 22430–5.

Lee, S., Kim, T., Lee, E., Lee, C., Kim, H., Rhee, H., Park, S., Son, H., Yu, S., Park, J., et al. (2020). Clinical course and molecular viral shedding among asymptomatic and symptomatic patients with SARS-CoV-2 infection in a community treatment center in the republic of Korea. JAMA Intern Med. 180, 1447–52.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med. 382, 1199–207.

Johansson, M.A., Quandelacy, T.M., Kada, S., Prasad, P.V., Steele, M., Brooks, J.T., Slayton, R.B., Biggerstaff, M., Butler, J.C., et al. SARS-CoV-2 transmission from people without COVID-19 symptoms. (2021). JAMA Netw Open. 4:e2035057.

Joinpoint Regression Program; Version 4.8.0; Statistical Methodology and Applications Branch; Surveillance Research Program; National Cancer Institute: Bethesda, MD, USA, 1 April 2020.

Kendall, M., Milsom, L., Abeler-Doerner, L., Wymant, C., Ferretti, L., Briers, M., Holmes, C., Bonsall, D., Abeler, J., Fraser, C. (2020). Epidemiological changes on the Isle of Wight after the launch of the NHS Test and Trace programme: A preliminary analysis. Lancet Digit. Health. 2, e658–e666.

Kimball, A., Hatfield, K.M., Arons, M., James, A., Taylor, J., Spicer, K., Bardossy, A.C., Oakley, L.P., Tanwar, S., Chisty, A., et al. (2020). 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. 69, 377–81.

Koo, J.R., Cook, A.R., Park, M., Sun, Y., Sun, H., Lim, J.T., Tam, C., Dickens, B.L. (2020). Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. Lancet Infect Dis. 20, 678–88.

Kucharski, A.J., Klepac, P., Conlan, A.J.K., Kissler, S.M., Tang, M.L., Fry, H., Gog, J.R., Edmunds, W.J. (2020). Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. Lancet Infect Dis. 20, 1151–60.

Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R.M., Sun, F., Jit, M., Munday, J.D., et al. (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. Lancet Infect. Dis. 20, 553–558.

Kuniya, T. (2020). Evaluation of the effect of the state of emergency for the first wave of COVID-19 in Japan. Infect. Dis. Model. 5, 580–587.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. 382, 1199–1207.

Linton, N., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A., Jung, S., Yuan, B., Kinoshita, R., Nishiura, H. (2020). Incubation period and other epidemiological characteristics of 2019 novel Coronavirus infections with right truncation: A statistical analysis of publicly available case data. J. Clin. Med. 9, 538.

Lipsitch, M., Joshi, K., Cobey, S.E. Comment on Pan A, Liu L, Wang C; et al. (2020). Association of Public Health Interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA J. Am. Med. Assoc. 323, 1915–1923.

Liu, Y., Funk, S., Flasche, S., Jit, M., Bosse, N.I., Gimma, A., Klepac, P., Russell, T.W., Sun, F., Rosello, A., et al. (2020). The contribution of pre-symptomatic infection to the transmission dynamics of COVID-2019. Wellcome Open Res.

McEvoy, D., McAloon, C.G., Collins, Á.B., Hunt, K., Butler, F., Byrne, A.W., Casey, M., Barber, A., Griffin, J., Lane, E.A., et al. (2020). The relative infectiousness of asymptomatic SARS-CoV-2 infected persons compared with symptomatic individuals: a rapid scoping review. medRxiv 2020.07.30.20165084.

Miyahara, R., Tsuchiya, N., Yasuda, I., Ko, Y.K., Furuse, Y., Sando, E., Nagata, S., Imamura, T., Saito, M., Morimoto, K., et al. (2021). Familial clusters of coronavirus disease in 10 prefectures, Japan, February–May 2020. Emerg Infect Dis. 27, 915–8.

Nishiura, H., Chowell, G. (2009). The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In Mathematical and Statistical Estimation Approaches in Epidemiology. (New York, USA: Springer), pp. 103–121. ISBN 9789048123124.

Nishiura, H., Chowell, G. (2014). Early transmission dynamics of Ebola virus disease (evd), West Africa, March to August 2014. Euro Surveill. 19, 1–6.

Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, SM., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A.R., et al. (2020). Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). Int J Infect Dis. 94, 154–5.

Nishiura, H., Linton, N.M., Akhmetzhanov, A.R. (2020). Serial interval of novel coronavirus (COVID-19) infections. Int. J. Infect. Dis. 93, 284–286.

Oran, D.P., Topol, E.J. (2020). Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. Ann Intern Med. 173, 362–7.

Oshitani, H., Expert Members of The National COVID-19 Cluster Taskforce at The Ministry of Health, Labour and Welfare, Japan. (2020). Cluster-based approach to coronavirus disease 2019 (COVID-19) response in Japan, from February to April 2020. Jpn J Infect Dis. 73, 491–3.

Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., et al. (2020). Association of public health interventions with the epidemiology of the Covid-19 outbreak in Wuhan, China. JAMA J. Am. Med. Assoc. 323, 1915–1923.

Petermann, M., Wyler, D. (2020). A pitfall in estimating the effective reproductive number Rt for COVID-19. Swiss Med. Wkly. 150, w20307.

R Foundation for Statistical Computing. (2020). R: A language and environment for statistical computing. https://www.r-project.org. Accessed on 3 March 2021.

