



Research Paper

What are the underlying transmission patterns of COVID-19 outbreak? An age-specific social contact characterization

Yang Liu^{a,b}, Zhonglei Gu^{a,b}, Shang Xia^{b,c,d,e}, Benyun Shi^{b,f}, Xiao-Nong Zhou^{b,c,d,e}, Yong Shi^{g,h}, Jiming Liu^{a,b,*}

^a Department of Computer Science, Hong Kong Baptist University, Hong Kong, China

^b HKBU-CSD & NIPD Joint Research Laboratory for Intelligent Disease Surveillance and Control, Hong Kong, China

^c National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research), Shanghai, China

^d Key Laboratory of Parasite and Vector Biology, National Health and Commission of the People's Republic of China, Shanghai, China

^e WHO Collaborating Center for Tropical Diseases, Shanghai, China

^f School of Computer Science and Technology, Nanjing Tech University, Nanjing, China

^g School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China

^h Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article History:

Received 18 March 2020

Revised 7 April 2020

Accepted 9 April 2020

Available online 18 April 2020

Keywords:

COVID-19

Underlying transmission patterns

Age-specific social contact patterns

Retrospective and prospective analysis

ABSTRACT

Background: COVID-19 has spread to 6 continents. Now is opportune to gain a deeper understanding of what may have happened. The findings can help inform mitigation strategies in the disease-affected countries.

Methods: In this work, we examine an essential factor that characterizes the disease transmission patterns: the interactions among people. We develop a computational model to reveal the interactions in terms of the social contact patterns among the population of different age-groups. We divide a city's population into seven age-groups: 0–6 years old (children); 7–14 (primary and junior high school students); 15–17 (high school students); 18–22 (university students); 23–44 (young/middle-aged people); 45–64 years old (middle-aged/elderly people); and 65 or above (elderly people). We consider four representative settings of social contacts that may cause the disease spread: (1) individual households; (2) schools, including primary/high schools as well as colleges and universities; (3) various physical workplaces; and (4) public places and communities where people can gather, such as stadiums, markets, squares, and organized tours. A contact matrix is computed to describe the contact intensity between different age-groups in each of the four settings. By integrating the four contact matrices with the next-generation matrix, we quantitatively characterize the underlying transmission patterns of COVID-19 among different populations.

Findings: We focus our study on 6 representative cities in China: Wuhan, the epicenter of COVID-19 in China, together with Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen, which are five major cities from three key economic zones. The results show that the social contact-based analysis can readily explain the underlying disease transmission patterns as well as the associated risks (including both confirmed and unconfirmed cases). In Wuhan, the age-groups involving relatively intensive contacts in households and public/communities are dispersedly distributed. This can explain why the transmission of COVID-19 in the early stage mainly took place in public places and families in Wuhan. We estimate that Feb. 11, 2020 was the date with the highest transmission risk in Wuhan, which is consistent with the actual peak period of the reported case number (Feb. 4–14). Moreover, the surge in the number of new cases reported on Feb. 12 and 13 in Wuhan can readily be captured using our model, showing its ability in forecasting the potential/unconfirmed cases. We further estimate the disease transmission risks associated with different work resumption plans in these cities after the outbreak. The estimation results are consistent with the actual situations in the cities with relatively lenient policies, such as Beijing, and those with strict policies, such as Shenzhen.

Interpretation: With an in-depth characterization of age-specific social contact-based transmission, the retrospective and prospective situations of the disease outbreak, including the past and future transmission risks, the effectiveness of different interventions, and the disease transmission risks of restoring normal social activities, are computationally analyzed and reasonably explained. The conclusions drawn from the study not only provide a comprehensive explanation of the underlying COVID-19 transmission patterns in China,

* Corresponding author at: Department of Computer Science, Hong Kong Baptist University, Hong Kong, China.
E-mail address: jiming@comp.hkbu.edu.hk (J. Liu).

but more importantly, offer the social contact-based risk analysis methods that can readily be applied to guide intervention planning and operational responses in other countries, so that the impact of COVID-19 pandemic can be strategically mitigated.

Funding: General Research Fund of the Hong Kong Research Grants Council; Key Project Grants of the National Natural Science Foundation of China.

© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Research in context

Evidence before this study

To timely control COVID-19, it is of vital importance to gain an in-depth understanding of the underlying transmission patterns among different populations throughout various phases of the COVID-19 outbreak. So far, efforts have been made in predicting the future cases based on the historical reported cases, analyzing the early epidemiology of COVID-19 using social media data, and modeling and investigating the effects of the case- and contact-isolation in containing the COVID-19 outbreak. However, it remains unclear how the disease transmits among the populations and what are the transmission patterns.

Added value of this study

To our knowledge, this is the first work that explicitly characterizes and quantifies the underlying transmission patterns of the COVID-19 outbreak. We show that the age-specific social contact patterns can accurately characterize the interactions among different groups of people, and thus provide explanations on the underlying disease transmission and associated risks in various phases of the outbreak. We analyze the situations in 6 representative cities in China. These cities are Wuhan, the epicenter of COVID-19 in China, and Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen, five cities situated in three major economic zones.

Implications of all the available evidence

This work has touched upon an important problem at a critical moment in time: COVID-19 has spread to more than 180 countries in 6 continents, and a deeper understanding of what may have happened in the outbreak is now overdue. Addressing this key question enables us to gain insights into the retrospective and prospective situations of the disease outbreak; this in turn will help further answer a series of questions in control and prevention of the disease; namely, how future risks and trends in different regions may evolve, how effective different intervention strategies can be in controlling the outbreak, and what may happen if people gradually return to schools and workplaces in the later stage of the outbreak at some point. Thus, through the prism of the outbreak in China, this work offers effective tools and insights to other countries or regions for their intervention planning and operational responses.

1. Introduction

The novel coronavirus disease (COVID-19) has spread widely at the global level [1]. According to the statistics from the Johns Hopkins Coronavirus Resource Center, until Apr. 9, 2020, there have been more than 1,500,000 confirmed cases found in more than 180 countries covering 6 continents [2]. In view of the severity of the disease spread, the World Health Organization (WHO) has officially declared COVID-19 as a pandemic [3].

In order to timely mitigate the impact of COVID-19 pandemic, there is an urgent need to understand the underlying transmission patterns among different populations throughout different phases of the COVID-19 outbreak [4]. By doing so, we can provide insights into what may have happened retrospectively and what can be anticipated prospectively of the disease outbreak, so as to further address a series of important questions that follow, such as how future risks and trends in different regions may evolve, how effective can different intervention strategies be in controlling the outbreak, and what may happen if people gradually return to schools and workplaces. More importantly, offering the methods of social contact-based risk analysis and demonstrating them in the case of COVID-19 outbreak in China can help other countries or regions in conducting similar studies and making subsequent intervention policies.

2. Methods

2.1. Scope of this study

We select 6 major cities in China for our study: Wuhan, Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen; their geographical locations and the disease situations (in terms of total case number from Dec. 2019 to Feb. 2020) are shown in Fig. 1. Wuhan was the epicenter of COVID-19 in China [5–7]. The other 5 cities are representative in that they are situated in the three most important economic zones in China, which contribute more than 40% of the national GDP. Specifically, Beijing and Tianjin are representing the Jing-Jin-Ji (Beijing-Tianjin-Hebei) Metropolitan Region in Northern China. Hangzhou and Suzhou are the major players in the Yangtze River Delta City Cluster in Eastern China. Shenzhen is the flagship in the Greater Bay Area in Southern China. Another important reason to select these cities for our study is that the population of these cities contains a large number of migrant workers and college students from other cities or provinces. The frequent human mobility largely increases the risk of imported cases, posing great challenges to the control and prevention of COVID-19, especially when people are gradually returning to workplaces and schools in a later stage.

