

ANALYSIS OF THE AGE-STRUCTURED EPIDEMIOLOGICAL CHARACTERISTICS OF SARS-COV-2 TRANSMISSION IN MAINLAND CHINA: AN AGGREGATED APPROACH

JUNYUAN YANG^{1,2,*}, GUOQIANG WANG^{1,2}, SHUO ZHANG^{1,2}, FEI XU³
AND XUEZHI LI⁴

Abstract. The novel coronavirus (SARS-Cov-2) has raged in mainland China for nearly three months resulting in a huge threat to people's health and economic development. According to the cumulative numbers of confirmed cases and deaths of SARS-COV-2 infection announced by the National Health Commission of China, we divided the human population into four subgroups including the adolescents group (0–19 yr old), the youth group (20–49 yr old), the middle-aged group (50–74 yr old) and the elderly group (over 75 yr old), and proposed a discrete age-structured SEIHRQ SARS-COV-2 transmission model. We utilized contact matrixes to describe the contact heterogeneities and correlations among different age groups. Adopting the Markov chain Monte Carlo (MCMC) algorithm, we identified the parameters of the model and fitted the confirmed cases from January 24th to March 31st. Through a more in-depth study, we showed that before January 28th (95% CI [Feb. 25th, Feb. 31st]), the effective reproduction number was greater than 1 and after that day its value was less than 1. Moreover, we estimated that the peak values of infection were 66 (95% CI [65,67]) for the adolescents, 3996 (95% CI [3957,4036]) for the young group, 14714 (95% CI [14692,14735]) for middle-aged group and 297 (95% CI [295,300]) for elderly people, respectively; the proportions of the final sizes of SARS-COV-2 infection accounted for less than 90% for each group. We found that under the current restricted control strategies, the most severe and high-risk group was middle-aged people aged between 50–74 yr old; without any prevention, the most severe and high-risk group had become the young adults aged 20–49 yr old.

Mathematics Subject Classification. 92D30, 93C15.

Received May 6, 2020. Accepted July 31, 2020.

1. BACKGROUND

The novel coronavirus pneumonia (SARS-COV-2) is a new champions league by linear single strand caused by chain RNA coronavirus pneumonia, which can lead to the Severe Acute Respiratory Syndrome (SARS), the

Keywords and phrases: Age structure, MCMC algorithm, peak value, basic reproduction number.

¹ Complex Systems Research Center, Shanxi University, ShanXi TaiYuan 030006, PR China.

² Shanxi Key Laboratory of Mathematical Techniques and Big Data Analysis on Disease Control and Prevention, Shanxi University, Taiyuan 030006, PR China.

³ Department of Mathematics, Wilfrid Laurier University, University Avenue, Waterloo N2L 3C5, Canada.

⁴ School of Mathematics and Science, Henan Normal University, Xinxiang 453000, PR China.

* Corresponding author: yjyang66@sxu.edu.cn

Middle East Respiratory Syndrome (MERS), and other serious diseases. The virus can be indirectly spread by respiratory droplets and directly transmitted by close contacts and as such it is highly contagious [19]. In December of 2019, SARS-COV-2 sorrowfully crept Wuhan, a center city of China. Since the outbreak coincided with the Chinese New Year, increased mobility and clustering gradually enhanced the spread of infection. Until March 17th, 2020, a cumulative of 80 881 confirmed cases and 3226 deaths had been reported in China [20]. After January 23rd, 2020, a series of control measures such as household quarantine, traffic restrictions and holiday extensions have been implemented across the country. Then, a national public health emergency with level I was initiated to mitigate SARS-COV-2 prevalence [21, 22]. It is urgent to capture the epidemiological characteristics of SARS-COV-2 and to develop effective control measures.

Recently, many scholars have used mathematical models to explore the transmission mechanisms and forecast the trends of SARS-COV-2 propagation [6, 7, 11, 15]. However, most of them assumed that individual contacts are uniformly mixed, and the contact patterns are homogeneous. Some literature [9, 10, 14] have showed that people at different age states at home, at work or in general communication are tied up with contact patterns. However, there is no evidence to support definite connections between SARS-COV-2 infection and individual ages. The data reported by the Chinese Center for Disease Control and Prevention [16] showed that SARS-COV-2 infection has large age spans, and the young adults and middle-aged individuals have high risk of infection. Though young and aged individuals generally have weak immune systems, individuals in these groups are not high-risk susceptibles. Among them, the proportion of infected individuals aged between 30–79 yr old is 86.6%, only 2.1% of confirmed cases were individuals aged between 0–19 yr old, and the proportion of individuals aged over 80 accounts for only less than 5% [12, 16]. As we know, mathematical models can effectively reveal the characteristics of the age distribution of SARS-COV-2 propagation, formulate control measures and optimally allocate medical resources.

The main contributions of this paper are as follows: First, we proposed an age-structured SEIHRQ model to describe age-related contact heterogeneities of SARS-COV-2 in mainland China. Second, we used the model to mimic the cumulative confirmed cases and cumulative deaths, calibrated the efficacy of the proposed model and recognized the high-risk susceptibles. Furthermore, we revealed the dynamics of SARS-COV-2 infection in different age structures by analyzing the effects of measures such as the city closure and home isolation on the spread of SARS-COV-2 in different age groups, and evaluating the effects of strict control policies.

2. METHODS

2.1. Data description and collection

All the daily cumulative documented cases on SARS-COV-2 infections from the National Health Commission of People’s Republic of China were collected for the modelling study and data fitting [20]. Since the diagnosed method has been changed and the clinically diagnosed cases had been counted in the cumulative documented cases on February 12th, 2020, the period of data for the epidemic curve fitting were collected from January 24th to March 31st [17]. Additionally, to obtain relatively reliable data, the sudden increment on February 12th was allocated to each day in the preceding week in proportion to the original increments of documented cases by day in these days.

2.2. Age-structured model

In order to construct the SARS-COV-2 transmission model, we divided the population into four age groups: 0–19 yr old, 20–49 yr old, 50–74 yr old, and 75 yr old and over, which characterize the spread heterogeneity of SARS-COV-2 in adolescents, young adults, middle-aged adults and elderly individuals. Let N_k ($k = 1, 2, 3, 4$) represent the total populations of 0–19 yr old, 20–49 yr old, 50–74 yr old, elderly people over 75 yr old, respectively. According to a generalized SEIR modeling approach, we divided the total population $N_k(t)$ in age group k into six different statuses: susceptible class, latent class, infectious class, hospitalized class, recovery class, and unsusceptible class, represented by $S_k(t)$, $E_k(t)$, $I_k(t)$, $H_k(t)$, $R_k(t)$, and $S_k^Q(t)$, respectively. The susceptible individuals in age group k contact c_{kj} times with the infected individuals in age group j . Then the

TABLE 1. Parameter definition and value.