Ryu, S., Ali, S.T., Jang, C., Kim, B., Cowling, B.J. (2020). Effect of nonpharmaceutical interventions on transmission of severe acute respiratory syndrome Coronavirus 2, South Korea, 2020. Emerg. Infect. Dis. 26, 2406–2410.

Sayampanathan, A.A., Heng, C.S., Pin, P.H., Pang, J., Leong, T.Y., Lee, V.J. (2021). Infectivity of asymptomatic versus symptomatic COVID-19. Lancet. 397, 93–4.

Scire, J., Nadeau, S., Vaughan, T., Brupbacher, G., Fuchs, S., Sommer, J., Koch, K.N., Misteli, R., Mundorff, L., Götz, T., et al. (2020). Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft. Swiss Med. Wkly. 150, w20271.

Tariq, A., Lee, Y., Roosa, K., Blumberg, S., Yan, P., Ma, S., Chowell, G. (2020). Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020. BMC Med. 18, 166.

The Royal Society. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. 2020. Available at: https://royalsociety.org/-/media/policy/projects/set-c/set-covid-19-R-estimates.pdf. Accessed 16 Mar 2021.

Wallinga, J., Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol. 160, 509–516.

Wallinga, J., Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. B Biol. Sci. 274, 599–604.

Yang, L., Dai, J., Zhao, J., Wang, Y., Deng, P., Wang, J. (2020). Estimation of incubation

period and serial interval of COVID-19: Analysis of 178 cases and 131 transmission chains in Hubei province, China. Epidemiol. Infect. 148, e117.

Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., et al. (2020). SARS-CoV-2 viral load in upper respiratory specimens of infected patients. N Engl J Med. 382, 1177–9.

