RESEARCH

Estimating COVID-19 cases infected with the variant alpha (VOC 202012/01): an analysis of screening data in Tokyo, January-March 2021

Hiroaki Murayama[1], Taishi Kayano[2] and Hiroshi Nishiura[2*]

Abstract

Background: In Japan, a part of confirmed patients' samples have been screened for the variant of concern (VOC), including the variant alpha with N501Y mutation. The present study aimed to estimate the actual number of cases with variant alpha and reconstruct the epidemiological dynamics.

Methods: The number of cases with variant alpha out of all PCR confirmed cases was estimated, employing a hypergeometric distribution. An exponential growth model was fitted to the growth data of variant alpha cases over fourteen weeks in Tokyo.

Results: The weekly incidence with variant alpha from 18–24 January 2021 was estimated at 4.2 (95% confidence interval (CI): 0.7, 44.0) cases. The expected incidence in early May ranged from 420–1120 cases per week, and the reproduction number of variant alpha was on the order of 1.5 even under the restriction of contact from January-March, 2021, Tokyo.

Conclusions: The variant alpha was predicted to swiftly dominate COVID-19 cases in Tokyo, and this has actually occurred by May 2021. Devising the proposed method, any country or location can interpret the virological sampling data.

Keywords: Statistical estimation, Mutation, Transmissibility, Coronavirus, Epidemiological model, Mathematical model

Background

The global pandemic of coronavirus disease (COVID-19), clinically represented by acute infection in upper and/or lower respiratory tract, has established since the emergence of severe respiratory syndrome coronavirus 2 (SARS-Cov-2) in December 2019, Wuhan, China. The widespread epidemiology of COVID-19 is featured by its substantial transmissibility with the estimated basic reproduction number, i.e., the average number of secondary cases generated by a single primary case in a fully susceptible population, ranging from 1.5–3.5 [1]. Once infected, the infection is known to involve greater number of deaths than seasonal influenza with the estimated infection fatality risk ranging from 0.4–3.6% [2, 3]. Due to the absence of specific preventive measures, many industrialized countries have implemented a series of non-pharmaceutical interventions which is in the present day referred to as Public Health and Social Measures, including self-isolation, social distancing, travel restrictions, or lockdown [4–6].

*Correspondence: nishiura.hiroshi.5r@kyoto-u.ac.jp
[2] Kyoto University School of Public Health, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
Full list of author information is available at the end of the article

Murayama *et al. Theor Biol Med Model* (2021) 18:13

Page 2 of 5

While such interventions have led the paramount impact on social and economic activities in many countries, the variant of concern (VOC) alpha, or the variant that is phylogenetically referred to as B.1.1.7 attracted a global attention [7, 8], rapidly replacing other variants due to 50–70% greater transmissibility and also featured by about 30% greater risk of death than others [9–12]. New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) in the United Kingdom (UK) identified the common mutation N501Y in the variant alpha [12]. The scientific fact of the increased transmissibility was officially reported as more transmissible in early December 2020, leading to an immediate ban of flights from the UK across the world, but the global spread has been underway, starting to be recognizable especially in European Union countries and the United States [13].

Japan intensified sequencing virus samples from late December both at border quarantine station and domestic testing centers, and also devised a real-time polymerase chain reaction (rt-PCR) technique to detect N501Y mutation as the screening method in each prefecture. VOC and other associated variants, including 501Y.V2 emerging from South Africa (now referred to as variant beta) and 501Y.V3 from Brazil (variant gamma), have then started to be detected in many parts of the country [14]. As of 18 June 2021, totals of 303 cases at the quarantine station and 19,453 domestically acquired cases with variant alpha have been confirmed. Given that only partial samples have been screened out of PCR confirmed cases, a method for estimating actual number of cases with variant alpha was called for. Employing simplistic mathematical models, this study aimed to estimate the actual number of PCR positive cases with variant alpha and reconstruct the epidemiological dynamics so that insights into the current and future prospects can be gained.

Materials and methods
Epidemiological data
Weekly number of PCR confirmed COVID-19 cases in Tokyo, January-March 2021 was analyzed. Not only confirmed case count but also numbers of samples screened for N501Y mutation by rt-PCR and positive samples were collected (see Online Supporting Material). The screening was performed via a simple random draw, analyzing viruses of diagnosed cases in an earlier period (e.g. confirmed $s=2$ weeks earlier than screening for N501Y). Our analysis was conducted for the dataset from Week 0 to 13, 2021, i.e. from the week starting with 28 December 2020 to that starting with 29 March 2021. The time interval was 7 days throughout the observation period.

