Statistics for COVID-19 Pandemic Data

Determination of optimal prevention strategy for COVID-19 based on multi-agent simulation

Satoki Fujita · Ryo Kiguchi · Yuki Yoshida · Yoshitake Kitanishi

Received: 20 October 2021 / Revised: 10 May 2022 / Accepted: 11 May 2022 / Published online: 14 June 2022
© The Author(s) under exclusive licence to Japanese Federation of Statistical Science Associations 2022

Abstract
This study proposes a direction for the utilization of multi-agent simulation (MAS) to consider an optimal prevention strategy for the spread of the coronavirus disease of 2019 (COVID-19) through a pandemic modeling example in Japan. MAS can flexibly express macroscopic phenomena formed through the interaction of micro-agents modeled to act autonomously. The use of MAS can provide a variety of recommendations for bringing a pandemic under control, even in the case of the COVID-19 pandemic, which has become more intense as of 2021. However, models that do not consider individual heterogeneity, such as analytical Susceptible–Exposed–Infectious–Recovered (SEIR) models, are often used as predictive models for infectious diseases and the main reference for decision-making. In this study, we show that by constructing a MAS that simulates a metropolitan city in Japan in a simple manner while considering the heterogeneity of age and other background information, we can capture the effects of various measures such as vaccinations on the spread of infections in a more realistic setting. Moreover, it is possible to offer various recommendations for optimal strategies to suppress a pandemic by combining reinforcement learning with MAS. This study explicates the potential of MAS in the development of strategies to prevent the spread of infection.

Ryo Kiguchi, Yuki Yoshida and Yoshitake Kitanishi contributed equally to this work.

Satoki Fujita
satoki.fujita@shionogi.co.jp
Ryo Kiguchi
ryo.kiguchi@shionogi.co.jp
Yuki Yoshida
yuuki.yoshida@shionogi.co.jp
Yoshitake Kitanishi
yoshitake.kitanishi@shionogi.co.jp

1 Data Science Department, Shionogi & Co. Ltd., Osaka, Japan
Keywords COVID-19 · Multi-agent simulation · Reinforcement learning · Heterogeneity

1 Introduction

The novel coronavirus disease of 2019 (COVID-19) caused by SARS-CoV-2, which originated in Wuhan City, Hubei Province, China, in December 2019, has now developed into a global pandemic. The development of vaccines and the inoculation of the public are rapidly advancing, and various countries are taking measures to suppress the spread of the infection by implementing lockdowns, requiring that people wear facemasks, and restricting movement. SARS-CoV-2 is a newly expressed virus, and its features differ from those of the annual influenza virus, including its infectiousness, infection route, and health risk; therefore, a new response guide is required. However, at present, especially in Japan, countermeasure policies are often taken based on the heuristic opinions of epidemiological experts. Because it is a new virus and it is sometimes difficult to see the interaction with environmental factors based on their opinions alone, it is preferable to have a more objective basis for decision-making. Thus, it is important to appropriately model the spread of infection and to evaluate the effects of various measures in advance to obtain prospective scenarios. Herein, we propose further utilization of multi-agent simulation (MAS) as a means of achieving this.

The epidemiological models commonly used are often expressed by a series of mathematical equations that are deterministic and have various system parameters, as represented by the analytical Susceptible–Exposed–Infectious–Recovered (SEIR) models (Kuniya 2020; Kurita et al., 2020). Although these models are useful in estimating the growth potential of infectious populations in a simple manner in the context of sparsely available information, there are concerns that each individual and population is treated uniformly, and heterogeneity cannot be considered. For example, the subject’s frequency of contact with other people, the contact location, and the severity of illness at the time of onset may vary depending on age; moreover, the risk of infection at home may also vary depending on the number and status of other family members in the household to which they each belong. Because such individual heterogeneity and diversity have a great impact on the spread of infectious diseases, it may be difficult to properly comprehend the effects of various measures such as regulatory strategies based on the results of a model if these considerations are missing from the model. From this perspective, it is considered more appropriate to use MAS, which models each individual and flexibly incorporates their heterogeneity. MAS is a simulation technique in which, instead of modeling the entire phenomenon itself, such as is usually performed in analytical SEIR models, the micro-elements (agents) constituting the phenomenon are each modeled and autonomously moved, thereby modeling the phenomena resulting from their interaction. Each agent affects its surroundings by acting on a regular basis, as it is influenced by environmental and other agent behaviors. Each of these agents behaves autonomously and interactions occur; hence, events in complex systems that are difficult to predict from individual elements can be reproduced.
Infectious disease-spreading phenomena are mainly those in which viruses are propagated by contact between humans (agent) and are caused by transmission; therefore, these phenomena are quite suitable as subjects of MAS. Vyklyuk et al. (2021) modeled the spatial distribution of COVID-19 by mobile cellular automata-based MAS, considering the differences and links between individual regions. MAS was then conducted to evaluate the impact of various factors, including movement regulations, resumption of operations, and publication of regulations, while also reproducing the characteristic conditions of some actual areas. Castro et al. (2021) scored each agent (person) for infection risk in response to a wide variety of attributes, such as occupation, health status, and purchasing power to model the spread of infectious diseases through social interactions, including agent heterogeneity and environmental diversity. The impact assessment of social distancing was conducted in conjunction with various micro-elements. Furthermore, Omae et al. (2020) evaluated the effect of the distribution of contact notification apps on the suppression of infection spread via parameters such as the app possession rate and the input rate on the virtual space reproduced by the MAS.

As can be seen from these cited studies, MAS can be used to flexibly capture and evaluate the effects of various environmental changes in virtual space. This is because the effects are modeled to actually act on each person (agent). By reflecting the effects of environmental changes on each agent from a micro-perspective and moving them (agents), the effect can be observed as a change in macro-phenomena. In analytical models such as SEIR models based on ordinary differential equations, the effect on macroscopic phenomena must be introduced into the model as it is in many situations. Miscalculation may result if the mechanisms of the phenomena are not deeply understood. In this regard, the use of highly flexible MAS is considered preferable when considering new measures to control the spread of infectious diseases, such as COVID-19.

