Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 21(1): 861–883.
DOI: 10.3934/mbe.2024036
Received: 17 September 2023
Revised: 15 November 2023
Accepted: 06 December 2023
Published: 21 December 2023

Research article

Quantifying the presymptomatic transmission of COVID-19 in the USA

Luyu Zhang[1], Zhaohua Zhang[1], Sen Pei[2], Qing Gao[3,4] and Wei Chen[4,5,6,*]

[1] LMIB and School of Mathematical Sciences, Beihang University, Beijing 100191, China

[2] Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY 10032, USA

[3] School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

[4] Zhongguancun Laboratory, Beijing 100194, China

[5] Institute of Artificial Intelligence, Beihang University, Beijing 100191, China

[6] Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China

* **Correspondence:** Email: chwei@buaa.edu.cn.

Abstract: The emergence of many presymptomatic hidden transmission events significantly complicated the intervention and control of the spread of COVID-19 in the USA during the year 2020. To analyze the role that presymptomatic infections play in the spread of this disease, we developed a state-level metapopulation model to simulate COVID-19 transmission in the USA in 2020 during which period the number of confirmed cases was more than in any other country. We estimated that the transmission rate (i.e., the number of new infections per unit time generated by an infected individual) of presymptomatic infections was approximately 59.9% the transmission rate of reported infections. We further estimated that at any point in time the average proportion of infected individuals in the presymptomatic stage was consistently over 50% of all infected individuals. Presymptomatic transmission was consistently contributing over 52% to daily new infections, as well as consistently contributing over 50% to the effective reproduction number from February to December. Finally, non-pharmaceutical intervention targeting presymptomatic infections was very effective in reducing the number of reported cases. These results reveal the significant contribution that presymptomatic transmission made to COVID-19 transmission in the USA during 2020, as well as pave the way for the design of effective disease control and mitigation strategies.

Keywords: epidemiology; nonlinear dynamics; presymptomatic transmission; covid-19; metapopulation model

1. Introduction

Severe acute respiratory syndrome coronavirus 2, the causative agent of COVID-19, spreads quickly and has established a global pandemic within a short period of time since it was first detected [1]. In the USA, despite wide implementation of non-pharmaceutical interventions after the initial outbreak in the spring of 2020, this disease resurged and continued spreading from mid 2020 throughout the year. This disease caused over 20 million confirmed infections by the end of 2020, which was more than any other countries [2–4]. A major challenge to the surveillance and control of this disease is the hidden transmission led by people with presymptomatic infections who are in the incubation period and do not show symptoms or mild symptoms for the moment [5–7]. It is urgent and necessary to understand the role of presymptomatic infections in disease transmission in the USA to make effective diseases control and mitigation strategies.

Presymptomatic infections have been discovered in many countries in the process of determining the epidemiological traceability of confirmed cases [8–13]. Many previous studies have focused on the contribution of presymptomatic infections to the secondary transmission, which is critical for determining the prioritization of various surveillance and control measures [14, 15]. Clinical studies suggested that the infectiousness of COVID-19 peaked on or before symptom onset [14]. Contact tracing studies showed that the proportion of transmission routes formed due to presymptomatic transmission varies greatly due to many confounding factors and could range from 6.4% to over 65% [14, 16–21]. A model-based study found that presymptomatic transmission could contribute a value as high as 0.9 to the basic reproduction number, which accounted for 46% of the contribution to the basic reproduction number. Therefore, presymptomatic transmission becomes a significant contributor to secondary transmission [11]. Epidemiological models have been used to quantify the presymptomatic transmission of COVID-19 [22, 23]. However, when quantifying the presymptomatic transmission in the United States of America in 2020, the success of these epidemiological models are impaired by the two following limitations. First, large-scale transmission between different states occurs frequently due to human mobility, which is not incorporated into existing models. Second, the disease transmission rate in different states varies due to heterogeneity in human contact, which is also not considered in existing models. In addition to the limitations in modeling, previous studies only focused on the presymptomatic transmission in the early stage of the COVID-19 pandemic. During the year 2020, the USA experienced three pandemic waves which caused the largest number of ascertained cases and deaths among all countries. Therefore, the role of presymptomatic transmission throughout the year of 2020 needs to be more comprehensively studied.