Parameters	Definition	Value	Standard deviation	Source
β_1	Transmission rate in adolescents	0.1061	0.00575	MCMC
β_2	Transmission rate in young adults	0.25	0.00572	MCMC
β_3	Transmission rate in middle age	0.32	0.0579	MCMC
β_4	Transmission rate in the elder age	0.1797	0.00573	MCMC
m	Detection rate	0.2	0.00578	MCMC
$1/q$	Latent period	5.2	-	[5]
$1/\gamma$	Course of the disease	7	-	[5, 16]
$1/\gamma_1$	Course of the disease	7	-	[16, 22]
α	The transfer rate from susceptibles to unsusceptibles	0.5	0.0788	MCMC
α_1	The transfer rate from unsusceptibles to susceptibles	0.15	0.00585	MCMC

average number of susceptibles S_k in age group k contacting the infected individuals in age group j is $c_{kj} \frac{I_j}{N_j}$, and the total contact number is $\sum_{j=1}^4 c_{kj} \frac{I_j}{N_j}$. Susceptible individuals in age group k infected by the total infected individuals in each age group j become latent individuals E_k at transmission probability β_k . A proportion of susceptible individuals may choose to self-quarantine (with rate α) to reduce the risk of infection. Unsusceptible individuals may drop off their awareness of self-protection and become susceptible individuals at rate α_1 . Latent individuals transferred to infected individuals after a $1/q$ -day incubation period. Infected individuals comprising of symptomatic individuals and asymptomatic individuals converted to recovered individuals R_k after $1/\gamma$ -day treatment or were quarantined in hospitals at rate k . The hospitalized individuals will either recover or die at rates γ_1 and d , respectively. The hospitalized individuals were absolutely quarantined, which completely cut off the transmission routes from susceptibles. The transformation between the above six states of SARS-COV-2 infection can be expressed by the following differential equations:

$$\left\{ \begin{array}{l} \frac{dS_k(t)}{dt} = -\beta_k S_k(t) \sum_{j=1}^4 c_{kj} \frac{I_j(t)}{N_j(t)} - \alpha S_k + \alpha_1 S_k^Q(t), \\ \frac{dE_k(t)}{dt} = \beta_k S_k(t) \sum_{j=1}^4 c_{kj} \frac{I_j(t)}{N_j(t)} - q E_k(t), \\ \frac{dI_k(t)}{dt} = q E_k(t) - (\gamma + m) I_k(t), \\ \frac{dH_k(t)}{dt} = m I_k(t) - (\gamma_1 + d) H_k(t), \\ \frac{dR_k(t)}{dt} = \gamma I_k(t) + \gamma_1 H_k(t), \\ \frac{dS_k^Q(t)}{dt} = \alpha S_k(t) - \alpha_1 S_k^Q(t), \end{array} \right. \quad k = 1, 2, 3, 4. \quad (2.1)$$

The biological significance and parameter values of model (2.1) were given in Table 1. Table 2 gave the initial values of model (2.1). The basic reproduction number, the average number of secondary cases produced by one infected individual during its infectious period in all complete susceptibles, is an important quantity to determine emerging diseases prevalence. Adopting the approach in [4, 13], we calculate the basic reproduction number of SARS-COV-2 with age structure as

$$\mathcal{R}_0 = \rho(FV^{-1}) \quad (2.2)$$

TABLE 2. Initial value of age stratification.

Initial value	0–19	20–49	50–74	75+	Source
Susceptible population(unit: 1000)	338081	639558	404420	48282	[18]
Latent population	38	3720	6912	184	MCMC
Standard deviation	0.0581	0.0547	0.0579	0.0578	MCMC
Infectious population	19	600	6720	92	MCMC
Standard deviation	0.058	0.0579	0.0057	0.0058	MCMC
Hospitalized population	19	100	960	92	[16, 22]
Unsusceptible population(unit: 1000)	16540	319779	202210	24141	$\alpha \times S_k$

TABLE 3. Total population of each age group (unit: 1000).

Age groups	$N_1(0-19)$	$N_2(20-49)$	$N_3(50-74)$	$N_4(75+)$
Population	338081	639558	404420	48282

with

$$F = \begin{pmatrix} \beta_1 c_{11} & \beta_1 \frac{c_{12} N_1}{N_2} & \beta_1 \frac{c_{13} N_1}{N_3} & \beta_1 \frac{c_{14} N_1}{N_4} \\ \mathbf{0} & \beta_2 \frac{c_{21} N_2}{N_1} & \beta_2 c_{22} & \beta_2 \frac{c_{24} N_2}{N_4} \\ \beta_3 \frac{c_{31} N_3}{N_1} & \beta_3 \frac{c_{32} N_3}{N_2} & \beta_3 c_{33} & \beta_3 \frac{c_{34} N_3}{N_4} \\ \mathbf{0} & \beta_4 \frac{c_{41} N_4}{N_1} & \beta_4 \frac{c_{42} N_4}{N_2} & \beta_4 c_{44} \\ \mathbf{0} & & \mathbf{0} & \end{pmatrix}, \quad (2.3)$$

and

$$V = \begin{pmatrix} qI_d & \mathbf{0} \\ -qI_d & (\gamma + m)I_d \end{pmatrix}, \quad (2.4)$$

where ρ is the spectral radius of matrix FV^{-1} , I_d represents a 4×4 identity matrix, and 0 represents a zero matrix with 4×4 elements. The numbers of N_k were taken from Table 3.

Generally, \mathcal{R}_0 is to estimate the severity of the onset of an outbreak. As the times evolves, the rates of susceptibles and infectives decreases as the rate of removed and quarantined people increase. However, the effective reproduction number, \mathcal{R}_e , is the average number of secondary infections produced by a newly infected individual at time t . For model (2.1), $\mathcal{R}_e = \rho(\bar{F}(t)V^{-1})$, where

$$\bar{F}(t) = \begin{pmatrix} \beta_1 \frac{c_{11} S_1}{N_1} & \beta_1 \frac{c_{12} S_1}{N_2} & \beta_1 \frac{c_{13} S_1}{N_3} & \beta_1 \frac{c_{14} S_1}{N_4} \\ \mathbf{0} & \beta_2 \frac{c_{21} S_2}{N_1} & \beta_2 \frac{c_{22} S_2}{N_2} & \beta_2 \frac{c_{24} S_2}{N_4} \\ \beta_3 \frac{c_{31} S_3}{N_1} & \beta_3 \frac{c_{32} S_3}{N_2} & \beta_3 \frac{c_{33} S_3}{N_3} & \beta_3 \frac{c_{34} S_3}{N_4} \\ \beta_4 \frac{c_{41} S_4}{N_1} & \beta_4 \frac{c_{42} S_4}{N_2} & \beta_4 \frac{c_{43} S_4}{N_3} & \beta_4 \frac{c_{44} S_4}{N_4} \\ \mathbf{0} & & \mathbf{0} & \end{pmatrix}. \quad (2.5)$$

Apparently, \mathcal{R}_e evolves over time, which has the property: if $\mathcal{R}_e < 1$, SARS-Cov-2 pandemic is self-limiting and said to be under control due to containment strategies. Undoubtedly, it is an important quantity to be qualified the effects of control measures.