Reinforcement learning is another tool that may be useful when considering policies for controlling the spread of infectious diseases. It is a machine learning technique that searches for behavioral policies that maximize a defined reward by repeating trial and error and optimizing the behavior in each state. In addition to evaluating the effects of various measures related to the spread of COVID-19 that are reproduced using MAS, it is hypothesized that by using reinforcement learning, it is possible to search for the optimal countermeasure strategy, such as in what order or at what timing measures should be implemented from the perspective of local government to obtain maximal effectiveness. In other words, action policy can be learned by reinforcement learning on the MAS so that measures can be dynamically changed according to the progression of infection, and the situation can be ideally controlled. Kompella et al. (2020) presented a methodology for optimizing regulatory mitigation policy by applying reinforcement learning to an agent-based pandemic simulator so as to not stress the healthcare system and to minimize impact on the economy. If such applications of reinforcement learning are used, it would be possible to create a regulatory strategy that could be used as a reference for decision-making based on realistic big data generated by MAS in the virtual space. It may be possible to break away from the current system, which relies completely on the heuristic implementation of regulatory policies.
As mentioned earlier, MAS has great potential for use in various decision-making scenarios regarding infectious disease countermeasures. However, it is difficult to say whether its utilization is progressing, especially in Japan, because there are still relatively few studies on COVID-19 using MAS, and regulatory policies are often heuristically decided. Therefore, in this study, Japanese metropolitan cities were represented using MAS to reproduce the spread of COVID-19 and evaluate the effect of deterrence measures. As an application example, we tested how the spread of infectious disease could be altered by different vaccination scenarios, with a focus on the heterogeneity of agents in terms of age, for vaccines that are currently being distributed and administered. In addition, in previous research on reinforcement learning (Kompella et al., 2020), experiments were conducted mainly using the simulator where parameters were set based on US statistical information and there were some parameters that are difficult to set in Japan due to the lack of reference data such as each detailed location setting. Therefore, it is hard to say that the current Japanese cities also can be simulated with the same simulator framework. In this study, we show an example of MAS construction for Japan, which is different from the framework and is mainly based on the contact rates according to age. Then, we searched for the optimal regulatory strategy using reinforcement learning in the MAS environment and confirmed its effectiveness.

The remainder of this paper is organized as follows. Section 2 describes the environment for the MAS based on a Japanese metropolitan city. Section 3 presents and analyzes the experimental results for the MAS. Finally, Sect. 4 concludes this paper.

2 Simulation environment

The simulator setup and the environment for reproducing the infectious disease spread are described in this section.

2.1 Agents and environment settings

First, we reproduce the movements of people in daily life and their contact pattern. We assume N agents and begin modeling them by allocating them to households. Because the environment in this study is based on a Japanese metropolitan city, the N people are divided by the ratios of the population 5-year age groups (0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–) based on the statistical information for Tokyo (Tokyo Metropolitan Government, 2020). They are approximately assigned to each household using information about household types by age group and the number of family members (Statistics Bureau, 2021). By reflecting data mainly from Tokyo on the settings about the distribution of family structures and contact rate in consideration of the age group, we imitate the large cities of Japan on the MAS. In the case imitating small and medium cities, it is expected that the decrease of the proportion of single-person households, the increase of the proportion of elderly households etc. will be reflected compared to large cities. At this time, following Kompella et al. (2020) and Omae et al. (2020), N = 1000, and several persons are removed owing to the effect of rounding to whole
numbers in the process of the assignment. The agent of the remaining persons is made to act autonomously in the simulation. Each agent has age classification and household identification information. Home, school, office, and others are assumed to be action areas (hereinafter, they are referred to as “Home area,” “School area,” “Office area,” and “Other area,” respectively), and they are assigned to a grid-like field (Fig. 1a). In the “Home area,” it is assumed that each household corresponds to a cell, and that the household members are located in their own household cell when they are at home, whereas the other three areas are defined as areas that each consists of multiple grid cells representing each action field. When agents are in one of those three areas other than the “Home area,” they can stay or move to other cell in that area with equal probability for each step. Then, the model is designed to mimic people’s daily movements (e.g., getting up and going to work, going home, and going to bed) by considering a step as an hour in the real world. Specifically, from 9:00 to 17:00, a certain percentage of people move to the “School area” and “Office area,” and from 17:00 to 19:00, a certain percentage of people move to the “Other area,” and other people are in their homes at “Home area.”

Subsequently, the concept of contact between agents is introduced here. Agents located in the same square are considered to have come into contact with each other. Because the home is set to one square (cell) for each household, contact is made with other household members at each step. Thus, in “Home area,” we set members to contact stochastically even if they belong to the same cell. If a rule is set here that a susceptible person can be infected when it comes into contact with infected people, the spread of infection can be reproduced (Fig. 1c).

In the above scenario, the average number of contacts per day in the simulation is determined by the probability of contact at home, the population density in each area (“School area,” “Office area,” “Other area”). Therefore, they are adjusted such that the average number of contacts in each area per day on the simulation approximately
matches the actual average number of contacts in each area on weekdays and weekends, as reported by Munasinghe et al. (2019) by age group. In addition, considering that people who normally come into contact in daily life do not change significantly in the school and office areas, the destination cell at the time of transitioning from home to each area is fixed for each agent, and movement within an area is also limited to the nearby cell. However, in “Other area,” the migration destinations are randomly determined from all cells in the area in every step (Fig. 1b).

2.2 Introduction of infection mechanisms

In the previous section, the simple daily contact mechanism with other people is constructed by determining the fields and the action rule of people in MAS. The infection mechanism is introduced here, considering state transitions based on the SEIR model, as also used by Omae et al. (2020). In other words, we assume a flow, as shown in Fig. 2, as the transition between states for each agent. The Susceptible state (S state) is a condition in which a person can be exposed to the virus and infected by being in contact with a person who has already been infected. When a person in the S state is actually infected by contact, they transition to the following Exposed state (E state). This condition represents an incubation condition that has not yet led to the onset of disease. At the symptom onset a few days after entering the E state, the state enters the Infectious state (I state) and, after a recovery period, it shifts to the Recovered state (R state). Because the number of patients experiencing recurrence is small in Japan, we do not consider the possibility of transition from the R state to

![Agent-based model (COVID-19 infection)](image)

