Chaos, Solitons and Fractals 188 (2024) 115490

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

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

Comprehensive risk management of infectious diseases

Jaewoo Hwang [a,1], Hyojung Lee [b,1], Sunmi Lee [c,1], Bongsoo Jang [d,1], Younghae Do [a,*], Ying-Cheng Lai [e,f]

[a] Department of Mathematics, Nonlinear Dynamics & Mathematical Application Center, Kyungpook National University, 41566, Daegu, Republic of Korea
[b] Department of Statistics, Kyungpook National University, 41566, Daegu, Republic of Korea
[c] Department of Applied Mathematics, Kyung Hee University, 17104, Yongin, Republic of Korea
[d] Department of Mathematical Science, Ulsan National Institute of Science and Technology, 44919, Ulsan, Republic of Korea
[e] School of Electrical, Computer and Energy Engineering, Arizona State University, 85287, Tempe, USA
[f] Department of Physics, Arizona State University, 85287, Tempe, USA

ARTICLE INFO

Dataset link: https://github.com/jaewoohwang/Agent-based-model

Keywords:
Risk management
Population density
Population mobility
Social distancing
Mass gathering events

ABSTRACT

The COVID-19 pandemic raised the need to prepare for the possibility of a new pandemic stemming from an unknown "Disease X." The extent to which an epidemic will spread depends on the complex interplay of various human and environmental factors. Previous studies focused on analyzing the effects of individual parameters on disease transmission. Based on empirical COVID-19 data from South Korea, we develop a comprehensive modeling framework incorporating the population density, inter-city human mobility, the location of the initial outbreak, social distancing, and mass gathering events, with the primary goal to assess the transmission risks at a quantitative level. Systematic computations reveal the emergence of a group structure among all possible spreading scenarios: they are organized into three distinct groups with well-defined boundaries. This group structure underscores the importance of individualized risk assessment strategies for cities based on their unique characteristics, leading to intervention policies tailored to their specific circumstances.

1. Introduction

When a new virus emerges and begins to spread, there is usually limited or no pre-existing immunity to it within the human population. In the absence of an effective vaccine, the virus has the potential to rapidly spread worldwide, leading to a pandemic. The first line of defense is often non-pharmaceutical interventions (NPIs) that serve as the primary public-health measures to mitigate the transmission of the novel virus, which include social distancing, close contact tracing, mask wearing, travel restriction, rapid diagnosis, and isolation of confirmed cases [1,2]. Factors such as the population density, inter-city mobility, the location of the initial outbreak, social distancing (SD), and mass gathering events (MGEs) play a crucial role in the extent and severity of infectious spread [3–5]. It is of paramount interest to assess, quantitatively and systematically, how different factors affect the risk of the disease transmission, taking into the unique population characteristics, and to develop effective intervention strategies.

Diverse factors affecting the spreading of an epidemic can be described using the global COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus (SARS-CoV-2). It started in Dec. 2019 and quickly spread globally [6,7]. In the absence of vaccination, NPIs were implemented to manage the transmission of COVID-19, and mass vaccination against SARS-CoV-2 commenced in America and Europe in Dec. 2020 to reduce high-risk populations [8]. SD was enforced by the Korean government since Feb. 2020 during the early outbreak of COVID-19, and social network analysis was conducted to estimate the effect of mask wearing campaign on the reduction of the size of outbreak in South Korea [9]. MGEs have the potential of spreading the disease rapidly and can lead to super-spreading events (SSEs). The crucial role of MGEs in the global propagation of diseases was reported in several countries [4,10–13], where it was found that MGEs from large public events to small family gatherings could lead to rapid spreading or even SSEs [14,15]. Significant international efforts were made to implement specific measures, risk assessment, surveillance, and event cancellations to prevent the spread of SARS-CoV-2 from the MGEs [16]. Also, the characteristics of COVID-19 have changed due to mutations over the past two years since 2019, highlighting the necessity of clarifying the clinical and epidemiological aspects to evaluate policy responses. In [17], they forecasted COVID-19 cases by

* Corresponding author.
E-mail addresses: jaewoo7932@knu.ac.kr (J. Hwang), hjlee@knu.ac.kr (H. Lee), sunmilee@khu.ac.kr (S. Lee), bsjang@unist.ac.kr (B. Jang), yhdo@knu.ac.kr (Y. Do), ying-cheng.lai@asu.edu (Y.-C. Lai).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.chaos.2024.115490
Received 3 June 2024; Received in revised form 1 September 2024; Accepted 3 September 2024

region using the compartmental model to estimate effective reproduction numbers under varying control-intervention intensities. Moreover, a comprehensive analysis of infection clusters is required, focusing on uncovering and examining key social factors related to age and regional patterns [18]. Other factors that contribute to the disease spreading include population density and mobility [19–21], where the density of inhabitants and the close contacts among people in urban areas are potential hot spots for rapid spread of COVID-19 and other infectious diseases such as influenza and severe acute respiratory syndrome (SARS) [22,23]. Previous studies highlighted the significance of movement between regions [24,25], which can introduce pathogens to new areas and facilitate their spread within and among the populations [3]. Outbreaks caused by MGEs were reported [4,10–13], and there were international efforts in attempting to mitigate the spread of COVID-19 from MGEs. For modeling disease transmission with MGEs, an earlier work considered several key factors [26]. Incidence of COVID-19 associated with MGEs in Spain was analyzed by using inverse probability weighted regression [27]. It is important to assess how factors such as the differences in the population density between two cities and the city of initial outbreak affect the spreading.

NPIs are among the best ways of controlling the spread of new virus when vaccine is not yet available, but prolonged application of NPIs can diminish the effect of the intervention, besides causing negative social and economic effects. While NPIs such as social distancing are effective in controlling outbreaks, they also have economic costs. This creates a challenge for policymakers who must balance disease control with economic impact. Therefore, it is argued that targeted interventions are necessary to balance outbreak control and economic impact mitigation in specific regions [28]. In addition, the previous study [29] analyzed the impact of COVID-19 variants on cost-effectiveness across age groups, considering vaccination efforts and NPIs in Republic of Korea. The transmission dynamics of an infectious disease depends on the strategies of NPIs that need to take into account factors such as the first outbreak place, city characteristics and travel. The effects of NPIs also depend on various conditions including the population density, the first outbreak place and travel among the cities. To make a decision to apply NPIs, an understanding of the risk of spreading under different circumstances is essential. For example, is it necessary to apply high SD level to a very low-density city?

The vast complexity of the COVID-19 pandemic made it difficult to determine effective preventive measures, especially in terms of the impact of physical distancing and vaccines on reducing the spread of the virus. The control strategies for COVID-19 include NPIs and vaccination programs. Different NPIs were employed to reduce COVID-19 cases, including social distancing, wearing of masks, temporary shutdowns in schools, reductions in social activities, and employment-related restrictions [7,30]. NPIs played an important role in controlling COVID-19 infections, especially when effective vaccines are unavailable. Mass vaccination against COVID-19 began in Europe at the end of 2021. Even after vaccination became widely available, many governments in Europe continued to impose limitations on social activities in tandem with their vaccination programs [3,23,31]. For informing the governmental policies and quantifying the pandemic dynamics, mathematical modeling is essential [22,32,33]. For example, agent-based modeling (ABM) can be implemented to simulate the transmission of COVID-19 with human behaviors taken into account, enabling the effects of various public health intervention method to be evaluated [34]. Agent-based modeling incorporating human mobility from mobile data was also exploited to assess the impact of NPIs in the Boston metropolitan area [35,36]. Furthermore, an ABM was employed to capture the early transmission dynamics of COVID-19. Without vaccines or treatments, contact tracing and case isolation are the most crucial interventions to prevent large outbreaks, underlining their significance in mitigating SSEs [37].

Previous studies did not take into account the combined impact of the crucial factors such as MGEs, intensity of NPIs, movement behaviour to assess the risk of outbreak of the infectious disease. The aim of this paper is to address the risk induced by multiple spreading factors. To achieve this goal, we introduce a measure to quantify the epidemic risk of spreading infectious disease for various scenarios and carry out a comprehensive comparison study of the risks of various scenarios. To obtain an appropriate setting for different scenarios, we assume that the mobility rate of the individuals is fixed and focus on the following four factors: (1) level of SD, (2) the sizes of MGEs, (3) the population density of the first outbreak city, and (4) the population density of the city where MGEs took place, where the density is categorized by two classes: high or low. The risk can then be calculated by the probabilities of various ways of transmission. For example, the probability of local transmission can be calculated from the number of infected people in the first outbreak city exceeds a threshold. Our specific approach is agent-based modeling where agent movements occur between two cities with high and low densities, respectively, based on empirical data of COVID-19 outbreaks from two major cities of South Korea [38,39]: Seoul with high-density and Daegu with low-density from Feb. 2020 to Oct. 2021. During this time period, two major outbreaks associated with MGEs occurred in each city, followed by an abrupt increase in the COVID-19 cases. In response, the South Korean government implemented different SD policies depending on the significance level of COVID-19. After an MGE, the SD level immediately enhanced to a higher level and super spreaders who have a high number of secondary infections were identified in both outbreaks. In our agent-based model, the SD levels and the sizes of MGEs are categorized into four different classes. Given a specific scenario, our modeling framework captures the main features of COVID-19 spreading in South Korea and allows the prioritizing factor affecting the spreading dynamics to be identified. For example, simulations confirm that an initial outbreak in a high-density city can lead to high risks, suggesting social distancing as an effective mitigation strategy in this case. However, when the initial outbreak occurs in a low population density region, a number of factors including human mobility, MGEs, and SD all can be important. Depending on other factors, MGEs in high-density areas can lead to significant spreading.