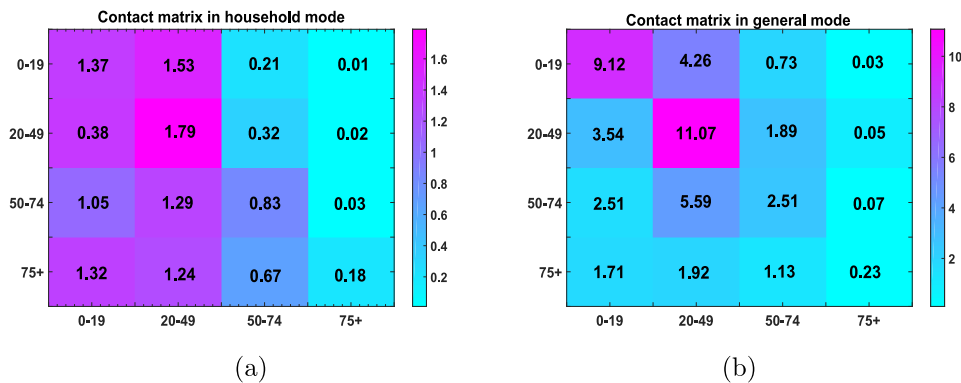


FIGURE 1. (a) The contact matrix of household mode. (b) The contact matrix of general mode.

2.3. Parameter estimation

According to the Chinese Population Statistic Yearbook [18], the numbers of population at each age group were listed in Table 2. Based on the introduction of SARS-COV-2 infection, the incubation period of SARS-COV-2 varies from 4 to 7 days. In this work, we chose an average incubation period of 5.2 days [5]. Moreover, the duration of SARS-COV-2 infection varies from 5 to 10 days and as such we assumed that the average duration of SARS-COV-2 propagation is 7 days.

As for age-structured epidemic models, the contact matrix plays a vital role in the description of age heterogeneities. However, the acquisition for such matrices were a little bit complected. Fortunately, we successfully caught such contact matrices in the household mode and in the general mode of 152 countries in [9]. Thanks to [9], we readily acquired a high-dimensional contact matrix of 16×16 elements with an interval of five years in length. In this paper, we mainly focus on the epidemiological characteristics of adolescents, young adults, middle aged and the elderly age groups of SARS-COV-2 transmission. Hence, we needed to reduce the high-dimensional contact matrix with 16×16 elements to a low-dimensional contact matrix with 4×4 elements by an appropriate approach. Indeed, the above objective can be achieved by the method in the appendix of [8]. The calculation details were enclosed in the Appendix B. By a manipulation, we obtained the associated contact matrices between different age groups in the household mode and the general mode in mainland China (see Fig. 1). It was easy to see that there were some significant differences under different exposure patterns. In the household mode, parents were more likely to interact with their children than contact with colleagues. In the general mode, people of their peers had more contact chances than those with other age groups. Since a series of strict control measures such as city closure, household quarantine and traffic restrictions came into effect on January 23 in mainland China, the contact pattern was approximated to the contact pattern in the household mode introduced before.

2.4. Simulation methods

Matlab 2017b (version 9.3.0.713579) developed by MathWorks was adopted for fitting data and statistic analysis. To fit the diagnosed cases and the solution of model (2.1), we combined Markov chain Monte Carlo (MCMC) numerical method and an adaptive Metropolis-Hastings (M-H) algorithm. First, we noted that the cumulative documented cases associated with the expression of model (2.1) were in form of

$$\bar{C}_I(T) = \sum_{t=1}^T \sum_{l=1}^4 mI_l(t), \quad T = 1, 2, \dots, 68, \quad (2.6)$$

and

$$\bar{C}_D(T) = \sum_{t=1}^T \sum_{l=1}^4 dH_l(t), \quad T = 1, 2, \dots, 68, \quad (2.7)$$

with some random perturbation errors

$$C_I(t) = \bar{C}_I(t) + \epsilon_I, \quad t = 1, 2, \dots, 68, \quad (2.8)$$

and

$$C_D(t) = \bar{C}_D(t) + \epsilon_D, \quad t = 1, 2, \dots, 68. \quad (2.9)$$

Let $C_j(1) = \bar{C}_j(1)$ and ϵ_j be assumed to follow the normal distribution $N(0, \sigma_j^2)$ ($j = I, D$) with $\sigma_I = 1000$ and $\sigma_D = 100$ which make the identified parameters located in a feasible domain. Denoting the parameter vector Q , we defined the combined maximum likelihood function associated with each of the independent cumulative documented data set $C = \{C_I(t), C_D(t)\}_{t=0}^{67}$ as follows

$$L(C|Q) = \prod_{t=1}^{67} \frac{1}{2\pi\sigma_I\sigma_D} \exp\left(-\frac{[C_I(t) - \bar{C}_I(t)]^2}{\sigma_I^2}\right) \exp\left(-\frac{[C_D(t) - \bar{C}_D(t)]^2}{\sigma_D^2}\right). \quad (2.10)$$

Then, we derived that the conditional posterior distribution of the parameters satisfies $P(Q|C) \propto L(C|Q)P(Q)$ and $P(Q|C) \propto L(C|Q)$ if we selected a symmetric proposal distribution in the Metropolis-Hastings Algorithm [2]. Furthermore, we adopted the random walk with the probability

$$\min\left\{1, \frac{P(\bar{Q}|C)}{P(Q|C)}\right\} \quad (2.11)$$

to update the identified parameters \bar{Q} .

We estimated the model parameters by maximizing the likelihood function defined by (2.10) as well as comparing the model quantities defined in (2.6) and (2.7). The MCMC numerical method with M-H algorithm was run for 50 000 iterations with a burn-in of the first 20 000 iterations. The 95% confidence intervals (95% CI) of estimates were obtained by conducting 1000 simulation samples and calculating the median and 95% CI.

3. RESULTS

3.1. Simulation results

Since the Chinese government has enforced a series of containment strategies, the contact patterns in different modes were substantially changed. If any strategy has not been implemented by the Chinese government, any of the Chinese may work, stay at home, go to school or go out leisure activities and in this case, the mode was defined as a general mode, which consists of work mode, household mode, school mode, and leisure mode. If a series of containment measures has been enacted, almost everyone has been quarantined at home and in this case, the mode was defined as the household mode.