Fig. 2 Agent state transition. When the agent is infected with a virus, it enters the Exposed state and is infectious. The infectivity becomes strongest before and after the onset. However, except for asymptomatic patients, it is assumed that isolation is properly performed at the time of onset and the infectivity is lost in the simulation.
the S state. In the case of the standard SEIR model, people in the I state are often infectious, but there are reports (He et al., 2020) that the infectivity becomes strongest approximately 3 days before to 5 days after the onset of COVID-19. Therefore, we make people in the E state infectious as well (Fig. 2). In addition, if a person in the I state is not asymptomatic, it is assumed that the person is properly isolated when the person enters the I state and that the infectivity in the simulation is lost. The probability of infection when in contact with an infectious person is proportional to the shape of the infectivity distribution before and after the day of onset, as indicated by He et al. (2020). In other words, when S, E, I, and R represent the susceptible, exposed, infectious, and recovered states, respectively, and $X_i^t$ is taken as the state of the $i$th agent at $t$ step, the probability of transitioning from an S state to an E state when the agent contacts others at $t$ step is as follows:

$$P(X_{i+1}^t = E | X_i^t = S, C_{i,j}^t = 1, T_{E \rightarrow I}^j, X_j^t, H_j^t) = \begin{cases} \max(1, \beta f(t - T_{E \rightarrow I}^j)), & X_i^t = 1 \text{and } H_j^t = 0, \\ 0, & \text{other.} \end{cases}$$

$C_{i,j}^t = 1$ indicates that the $i$th agent contacted the $j$th agent at $t$ step, and $C_{i,j}^t = 0$ indicates that it did not. $T_{E \rightarrow I}^j$ represents the step when the $j$th agent enters the I state from the E state. $H_j^t = 1$ indicates that the $j$th agent is in a symptomatic I state and is isolated at step $t$ and $H_j^t = 0$ indicates that it is not. In addition, $\beta$ represents base infectivity, and $f(t - T_{E \rightarrow I}^j)$ represents the distribution of the timing at which the secondary infection occurs through the $j$th agent, and is centered on the onset step $T_{E \rightarrow I}^j$ of the $j$th agent (the gamma distribution was set based on He et al. (2020)). Then, the probability of transition from the E state to the I state is as follows:

$$P(X_{i+1}^t = I | X_i^t = E, T_{E \rightarrow I}^i) = \begin{cases} 1, & t + 1 = T_{E \rightarrow I}^i, \\ 0, & \text{other.} \end{cases}$$

$T_{S \rightarrow E}^i$ represents the step when the $i$th agent is infected and shifts from the S state to the E state; the step $T_{E \rightarrow I}^i$ when the $i$th agent develops and enters the E state is determined to be $T_{E \rightarrow I}^i = T_{S \rightarrow E}^i + \lambda$, and the period $\lambda$ in the E state is generated from the gamma distribution ($\lambda \sim \text{Gamma}(k, \theta)$). For the mean ($k\theta$) and variance ($k\theta^2$) of the gamma distribution, the values for the incubation period reported by the National Institute of Infectious Diseases (2021) were used. Although there are some differences among people, after $k\theta$ hours on average from entering the E state, the state changes from the E state to the I state. Furthermore, when the agent enters the I state, the severity attribute is assigned to the agent. Specifically, asymptomatic, mild, moderate, and severe cases are used as symptoms (Fig. 2), and those in age group $l$ are assigned probabilities of $p_{\text{Asymptomatic}}^l$, $p_{\text{Mild}}^l$, $p_{\text{Moderate}}^l$, and $p_{\text{Severe}}^l$, respectively ($p_{\text{Asymptomatic}}^l + p_{\text{Mild}}^l + p_{\text{Moderate}}^l + p_{\text{Severe}}^l = 1$). In other words, the fact that the symptom severity of COVID-19 when an agent is in the I state depends on age and
the characteristics of COVID-19, which tend to become more severe in the elderly, are considered to reflect the effects of age heterogeneity on the simulation of infection spread and bring it closer to reality. For the values of those parameters here, the ratio values reported by National Center for Global Health and Medicine (2020) and Alene et al. (2021) are used (the asymptomatic rate is 25% regardless of age category). The probability of transition from the I state to the R state, which is the final transition, is as follows:

\[ P\left( X_{i,t+1} = R | X_{i,t} = I, T_{i \rightarrow R}^i \right) = \begin{cases} 1, t + 1 = T_{i \rightarrow R}^i, \\ 0, \text{other}. \end{cases} \]

\( T_{i \rightarrow R}^i \) represents the step at which the \( i \)th agent has been infected and has transitioned from the I state to the R state, as defined in \( T_{i \rightarrow R}^i = T_{S \rightarrow E}^i + \gamma^i \). \( \gamma^i \) is uniformly set to 15 days if the severity of the \( i \)th agent is severe, and 10 days otherwise, based on the discharge criteria for COVID-19 patients set by the Japanese Government (Ministry of Health, Labour and Welfare, 2021).

By adding the aforementioned settings to the MAS environment configured in Sect. 2.1, it is possible to simulate the spread of COVID-19. However, because there are no data that can be referred to regarding the base infectivity \( \beta \), for the sake of simplicity, the value of \( \beta \) is tentatively set so that the cumulative number of infected persons after the spread and convergence of infection in the MAS, which is when there are no more people in the E or I state, would be close to the predicted cumulative number of infected people in Tokyo (Kurita et al., 2020). Therefore, the value \( \beta \) is adjusted so that approximately 60% of the total population will experience infection. The parameters used in the MAS in this study are summarized in the Appendix. Section 3 presents case studies for controlling the spread of infection using the MAS designed in this section.

### 3 Case studies

Case studies involving examining a vaccine strategy using MAS alone and searching for an optimal regulatory strategy using MAS in combination with reinforcement learning are presented in this section. The extent of the spread of an infectious disease depends largely on each person’s resistance to the virus and the frequency and circumstances of contact with others. Therefore, as realistic and effective countermeasures that can control each, vaccination and influential government regulation are focused here. In fact, even in the real world, these two countermeasure categories are the most important keys in suppressing COVID-19.

#### 3.1 Vaccination strategy focusing on inoculation age

We explore appropriate measures to prevent the spread of COVID-19 infections using the MAS environment presented in the previous section. Developing and administering vaccines are strongly required for the population to carry out their daily life in a
way similar to that before the spread of COVID-19, without implementing measures to restrict action, such as outdoor restrictions. In Japan, the acquisition and distribution of the Pfizer and Moderna vaccines were being promoted in early 2021, and administration of these vaccines begun, preferentially for medical workers, the elderly, and people with comorbidities.

