Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

International Journal of Infectious Diseases 105 (2021) 286–292

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Localized end-of-outbreak determination for coronavirus disease 2019 (COVID-19): examples from clusters in Japan

Natalie M. Linton[a,b], Andrei R. Akhmetzhanov[a,c], Hiroshi Nishiura[a,b,*]

[a] Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome, Kita-ku, Sapporo-shi, Hokkaido, 060-8638, Japan
[b] Kyoto University School of Public Health, Yoshidakonoecho, Sakyoku, Kyoto, 606-8501, Japan
[c] College of Public Health, National Taiwan University, 17 Xu-Zhou Road, Taipei, 10055, Taiwan

ARTICLE INFO

Article history:
Received 28 November 2020
Received in revised form 25 February 2021
Accepted 25 February 2021

Keywords:
epidemic
mathematical model
extinction
epidemiology
elimination
transmission dynamics

ABSTRACT

Objectives: End-of-outbreak declarations are an important component of outbreak response because they indicate that public health and social interventions may be relaxed or lapsed. Our study aimed to assess end-of-outbreak probabilities for clusters of coronavirus disease 2019 (COVID-19) cases detected during the first wave of the COVID-19 pandemic in Japan.

Methods: A statistical model for end-of-outbreak determination, which accounted for reporting delays for new cases, was computed. Four clusters, representing different social contexts and time points during the first wave of the epidemic, were selected and their end-of-outbreak probabilities were evaluated.

Results: The speed of end-of-outbreak determination was most closely tied to outbreak size. Notably, accounting underascertainment of cases led to later end-of-outbreak determinations. In addition, end-of-outbreak determination was closely related to estimates of case dispersion$k$ and the effective reproduction number R_e. Increasing local transmission ($R_e > 1$) leads to greater uncertainty in the probability estimates.

Conclusions: When public health measures are effective, lowerR_e (less transmission on average) and larger k (lower risk of superspreading) will be in effect, and end-of-outbreak determinations can be declared with greater confidence. The application of end-of-outbreak probabilities can help distinguish between local extinction and low levels of transmission, and communicating these end-of-outbreak probabilities can help inform public health decision making with regard to the appropriate use of resources.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic has been the most fatal and disruptive biological disaster in recent history. As of 24 February 2021, COVID-19 had been diagnosed in over 113 million people and associated with more than 2.5 million deaths. Travel restrictions, school closures, cancellation of public events, and other public health and social measures implemented to curb disease transmission have deeply affected lives and livelihoods worldwide. In the midst of the disaster, mathematical modeling has served prominently in informing pandemic response as outbreaks have erupted and grown (Anderson et al., 2020; Ferguson et al., 2020), but the field can also provide insight into the transmission dynamics of outbreaks as they come to an end (Thompson et al., 2019; Lee and Nishiura, 2019). Control of COVID-19 has proven difficult, and timely, localized information regarding which outbreaks are growing and which are likely to end is needed to inform control measures. Here, we explain a method to estimate end-of-outbreak probabilities at localized levels, allowing for evidence-based decision making around the scaling-back of public health and social response measures in real time. Our study demonstrates the applicability of this method in relation to clusters of cases, using several examples from Japan.

Early research into the transmission dynamics of COVID-19 indicated that spread of the causal virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was highly overdispersed (Endo et al., 2020; Lloyd-Smith et al., 2005). The degree of overdispersion is quantified by the dispersion parameter k, which describes the variance in the distribution of the number of secondary cases infected by a typical primary case. Lower values of

* Corresponding author at: Kyoto University School of Public Health, Yoshida-konoecho, Sakyoku, Kyoto, 606-8501, Japan.
E-mail addresses: nlinton@hyg.med.kyoto-u.ac.jp (N.M. Linton), akhmetzhanov@ntu.edu.tw (A.R. Akhmetzhanov), nishiura.hiroshi.5r@kyoto-u.ac.jp (H. Nishiura).

https://doi.org/10.1016/j.ijid.2021.02.106

k represent greater variance and thereby a propensity towards superspreading — the phenomenon by which some infected individuals transmit a pathogen to large numbers of secondary cases, while most do not infect others (Lloyd-Smith et al., 2005). Larger values of k indicate less dispersion, and a Poisson distribution is obtained as a special case of the negative binomial distribution as $k \to \infty$ (Lloyd-Smith, 2007). During the pandemic, k for COVID-19 has been reported in the range of 0.1–0.6 (Endo et al., 2020; Bi et al., 2020; Riou and Althaus, 2020; Tariq et al., 2020), indicating that most secondary infections are caused by a small fraction of primary cases, and therefore that superspreading events can fuel disease transmission.