Our systematic calculation and analysis of the epidemic risks for all possible spreading scenarios reveal the emergence of a group structure among all possible spreading scenarios: they are organized into three distinct groups with well-defined boundaries. The significance of this finding is that it underscores the importance of individualized risk assessment strategies for cities based on their unique characteristics, leading to tailored intervention policies adeptly suited to their specific circumstances. To our knowledge, in the existing literature on infectious disease spreading and interventions, no such a group structure associated with the risks of various epidemic scenarios at a quantitative level was reported.

2. Methods

2.1. Construction of mathematical model and empirical data

Agent-based model with population movements. In South Korea, there were two instances of initial COVID-19 outbreaks attributed to MGEs: the one occurred in Daegu in Feb. 2020 followed by a second outbreak in Seoul in Aug. 2020. Both Daegu and Seoul are major cities in the country with massive daily population movements between them. The urban population density of Seoul is nearly four times higher than that of Daegu. Fig. 1 displays empirical COVID-19 outbreak data gathered from the two cities. Accordingly, our model has two cities, where the population density of one city is four times higher than that of the other, and there can be significant human movements between the two cities.

In the agent-based model, each city represents a region of size L in the two-dimensional physical space, one with N_1 and another with N_2 agents. The population movements between the two cities are

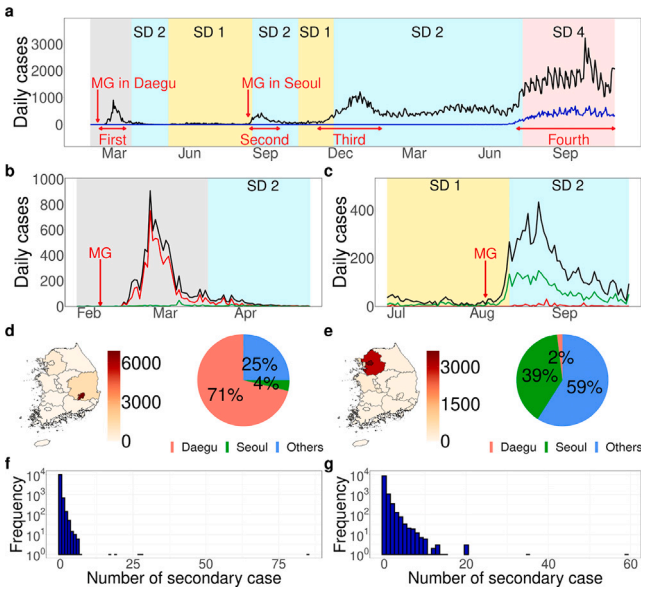

Fig. 1. Description of empirical COVID-19 data in South Korea. **a** The time series of the daily COVID-19 confirmed cases (black) and delta variant cases (blue), in South Korea from Feb. 1, 2020 to Oct. 31, 2021. The shaded colors represent the levels of SD. **b, c** The first and second COVID-19 outbreaks caused by MGEs in different periods, from Feb. 1 to Apr. 30, 2020 and from Jul. 1 to Sep. 30, 2020, respectively. The black, red, and green traces represent the total number of cases in South Korea and the number of COVID-19 cases in Daegu and Seoul cities, respectively. **d, e** Geographical distributions of the total cases corresponding to **b, c** respectively. **f, g** Distributions of secondary cases due to the MGEs, corresponding to **b, c** respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

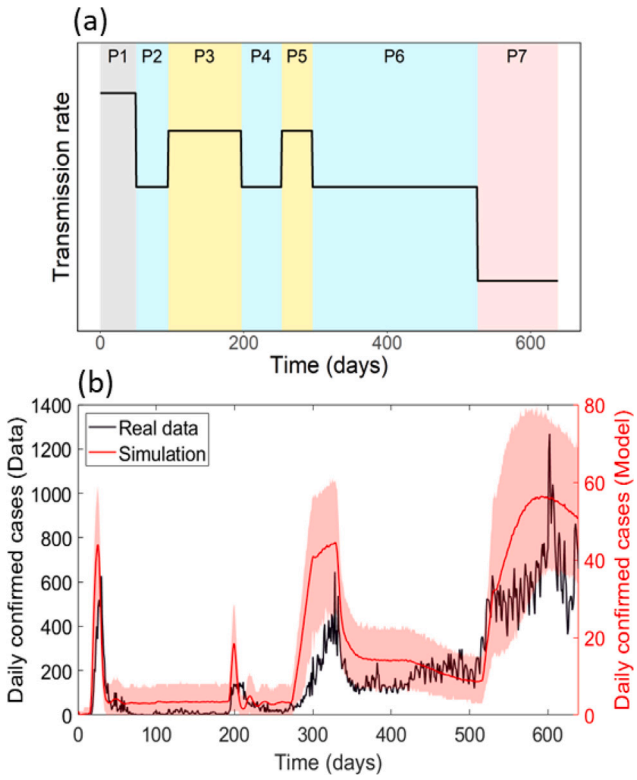

Fig. 2. Schematic illustration of the disease transmission dynamics in the agent-based model. **(a)** A time-varying transmission rate function, where the black solid line represents the transmission rate according to the levels of SD, and the shaded regions mark the levels of SD (see Supplementary Tab. I for more details). **(b)** A comparison between the real data and the simulation results, where the black trace is a time series of the daily confirmed COVID-19 cases, which is the sum of all confirmed cases in Daegu and Seoul from Feb. 1, 2020 to Oct. 31, 2021. The red curve is the average confirmed cases estimated from 1000 independent simulations of the agent-based model, and the error bars correspond to the 95% confidence interval. Note that, due to the different population scales, the left and right vertical axes are for empirical data and simulation results, respectively.

characterized by the mobility rate R_p - the fraction of individuals in movement. Due to the difference in the population densities of the two cities, the ratio between the number of travelers from one city to another and that in the opposite direction is given by the coefficient N_1/N_2. Specifically, at each simulation time step Δt, randomly chosen $R_p N_1$ agents in the low-density city move to the high-density city, and $(N_1/N_2) \times R_p N_2$ agents move in the opposite direction. In each city, the agents uncorrelated random walks: $(x_{t+1}, y_{t+1}) = (x_t, y_t) + (r\cos\theta, r\sin\theta)$, where (x_t, y_t) represents the position of an agent at time t, and the radius r and angle θ follow a multivariate normal and a uniform distribution, respectively. Individuals are susceptible to either the original COVID-19 strain or the delta variant. The transmission process is described by the compartmental SEIR model [40], where an agent at time t can be in one of seven states: susceptible (S), exposed (E), infectious (I), removed (R), vaccinated (V), exposed by variants (E_D), and infectious by variants (I_D). The state evolves according to two sets of rules, described as follows.

Rule #1: A susceptible agent S can be exposed to the disease from an infectious agent I at the transmission rate β. An exposed agent E is a newly infected agent without transmissibility, but will be in the infectious state I after a latent period. An agent I can transmit the disease to other agents, and then become a removed state R after an infectious period. Both the latent and infectious time periods follow the gamma distribution with a specific set of parameters [41,42]. Removed agents R are those who have recovered from the infectious disease or died. Vaccinated agents V have a higher protection rate ρ from the disease than that for the susceptible individuals. When a susceptible agent S contacts a virus carrier in the state I, *i.e.*, when the distance between S and I is less than the time-varying infection radius $\epsilon(t)$, the susceptible agent is exposed to a time-varying transmission rate $\beta(t)$. Especially, if a susceptible agent S contacts n number of carriers at time t, the transmission rate is given by [43] $1-(1-\beta(t))^n$.

Rule #2: After mid-2021, the delta variant, SARS-CoV-2 B.1.617.2, became the most dominant type of variants in Korea. To take into account different transmission types of COVID-19 variants, we introduce two states: delta-exposed and delta-infectious states, denoted as E_D and I_D, respectively, after t_D, the time that the first delta variant emerged. In the model simulation, an exposed agent at $t = t_D$ is randomly chosen and becomes the first carrier of the delta variant after a latent period. Note that there are no differences between preexisting virus carriers and delta variant carriers except that the latter have an increased transmission rate of $r_D\beta$, where r_D denotes the increased transmission rate caused by the delta variant [44,45]. If an susceptible agents simultaneously contacts n_1 number of preexisting virus carriers and n_2 number of delta variant carriers, the conditional probability P that S gets exposed to delta variant will be

$$P = \frac{n_2 r_D \beta}{n_1 \beta + n_2 r_D \beta}.$$

2.2. Mass gathering events, social distancing & vaccination

Mass gathering events. An MGE, a high concentration of individuals at a specific location, can be modeled based on the concept of a *gathering zone*. MGEs take place in this zone, which attract agents including at least one infectious individual. In our simulation, the gathering zone is designated as a square area with periodic boundary conditions. When MGEs occur, agents are randomly chosen to relocate to the gathering zone, where they are uniformly distributed. In the

zone, infectious agents have the opportunity to transmit the virus to susceptible individuals. Once an MGE finishes, the agents return to their original locations within the city. To account for the heightened transmission risks and safety concerns arising from crowd density during an MGE, two distinct time scales are needed: (1) a macro-time scale (Δt), typically corresponding to a day, which governs the agent movement within each city, and (2) a micro-time scale (δt) that dictates the agents' activities within the gathering zone. In our simulation, 50 micro-times is equal to one macro time, i.e., $\Delta t = 50\delta t$. The parameters and configurations associated with MGEs are outlined in Supplementary Tabs. II & III. MGEs are categorized into four risk levels: $0-3$, with the classification based on the size of the population participating in the MGEs.