Many local governments had been vaccinating in order of decreasing age to suppress the severity of illness, but some local governments were trying to prioritize vaccination for young people because they have a wide range of activities and it is easy for them to become a source of infection. In view of this situation, we investigate how the infection situation differs depending on the different priorities for vaccination according to age. Considering that there are up to about 300,000 vaccinations per day in Tokyo, which represents about 2% of the population of Tokyo, it is assumed that 20 agents are vaccinated every day based on the scale of the MAS. Specifically, at the end of day 1 of the simulation, 20 people are given a vaccine at the same time and, as an effect of the vaccinations, a vaccine efficacy of 64% is applied, based on the information on vaccine efficiency provided by Ministry of Health (Israeli) (2021). That is, the base infectivity to the vaccinated agent $\beta_v$ is calculated as $\beta_v = (1 - 0.64)\beta$. The spread of infection when the vaccination is not considered (Fig. 3a) and when the daily vaccination is begun at the start of the simulation (Fig. 3b), as described earlier, is shown. The mean evolution of the number of individuals in each state of S, E, I, and R for the 50-trial run is depicted, along with the 95% confidence interval at each time point. The number of infected individuals at step 0 is set to 10, and the individuals are assigned to the E state. The order of vaccination is set to be completely random. It can be confirmed that the reduction in the susceptible population in the case that includes vaccination occurs earlier than in the case without vaccination, and that the infection spread converges relatively quickly and the number of infected individuals is much lower.

![Fig. 3 Spread of infection. a Changes in the number of agents in each state in the case where vaccination is not considered. b Changes in the number of agents in each state in the case where vaccination is considered. Every day, 20 random people are vaccinated, and from the next day onwards, their infection rate is 64% lower than before. (The average at each time point is drawn (line plot) together with the 95% confidence interval (shadow part) by running the simulation 50 trials in (a), (b))](image-url)
We then observe how the transition and the number of infected patients differ depending on the vaccination priority. To do this, we assume the following four scenarios for choosing which 20 individuals to inoculate per day. 1: Random vaccination; 2: vaccination of individuals from the oldest age categories; 3: vaccination of individuals from the youngest age categories; 4: vaccination of individuals over 60 years of age (60–), then individuals in their 20–30 s (20–39), and finally everyone else (0–19, 40–59). Currently, many local governments are promoting vaccination with a strategy such as scenario 2, but for areas where young people gather, a vaccination strategy like scenario 4 is being considered.

Figure 4 shows the number of individuals with symptoms of mild severity, moderate severity, and severe severity for each scenario. 50-trial simulations were run, and their average transitions were depicted with 95% confidence intervals for each time point. Although the number of days to convergence is not considerably different between
the four scenarios, the number of patients, particularly those with mild severity, can be seen to differ.

To make it easier to compare the outcomes due to different scenarios, the cumulative number of patients in each scenario was drawn using a box plot for those with mild severity, those with moderate severity, those with severe severity, and the sum of all severities (Fig. 5). The number of patients with mild severity (Fig. 5a) was relatively high in scenario 2, that is, the vaccination scenario in descending order of age. This may be attributed to the fact that the postponement of vaccination for young individuals made infection more likely among young individuals, as well as the lower risk of severity when young individuals are infected with COVID-19. Therefore, the number of mild symptoms was relatively low in scenario 3, where vaccination was conducted in young individuals. However, that number in scenario 3 does not differ significantly from the number in the randomly inoculated scenario 1. The number of people having

![Box plot of the cumulative number of patients for each vaccination scenario](image)

**Fig. 5** Box plot of the cumulative number of patients for each vaccination scenario: a mild severity; b moderate severity; c severe severity; d mild or moderate or severe severity—total of (a–c). (The samples for the box plots in (a–d) are obtained by running the simulation 50 trials for each vaccination scenario)
symptoms of moderate severity (Fig. 5b) did not differ greatly between scenarios. The number of people having symptoms of moderate severity in scenario 3 is distributed in a relatively large number band because the median is the highest of the four scenarios. This may be attributed to the higher probability of infection in the elderly when young individuals are given vaccination priority and the high likelihood of moderate severity in older individuals.

For the number of patients with severe severity (Fig. 5c), it was distributed lower in scenario 4 than in scenarios 2 and 3. This may be due to the reduction of the overall infection scale by suppressing the transmission of infection among individuals in their 20s (20–29), who have a relatively large number of contacts in school and office settings, through vaccination of individuals in their 20–30s after the completion of the vaccination of people in their 60s (scenario 4), who are more susceptible to aggravation. In scenarios 2 and 4, although individuals 60 years of age and older completed vaccination at an early stage and had some antibodies, they rarely entered the “School area” or “Office area,” and they did not contribute to herd immunity in those areas. In scenario 2, the immunization in “Other area” may have been enhanced by vaccinating older people at an early stage, but it can be inferred that during this time when vaccinating older people is being promoted, the spread of infection in “School area” and “Office area” had progressed, and the risk of severity was eventually reduced. As a consequence, for the sum of the mild, moderate, and severe cases, the value is distributed higher in scenario 2 than in other scenarios (Fig. 5d). From the aforementioned results, it is suggested that attention should not only be paid to the severity risk based on biological information such as the age and health status of individuals, but also to the prevention of the infection spread in each area. Therefore, it may become important for the vaccination strategy to control the number of severely ill patients.

As described earlier, in the MAS, changes in the influence on individuals, such as vaccination, can be flexibly incorporated by changing the parameters and the behavior pattern of the agents; moreover, the effect of measures can be sufficiently studied in virtual space. In this study, we divided areas into relatively simple and basic areas, such as the school area, office area, and other areas. However, by changing the structure of the simulation according to the characteristics of the region to be examined, an investigation according to various places and situations can be carried out. For example, using areas such as bars and clubs and reproducing the movement of people during the night period would allow us to model the relationship between the spread of infection and the nightlife districts. The superiority of strategies such as scenario 4 would be clearly evident when considering vaccination strategies in areas where youth infections are expected to spread.