In Japan, the propensity of COVID-19 towards superspreading was addressed by including the prevention, detection, and suppression of clusters — groups of cases linked by a common place and time — as a key component of the national response (Japan National COVID-19 Cluster Taskforce, 2020). Additional public health and social measures took into consideration commonalities in transmission settings between clusters (Furuse et al., 2020) — most prominently in the form of a nationwide messaging campaign encouraging residents to avoid the 'three Cs': closed spaces with poor ventilation, crowded places, and close-contact situations (Japan National COVID-19 Cluster Taskforce, 2020; Prime Minister's Office of Japan and Ministry of Health Labour and Welfare, 2021). Although Japan had reported relatively low levels of epidemic growth and fatalities during the first 6 months of the pandemic compared with other developed nations in Europe and North America, it faced a resurgence of cases during the summer months.

Current World Health Organization (WHO) guidance focuses on assigning levels of risk based on epidemiological, health-system, and surveillance criteria from trends and other descriptive data to inform the loosening and re-tightening of response activities (World Health Organization (WHO), 2020a). The guidance indicates that an outbreak can be considered controlled if the effective reproduction number R_e — the average number of secondary cases produced per primary case in the presence of interventions and immune individuals in a given time period — is maintained below the threshold value of 1 for at least 2 weeks. While this distinguishes well between epidemic growth and decline, this method of surveillance does not differentiate between whether an outbreak will end entirely or will continue as stuttering chains of transmission (i.e. $0 < R_e < 1$) before potentially resurging ($R_e > 1$) (Blumberg and Lloyd-smith, 2013). Other COVID-19 guidance that addresses end-of-outbreak declarations relates to the incubation period-based guidelines used for other directly transmittable diseases, such as measles and Ebola virus disease. For those diseases, two times the maximum incubation period — the time from infection to symptom onset — since the last possible date of exposure to a source of infection within the outbreak is used to determine the time until the end of an outbreak can be declared (Public Health Agency of Canada, 2020; World Health Organization (WHO), 2020b). Despite widespread use, this method has previously been proven flawed, and is particularly vulnerable when local surveillance systems are weak (Thompson et al., 2019; Lee and Nishiura, 2017).

An alternative, statistically based method is to leverage parameters that describe transmission dynamics underlying the epidemic curve, in order to estimate the probability that one or more cases will be reported going forward in time (Nishiura et al., 2016). This more rigorous basis for end-of-outbreak determination can also be adjusted to account for additional factors, such as case underascertainment and changing transmission dynamics. Including this statistical analysis adds to the information available to politicians and public health officials when making decisions regarding response needs. Here, we present examples of where this method was used to assess end-of-outbreak probabilities for clusters in Japan during the first wave of the pandemic.

Methods

Data collection and cluster selection

Epidemiological data were collected from the official case reports published online by reporting jurisdictions (prefectures and some cities) within Japan. In some instances, the data were supplemented and additionally verified using information from press briefings from the reporting jurisdictions. The collected data included date of onset, date of report, and linkage information used to support the grouping of cases into clusters. A dataset, including dates and cluster information, is included in the supplementary materials.

For this study, four clusters representing different social contexts and time points during the epidemic were selected. The first cluster occurred early in the pandemic in Aichi Prefecture, with the first case reported on February 14, 2020 (Ministry of Health Labour and Welfare (MHLW), 2020). This involved chains of transmission related to fitness gym use and social contact between cases. The second cluster was identified in Kyoto Prefecture in March, and was initiated by importation of cases exposed to SARS-CoV-2 in Europe. Subsequent domestic transmission occurred during a series of events attended by university students and other contacts. The third and fourth clusters were linked to nosocomial transmission in a medical facility and a senior care facility in Hokkaido Prefecture, with cases reported during April and May, respectively. Clusters included all secondary cases arising from chains of transmission linking back to the original sources of common exposure.

Statistical model

In the model used, each cluster is given by the epidemic curve represented at the time of report t, with onset dates $t_i \leq t$ for an epidemiologically defined group of cases $i = \{1, \ldots, M\}$. The probability that one or more new cases $X(t)$ will be reported on day t is written as follows: (Nishiura et al., 2016)

$$\mathrm{Prob}(X(t) > 0) = 1 - \prod_{i=1}^{M} \sum_{y=0}^{\infty} p_y [F(t - t_i)]^y \quad (1)$$

where p_y is the probability that y secondary cases arise from a given primary case i, following a negative binomial distribution with the mean R_e and variance $R_e(1 + \frac{R_e}{k})$, with k the dispersion parameter.