Social distancing. Since the initial COVID-19 outbreak in Feb. 2020, the Korean government has implemented NPIs including SD measures and the mandatory mask wearing, as shown in Fig. 2**(a)**. SD means maintaining a greater physical distance from others than usual to minimize potential exposure and reduce transmission. We categorize SD into four levels based on guidelines of the Korean government: "Without = S_0" (SD level less than 1), "Low = S_1" (SD level between 1 and 1.5), "Intermediate = S_2" (SD level between 2 and 2.5), and "High = S_3" (SD level between 3 and 4). We assume that the disease transmission rate $\beta(t)$ depends on the SD level. Specifically, for without, low, intermediate, and high levels, the rate $\beta(t)$ is β_0, $0.9\beta_0$, $0.75\beta_0$, $0.5\beta_0$, respectively.

To account for the reduced effect of the intervention due to the prolonged SD, we introduce the parameter of fatigue rate. For each city, we define the infection radius function as $\epsilon_i(t) = \epsilon_0\lambda_i(t)^{-(1-\tau(t))}$, where ϵ_0 is the baseline of the radius, $\lambda_i(t)$ is the weekly moving average of daily confirmed cases in city i from $t-7$ to t, and $\tau(t) \in [0, 1]$ is a non-negative step function modeling public fatigue caused by the prolonged interventions [38,46]. To avoid the explosion of the infection radius, we set $\epsilon_i(t) = \epsilon_0$ for $\lambda_i(t) < 1$.

Vaccination. Vaccination against COVID-19 commenced in Korea on Feb. 26, 2021, stipulated by the Korea Disease Control and Prevention Agency (KDCA) [47]. To integrate the impact of vaccination into our agent-based model, we introduce four parameters: t_V, ϕ, κ, ρ, whose values are listed in Supplementary Table 2. In particular, t_V signifies the date when the second vaccine dose was administered, which, in this case, is set to Mar. 6, 2021. We assume that it takes a two-week period for individuals to develop immunity from the vaccination when its coverage (κ) is limited to 95%. Susceptible individuals are chosen at random to receive the vaccine at the rate ϕ. Additionally, we account for a reduction in the transmission rate due to the protective effect of vaccination: $(1-\rho)\beta$.

Parameter estimation. We conducted simulations for two MGEs spanning from Feb. 1, 2020, to Oct. 31, 2021. The initial MGE took place in Daegu in Feb. 2020, followed by a second event in Seoul in Aug. 2020 (details in the Supplementary Tab. IV). To estimate the pertinent parameters, we performed multiple Monte-Carlo simulations to determine the most suitable parameter configuration. Fig. 2**(b)** shows that the simulation results agree well with the real-world data, justifying the choice of the model parameters pertinent to MGEs and SD.

Process of simulation. To obtain data, we simulate it at least 10,000 times for each scenario (S_i, M_j^k, O_m). Our basic assumption for each simulation is that (1) when exposed or infected agents do not exist, our simulation is terminated, because our simulation reaches an equilibrium, and (2) the maximal simulation time ($= 1000\Delta t$) is fixed. For extreme case, i.e., with higher MGE and population density, most of population ($\geq$70%) is infected over the maximal simulation time, which means that we may consider all population is infected in the epidemic sense. But, except extreme scenario, we confirmed that our simulation was terminated before the maximal simulation time.

Data. The empirical data on which our study was based was from the COVID-19 data base by the Korea Disease Control and Prevention Agency (KDCA), which includes the daily number of confirmed cases, regional infection rates, routes of transmission, and variant types in South Korea [47].

2.3. Quantification of spreading: Epidemic risk and effective reproduction number

Probabilities of no, local, influential and wide transmissions. To examine the impact of infectious transmission between the two cities, the concepts of "local" and "influential" transmissions are important. The two transmission types can be distinguished by setting a threshold in terms of the number of infected individuals, which we set as 2% of the total population for each city. That is, when there is an outbreak in city A, we classify it as a "local transmission" if the count of infected individuals in the city surpasses the threshold. Likewise, if the number of infected individuals in city B exceeds the threshold, we categorize it as an "influential transmission". These classifications form the basis for defining the probabilities associated with both local and influential transmissions:

$$P_L = \frac{\text{Number of local transmission occurrences}}{\text{Total number of simulations}},$$

$$P_I = \frac{\text{Number of influential transmissions occurrences}}{\text{Total number of simulations}}.$$

Furthermore, it is possible to define a probability P_W to represent the concurrent occurrence of both local and influential transmissions, referred to as *wide transmission*. Similarly, we define the probability of *no transmission*, denoted as P_N, which pertains to the situation where the cumulative number of infected individuals in each city does not surpass the specified threshold.

Epidemic risk. For a complex epidemic environment, we define the *risk for spreading of infectious disease* R_{isk} as

$$R_{isk} = P_L\langle R_L\rangle + P_I\langle R_I\rangle - P_W\langle R_W\rangle, \tag{1}$$

where $\langle R_j\rangle$ $(j \in L, I, W)$ are the average ratios of the total number of infected individuals in the two cities relative to the combined population of the two cities for different modes of transmission. The component associated with "wide transmission" is taken away because the "local" and "influential" transmissions contain elements of "wide transmission". When both "local" and "influential"transmissions occur, the risk of propagation can vary based on the number of infected individuals involved. The measure of infectious spreading risk in Eq. (1) serves to gauge the potential severity of the spreading.

Effective reproduction number. The basic reproduction number R_0, the expected number of secondary cases resulting from a single infected individual in a population of susceptible individuals, is a critical quantity characterizing the potential spread of an emerging infectious disease. For $R_0 < 1$, the disease is not expected to propagate significantly. Conversely, for $R_0 > 1$, chances are that the disease will spread within the population [48,49]. A more practical metric is the effective reproduction number R_t, which represents the expected number of secondary cases arising from a primary case infected at a specific time t. The impact of interventions on disease transmission can be assessed by R_t, which has been used widely in public health policies and strategies. Computationally, R_t can be calculated as the number of new infections at time t, denoted as $I(t)$, divided by the number of infectious individuals $I(t-1)$ at the preceding time step $t-1$. Numerical fluctuations can arise due to the small denominator and the consideration of the infectious period for exposed individuals. As a remedy, we define the effective reproduction number R_t as

$$R_t = \begin{cases} r_0 \times \frac{I(t)}{I(t-1)}, & \text{if } I(t-1) > I_c, \\ 0, & \text{if } I(t-1) \leq I_c, \end{cases} \tag{2}$$

where r_0 and I_c are the mean of the infectious period γ and the critical number of the infected individuals, respectively. We set $r_0 = 7$ and $I_c = 1$.

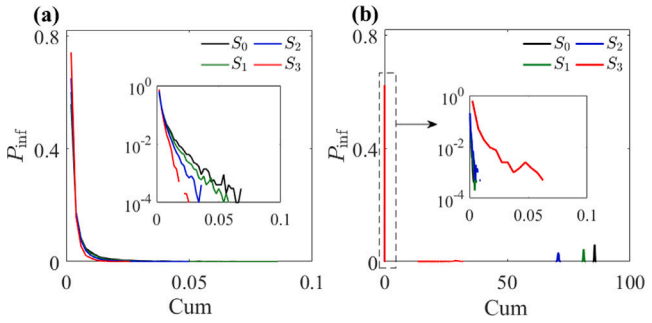

Fig. 3. Effects of SD. **(a,b)** Probability distribution $P_{inf}(x)$ of the cumulative cases (Cum) of infection for different SD levels at the low- and high-density density city, respectively. The distribution in **(a)** can be fitted by an exponential function while that for the high-density city is approximately Gaussian. The dashed box in **(b)** is zoomed in for different scales.

3. Results

3.1. Effects of SD and MGEs on epidemic spreading in absence of population movement

Probability of cumulative cases, $P_{inf}(x)$. We classify four different SD level: S_0 = no SD, S_1 = low level of SD, S_2 = intermediate SD level and S_3 = high SD level. For each SD level, we calculate $P_{inf}(x)$, the probability distribution of the cumulative infected population in both the low- and high-density cities, which can be seen in Fig. 3. The distribution is approximately exponential-type for the low-density city and can be fitted by a Gaussian-type curve of mean μ_m and standard deviation s_m for the high-density city:

$$P_{inf}(x) \sim \begin{cases} \exp(-\lambda_S x), & \text{for small } x, \\ \exp[-(x-\mu_m)^2/2s_m^2], & \text{for large } x, \end{cases} \quad (3)$$

as shown in Fig. 3**(a)** and **(b)**, respectively. For the low-density city, SD has little effect on the spreading and a higher level of SD can reduce the probability of cumulative cases (Cum) but only slightly. For the high-density city, the probability of a large number of infected people decreases from 80% to 76%, 67%, 28% by increasing SD-level, $S_0 \to S_3$, i.e., the number of infections can be reduced by almost three times by increasing the SD level from low to high. For a city with high population density, SD as an NPI can then be effective in preventing the disease spreading.