Appendix

R code for Section 1:

```
rm(list=ls())
libraries = c("dplyr","magrittr", "ggplot2", "readr")
for(x in libraries) { library(x,character.only=TRUE,warn.conflicts=FALSE) }

#effective reproductive number of COVID19 in Osaka city

#--- data ---
osaka <- read.csv(file = "osaka_30Sep2020.csv")
osaka %<>% select(day, case)

#[Figure] create epidemic curve
osaka  %<>% mutate(date = seq(as.Date('2020-02-17'), as.Date('2020-09-29'), by = '1 day'))
ggplot(data=osaka)+
  geom_bar(mapping = aes(x=date, y=case),
           stat = "identity")+
  theme_classic() + #background will be white
  scale_x_date(date_labels = "%m/%d", date_breaks = "1 week")  +
  theme(axis.text.x = element_text(angle = 60, hjust = 1))+
  labs(x="Date of illness onset",y="Number of daily incidence")

# secondary transmission function
# from He et al. from Nature medicine
st_ <- function(s){
  dgamma(s, 20.516508,  1.592124)
}

# data frame for secondary transmission
# set day 12 as the oldest day relative to the onset of illness
df <- data.frame(day = 0:(max(osaka$day) + 12))

#secondary transmission vector on day
stv <- sapply(0:max(df$day), st_) #sapply will provide vector

#data frame with secondary transmission risk
df %<>% mutate(ST = stv)
head(df)

#calculate first convulsion between cases and frequency of secondary transmission
conv = c()
for (x in 1:nrow(osaka)) {
  conv = c(conv,sum(osaka$case[(x+12):1]*df$ST[1:(x+12)])) }

#data frame with the results of first convulsion
osaka %<>% do(cbind(.,conv1=conv))
head(osaka)

#delete rows with NA: the last 12 rows will be deleted
osaka %<>% na.omit
tail(osaka)
```

```
# log-normal distribution for incubation period
# from Li et al NEJM 2020
# lognormal mean = 5.2; 95% CI = c(4.1, 7.0)
# logmean = 1.434065
# logsd = 0.6612
ip = function(s){
  dlnorm(s, meanlog = 1.434065, sdlog = 0.6612, log = FALSE)}

#data frame with frequencies of incubation period
osaka %<>% mutate(f = sapply(0:max(osaka$day), ip))
tail(osaka) #to get the number of row

#go to code for each piece-wise model

#5-day model ----------------------------------------------------------------
#get the number of Rt to be estimated
print(214%/%5)
print(214%%5) #total 42 Rts will be estimated with the last having 9 as an interval

#log-likelihood function for piece wise model
llik.pc <- function(prm, prediction=FALSE){
  R1=prm[1]; R2=prm[2]; R3=prm[3]; R4=prm[4]; R5=prm[5]; R6=prm[6]; R7=prm[7];
  R8=prm[8]; R9=prm[9]; R10=prm[10];R11=prm[11]; R12=prm[12]; R13=prm[13]; R14=prm[14];
  R15=prm[15]; R16=prm[16]; R17=prm[17]; R18=prm[18]; R19=prm[19]; R20=prm[20];
R21=prm[21];
  R22=prm[22];R23=prm[23]; R24=prm[24]; R25=prm[25]; R26=prm[26]; R27=prm[27]; R28=prm[28];
  R29=prm[29]; R30=prm[30]; R31=prm[31]; R32=prm[32]; R33=prm[33];R34=prm[34];R35=prm[35];
  R36=prm[36]; R37=prm[37]; R38=prm[38]; R39=prm[39]; R40=prm[40]; R41=prm[41]; R42=prm[42]
  Rt <- c(rep(R1, 5), rep(R2, 5), rep(R3, 5), rep(R4, 5), rep(R5, 5), rep(R6, 5),
          rep(R7, 5), rep(R8, 5), rep(R9, 5), rep(R10,5),rep(R11, 5), rep(R12, 5),
          rep(R13, 5), rep(R14, 5), rep(R15, 5), rep(R16, 5), rep(R17, 5), rep(R18, 5),
          rep(R19, 5), rep(R20, 5), rep(R21, 5), rep(R22,5),rep(R23, 5), rep(R24, 5),
          rep(R25, 5), rep(R26, 5), rep(R27, 5), rep(R28, 5), rep(R29, 5), rep(R30, 5),
          rep(R31, 5), rep(R32, 5), rep(R33, 5), rep(R34,5),rep(R35, 5), rep(R36, 5),
          rep(R37, 5), rep(R38, 5), rep(R39, 5), rep(R40, 5), rep(R41, 5), rep(R42, 9))
  conv = c()
  for (x in 1:nrow(osaka)) {
    conv = c(conv,sum(osaka$conv1[x:1]*(Rt[x:1]*osaka$f[1:x])))}
  #data frame with the second convolution
  osaka %<>% do(cbind(.,epc=conv))

  if (prediction) {
    osaka %>% select(epc) %>% return
  } else {
    osaka %>%
      summarize(loglk = sum(case[2:nrow(osaka)]*log(epc[2:nrow(osaka)])-epc[2:nrow(osaka)]-
lfactorial(case[2:nrow(osaka)]))) %>%
      .$loglk -> llik #exclude n=1 because the value of log is infinite
    return(-llik)}}

#initial values
prm <- rep(2, 42)

#Optimize by L-BFGS-B
```

```
sol = optim(par=prm,fn=llik.pc,
            method="L-BFGS-B",lower=rep(0,42),control = list(maxit=10000),
            hessian=TRUE)
pars = sol$par
sol

#data frame with the estimated Rt
Rt <- data.frame(Rt = c(rep(sol$par[1], 5),
                        rep(sol$par[2], 5),
                        rep(sol$par[3], 5),
                        rep(sol$par[4], 5),
                        rep(sol$par[5], 5),
                        rep(sol$par[6], 5),
                        rep(sol$par[7], 5),
                        rep(sol$par[8], 5),
                        rep(sol$par[9], 5),
                        rep(sol$par[10], 5),
                        rep(sol$par[11], 5),
                        rep(sol$par[12], 5),
                        rep(sol$par[13], 5),
                        rep(sol$par[14], 5),
                        rep(sol$par[15], 5),
                        rep(sol$par[16], 5),
                        rep(sol$par[17], 5),
                        rep(sol$par[18], 5),
                        rep(sol$par[19], 5),
                        rep(sol$par[20], 5),
                        rep(sol$par[21], 5),
                        rep(sol$par[22], 5),
                        rep(sol$par[23], 5),
                        rep(sol$par[24], 5),
                        rep(sol$par[25], 5),
                        rep(sol$par[26], 5),
                        rep(sol$par[27], 5),
                        rep(sol$par[28], 5),
                        rep(sol$par[29], 5),
                        rep(sol$par[30], 5),
                        rep(sol$par[31], 5),
                        rep(sol$par[32], 5),
                        rep(sol$par[33], 5),
                        rep(sol$par[34], 5),
                        rep(sol$par[35], 5),
                        rep(sol$par[36], 5),
                        rep(sol$par[37], 5),
                        rep(sol$par[38], 5),
                        rep(sol$par[39], 5),
                        rep(sol$par[40], 5),
                        rep(sol$par[41], 5),
                        rep(sol$par[42], 9)))

#combine two data-frames with respect to column
osaka %<>% do(cbind(.,Rt))

#construct 95% CI of Rt
```

```
#get variance-covariance matrix 繰□from hessian
vc = solve(sol$hessian[1:42, 1:42])
sd = sqrt(diag(vc))

#sampling of Rt from multivariate normal distribution
sample = 1e3
hess_sam = MASS::mvrnorm(n=sample, mu=sol$par[1:42],
                         Sigma=solve(sol$hessian[1:42, 1:42]),
                         tol=1e-06, empirical=FALSE, EISPACK=FALSE)
hess_sam %>% as.data.frame -> df_hess