The function $F(.)$ defines the cumulative distribution function (CDF) of the serial interval f, which is backprojected using the time delay from illness onset to report h, and defined by the convolution:

$$F(t - t_i) = \sum_{s=2}^{t-t_i} \sum_{\tau=1}^{s-1} f(s - \tau) h(\tau) \quad (2)$$

Outbreak extinction is determined once the estimate for the probability of observing one or more additional cases $\mathrm{Prob}(X(t) > 0)$ becomes lower than a given threshold. The day the probability estimate drops below the threshold is, in effect, the day the outbreak would be declared over. Selection of a threshold value depends on whether the goal is to minimize the observation period (higher threshold) or minimize the risk that undetected cases may exist and become detected following the end of outbreak declaration (lower threshold). In this study, we examined

a 5% threshold, which translates to a 5% risk that the end-of-outbreak declaration would be preemptive, and case(s) would be detected following the declaration (Lee and Nishiura, 2019).

Missing dates of illness onset

For some cases, the date of onset was unavailable. Either the case did not give permission for disclosure, the information was not collected, or the case was asymptomatic at the time of detection and did not later report symptoms. The latter scenario can therefore represent either presymptomatic cases with no follow-up reports or cases who were completely asymptomatic until recovery.

We approached the missing dates of illness onset in two ways. First, we excluded cases with no reported onset date and calculated the probability of new cases based solely on the existing epidemic curve of illness onsets, hereafter referred to as the 'reported' dataset. Second, we sampled the reporting delays and subtracted their values from the known report dates for all cases with no available date of illness onset to obtain a proxy onset date, hereafter referred to as the 'imputed' dataset.

Underascertainment of cases

Despite efforts to obtain high-quality surveillance data through contact tracing and testing, it is still likely that cases remain underascertained. Previous reports have suggested 9.2–44.4% case ascertainment, using data on Japanese evacuees from the original epicenter of Wuhan, China and laboratory testing conducted in Japan during January and February 2020, respectively (Nishiura et al., 2020a; Omori et al., 2020). For clusters with relatively stable and captive populations (e.g. medical centers and senior homes), ascertainment is likely to be higher due to intensive contact tracing on a focused population, although as chains of transmission move away from the common exposure setting, ascertainment will approach that of the general population. Clusters based on social contact linked to > 1 common exposure setting (e.g. multiple restaurants, gyms, or other venues) are more likely to have lower levels of case ascertainment; however, in Japan it is expected that ascertainment for cases related to a cluster would be higher than for the general population due to targeted case finding.

To address likely case underascertainment, we sampled from a binomial distribution with probability of success $p = 1 - q$, where q was the ascertainment rate. We assumed that underascertained cases could only exist within one serial interval (i.e. 5 days (Nishiura et al., 2020b)) from the date of onset of the last reported case because the contact tracing team was unlikely to miss cases from two consecutive generations. A number of unreported cases u_τ at day τ can be then inferred, using the following observational model:

$$i_\tau \sim \text{Binomial}(\text{size} = u_\tau + i_\tau, \text{prob} = p) \quad 0 \leq u_\tau \leq U, \quad \tau \leq \max(t_i) + 5 \tag{3}$$

where U is a maximally possible number of unreported cases, which was assigned to 50 in our simulations.

Parameter selection and statistical analysis

The model was applied using data on the epidemic curves for individual clusters, accounting for missing dates of onset and varying levels of case ascertainment, as described previously. The applied parameter values are shown in Table 1. R_e was explicitly varied between 0.5, 1.5, and 3, although an estimate of the local time-varying effective reproduction number (R_t) would be a sensible option when conducting analyses in real time. Otherwise, we accounted for parameter uncertainties via resampling. Estimates for k were drawn from a positive half-normal distribution using mean and SD values from published studies (Bi et al., 2020; Tariq et al., 2020; Zhang et al., 2020). Other distributions were also resampled from their empirical distributions. The serial interval distribution was previously reported by Nishiura et al. (Nishiura et al., 2020b), and the reporting delay was estimated from all cases reported in Japan through to the end of May, using doubly interval-censored methods described elsewhere (Nishiura et al., 2020b; Linton et al., 2020). The reporting delay was estimated from cases reported through to the end of May, to coincide with the time interval during which cases for the four clusters were reported. Analysis was implemented using R 4.0.3 and CmdStan 2.26.1 (R Development Core Team, 2019). Reproducible code for this study is available on GitHub at https://github.com/nlinton/covid19_eoo.