The strength of MAS is that it can incorporate specific response strategies tailored to a situation with micro-elements. It can model them in virtual space and confirm macroscopic effects. In addition, MAS can be used in combination with reinforcement learning to provide a reference or alternative response strategy. Its applications are illustrated in the next section.
3.2 Dynamic regulatory strategies

If an effective vaccine is already available, it is expected that it will be possible to acquire population immunity and suppress the spread of infection; if none is available, limitations such as outdoor movement and travel control are required, as in the early stage of the COVID-19 pandemic. The question is how strictly and for how long these measures should be implemented to effectively suppress the pandemic. Strict action limitations, such as lockdowns, certainly reduce the frequency of people contact and reduce the number of new infections. However, when action restrictions are imposed, restaurants and retail businesses cannot do business, and therefore the economy suffers. In addition, because freedom of movement is restricted and a state of depression continues, it is not preferable that movement regulations with strong enforcement continue for a long period for the sake of mental health. In other words, it is necessary to search for optimal regulatory strategies that do not impose regulations too strictly, but still limit the pandemic. More specifically, it is rational to dynamically switch the regulation level so that strict behavioral measures are taken when the risk of infection spread is high, and regulations are relaxed as the risk decreases. In reality, the timing of such tightening and relaxation of regulatory levels is often based on the knowledge of epidemiological experts and trial and error. However, the current situation is the situation in which people’s behavioral changes are intricately intertwined, and it is difficult to understand it only by such heuristic measures. Therefore, it is likely that action regulations will be extended beyond the scheduled period, or that regulations will be relaxed during periods when, from a retrospective point of view, they need to be strengthened. To avoid such situations, in addition to a MAS that can flexibly reflect situation changes in the simulation, reinforcement learning that learns the optimal behavior strategy by trial and error may be key. By combining reinforcement learning with a virtual world constructed using MAS, it may be possible to determine the best measures to take for a given situation, with a view to maintaining the economic and social status quo. Here, reinforcement learning is used to search for regulatory strategies, from the perspective of local governments, regarding when and how much restriction should be placed to effectively control the spread of infection while minimizing the economic damage.

As for incorporating reinforcement learning into the framework, the form shown in (Fig. 6) is assumed. The local government takes the role of the Agent in reinforcement learning, and the society in which people live takes the role of the Environment. Action is equivalent to enforcing regulations, and the information on societal conditions after being affected by that Action is returned as State. In addition, Reward is also returned as an indicator to evaluate the goodness of the Action, which is how much it contributed to the effective control of the pandemic. Thus, the Agent (local government) explores effective infection control strategies so that the sum of rewards returned for each action is maximized. More specifically, the setting content of each item in this experiment is described.

First, as the Environment, we take the society of a Japanese metropolitan city, and consider the MAS environment used in Sect. 2, as mentioned earlier. Then, the Agent (in reinforcement learning) assumed is the local government. Subsequently,
Fig. 6 Reinforcement learning for optimizing regulation strategy. Agent, Environment, Action, and State in reinforcement learning for optimizing regulation strategy correspond to the following. (Agent: local government, Environment: society, Action: issuance of regulations, State: infection spread status in society). Reward is set according to the purpose the regulatory content that the local government can take as action is set up. As regulation options, the four stages in Table 1 were assumed with reference to the stage classification of the infection status (Office for Novel Coronavirus Disease Control, 2020), which is notified to the government of each prefecture in Japan. As the stage advances, regulatory conditions are added and become stricter. The contents of each regulatory condition are set to be very simple.

The situation where Stage 1 should be enforced in the real world corresponds to a situation in which infections occur sporadically, there is no trouble in the medical system, and no regulations are in place. Stage 2 corresponds to the response when the number of infected people is gradually increasing; when contact with infected people is discovered, isolation at home is imposed. More specifically, if a person experiences illness onset (enters the I state) and is not asymptomatic, all those who had contact

| Stage | Home isolation from the time of detection of contact with infected person | Going out restrictions (decrease in rate of moving to “Other area”) | Telework promotion (decrease in rate of moving to “Office area”) | School closure (no one moves to “School area”) |
|-------|-------------------------------------------------------------------------|------------------------------------------------------------------|-----------------------------------------------------------------|-----------------------------------------------|
| Stage 1 | False | None | None | False |
| Stage 2 | True | None | None | False |
| Stage 3 | True | 20% | 20% | False |
| Stage 4 | True | 50% | 20% | True |
with that infected person in the “School area” or “Office area” over the previous two days are isolated at home for 10 days. Each period (2 days, 10 days) is approximately set based on the general response in Japan. Assuming that if someone has an onset of disease in the office or school, that information will most likely be passed on to those who have come into contact with them in the past few days. Only the contact history in areas corresponding to school and office is referred to, as we decided that it would be difficult to identify the people who made contact in other areas because the distribution volume of contact apps is not large.

Stage 3 corresponds to the response at the stage when the number of infected people increases rapidly, and major problems may occur in the medical care system. In addition to the regulations of Stage 2, going out regulations and telework promotion are imposed. Regarding going out restrictions, the frequency of transitions to the “Other area” is reduced by 20% on average compared with no restriction. Regarding the promotion of telework, about 20% of people considered that they were in an environment where they could telework, and that they were at home instead of transitioning to the “Office area”. This figure of 20% was approximately set with reference to the situation in the early days of the COVID-19 epidemic in Japan. The strictest stage, Stage 4, corresponds to measures to prevent the explosive spread of infection and malfunctioning of the medical care provision system, the further tightening of restrictions on going out (transition to the “Other area”), and reducing to zero the probability of transition to the “School area” by closing schools. To make it easy to understand the effect of each stage, the transition of the number of people in each state when each stage is not switched is shown in Fig. 7. (Changes in the average number of people in each state for 50 trials is shown together with the confidence interval at each time point. Ten random people are in the E state at the initial time point.)

The purpose of this study is to determine the timing for switching the stage and which stage should be switched to lead to an effective infection spread control strategy using reinforcement learning. Agent (local government) selects from the aforementioned four stages at regular intervals as Action. This time, because a certain period for the measures to take effect is needed and the measures will not likely change within a short time, such as 1 week, we let the agent (in reinforcement learning) select the stage at intervals of two weeks from step 0. In other words, we repeat a cycle of selecting which stage regulation to apply (Action) in consideration of the current infection status (State) and receiving Reward and new State two weeks after applying it. The observable number of people who are considered to be in the S state and the number of new symptomatic infections in the previous week are considered as the State for understanding the status of infection spread, where they correspond to the total population, excluding people who have experienced symptomatic infection (people who have experienced an asymptomatic severity I state), and the number of people who have entered the I state that is not asymptomatic in the past week, respectively.

