medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Estimating SARS-CoV-2 reproduction number by infection location in Japan

Junko Kurita[1)], Takahide Hata[2)], Tamie Sugawara[3)], Yasushi Ohkusa[3)] , Atsuko Hata[2)]

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

2) Department of Pediatrics, Kitano Hospital, Tazuke Kofukai Medical Research Institute, Osaka, Japan

3) National Institute of Infectious Diseases, Tokyo, Japan

Corresponding author: Atsuko Hata, ahata@kitano-hp.or.jp

Keywords: COVID-19, infectiousness by age class, school, reproduction number, infection place

ICMJE Statement

Contributors AH was responsible for the coordination of the study. JK and TH set data. YO developed the model and ST illustrated the results. All authors contributed to the writing of the final manuscript.

Keywords: COVID-19, infection location, reproduction number, restaurant, school, hospital, home

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Background: COVID-19 infectiousness might differ by infection location. Nevertheless, no such study of infectiousness has been reported.

Object: The study objective was estimation of the reproduction number by infection location.

Method: Patients who infected no one were ignored because their reliability might be lower than that of patients who infected more than one person. On the assumption that the histogram follows an exponential distribution, we estimated the reproduction number from the histogram of the number of people infected by the same patient.

Results: Night entertainment venues showed the greatest infectiousness, followed by facilities for elderly people and hospitals. Nursery schools and workplaces were followed by homes, with the lowest infectiousness.

Discussion and Conclusion: Countermeasures under the second declaration of emergency status targeted restaurants. However, infectiousness at restaurants was not high. Comparable to those of universities and karaoke, and not significantly different from homes: the least infectious location studied.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Since the emergence of COVID-19 in December, 2019 in Wuhan, China, reproduction numbers have been estimated several times. Some of the earliest studies conducted in Wuhan [1–3] estimated R_0 for COVID-19 as 2.24–3.58. Even in Japan, early research [4] estimated R_0 as 2.049 (95% confidence interval (CI) [2.403, 2.557]). However, these reproduction numbers were for whole populations. Reproduction numbers by location of infection are less known, but infectiousness probably differs among infected places. For instance, countermeasures under the second emergency status declaration on January 7, 2021 clearly stipulate that restaurants close earlier than eight o'clock p.m. This policy was based on an inference that infectiousness at restaurants was higher than at other areas. By contrast, nursery schools and schools were not required to close as a countermeasure, although they had been closed under the first emergency declaration from April 8 to May 24, 2020. The objective of this study was confirmation of differences in infectiousness by infection location.

A study conducted to estimate infectiousness in the earlier stage of the outbreak in Japan included patients who were not reported as having infected someone [5]. They estimated a very small reproduction number, 0.6, as of the end of February in Japan. Although they did not designate it as R_0, they referred to it as the average number of

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

secondary infections. Such a low number indicates that the outbreak of COVID-19 was self-limited. Therefore, any intensive infection control such as school closure or restriction against going out is expected to be unnecessary. The authors of that report apparently misunderstood the meaning of patients who were not reported as having infected someone. They might have been severely underestimated at that time. Therefore, people they infected might have been found and reported. Alternatively, investigation of them cannot simply reveal who had been infected by them. Therefore, we proposed another method of estimating infectiousness that excluded information of patients who were reported as not having infected anyone [6].

When we applied our proposed procedure for the present study to data obtained from an earlier study, we obtained a figure of 4.4273. Its 95% CI was [3.6000, 5.3364]: more than six times greater than the original estimate. That finding was comparable to our results obtained for infections from adults to elderly people and from elderly people to adults. They apparently underestimated R_0. Therefore, the chosen infection-control policy was misguided, with insistence on contact tracing.

Method

We adopted a similar method to estimate infectiousness by location of infection, as

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

in our earlier study, which investigated infectiousness by age of the infected person and age class inferred from the infection source [6].

We chose to examine major places where people were being infected: homes, hospitals, facilities for elderly people, workplaces, schools, nursery schools, universities, restaurants, night entertainment venues, and karaoke. For this study, the School category does not include nursery schools or universities, but does include kindergartens, elementary schools, junior high schools, and high schools. In addition, the Restaurant category does not include night entertainment venues or karaoke.

