

Potential Biases of the Transmission Risks of COVID-19 estimated by Contact Tracing Surveys in Japan

Tsubasa Ito⁽¹⁾, Takahiro Otani⁽²⁾, Tatsuhiko Anzai⁽³⁾, Takashi Okumura⁽⁴⁾, Kunihiro Takahashi⁽³⁾

(1) Faculty of Public Policy, Graduate School of Public Policy, Hokkaido University, Hokkaido, Japan

(2) Department of Public Health, Graduate School of Medical Sciences, Nagoya City University, Aichi, Japan

(3) Department of Biostatistics, M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan

(4) Health Administration Center, Kitami Institute of Technology, Hokkaido, Japan

CORRESPONDING AUTHOR: Tsubasa Ito, Faculty of Public Policy, Graduate School of Public Policy, Hokkaido University, Hokkaido, Japan. E-mail: tito@econ.hokudai.ac.jp, ORCID 0000-0001-8021-9907

SUMMARY

Introduction: Contact tracing surveys are being conducted to identify and isolate close contacts of an identified patient to reduce the spread of coronavirus disease (COVID-19). However, the estimates of risk indexes based on information obtained from the surveys and normally used in practice can have biases comparing with true magnitude of risks of infection and spread.

Method: We evaluated whether the estimates of the risk indexes obtained from information of the active epidemiological surveillance, contact tracing surveys in Japan, are suitable for quantitative assessment of the risk factors of COVID-19, using pseudo data via a simulation study. We discussed two types of risks considered in the issue of infectious disease, the probability of infection and that of spreading, and the estimates of these risks.

Results and Discussion: A naive method to estimate the risks of infection and spreading of COVID-19 is to calculate the ratio of infected patients to close contacts and the ratio of patients who infected others to all the confirmed patients, respectively. However, these estimates could possibly have significant biases and result in being ineffective for both the exploration and the quantitative assessment of the risk factors in the following ordinary cases: a person contacts closely with many confirmed patients, or a confirmed patient contact closely with many people. Then, some steps are needed to reduce such possible biases for the estimation the risks of both the infection and spreading of COVID-19.

Keywords: *spatio-temporal epidemiology; COVID-19; active epidemiological surveillance; evidence-based policy-making; infection risk; spreading risk; simulation study.*

INTRODUCTION

The spread of the coronavirus disease (COVID-19) still can threaten the global health, and the recent (1 April, 2023) World Health Organization (WHO) COVID-19 dashboard has reported approximately 763 million confirmed cases and 7 million deaths globally [1].

Many countries use contact tracing as one of the most powerful public health interventions. The common purpose of these tracings is to identify and isolate individuals who may have been infected due to close contact with an identified patient, to prevent the infectious disease from being transmitted further [2, 3].

In Japan, a bidirectional contact tracing called “active epidemiological surveillance” is being conducted [4]. The survey investigates not only the close contacts of a confirmed patient but also other patients who possibly infected the patient, i.e. sources of infection, and then successive contact tracings are iterated to identify additional patients and close contacts related to the patient [5].

Many researchers have investigated COVID-19 using information from contact tracings, which have enabled us to identify risk factors for infection and spread. For example, it is well-known that being in a closed and poorly ventilated environment causes higher secondary transmissions of COVID-19 than

DOI: [10.54103/2282-0930/20757](https://doi.org/10.54103/2282-0930/20757)

Accepted: 16th June 2023

being in an open and well-ventilated environment [6, 7]. Furthermore, in Japan, the data from active epidemiological surveillance have been used to implement COVID-19 measures, and some new phrases have been coined, such as “stay away from the three Cs (closed spaces, crowded places, close-contact settings)” and “five situations that increase the risk of infection (social gatherings involving alcohol consumption, big feasts in large groups, conversation without a mask, living together in a small limited space, and moving to different locations)” [8, 9], which has been considered to contribute significantly to restrain the pandemic.