According to the cumulative number of confirmed cases reported by the Health Committee of China from January 24th to March 31th [20], the fitting process was achieved by using the MCMC numerical method. According to the MCMC method with MH algorithm, the basic reproduction number was evaluated to be 1.6811 (95% CI [1.6782, 1.6840]). Figure 2 showed that the model was a good fit for the cumulative number of SARS-COV-2 documented cases and confirmed deaths with the 95% confidence intervals. To calibrate the rationality of model (2.1), we picked up cumulative reported cases and cumulative deaths from April 1st to 13th

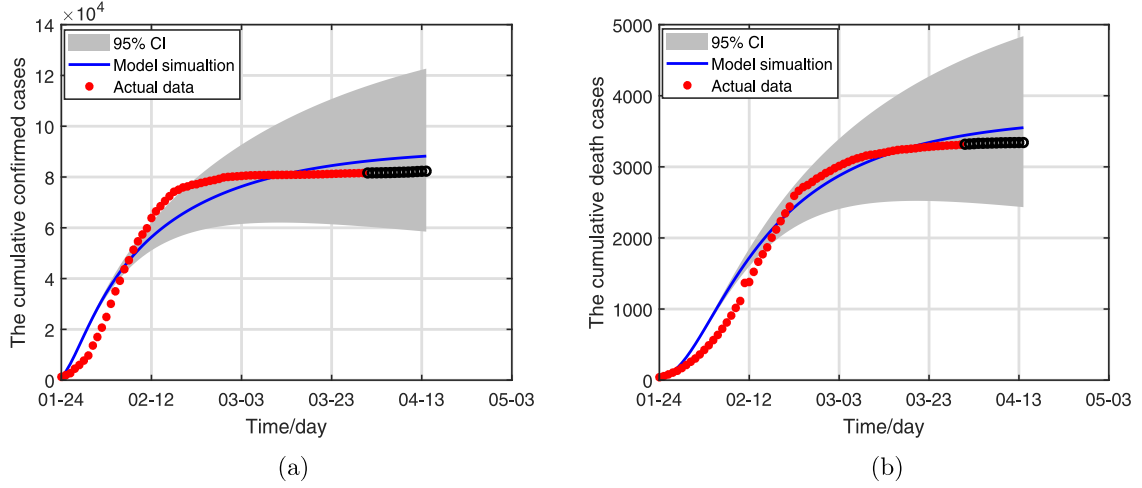


FIGURE 2. (a) The fitting results of the confirmed cases and the solution of model (2.1). (b) The fitting results of the deaths and the solution of model (2.1). The red dots denote the cumulative documented cases and deaths of SARS-COV-2; the blue line denotes the solution of model (2.1); the grey areas represent the 95% confidence interval (CI); the black circles represent the cumulative confirmed cases and deaths from April 1st to 13th.

TABLE 4. Biological biomass of the SARS-COV-2 in different age groups under different modes.

Biomasses	0-19 Y/O		20-49 Y/O		50-74 Y/O		75+ Y/O	
	Household	General	Household	General	Household	General	Household	General
Peak value	66	3854800	3996	100960000	14714	41605000	297	3612600
95% CI	[65,67]	[3553400,3606400]	[3957,4036]	[100520000,101410000]	[14692,14735]	[55122000,55713000]	[295,300]	[3582900,3642200]
Peak time	Feb.14th	Apr.14	Feb.13th	Apr.5th	Feb.3rd	Apr.7th	Feb.10th	Apr.12th
95% CI	Feb.[13th,15th]	Apr.[13th,15th]	Feb.[11th,15th]	Apr.[3rd,7th]	Feb.[1st,5th]	Apr.[5th,9th]	Feb.[9th,11th]	Apr.[11th,13th]
Duration	84 days	207 days	140 days	243 days	138 days	242 days	107 days	231 days
95% CI	[77,91]	[184,230]	[137,142]	[240,245]	[135,140]	[240,244]	[101,112]	[228,234]
Final size	460	20778000	24698	517860000	59287	299210000	1815	20769000
95% CI	[399,521]	[20678000,20868000]	[24509,24888]	[507860000,50875000]	[59091,59482]	[289210000,309210000]	[1751,1878]	[19689000,21689000]

to take a comparison with the solution of model (2.1). From Figure 2, we found that the validating cases were closely consistent with the overall trend of the model and they were allocated in the 95% confidence intervals of the solution of model (2.1), which indicated that model (2.1) captured the epidemic trend of SARS-COV-2 fairly well in mainland China. Other than that, further information was available by evaluating model (2.1). First, the number of SARS-COV-2 patients arrived at a peak value of 18072 (95% CI [17950,18194]) on Feb, 5th (95% CI [Feb.4, Feb.6]); Second, we obtained the similar biomasses including the peak values and the peak arrival times of SARS-COV-2 infection at different age stages by studying model (2.1) (see Tab. 4 and Fig. 3). Finally, as shown in Figure 4, in the household mode, the number of confirmed patients with SARS-COV-2 in the 50–74 yr old segment accounted for 67.44%; the proportion of confirmed illnesses in 0–19 yr old group was only less than 1%; elderly patients of 75 yr old and over accounted for nearly 3%.

3.2. Sensitivity analysis

The basic reproduction number plays a key role in determining the prevalence of a disease. In general, if the basic reproduction number is less than 1, the disease is eradicated; otherwise, the disease erupts. In order to calculate the basic reproduction number of SARS-COV-2 infection in mainland China, we employed the identified parameters in the previous section (see Tabs. 1 and 3) and formula (2.2). As shown in Figure 5a, the effective

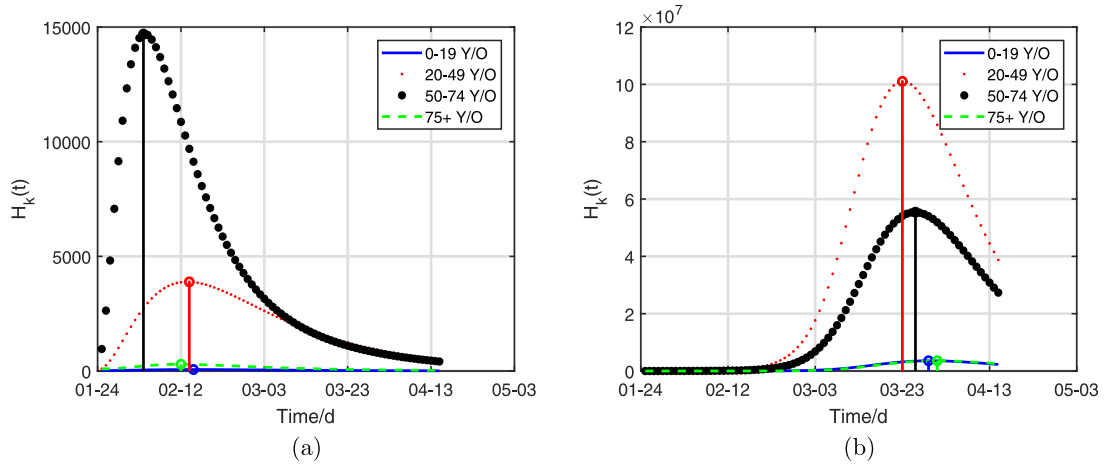


FIGURE 3. (a) The trends of SARS-COV-2 infection of different age groups in the household mode. (b) The trends of SARS-COV-2 infection of different age groups in the general mode. The blue solid lines denote confirmed cases aged between 0 and 19. The red dot lines denote the confirmed cases aged between 20 and 49. The black dot lines denote the diagnosed cases aged between 50 and 74. The green dot lines represent the documented cases over the age of 75.