colnames(df_hess) = c("R1","R2","R3","R4","R5","R6","R7","R8", "R9","R10","R11","R12",
"R13",
                      "R14","R15","R16","R17","R18","R19","R20","R21",
"R22","R23","R24","R25", "R26",
                      "R27","R28","R29","R30","R31","R32","R33","R34",
"R35","R36","R37","R38", "R39",
                      "R40","R41","R42")
df_hess %>% head

# selecting samples within 95% range
data.frame(
  R1=filter(df_hess,R1 > quantile(R1, 0.025) & R1 < quantile(R1, 0.975))$R1,
  R2=filter(df_hess,R2 > quantile(R2, 0.025) & R2 < quantile(R2, 0.975))$R2,
  R3=filter(df_hess,R3 > quantile(R3, 0.025) & R3 < quantile(R3, 0.975))$R3,
  R4=filter(df_hess,R4 > quantile(R4, 0.025) & R4 < quantile(R4, 0.975))$R4,
  R5=filter(df_hess,R5 > quantile(R5, 0.025) & R5 < quantile(R5, 0.975))$R5,
  R6=filter(df_hess,R6 > quantile(R6, 0.025) & R6 < quantile(R6, 0.975))$R6,
  R7=filter(df_hess,R7 > quantile(R7, 0.025) & R7 < quantile(R7, 0.975))$R7,
  R8=filter(df_hess,R8 > quantile(R8, 0.025) & R8 < quantile(R8, 0.975))$R8,
  R9=filter(df_hess,R9 > quantile(R9, 0.025) & R9 < quantile(R9, 0.975))$R9,
  R10=filter(df_hess,R10 > quantile(R10, 0.025) & R10 < quantile(R10, 0.975))$R10,
  R11=filter(df_hess,R11 > quantile(R11, 0.025) & R11 < quantile(R11, 0.975))$R11,
  R12=filter(df_hess,R12 > quantile(R12, 0.025) & R12 < quantile(R12, 0.975))$R12,
  R13=filter(df_hess,R13 > quantile(R13, 0.025) & R13 < quantile(R13, 0.975))$R13,
  R14=filter(df_hess,R14 > quantile(R14, 0.025) & R14 < quantile(R14, 0.975))$R14,
  R15=filter(df_hess,R15 > quantile(R15, 0.025) & R15 < quantile(R15, 0.975))$R15,
  R16=filter(df_hess,R16 > quantile(R16, 0.025) & R16 < quantile(R16, 0.975))$R16,
  R17=filter(df_hess,R17 > quantile(R17, 0.025) & R17 < quantile(R17, 0.975))$R17,
  R18=filter(df_hess,R18 > quantile(R18, 0.025) & R18 < quantile(R18, 0.975))$R18,
  R19=filter(df_hess,R19 > quantile(R19, 0.025) & R19 < quantile(R19, 0.975))$R19,
  R20=filter(df_hess,R20 > quantile(R20, 0.025) & R20 < quantile(R20, 0.975))$R20,
  R21=filter(df_hess,R21 > quantile(R21, 0.025) & R21 < quantile(R21, 0.975))$R21,
  R22=filter(df_hess,R22 > quantile(R22, 0.025) & R22 < quantile(R22, 0.975))$R22,
  R23=filter(df_hess,R23 > quantile(R23, 0.025) & R23 < quantile(R23, 0.975))$R23,
  R24=filter(df_hess,R24 > quantile(R24, 0.025) & R24 < quantile(R24, 0.975))$R24,
  R25=filter(df_hess,R25 > quantile(R25, 0.025) & R25 < quantile(R25, 0.975))$R25,
  R26=filter(df_hess,R26 > quantile(R26, 0.025) & R26 < quantile(R26, 0.975))$R26,
  R27=filter(df_hess,R27 > quantile(R27, 0.025) & R27 < quantile(R27, 0.975))$R27,
  R28=filter(df_hess,R28 > quantile(R28, 0.025) & R28 < quantile(R28, 0.975))$R28,
  R29=filter(df_hess,R29 > quantile(R29, 0.025) & R29 < quantile(R29, 0.975))$R29,
  R30=filter(df_hess,R30 > quantile(R30, 0.025) & R30 < quantile(R30, 0.975))$R30,
  R31=filter(df_hess,R31 > quantile(R31, 0.025) & R31 < quantile(R31, 0.975))$R31,
  R32=filter(df_hess,R32 > quantile(R32, 0.025) & R32 < quantile(R32, 0.975))$R32,
```

```
  R33=filter(df_hess,R33 > quantile(R33, 0.025) & R33 < quantile(R33, 0.975))$R33,
  R34=filter(df_hess,R34 > quantile(R34, 0.025) & R34 < quantile(R34, 0.975))$R34,
  R35=filter(df_hess,R35 > quantile(R35, 0.025) & R35 < quantile(R35, 0.975))$R35,
  R36=filter(df_hess,R36 > quantile(R36, 0.025) & R36 < quantile(R36, 0.975))$R36,
  R37=filter(df_hess,R37 > quantile(R37, 0.025) & R37 < quantile(R37, 0.975))$R37,
  R38=filter(df_hess,R38 > quantile(R38, 0.025) & R38 < quantile(R38, 0.975))$R38,
  R39=filter(df_hess,R39 > quantile(R39, 0.025) & R39 < quantile(R39, 0.975))$R39,
  R40=filter(df_hess,R40 > quantile(R40, 0.025) & R40 < quantile(R40, 0.975))$R40,
  R41=filter(df_hess,R41 > quantile(R41, 0.025) & R41 < quantile(R41, 0.975))$R41,
  R42=filter(df_hess,R42 > quantile(R42, 0.025) & R42 < quantile(R42, 0.975))$R42
) -> df_perturb

# if selected samples are numerically negative, we truncate them to zero
df_perturb[df_perturb<0] = 0
df_perturb %>% head

#create data frame for lower CI of Rt
outputmin = data.frame(df_perturb %>% summarize_all(min) %>% as.numeric)
colnames(outputmin) <- c("lowerCI")

outputmin <- data.frame(lowerCI=c(rep(outputmin$lowerCI[1], 5),
                                  rep(outputmin$lowerCI[2], 5),
                                  rep(outputmin$lowerCI[3], 5),
                                  rep(outputmin$lowerCI[4], 5),
                                  rep(outputmin$lowerCI[5], 5),
                                  rep(outputmin$lowerCI[6], 5),
                                  rep(outputmin$lowerCI[7], 5),
                                  rep(outputmin$lowerCI[8], 5),
                                  rep(outputmin$lowerCI[9], 5),
                                  rep(outputmin$lowerCI[10], 5),
                                  rep(outputmin$lowerCI[11], 5),
                                  rep(outputmin$lowerCI[12], 5),
                                  rep(outputmin$lowerCI[13], 5),
                                  rep(outputmin$lowerCI[14], 5),
                                  rep(outputmin$lowerCI[15], 5),
                                  rep(outputmin$lowerCI[16], 5),
                                  rep(outputmin$lowerCI[17], 5),
                                  rep(outputmin$lowerCI[18], 5),
                                  rep(outputmin$lowerCI[19], 5),
                                  rep(outputmin$lowerCI[20], 5),
                                  rep(outputmin$lowerCI[21], 5),
                                  rep(outputmin$lowerCI[22], 5),
                                  rep(outputmin$lowerCI[23], 5),
                                  rep(outputmin$lowerCI[24], 5),
                                  rep(outputmin$lowerCI[25], 5),
                                  rep(outputmin$lowerCI[26], 5),
                                  rep(outputmin$lowerCI[27], 5),
                                  rep(outputmin$lowerCI[28], 5),
                                  rep(outputmin$lowerCI[29], 5),
                                  rep(outputmin$lowerCI[30], 5),
                                  rep(outputmin$lowerCI[31], 5),
                                  rep(outputmin$lowerCI[32], 5),
                                  rep(outputmin$lowerCI[33], 5),
                                  rep(outputmin$lowerCI[34], 5),
```

```
                                    rep(outputmin$lowerCI[35], 5),
                                    rep(outputmin$lowerCI[36], 5),
                                    rep(outputmin$lowerCI[37], 5),
                                    rep(outputmin$lowerCI[38], 5),