The risks of infection and spread are generally defined as the probability that a closed contact becomes infected and the probability that a patient infects the other, respectively. Then, in the standard epidemiological investigations, these risks are typically estimated by the ratio of the number of confirmed positive patients to the number of identified close contacts and the ratio of the number of patients who are identified to infect others to the number of confirmed patients as discussed in the next section. Although these simple estimates are practical, it is likely that they have biases and could not reflect the true magnitude of risks of infection and spread. For example, if a confirmed patient with some risk factor (e.g., without a mask) becomes close contacts with someone and they develop COVID-19, the confirmed patient will be seen as the spreader even if they actually did not infect others. Then, if one person becomes contact with a number of patients with or without symptoms, which is likely enough, that risk factor would be overestimated.

Thus, this study aimed to investigate how these estimates of the risk indexes based on information obtained by the active epidemiological surveillance could have biases comparing with the true magnitude of the risk factors for COVID-19 via simulation studies under some situations reflecting how people come into contact with each other. Assessment of potential biases of risk estimates helps improve the estimations of risk of infection and spread, which tells how we should act to prevent COVID-19 from spreading, recognizing the limitations of the risk estimates technically feasible.

METHODS

Simulation study

In this section, by using simulation studies we investigated the performances of the estimates of risks for quantitative risk assessment under two scenarios.

In the first scenario, to generate pseudo data, we followed the form of the active epidemiological surveillance in Japan. This investigation targets patients, suspected disease carriers, and their close contacts [10]. From this investigation, we can extract spatiotemporal information about the behaviors of patients (when, where, with whom, what and how they did). Then, we supposed several activities, such as watching sports in a stadium or seeing a live concert in a venue, and we set the population size, not including the infected individuals who were unaware of their condition, to 10,000. We assumed that the prevalence rate of COVID-19 in Japan is approximately 2.0% based on the estimate of the average detection rate of COVID-19 [11]. Thus, the number of patients was set to 200. We considered the case in which an individual participates in an event alone, with some or many people. One patient was in close contact with at least 1 person or 3 or 5 people, thus, there are three values for the number of close contacts: 200, 600, or 1,000 close contacts. We also considered patterns concerning how large the flow of people is. Thus, three patterns were established for the number of patients a close contact comes in close contact with on average: 1, 1.2, and 1.5. These settings can be interpreted as follows. When the number of patients per close contact on average was 1, the patients can be considered to be distributed uniformly and the infections occur globally, as shown in Figure 1.

In contrast, when the average number of patients per close contact was 1.2, we randomly selected a patient with whom a close contact has come in close contact. For example, as a result of random selection, if close contact 1 came in close contact with patient 2 in addition to patient 1, we assume that close contacts 2 and 3 also came in close contact with patient 2. In this case, a number of patients were concentrated in

Figure 1: Scenario in which the infection occurs globally (the number of patients per close contact is fixed) (left panel), and illustration for this scenario (right panel).