Let $x_{i,j}$ represent the number of cases in which j patients were infected secondarily in place i. Because we do not know the probability by which a patient infected one person, the probability that a person infected two or more people was assumed to follow an exponential distribution as p_i, p_{ij}^2, p_{ij}^3, and so on. Then $R_{ij=}\ p_{ij}+2p_i^2+3p_{ij}^3+\ldots=\Sigma_{k=1}\, k\, p_{ij}^{\ k}=p_{ij}/(1-p_{ij})^2$.

We observed an estimator of $p_{i,}$ as $x_{i,1}/N_i$, where N_i, represents an unknown total number of age class i patients who were infected by age class j patients. Similarly, p_i^2 was estimated in general as $x_{i,2}/N_i$ and $p_i^m= x_{i,m}/N_i$. By log transformation, we have $m\ log\ p_i= log\ x_{i,m} - log\ N_i$ (m=1, 2,…M) where M stands for the maximum number of secondary infections. Therefore, we obtain an estimator of p as an estimated coefficient

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of regression of $log\ x_{i,m}$ on m using ordinary least squares method. In addition, N_i^* was obtained from an exponential transform for the estimated constant term.

The confidence interval (CI) of $R_{i,}^*$ was obtained using a bootstrapping procedure for the distribution of $\{x_{i,m}\,(m=1, 2, \ldots L_i)\}$, where $L_{i,}$ stands for the maximum number of non-zero secondary infection [4]. In addition to usual bootstrapping, we conducted it with special consideration for the case of $x_{i,m}=0$ $(m=1, 2, \ldots L_i)$. These cases were ignored in estimation despite including much information. We bootstrapped for the distribution of $\{x_{i,m}+1\ (m=1, 2, \ldots L_i)\}$ and produced an estimate using $\max[0.001, \{x_{i,m}+1\}^b-1](m=1,2, \ldots L_i)$, where superscript b denotes a bootstrapped series and 0.001 was a small number instead of 0.

Based on the j-th bootstrapped distribution $\{x_{i,m}\,(m=1, 2,\ldots)\}^j$, we can obtain $R_{i,j}^*$. We repeated this procedure one million times, thereby obtaining one million bootstrapped $R_{i,j}^*$. We sorted these variables. The duration from $R_{i,25000}^*$ to $R_{i,975000}^*$ is expected to be 95% CI of R_i^*.

All information used for this study was obtained from reports of the Ministry of Health, Labour and Welfare [7] and local governments. The study period extended from January 15, when the initial case was detected in Japan, to the end of July.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Ethical considerations

All information used for this study has been published elsewhere [7]. There is therefore no ethical issue related to this study. We inferred significance at the 5% level.

Results

Through the end of July, 36,431 patients had been confirmed in Japan. From those, after excluding asymptomatic cases, cases of people presumed to have been infected in foreign countries, and cases for which no onset date was available, we were left with 30,780 cases. Of those cases, after excluding cases for which the infection source was unknown, and cases for which the age of patients and sources of infection were unavailable, we were left with 5383 cases. Of those, 4886 cases were identified as infection sources. These 4886 cases were analyzed for this study.

Figure 1 presents a histogram of cases by the number of secondary infections at home. Figure 2 depicts infection cases related to hospitals, facilities for elderly persons, or workplaces. It is noteworthy that three cases showing 20 secondary infections in a hospital represented more than 20 secondary infections. Figure 3 portrays those in schools, nursery schools, and universities. Figure 4 depicts those in restaurants, night entertainment venues, and karaoke.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Estimation results of R_i are presented in Table 1. Regarding median values, hospitals were found to be the highest, followed by universities and facilities for elderly people. Night entertainment venues were the lowest, followed by nursery schools, schools, and workplaces.

However, except for homes and hospitals, the lower bound of 95% CI of all other sites was less than one. In other words, their infectiousness was not significantly different from one. Therefore, their infectiousness at hospitals and homes was considerably higher than one.

Discussion

We used a procedure to estimate the case distributions among numbers of infected cases developed in our earlier study [6]. Although infected cases or unlinked cases for which the infection source was unknown represented a majority of cases, the procedure we used ignores information those cases because it was less credible. However, information about patients who were reported as having infected someone was more reliable than others because, at least, they had been investigated by public health authorities.

Results demonstrated that the estimated infectiousness at hospitals and homes was significantly greater than one. Infectiousness at facilities for elderly people was

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

marginally higher than one. Infectiousness at the other considered places was not significantly higher than one. Particularly, the estimated infectiousness in restaurants was not high. Therefore, rather than restaurants, countermeasures for COVID-19 should specifically examine hospitals, some other considered places, or homes.