2.2. Data sources

The data used in our study include:

- (1) The daily confirmed cases from Dec. 8, 2019 to Feb. 29, 2020 in Wuhan, Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen, which were accessed and collected from the websites of the Health Commission of Hubei Province [8], the Beijing Municipal Health Commission [9], the Tianjin Municipal Health Commission [10], the Hangzhou Municipal Health Commission [11], the Suzhou Municipal Health Commission [12], and the Shenzhen Municipal Data Open Platform [13], respectively.
- (2) The demographic data of Wuhan [14], Beijing [15], Tianjin [16], Hangzhou [17], Suzhou [18], and Shenzhen [19].

2.3. Age-specific social contact characterization

The underlying transmission patterns of COVID-19 among different populations are difficult to characterize because they are complex and

related to various observations and disease-related factors, including the number of confirmed cases, the potential risks brought by unconfirmed cases, the distribution of different case categories (indigenous/imported) in different regions/cities, the population distribution of different age-groups, the social contact patterns in different settings (e.g., households, schools, workplaces, and public places), the extent of interventions implemented in different regions/cities, etc. To address this challenging issue in a fundamental way, we examine an essential factor that characterizes the disease transmission patterns: the interactions among people [20,21]. Specifically, we examine the interactions in terms of the social contact patterns among the population of different age-groups. To characterize the age-specific social contact-based transmission, we divide a city's population into seven age-groups: 0–6 years old (children); 7–14 (primary and junior high school students); 15–17 (high school students); 18–22 (university and college students); 23–44 (young/middle-aged people); 45–64 years old (middle-aged/elderly people); and 65 or above (elderly people). The population in each of the seven groups has its own specific social circles, gathering places, or activity patterns. Meanwhile, we consider four representative settings of social contacts that may cause the disease spread: (1) individual households, which may lead to the transmission within families; (2) schools, including primary/high schools as well as colleges and universities, which may cause the spread among students and teachers; (3) various physical workplaces, which may affect in-office and outside workers; and (4) public places and communities, such as stadiums, markets, squares, and organized tours, where the spread within a dense population may arise.

Let G_1 – G_7 be the seven age-groups: 0–6, 7–14, 15–17, 18–22, 23–44, 45–64, and 65 or above, respectively. Then the contact frequencies between an individual from G_i and an individual from G_j ($i, j = 1, \dots, 7$) under the settings of Households, Schools, Workplaces, and Public/community, denoted by $C_{ij}^H, C_{ij}^S, C_{ij}^W,$ and C_{ij}^P , respectively, are calculated as follows:

$$C_{ij}^H = \frac{C_{ij}^H}{P_i P_j}, C_{ij}^S = \frac{C_{ij}^S}{P_i P_j}, C_{ij}^W = \frac{C_{ij}^W}{P_i P_j}, C_{ij}^P = \frac{C_{ij}^P}{P_i P_j}, \quad (1)$$

$$C^H = [C_{ij}^H]_{7 \times 7}, C^S = [C_{ij}^S]_{7 \times 7}, C^W = [C_{ij}^W]_{7 \times 7}, C^P = [C_{ij}^P]_{7 \times 7}, \quad (2)$$

where $C_{ij}^H, C_{ij}^S, C_{ij}^W,$ and C_{ij}^P denote the total number of contacts between individuals from G_i and those from G_j under the settings of Households, Schools, Workplaces, and Public/community, respectively, P_i and P_j denote the population of G_i and G_j , and $C^H, C^S, C^W,$ and C^P denote the 7×7 social contact matrices. In Eq. (1), we use demographic data to calculate P_i ($i = 1, \dots, 7$). For $C_{ij}^H, C_{ij}^S, C_{ij}^W,$ and C_{ij}^P , as the city-specific data of social contacts between age-groups is unavailable, we adopt a computational method [20] to estimate them. The appropriateness of using such a computational method for social-contact estimation in data-scarce situations has been validated²⁰: the estimated $C^H, C^S, C^W,$ and C^P are consistent with the results from a real-world social contact survey [22] in terms of the strong assortativeness and the appearance of similar secondary diagonal contact patterns.

Next, we represent the overall age-specific social contact matrix as a linear combination of the above four matrices [21]:

$$C = r_H C^H + r_S C^S + r_W C^W + r_P C^P, \quad (3)$$

where $r_H, r_S, r_W, r_P \geq 0$ are the weights of matrices $C^H, C^S, C^W,$ and C^P , respectively, and satisfy that $r_H + r_S + r_W + r_P = 1$. According to Xia et al.'s work [21], the initial weights of four social contact matrices in our study are set as: $r_H = 0.31, r_S = 0.24, r_W = 0.16, r_P = 0.29$. It should be pointed out that similar settings have also been utilized to simulate the contact matrices in other studies, e.g., the similar weight for the household matrix has been used to calculate the contact matrix for Varicella and Parvovirus B19 [23]. The results in Fig. 3 show that our model with the above parameter settings can adequately capture the disease trends in different cities; our sensitivity analysis also confirms that the developed model (to be described below) is relatively robust to the parameter settings.

With the overall age-specific social contact matrix C , we can characterize the disease transmission pattern using the next-generation matrix K , [24]:

Region	City	Case No.	Regional Rank (Case No.)
Hubei Province	Wuhan (WH)	49426	1 (in China)
Jing-Jin-Ji Region	Beijing (BJ)	414	1
	Tianjin (TJ)	136	2
Yangtze River Delta	Hangzhou (HZ)	169	4
	Suzhou (SZ)	87	11
Greater Bay Area	Shenzhen (SZ)	418	1

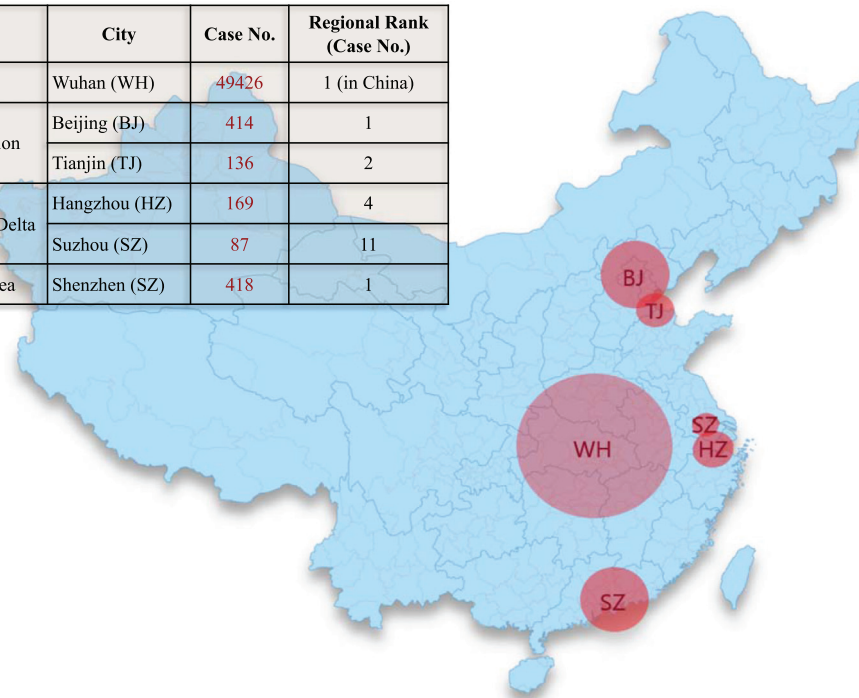


Fig. 1. The geographical locations and the disease situations (total number of cases from Dec. 2019 to Feb. 2020) of six major cities selected in this study: Wuhan, Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen.