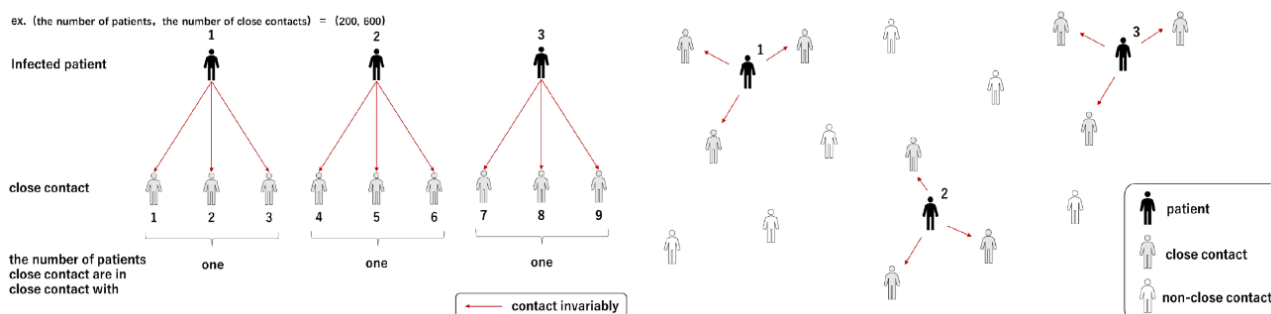
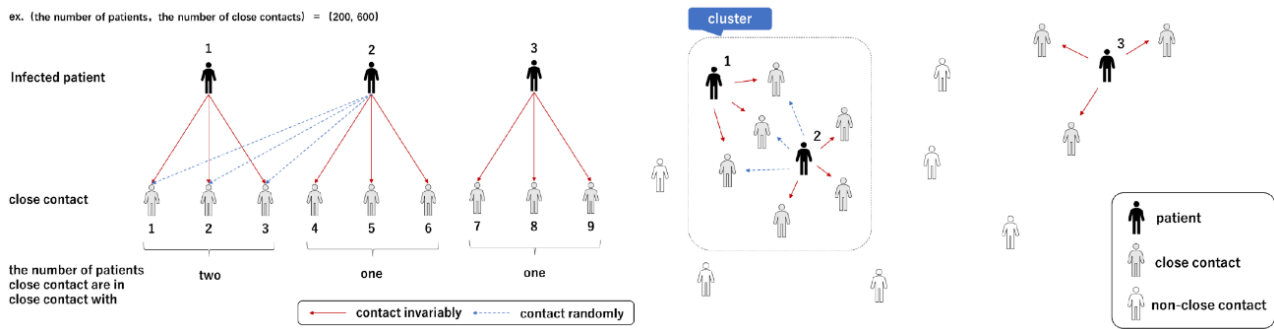


Figure 2: Scenario in which the infection occurs locally (left panel), and illustration for this scenario (right panel).



one portion and the infections occurred locally; this formed a cluster, as shown in Figure 2.

Moreover, as the average number of patients per close contact increased to 1.5, more close contacts were more likely to come in close contact with a common patient, and the size of the cluster became larger.

In this study, we considered scenarios in which individuals are exposed if they did not wear a mask. For simplicity, the probability of patients who do not wear a mask and the probability that a close contact did not wear a mask were set to 0.5. When patients did not wear a mask, the probability that their close contacts who also did not wear a mask would be infected was set to 4/9, while the probability that their close contacts with masks would be infected was set to 2/9. Contrarily, when patients wear a mask, the probability that their close contacts without masks would be infected was set to 2/9, while the probability that their close contacts with masks would be infected was set to 1/9. Therefore, the true RR of the infection and spreading risks was 2.0, and the true OR of the infection and spreading risks was 2.5. If individuals were in close contact with many patients and they were infected, it is generally difficult to infer who infected the individual. Therefore, in such cases, we considered that the individual was infected from all patients with whom they were in close contact.

In the second scenario, the simulation setting is almost same with that of the first scenario. In this time, however, the average number of close contacts per patient were not fixed, and we considered six patterns for it, that is, 1.5, 2.0, 3.5, 4.5, 5.0 and 6.0, and we randomly selected people who became in close contact with each patient such that all the patients were in close contact with at least one person. Then, the interpretation of this setting is similar to that of the case in which the number of patients close contacts are in close contact with on average is 1 in the first scenario. In this time, however, the number of close contacts a patient is in close contact with varies depending on the patient, and then this setting is closer to the realistic situation.

Statistical Analysis

We estimated the risks of both the infection and spreading of COVID-19 associated with human behaviors, which were whether they wore masks in a crowd of people in our simulation setting.

In each scenario, we estimated the infection risk and the spreading risk respectively by

$$\text{infection risk} \approx \frac{\text{the number of confirmed positive patients}}{\text{the number of identified close contacts}}$$

and

$$\text{spreading risk} \approx \frac{\text{the number of patients who are identified to infect others}}{\text{the number of confirmed patients}}$$

for both the exposed and unexposed groups. The exposed group comprises patients or close contacts who practiced risky behaviors that increased the likelihood of developing an infection, such as not wearing masks. Then, we calculated the relative risk (RR) or odds ratio (OR) as the ratio of the risk of the two groups for both the infection risk and the spreading risk. Moreover, we constructed 95% confidence intervals (CI) of each RR and OR. Then, based on 2,000 simulation runs, we computed the averaged values of the RR and OR of the infection and spreading risks, the standard deviations of the RR and OR, and the averaged values of coverage probabilities (CPs) of 95% confidence intervals (CI) of the infection and spreading risks.