Note that to determine a Gaussian-type distribution for large x shown as in Fig. 3**(b)**, we may use Kolmogorov–Smirnov test [50,51] and whose p-values for S_0, S_1 and S_2 are 0.57, 0.96 and 0.06. It means that $P_{inf}(x)$ for S_0 and S_1 satisfy a Gaussian distribution, but for S_2, not Gaussian. That is, due to stronger SD, Gaussian distribution is destroyed. It means that SD as an NPI can be effective in preventing the disease spreading. The significance of having two distribution types for different scales is that higher population density can trigger new, large-scale outbreaks. That is, population density is one of key factors in epidemic spreading.

MGE-effect in an isolated city. The size of MGEs can be categorized into four groups in terms of the number of attendees: 0%, 0.5%, 1% and 1.5% of the total population of the city (denoted as M_0, M_1, M_2, M_3). Fig. 4**(a)** and **(b)** show $P_{inf}(x)$ for the four MGE sizes at the low- and high-density city, respectively, which can be approximately fitted by Eq. (3). For the low-density city, as the MGE size increases from M_0 to M_3, the probability of an outbreak increases from 0% to 29% and the number of infected individuals increases from 1% to 8%, indicating that an MGE can lead to a large-scale outbreak. For the high-density city, for the same increase in the MGE level, the number of infected individuals changes from 85% to 88%, indicating that MGEs have little effect on the spreading once an outbreak has already occurred. Fig. 4**(c)** and **(d)** show a representative time series of the daily confirmed cases and the effective reproduction number R_t for the fixed SD level S_0 and MGE size M_1, for the low- and high-density city, respectively. When MGE is occurred, the value of R_t increases dramatically and becomes greater than 1, suggesting MGEs as a likely trigger for a massive increase in the new infections. Fig. 4**(e)** and **(f)** show the mean and variance of P_{inf} for different SD levels and MGE sizes. For a fixed MGE size, the mean and variance decreases with the SD level, implying that SD is effective at reducing spreading even if MGEs occur, especially for the high-density city [Fig. 4**(f)**]. For the low-density city, a large MGE size can cause serious transmission. However, for the high-density city, MGEs do not have a significant affect on the spreading, because of the high population density as a strong contributing factor to disease transmission. Fig. 4**(g) (h)** show the probability of the local transmission, denoted as P_L, for different SD levels and MGE sizes. For the low-density city, P_L increases with the MGE size and decrease with the SD level. For the high-density city, the MGE size has little effect on P_L, except for a high SD level. Overall, high population density represents a potential risk factor for infectious disease transmission.

3.2. Epidemic risks with population movements

The results presented so far are for the setting where there are no population movements between the low- and high-density cities. We now address the effects of population mobility on epidemic spreading. To be concrete, we assume that the population of the high-density city (**H**) is four times that of the low-density city (**L**). In addition to studying the effects of NPIs and mobility rate on the transmission dynamics, another key factor is the city in which the infectious disease emerges for the first time. Because of the large number of possible scenarios, it is necessary to have a labeling scheme. We introduce the following notation to label the various scenarios:

$$(S_i, M_j^k, O_m),$$

where S_i is for SD level ($i = 0, 1, 2, 3$), M_j^k represents the size j of MGE occurred in the k-density city ($j = 0, 1, 2, 3$ and $k = L, H$), O_m indicates the city of the first outbreak ($m = L, H$), and M_0 indicates that MGE does not occur. Altogether, there are 56 distinct scenarios.

Effects of SD. In the presence of population movements, the size and speed of epidemic spreading will depend on the population density of the two cities and the specific city in which the first outbreak occurs. To be concrete, we assume that there are no MGEs and fix the population mobility rate at $R_p = 0.0004$.

Suppose that the first outbreak of an infectious disease begins in the low-density city and population movement into the high-density city is allowed. This scenario is labeled as (S_i, M_0, O_L). Fig. 5**(a)** shows P_{inf}, which is the probability distribution of the cumulative infected population in both the low- and high-density cities. For a small number of cumulative cases, P_{inf} is approximately exponential with the total probability about 99%. The probability for a large number of cumulative cases is less than 1%: for the four SD levels (S_0 to S_3), the probability of large-scale infection is about 0.64%, 0.49%, 0.38% and 0.16%, respectively. These behaviors suggest that, if the initial outbreak occurs in the low-density city, a large scale outbreak is not likely. Quantitatively, the values of risk measure (see **Methods**) for the four SD levels are close to zero:

$$R_{isk}(S_0, M_0, O_L) \approx 0.0047, \qquad R_{isk}(S_1, M_0, O_L) \approx 0.0035,$$
$$R_{isk}(S_2, M_0, O_L) \approx 0.0018, \qquad R_{isk}(S_3, M_0, O_L) \approx 0.0001,$$

indicating that, for the scenario (S_i, M_0, O_L), SD is ineffective in spite of the population movements.

What control strategy is effective for this scenario? Fig. 5**(b)** shows the time series of the daily confirmed cases and the effective reproduction number for each city. Due to the population movement from the

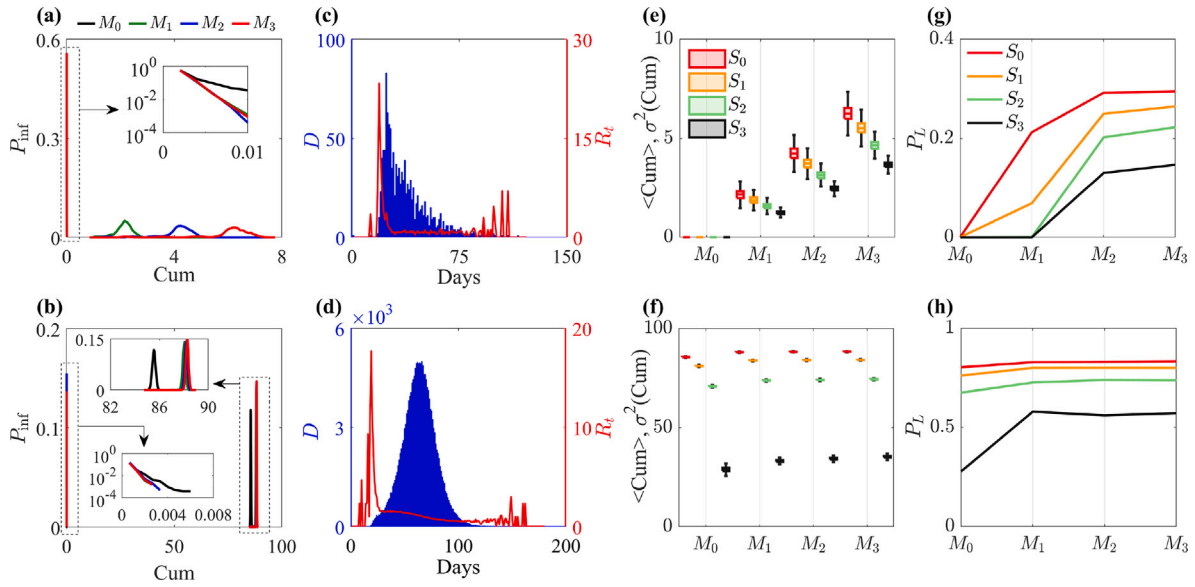

Fig. 4. Effects of MGEs. The first and second rows are for the low- and high-density cities, respectively. **(a,b)** For fixed SD level S_0, the probability distribution of cumulative cases for different MGE sizes. **(c,d)** For fixed MGE size (= M_1) and SD level S_0, a representative time series of daily confirmed cases D and the effective reproduction number R_t. **(e,f)** Mean $\langle \text{Cum} \rangle$ and variance $\sigma^2(\text{Cum})$ of cumulative cases for different MGE sizes and SD levels. **(g,h)** The probability of local transmission P_L. The dashed box in **(a,b)** is zoomed in for different scales.

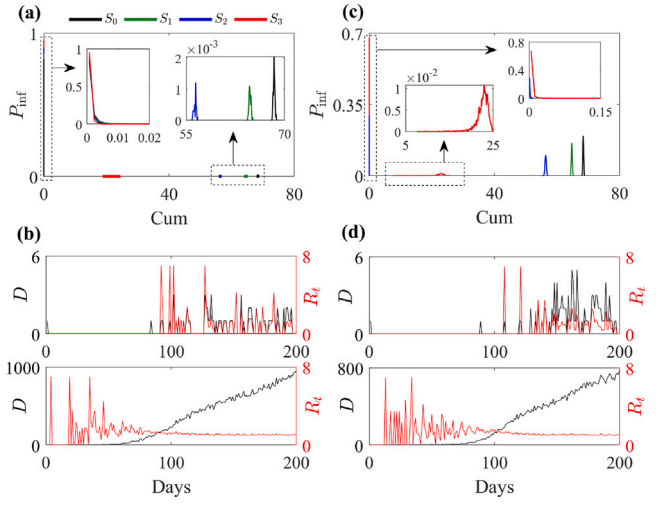

Fig. 5. SD effects with population movement. The population mobility rate is set to be $R_p = 0.0004$. **(a)** For an initial outbreak in the low-density city, the probability distribution of cumulative infection, $P_{\text{inf}}(x)$, for different SD levels. **(b)** For the fixed SD level S_0, representative time series of daily confirmed cases D (black) and its effective reproduction number R_t (red) for both the low-density (top) and high-density (bottom) cities. **(c)** For an initial outbreak in the high-density city, the distribution $P_{\text{inf}}(x)$ for different SD levels. **(d)** For the fixed SD level S_0, representative time series as in **(b)**. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

low- to high-density city, the effective reproduction number is greater than one in the long run. There can be rapidly spreading events in both cities, albeit with a low probability. For this scenario, travel restriction can be effective at preventing large-scale outbreaks.