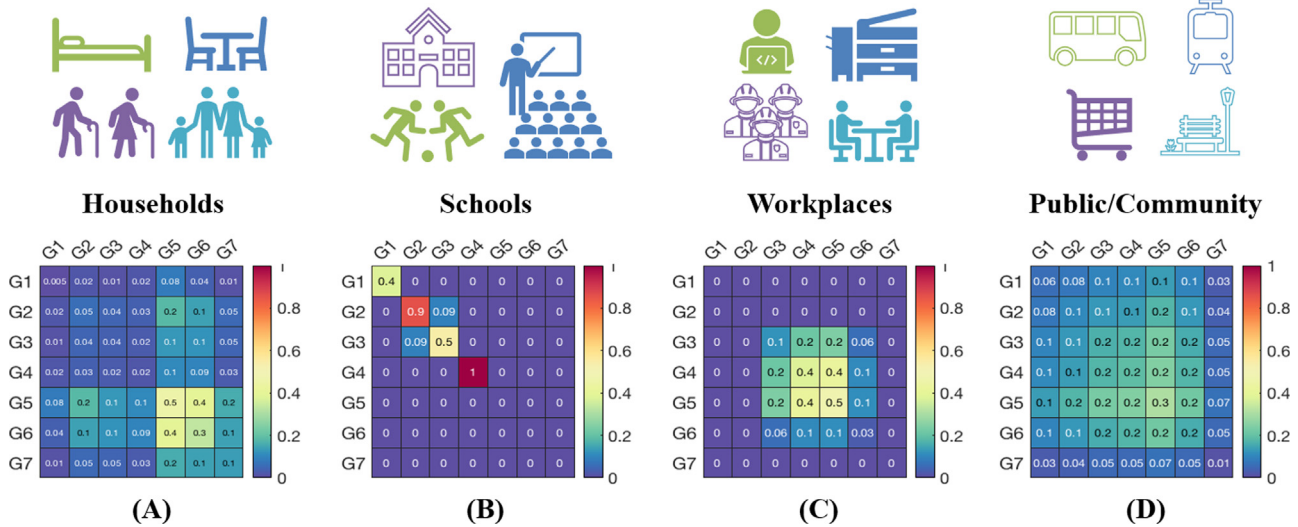


Fig. 2. Measurement of the intensity of social contacts among seven age-groups (G_1 : 0–6; G_2 : 7–14; G_3 : 15–17; G_4 : 18–22; G_5 : 23–44; G_6 : 45–64; and G_7 : 65 or above) in four major settings: (A) households; (B) schools; (C) workplaces; and (D) public/community, in Wuhan. The contact patterns in these four settings are consistent with common social behaviors observed in a typical society. Specifically, as shown in (A), the majority of the social contacts within households occur across different generations. (B) demonstrates that the main social contact in schools centers around kids in the same age-group. As depicted in (C), workplaces are dominated by social contacts among young adults and the middle-aged adults. In (D), the social contacts are more diverse when people are in public places.

$$K_t = \left(\frac{\mu}{\gamma}\right) S_t B C A, \tag{4}$$

$$I_{t+1} = K_t I_t,$$

where S_t , B , A , and K_t are 7×7 matrices and I_t is a 7×1 vector. Specifically, S_t , B , and A are diagonal matrices, with the diagonal elements s_{ii}^t , b_{ii} , and a_{ii} ($i = 1, \dots, 7$) being the size of susceptible population in G_i at the t -th generation of the disease infection, the individual susceptibility in G_i , and the infectivity of infected individuals in G_i , respectively. The i -th element in vector I_t denotes the number of infectious individuals in G_i at the t -th generation of the disease infection. By referring to Li et al.'s work [6], we set the reproduction number $R_0 = 2.2$. For the recovery rate γ , we calculate it as follows: First, according to the definition of recovery rate [25], it is the reciprocal of the duration of being infectious (i.e., $\gamma = 1/\text{infectious period}$). Then, according to Svensson's work [26], the infectious period is equal to the mean generation time minus the mean latent period, so we have $\gamma = 1/(\text{mean generation time} - \text{mean latent period})$. Further, as pointed out in Binti Hamzah et al.'s work [27] and Liu et al.'s work [28], the mean generation time is 7.5 days. Moreover, Wu et al.'s work [7] indicates that the mean incubation period is 5.2 days for COVID-19. As there is not precise infection date for those patients to estimate the mean latent period, we use the mean incubation period to approximate the mean latent period. Therefore, the recovery rate is estimated as $\gamma = 1/(7.5 - 5.2)$. For the infectivity, we set $a_{ii} = 1.0$ (for $i = 1, \dots, 7$) according to Xia et al.'s work [21]. For the susceptibility b_{ii} , as it represents the probability of being infected when a susceptible individual is exposed to infectious contacts, we estimate it as follows: For each G_i , we first calculate its infected population ratio r_i by dividing the number of infected cases in G_i by P_i , i.e., $r_i = n_i/P_i$. With r_i calculated for all 7 age-groups, we then obtain a multiplier, $1/\min\{r_1, \dots, r_7\}$, through normalizing the smallest r_i to 1, and inflate all other infected population ratios by $1/\min\{r_1, \dots, r_7\}$. Then we estimate the susceptibility as $b_{ii} = r_i/\min\{r_1, \dots, r_7\}$. As different cities have different numbers of infected cases and different population sizes, they will have different susceptibilities. Specifically, we have: $b_{11} = 1.00$, $b_{22} = 1.23$, $b_{33} = 35.33$, $b_{44} = 21.09$, $b_{55} = 13.18$, $b_{66} = 42.16$, and $b_{77} = 97.48$ for Wuhan; $b_{11} = 9.08$, $b_{22} = 1.00$, $b_{33} = 10.67$, $b_{44} = 4.15$, $b_{55} = 10.26$, $b_{66} = 14.58$, and $b_{77} = 17.90$ for Beijing; $b_{11} = 1.00$, $b_{22} = 1.86$, $b_{33} = 1.66$, $b_{44} = 3.14$, $b_{55} = 5.24$, $b_{66} = 9.86$, and $b_{77} = 11.94$ for Tianjin; $b_{11} = 1.06$, $b_{22} = 1.87$, $b_{33} = 1.00$, $b_{44} = 1.78$, $b_{55} = 5.46$, $b_{66} = 9.38$, and $b_{77} = 4.51$ for Hangzhou; $b_{11} = 1.07$, $b_{22} = 1.18$,

$b_{33} = 1.57$, $b_{44} = 1.00$, $b_{55} = 3.97$, $b_{66} = 7.20$, and $b_{77} = 3.14$ for Suzhou; and $b_{11} = 3.57$, $b_{22} = 4.72$, $b_{33} = 1.00$, $b_{44} = 2.76$, $b_{55} = 3.49$, $b_{66} = 17.59$, and $b_{77} = 32.51$ for Shenzhen.

The disease infection dynamics computed using Eqs. (3) and (4) corresponds to the situation without any intervention. To take the effect of intervention into consideration, it is important for us to further decrease r_H , r_S , r_W , r_P in Eq. (3), i.e., the weights of different social contact matrices accordingly. Similarly, if we consider different work resumption plans, we will need to increase these weights proportional to the rate of work resumption. Specifically, we reduce r_W from its original value to 0 as of Jan. 23 (the starting date of implementing stringent public health control policies). Moreover, we gradually recover its value from the starting date of our work resumption plans to reflect the effect of "back-to-work" policies. We apply the similar rationale to r_S and r_P . For r_H , as the public social distancing policies would increase social contacts within households, we increase the value of r_H starting from Jan. 23, and gradually reduce to its original value once the "back-to-work" policy kicks in.