Results

The results of the first scenario of the simulation study are reported in Tables 1 and 2. The RR and OR of the infection and spreading risks performed the best when both the number of patients and the number of close contacts were 200 and the average number of patients per close contact was 1. [RR=2.08, CP of its 95% CI is 95.6%, OR=2.67 and CP of its 95% CI is 95.6% for infection risk, and RR=2.11, CP of its 95% CI is 94.9%, OR=2.71 and CP of its 95% CI is 94.7% for spreading risk]

On the other hand, when both the number of patients and the number of close contacts were 200 and the average number of patients per close contact increased to 1.5, for example, $RR=1.94$, the CP of its 95% CI is 93.5%, $OR=2.71$ and the CP of its 95% CI is 95.1% for infection risk and, for $RR=1.37$, the CP of its 95% CI is 27.6%, $OR=1.85$ and the CP of its 95% CI is 78.6% for spreading risk.

Moreover, when the number of close contacts increased to 1,000 for example and the average number of patients per close contact was 1, for $RR=2.02$, the CP of its 95% CI is 94.6%, $OR=2.53$ and the CP of its 95% CI is 94.6% for infection risk and, for $RR=1.46$, the CP of its 95% CI is 9.8%, $OR=4.86$ and the CP of its 95% CI is 64.5% for spreading risk.

The results of the second scenario are reported in Tables 3 and 4. The values of RR of infection risk and its CP of 95% CI corresponding to the average number of close contacts per patient, 1.5, 2.0, 3.5, 4.5, 5.0 and 6.0, are $(RR, CP)=(2.07, 94.8\%), (2.04, 94.8\%), (2.01, 94.9\%), (2.00, 96.1\%),$

$(2.00, 95.8\%)$ and $(2.00, 93.7\%)$, respectively. The values of OR of infection risk and its CP of 95% CI corresponding to the average number of close contacts per patient, 1.5, 2.0, 3.5, 4.5, 5.0 and 6.0, are $(OR, CP)=(2.64, 95.0\%), (2.59, 94.4\%), (2.55, 94.9\%), (2.54, 95.6\%), (2.54, 95.3\%)$ and $(2.54, 94.0\%)$, respectively. On the other hand, the values of RR of spreading risk and its CP of 95% CI corresponding to the average number of close contacts per patient, 1.5, 2.0, 3.5, 4.5, 5.0 and 6.0, are $(RR, CP)=(1.90, 92.5\%), (1.78, 86.1\%), (1.54, 39.7\%), (1.42, 9.5\%), (1.37, 2.9\%)$ and $(1.30, 0.0\%)$, respectively. The values of OR of spreading risk and its CP of 95% CI corresponding to the average number of close contacts per patient, 1.5, 2.0, 3.5, 4.5, 5.0 and 6.0, are $(RR, CP)=(2.64, 95.8\%), (2.70, 95.8\%), (3.05, 92.4\%), (3.34, 90.6\%), (3.50, 90.0\%)$ and $(4.03, 87.2\%)$, respectively. Then, we can see that when the average number of close contacts per patient is large, the CI of RR for spreading risk did not work at all.

Table 1: The averaged values of the relative risk (RR), odds ratio (OR) and coverage probabilities (CP) of their 95% confidence intervals of the infection risk, the standard deviation in parentheses and α is the number of patients with whom a close contact becomes close contact on average.