Ethical considerations

Our study analyzed publicly available data, which had already been de-identified upon press release. The study was approved by the Medical Ethics Board of the Graduate School of Medicine, Kyoto University (R2673).

Results

Characteristics of the four clusters are shown in Table 2. The Aichi fitness gyms cluster had the smallest number of cases ($n = 40$) while the Hokkaido senior care facility cluster was the largest ($n = 94$). The average age was lowest for the Kyoto cluster, which was associated with parties attended primarily by university students, while the average age was highest for the senior care facility cluster. Three-quarters of the Hokkaido cancer center cluster were female, while the female-to-male ratio for the gyms and university-related parties cluster were nearly evenly split between males and females. The proportion of cases with no reported date of onset ranged from 2.5% for the Aichi gyms cluster to 46.5% for the Kyoto university-related parties cluster. The time between first onset and last onset within each cluster ranged from 22 to 43 days.

The delay from onset to prefecture report date for cases reported between when the first case was detected in January and May 31, 2020 was estimated at 7.2 days (95% credible interval [CrI]: 7.1–7.3 days), using the best-fit gamma distribution (Table S1 and Figure S1). Figure 1 depicts the probability distributions for observing additional cases by cluster, varyingly accounting for asymptomatic cases and missing dates of onset (imputed dataset) and underascertainment of cases.

Table 1
Parameters used in the statistical model.

Description	Distribution	Examined values	References
Offspring distribution	Negative binomial	R_e: 0.5, 1.5, 3.0 k (SD): 0.11 (0.05), 0.25 (0.19), 0.58 (0.26)	Park et al. (2020) and Bi et al. (2020); Tariq et al. (2020); Zhang et al. (2020)
Serial interval	Weibull	Mean: 4.8 (SD: 2.3)	Nishiura et al. (2020b)
Reporting delay	Gamma	Mean: 7.2 (SD: 4.7)	Estimated

R_e: effective reproduction number; k: overdispersion parameter; SD: standard deviation.

Table 2
Descriptive information regarding the four case clusters.

Characteristics	Fitness gyms	University parties	Senior care	Cancer center
Source of index case infection	Imported	Imported	Domestic	Domestic
Number of cases	40	72	94	92
Age in years,[a] mean (range)	59 (20–80)	30 (0–70)	72 (20–100)	50 (20–80)
Female (%)	19 (47.5)	33 (45.8)	71 (75.5)	62 (67.4)
Date of onset not reported (%)	1 (2.5)	33 (45.8)	34 (36.2)	24 (26.1)
First to last onset	35 days	22 days	35 days	43 days

[a] Ages were reported in deciles. Source of index case infection 'imported' indicates that the index case(s) for the cluster reported international travel and were likely infected while abroad.