In this study, we developed a state-level metapopulation SEPAIR model to simulate the spread of COVID-19 in the USA from February to December in 2020 (Figure 1). This metapopulation SEPAIR model incorporates many important factors that impact disease spread including heterogeneity in human contact and human mobility. We quantified the contagiousness and prevalence of presymptomatic infections during the continuous development of COVID-19 in each state [23, 24]. Subsequently, we studied the contributions of different transmission routes including presymtomatic transmission to the spread of the disease in each state. Finally, we performed counterfactual simulation to quantify the effect of controlling presymptomatic transmission on the COVID-19 disease spread.

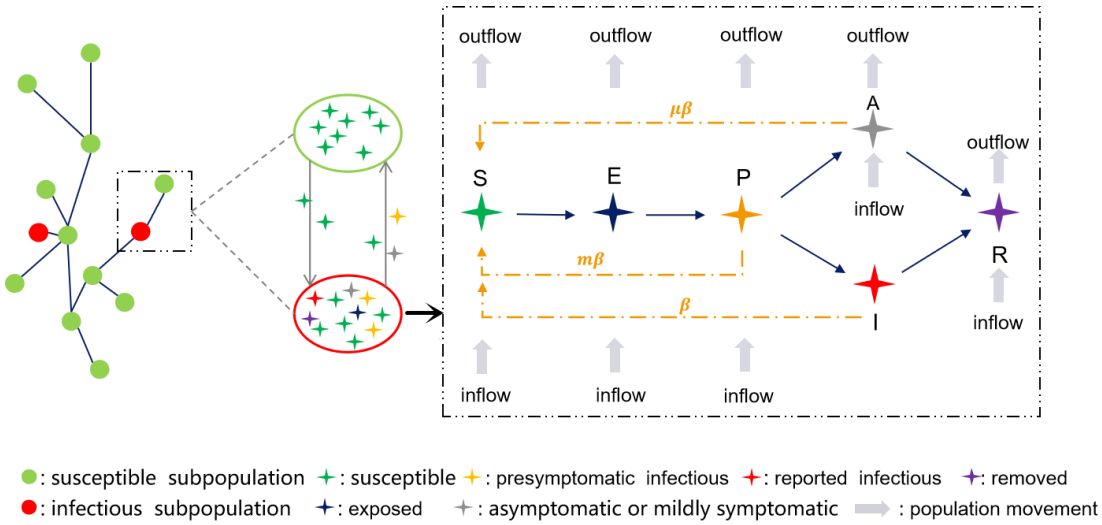

Figure 1. Metapopulation SEPAIR model. A red subpopulation with infected individuals may spread the disease to a green uninfected subpopulation through population movement. This allows community transmission to occur in the green (susceptible) subpopulation after a period of time. In the SEPAIR model, there are six compartments: susceptible (S), exposed (E), presymptomatic infectious (P), asymptomatic or mildly symptomatic infectious (A), reported infectious (I) and removed (R). Susceptible individuals can be infected with presymptomatic infections, asymptomatic or mildly symptomatic infections or reported infections to become exposed individuals. After latent period Z_1, an exposed individual becomes a presymptomatic infected individual, who does not show symptoms but is contagious. After a contagious incubation period Z_2, a presymptomatic individual becomes a reported individual, or an asymptomatic or mildly symptomatic individual. β is the transmission rate of reported infection; m and μ represent the relative transmission rates of presymptomatic infection and asymptomatic or mildly symptomatic infection, respectively.

2. Methods

We have developed a SEPAIR metapopulation network model based on commuting data between different states and human mobility data to characterize the large-scale spatial transmission of COVID-19 [25] (Figure 1). In this model, the transmission of COVID-19 throughout the day and night is described as a discrete Markov process, as given by Eqs (A1)–(A12) in the Appendix. We divided the population into susceptible individuals S, exposed individuals E (i.e., infected but no symptoms and not infectious), presymptomatic cases P (i.e., infected, no symptoms but infectious), reported cases I, asymptomatic or mildly symptomatic cases A and removed cases R. Susceptible individuals S can be infected by presymptomatic individuals P, asymptomatic or mildly symptomatic individuals A and reported individuals I to become exposed individuals E. β_i (the initial prior range is 0.01 to 2.5) is the transmission rate of reported cases in state i, and m (the initial prior range is 0.1 to 1.0) and μ (the initial prior range is 0.1 to 1) represent the relative transmission rates of presymptomatic cases and asymptomatic or mildly symptomatic cases, respectively. After an average latent period Z_1 (the initial prior range is 2 to 5), an exposed individual becomes a presymptomatic individual who

is contagious but does not show symptoms. After a contagious incubation period Z_2 (the initial prior range is 2 to 5), α_i (the initial prior range is 0.03 to 1), i.e., the fraction of presymptomatic individuals in state i becomes reported individuals, and the rest of the presymptomatic individuals become mildly symptomatic or asymptomatic cases. The incubation period contains two stages: latent period and contagious incubation period, denoted as Z1+Z2. D (the initial prior range is 2 to 5) is the average duration of the infectious period after which both reported individuals I and asymptomatic or mildly symptomatic individuals A recover or die and then become removed individuals R. The priors of the parameters in our model was taken from [24].