                                    rep(outputmin$lowerCI[39], 5),
                                    rep(outputmin$lowerCI[40], 5),
                                    rep(outputmin$lowerCI[41], 5),
                                    rep(outputmin$lowerCI[42], 9)))

#create data frame for upper CI of Rt
outputmax = data.frame(df_perturb %>% summarize_all(max) %>% as.numeric)
colnames(outputmax) <- c("upperCI")

outputmax <- data.frame(upperCI=c(rep(outputmax$upperCI[1], 5),
                                    rep(outputmax$upperCI[2], 5),
                                    rep(outputmax$upperCI[3], 5),
                                    rep(outputmax$upperCI[4], 5),
                                    rep(outputmax$upperCI[5], 5),
                                    rep(outputmax$upperCI[6], 5),
                                    rep(outputmax$upperCI[7], 5),
                                    rep(outputmax$upperCI[8], 5),
                                    rep(outputmax$upperCI[9], 5),
                                    rep(outputmax$upperCI[10], 5),
                                    rep(outputmax$upperCI[11], 5),
                                    rep(outputmax$upperCI[12], 5),
                                    rep(outputmax$upperCI[13], 5),
                                    rep(outputmax$upperCI[14], 5),
                                    rep(outputmax$upperCI[15], 5),
                                    rep(outputmax$upperCI[16], 5),
                                    rep(outputmax$upperCI[17], 5),
                                    rep(outputmax$upperCI[18], 5),
                                    rep(outputmax$upperCI[19], 5),
                                    rep(outputmax$upperCI[20], 5),
                                    rep(outputmax$upperCI[21], 5),
                                    rep(outputmax$upperCI[22], 5),
                                    rep(outputmax$upperCI[23], 5),
                                    rep(outputmax$upperCI[24], 5),
                                    rep(outputmax$upperCI[25], 5),
                                    rep(outputmax$upperCI[26], 5),
                                    rep(outputmax$upperCI[27], 5),
                                    rep(outputmax$upperCI[28], 5),
                                    rep(outputmax$upperCI[29], 5),
                                    rep(outputmax$upperCI[30], 5),
                                    rep(outputmax$upperCI[31], 5),
                                    rep(outputmax$upperCI[32], 5),
                                    rep(outputmax$upperCI[33], 5),
                                    rep(outputmax$upperCI[34], 5),
                                    rep(outputmax$upperCI[35], 5),
                                    rep(outputmax$upperCI[36], 5),
                                    rep(outputmax$upperCI[37], 5),
                                    rep(outputmax$upperCI[38], 5),
                                    rep(outputmax$upperCI[39], 5),
                                    rep(outputmax$upperCI[40], 5),
                                    rep(outputmax$upperCI[41], 5),
```

```
                                rep(outputmax$upperCI[42], 9)))

#combine the data-frames of lower CI and upper CI
osaka %<>% do(cbind(.,lowerCI=outputmin$lowerCI, upperCI=outputmax$upperCI))

#[Figure] plot Rt and its CIs along with epidemic curve
second_rate <- 40 #scaling enabling both incidence and Rt on the same figure
df <- data.frame(Cases = osaka$case, date = osaka$date, lowerCI =
osaka$lowerCI*second_rate, upperCI = osaka$upperCI*second_rate, Rt = osaka$Rt*second_rate)
ggplot(data=df)+
  geom_bar(mapping = aes(x=date, y=Cases),
           stat = "identity")+
  theme_classic() + #background will be white
  scale_x_date(date_labels = "%m/%d", date_breaks = "1 week") +
  theme(axis.text.x = element_text(angle = 60, hjust = 1))+
  labs(x="Date of illness onset",y="Incidence (cases)")  +
  geom_line(data = df, aes(x=date, y=Rt)) +
  geom_line(data = df, aes(x=date, y=lowerCI), linetype = "dashed")+
  geom_line(data = df, aes(x=date, y=upperCI), linetype = "dotted")+
  scale_y_continuous(
    limits = c(0,4300),
    sec.axis = sec_axis(~ . / second_rate, name = "Rt")
  ) +
  geom_line(data = df, aes(x=date, y=40))+ #add horizontal line indicating Rt of one
  coord_cartesian(ylim = c(0,200)) #zoom to exclude large Rt

#observed cases and estimated cases based on the model
#sampling of Rt from multivariate normal distribution
sample <- 1e3 #1000 samples of parameters
random_para = MASS::mvrnorm(n=sample, mu=sol$par[1:42],
                            Sigma=solve(sol$hessian[1:42, 1:42]),
                            tol=1e-06, empirical=FALSE, EISPACK=FALSE)
random_para %>% as.data.frame -> sample_para

sample_para[sample_para<0] = 0 #negative values will be handled as zero

sample_para <- as.data.frame(t(sample_para))

#1000 samples of expected cases based on the model
epcsample <- llik.pc(prm=sample_para[,1], prediction=TRUE)
for (i in 2:ncol(sample_para)) {
  epcsample  %<>% do(cbind(.,llik.pc(prm=sample_para[,i], prediction=TRUE))) }

#1000 random samples from Poisson distribution
rpois_ <- function(x){rpois(n=1, x)}
randompois <- as.data.frame(sapply(epcsample[,1], rpois_))
for (i in 2:ncol(epcsample)) {
  randompois  %<>% do(cbind(., sapply(epcsample[,i], rpois_)))} #predicted cases as column

randompois <- as.data.frame(t(randompois))

#get lower CI for cases
quantile_min <- function(x){quantile(
  x, probs = 0.025)}
```

```
lowercase <- sapply(randompois, quantile_min)

#get upper CI for cases
quantile_max <- function(x){quantile(
  x, probs = 0.975)}
uppercase <- sapply(randompois, quantile_max)

#data frame of 95%CI for cases
lowerupper <- data.frame(lowercase, uppercase)

# data frame of estimated cases
epc <- as.data.frame(llik.pc(prm = sol$par, prediction = TRUE))

#data frame with estimated cases and 95% CI
osaka %<>% do(cbind(.,epc))
osaka %<>% mutate(lowercase = lowerupper$lowercase)
osaka %<>% mutate(uppercase = lowerupper$uppercase)

#[Figure] plot of estimated and reported cases
ggplot(data=osaka)+
  geom_point(mapping = aes(x=date, y=case),
             stat = "identity")+
  theme_classic() + #background will be white
  scale_x_date(date_labels = "%m/%d", date_breaks = "2 week") +
  labs(x="Date of illness onset",y="Estimated and reported cases")  +
  geom_line(data = osaka, aes(x=date, y=epc)) +
  geom_line(data = osaka, aes(x=date, y=lowercase), linetype = "dotted")+
  geom_line(data = osaka, aes(x=date, y=uppercase), linetype = "dotted")
```