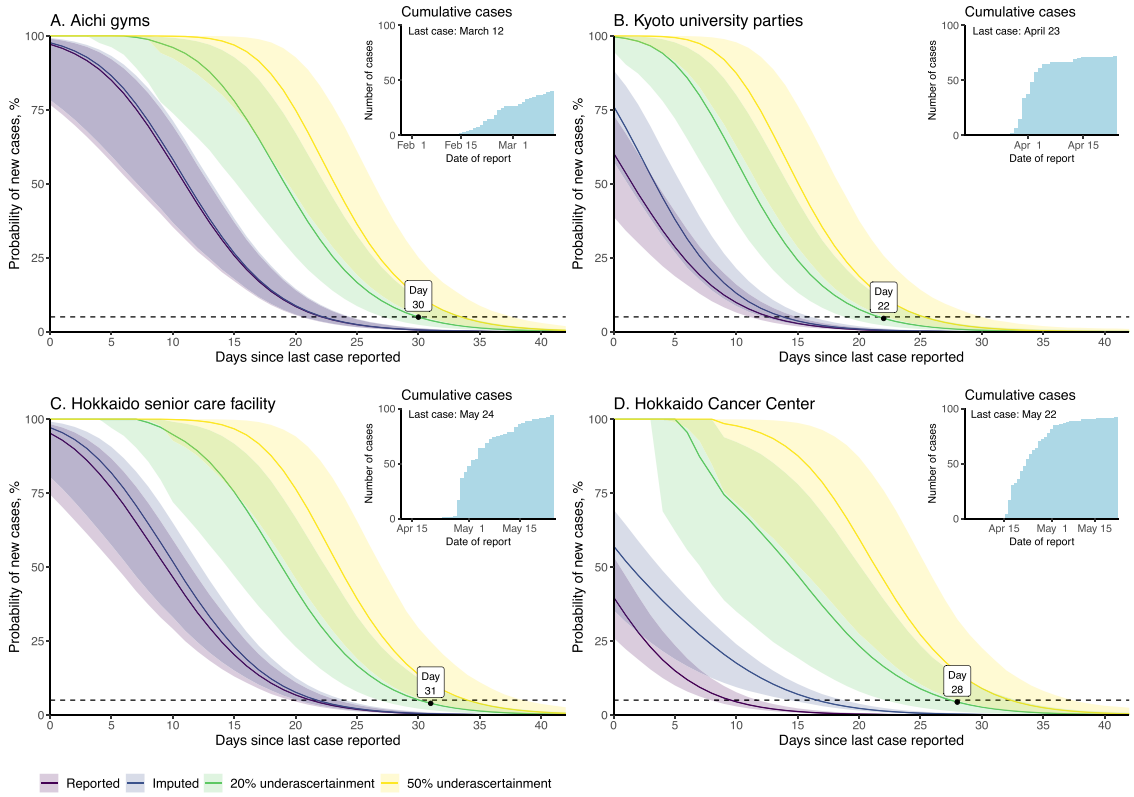

Figure 1. End-of-outbreak probabilities for four coronavirus disease 2019 (COVID-19) case clusters in Japan.
Each subfigure begins on the last date of onset within the cluster. All plots assume $R_e = 0.5$ and $k = 0.25$. Lines are median values and shaded areas are 95% credible intervals (CrI) for the datasets. Purple represents the datasets including only reported dates of onset; indigo represents the datasets including imputed dates of onset; yellow represents the datasets accounting for 20% underascertainment of cases; green represents the datasets accounting for 50% underascertainment of cases. The horizontal line represents the threshold for 5% probability of failure of the model. Cumulative case counts over time for each cluster are shown in the inset figures.

Table 2 presents the estimated observation period in days for each cluster, based on the varying parameter values. End-of-outbreak determination was most closely tied with the size of the outbreak. The datasets with more cases (accounting for missing onsets and underascertainment of cases) reached the proscribed probability thresholds at later dates compared with the datasets based on the original epidemic curve (the reported dataset). Studies on Ebola virus disease likewise found that lower ascertainment of cases indicated that more time is needed to be sufficiently certain that the end of an outbreak is declared appropriately (Thompson et al., 2019; Lee and Nishiura, 2019). Moreover, a larger R_e consistently resulted in slightly longer observation periods compared with smaller R_e values (Table 3 and Figures S2.1–2.4). The CrIs were widest for combinations of large R_e and small k, indicating greater uncertainty over whether the outbreak had truly ended. The probabilities for some of the upper 95% CrIs never dropped below the threshold values within the 42-day periods examined.

Discussion

As guidance continues to be developed for COVID-19 responses, it is important to incorporate insights from statistical modeling for declining phases of the pandemic (Thompson et al., 2020). In this regard, localized, end-of-outbreak declarations are a valuable component of outbreak responses because they indicate that public health interventions may be relaxed or lapsed. Statistical models that can provide insight into levels of certainty of the decline of an outbreak are useful for dynamically optimizing the timing of such declarations.

Our application of a transmission characteristic-based statistical model to COVID-19 clusters found that confidence in end-of-outbreak determination was most closely tied to estimates of R_e and, which in turn reflected local levels of control. If public health measures are effective, we would expect to see lower R_e (less transmission on average) and larger k (lower risk of superspreading due to measures targeted to prevent superspreading

Table 3
Length of the observation period in days from last date of report of a case to estimated end of transmission for four coronavirus disease 2019 (COVID-19) case clusters in Japan.