2.4. Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the Article, or the decision to submit for publication. All authors had full access to all the data in the study and were responsible for the decision to submit the Article for publication.

3. Results

3.1. Social contact-based transmission characterization

As can be seen in Fig. 2, the distribution of age-groups involving relatively intensive contacts in households and public/communities is rather scattered, and thus it is easy to cause the disease spread among different age-groups in these two settings. This is consistent with the observation that the transmission of COVID-19 in the early stage mainly took place in public places and families. In contrast, the distribution of age-groups with intensive contacts in schools and workplaces are relatively concentrated. Moreover, the composition of people in these two settings is relatively stable, making the management easier than that in public places or communities. Because most of the schools and workplaces were closed before the Chinese

Spring Festival and have not been reopened or resumed yet, the scale of the COVID-19 outbreak in these two settings is relatively limited. However, if normal educational and economic activities are to be resumed, a large number of students and staff will gather in these two settings, which may present a real challenge to the control and prevention of COVID-19 infection in these concerned places.

3.2. Retrospective analysis of the disease outbreak

For different cities, the transmission patterns of COVID-19 might be different. In Wuhan, the cases were mainly indigenous. In the other 5 cities, the cases might be either indigenous or imported from Hubei. Therefore, for these 5 cities, we need to take both indigenous cases and imported cases into consideration when investigating the transmission patterns among populations. To model the potential local transmission risk caused by the imported cases, we use the following approach: First, for each confirmed case, we identify if it is imported or indigenous according to the information provided by the Municipal Health Commission [8–13]. If the case is an imported case, we consider its potential risk in bringing in local transmission. According to Li et al.'s results [6], the mean serial interval is 7.5 days, so we assume that for each imported case, from the day of arrival to the day of hospitalization, he/she could infect 1/7.5 person per day. We apply the same principle to all imported cases to estimate their potential infections. Those cases infected by the imported ones are considered as potential cases in our study. The confirmed cases and potential cases together constitute the disease transmission risk, i.e., the focus of the following retrospective and prospective analyses.

With the age-specific social contact-based transmission modeling, we are able to describe and explain what may have happened retrospectively and what can be anticipated prospectively of the COVID-19 outbreak. Fig. 3 shows the estimation on the trends of disease infection and the transmission risks associated with different work resumption plans based on the social contact patterns and reported cases. From the results of Wuhan (Fig. 3(A)), we can observe that the situation without any interventions (the brown line) is estimated to be much severer than that with interventions (the blue line), indicating the effectiveness of the interventions implemented in Wuhan. Here the interventions refer to various social distancing measures, including quarantine of patients, closure of workplaces and schools, suspension of public transportation, and requirement for people to wear masks [29–31].

It can also be observed from Fig. 3(A) that the date with the estimated peak number of the new cases was Feb. 11, which is consistent with the actual situation: the number of reported cases reached the peak during Feb. 4–14. Moreover, note that there was a sharp increase in the number of confirmed cases on 12 and 13 Feb. This is because that the National Health Commission of China adopted a new case definition in Hubei province. In the new case definition, the clinically diagnosed cases (suspected cases with pneumonic-type imaging characteristics) were included in the newly reported cases. To put it in the right context, the observed surge in the number of reported cases does not imply a large number of cases were found on 12 and 13 Feb, but an inflow of accumulated clinically confirmed cases from the past few days. As can be observed from Fig. 3(A), the result of our model (the gap between the blue line and the red bars) offers a reasonable explanation of the not-yet-reported cases over the period. Specifically, the model is not designed to provide a mathematical estimation/prediction that fits exactly to the number of reported cases, but to present an estimation of the risk to the community if certain measures are or are not exercised. One reasonable assumption included in the model is the consideration of unreported cases, as it is nearly impossible to timely capture all new cases given limited resources and the fixed capacity of the medical system. For this reason, our model estimates a case number larger than that of the reported confirmed cases, with the excessive number representing the potential risks that are not yet identified as confirmed cases.

3.3. Prospective analysis of disease transmission risks and economic impacts

The COVID-19 pandemic has hit the global economy by a storm. As the public health crisis escalates, countries have responded by enforcing social distancing measures, such as the closure of public venues and reduced working hours, to reduce the chance of contracting the highly contagious virus in a social setting. At the same time, this will inevitably lead to a massive decline in business activities, causing unprecedented economic loss to the countries. When the COVID-19 outbreak is contained, as in the case of Wuhan, China, it is foreseeable that the countries will need to think about how to safely resume social activities and bring work and life back to normal, as any pre-mature resumption of social contacts could potentially cause a rebound (second wave) in new infection cases. In the light of the pressing needs to provide a scientific ground for systematically planning the resumption of social/business activities near the end of the outbreak, we present a prospective analysis of different work resumption plans, which can enable us to assess not only the respective economic implications of the plans, but more importantly, the levels of disease transmission risks associated with the corresponding plans.

Specifically, to further understand what can be anticipated prospectively of the disease outbreak, we analyze what may happen if the social/business activities gradually restore from the strong control and isolation to the normal situation (including public and work places). We analyze the disease transmission risks associated with different work resumption plans in Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen, respectively. Since Wuhan was in a serious situation during the outbreak, its work resumption may take longer than the other 5 cities; its detailed plans for the resumption and associated risks are discussed in the Supplementary as shown in Fig. S1. We conduct our prospective study on 2 sets of different work resumption plans (Plans A1–A3 and Plans B1–B3) and accordingly examine their associated risks of disease transmission. Plans A1–A3 resume work when the disease transmission is well under control, i.e., the number of new reported cases is about to become zero. Specifically, Plans A1–A3 start on Feb. 17 (Monday) and finish the resumption in 1 week, 1/2 months, and 1 month, respectively. Plans B1–B3 are stricter than Plans A1–A3; they resume work when the number of new reported cases has been zero for three consecutive days. Therefore, Plans B1–B3 start on Feb. 24 (Monday) and, similarly, finish the resumption in 1 week, 1/2 months, and 1 month, respectively.

In order to parameterize our model for risk prediction with different work resumption plans, we estimate the percentage of work ongoing or recovered in each of those cities at the time when the resumption plans start. For Beijing, according to a document issued by the municipal government [32,33], eligible companies can resume work from Feb. 3. Therefore, we set a weekly increase of 10% in the resumption of work in Beijing from Feb. 3 until the resumption plan begins. For Tianjin, we use the same resumption settings as Beijing because both cities are situated in the Jingjinji Metropolitan Region. For Hangzhou, according to the municipal government regulations [34], general enterprises shall not resume work before 23:59 on Feb. 9. Meanwhile, according to the electricity consumption statistics of the State Grid Corporation of China, about 20% of enterprises in Zhejiang Province generated electricity consumption on Feb. 10 [35]. By further taking into account the situation of home and remote office, we set Hangzhou's work resumption rate on Feb. 10 to be 10%, and from Feb. 11 until its resumption plan begins, the weekly work resumption rate will be 10%. For Suzhou, we use the same resumption settings as Hangzhou as both are in the Yangtze River Delta Economic Zone. Last but not the least, for Shenzhen, according to the municipal government regulations, general enterprises may not resume work before 24:00 on Feb. 9 [36]. Therefore, we set Shenzhen's weekly work resumption rate to be 10% from Feb. 10 until its resumption plan begins.