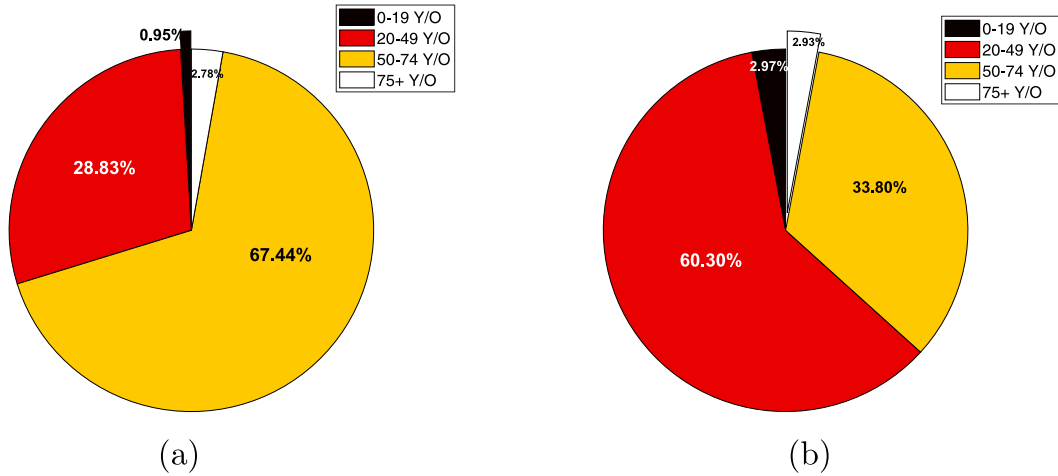


FIGURE 4. (a) The pie chart of the percentage of the epidemic final size in the household mode. (b) The pie chart of the percentage of the epidemic final size in the general mode.

reproduction number was a monotone decreasing function associated with time t , and the effective reproduction number of SARS-COV-2 infection in mainland China on April 13th is about 0.8298. This manifested that the number of people infected by SARS-COV-2 is gradually decreasing, and the situation took a favorable turn. In addition, from Table 4, it can be found that the final size at each age group in household mode has been significantly lowered than that value in the general case, which indicated that the current measures such as city closure and home isolation are very effective.

Figure 3 and Table 4 showed that when control measures are in place, the peak of the epidemic will arrive earlier. Of course, the peak value of the epidemic with control was much smaller than that without control measures. Without strict control measures, infection may increase by a factor of 50 thousands in population

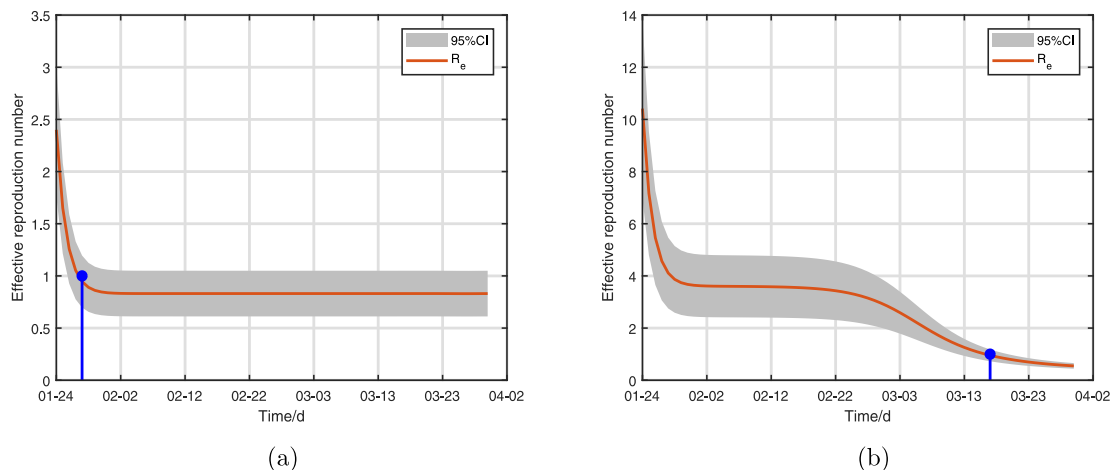


FIGURE 5. (a) The effective reproduction number in the household mode of SARS-COV-2 with 95% CI. (b) The effective reproduction number in the general mode of SARS-COV-2 with 95% CI. The solid line denotes the average effective reproduction number; the grey area denotes 95% CI.

between the ages of 0 and 19, the peak value may increase by a factor of 25 thousands in population between the ages of 20 and 49, the peak value may increase by a factor of 2 thousands in population between the ages of 50 and 57, and the risk may increase by a factor of 12 thousands in population at the ages of 75 or older. These control measures advanced the peaks of the infection. For example, for population group aged 0–19, the peak was 60 days ahead. For population group aged 20–49, the peak was 50 days ahead. For population group aged 50–74, the peak was 64 days ahead, and for population group aged 75 or older, the peak was 62 days ahead. In the absence of control measures, people in the 20–49 and 50–74 age groups were most vulnerable. Under the current control strategies, individuals in the 50–74 age group had the greatest risk of infection compared to individuals in other age groups.

4. DISCUSSION

In this paper, we established an age-structured SARS-COV-2 infection model based on the compartmental theory. The age-structured heterogeneity has been addressed by a certain contact matrix obtained from [9]. Actually, such an approach not only reduces the numbers of identified parameters, but also describes the correlations between different age groups. Although we had few data information in each age group about SARS-COV-2 infection in China, we successfully mimicked the cumulative confirmed cases with the solution of model $\sum_{l=1}^4 mI_l(t)$ and used the model with identified parameters to explore the dynamics of SARS-COV-2 infection in each age group. Our investigation showed that the middle-aged group in the household mode and the young group in the general mode have the highest-risk of infection, respectively.