Mathematical model
The data were generated via hypergeometric sampling process. That is, the observed data were considered as resulting from random sampling trials, and the presence of the variant alpha was assumed to follow a hypergeometric distribution. Exploiting the distribution, we would like to understand how many variant alpha cases (i_t) there were in week t out of the total of PCR confirmed cases n_{t-s} where s is the delayed number of weeks for screening ($s=2$ for the following analyses). Suppose that m_t samples were screened by rt-PCR in week t, the probability of identifying k_t positive variant alpha cases in the screening samples is

$$\Pr\left(X_t = k_t; n_{t-s}, i_t, m_t\right) = f(k_t) = \frac{\binom{i_t}{k_t}\binom{n_{t-s} - i_t}{m_t - k_t}}{\binom{n_{t-s}}{m_t}} \tag{1}$$

Using Eq. (1) as the likelihood with known datasets n_{t-s}, m_t and k_t, we estimated i_t with the exact 95% confidence interval (CI) derived from the hypergeometric distribution through maximum likelihood method. It should be noted that week t in our study represents the week of screening testing for N501Y mutation, not the week of PCR confirmation of cases. To interpret the epidemiological dynamics by the week of diagnosis, two weeks must be subtracted.

Subsequently, assuming an exponential growth for i_t, with the daily growth rate r, we consider an exponential growth model $i(t) = i_0\exp(rt)$ for continuous time t where i_0 is the initial value at $t=0$. Integrating $i(t)$ for Δj days, i.e., the length of week (i.e., 7 days), we obtain $I_1 = i_0(\exp(r\Delta j)\text{-}1)/r$ for week 0. The same solution for week 1 is $I_2 = i_0 \exp(r\Delta j) (\exp(r\Delta j)\text{-}1)/r$, and we have $I_3 = i_0 \exp(2r\Delta j) (\exp(r\Delta j)\text{-}1)/r$ for week 2. Generalizing it, we have $E(i_t) = i_0\exp(rw\Delta j)(\exp(r\Delta j)\text{-}1)/r$ for week w where i_0 is the initial value in week 0, r is the daily growth rate and Δj is the length of week. We substituted i_t in Eq. (1) by the expectation $E(i_t)$ and estimated those unknown parameters i_0 and r by minimizing the negative logarithm of the likelihood. The future forecast by Week 18 was obtained with the uncertainty bound calculated from the dataset by Week 13, employing parametric bootstrap method with resampling of parameters i_0 and r for 1,000 times.

Ethical considerations
The datasets used in this study are publicly available, and each of the cases is deidentified before the analysis. Therefore, the present study did not require ethical approval.