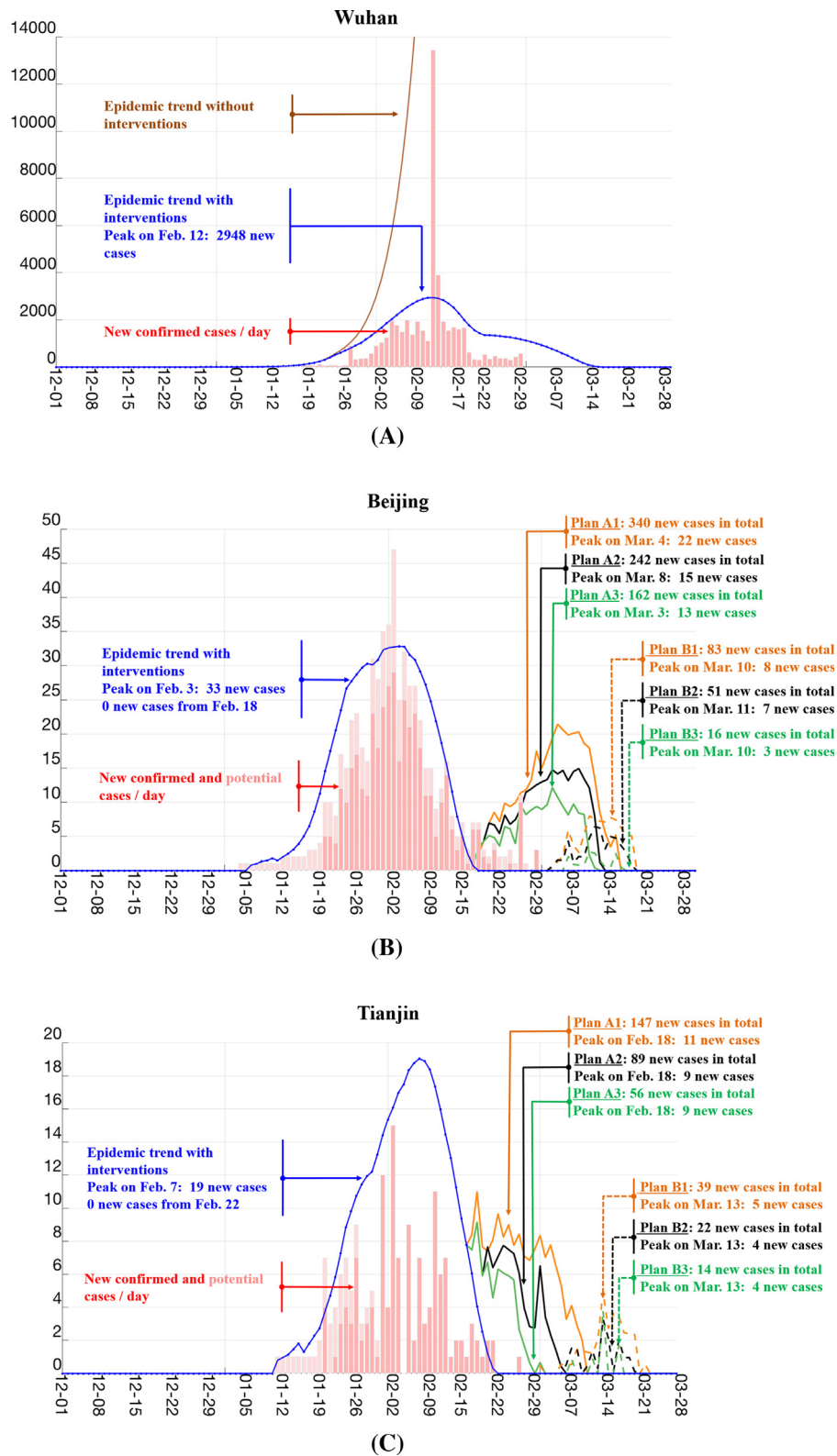


Fig. 3. Estimation on the trends of disease infection and transmission risks associated with different work resumption plans based on the social contact patterns and reported cases. (A): The estimated disease trends without any interventions (the brown line) and with interventions (the blue line) in Wuhan. The newly confirmed cases reported every day are shown in red bars. (B–F): The estimated disease trends with interventions (the blue line) and the transmission risks associated with different work resumption plans in Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen. The dark red bars denote the newly confirmed cases reported every day while the light red bars denote the potential cases locally infected by the imported cases, which are estimated according to the mean serial interval of 7.5 days [6]. Plans A1–A3 refer to the plans that start on Feb. 17 (Monday) and finish the resumption in 1 week, 1/2 months, and 1 month, respectively. Plans B1–B3 refer to the plans that start on Feb. 24 (Monday) and finish the resumption in 1 week, 1/2 months, and 1 month, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

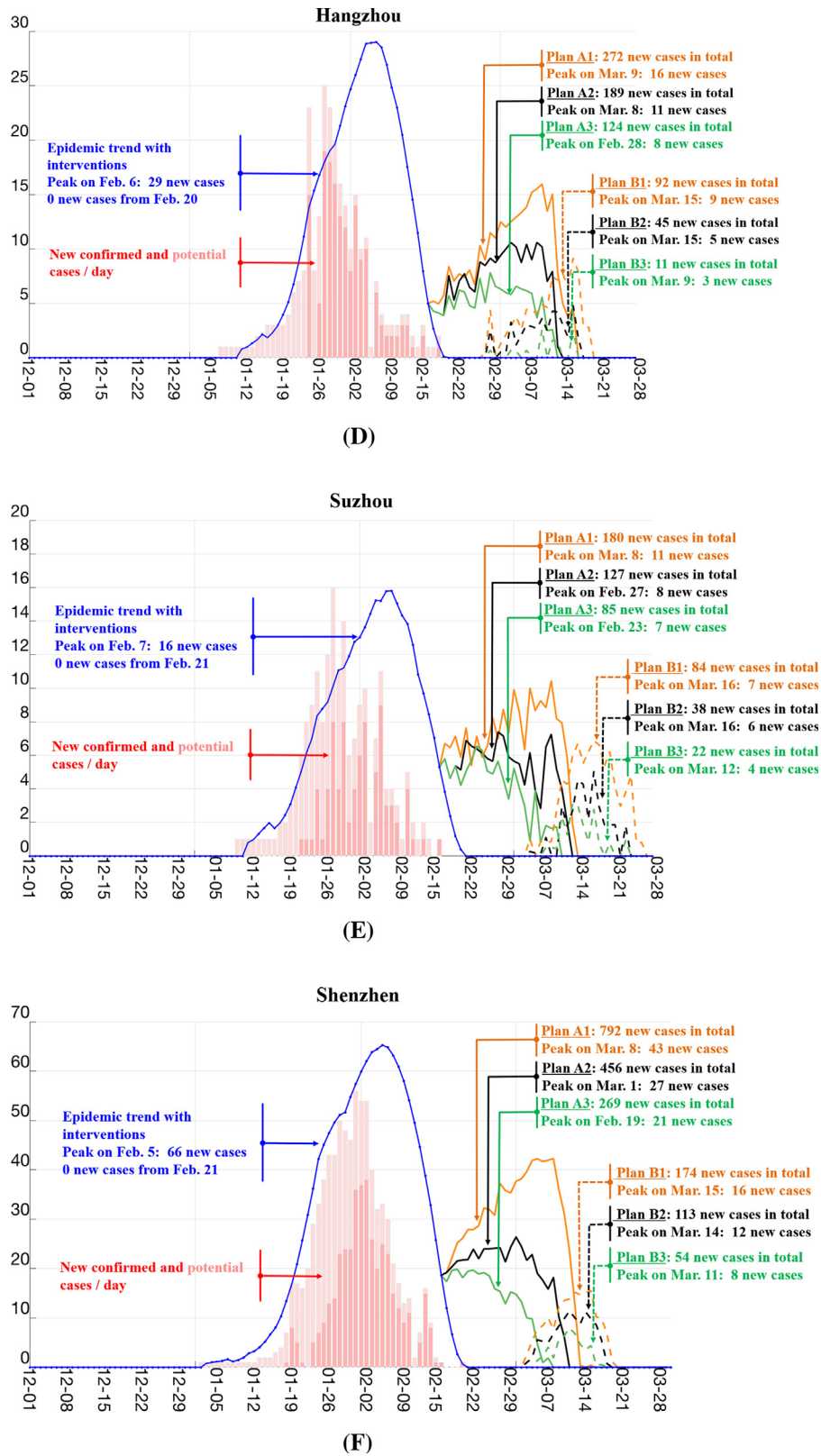


Fig. 3 Continued.