It is noteworthy that infectious areas found from the present study do not represent a hot spot at which numerous people were infected. The total number of people infected in a type of place represents the product of infectiousness and people who are infectious visiting and staying at a place. For example, although infectiousness at homes was less than at other places, a huge number of patients stayed at home and shared contact with family members. For those reasons, one would expect that the number of people infected at home would be quite larger than at other places: and it was. When interpreting the obtained results, one must be reminded that infectiousness represents an average number of secondarily infected people per infectious person.

We have examined advanced bootstrapping procedures with special consideration for some particle numbers of secondary infection recording zero cases. For estimation in the present study, information about the number of secondary infections was ignored because log transformation of the number of cases was used. However, the likelihood of one case at a particular number of secondary infections actually leading to zero cases

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

was probably less but an almost comparable likelihood to that of one case at a particular number of secondary infections actually recording one case in a bootstrapping procedure. Therefore, we treat those numbers of secondary infections recording zero cases with special consideration.

The present study has some limitations. First, because infectiousness in all places were not significantly different as results, data might be insufficient to do our procedure. When we accumulate the data, it might be solved partially.

Second, because of data limitations, we cannot analyze characteristics such as those of patients or hospital staff, residents or staff at a facility for elderly persons, or students and teachers at a school. For example, infectiousness among students in school or among kids in nursery school, or of medical staff to patients are probably very important factors to control the outbreak. To resolve that difficulty to some degree, data accumulation is expected to be necessary in the near future.

Thirdly, seasonality of infectiousness might be fundamentally important, as it has come to be for influenza. Because data used for this study were accumulated through July, we are unable to evaluate them. In winter, data must also be analyzed similarly. Risk related to location must be evaluated.

Conclusion

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

This study demonstrated that effective reproduction numbers at restaurants were not high. Results show that they were comparable to data from universities or karaoke and were not significantly different from data related to infection at home: the least infectious place. Therefore, countermeasures taken under the second emergency status declaration targeting infection at restaurants might not be based on evidence. We can find no significant difference in infectiousness among the places considered.

The present study is based on the authors' opinions: it does not reflect any stance or policy of their professionally affiliated bodies.

Acknowledgments

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

Ethical considerations

All information used for this study was collected under the Law of Infection Control, Japan and published data was used. There is therefore no ethical issue related to this study.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Competing Interest

No author has any conflict of interest, financial or otherwise, to declare in relation to this study.

Reference

1. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Lou Y, Gao D, Yang L, He D, Wang M. Preliminary Estimation of the Basic Reproduction Number of Novel Coronavirus (2019-nCoV) in China from 2019 to 2020: A Data-Driven Analysis in the Early Phase of the Outbreak. Int J Infect Dis 2020 ;92:214-7. doi: 10.1016/j.ijid.2020.01.050.

2. Liu Y, Gayle AA, Wilder-Smith A, Rockly J. The reproductive number of COVID-19 is higher than SARS coronavirus. J Travel Med. 2020.DOI: 10.1093/jtm/taaa021

3. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Coronavirus disease-2019 (COVID-19): The Epidemic and the Challenges. Int J Antimicrob Agents. doi:10.1016/j.ijantimicag.2020.105924

4. Sugishita Y, Kurita J, Sugawara T, Ohkusa Y. Effects of voluntary event

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

cancellation and school closure as countermeasures against COVID-19 outbreak in Japan. Plos one 2020. https://doi.org/10.1371/journal.pone.0239455