		RR	CP for RR	OR	CP for OR
	$\alpha = 1$	2.08 (0.77)	95.6%	2.67 (0.99)	95.6%
(200,200)	$\alpha = 1.2$	2.05 (0.74)	95.2%	2.73 (0.99)	95.6%
	$\alpha = 1.5$	1.94 (0.67)	93.4%	2.71 (0.96)	95.1%
	$\alpha = 1$	2.03 (0.57)	95.4%	2.56 (0.73)	95.0%
(200,600)	$\alpha = 1.2$	1.97 (0.53)	94.8%	2.56 (0.70)	95.4%
	$\alpha = 1.5$	1.89 (0.51)	91.8%	2.59 (0.70)	95.0%
	$\alpha = 1$	2.02 (0.49)	94.6%	2.53 (0.63)	94.6%
(200,1000)	$\alpha = 1.2$	1.96 (0.47)	94.1%	2.54 (0.62)	95.0%
	$\alpha = 1.5$	1.87 (0.47)	88.1%	2.56 (0.61)	94.8%

Table 2: The averaged values of the relative risk (RR), odds ratio (OR) and coverage probabilities of their 95% confidence intervals (CP) of the spreading risk, the standard deviation in parentheses and α is the number of patients with whom a close contact becomes close contact on average.

		RR	CP for RR	OR	CP for OR
	$\alpha = 1$	2.11 (0.81)	94.9%	2.71 (1.03)	94.7%
(200,200)	$\alpha = 1.2$	1.64 (0.71)	78.9%	2.14 (0.87)	89.9%
	$\alpha = 1.5$	1.37 (0.82)	27.6%	1.85 (0.92)	78.6%
	$\alpha = 1$	1.69 (0.62)	70.9%	3.46 (1.21)	85.1%
(200,600)	$\alpha = 1.2$	1.35 (0.81)	2.6%	2.98 (1.08)	94.6%
	$\alpha = 1.5$	1.15 (0.92)	0.0%	2.99 (1.30)	95.4%
	$\alpha = 1$	1.46 (0.75)	9.8%	4.86 (1.76)	64.5%
(200,1000)	$\alpha = 1.2$	1.19 (0.90)	0.0%	4.95 (2.03)	88.0%
	$\alpha = 1.5$	1.06 (0.97)	0.0%	- (-)	-

Table 3: ncc is mean of the number of close contacts, “mean” is mean of the number of close contacts per patient and “sd” is its standard deviation. Averaged values of the relative risk (RR), odds ratio (OR), the standard deviations in parentheses and coverage probabilities (CP) of their 95% confidence intervals of the infection risk.

ncc	(mean, sd)	RR	CP for RR	OR	CP for OR
297.6	(1.5, 0.70)	2.07 (0.71)	94.8%	2.64 (0.92)	95.0%
394.2	(2.0, 1.05)	2.04 (0.65)	94.8%	2.59 (0.83)	94.4%
678.1	(3.5, 1.58)	2.01 (0.54)	94.9%	2.55 (0.83)	94.9%
863.2	(4.5, 1.87)	2.00 (0.51)	96.1%	2.54 (0.65)	95.6%
953.0	(5.0, 2.00)	2.00 (0.49)	95.8%	2.54 (0.63)	95.3%
1131.9	(6.0, 2.24)	2.00 (0.47)	93.7%	2.54 (0.61)	94.0%

Table 4: ncc is mean of the number of close contacts, “mean” is mean of the number of close contacts per patient and “sd” is its standard deviation. Averaged values of the relative risk (RR), odds ratio (OR), the standard deviations in parentheses and coverage probabilities (CP) of their 95% confidence intervals of the spreading risk.

ncc	(mean, sd)	RR	CP for RR	OR	CP for OR
297.6	(1.5, 0.70)	1.90 (0.67)	92.5%	2.64 (0.94)	95.8%
394.2	(2.0, 1.05)	1.78 (0.62)	86.1%	2.70 (0.92)	95.8%
678.1	(3.5, 1.58)	1.54 (0.71)	39.7%	3.05 (1.07)	92.4%
863.2	(4.5, 1.87)	1.42 (0.77)	9.5%	3.34 (1.20)	90.6%
953.0	(5.0, 2.00)	1.37 (0.80)	2.9%	3.50 (1.28)	90.0%
1131.9	(6.0, 2.24)	1.30 (0.84)	0.0%	4.03 (1.56)	87.2%

DISCUSSION

This study evaluated the estimates of the risks of infection and spreading of COVID-19 obtained using information from an active epidemiological surveillance.