Here a subpopulation is defined as a pair of states in the United States of America between which there is human mobility based on the data. Then, the 51 states and regions were divided into 1367 subpopulations. Every node in the metapopulation network represents a subpopulation and two nodes are connected if there is human mobility between them. Here, two types of population mobility are considered, i.e., daily commuting mobility and random visit mobility. The population mobility caused by the relatively fixed level of work commuting is defined as the commuting population mobility. N_{ij} denotes the number of people in the subpopulation who live in the state j and work in the state i. During the daytime, they travel from the state j to the state i to participate in local transmission in the state i, and they return to the state j at night to participate in disease transmission in the state j. N_{ii} is the number of individuals in the total population who both live and work in state i. They stay in state i for local disease transmission throughout the day and night. The random flow, which comprises travel, visits, etc., is defined as the random population mobility. The number of random visitors between subpopulations i and j is proportional to the average number of commuters between them in our model, denoted as $\theta\bar{N}_{ij}$, in which $\bar{N}_{ij} = \left(N_{ij} + N_{ji}\right)/2$ and θ is an adjustable parameter [26]. The resident travel data set encompassing 51 states and regions in the USA was taken from the Foursquare Labs Inc laboratory where the time, duration and location of users are determined by mobile device signals [27]. The real-time population flow data detailed the flow of people in 50 states and Washington, D.C. from February 21, 2020 to December 11, 2020. With the implementation of disease prevention and control measures, population mobility had decreased, and interstate commuting data before the pandemic could no longer reflect population mobility between pairs of states. We used the population mobility data from March 1, 2020 as a benchmark, and we calculated the population mobility between pairs of states following the initiation of the disease prevention and control measures by applying the proportion of the relative decrease in population mobility θ (the initial prior range is 0.01 to 0.3) to the data after March 15.

We calibrated the transmission model to state-level incidence data reported from February 21, 2020 to December 11, 2020 and death data reported from February 21, 2020 to December 26, 2020. We inferred the time-varying parameters of the model by using a sequential data assimilation technique, i.e., the ensemble-adjusted Kalman filter algorithm, which has been widely applied for the prediction of infectious diseases such as influenza [28–30]. Compared to other particle filter methods, it demonstrates better performance on high-dimensional systems [26, 31–33]. The details on the implementation of the ensemble-adjusted Kalman filter algorithm can be found in previous studies [27, 28]. In addition, we conducted counterfactual simulations to quantify the impact of controlling presymptomatic transmission through non-pharmaceutical interventions on COVID-19 disease spread. Specifically, we first employed data ranging from February 21, 2020 to December 11, 2020 to estimate the original values of the relative transmission rates of presymptomatic cases m. Then, we multiplied the values of m by 0.5, 0.3 and 0.1, respectively while keeping the values of other

parameters unchanged. Using these updated values of parameters, we predicted the number of daily new reported cases. Here, we quantify the effect of controlling presymptomatic transmission on the COVID-19 disease spread for the United States of America and five representative states.

3. Results

3.1. Model calibration

We first inferred the parameters of the SEPAIR metapopulation model based on the ensemble-adjusted Kalman filter algorithm. The inferred epidemiological parameters for the period from February 21, 2020 to December 11, 2020 are shown in Table 1. Note that these parameters changed over time. We used the inferred parameters to calculate the basic reproduction number $R_t = \alpha\beta D + (1-\alpha)\mu\beta D + m\beta Z_2$, which was the largest eigenvalue of the next-generation matrix [24, 34]. We further calculated the effective reproduction number $R_e = (\alpha\beta D + (1-\alpha)\mu\beta D + m\beta Z_2)\frac{S}{N}$ for the study period from February 21, 2020 to December 11, 2020. The effective reproduction number gradually decreased from 4.45 (95% CI: [4.24–4.67]) at the beginning of 2020 to 1.35 (95% CI: [1.28–1.43]) at the end of the year. However, as of December 11, 2020 except for Hawaii, Indiana, Minnesota, Montana, Oklahoma, Wyoming and Vermont, the effective reproduction number was still larger than 1 for all states, which indicated that the COVID-19 in the USA continued under the current disease control measures. We further found that, during our study period, the average transmission rate of the presymptomatic infections was 59.94% (95% CI: [58.08%–61.72%]) the average transmission rate of reported cases (Table 1). Moreover, the average transmission rate of the asymptomatic or mildly symptomatic cases was 30.52% (95% CI: [29.73%–31.68%]) the average transmission rate of reported cases. We also obtained that the average incubation period was about 7.944 days (95% CI: [7.794–8.076]), and that the average contagious incubation period during which an individual is infectious but without symptoms was as long as 4.285 days (95% CI: [4.223–4.314]). Since the presymptomatic infections were difficult to identify, the high contagiousness and long incubation period of presymptomatic infection in individuals with SARS-CoV-2 made the prevention and control of COVID-19 rather challenging [35, 36].