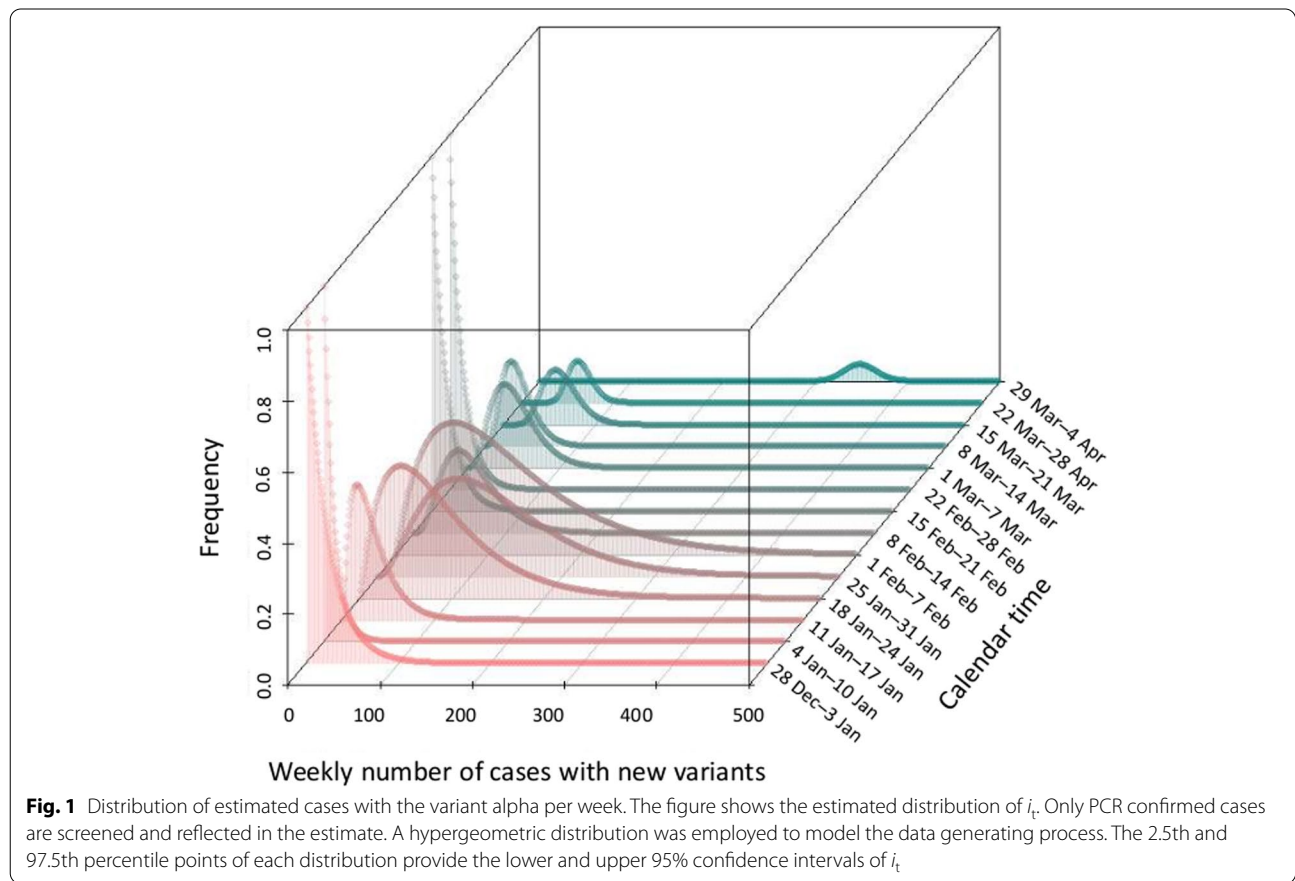Murayama *et al. Theor Biol Med Model*　　(2021) 18:13

Page 3 of 5

Results

Online supporting material shows the empirical data from Week 0 to 13. While no variant was detected from Week 0 to 2, there was 1 positive result each for Week 3 and 4, respectively. These are based on weekly screening testing of 333.6 samples on average, derived from a total of 70, 272 PCR confirmed cases during this period in Tokyo. Namely, from Week 0 to 13, weekly samples of 180, 482, 371, 285, 246, 135, 367, 241, 188, 178, 208, 250, 509, 1030 cases undertook screening. Analyzing the hypergeometric data generating process, the estimated PCR positive variant alpha cases were 0 cases (95% CI: 0, 2) for Week 0 and 1, and 16 cases (95% CI: 1, 89) and 44 cases (95% CI: 2, 243), respectively, for Week 2 and 3 (Fig. 1).

Figure 1 purely rests on empirical data alone. However, imposing an exponential growth assumption, the estimated PCR positive variant alpha cases were 0 (95% CI: 0.1, 0.1), 1.8 (95% CI: 0.9, 9.9), 2.5 (95% CI: 0.6, 15.5), 3.1 (95% CI: 0.8, 38.8) and 4.2 (95% CI: 0.7, 44.0) cases, respectively, for Week 0–4 (Fig. 2A). By Week 13, the incidence was estimated to have increased to 107.7 (95% CI: 72.6, 151.3) cases per week.

Parameter i_0 was estimated as 0.09 (95% CI: 0.09, 0.10) and exponential growth rate r was estimated to be 0.05 (95% CI: 0.05, 0.06). The latter implies that, assuming that the mean generation time is $T_g = 5$ days [15], the reproduction number of variant alpha under the voluntary lockdown period in Tokyo was 1.5 (95% CI: 1.4, 1.5) and 1.6 (95% CI: 1.5, 1.6), respectively, for exponentially distributed and constant generation time (i.e. by using $1 + rT_g$ and $\exp(rt)$ as the estimator). Figure 2B shows the future forecast based on the parameterized model. The expected weekly incidence of variant alpha in Week 18, i.e., from 3–9 May, was 730 cases (95% CI: 420, 1120).