Cluster			A	B	C	D
Parameters	R_e	k	Probability threshold (5%)			
Reported onsets	0.5	0.11	23	13	22	10
		0.25	23	13	22	10
		0.58	23	13	22	10
	1.5	0.11	27	17	26	14
		0.25	27	17	26	14
		0.58	27	17	26	14
	3.0	0.11	29	20	29	17
		0.25	29	20	28	17
		0.58	29	20	28	17
Imputed onsets	0.5	0.11	23	14	22	17
		0.25	23	14	22	17
		0.58	23	14	22	17
	1.5	0.11	27	18	26	21
		0.25	27	18	26	21
		0.58	27	18	26	22
	3.0	0.11	30	21	29	24
		0.25	29	21	29	24
		0.58	29	21	29	24
20% ascertainment	0.5	0.11	30	22	30	28
		0.25	30	22	30	28
		0.58	30	22	31	28
	1.5	0.11	35	27	35	33
		0.25	35	26	35	32
		0.58	34	26	35	32
	3.0	0.11	–	–	–	–
		0.25	38	29	38	35
		0.58	37	29	37	34

Cluster A: fitness gyms cluster; Cluster B: university parties cluster; Cluster C: senior care facility; Cluster D: cancer center cluster. Observation days are reported relative to the last date of report of a case in the cluster (day 0). Cells with '– 'did not reach the proscribed threshold probability within the 42-day period analyzed.

events), and have more confidence in declaring the end of the outbreak. When considering end-of-outbreak scenarios we would generally anticipate $R_e < 1$. However, if no or lax public health interventions are implemented, R_e may be > 1. Under these circumstances, there is less certainty as the end-of-outbreak probabilities approach zero (see supplementary figures). Under some assumptions of R_e and k, we estimated that end-of-outbreak declarations could be made before 28 days (two times the approximate maximum incubation period for COVID-19) (Public Health Agency of Canada, 2020; Linton et al., 2020) had passed from the date of last reporting of any case (which was often the same day the case was isolated). Using these estimates would potentially allow for earlier end-of-outbreak declarations, leading to saved resources.

Parameters such as reporting delay and serial interval can also vary throughout the epidemic. When surveillance is heightened the reporting delay may be shorter than when surveillance systems are overwhelmed. Similarly, nonpharmaceutical interventions, such as contact tracing, isolation, and physical distancing, change contact patterns and limit the time during which an infectious case may be in contact with susceptible individuals, shortening the serial interval (Ali et al., 2020). Although these possible variations were not accounted for in this study, they can be incorporated if deemed to be of value to inform decisions regarding the continuation of public health and social response measures.

The size and scale of the epidemic curves used in our analyses depended on case, cluster, and outbreak definitions, as well as case ascertainment by the surveillance system. When a cluster definition is limited to cases linked directly to a location or activity (e.g. a hospital or an event) then the cluster size will be smaller than if the cluster includes all secondary infections of household members and other contacts not directly related to the transmission event(s) that defined the cluster. Likewise, when the cases are limited to those whose samples test positive (i.e. via polymerase chain reaction [PCR] or antigen testing) the scope and scale of the epidemic curve will be smaller than if other probable cases were included in the outbreak case definition. The clusters reported here include original cluster cases as well as all subsequent cases in chains of transmission reported by local public health jurisdictions. It is possible that some cases associated with the cluster were missed or incorrectly attributed; however, we repeatedly reviewed the data to minimize these possibilities. Our analyses accounting for possible underascertainment of cases likewise showed that, when accounting for missed cases, the observation period before reaching our 5% threshold was extended.

Up to the end of May 2020, only PCR-positive cases were included in the case definition for COVID-19 cases in Japan. Infected individuals may not have been tested if they were never suspected of being a case or did not meet testing criteria (Ministry of Health Labour and Welfare (MHLW), 2021). In addition, PCR sensitivity is less than perfect, reducing to around 70% more than 1 week after symptom onset (Miller et al., 2020), so some infected individuals may have received a false negative test result if their viral load at the time they were tested was insufficient to trigger a positive result. Sequentially repeated PCR testing in Japan for persons with persistent symptoms and/or new onset of symptoms after initially being tested while asymptomatic has identified cases that were initially PCR negative but epidemiologically linked to other cases, as has been seen elsewhere (Ai et al., 2020).

Importation of cases is not accounted for in this method. Defining outbreaks based on the epidemiological linkage of cases to at least one common source of exposure (i.e. clusters) necessarily precludes inclusion of new sources of infection (i.e. importation). A new case linked to, for example, a physical location

that was a common source of exposure for cases in a cluster may represent an importation event rather than a continuation of the outbreak/cluster, unless there is clear epidemiological link (e.g. close contact or physical proximity during a given timeframe) between the newly detected case and the cluster. When applying this method to outbreaks defined by a geographic region with free-flow borders to other regions with active transmission — as was the cases for prefectures in Japan during the first wave of the pandemic — importation of one or more cases before the outbreak would simply add to the existing epidemic curve and feed into the end-of-outbreak probability calculations. Likewise, exportation of cases is not accounted for in this method, because any case epidemiologically linked to the cluster is included, regardless of the geographic boundaries within Japan. However, possible exportation across international borders is not accounted for. Even when considering clusters, if local public health jurisdictions minimize publicly shared information (as was seen in later stages of the pandemic), some links may be missed.