Table 1. The inferred epidemiological parameters from February 21, 2020 to December 11, 2020.

Parameter	**Median (95% CI)**
Average reporting rate (α)	0.460 (0.456, 0.464)
Average transmission rate (β)	0.554 (0.552, 0.555)
Latent period (Z_1, days)	3.659 (3.571, 3.753)
Contagious incubation period (Z_2, days)	4.285 (4.223, 4.314)
Infectious period (D, days)	3.951 (3.820, 4.041)
The relative transmission rate of presymptomatic cases (m)	0.599 (0.580, 0.617)
The relative transmission rate of asymptomatic or mildly symptomatic cases (μ)	0.305 (0.297, 0.317)
Mobility factor (θ)	0.152 (0.147, 0.159)

3.2. Prevalence of presymptomatic infections

To evaluate the efficacy of the mitigation and control measures, we estimated the prevalence of presymptomatic infections in the ongoing spread of COVID-19 during the study period. Here, we measured the numbers of active infections in the classes of presymptomatic infections, reported infections and asymptomatic or mildly symptomatic infections. Nationwide, the estimated number of presymptomatic infections had three peaks during 2020, which appeared simultaneously with the three pandemic waves (Figure 2(a)). The number of active presymptomatic infections was consistently more than that in other classes during the study period. At the time of the first peak, the number of presymptomatic infections was even about 11.2371 times (95% CI: [10.8051–11.6691]) the number of reported infections. In the autumn and winter seasons of the pandemic, the number of presymptomatic infections continued on an upward trend and reached about 1663883 (95% CI: [1619093–1708671]) on December 11. The average proportion of presymptomatic infections among all active infections continued to exceed 50%, and peaked at 68.87% (95% CI: [68.03%–69.71%]) in the initial stage of the pandemic in spring (Figure 2(b)). We further studied five representative states, i.e., California (CA), Texas (TX), Florida (FL), Illinois (IL) and New Jersey (NJ), which had large numbers of reported infections and were located in different parts of the USA. The number of presymptomatic infections in California, Texas and Florida had three clear peaks during the three pandemic waves. However, the numbers of presymptomatic infections in Illinois and New Jersey did not have a peak during the second pandemic wave. The numbers of presymptomatic infections in Illinois and New Jersey indicated pronounced peaks in the spring, with values of about 42402 (95% CI: [41121–43684]) and 166481 (95% CI: [162062–170899]) (Figure 2(a)). The numbers of presymptomatic infections in Texas and Florida revealed pronounced peaks in early July during the second pandemic wave. Texas reached 103566 (95% CI: [100757–106374]), and Florida reached 60841 (95% CI: [58897–62786]) (Figure 2(a)).

California, the state with the most reported cases in the nation, saw three peaks of presymptomatic infections in early April, mid-July and winter, reaching 33049 (95% CI: [31473–34526]), 46253 (95% CI: [44946–47559]) and 222416 (95% CI: [215327-229506]), respectively (Figure 2(a)). Although the number of presymptomatic infections varied by state, presymptomatic infections still constituted the largest proportion among all classes of active infectious cases; it reached the highest proportion during the spring pandemic wave which is similar to the situation of the whole country (Figure 2(b)).