The basic reproduction number was estimated to be $\mathcal{R}_0 = 1.6811$ with 95% confidence intervals from 1.6782 to 1.6840. This means that on average each infectious individual had infected 1.6811 susceptibles. Generally, the disease will increase exponentially if \mathcal{R}_0 was greater than one, whereas it can be eradicated if \mathcal{R}_0 was less than one under certain control measures. Indeed, although \mathcal{R}_0 of SARS was estimated to be around 3 [1], the outbreaks of SARS were successfully controlled by isolating infectious individuals and some carefully implemented control strategies. Facing SARS-COV-2 infection, the Chinese government has adopted the familiar control strategies as the prevention experiences as SARS in 2003. The effects of household isolation and traffic restrictions have been become more and more obvious over time. On January 28th, the value of the effective reproduction number \mathcal{R}_e was about 1 followed by a decrease in \mathcal{R}_0 , while that value of \mathcal{R}_e tended to 1 till

March 16th, which implied that such policies are effective in slowing down SARS-COV-2 infection in mainland China. It is worth noting that the underlying reproduction number was smaller than those values based on the earlier cumulative data, but it was well in accordance with the value of 2.2 (95% CI [1.4,3.9]) by the statistic analysis [3, 5].

Table 4 and Figure 4 reflected that under the current strategies, the duration of SARS-COV-2 infection in adolescents has been shrunk almost four months, the length of SARS-COV-2 infection for young group has been reduced almost three months, the duration for middle-aged group has been shrunk 100 days and that value for elderly people have been reduced 120 days. Meanwhile, under the current measures, the final sizes of SARS-COV-2 infection have been reduced 90% for each group compared to those values with on control strategy. Our comparative results suggested that these control measures were more effective in populations aged 0–19 yr because their risk of contact with infectives was significantly reduced compared with other age groups.

Although the social distancing and age structured epidemiological characteristics of SARS-COV-2 have been addressed in this project, the further precise information about each age group and region heterogeneities of our estimate was not well-fitted because of lacking of time series diagnosed cases in each age group. The absolute accuracy of the estimated outcomes should be validated by those epidemiological information in detail in the future. Some evidence showed that SARS-COV-2 infection has some certain link with climate and how these aspects affect the SARS-Cov-2 transmission has not been addressed. Additionally, other factors, including the improvements of medical sources, detection methods, antiviral treatments *etc.*, will be introduced to study the potential dynamics of SARS-COV-2 infection. Individual-level heterogeneities, considering nosocomial infection and clustering infection should also be addressed by some individual-level or individual-agent models in future study.

5. CONCLUSION

By analyzing and fitting the cumulative documented data of SARS-COV-2 infection in mainland China from January 24th to April 13th, 2020, an aged-structured SEIHRQ model was constructed by dividing the population into four groups: adolescents group (0–19 yr old), young adults group (20–49 yr old), middle-aged group (50–74 yr old) and elderly aged group (75 yr old and over). The parameters and initial values of the model were identified by the MCMC algorithm, the effective basic reproduction number of the model was calculated, and the age-related characteristics of the SARS-COV-2 infection under different policies were analyzed. Through careful analysis, we found that the young group and middle-aged group have high-risk of infection. Under the policy of the household isolation and traffic restrictions, the middle-aged individuals have higher risk of infection than other age groups. Without any measure, the young group had the highest risk of being infected with SARS-COV-2.

The Chinese government has implemented a series of strict control measures on January 23rd. Those policies, aiming to ensure early detection, early quarantine, and early treatment, have updated over time. That accounted for the uptake value of the effective reproduction number of \mathcal{R}_e . Figure 5a showed that the effective reproduction number of SARS-COV-2 infection in mainland China on January 28th (95% CI [Jan.25th, Jan.31st]) reached 1, followed by a continuous decrease. That is to say, the effective reproduction number was less than 1 after January 28th, 2020. The above analysis showed that non-pharmaceutical interventions were in favor of reducing the magnitude of the epidemic peaks, delaying the arrival time of the peak, and thus combating SARS-COV-2 prevalence in mainland China. Consequently, lowering the peak size and delaying the arrival time of the peak greatly alleviate the acute pressure on the health-care system.

Acknowledgements. The authors would like to thank editors's and referees's valuable comments and suggestions for improving the qualitative of this paper. This work was supported by the National Natural Science Foundation of China (Nos. 61573016, 11771017) and the New Coronavirus (COVID-19) Prevention and Control Project of Shanxi Province (No. 202003D31011 / GZ).

APPENDIX A. ORIGINAL CONTACT MATRICES

For a discrete age-structured SARS-COV-2 transmission model (2.1), the contact matrix plays an important role in the description of age heterogeneities and correlations. As shown in Section 2.3, we directly acquired the two contact matrices with 16×16 elements in the household and general modes from [9], listed in Tables A.1 and A.2. Due to lack of detailed information on SARS-COV-2 infection in mainland China, we just divided the population into four age groups (0-19, 20-49, 50-74, ≥ 75 yr old). We were primarily concerned with the risk of contact and possible onward SARS-COV-2 spread among the adolescents, young group, middle-aged group and elderly people. Thus, we need to adopt an appropriate approach to aggregate the matrices A.1 and A.2 into two 4×4 matrices. Following the steps proposed by Meltzer *et al.* [8], we gave the detailed calculation processes as follows.

TABLE A.1. Contact matrix in the general mode.

1.41	0.69	0.36	0.24	0.48	0.74	0.88	0.84	0.51	0.29	0.25	0.25	0.16	0.10	0.06	0.03
0.56	6.22	0.98	0.31	0.22	0.51	0.82	0.93	0.81	0.36	0.19	0.20	0.14	0.09	0.04	0.03
0.23	1.74	7.65	0.85	0.38	0.31	0.46	0.78	0.98	0.53	0.23	0.15	0.08	0.07	0.05	0.04
0.15	0.40	3.07	11.84	1.76	0.79	0.59	0.91	1.15	1.06	0.47	0.22	0.09	0.06	0.03	0.02
0.35	0.28	0.38	3.19	5.99	2.29	1.36	1.24	1.16	1.39	0.85	0.49	0.15	0.06	0.06	0.05
0.61	0.36	0.21	0.91	2.56	3.72	1.89	1.54	1.31	1.10	0.96	0.56	0.20	0.06	0.03	0.02
0.66	1.00	0.71	0.51	1.19	1.85	2.76	1.93	1.48	1.11	0.77	0.61	0.25	0.09	0.04	0.04
0.75	1.28	1.10	0.85	0.90	1.52	1.86	3.25	2.23	1.37	0.87	0.56	0.32	0.18	0.10	0.03
0.51	1.02	1.32	1.56	1.18	1.35	1.70	2.08	3.19	1.78	1.09	0.46	0.28	0.15	0.09	0.05
0.31	0.78	0.90	1.95	1.17	1.14	1.26	1.54	1.70	2.22	1.10	0.55	0.20	0.10	0.08	0.08
0.28	0.61	0.84	1.17	1.02	1.28	1.12	1.14	1.53	1.62	1.49	0.78	0.24	0.08	0.05	0.06
0.65	1.00	0.80	1.07	0.93	1.41	1.43	1.16	1.29	1.06	1.13	1.44	0.50	0.19	0.09	0.07
0.57	0.58	0.40	0.57	0.52	0.72	0.84	0.95	0.76	0.61	0.48	0.67	0.77	0.30	0.14	0.05
0.34	0.53	0.44	0.24	0.31	0.39	0.57	0.67	0.59	0.33	0.26	0.38	0.37	0.61	0.16	0.06
0.16	0.47	0.43	0.40	0.19	0.30	0.28	0.57	0.62	0.48	0.28	0.27	0.37	0.33	0.47	0.16
0.28	0.39	0.57	0.48	0.18	0.18	0.25	0.39	0.44	0.47	0.39	0.24	0.14	0.20	0.16	0.23

TABLE A.2. Contact matrix in the household mode.