We now study the scenarios (S_i, M_0, O_H), where an initial outbreak occurs in the high-density city and population movement into the low-density city is allowed. Fig. 5**(c)** shows P_{inf} in both cities for different SD levels. The distribution is approximately Gaussian. The probabilities of having a large number of cumulative cases (a large-scale outbreak) for the four SD levels are high: 81%, 77%, 66% and 28%, respectively, with the corresponding risk measures:

$$R_{isk}(S_0, M_0, O_H) \approx 0.5532,\ R_{isk}(S_1, M_0, O_H) \approx 0.4942,$$
$$R_{isk}(S_2, M_0, O_H) \approx 0.3738,\ R_{isk}(S_3, M_0, O_H) \approx 0.0642.$$

These results indicate that SD as an NPI can be effective because a high level of SD can lead to a three-fold reduction in the infection and the spreading risk can be reduced by at least eight times. Fig. 5**(d)** shows time series of daily confirmed cases and its effective reproduction number in both cities. It takes a long time for the reproduction number to exceed one due to the high population density, after which the disease spreads to the low-density city due to human mobility. Overall, if the initial outbreak occurs in the high-density city, the likelihood of a large-scale epidemic spreading will be high. Compared with the case where the initial outbreak occurs in the low-density city, a large-scale outbreak is 100 times more likely: $R_{isk}(S_0, M_0, O_H)/R_{isk}(S_0, M_0, O_L) > 100$.

Effects of MGEs. We fix the population mobility rate at $R_p = 0.0004$ and assume SD level S_1. Consider the scenario where the first outbreak and MGEs occur in the low-density city, denoted as (S_1, M_j^L, O_L). Fig. 6**(a)** shows the distribution P_{inf} for different sizes of MGEs. As the size of the MGE increases ($M_0 \to M_3$), the probability for more than 60% of the total population to be infected increases as 0.5% → 24% → 26% → 27%. Fig. 6**(b)** shows the probabilities of the local, influential and wide transmission, where $P_L \approx P_I \approx P_W$. The values of the risk measure R_{isk} can be obtained from Fig. 6**(c)** as

$$R_{isk}(S_1, M_0, O_L) \approx 0.0032,\ R_{isk}(S_1, M_1^L, O_L) \approx 0.1574,$$
$$R_{isk}(S_1, M_2^L, O_L) \approx 0.1732,\ R_{isk}(S_1, M_3^L, O_L) \approx 0.1796.$$

In this case, MGEs have a significant effect on the probability of occurrence of large-scale transmission. In fact, the risk associated with the scenario (S_1, M_3^L, O_L) is less than half of that of the scenario (S_2, M_0, O_H): $R_{isk}(S_2, M_0, O_H) \approx 0.3738$ and $R_{isk}(S_1, M_3^L, O_L) \approx 0.1796$.

We now consider the scenario where the first outbreak occurs in the high-density city: (S_1, M_j^H, O_H). Fig. 6**(d)** shows the P_{inf}, which is similar to the distribution in Fig. 6**(a)**. As the MGE size increases from zero to three, the probability that more than 60% of the total population is infected is approximately constant: 76%, 78%, 79%, 79%, indicating that, in this case, MGEs have little effect on enhancing the disease transmission. Fig. 6**(e)** shows the transmission probabilities P_L,

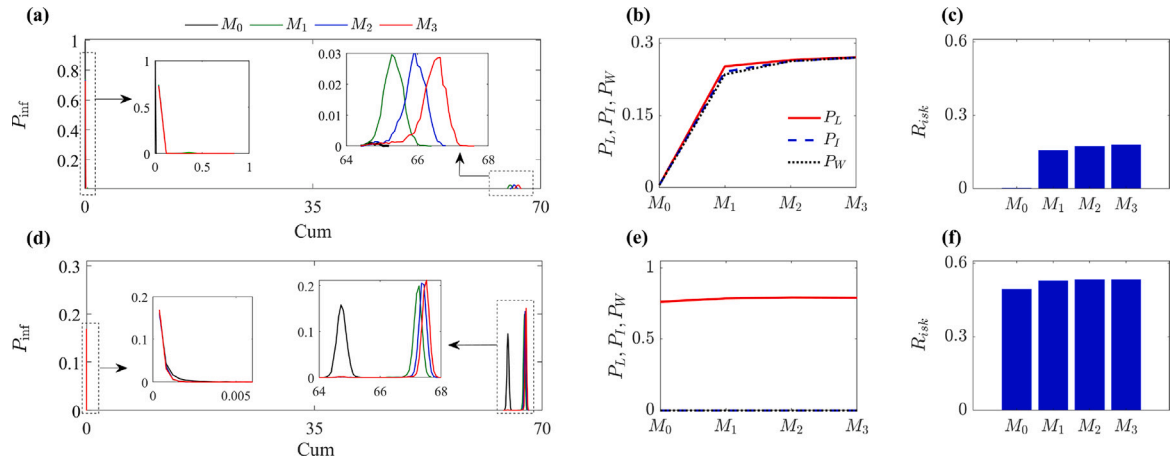

Fig. 6. Effects of MGEs with population movements. For the scenario (S_1, M_j^L, O_L), **(a)–(c)** Distribution $P_{\inf}$, the probabilities of P_L, P_I and P_W, and the risk measure R_{isk} for different sizes of MGEs, respectively. **(d)–(f)** Same as **(a)–(c)** but for the scenario (S_1, M_j^H, O_H).

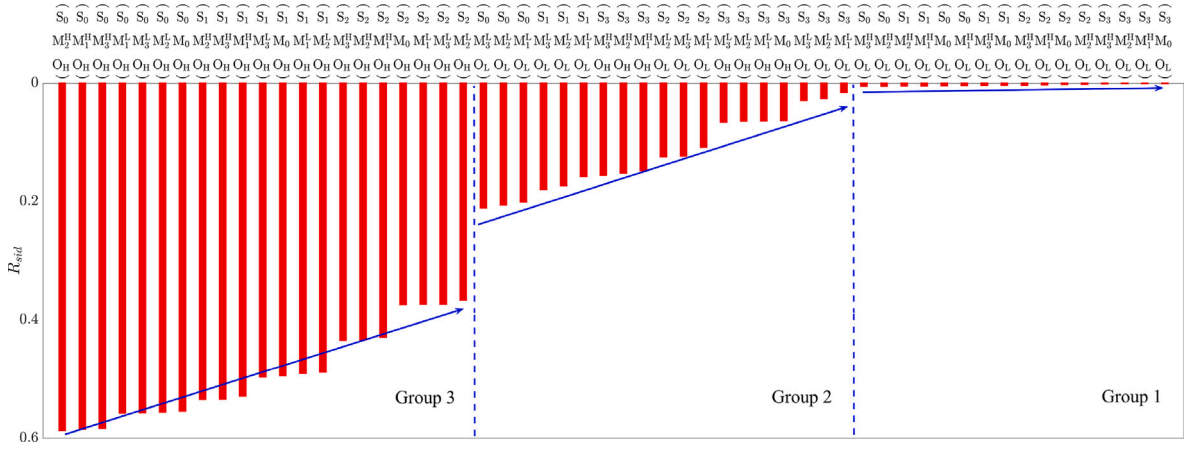

Fig. 7. Group structure of risks for all spreading scenarios. Shown is a systematic display of the values of the rise measure R_{isk} for all 56 spreading scenarios for mobility rate $R_p = 0.0004$, which can be divided into three groups (separated by the two blue vertical dashed lines): left, center and right. The blue arrows indicate the decreasing trend of the risk for different scenarios. The group structure is robust, which holds for other values of R_p (see Supplementary Materials). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

P_I, and P_W, where the local transmission probability is relatively high ($\geq$0.76), but the other two probabilities are near zero, indicating that the disease spreads mostly locally due to the high population density. The risk values R_{isk} for (S_1, M_j^H, O_H) can be obtained from Fig. 6(**f**):

$$R_{isk}(S_1, M_0, O_H) \approx 0.4941, \quad R_{isk}(S_1, M_1^H, O_H) \approx 0.5288,$$
$$R_{isk}(S_1, M_2^H, O_H) \approx 0.5343, \quad R_{isk}(S_1, M_3^H, O_H) \approx 0.5338,$$

indicating a high risk, even though the effects of MGEs on spreading is quite insignificant.