As shown in Fig. 3(B-F), Plans B1–B3 represent a stricter work resumption policy; they start 1 week later than Plans A1–A3. The estimation of disease transmission risks is consistent with actual situations. For example, Beijing implemented a relatively lenient policy on the early resumption of work, and thus has several new cases

reported every day during the past two weeks. This is consistent with our estimated risk trend of Beijing with the plans A1–A3. In contrast, Shenzhen still strictly controlled the resumption of work, so there is no new case reported during Feb. 24–29 [13]. This is also in line with our estimation on Shenzhen’s risk with stricter plans B1-B3.

Table 1

The disease transmission risk (in terms of estimated new cases during the work resumption period) and the estimated Year-over-Year GDP growth (%) in the first half of 2020 ($YoY_{2020} = \frac{\Delta Y_{2020} - \Delta Y_{2019}}{\Delta Y_{2019}} \times 100\%$, with ΔY_{2020} and ΔY_{2019} denoting the estimated GDP growth in the first half of 2020 and the actual GDP growth in the first half of 2019, respectively) of 6 different work resumption plans in 5 cities (Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen). Here ΔY_{2020} of each city is calculated based on ΔY_{2019} and the proportional GDP loss caused by the corresponding percentage of workplace closure in different work resumption plans.

	Beijing		Tianjin		Hangzhou		Suzhou		Shenzhen	
	New cases	YoY ₂₀₂₀	New cases	YoY ₂₀₂₀	New cases	YoY ₂₀₂₀	New cases	YoY ₂₀₂₀	New cases	YoY ₂₀₂₀
Plan A1	340	-9.9%	147	-9.9%	272	-10.1%	180	-10.1%	792	-10.7%
Plan A2	242	-11.7%	89	-11.7%	189	-11.9%	127	-11.9%	456	-12.6%
Plan A3	162	-15.0%	56	-15.0%	124	-15.2%	85	-15.2%	269	-16.4%
Plan B1	83	-12.6%	39	-12.6%	92	-12.8%	84	-12.8%	174	-13.8%
Plan B2	51	-14.1%	22	-14.1%	45	-14.4%	38	-14.4%	113	-15.5%
Plan B3	16	-17.0%	14	-17.0%	11	-17.3%	22	-17.3%	54	-18.8%

Table 1 summarizes the disease transmission risks as well as the estimated Year-over-Year GDP growth (%) in the first half of 2020 with respect to different work resumption plans in Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen. As can be noted, Plan B3 resumes the work as late as possible and completes the resumption as slow as needed, and thus minimizes the disease transmission risk. The plan has the least expected GDP growth. Alternatively, it may also be practically desirable to gradually bring the work back to normal, while keeping all the necessary control measures possible to eliminate any potential disease transmission. In such a case, Plan B1 could be adopted. It achieves risk mitigation and gradual work and life recovery at the same time.

3.4. Sensitivity to parameter variations

We conduct the sensitivity study to examine variations of the analytic results with respect to variations in different age-groups and

various social contact patterns. Specifically, we analyze the sensitivity of the estimated disease trends with respect to changes in the infectivity matrix **A**, the individual susceptibility matrix **B**, and the contact matrix **C**. Note that since both **A** and **B** are diagonal matrices, the impact to the next-generation matrix **K** will be identical if **A** and **B** change in the same scale, we only conduct the sensitivity analysis on **A** as a representative. Each time, we change the diagonal value in **A** for one specific age-group while keeping other age-groups' value fixed. By doing so, we can investigate the impact of different age-groups on our results. For the contact matrix **C**, each time we change the weight for one of the four matrices C^H , C^S , C^W , and C^P , while keeping the weights of the other three unchanged. By doing so, we can observe the impact of different social contact patterns on the results.

Fig. 4 shows variations in the estimated disease trends corresponding to variations in (A) different age-groups and (B) various social contact patterns. The trends in Fig. 4 are measured by the total number of confirmed and potential cases. From Fig. 4(A) we can

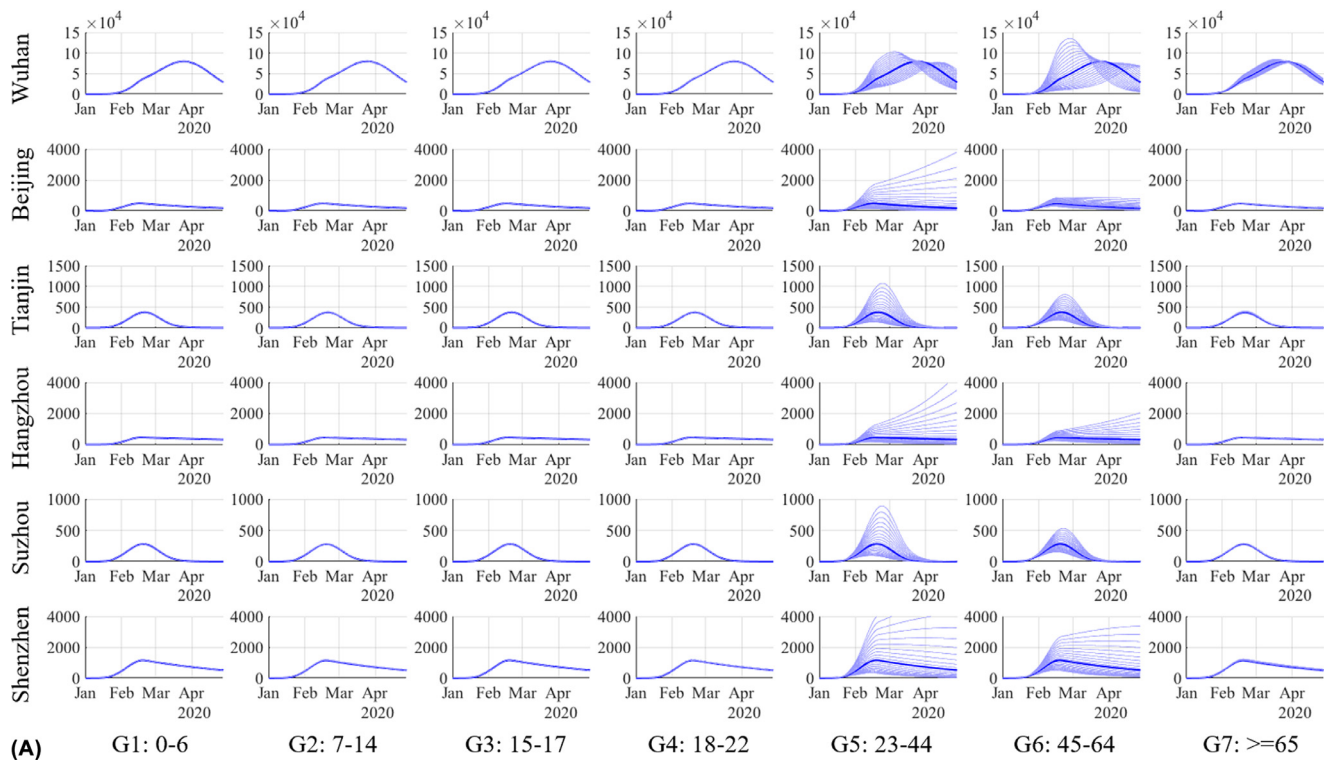


Fig. 4. Variations in the estimated disease trends corresponding to variations in the infectivity matrix **A** and the contact matrix **C**. Here the trends are measured by the total number of confirmed and potential cases. (A): Trend variations with the variations (+/-10%) of a_{ii} ($i = 1, \dots, 7$), the infectivity in age-group G_i , in Wuhan, Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen. Each row represents a city and each column represents an age-group. (B): Trend variations with the variations (+/-10%) of $r_{H, S, W, P}$ in Wuhan, Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen. Each row represents a city and each column represents a social contact setting.