0.50	0.39	0.22	0.11	0.21	0.38	0.46	0.42	0.21	0.07	0.08	0.06	0.04	0.02	0.01	0.01
0.24	0.71	0.37	0.15	0.05	0.21	0.45	0.46	0.35	0.13	0.06	0.06	0.03	0.02	0.01	0.01
0.14	0.38	1.14	0.35	0.07	0.06	0.17	0.37	0.44	0.19	0.08	0.03	0.02	0.02	0.02	0.01
0.09	0.16	0.42	1.10	0.23	0.07	0.04	0.23	0.40	0.40	0.20	0.06	0.02	0.02	0.01	0.01
0.23	0.12	0.14	0.51	1.37	0.37	0.08	0.04	0.19	0.46	0.31	0.14	0.03	0.01	0.01	0.01
0.44	0.18	0.07	0.13	0.32	0.95	0.23	0.04	0.03	0.12	0.24	0.14	0.07	0.02	0.00	0.01
0.49	0.57	0.30	0.06	0.09	0.25	0.79	0.23	0.08	0.03	0.06	0.09	0.08	0.02	0.01	0.01
0.55	0.86	0.78	0.35	0.06	0.07	0.23	0.97	0.21	0.06	0.03	0.04	0.07	0.04	0.02	0.01
0.37	0.71	0.92	0.71	0.21	0.07	0.14	0.25	0.90	0.21	0.07	0.02	0.05	0.05	0.02	0.02
0.21	0.43	0.61	0.83	0.50	0.18	0.06	0.13	0.22	0.79	0.20	0.05	0.02	0.02	0.02	0.04
0.22	0.19	0.31	0.36	0.36	0.25	0.11	0.05	0.09	0.16	0.46	0.12	0.03	0.01	0.01	0.03
0.51	0.52	0.33	0.42	0.37	0.48	0.39	0.14	0.07	0.16	0.25	0.62	0.16	0.06	0.01	0.03
0.49	0.45	0.29	0.27	0.19	0.27	0.36	0.27	0.13	0.07	0.11	0.20	0.47	0.12	0.02	0.01
0.30	0.45	0.40	0.20	0.15	0.18	0.29	0.38	0.29	0.09	0.08	0.11	0.14	0.42	0.07	0.01
0.15	0.44	0.39	0.29	0.06	0.16	0.13	0.31	0.28	0.22	0.12	0.07	0.09	0.11	0.27	0.07
0.25	0.34	0.54	0.43	0.12	0.11	0.13	0.27	0.31	0.31	0.31	0.14	0.05	0.09	0.07	0.18

APPENDIX B. AGGREGATE APPROACH

Let $a_{ij}, i, j = 1, \dots, m$, be the elements of a known matrix, where i, j refer to row and columns, respectively, and m is the number of age groups in the known matrices A.1 and A.2. Let us aggregate the known matrices as the required matrices as introduced in [8]. Denote the modified matrix of $D = (d_{fg})$ with $f, g = 1, \dots, n$. Then the infer and super subindexes of the revised matrices satisfies

$$j_{inf} = l(g), \quad j_{sup} = u(g).$$

- 1. Let the contact rate between someone in group i and another individual in group g satisfy

$$d_{ig} = \sum_{j=l(g)}^{u(g)} a_{ij}. \quad (\text{B.1})$$

For the sake of convenience, we gave an example with matrix A.1 to show this calculation process. If we wanted to calculate the average number of contacts between an individual in 0–19 age group (in the first four columns) and another individual in 0–4 age group (in the first one row), we defined it as

$$d_{11} = \sum_{j=l(1)}^{u(1)} a_{1j}, \quad j = 1, 2, 3, 4, \quad (\text{B.2})$$

i.e.,

$$d_{11} = \sum_{j=1}^4 a_{1j} = 2.70. \quad (\text{B.3})$$

Repeating the above processes, we aggregated the Table A.1 into the transpose of Table B.1.

After further aggregations, 0–4, 5–9, 10–14 and 15–19 age groups were aggregated into a single age group (0–19 age group), as shown in Table B.1.

- 2. Next, we tried to aggregate the row vectors with 16 dimensions into the associated row vectors with 3 dimensions. Assuming that Chinese population in the age group i is N_i , we calculated the population-weighted contacts of $N_i d_{ig}$ and then obtained the total contact rates between groups f and g . The above processes were given by

$$Y_{fg} = \sum_{i=l(f)}^{u(f)} N_i d_{ig}, \quad (\text{B.4})$$

and

$$Y_{gf} = \sum_{i=l(g)}^{u(g)} N_i d_{if}. \quad (\text{B.5})$$

Using equation (B.4), we calculated Y_{11} (the total number of contacts made between 0–19 and 0–19 age groups), Y_{12} (the total number of contacts made between 0–19 and 20–49 age groups) and Y_{21} (the total

TABLE B.1. 16-group by 4-group matrix in the general mode.

2.70	8.08	10.48	15.46	4.20	2.09	2.88	3.98	4.41	3.95	2.90	3.53	2.12	1.56	1.47	1.71
3.73	3.65	3.44	6.27	13.43	12.13	10.31	11.13	11.29	9.02	7.71	7.28	4.40	2.88	2.44	1.92
0.82	0.66	0.58	0.87	1.62	1.80	1.77	2.02	2.08	2.04	2.64	3.34	2.37	1.78	1.72	1.13
0.03	0.03	0.04	0.02	0.05	0.02	0.04	0.03	0.05	0.08	0.06	0.07	0.05	0.06	0.16	0.23

TABLE B.2. China's population distribution.

0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	≥ 75
85041	86545	83692	82803	88576	104315	127215	96291	99881	123280	120656	93967	76975	71445	41377	48282

TABLE B.3. Total number of contact matrix (General).