Lastly, further analyses describing transmissivity of the virus are needed to improve understanding of the most plausible range for these critical values. In addition, when more information is available regarding the different transmission routes of COVID-19 (i.e. airborne, droplet spread, or contact with contaminated fomites) the model could potentially be updated to account for differences in these routes, as was previously done for modeling the flare-ups of Ebola virus disease due to sexual transmission from male survivors (Lee and Nishiura, 2019).

Other methods for statistical end-of-outbreak determination have recently been proposed, and provide alternative options for examining cases using geographically based outbreak definitions (Parag et al., 2020; Hart et al., 2019). The focus of this study on clusters may be different from typical geographically based analyses, but we believe that focusing on this scale can be meaningful to decision makers dealing with individual clusters, such as officials for the involved local health jurisdictions (i.e. cities and prefectures) as well as the facilities (hospitals, senior homes, gyms, schools, etc.) where cases have been identified.

In summary, this study incorporated a reporting delay distribution into a model for end-of-outbreak probability estimation, and applied this method to clusters in the COVID-19 epidemic in Japan. In doing so, it provides estimates of the probability that the outbreak will continue in real time. Communicating these probabilities can inform public health decision making regarding the appropriate use of resources when transmission has declined for a given outbreak.

Conflict of interest

The authors have no conflicts of interest to disclose.

Funding

NML received a Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) graduate scholarship. HN received funding from a Health and Labor Sciences ResearchGrant (19HA1003, 20CA2024, and 20HA2007), the Japan Agency for Medical Research and Development (AMED; JP19fk0108104 andJP20fk0108140), 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 funders were not involved in the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit for publication.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijid.2021.02.106.

References

Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 2020;2019:200642, doi:http://dx.doi.org/10.1148/radiol.2020200642.

Ali ST, Wang L, Lau EHY, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science (80–) 2020;9004(July), doi:http://dx.doi.org/10.1126/science.abc9004 eabc9004.

Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic?. Lancet 2020;395(10228):931–4, doi:http://dx.doi.org/10.1016/S0140-6736(20)30567-5.

Bi Q, Wu Y, Mei S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis 2020;3099(20):1–9, doi:http://dx.doi.org/10.1016/S1473-3099(20)30287-5.

Blumberg S, Lloyd-smith JO. Inference of R0 and transmission heterogeneity from the size distribution of stuttering chains. PLoS Comput Biol 2013;9(5):1–17, doi:http://dx.doi.org/10.1371/journal.pcbi.1002993.

Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res 2020;(May):1–8, doi:http://dx.doi.org/10.12688/wellcomeopenres.15842.1.

Ferguson NM, Laydon DJ, Nedjati-Gilani GL, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. 2020.

Furuse Y, Sando E, Tsuchiya N, et al. Clusters of coronavirus disease in communities, Japan, January–April 2020. Emerg Infect Dis 2020;26(9), doi:http://dx.doi.org/10.3201/eid2609.202272.

Hart WS, Hochfilzer LFR, Cunniffe NJ, Lee H, Nishiura H, Thompson RN. Accurate forecasts of the effectiveness of interventions against Ebola may require models that account for variations in symptoms during infection. Epidemics 2019;100371, doi:http://dx.doi.org/10.1016/j.epidem.2019.100371.

Japan National COVID-19 Cluster Taskforce. Cluster-based approach to coronavirus disease 2019 (COVID-19) response in Japan — February–April 2020. Jpn J Infect Dis 2020;, doi:http://dx.doi.org/10.1038/pr.2014.160.

Lee H, Nishiura H. Recrudescence of Ebola virus disease outbreak in West Africa, 2014–2016. Int J Infect Dis 2017;64:90–2, doi:http://dx.doi.org/10.1016/j.ijid.2017.09.013.

Lee H, Nishiura H. Sexual transmission and the probability of an end of the Ebola virus disease epidemic. J Theor Biol 2019;471:1–12, doi:http://dx.doi.org/10.1016/j.jtbi.2019.03.022.