R code for Section 2 (Tokyo cluster):

```
### Code for yakatabune and sagamihara###
###   26 Dec 2021   ###
rm(list=ls())
library(BB)
library(Bhat)
library(MASS)
offspring_data <- read.csv("yakatasagami_offspring.csv", header=T)

## setting ##
offspring_data <- na.omit(offspring_data)
r <- offspring_data[,1] # offspring
sym <- offspring_data[,2] # symptomatic status(0:asymptomatic, 1:symptomatic)

## likelihood function ##
nlogl <-function(x){
         z=0
         ## parameters for estimation
         R0=x[1]
         v=x[2] # parameter for relative effect of asymptomatic on R0
         #beta=x[3]

         ## offspring ##
         for(i in 1:nrow(offspring_data)){
                  if(sym[i]==1){
                           R=R0
                  }else
                    {
                           R=v*R0}
           #z=z-log(dpois(r[i],R))
                  z=z-log(dgeom(r[i],1/(R+1)))
                  #z=z-log(dnbinom(r[i],beta,beta/(beta+R)))
         }
         print(z)
         return(z)
}

## statistical estimation ##
## for poisson or geometric model
param <-list(label=c("R", "v"),est=c(1.6869878,
0.4350158),low=c(0.01,0.01),upp=c(20,0.999))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*2

## for negative binomial model
param <-list(label=c("R", "v","beta"),est=c(3.6869878, 0.750158,
0.6521904),low=c(0.01,0.01,0.01),upp=c(20,0.999,10))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*3
```