From the simulation results, we can deduce the following findings. At first, we fix the number of patients and that of close contacts, and varies the number of the patients per close contact. (ex. We see the case where the number of patients is 200 and the number of close contacts is 600, and varies α from 1 to 1.5 in Table 1.) Then, we see that the performances of the both RR and OR of the infection risk were stable for most cases. However, in the case where the number of patients is 200, the number of close contacts is 1000 and the number of the patients per close contact on average is 1.5, RR has a little downward bias and the CP of 95% confidence interval is much smaller than the nominal confidence level. This downward bias might come from the fact that when a person is in close contact with many patients, they are more likely to be infected from one of the patients even if the patient wore a mask, which increases the infection risk of unexposed group. For the spreading risk, their RR had downward biases which became large as the average number of patients per close contact. Concerning the OR, slightly downward biases arose when the number of close contacts was 200 and large upward biases

arose when the number of close contacts was 600 or 1000. Consequently, their Cis did not achieve the nominal confidence level at all. These different directions of large biases might be explained by the same reason. An individual comes in close contact with many patients, the patients are more likely to be classified in the group of patients who infected others no matter whether they actually infected others. This increases the spreading risk for the unexposed group of patients, which causes underestimation of the RR of spreading risk. Simultaneously, the estimates of the risk of spreading risk for the exposed group were close to 1, which causes the overestimation of OR.

Next, we fix the number of the patients per close contact, and varies the number of close contacts. (ex. We see the case where α is 1, and varies the number of close contacts from 200 to 1000 in Table 1.) Then, the performances of the RR and OR of the infection risk were considerably stable, but for the spreading risk, large downward biases occurred on the RR and large upward biases in the OR were observed as the number of close contacts increased. These findings are consistent with the results of the second scenario. Both the RR and OR for the infection risk were accurate enough, because almost all close contacts came in close contact with one patient (Table 3), and the RR and OR for the spreading risk were not accurate enough as the average number of close contacts per patient increased (Table 4). These downward biases

might be because when patients come in contact with many people, they are more likely to infect one of their close contacts even if they wore a mask, and this results in an increase in the spreading risk for the unexposed group of patients.

All these poor performances of RR and OR might be reasonable because these estimates are only based on the number of close contacts. However, we might be able to evaluate the direction of biases in the estimates of infection or spreading risks by investigating the average number of patients per close contact or the number of close contacts per patient, which might be feasible by using the methods which evaluate the infection risk using the locational information of mobile phones proposed by [12] and [13] for example.

Practically, there are other factors that can cause biases in risk estimators. For example, we cannot exactly trace all close contacts using the active epidemiological surveillance [14], which causes a selection bias. This is because the travel recall of patients may be inaccurate or indistinct, some patients may be uncooperative during the investigation, patients with subjective symptoms fail to present at medical institutions, and patients with no or mild symptoms do not perceive as having the infection. Therefore, the reported number of close contacts is usually smaller than the true number. The reported number of patients is also smaller than the true number, because it has been shown that a significant proportion of patients are asymptomatic [15]. These factors cause bias in the estimation of the infection risk.

Lastly, the prevalence rate in the group of close contacts tends to be high; therefore, the OR computed from case-controlled studies using information from an active epidemiological surveillance might be higher than the actual RR.

Limitations

In this study, we could not consider all situations that could actually occur in the simulation setting and could only consider certain simple situations. In our simulation setting, the probabilities that two individuals will come in close contact with each other are the same for all patients and close contacts, though these probabilities depend on many confounding factors. Moreover, in our simulation setting, all patients and their close contacts were completely matched, though such exact tracing is impossible. In these more complicated situations, it can be expected that the RR and OR will not be as accurate.