Discussion

The present study exploited the hypergeometric distribution to estimate the incidence of PCR positive variant alpha cases in Tokyo [12]. Using this simplistic method, we have successfully demonstrated that the estimation task can be simplified and easily integrated into the epidemiological surveillance practice. We have shown that the weekly incidence of variant alpha has at least exceeded 10 cases by mid-February, which was perhaps too late to consider possible elimination, and the
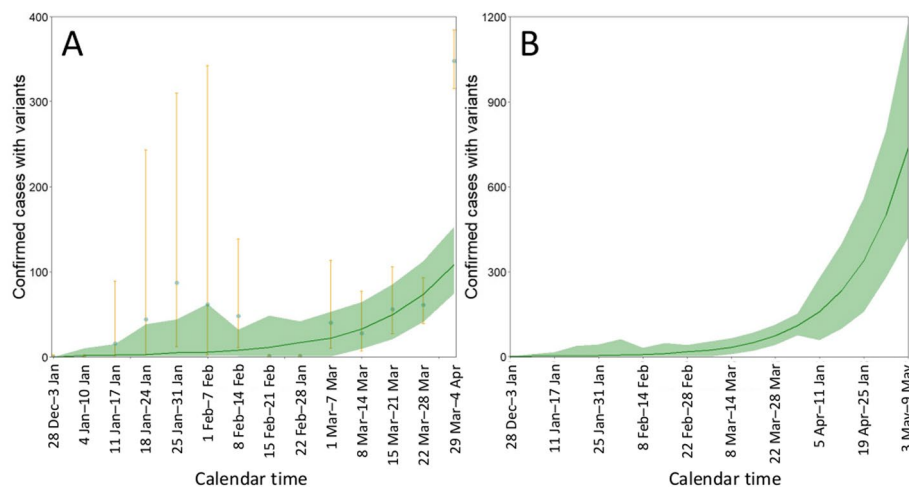

Fig. 1 Distribution of estimated cases with the variant alpha per week. The figure shows the estimated distribution of i_t. Only PCR confirmed cases are screened and reflected in the estimate. A hypergeometric distribution was employed to model the data generating process. The 2.5th and 97.5th percentile points of each distribution provide the lower and upper 95% confidence intervals of i_t

Murayama *et al. Theor Biol Med Model* (2021) 18:13

Page 4 of 5

Fig. 2 Weekly number of estimated PCR confirmed cases with variant from week 0 to 13 and weekly prediction to week 18. **A** The weekly number of estimated PCR confirmed cases with variant from 28 December 2020 to 29 March 2021. The light blue dot represents sample estimates from empirical data, with the exact uncertainty bound (95% confidence intervals) represented by yellow error bars. The green line shows the simulated mean from the exponential growth model with bootstrap resampling experiments ($n = 1000$ times). The green shade shows the 95% confidence intervals derived from the parametric bootstrap method. **B** The weekly prediction of confirmed cases with variant. The green line shows the simulated mean from the exponential growth model with bootstrap resampling experiments ($n = 1000$ times). The green shade shows the 95% prediction intervals derived from the parametric bootstrap method. The number of PCR testing per week was assumed to be 10,000

estimated reproduction number was on the order of 1.5 even under the restriction of contact. From very limited partial sample data, the present study warned that the variant alpha was expected to swiftly dominate cases soon in Tokyo. Unfortunately, the replacement has actually occurred by May 2021 across Japan.

Devising the proposed method, any other country or location can interpret the partial virological sampling data [11]. Depending on the estimate, stringent countermeasures may be considered, if aiming to fully contain the epidemic and bring the transmission of variant alpha cases under control. To do so, our simplistic method indicates that the numbers of screened samples and positive count (especially, positive cases without an apparent epidemiological link) must be regularly surveyed and reported.