R code for Section 2 (Kyoto cluster):

```
#### main analysis ####
rm(list=ls())
library(BB)
library(Bhat)
library(MASS)
serial_data <- read.csv("kyoto_serial_interval_1.csv", header=T)
offspring_data <- read.csv("kyoto_offspring_1.csv", header=T)

## setting ##
serial_data <- na.omit(serial_data)
t <- serial_data[,1] # serial interval
r <- offspring_data[,1] # offspring
sym <- offspring_data[,2] # symptomatic status(0:asymptomatic, 1:symptomatic)

## discretization ##
dlnormd <- function(x,alpha,beta){
          y=(plnorm(x+1,alpha,beta)-plnorm(x,alpha,beta))
          return(y)
}
dexpd <- function(x,alpha){
          y=(pexp(x+1,alpha)-pexp(x,alpha))
          return(y)
}
dgammad <- function(x,alpha,beta){
          y=(pgamma(x+1,alpha,beta)-pgamma(x,alpha,beta))
          return(y)
}
dweibulld <- function(x,alpha,beta){
          y=(pweibull(x+1,alpha,beta)-pweibull(x,alpha,beta))
          return(y)
}

f <- dlnormd(1:200,1.43,0.66) # incubation period
m <- 6

## estimation of parameters of pdf of infectiousness ##
nlogl <-function(x){
  z=0
  ## parameters for estimation
  alpha1=x[1] # parameters for h (h is the distribution of infectiousness relative to
disease-age)
  alpha2=x[2] # parameters for h
  h <- dgammad(1:200,alpha1,alpha2) #the distribution of infectiousness relative to
disease-age
  #h <- dweibulld(1:200,alpha1,alpha2)

  ## serial interval ##
  for(i in 1:nrow(serial_data)){
    s <- numeric(100)
    for(j in 1:100){
      tau=1
      while(tau<j){
        s[j]=s[j]+h[tau]*f[j-tau]
```

```
        tau=tau+1
      }
    }
    s=s/sum(s) # normalization
    z=z-log(s[t[i]+m])
  }

  print(z)
  return(z)
}

param <-list(label=c("alpha1","alpha2"),est=c(5.9097840,
0.9811242),low=c(0.01,0.01),upp=c(200,10))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*2

## estimation of relative infection (negative-binomial and constant model) ##
nlogl <-function(x){

          z=0
          ## parameters for estimation
          R0=x[1]
          v=x[2] # parameter for relative effect of asymptomatic on R0
          beta=x[3]
          #delta=x[4] # the exponential rate of reproduction number

          ## offspring ##
          for(i in 1:nrow(offspring_data)){
            if(sym[i]==1){
              R=R0
              #R=R0*exp(-(delta*d1[i]))
            }else
            {
              R=v*R0}
            #R=v*(R0*exp(-(delta*d1[i])))}

            z=z-log(dnbinom(r[i],beta,beta/(beta+R)))
          }
          print(z)
          return(z)
}

param <-list(label=c("R", "v", "beta"),est=c(1.6869878, 0.4350158,
0.2521904),low=c(0.01,0.01,0.01),upp=c(20,0.999,10))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*3

## estimation of relative infection (negative-binomial and time-dependent model) ##
offspring_data <- na.omit(offspring_data)
d1 <- offspring_data[,4] # transmission day as a calendar date
nlogl <-function(x){

  z=0
  ## parameters for estimation
```

```
  R0=x[1]
  v=x[2] # parameter for relative effect of asymptomatic on R0
  beta=x[3]
  delta=x[4] # the exponential rate of reproduction number

  ## offspring ##
  for(i in 1:nrow(offspring_data)){
    if(sym[i]==1){
      #R=R0
      R=R0*exp(-(delta*d1[i]))
    }else
    {
      #R=v*R0}
    R=v*(R0*exp(-(delta*d1[i])))}

    z=z-log(dnbinom(r[i],beta,beta/(beta+R)))
  }
  print(z)
  return(z)
}

param <-list(label=c("R", "v", "beta", "delta"),est=c(1.6869878, 0.4350158, 0.2521904,
0.05),low=c(0.01,0.01,0.01,0.01),upp=c(40,0.999,10,0.99))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*4

#sensitivity analysis of k on R0 and v
nlogl <-function(x){

  z=0
  ## parameters for estimation
  R0=x[1]
  v=x[2] # parameter for relative effect of asymptomatic on R0
  beta= 10

  ## offspring ##
  for(i in 1:nrow(offspring_data)){
    if(sym[i]==1){
      R=R0
    }else
    {
      R=v*R0}

    z=z-log(dnbinom(r[i],beta,beta/(beta+R)))
  }
  print(z)
  return(z)
}

param <-list(label=c("R", "v"),est=c(1.6869878,
0.4350158),low=c(0.01,0.01),upp=c(20,0.999))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*2
```

```
### estimation of proportion of symptomatic cases depending on primary symptomatic status
###
rm(list=ls())
library(Bhat)
offspring_data <- read.csv("kyoto_secondary.csv", header=T)

## setting ##
sym <- offspring_data[,3] # symptomatic status(0:asymptomatic, 1:symptomatic)
rs <- offspring_data[,4] # number of symptomatic secondary cases
ra <- offspring_data[,5] # number of asymptomatic secondary cases

## likelihood function for offspring (negative binomial model) ##
nlogl <-function(x){

  z=0
  ## parameters for estimation
  R1=x[1] #reproductive number of secondary cases by primary symptomatic case
  v=0.2 #reduction of secondary cases by primary asymptomatic case
  p=x[2] # proportion of symptomatic cases among secondary cases due to primary symptomatic
case