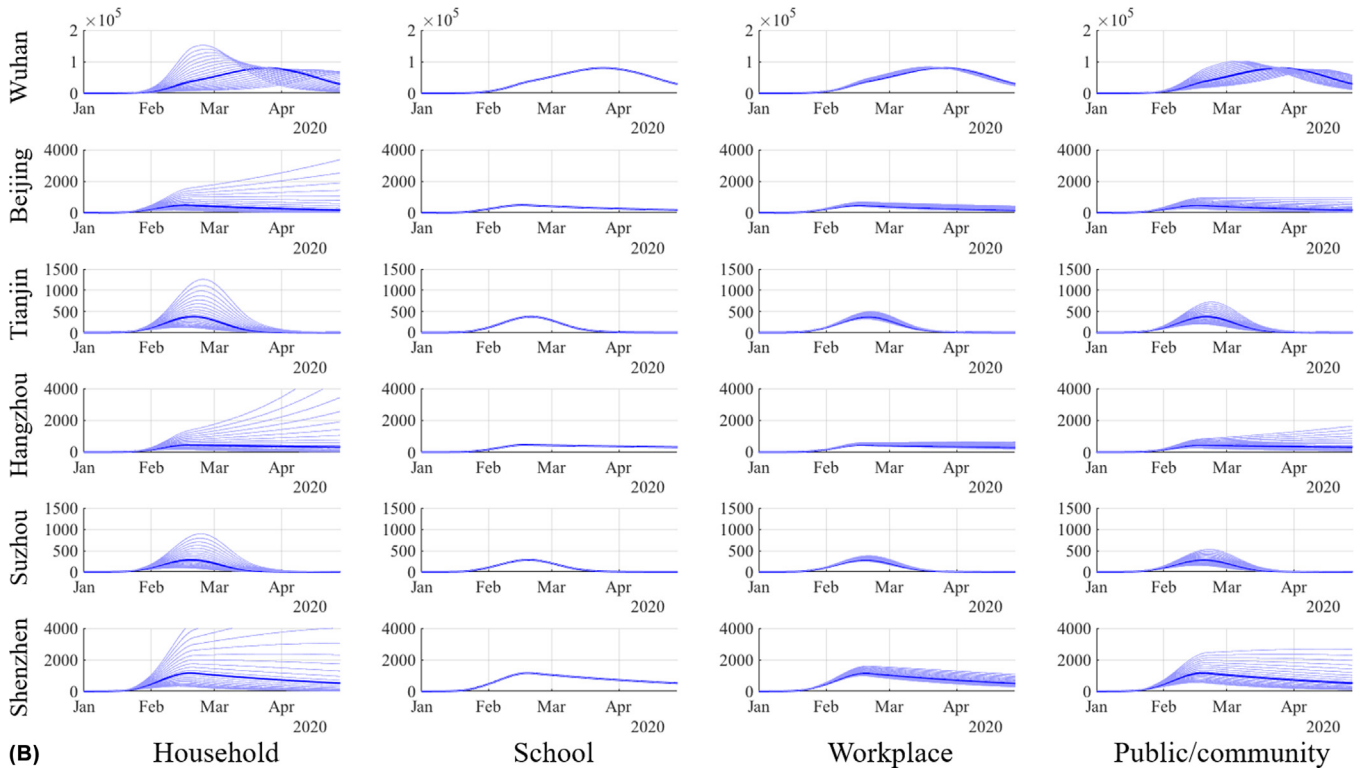


Fig. 4 Continued.

observe that the disease trends in all six cities are relatively sensitive to the variations in the infectivity in G5 (23–44) and G6 (45–64). There are two main reasons. First, both the population size and the case number in these two age-groups are relatively large. Second, people in these two groups are more frequently engaged in social activities than those in other age-groups. Therefore, a slight variation on the infectivity in these two groups might cause relatively large variations in the disease trends. From Fig. 4(B) we can observe that the variations of the disease trends are more obvious in households than in schools, workplaces, and public places/communities, which is consistent with our observations in Fig. 2. The reason is that China has already implemented strong social distancing strategies, such as the closure of schools and workplaces, and thus the variations of social contact intensities in schools and workplaces have relatively limited impact on the disease trends. Note that even various intervention strategies have been deployed, the disease trends can still be relatively sensitive to the variations of social contact intensities in public places and communities. The key implication, as revealed by Fig. 4(B), is to maintain proper social distance in public/community, even when school and business activities are resumed. Moreover, there are some intrinsic consistencies between the results in Fig. 4(A) and (B). Note that the group in Fig. 4(A) with the highest sensitivity is G5, in the age of 23–44. People of this age are generally the family breadwinner and play the most active role in social activities. If they are infected, they will bring a huge risk to their family members in households and their friends and acquaintances in communities. This explains the observation in Fig. 4(A), i.e., the variation of infectivity in G5 brings the highest impact on the disease trends, and thus reminds the public health workers to pay special attention to this group of population when implementing control and prevention strategies in a later stage.

4. Discussion

In this study, we demonstrated the importance of characterizing the underlying transmission patterns among different populations for the

purpose of understanding the COVID-19 outbreak in China from an epidemiological perspective. With an in-depth characterization of the age-specific social contact-based transmission, we conducted the retrospective and prospective analyses of the disease outbreak, including the past and future disease transmission trends, the effectiveness of different interventions, and the disease transmission risks of restoring normal social activities. We focused on 6 representative cities in China; the conclusions drawn from the study not only provide a comprehensive explanation of the underlying COVID-19 transmission patterns in China, but more importantly, offer the contact-based risk analysis methodology that can readily be applied to guide intervention planning and operational responses in other countries, so as to effectively control the COVID-19 pandemic.

In this study, the analysis was conducted on 6 cities in China, including Wuhan, Beijing, Tianjin, Hangzhou, Suzhou, and Shenzhen. On the one hand, the selected cities were representative in terms of the severity of the COVID-19 outbreak in the respective region during the time that the study was conducted and their economic impact in China. On the other hand, they presented different characteristics in several important aspects: First, the categories of confirmed cases in these cities were different. For Wuhan, most of the cases were indigenous cases; while in the other 5 cities, a large portion of the cases was imported cases (from Hubei Province). Second, the distributions of populations in different age-groups were different in these cities. Third, the levels of social distancing interventions and the work resumption plans implemented in different cities were different. The above differences made the scenarios in different cities quite different and intriguing. The retrospective and prospective analyses conducted on these cities show that our results are consistent with the real situations of the corresponding cities, validating the model's generalization ability given different real-world contexts. Importantly, it should be noted that, although the numerical results derived from the 6 cities in this study may not be the same as those in other countries, the developed methods are general at the methodological level and the idea of using age-specific contacts to characterize the disease transmission patterns is instructive in understanding, and hence

planning corresponding interventions in, the situations of the disease outbreaks in those countries. When applying the developed methodology to the wider global population, country/region-specific scenarios and settings, such as case categories, distribution of age-specific population, working environment and hours, and interventions and work resumption plans, should be incorporated to provide better tailor-made parameterization, thus making the retrospective and prospective analyses more situation-specific and informative.