Next, the Reward is determined. The maximum socially acceptable spread of infection is defined as the range in which the number of COVID-19 patients requiring hospitalization does not exceed the number of hospital beds. In this experiment, patients with an illness of moderate or severe severity are designated as patients requiring hospitalization (hereinafter referred to as inpatients), and we aim to control the
Fig. 7 Spread of infection for each regulation stage. Changes in the number of agents in each state in the case where a Stage 1 in Table 1, b Stage 2 in Table 1, c Stage 3 in Table 1, or d Stage 4 in Table 1 is enforced for the entire period. (The average at each time point is drawn (line plot) together with the 95% confidence interval (shadow part) by running the simulation 50 trials in (a–d)).

infection status so that the number of those inpatients does not exceed five. It is tentatively set to five by matching the target number of beds for COVID-19 in Tokyo (House of Councilors, 2020) with the simulation scale. This time, we would like to find an efficient regulatory strategy that can control the spread of infection so that it falls within the aforementioned allowable range, while also suppressing economic loss as much as possible. Therefore, as Reward, the sum of $r = -a \max \left( \frac{n_h-n_{\text{bed}}}{n_{\text{bed}}}, 0 \right) - b \frac{(\text{Stage}-1)^p}{\max_{\text{stage}}(\text{Stage}-1)^p}$ on each day until the next Action is selected is used. $n_h$ represents the number of inpatients, and $n_{\text{bed}}$ represents the maximum number of beds, that is, the medical capacity ($n_{\text{bed}} = 5$ in this experiment). Stage represents the regulatory stage in force on that day (Stage $\in \{1, 2, 3, 4\}$). $a$, $b$, and $p$ are positive constant parameters that adjust the weight of each element considered as Reward (set $a = 0.4$, $b = 0.1$, $p = 2$ in this experiment). This Reward was also used in Kompella et al. (2020), and if the number of hospital beds is insufficient, the loss due to the first term of $r$ will increase. By contrast, if regulations are tightened to suppress the spread of infection, the loss due to the second term will increase. Therefore, the more loosely regulated the strategy,
the greater the reward obtained, as long as the number of patients requiring hospitalization does not exceed the number of hospital beds. A simple Q-learning algorithm (Watkins & Dayan, 1992) was used as the learning algorithm, and each continuous value given as State was discretized into 10 divisions. We define one episode as 20 simulated weeks starting from the situation where the number of initial infections in the E state was 10, and the episodes were repeated 3000 times to train the regulation policy.

In addition to the regulation policy learned with Q-learning, two predetermined policies, A and B, are prepared to examine the features of the learned policy. Policy A is determined based on the guideline of the stages in Office for Novel Coronavirus Disease Control (2020), and the stage change is judged on a daily basis so that the behavior can be drawn more faithfully according to the guideline. If the number of inpatients exceeds 1/5 of the maximum number of beds at the end of the day, the regulation stage is switched to Stage 3, and if it exceeds 1/2 of the maximum number of beds at the end of the day, the regulation stage is switched to Stage 4. If this does not apply, Stage 2 is set, and if new symptomatic patients do not appear for a certain period (set to 2 weeks), Stage 1 is set from the following day. Policy B is a strategy that continues Stage 4 in the case where regulations are tightened. However, the stage is set to be lowered by one after a certain period (set to 2 weeks) of no new symptomatic patients. If a symptomatic patient appears on the way, the regulation stage is switched to Stage 4. Simulations of regulation switching as appropriate based on Policy A, Policy B, and the Policy learned by reinforcement learning (RL) are performed over 50 trials, and the changes in the average number of inpatients (Hospitalized), E state populations, and I state populations are shown with 95% confidence intervals at each time point (Fig. 8). The horizontal dotted line represents the medical capacity (= 5).

Under Policy A (Fig. 8a), the number of infected people is relatively large among the three policies, and there are times when the number of Hospitalized people slightly exceeds the medical capacity. As for Policy B (Fig. 8b), because it is in the strictest stage for a long time, it reflects that and converges earlier; moreover, it does not exceed the medical capacity. Under Policy RL (Fig. 8c), the transition of Hospitalized people is within the range that does not exceed the medical capacity, and there are longer

![Fig. 8 Spread of infection when each policy is applied. Changes in the number of agents in each state in the case where the stage is dynamically switched based on a Policy A, b Policy B, c Policy RL. (Hospitalized state is defined as the I state of moderate or severe severity.) (The average at each time point is drawn (line plot) together with the 95% confidence interval (shadow part) by running the simulation 50 trials in (a–c))](image-url)
periods when Hospitalized people exist than under Policy B. In other words, it can be said that this is the result of loosening regulations as appropriate based on the learning content in reinforcement learning.

Considering the infection situation, the regulation stage was selected so as not to burden the economy. For a more specific view, the burden on medical care and the burden on economics are simply defined and viewed for each policy. First, the number of inpatients exceeding \( n_{\text{bed}} \) is calculated as \( l_1 = \max(n_h - n_{\text{bed}}, 0) \) on each day, and the burden on medical care (Medical Burden) is defined as their sum (\( \sum l_1 \)) over one episode. Then, \( l_2 = (\text{Stage} - 1)^p \) used in the reward for reinforcement learning is calculated on each day, and the burden on the economy (Economic Burden) is defined as their sum (\( \sum l_2 \)) over one episode.

For each policy, the box plot of the burdens calculated in each trial is shown in (Fig. 9). Compared to the other two policies, Policy A is a policy in which the regulation is weaker in the earliest situation when the initial infected people are confirmed, and the regulation is strengthened immediately after the spread of infection becomes more serious. Therefore, the measures to suppress the spread of infection have been delayed, and as a result, the infection has spread more than the other two and the burden on medical care has increased (Fig. 9a). In Policy B, the spread of infection is suppressed because the regulations are strict from the beginning, but the period of such strict regulations is relatively longer than the other two, and the burden on economics has

---

**Fig. 9** Box plots showing the magnitude of burden when each policy is applied. **(a)** Medical burden representing the degree by which the medical capacity is exceeded in each policy. **(b)** Economic burden representing the strictness of the regulatory stage enforced in each policy. (The samples for the box plots in (a), (b) were obtained by running the simulation 50 trials for each policy)
increased (Fig. 9b). In Policy RL found by learning by trial and error, the regulation is tightened at the initial stage like Policy B, but the deregulation from there is performed at an earlier timing than Policy B. Then, by lowering the regulation level relatively slowly until the complete deregulation, the regulatory adjustment is made well so that the spread of infection is suppressed and the impact on the economy is not great. Automatically finding the best timing for regulatory switching is a strength that fixed policies do not have, and by setting Reward based on the desired goal in this way, even on MAS where micro factors such as age are intricately intertwined like this time, reinforcement learning can provide useful strategies that can be used as a reference.