  q=x[3] # proportion of symptomatic cases among secondary cases due to primary
asymptomatic case
  k=x[4] # dispersion parameter

  ## offspring ##
  for(i in 1:nrow(offspring_data)){
    if(sym[i]==1){
      R0=R1
      sp=p
    }else
    {
      R0=v*R1
      sp=q
    }

    z=z-log(dnbinom(rs[i]+ra[i],k,prob=k/(k+R0)))-log(dbinom(rs[i], rs[i]+ra[i], sp))
  }
  print(z)
  return(z)}

param <-list(label=c("R1", "p", "q", "k"),est=c(1.6869878, 0.2, 0.3,
0.2521904),low=c(0.01,0.01, 0.01,0.01),upp=c(20,0.999,0.999,10))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*4


#### supplemental analysis ####
rm(list=ls())
library(BB)
library(Bhat)
library(MASS)
serial_data <- read.csv("kyoto_serial_interval_1.csv", header=T)
offspring_data <- read.csv("kyoto_offspring_1.csv", header=T)
```

```
## setting ##
serial_data <- na.omit(serial_data)
offspring_data <- na.omit(offspring_data)
t <- serial_data[,1] # serial interval
k <- serial_data[,2] # isolation day of primary case relative to illness onset
r <- offspring_data[,1] # offspring
sym <- offspring_data[,2] # symptomatic status(0:asymptomatic, 1:symptomatic)
k1 <- offspring_data[,3] # isolation day relative to illness onset
d1 <- offspring_data[,4] # onset of infectiousness as a calendar date
m <-6 # a known disease-age (usually before onset of disease) when infected individual
acquires infectiousness.
# the value should be chosen so that (m + the smallest serial interval) will be more than
1; otherwise log will be infinite

## discretization ##
dlnormd <- function(x,alpha,beta){
  y=(plnorm(x+1,alpha,beta)-plnorm(x,alpha,beta))
  return(y)
}
dexpd <- function(x,alpha){
  y=(pexp(x+1,alpha)-pexp(x,alpha))
  return(y)
}
dgammad <- function(x,alpha,beta){
  y=(pgamma(x+1,alpha,beta)-pgamma(x,alpha,beta))
  return(y)
}
dweibulld <- function(x,alpha,beta){
  y=(pweibull(x+1,alpha,beta)-pweibull(x,alpha,beta))
  return(y)
}

# log-normal distribution for incubation period
# from Li et al NEJM 2020
# lognormal mean = 5.2; 95% CI = c(4.1, 7.0)
# logmean = 1.434065
# logsd = 0.6612
f <- dlnormd(1:200,1.43,0.66)

## likelihood function ##
nlogl <-function(x){

  z=0
  ## parameters for estimation
  R0=x[1]
  epsilon=x[2]
  v=x[3] # parameter for relative effect of asymptomatic on R0
  alpha1=x[4] # parameters for h (h is the distribution of infectiousness relative to
disease-age)
  alpha2=x[5] # parameters for h
  beta=x[6]
  delta=x[7] # the exponential rate of reproduction number
  h <- dgammad(1:200,alpha1,alpha2) #the distribution of infectiousness relative to
disease-age
```

```
  #h <- dweibulld(1:200,alpha1,alpha2)

  ## serial interval ##
  for(i in 1:nrow(serial_data)){
    ## distribution of g (g is the distribution of infectiousness accounting for
isolation)##
    g <- NULL
    if(k[i]==999){
      g=h
    }else{
      for(j in 1:100){
        if(j<(k[i]+m)){
          g[j]=h[j]/(pgamma(k[i]+m,alpha1,alpha2)+(1-epsilon)*(1-
pgamma(k[i]+m,alpha1,alpha2)))
        }else{
          g[j]=(1-epsilon)*h[j]/(pgamma(k[i]+m,alpha1,alpha2)+(1-epsilon)*(1-
pgamma(k[i]+m,alpha1,alpha2)))
        }
      }
    }
    g=g/sum(g) # normalization
    ## distribution of biased s (s is the serial interval) ##
    s <- numeric(100)
    for(j in 1:100){
      tau=1
      while(tau<j){
        s[j]=s[j]+g[tau]*f[j-tau]
        tau=tau+1
      }
    }
    s=s/sum(s) # normalization
    z=z-log(s[t[i]+m])
  }

  ## generation interval##
  l <- numeric(100)
  for (j in 1:100) {
    tau=1
    while (tau<j) {
      l[j]=l[j]+h[j-tau]*f[tau]
      tau=tau+1
    }
  }
  l=l/sum(l) # normalization


  ## offspring ##
  for(i in 1:nrow(offspring_data)){
    if(k1[i]==999){
      R=R0
      #R=R0*exp(-(delta*d1[i]))
    }else
    {
      R=R0*(pgamma(k1[i]+m,alpha1,alpha2)+(1-epsilon)*(1-pgamma(k1[i]+m,alpha1,alpha2)))
```

```
      #R=R0*exp(-(delta*d1[i]))*(pgamma(k1[i]+m,alpha1,alpha2)+(1-epsilon)*(1-
pgamma(k1[i]+m,alpha1,alpha2)))
    }
    if(sym[i]==0) { # asymptomatic
      R=v*R0*(sum(l[1:k1[i]])+(1-epsilon)*(1-sum(l[1:k1[i]])))}
    #R=v*(R0*exp(-(delta*d1[i])))*(sum(l[1:k1[i]])+(1-epsilon)*(1-sum(l[1:k1[i]])))}

    z=z-log(dnbinom(r[i],beta,beta/(beta+R)))
  }
  print(z)
  return(z)
}

## statistical estimation ##
## for negative binomial model
param <-list(label=c("R","epsilon", "v","alpha1","alpha2", "beta"),est=c(1.6869878,
0.78985, 0.4350158, 5.9097840, 0.9811242,
0.2521904),low=c(0.01,0.001,0.01,0.01,0.01,0.01),upp=c(20,0.999,0.999, 10,10,10))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*6

## for negative binomial and time-dependent reproductive number model
param <-list(label=c("R","epsilon", "v","alpha1","alpha2", "beta",
"delta"),est=c(1.6869878, 0.6478985, 0.4350158, 5.9097840, 0.9811242, 0.2521904,
0.1),low=c(0.01,0.001,0.01,0.01,0.01,0.01,0.01),upp=c(20,0.999,0.999, 100,10,10,0.99))
fit <- dfp(param, nlogl)
AIC <- 2*fit$fmin+2*7
```