As COVID-19 is a newly emerging infectious disease, we are still in the process of gaining more knowledge and understanding of its transmission patterns. As a result, the parameters estimated based on the current understanding might not be as adequate or precise as those in some of the well-understood diseases, such as seasonal influenza. Therefore, one of our future research directions is to continue investigating the characteristics of the disease, from both epidemiological and computational perspectives, so as to parameterize the model in a more accurate way. Further, in this study, we have modeled the underlying transmission of COVID-19 outbreak by considering the age-specific social contact patterns. It should be pointed out that there also exist other disease-related factors that might affect the disease transmission patterns, such as the cross-region mobility of the population and the environmental factors. We plan to incorporate these disease-related factors into the model, thus making our analysis more comprehensive. Moreover, the current study focuses on the representative cities in China. However, it will be desirable to conduct further analyses on a global scale. In this regard, the general methodology provided in this study can readily be applied, while considering country/region-specific social, demographic, and epidemiological characteristics, such as infection-related social contact patterns [37]. To further generalize and transfer our research, we plan to collaborate with researchers and practitioners around the world to conduct the corresponding analyses for other countries/regions.

Data sharing

The modeling/analytics tools of this study will be made publicly available on: <http://aic.hkbu.ai/>.

Declaration of competing interest

We declare no competing interests.

Acknowledgments

The authors would like to specially thank K. Guo from the Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, for her helpful discussions in estimating the impact of COVID-19 on Chinese GDP. J. Liu acknowledges support from the Hong Kong Research Grant Council (under grants: [RGC/HKBU12202415](#) and [RGC/HKBU12201318](#)). Y. Shi acknowledges support from the National Natural Science Foundation of China (under grants [71932008](#) and [91546201](#)).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.eclinm.2020.100354](https://doi.org/10.1016/j.eclinm.2020.100354).

References

- [1] Hui DS, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health – the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* 2020;91:264–6 February.
- [2] Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html>.
- [3] World Health Organization, *Coronavirus disease 2019 (COVID-19) situation* 69. March 29 (2020). <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200329-sitrep-69-covid-19.pdf>.
- [4] Strategy and Policy Working Group for NCIP Epidemic Response. Chinese center for disease control and prevention, urgent research agenda for the novel coronavirus epidemic: transmission and non-pharmaceutical mitigation strategies. *Chin J Epidemiol* 2020;41(2):135–8 [in Chinese] http://chinaepi.icdc.cn/zhlxbx/ch/reader/create_pdf.aspx?file_no=001&flag=1&journal_id=zhlxbx&year_id=2020.
- [5] World Health Organization, *Report of the WHO-china joint mission on coronavirus disease 2019 (COVID-19)*. February 28 (2020). <https://reliefweb.int/sites/reliefweb.int/files/resources/who-china-joint-mission-on-covid-19-final-report.pdf>
- [6] Li Q, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020;382:1199–207.
- [7] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020 January 31. doi: [10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
- [8] Health Commission of Hubei Province: <http://wjw.hubei.gov.cn/>.
- [9] Beijing Municipal Health Commission: <http://wjw.beijing.gov.cn/>.
- [10] Tianjin Municipal Health Commission: <http://wsjk.tj.gov.cn/>.
- [11] Hangzhou Municipal Health Commission: <http://wsjkw.hangzhou.gov.cn/>.
- [12] Suzhou Municipal Health Commission: <http://wsjkw.suzhou.gov.cn/>.
- [13] Shenzhen Municipal Data Open Platform: https://opendata.sz.gov.cn/data/dataset/toDataDetails/29200_01503668.
- [14] Wuhan Statistical Yearbook 2018: <http://tjj.hubei.gov.cn/tjsj/sjksxc/tjnj/gszjt/whs/>.
- [15] Beijing Municipal Bureau of Statistics: <http://tjj.beijing.gov.cn/>.
- [16] Tianjin Municipal Bureau of Statistics: <http://stats.tj.gov.cn/>.
- [17] Hangzhou Municipal Bureau of Statistics: <http://tjj.hangzhou.gov.cn/>.
- [18] Suzhou Municipal Bureau of Statistics: <http://tjj.suzhou.gov.cn/>.
- [19] Shenzhen Municipal Bureau of Statistics: <http://tjj.sz.gov.cn/>.
- [20] Fumanelli L, Ajelli M, Manfredi P, Vespignani A, Merler S. Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Comput Biol* 2012;8(9):e1002673.
- [21] Xia S, Liu J, Cheung W. Identifying the relative priorities of subpopulations for containing infectious disease spread. *PLoS ONE* 2013;8(6):e65271.
- [22] Mossong J, Hens N, Jit M, Beutels P, Auranen K, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 2008;5(3):e74.
- [23] Iozzi F, Trusiano F, Chinazzi M, Billari FC, Zagheni E. Little Italy: an agent-based approach to the estimation of contact patterns– fitting predicted matrices to serological data. *PLoS Comput Biol* 2010;6(12):e1001021.
- [24] Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. *J R Soc Interface* 2010;7(47):873–85.
- [25] Hethcote HW. The mathematics of infectious diseases. *SIAM Rev* 2000;42(4):599–653.
- [26] Svensson A. A note on generation time in epidemic models. *Math Biosci* 2007;208:300–11.
- [27] Binti Hamzah FA, Lau C, Nazri H, Ligo DV, Lee G, Tan CL. CoronaTracker: world-wide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ* 2020;19 March. doi: [10.2471/BLT.20.255695](https://doi.org/10.2471/BLT.20.255695).
- [28] Liu T, et al. Time-varying transmission dynamics of novel coronavirus pneumonia in China. *bioRxiv* 2020. doi: [10.1101/2020.01.25.919787](https://doi.org/10.1101/2020.01.25.919787).
- [29] Hellewell J, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* 2020;February 28. doi: [10.1016/S2214-109X\(20\)30074-7](https://doi.org/10.1016/S2214-109X(20)30074-7).
- [30] Epidemiology Working Group for NCIP Epidemic Response. Chinese center for disease control and prevention, the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Chin J Epidemiol* 2020;41(2):145–51 [in Chinese] http://chinaepi.icdc.cn/zhlxbx/ch/reader/create_pdf.aspx?file_no=004&flag=1&journal_id=zhlxbx&year_id=2020.
- [31] Sadique MZ, Adams EJ, Edmunds WJ. Estimating the costs of school closure for mitigating an influenza pandemic. *BMC Public Health* 2008;8(135).
- [32] http://www.gov.cn/xinwen/2020-01/31/content_5473425.htm (Accessed on April 1, 2020)
- [33] <http://www.dehenglaw.com/CN/tansuocontent/0008/017738/7.aspx?MID=0902> (Accessed on April 1, 2020)
- [34] http://www.hangzhou.gov.cn/art/2020/2/9/art_1256295_41893739.html (Accessed on April 1, 2020)
- [35] <http://energy.people.com.cn/n1/2020/02/13/c71661-31585079.html> (Accessed on April 1, 2020)
- [36] http://www.sz.gov.cn/szst2010/yqfk2020/szxd/zczy/zcwj/fgzj/content/post_6728851.html (Accessed on April 1, 2020)
- [37] Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, Beutels P, Hens N. A systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology* 2019;30(5):723–36.