There was little State information this time because the experiment was conducted in a very simple setting; hence, we performed reinforcement learning with Q-learning. However, when applying reinforcement learning in the form of this time to a concrete real world study, it is thought that more detailed and diverse State information is required. However, even in such cases where a large amount of State information must be processed, learning can be performed by applying deep reinforcement learning, and MAS can be used as a learning environment. When various elements are intricately related, as with COVID-19, and it is difficult to manage using only a heuristic strategy, it may be an option to use a strategy found by reinforcement learning in a virtual world (MAS).

4 Conclusion

In this study, we constructed a simulation of COVID-19 spread in a MAS environment that represented a Japanese metropolitan city by considering differences in background information such as age group and household structure. It shows an example of the MAS framework for the spread of COVID-19 in Japan. Among various micro information, the MAS setting is made with particular consideration for the interaction with age information, which is an important key for COVID-19 infection. Therefore, it is possible to more realistically grasp and examine the spread of infection itself and the effects of age-related measures. Through experimentation, we showed that the flexibility of MAS modeling could be utilized for strategic studies to prevent the spread of infection. In particular, by observing how the changes in the spread of infection differ depending on the scenario of the vaccination priority based on age group, it was shown that MAS could examine strategies that consider detailed micro-information. In addition, we showed that the optimization of dynamic prevention strategies can be tackled in virtual space by combining reinforcement learning with MAS using a case study of determining a regulatory strategy that also considers economic impact. While many studies are based on analytical models in response to the spread of COVID-19, this study presents the direction of MAS utilization. MAS can reproduce complex systems that are realistic by considering the interaction of micro-elements, which is difficult for analytical models. Therefore, it should be further utilized in the study of countermeasures for the spread of infectious diseases such as COVID-19, which is a phenomenon that occurs owing to the interaction of people and the environment.
It should be noted that this study has several limitations. First, we are considering only the heterogeneity which is relatively populous in each class in the real world, such as age group this time. Therefore, there is no problem even if the number of agents is about 1000, but it is necessary to prepare more agents when considering micro-variation in finer grain size such as detailed classification of occupations and further area division. In addition, the mechanisms and nature of COVID-19 viruses, such as the presence of super spreaders and the appearance of new strains, are not considered, and only basic properties are included in the simulations.

Furthermore, although some assumptions are based on the global situation before or during the spread of COVID-19 infection, considering that people’s work styles and services will be different after the infection has converged, and the assumptions could change significantly, it is difficult to actually use the results of this example in the actual study of measures for a real city. However, it is possible to get closer to a more realistic model that can be used for such studies. For that purpose, along with running the simulation at a larger scale, additional settings, such as a more detailed classification of areas, an introduction of differences in people’s movements within each area, consideration of migrants from outside the prefecture, and subdivision of agent occupations, are required. In addition, as new knowledge about COVID-19 is obtained, it is possible to make the phenomena reproduced by MAS more realistic and truer by incorporating them into MAS as needed. The merit of MAS lies in the fact that the discovery of these new mechanisms and changes in social conditions can be incorporated into the virtual world in a relatively flexible and dynamic manner. If we form a base MAS environment, we do not need to build a model from the ground up every time we run a new study, and by updating parameters and adding settings as needed according to changes in the situation and the times it will be possible to immediately respond to detailed studies. If the phenomena can be reproduced without delay in this way, it will be possible to immediately derive the outlook for optimal infection control measures mechanically using reinforcement learning. This will be of great help for decision-making. In addition, regarding reinforcement learning, after learning is done in the base MAS environment, even if the MAS parameters change slightly according to changes in the actual environment and as long as the tasks of reinforcement learning are similar, the information learned before can be used as the initial value to help improve the efficiency and speed of new learning. In other words, transfer learning within reinforcement learning is required. Based on these findings, we believe that it is necessary to construct a large-scale MAS that precisely reflects all agent information, such as behavior patterns, contact status, household information, and occupation, in each region of Japan that can be used as the basis of a study model for future pandemics. In addition, we would like to develop a reinforcement learning system that could learn the optimal policy on the MAS and mechanically present every decision, such as the timing of self-restraint request issuance and the request period, the strictness of regulation, school closure periods, shortened business periods, and deregulation orders, to improve the current COVID-19 situation and to prepare for a new pandemic that may occur in the future. This research is planned as future work.
| Parameter | Value | Source |
|-----------|-------|--------|
| Age category rate (0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–) | Based on Tokyo population age distribution | Tokyo Metropolitan Government (2020) |
| Household id (family type of household) | Based on Tokyo household distribution | Statistics Bureau (2021) |
| Simulator steps in 1 day | 24 (Home area: 14; School/Office/Home area: 8; Other/Home area: 2) | |
| Area size (the number of cells in each area) | School area: 200 | |
| | Office area: 1200 | |
| | Other area: 400 | |
| Rate of movements to each area by age category (0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–) | School area (weekday): [0.15, 0.7, 0.82, 0.85, 0.43, 0.08, 0.01, 0.01, 0.02, 0.01, 0., 0.01, 0., 0., 0.0, 0.] | |
| | School area (weekend): [0., 0.17, 0.35, 0.45, 0.3, 0.7, 0.04, 0.02, 0.02, 0.01, 0., 0., 0., 0.] | |
| | Office area (weekday): [0., 0., 0., 0.04, 0.31, 0.91, 0.58, 0.7, 0.71, 0.73, 0.71, 0.78, 0.51, 0.24, 0.02] | |
| | Office area (weekend): [0., 0.0, 0., 0.08, 0.23, 0.32, 0.39, 0.42, 0.21, 0.35, 0.45, 0.37, 0.29, 0.29, 0.01] | |
| | Other area (weekday): [0.69, 0.96, 0.53, 0.28, 0.26, 0.27, 0.53, 0.41, 0.43, 0.38, 0.42, 0.5, 0.59, 0.71, 0.74] | Values are adjusted so that the average number of contacts in each area per day is close to the reported number in (Munasinghe et al., 2019) |
| | Other area (weekend): [0.39, 0.83, 0.52, 0.52, 0.49, 0.56, 0.45, 0.56, 0.55, 0.5, 0.55, 0.58, 0.61, 0.75, 0.68] | |
| Contact rates at Home area by age category (0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–) | (weekday): [0.07, 0.12, 0.13, 0.15, 0.15, 0.17, 0.10, 0.14, 0.16, 0.17, 0.17, 0.16, 0.16, 0.14, 0.15] | |
| | (weekend): [0.06, 0.09, 0.11, 0.13, 0.15, 0.15, 0.09, 0.11, 0.12, 0.14, 0.15, 0.13, 0.16, 0.15, 0.15] | |
| λ: incubation period (period of E state) | λ ~ Gamma(k, θ) | National Institute of Infectious Diseases (2021) |
| | k: shape parameter | |
| | θ: scale parameter | |
| | kθ^2 = 4.82 | |
| | kθ^2 = 2.712 | |
| f(t): distribution of secondary infection | f(t) = dGamma(t/24 + 12.27; 20.52, 0.63) | He et al. (2020) |
| | dGamma(t; k, θ): probability density function of gamma distribution | |
| | Gamma(k, θ) | |
Table 2 (continued)