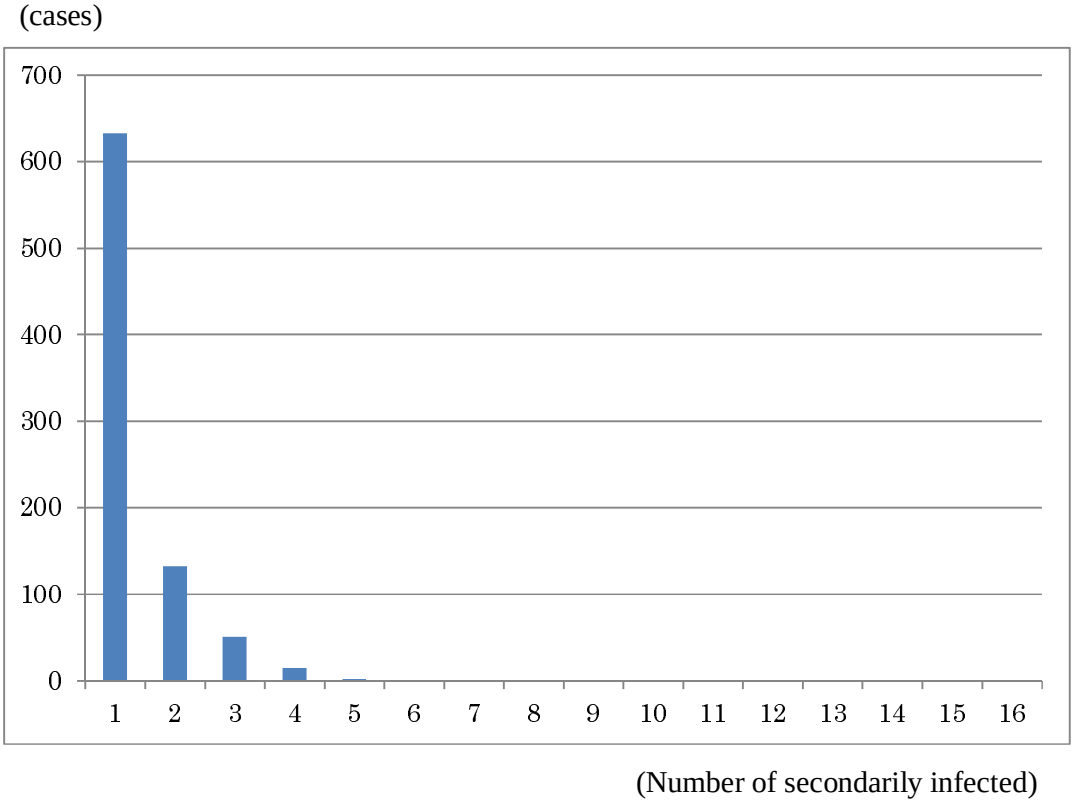
5. H Nishiura, H Oshitani, T Kobayashi, T Saito, T Sunagawa, T Wakita, MHLW COVID-19 Response Team. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). doi: https://doi.org/10.1101/2020.02.28.20029272

6. Kurita J, Hata T, Sugawara T, Ohkusa Y, Hata A. An Estimation of Reproduction Number of SARS-CoV-2 by Age Class for Age Classes in Japan. https://www.medrxiv.org/content/10.1101/2021.01.14.21249854v1.full

7. Japan Ministry of Health, Labour and Welfare. Press Releases. https://www.mhlw.go.jp/stf/newpage_10723.html (in Japanese) [accessed on Deember 10, 2020]

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

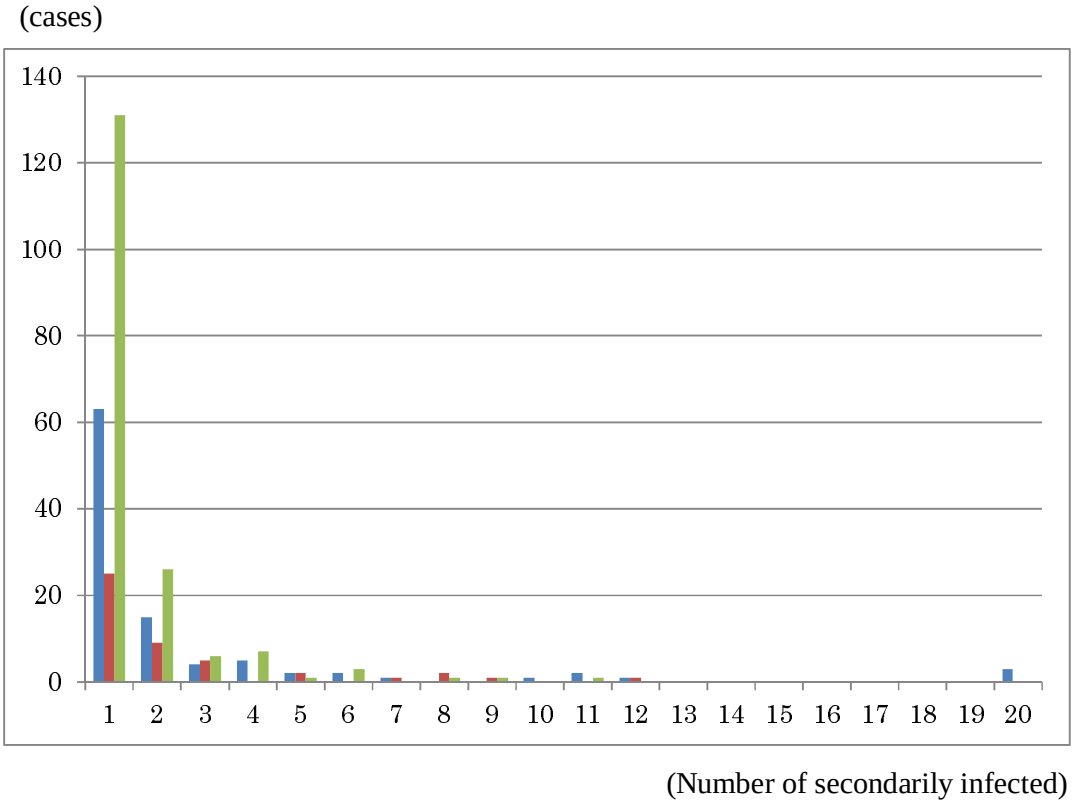
Figure 1: Histogram showing the numbers of infected cases at home.