3.3. Spreading risk management

From the perspective of risk management, the results presented so far indicate that SD is not the only tool to prevent spreading and MGEs are not always the leading cause of transmission. To obtain a comprehensive picture of the risks associated with different spreading scenarios, we calculate the values of the risk measure R_{isk} for all 56 scenarios for a fixed population mobility rate (e.g., $R_p = 0.0004$), as shown in Fig. 7. Depending on the values of R_{isk}, the scenarios can be distinguished into three groups, which are separated by $R_{isk}^1 \approx 0.015$ and $R_{isk}^2 \approx 0.366$, as indicated by the two vertical dashed lines in Fig. 7. More specifically, Groups 1, 2, and 3 are associated with low, medium, and high risks, respectively, where the highest risk values for Groups 1–3 are approximately near zero, 0.2, and 0.6, respectively. A common feature of the scenarios in Group 3 is that the population density of the first outbreak location is high, and MGEs generally increase the risk. Fig. 7 provides guidance for managing the spreading risk. For example, by increasing the SD level as an NPI, a scenario in Group 3 can become one in Group 2:

$$R_{isk}(S_1, M_3^H, O_H) \in \text{Group 3} \rightarrow R_{isk}(S_3, M_3^H, O_H) \in \text{Group 2}.$$

Even for the scenarios belonging to the same group, imposing a higher SD level can reduce the risk, e.g., $R_{isk}(S_1, M_1^L, O_L) \approx 0.1574 \in$ Group 2 and $R_{isk}(S_3, M_3^L, O_L) \approx 0.0294 \in$ Group 2.

To better understand the roles of SD and MGEs in spreading for different scenarios, we consider four city combinations: (M_j^H, O_H), (M_j^L, O_H), (M_j^H, O_L) and (M_j^L, O_L), where the first indicates the city in which MGEs occur with size j and the second is the city in which the first outbreak occurs, and calculate the risks for each combination. The results for the four combinations are presented in Fig. 8(**a–d**), respectively.

For (M_j^H, O_H), the range of the risks is approximately $(0.148, 0.587)$, as shown in Fig. 8(**a**). Except for the scenario with the high SD level, (S_3, M_j^H, O_H) all others (S_i, M_j^H, O_H) belong to Group 3. The implication is that controlling MGEs cannot be effective for the combination (M_j^H, O_H) and only a high SD level can move the scenarios into Group 2. The reason is quite intuitive: living in a high-density city is similar

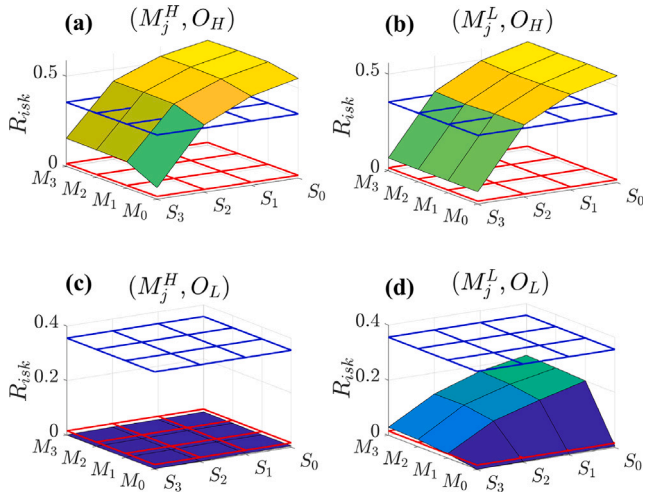

Fig. 8. Illustration of the roles of SD and MGEs for different scenarios. The risks associated with four city combinations are exemplified for different SD levels and MGE sizes: (M_j^H, O_H), (M_j^L, O_H), (M_j^H, O_L) and (M_j^L, O_L), where the first notation indicates the city in which the MGEs occur with size j and the second notation is the city in which the first outbreak occurs.

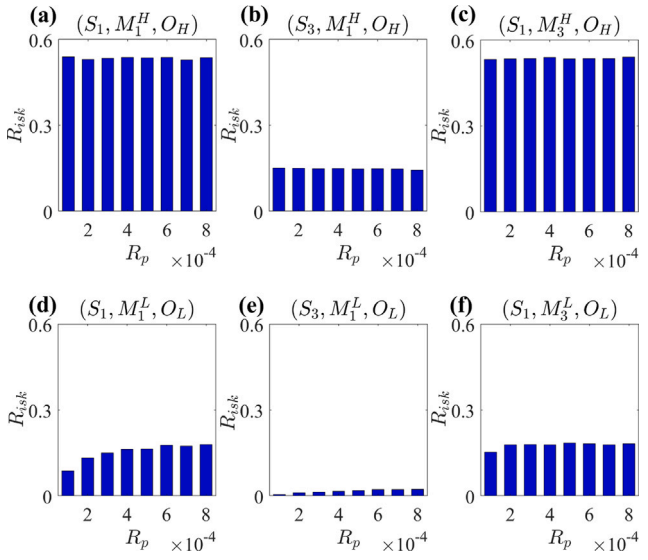

Fig. 9. Effects of human mobility. **(a)–(c)** The risk R_{isk} for the scenarios associated with (M_j^H, O_H) versus the mobility rate R_p for different SD levels and MGE sizes. **(d)–(f)** The corresponding results for the scenarios associated with (M_j^L, O_L).

to experiencing MGEs, so a high SD level is necessary to suppress the spreading.

For (M_j^L, O_H), the range of the risks is approximately $(0.064, 0.558)$, as shown in Fig. 8(**b**). Similar to (M_j^L, O_H), the scenarios (S_i, M_j^L, O_H) belong to Group 3 except for the scenario (S_3, M_j^L, O_H) with a high SD level. The first outbreak in the high-density city is still the leading cause of spreading, and MGEs in the low-density city have little effect on the spreading.

For (M_j^H, O_L): the range of risks is approximately $[0.0001, 0.0005]$, as shown in Fig. 8(**c**). By these scenarios, MGEs have occurred in the high-density city before producing possible super spreaders due to the first outbreak in the low-density city. The scenarios belong to Group 1, and NPIs have little effect on the spreading dynamics. Note that, if some super spreaders produced in the low-density city visit the high-density city, then the scenarios associated with (M_j^H, O_L) will become those with (M_j^H, O_H), rendering effective travel restriction.

For (M_j^L, O_L): the range of risks is approximately $[0.016, 0.211]$, as shown in Fig. 8(**d**). The scenarios without MGEs, *i.e.*, (S_i, M_0, O_L), belong to Group 1, and all others belong to Group 2. Depending on the implementation of NPIs, the risk difference between the spreading scenarios is almost 13 times. For these scenarios, restricting MGEs can be effective.

3.4. Effects of population mobility

The main result – the emergence of three distinct groups associated with different epidemic risks as illustrated in Fig. 7, was obtained for the fixed population mobility rate $R_p = 0.0004$. Will a different mobility rate change the phenomenon? Note that, for $R_p = 0.0004$, the scenarios associated with (M_j^H, O_H) and (M_j^L, O_H) have similar dynamical behaviors of spreading while the scenarios associated with (M_j^H, O_L) have low risks. It thus suffices to focus on the scenarios associated with (M_j^H, O_H) and (M_j^L, O_L) to examine the effects of different R_p values on the risk.

We first study the scenarios where the first outbreak occurs in the high-density city. Fig. 9(**a**) shows $R_{isk}(S_1, M_1^H, O_H)$ versus R_p, where the risks for the two cities are approximately the same, indicating that the value of R_p does not affect the spreading dynamics. The reason is that, for the scenarios associated with (M_j^H, O_H), the risks are already high, making high population density as the major cause of the spreading, regardless of the value of R_p. As the SD level increases, the following transition occurs: $(S_1, M_1^H, O_H) \rightarrow (S_3, M_1^H, O_H)$. Fig. 9(**b**) shows that the risk is reduced by more than one-third, but it is still independent of the value of R_p, indicating that traveling between the two cities has little effect on the spreading, for the same reason that the epidemic has already occurred in the high-density city and increasing the SD level can reduce the risk. For large MGE sizes, *i.e.*, $(S_1, M_1^H, O_H) \rightarrow (S_1, M_3^H, O_H)$, the risks are high but independent of the mobility rate, as shown in Fig. 9(**c**), which is similar to that in Fig. 9(**a**), indicating again that the population density of the first outbreak city is a key spreading factor and SD can be effective at suppressing the spreading.

If the first outbreak occurs in the low-density city (scenarios (S_1, M_1^L, O_L)), a higher mobility rate can increase the risk, as illustrated in Fig. 9(**d**). Applying stronger SD to induces the transition $(S_1, M_1^L, O_L) \rightarrow (S_3, M_1^L, O_L)$. For these scenarios, the risk is significantly lower but it increases with the mobility rate, as shown in Fig. 9(**e**). That is, SD can be effective at mitigating the spreading even for more intense population movements. When the MGE size increases to induce the transition $(S_1, M_1^L, O_L) \rightarrow (S_1, M_3^L, O_L)$, the risk will increase even for small mobility rate, as shown in Fig. 9(**f**).

Table 1
Comparison of two major COVID-19 Mass Gathering Events.