| Parameter                                      | Value                                           | Source                                      |
|------------------------------------------------|-------------------------------------------------|---------------------------------------------|
| Rate of asymptomatic patients                  | 0.25                                            | Alene et al. (2021)                         |
| Rate of severity by age category in symptomatic patients (0–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70-) | mild: [0.97, 0.95, 0.92, 0.87, 0.71, 0.65, 0.49, 0.38] moderate: [0.03, 0.05, 0.05, 0.11, 0.23, 0.27, 0.36, 0.50] severe: [0., 0., 0.02, 0.02, 0.05, 0.09, 0.15, 0.12] | National Center for Global Health and Medicine (2020) |
| $\gamma^i$: recovery period for $i$ th agent (period of I state) | (severity of $i$ th agent is severe): $15 \times 24$ (steps) (severity of $i$ th agent is not severe): $10 \times 24$ (steps) | Ministry of Health, Labour and Welfare (2021) |
| $\beta$: base infectivity                      | 0.5                                             | Value is adjusted so that the cumulative number of infected patients is close to the predicted number in (Kurita et al., 2020) |
| Vaccine efficacy                                | 0.64                                            | Ministry of Health (Israeli) (2021)          |

The main parameters used in the simulation (Parameter), their set values (Value), and reference data (Source) are listed. Values given as vectors are age-stratified with values corresponding to age category in Parameter column.

Implementation of multi-agent simulation was done on Python 3.7.7 using mesa (https://mesa.readthedocs.io/en/master/tutorials/intro_tutorial.html). Line plots and box plots were drawn using the Python data visualization library Seaborn (seaborn.relplot and seaborn.boxplot, respectively).

Declarations

Conflict of Interests Satoki Fujita is a full-time employee of, and holds stock in, Shionogi & Co., Ltd. Ryo Kiguchi is a full-time employee of, and holds stock in, Shionogi & Co., Ltd. Yuki Yoshida is a full-time employee of, and holds stock in, Shionogi & Co., Ltd. Yoshitake Kitanishi is a full-time employee of, and holds stock in, Shionogi & Co., Ltd.

Appendix Parameter settings

We show the parameter values used in this simulation study in Table 2.

References

Alene, M., Yismaw, L., Assemie, M. A., Ketema, D. B., Mengist, B., Kassie, B., & Birhan, T. Y. (2021). Magnitude of asymptomatic COVID-19 cases throughout the course of infection: A systematic review and meta-analysis. PLoS One, 16(3), e0249090. https://doi.org/10.1371/journal.pone.0249090
Castro, B. M., de Melo, Y. A., Dos Santos, N. F., da Costa Barcellos, A. L., Choren, R., & Salles, R. M. (2021). Multi-agent simulation model for the evaluation of COVID-19 transmission. *Computers in Biology and Medicine, 136*, 104645.

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine, 26*, 672–675.

House of Councillors. (2020). The National Diet of Japan: Efforts to secure beds in response to new coronavirus infections. (accessed 2021–08–15)

Kompella, V., Capobianco, R., Jong, S., Browne, J., Fox, S., Meyers, L., Wurman, P., & Stone, P. (2020). Reinforcement learning for optimization of COVID-19 mitigation policies. In: *AAAI Fall Symposium on AI for Social Good*.

Kuniya, T. (2020). Prediction of the epidemic peak of coronavirus disease in Japan, 2020. *Journal of Clinical Medicine, 9*(3), 789.

Kurita, J., Sugawara, T., & Ohkusa, Y. (2020). Forecast of the COVID-19 outbreak and effects of self-restraint in going out in Tokyo, Japan. *Medrxiv Preprint*. https://doi.org/10.1101/2020.04.02.20051490

Ministry of Health, Labour and Welfare. (2021). Regarding the handling of discharge and employment restrictions for patients with new coronavirus infection in the Act on Prevention of Infectious Diseases and Medical Care for Patients with Infectious Diseases. https://www.mhlw.go.jp/content/000745527.pdf. (accessed 2021–08–16)

Ministry of Health (Israeli). (2021). Decline in vaccine effectiveness against infection and symptomatic illness. (accessed 2021–08–15)

Munasinghe, L., Asai, Y., & Nishiura, H. (2019). Quantifying heterogeneous contact patterns in Japan: A social contact survey. *Theoretical Biology and Medical Modelling*. https://doi.org/10.1186/s12976-019-0102-8

National Center for Global Health and Medicine. (2020). About the interim report on COVID-19 registry research. Media study session. (accessed 2021–08–15)

National Institute of Infectious Diseases. (2021). IASR Vol. 42, No.6 (No. 496), 131–2. https://www.niid.go.jp/nIID/images/idsc/IASR/42/496.pdf. (accessed 2021–08–15)

Office for Novel Coronavirus Disease Control, Cabinet Secretariat, Government of Japan. (2020). Indicators and guidelines for implementing measures in response to future changes in infection status. (accessed 2021–08–15)

Omae, Y., Toyotani, J., Hara, K., Gon, Y., & Takahashi, H. (2020). Effectiveness of the COVID-19 Contact-Confirming Application (COCOA) Based on a Multi-Agent Simulation. *arXiv preprint*. https://arxiv.org/abs/2008.13166

Statistics Bureau. (2021). *Ministry of Internal Affairs and Communications: e-Stat*. https://www.e-stat.go.jp/. (accessed 2021–08–15)

Tokyo Metropolitan Government. (2020). *Statistics of Tokyo*. (accessed 2021–08–15)

Vyklyuk, Y., Manylich, M., Škoda, M., Radovanović, M. M., & Petrović, M. D. (2021). Modeling and analysis of different scenarios for the spread of COVID-19 by using the modified multi-agent systems—Evidence from the selected countries. *Results in Physics, 20*, 103662.

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning, 8*(3), 279–292.