Note: Bars represent numbers of people infected at home.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

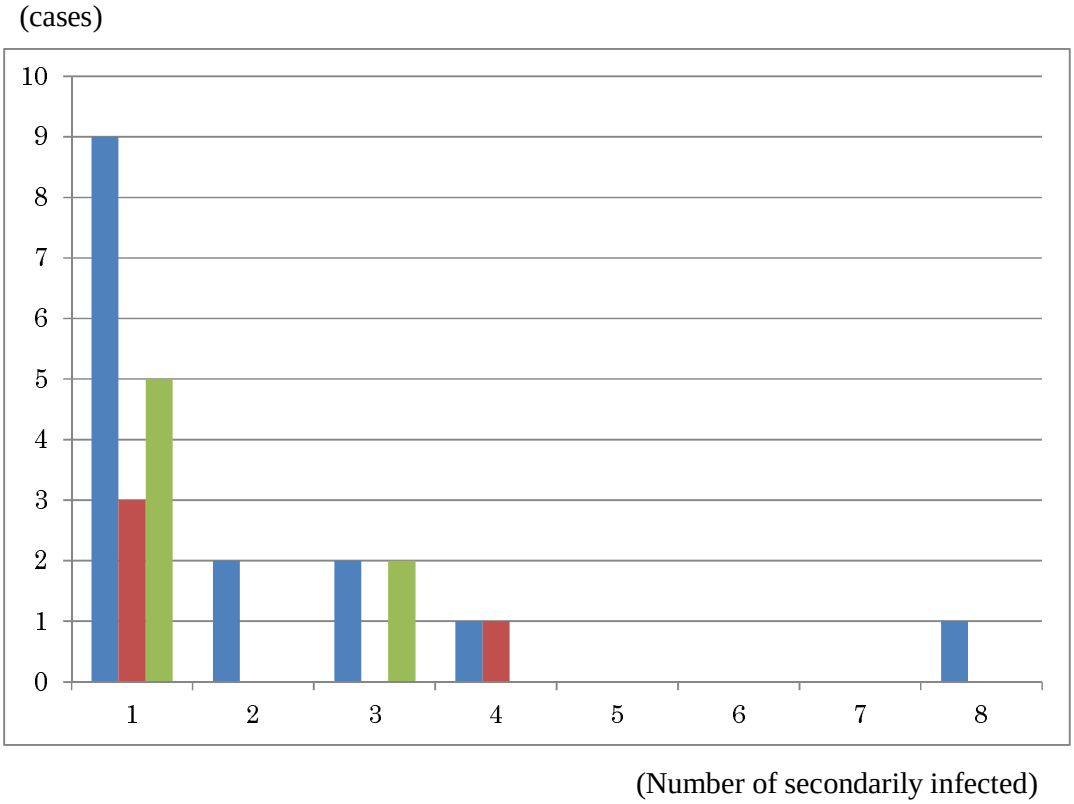
Figure 2: Histogram showing numbers of people infected at hospitals, facilities for elderly people, and workplaces.