Several limitations must be acknowledged. First, the sample size was very limited in the present study. Especially, empirical observation result during early weeks did not involve any positive screening samples, and exponential growth assumption during the corresponding phase needs to be validated more in the future. At least, we employed the most appropriate statistical inferential approach and attempted to make the best use of the data with uncertainty bound, i.e., the exact and bootstrap-based confidence intervals. Second, our geographic subject was restricted to Tokyo, and we have yet to analyze the dataset in other locations explicitly. In the biggest metropolitan city, we have demonstrated that the variant alpha is perhaps about to replace other strains. Third,

the estimated case count is calculated out of all PCR confirmed cases, and there must be greater number of infections with variant alpha.

Despite the abovementioned limitations, we believe that the proposed method successfully established a method for scientifically monitoring the emergence and propagation of the variant alpha. The escaping process of variant alpha cases from entry screening should also be explored to understand the invading dynamic process of the variant.

Conclusions

The present study offered a method to estimate the actual number of the variant alpha cases out of all PCR confirmed cases. Estimates implied that the variant alpha possesses substantial transmissibility, and have a potential to swiftly dominate COVID-19 cases in Japan.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12976-021-00146-x.

Additional file 1.

Murayama *et al. Theor Biol Med Model*   (2021) 18:13

Page 5 of 5

Availability of data and materials

All data generated or analyzed during this study are included in the online supporting material.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that co-author H. Nishiura is the Editor-in-Chief of Theoretical Biology and Medical Modelling. This does not alter the authors' adherence to all the Theoretical Biology and Medical Modelling policies on sharing data and materials.

Author details

[1]School of Medicine, International University of Health and Welfare, Kozuno-mori 4-3, Narita City, Chiba 286-8686, Japan. [2]Kyoto University School of Public Health, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

References

1. Jung SM, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, Kobayashi T, Kinoshita R, Nishiura H. Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. J Clin Med. 2020;9(2):523. https://doi.org/10.3390/jcm9020523.
2. Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. Euro Surveill. 2020;25(12):2000256.
3. Kobayashi T, Jung SM, Linton NM, Kinoshita R, Hayashi K, Miyama T, Anzai A, Yang Y, Yuan B, Akhmetzhanov AR, Suzuki A, Nishiura H. Communicating the risk of death from novel coronavirus disease (COVID-19). J Clin Med. 2020;9(2):580. https://doi.org/10.3390/jcm9020580.
4. Askitas N, Tatsiramos K, Verheyden B. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. Sci Rep. 2021;11(1):1972.
5. Brett TS, Rohani P. Transmission dynamics reveal the impracticality of COVID-19 herd immunity strategies. Proc Natl Acad Sci U S A. 2020;117(41):25897–903.
6. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020;584(7820):257–61.
7. Public Health England. Investigation of SARS-CoV-2 variants of concern in England, teaching briefing 6. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961299/Variants_of_Concern_VOC_Technical_Briefing_6_England-1.pdf. Accessed 18 Feb 2021.
8. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et.al. Preliminary genomic characterisation of an emergent SARS-Cov-2 lineage in the UK defined by a novel set of spike mutations. Virological. org. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563. Accessed 18 Feb 2021.
9. Knock ES, Whittles LK, Lees JA, Perez-Guzman PN, Verity R, FitzJohn RG, et al. The 2020 SARS-CoV-2 epidemiology in England: key epidemiological drivers and impact of interventions. medRxiv. 2021.01.11.21249564.
10. Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill. 2021;26(1):2002106.
11. ZhaoLouCaoZhengChongChen SJLHMKCZ, et al. Quantifying the transmission advantage associated with N501Y substitution of SARS-CoV-2 in the United Kingdom: an early data-driven analysis. J Travel Med. 2021;28:taab011.
12. New and Emerging Respiratory Virus Threats Advisory Group. NERVTAG note on B.1.1.7 severity for SAGE 77. NERVTAG - COVID-19 Public statements.
13. Du Z, Wang L, Yang B, Ali ST, Tsang TK, Shan S, et al. International risk of the new variant COVID-19 importations originating in the United Kingdom. medRxiv. 2021.01.09.21249384.
14. National Institute of Infectious Diseases, Japan. Report 6: new variant strains of SARS-CoV-2. 2021. https://www.niid.go.jp/niid/ja/diseases/ka/corona-virus/2019-ncov/10169-covid19-35.html. Accessed 18 Feb 2021.
15. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis. 2020;93:284–6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.