CONCLUSION

In the future, more complicated simulation studies should be conducted, specifically, those that consider the movement of people. The simple simulations performed in this study, however, suggested some

issues of the estimates of the risks of infection and spread of COVID-19. There were a few possible biases in the estimates of infection risks. The RR and OR based on these estimates are useful in the identification and quantitative assessment of the risk factors for infection associated with human behaviors. However, there were several possible biases in the estimates of spreading risks. Thus, we cannot use the RR and OR based on the estimates of the spreading risk for that purpose. Moreover, for more complicated situations, we might not be able to use the estimates of both infection and spreading risks if we simply calculate the estimates of the risks using the information from an active epidemiological surveillance. In this case, we should collect more high-quality information such as the number of times of contacts in contact tracing surveys.

FUNDING

This research was supported by the Japan Agency for Medical Research and Development (AMED) under Grant Number JP20he0622042.

REFERENCES

1. World Health Organization. Who coronavirus disease (covid-19) dashboard, 2023.
2. J Cohen and Kupferschmidt K. Countries test tactics in 'war' against covid-19. *Science*, 367:1287–1288, 2020.
3. Emeline Han MSc, Melisa Mei Jin Tan MSc, Eva Turk PhD, Devi Sridhar PhD, Gabriel M Leung MD Kenji Shibuya DrPH, Nima Asgari MD Juhwan Oh PhD, Alberto L Garcia-Basteiro PhD, Johanna Hanefeld PhD, Alex R Cook PhD, Li Yang Hsu MBBS, Yik Ying Teo PhD, David Heymann DTM&H, Helen Clark MA, Martin McKee DSc, and Helena Legido-Quigley PhD. Lessons learnt from easing covid-19 restrictions: an analysis of countries and regions in asia pacific and europe. *Lancet, Health Policy*, 396:1525–1534, 2020.
4. Ministry of Health, Labor, and Welfare, Japan. Preventing the spread of covid-19 by targeting disease clusters, 2020.
5. Sadamori Kojaku, Laurent Hebert-Dufresne, Enys Mones, Sune Lehmann, and Yong-Yeol Ahn. The effectiveness of backward contact tracing in networks. *nature physics*, 2021.
6. Hiroshi Nishiura, Hitoshi Oshitani, Tetsuro Kobayashi, Tomoya Saito, Tomimasa Sunagawa, Tamano Matsui, Takaji Wakita, MHLW COVID-19 Response Team, and Motoi Suzuki. Closed environments facilitate secondary transmission of coronavirus disease 2019 (covid-19). *medRxiv*, 32(12), 2020.
7. Valentyn Stadnytskyi, Christina E. Bax, Adriaan Bax, and Philip Anfinrud. The airborne lifetime of small speech droplets and their potential importance in

- sars-cov-2 transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (22), 2020.
8. Ministry of Health, Labour and Welfare, Japan. Avoid the "three cs", 2021.
 9. Ministry of Health, Labour and Welfare, Japan. "5 situations" that increase the risk of infection, 2021.
 10. National Institute of Infectious Diseases. Guidelines for active epidemiological investigation in patients with novel coronavirus infection (tentative version), 2020.
 11. Sebastian Vollmer. Average detection rate of sars-cov-2 infections has improved since our last estimates but is still as low as nine percent on march 30th, 2020.
 12. Junko Ami, Kunihiro Ishii, Yoshihide Sekimoto, Hiroshi Masui, Ikki Ohmukai, Yasunori Yamamoto and Takashi Okumura. Computation of infection risk via confidential locational entries: A precedent approach for contact tracing with privacy protection. *IEEE Access*, 9: 87420 – 87433, 2021
 13. Ikki Ohmukai, Yasunori Yamamoto, Maori Ito and Takashi Okumura. Tracing patients' PLOD with mobile phones: Mitigation of epidemic risks through patients' locational open data. *IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE-2020)*, June 2020. arXiv:2003.06119.
 14. Matt J Keeling, Deirdre Hollingsworth, and Jonathan M Read. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (covid-19). *Journal of Epidemiology and Community Health*, 74:861–866, 2020.
 15. Daniel P Oran and Eric J Topol. The proportion of sars-cov-2 infections that are asymptomatic : A systematic review. *Annals of internal medicine*, 2021.