Age groups	Total contacts per day			
	Age groups			
	0-19	20-49	50-74	≥ 75
0-19	3084620	1440497	248043	11108
20-49	2265916	7077732	1209677	30222
50-74	1016401	2259287	1013638	28463
≥	82604	92562	54566	11333

number of contacts made between 20-49 and 0-19 age groups), respectively given by

$$\begin{aligned}
 Y_{11} &= \sum_{i=l(1)}^{u(1)} N_i d_{i1} = 85041 \times 2.70 + 86545 \times 8.08 + 83692 \times 10.48 + 82803 \times 15.46 \\
 &= 3084620,
 \end{aligned}$$

$$\begin{aligned}
 Y_{12} &= \sum_{i=l(1)}^{u(1)} N_i d_{i2} = 85041 \times 3.73 + 86545 \times 3.65 + 83692 \times 3.44 + 82803 \times 6.27 \\
 &= 1440497,
 \end{aligned}$$

and

$$\begin{aligned}
 Y_{21} &= \sum_{i=l(2)}^{u(2)} N_i d_{i1} = 88576 \times 4.20 + 104315 \times 2.09 + 127215 \times 2.88 + 96291 \times 3.98 \\
 &\quad + 99881 \times 4.41 + 123280 \times 3.95 \\
 &= 2265916.
 \end{aligned}$$

According to [18], the population distributions ($N_i, i = 1, \dots, 16$) are listed in Table B.2.

Repeating the above calculation, the total number of contact matrices with 4×4 elements in the general mode was listed in Table B.3.

- 3. Step 2 gave the whole aggregated contact matrix. Now we were concerned with the matrix that gives the number of contacts per person in the relevant age group in every day. Thus, we need to divide the

TABLE B.4. The aggregated matrix in the general model.

Age groups	Average numbers of contacts per day			
	Age groups			
	0-19	20-49	50-74	≥ 75
0-19	9.12	4.26	0.73	0.03
20-49	3.54	11.07	1.89	0.05
50-74	2.51	5.59	2.51	0.07
≥ 75	1.71	1.92	1.13	0.23

TABLE B.5. The aggregated matrix in the household model.

Age groups	Average numbers of contacts per day			
	Age groups			
	0-19	20-49	50-74	≥ 75
0-19	1.37	1.53	0.21	0.01
20-49	1.38	1.79	0.32	0.02
50-74	1.05	1.29	0.83	0.03
≥ 75	1.32	1.24	0.67	0.18

total contact numbers by the number of individuals in the associated age group, which satisfied

$$e_{fg} = \frac{\sum_{i=l(g)}^{u(g)} N_i d_{if}}{\sum_{i=l(g)}^{u(g)} N_i}. \quad (\text{B.6})$$

Obviously, e_{fg} denoted the rate at which an individual in age group f contacted with anyone in age group g per unit time.

For an instance,

$$e_{11} = \frac{3084621}{85041 + 86545 + 83692 + 82803} = 9.12, \quad (\text{B.7})$$

$$e_{12} = \frac{1440497}{85041 + 86545 + 83692 + 82803} = 4.26, \quad (\text{B.8})$$

$$e_{21} = \frac{2265916}{88576 + 104315 + 127215 + 96291 + 99881 + 123280} = 3.53. \quad (\text{B.9})$$

Doing the similar algebras, we addressed the aggregated matrix in the general mode as shown in Table B.4 and Figure 1b.

Redoing steps 1-3, we readily got the aggregated matrix in the household mode as shown in Table B.5 and Figure 1a.

REFERENCES

- [1] C.T. Bauch, J.O. Lloyd-Smith, M.P. Coffee and AP Galvani, Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. *Epidemiology* **16** (2005) 791-801.
- [2] D. Gamerman and H.F. Lopes, Markov chain Monte Carlo: stochastic simulation for Bayesian inference. Chapman and Hall/CRC, Boca Raton (2006).
- [3] A.J. Kucharski, T.W. Russell, C. Diamond, *et al.*, Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* (2020). [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4).

- [4] X.Z. Li, J.Y. Yang and M. Martcheva, Age structured epidemic modelling, *Interdisciplinary Applied Mathematics*. Cham, Switzerland (2020) 11, 6330.
- [5] Q. Li, X.H. Guan, P. Wu, *et al.*, Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England J. Med.* **382** (2020) 1199–1207.
- [6] Z. Liu, P. Magal, O. Seydi and G. Webb, Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. Preprint medRxiv (2020). <https://arxiv.org/abs/2002.12298>.
- [7] X. Luo, S. Feng, J. Yang, *et al.*, Analysis of potential risk of COVID-19 infections in China based on a pairwise epidemic model. (2020). <https://www.preprints.org/manuscript/202002.0398/v1>.
- [8] M.I. Meltzer, M. Gambhir, C.Y. Atkins and D.L. Swerdlow, Standardizing scenarios to assess the need to respond to an influenza pandemic. *Clin. Infect. Dis.* **60** (2015) S1–8.
- [9] K. Prem, A.R. Cook and M. Jim, Projecting social contact matrices in 152 countries using contact surveys and demographic data, *Plos. Comput. Biol.* **13** (2017) e1005697.
- [10] J.M. Read, J. Lessler, S. Riley, S. Wang, L.J. Tan, K.O. Kwok, Y. Guan, C.Q. Jiang and D.A.T. Cummings, Social mixing patterns in rural and urban areas of southern China. *Proc. R. Soc. B* **281** (2014) 20140268.
- [11] P. Shao and Y.G. Shan, Beware of asymptomatic transmission: study on 2019-nCov prevention and control measures based on SEIR model (2020). <https://doi.org/10.1101/2020.01.28.923169>.
- [12] The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 Novel coronavirus disease (COVID-19)-China, 2020. *CCDC Weekly* **2** (2020) 113–122.
- [13] P. Van Driessche and J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180** (2002) 29–48.
- [14] J. Wallinga, P. Teunis and M. Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am. J. Epidemiol.* **164** (2006) 936–944.
- [15] S.L. Zhao and H. Chen, Modelling the epidemic dynamics and control of COVID-19 outbreak in China, *Quant. Biol.* (2020). <https://doi.org/10.1007/s40484-020-0199-0>.
- [16] Chinese Control Disease Center and Prevention. The trends and risk estimation of COVID-19. (2020). <http://www.chinacdc.cn/jkzt/crb/zl/szkb11803/jszl11811/202001/P020200127544648420736.pdf>.
- [17] Official news on the change of diagnosed approaches issued by the Chinese government (in Chinese). (2020). <http://www.gov.cn/zhengce/zhengceku/2020-02/09/content5476407.htm>.
- [18] Chinese Population Statistic Yearbook 2019. <http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm>.
- [19] 2019 novel coronavirus (in Chinese). <https://baike.baidu.com/item/2019novelcoronavirus/24267858?fr=aladdin>.
- [20] The National Health Commission of the People’s Republic of China. <http://www.nhc.gov.cn/xcs/yqtb/listgzbd.shtml>.
- [21] China News. Mayor of Wuhan: More than 5 million people left Wuhan. 2020. <http://ent.chinanews.com/sh/2020/01-26/9070484.shtml>.
- [22] Chinese Control Disease Center and Prevention. <http://www.chinacdc.cn/jkzt/crb/zl/szkb11803/jszl11809>.