Category	Daegu MGE	Seoul MGE	References
Cluster	Shincheonji church	Sarang Jeil church	
Level of SD	S_0	S_2	[52]
Level of MGE	M_3	M_1	
Cumulative cases of entire Korea	8132	8556	[53,54]
Cumulative cases of MGE	5006	1136	[53,54]
Proportion of MGE cases	61.6%	13.3%	[53,54]
Scenario	(S_0, M_3^L, O_L)	(S_2, M_1^H, O_H)	
Group for R_{isk}	**Group 2**	**Group 3**	

3.5. Applications of risk management

In South Korea, during the pandemic several mass infections emerged: two major events were the Shincheonji Church cluster in Daegu from Feb. 18, 2020 to March 15, 2020 (referred to as the Daegu MGE) and the Sarang Jeil Church cluster in Seoul from Aug. 12, 2020 to

Sep. 3, 2020 (referred to as the Seoul MGE). Fig. 1**a** illustrates the time-series data of these two significant infection events. The Shincheonji Church cluster had triggered a significant outbreak concentrated in Daegu/Gyeongbuk, as shown in Fig. 1**b**. A total of 5006 confirmed COVID-19 cases were related to the Daegu MGE [53], as shown in Fig. 1**b**. Those from Daegu MGE accounted for 61.6% among the total COVID-19 cases, as illustrated in Fig. 1**b**, indicating that the COVID-19 outbreak was concentrated mainly in Daegu/Gyeongbuk. The Seoul MGE is depicted in Fig. 1**c**, where a total of 1136 COVID-19 cases were linked to the Sarang Jeil Church [54]. The Seoul MGE cases represented only 13.3% of the total patients, suggesting that there were more patients reported in other regions and explaining the sporadic and continuous transmission across a nationwide spread.

Table 1 provides a summary of the two scenarios for the Daegu and Seoul MGEs [53,54]. The scenario for Daegu MGE is represented by (S_0, M_3^L, O_L). If SD had been enhanced, transitioning from (S_0, M_3^L, O_L) to (S_3, M_3^L, O_L) would have resulted in a significant decrease in R_{isk}, moving it to the lowest rank within Group 2. Furthermore, if there had been strong regulations for mobility-generated exposure, it would have shifted from (S_0, M_3^L, O_L) to (S_0, M_0^L, O_L), reducing it to Group 1. However, the scenario for the Seoul MGE corresponds to (S_2, M_1^H, O_H), placing it in the middle-risk category, Group 3, as depicted in Fig. 7.

Despite the Seoul MGE having a relatively higher level of social distancing, a lower level of movement guidance enforcement, and fewer cumulative cases compared to the Daegu MGE, it still falls into a higher risk category. This is primarily due to the initial outbreak and the MGE occurred in Seoul, which has a high population density, thus increasing the potential risk of disease transmission within the city. Indeed, the transmission impact originating from the Sarang Jeil Church was significant, as shown in Fig. 1**a** & **c**. The number of cases from the Seoul MGE involved extensive spread to other regions, resulting in sporadic infections that continued over time, making the full scale of the outbreak difficult to assess. If SD were enhanced, transitioning from (S_2, M_1^H, O_H) to (S_3, M_1^H, O_H) would significantly reduce R_{isk} within Group 3. Moreover, if both SD and mobility-generated exposure regulations were controlled together, it could lead to a substantial reduction, potentially to the lower ranks of Group 2. This could demonstrate that SD was significantly more effective than MGE regulation, and even more so when both measures were combined together.

4. Conclusion and discussion

The COVID-19 pandemic has raised the need to prepare for the next possible pandemic triggered by some unknown pathogen. It is quite likely that a vaccine will not be available at the initial outbreak, so NPIs are necessary, as the extent of any disease transmission depends on diverse epidemic environmental factors such as the population density, mobility rates, the location of the initial outbreak, social distancing, mass gathering events, etc. In this paper, employing agent-based modeling, we developed a comprehensive framework to assess, quantitatively, the effects of these factors on epidemic spreading and the risks associated with distinct scenarios defined according to the environmental factors. Our concrete setting is agent movements between two cities with different population densities. The foundation of our model construction was the empirical data gathered during the COVID-19 outbreaks from Feb. 2020 to Oct. 2021 in two cities in South Korea: Seoul with a very high population density and Daegu with a relatively low population density (see **Methods**). Various epidemic environmental factors were incorporated into the model. Based on the probabilities pertinent to the local, influential, and wide transmissions, we introduced a quantitative measure to characterize the epidemic risk associated with different scenarios.

We first focused on the impacts of SD and MGEs in the two cities. In the low-density city, MGEs and SD have predictable effects on spreading and outbreak reduction, MGEs tend to facilitate disease transmission while SD can effectively curb the epidemic, as expected. In the high-density city, the disease transmission can be substantial regardless of the scale of MGEs, suggesting that MGEs do not have a significant impact on the outbreak, primarily because the high population density itself acted as a powerful driver of disease spread. In this case, SD can be highly effective as a vital approach to containing the spreading. Our analysis underscores that the effectiveness of NPIs such as SD and MGEs depends on the population density of the city. We then studied the effects of population movements between the two cities on the epidemic associated with all scenarios and discovered that the scenarios can be divided into three groups associated with distinct risks. In particular, individuals in the high-density city already face an inherent risk of outbreaks due to the high population density. Our analysis revealed that the most precarious scenario occurs when the initial outbreak takes place in the high-density city, as expected. For these scenarios, MGEs have little effect on the spreading, making SD the primary means of risk reduction. Our results reaffirms, at a quantitative level, that among the myriad factors influencing epidemic spreading, high population density is the most critical factor.

We discuss our findings on risk management of epidemic spreading with respect to previous relevant results. Past research emphasized the importance of implementing NPIs as well as population density and movements on the epidemic risk [55–57]. For example, the population density was identified as a crucial factor in estimating the total number of infection cases at the country level [55], highlighting the necessity of incorporating population density into transmission models. Another study [56] underscored the importance of a city's population density and social contact patterns, particularly in high-density cities, where moderate physical distancing measures can effectively control the extent of infection. An analysis of the territorial differences in the spread of COVID-19 and the associated excess mortality in Europe and the US was carried out [57], revealing the importance of territorial characteristics in shaping effective public health policy responses at the regional level. These studies highlight the multifaceted nature of infectious disease spread and vast complexity in devising effective control measures. In comparison, we have provided a comprehensive analysis encompassing all the factors and the effects of their interplay on spreading, which include NPIs, MGEs, population density of the first outbreak, and population movement, with an emphasis on the importance of tailoring effective control interventions to prevent disease spread according to the specific characteristics of each region. In particular, our study suggests that each city should conduct a thorough assessment of its own risk profile based on its distinct characteristics and develop quarantine policies tailored to its unique circumstances.

Taken together, an extensive investigation of the risk of disease spread as in our work is necessary to manage epidemic risks when implementing NPIs. Our research represents a foundational framework for the management of risk across a spectrum of epidemic scenarios, facilitating the application of multiple NPIs, and provides a quantitative base for health authorities and governments to make informed decisions to safeguard public health in the face of potential infectious disease outbreaks.

CRediT authorship contribution statement

Jaewoo Hwang: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Hyojung Lee:** Writing – original draft, Investigation, Formal analysis, Conceptualization. **Sunmi Lee:** Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. **Bongsoo Jang:** Writing – original draft, Investigation, Formal analysis, Conceptualization. **Younghae Do:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Ying-Cheng Lai:** Writing – review & editing, Writing – original draft, Supervision, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets and codes are available from https://github.com/jaewoohwang/Agent-based-model.

Acknowledgments

We thank the Korea Disease Control and Prevention Agency (KDCA) for providing the epidemiological data. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (Nos. NRF-2022R1A5A1033624 & 2022R1A2C3011711). The work at Arizona State University was supported by the Air Force Office of Scientific Research, United States through Grant No. FA9550-21-1-0438.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chaos.2024.115490.