Note: Blue bars represent the number of the infected cases at hospitals. Orange bars represent those at facilities for elderly people. Gray bars represent those at workplaces. Infections at hospitals include cases in which 21, 34, or 57 were secondarily infected. In the figure, these three cases were added together as 20 secondarily infected.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

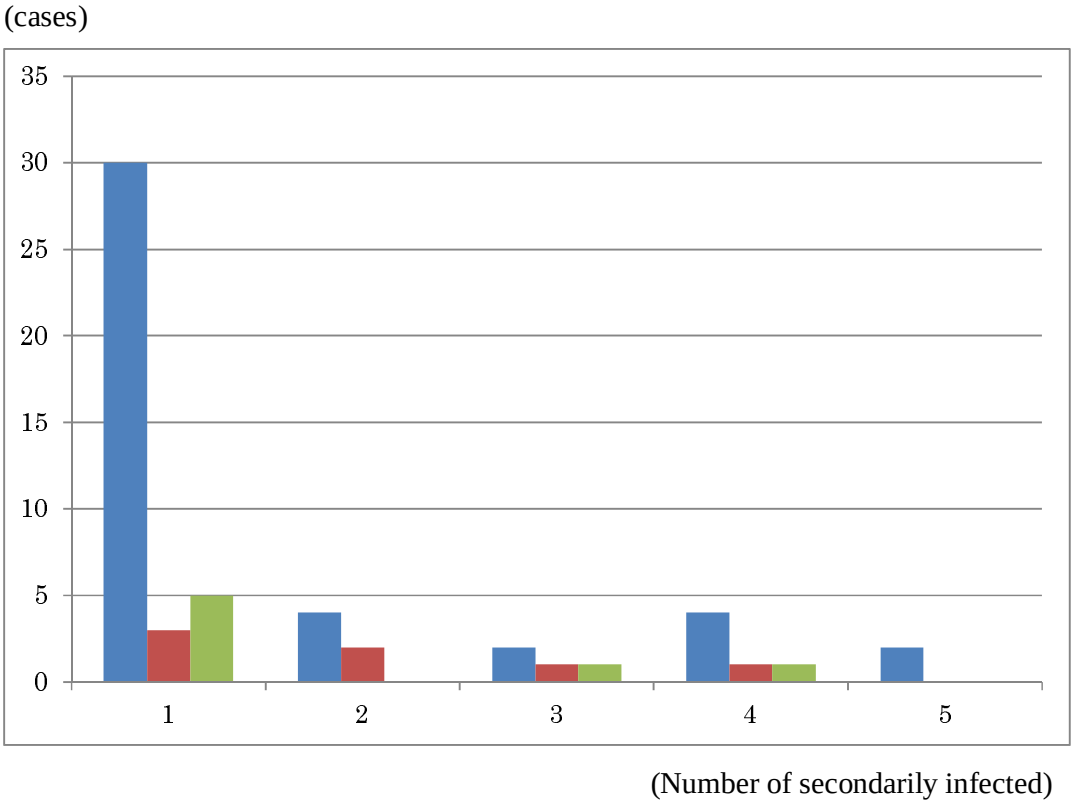
Figure 3: Histogram showing the numbers of people infected at schools, nursery schools, and universities.



Note: Blue bars represent the number of the infected cases in school, Orange bars represent those at nursery school. Gray bars represent those at university. Schools do not include nursery schools or universities, but include kindergartens, elementary, junior high, and high schools.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 4: Histogram showing numbers of people infected at restaurants, night entertainment venues, and karaoke.



Note: Blue bars represent the number of people infected at restaurants. Orange bars represent those infected at night entertainment venues. Gray bars represent those infected at karaoke. Restaurants do not include night entertainment venues or karaoke.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.13.21255296; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1: Estimated results of effective reproduction number by infection location

	Median	95% CI	
		Lower bound	Upper bound
Home	2.4752	1.2943	6.7998
Hospital	7.0849	3.2501	28.9211
Facility for elderly people	3.2267	0.9927	29.4718
School	1.2331	0.3160	39.8456
Nursery School	1.1580	0.0453	67.5018
University	3.2321	0.20114	48.2487
Restaurant	2.9824	0.1462	17.2831
Night entertainment	0.6481	0.0374	67.5018
Karaoke	2.0143	0.0391	42.4974
Workplace	1.3161	0.4993	6.7833

Note: “School” includes kindergartens, elementary schools, junior high schools, and high schools. “Restaurant” excludes “Night entertainment” and “Karaoke.”