Linton NM, Kobayashi T, Yang Y, et al. 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 2020;9(538), doi:http://dx.doi.org/10.3390/jcm9020538.

Lloyd-Smith JO. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PLoS One 2007;2(2):1–8, doi:http://dx.doi.org/10.1371/journal.pone.0000180.

Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature 2005;438(7066):355–9, doi:http://dx.doi.org/10.1038/nature04153.

Miller TE, Garcia Beltran WF, Bard AZ, et al. Clinical sensitivity and interpretation of PCR and serological COVID-19 diagnostics for patients presenting to the hospital. FASEB J 2020;34(10):13877–84, doi:http://dx.doi.org/10.1096/fj.202001700RR.

Ministry of Health Labour and Welfare (MHLW). COVID-19 clusters in Japan. 2020. https://www.mhlw.go.jp/content/10900000/000606691.pdf.

Ministry of Health Labour and Welfare (MHLW). COVID-19 situation and MHLW response (August 10). https://www.mhlw.go.jp/stf/newpage_12910.html. [Accessed August 10, 2020].

Nishiura H, Miyamatsu Y, Mizumoto K. Objective determination of end of MERS outbreak, South Korea, 2015. Emerg Infect Dis 2016;22(1):146–8, doi:http://dx.doi.org/10.3201/eid2201.151383.

Nishiura H, Kobayashi T, Yang Y, et al. The rate of underascertainment of novel coronavirus (2019-nCoV) infection: estimation using Japanese passengers data on evacuation flights. J Clin Med 2020a;(February):2019–21, doi:http://dx.doi.org/10.3390/jcm9020419.

Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020b;113332, doi:http://dx.doi.org/10.1016/j.ijid.2020.02.060.

Omori R, Mizumoto K, Nishiura H. Ascertainment rate of novel coronavirus disease (COVID- 19). Int J Infect Dis 2020;96:673–5, doi:http://dx.doi.org/10.1016/j.ijid.2020.04.080.

Parag KV, Donnelly CA, Jha R, Thompson RN. An exact method for quantifying the reliability of end-of-epidemic declarations in real time. PLoS Comput Biol 2020;16(11)e1008478, doi:http://dx.doi.org/10.1371/journal.pcbi.1008478.

Park M, Cook AR, Lim JT, Sun Y, Dickens BL. A systematic review of COVID-19 epidemiology based on current evidence. J Clin Med 2020;9(4):967, doi:http://dx.doi.org/10.3390/jcm9040967.

Prime Minister's Office of Japan, Ministry of Health Labour and Welfare. Avoid the' three Cs'! https://www.mhlw.go.jp/content/3CS.pdf.

Public Health Agency of Canada. Interim guidance: public health management of cases and contacts associated with novel coronavirus disease 2019 (COVID-19). Gov Canada; 2020. p. 1–19. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/interim-guidance-cases-contacts.html.

R Development Core Team. R: a language and environment for statistical computing. 2019. https://www.r-project.org/.

Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance 2020;25(4):1–5, doi:http://dx.doi.org/10.2807/1560-7917.ES.2020.25.4.2000058.

Tariq A, Lee Y, Roosa K, et al. Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020. BMC Med 2020;18(1):166, doi:http://dx.doi.org/10.1186/s12916-020-01615-9.

Thompson RN, Hollingsworth TD, Isham V, et al. Key questions for modelling COVID-19 exit strategies. Proceedings Biol Sci 2020;287(1932):20201405, doi:http://dx.doi.org/10.1098/rspb.2020.1405.

Thompson RN, Morgan OW, Jalava K. Rigorous surveillance is necessary for high confidence in end-of-outbreak declarations for Ebola and other infectious diseases. Philos Trans R Soc B Biol Sci 2019;374(1776), doi:http://dx.doi.org/10.1098/rstb.2018.0431.

World Health Organization (WHO). Public health criteria to adjust public health and social measures in the context of COVID-19 (12 May 2020). 2020.

World Health Organization (WHO). WHO Recommended Criteria for Declaring the End of the Ebola Virus Disease Outbreak. 2020. https://www.who.int/who-documents-detail/who-recommended-criteria-for-declaring-the-end-of-the-ebola-virus-disease-outbreak.

Zhang Y, Li Y, Wang L, Li M, Zhou X. Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a metropolis of China. Int J Environ Res Public Health 2020;17(10), doi:http://dx.doi.org/10.3390/ijerph17103705.