References

[1] Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Whittaker C, Zhu H, Berah T, Eaton JW, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020;584(7820):257–61.
[2] Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S, Boonyasiri A, Cucunuba Perez Z, Cuomo-Dannenburg G, et al. Report 9: impact of non-pharmaceutical interventions (nPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College London; 2020.
[3] Aleta A, Martín-Corral D, Bakker MA, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, Dean NE, Halloran ME, Longini Jr IM, et al. Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. Proc Natl Acad Sci USA 2022;119(26):e2112182119.
[4] Che Mat NF, Edinur HA, Abdul Razab MKA, Safuan S. A single mass gathering resulted in massive transmission of COVID-19 infections in Malaysia with further international spread. J Travel Med 2020;27(3):taaa059.
[5] Kanaujiya AK, Tiwari V. The statistical analysis to find correlation between mass gathering event and COVID-19 pandemic outbreak: The prayagraj magh mela 2021 experience. Proc Natl Acad Sci USA 2023;93(1):107–11.
[6] Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, Giattino C, Rodés-Guirao L. A global database of COVID-19 vaccinations. Nat Hum Behav 2021;5(7):947–53.
[7] Gozzi N, Tizzoni M, Chinazzi M, Ferres L, Vespignani A, Perra N. Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in santiago de Chile. Nature Commun 2021;12(1):2429.
[8] World Health Organization, et al. COVID-19 weekly epidemiological update. 108. 2022, 7september 2022.
[9] Kim N, Kang SJ, Tak S. Reconstructing a COVID-19 outbreak within a religious group using social network analysis simulation in Korea. Epidemiol Health 2021;43.
[10] Quadri SA. COVID-19 and religious congregations: Implications for spread of novel pathogens. Int J Infect Dis 2020;96:219–21.
[11] Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. J Travel Med 2020;27(3):taaa038.
[12] Ebrahim SH, Memish ZA. COVID-19–the role of mass gatherings. Travel Med Infect Dis 2020;34:101617.
[13] Ahammer A, Halla M, Lackner M. Mass gatherings contributed to early COVID-19 mortality: Evidence from US sports. Contemp Econ Policy 2023.
[14] Kwok KO, Chan HHH, Huang Y, Hui DSC, Tambyah PA, Wei WI, Chau PYK, Wong SYS, Tang JWT. Inferring super-spreading from transmission clusters of COVID-19 in Hong Kong, Japan, and Singapore. J Hosp Infect 2020;105(4):682–5.
[15] Feltham E, Forastiere L, Alexander M, Christakis NA. Mass gatherings for political expression had no discernible association with the local course of the COVID-19 pandemic in the USA in 2020 and 2021. Nat Hum Behav 2023;7(10):1708–28.
[16] World Health Organization, et al. Key planning recommendations for mass gatherings in the context of COVID-19: interim guidance, 29 May 2020. Technical report, World Health Organization; 2020.
[17] Lee H, Jang G, Cho G. Forecasting COVID-19 cases by assessing control-intervention effects in Republic of Korea: a statistical modeling approach. Alex Eng J 2022;61(11):9203–17.
[18] Jeon J, Han C, Kim T, Lee S. Evolution of responses to COVID-19 and epidemiological characteristics in South Korea. Int J Environ Res Public Health 2022;19(7):4056.
[19] Oka T, Wei W, Zhu D. The effect of human mobility restrictions on the COVID-19 transmission network in China. PLoS One 2021;16(7):e0254403.
[20] Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Pastore y Piontti A, Mu K, Rossi L, Sun K, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020;368(6489):395–400.
[21] Kephart JL, Delclòs-Alió X, Rodríguez DA, Sarmiento OL, Barrientos-Gutiérrez T, Ramirez-Zea M, Quistberg DA, Bilal U, Roux AVD. The effect of population mobility on COVID-19 incidence in 314 latin American cities: a longitudinal ecological study with mobile phone location data. Lancet Digit Health 2021;3(11):e716–22.
[22] Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J. Mobility network models of COVID-19 explain inequities and inform reopening. Nature 2021;589(7840):82–7.
[23] Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, Liao Q, Kwan MYW, Lee SL, Chiu SS, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020;5(5):e279–88.
[24] Bergman NK, Fishman R. Correlations of mobility and Covid-19 transmission in global data. PLoS One 2023;18(7):e0279484.
[25] Leung WTM, Meeyai A, Holt HR, Khieu B, Chhay T, Seng S, Pok S, Chiv P, Drake T, Rudge JW. Social contact patterns relevant for infectious disease transmission in Cambodia. Sci Rep 2023;13(1):5542.
[26] Chowell G, Nishiura H, Viboud C. Modeling rapidly disseminating infectious disease during mass gatherings. BMC Med 2012;10(1):1–12.
[27] Domènech-Montoliu S, Pac-Sa MR, Vidal-Utrillas P, Latorre-Poveda M, Del Rio-González A, Ferrando-Rubert S, Ferrer-Abad G, Sánchez-Urbano M, Aparisi-Esteve L, Badenes-Marques G, et al. Mass gathering events and COVID-19 transmission in Borriana (Spain): A retrospective cohort study. PLoS One 2021;16(8):e0256747.
[28] Ryu H, Abulali A, Lee S. Assessing the effectiveness of isolation and contact-tracing interventions for early transmission dynamics of COVID-19 in South Korea. IEEE Access 2021;9:41456–67.
[29] Cho G, Kim YJ, Seo SH, Jang G, Lee H. Cost-effectiveness analysis of COVID-19 variants effects in an age-structured model. Sci Rep 2023;13(1):15844.
[30] Giordano G, Colaneri M, Di Filippo A, Blanchini F, Bolzern P, De Nicolao G, Sacchi P, Colaneri P, Bruno R. Modeling vaccination rollouts, SARS-CoV-2 variants and the requirement for non-pharmaceutical interventions in Italy. Nature Med 2021;27(6):993–8.
[31] Williams SN, Dienes K, Jaheed J, Wardman JK, Petts J. Effectiveness of communications in enhancing adherence to public health behavioural interventions: a COVID-19 evidence review. Philos Trans R Soc A 2023;381(2257):20230129.
[32] Leung K, Wu JT, Leung GM. Real-time tracking and prediction of COVID-19 infection using digital proxies of population mobility and mixing. Nature Commun 2021;12(1):1501.
[33] Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Med 2020;26(6):855–60.
[34] Wang Y, Xiong H, Liu S, Jung A, Stone T, Chukoskie L. Simulation agent-based model to demonstrate the transmission of COVID-19 and effectiveness of different public health strategies. Front Comput Sci 2021;3:642321.
[35] Aleta A, Martin-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, Dean NE, Halloran ME, Longini Jr IM, Merler S, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat Hum Behav 2020;4(9):964–71.
[36] Al-Bazi A, Madi F, Monshar AA, Eliya Y, Adediran T, Khudir KA. Modelling the impact of non-pharmaceutical interventions on COVID-19 exposure in closed-environments using agent-based modelling. Int J Healthc Manag 2023. http://dx.doi.org/10.1080/20479700.2023.2189555.
[37] Lee H, Abdulali A, Park H, Lee S. Optimal region-specific social distancing strategies in a complex multi-patch model through reinforcement learning. Math Comput Simulation 2024. http://dx.doi.org/10.1016/j.matcom.2024.06.013.
[38] Seong H, Hyun HJ, Yun JG, Noh JY, Cheong HJ, Kim WJ, Song JY. Comparison of the second and third waves of the COVID-19 pandemic in South Korea: Importance of early public health intervention. Int J Infect Dis 2021;104:742–5.
[39] Bae TW, Kwon KK, Kim KH. Mass infection analysis of COVID-19 using the SEIRD model in Daegu-Gyeongbuk of Korea from Apr. to May, 2020. J Korean Med Sci 2020;35(34):e317.
[40] Li MY, Muldowney JS. Global stability for the SEIR model in epidemiology. Math Biosci 1995;125(2):155–64.
[41] Xin H, Li Y, Wu P, Li Z, Lau EHY, Qin Y, Wang L, Cowling BJ, Tsang TK, Li Z. Estimating the latent period of coronavirus disease 2019 (COVID-19). Clin Infect Dis 2022;74(9):1678–81.
[42] Davies NG, Klepac P, Liu Y, Prem K, Jit M, Eggo RM. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nature Med 2020;26(8):1205–11.

[43] Wang W, Tang M, Yang H, Do Y, Lai YC, Lee G. Asymmetrically interacting spreading dynamics on complex layered networks. Sci Rep 2014;4(1):5097.
[44] Antonini C, Calandrini S, Bianconi F. A modeling study on vaccination and spread of SARS-CoV-2 variants in Italy. Vaccines 2021;9(8):915.
[45] Hagen A. How dangerous is the delta variant (b. 1.617. 2). 2021, mBio.
[46] Petherick A, Goldszmidt R, Andrade EB, Furst R, Hale T, Pott A, Wood A. A worldwide assessment of changes in adherence to COVID-19 protective behaviours and hypothesized pandemic fatigue. Nat Hum Behav 2021;5(9):1145–60.
[47] Korea disease control and prevention agency. available at: http://ncov.mohw.go.kr/.
[48] Liu QH, Ajelli M, Aleta A, Merler S, Moreno Y, Vespignani A. Measurability of the epidemic reproduction number in data-driven contact networks. Proc Natl Acad Sci USA 2018;115(50):12680–5.
[49] Guerra FM, Bolotin S, Lim G, Heffernan J, Deeks SL, Li Y, Crowcroft NS. The basic reproduction number (R_0) of measles: a systematic review. Lancet Infect Dis 2017;17(12):e420–8.
[50] Drezner Z, Turel O, Zerom D. A modified Kolmogorov–Smirnov test for normality. Comm Statist Simulation Comput 2010;39(4):693–704.
[51] Lilliefors HW. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. J Amer Statist Assoc 1967;62(318):399–402.
[52] Korea disease control and prevention agency. Social distancing implementation history. 2024, Available at:https://www.data.go.kr/data/15106451/fileData.do (accessed on 2024).
[53] Korea disease control and prevention agency. Coronavirus infectious disease-19 (COVID-19) – domestic occurrence status (mar. 15 2020regular briefing). 2024, Available at:https://www.mohw.go.kr/board.es?mid=a10503010100&bid=0027&cg_code= (accessed on 2024).
[54] Korea disease control and prevention agency. Coronavirus infectious disease-19 (COVID-19) – domestic occurrence status (sep. 3 2020regular briefing). 2024, Available at:https://www.mohw.go.kr/board.es?mid=a10503010100&bid=0027&cg_code= (accessed on 2024).
[55] Wong DWS, Li Y. Spreading of COVID-19: Density matters. PLoS One 2020;15(12):e0242398.
[56] Huang B, Wang J, Cai J, Yao S, Chan PKS, H-w Tam T, Hong Y-Y, Ruktanonchai CW, Carioli A, Floyd JR, et al. Integrated vaccination and physical distancing interventions to prevent future COVID-19 waves in Chinese cities. Nat Hum Behav 2021;5(6):695–705.
[57] Natale F, Iacus SM, Conte A, Spyratos S, Sermi F. Territorial differences in the spread of COVID-19 in European regions and US counties. PLoS One 2023;18(2):e0280780.