3.3. Contribution of presymptomatic transmission to disease spread

In our metapopulation SEPAIR model, susceptible individuals can be infected by active infections and become the exposed individuals. To quantify the contribution of presymptomatic transmission to daily new infections, we measured the expected number of susceptible individuals infected by active infections in different classes (i.e., presymptomatic infections, reported infections and asymptomatic or mildly symptomatic infections), as well as their corresponding proportions every day. Nationwide, the proportion of daily new infections from presymptomatic transmission consistently exceeded 52% during the study period (P/(P+A+I)), which implies that presymptomatic infection was a crucial and stable source of daily new infections, and that the existing disease control measures were uneffective in reducing presymptomatic transmission (Figure 3(a)). Nationwide, the proportion of daily new infections from presymptomatic transmission before mid-June was far greater than that from active

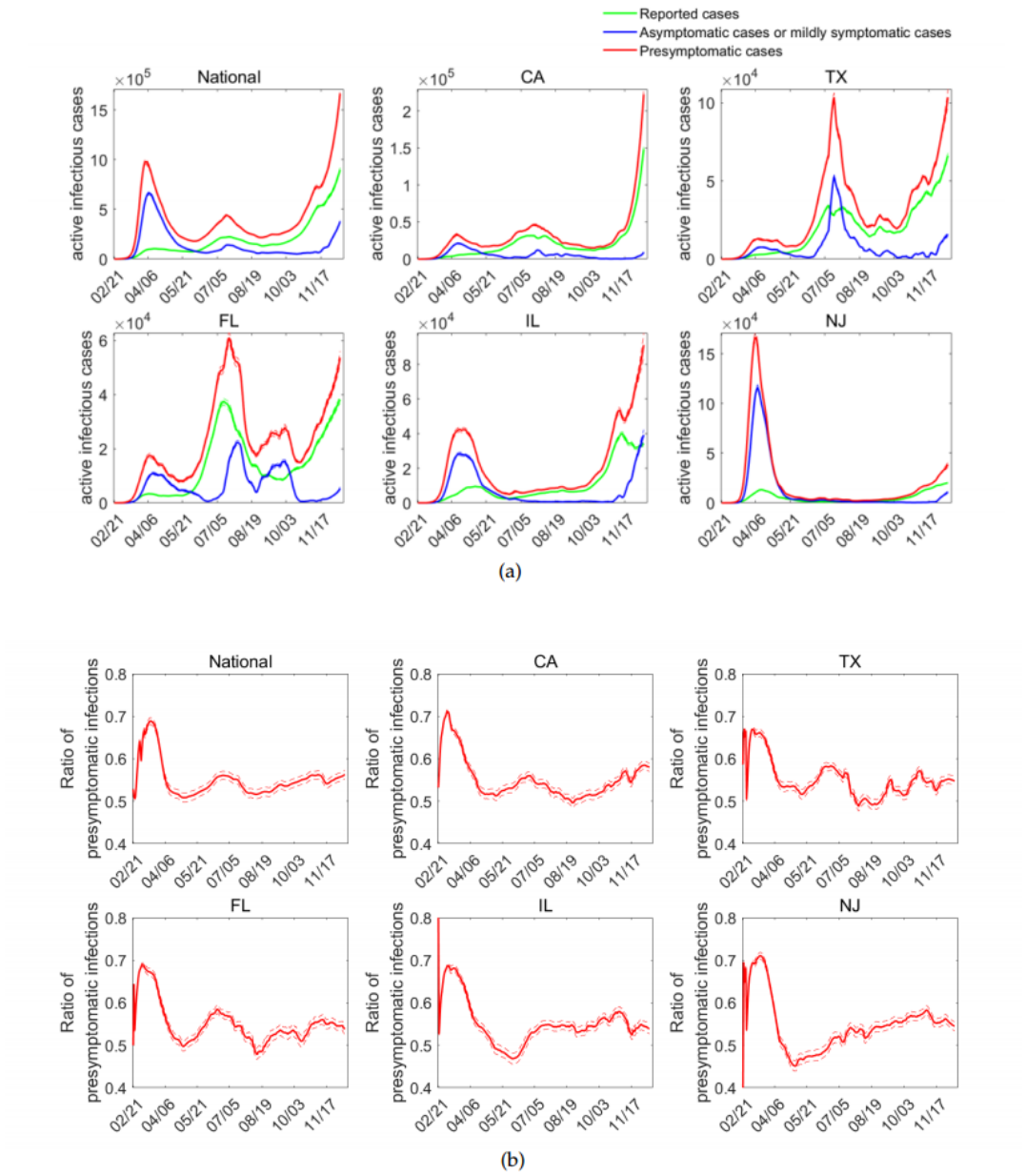

Figure 2. (a) The number of active infections nationwide (National) and in California (CA), Texas (TX), Florida (FL), Illinois (IL) and New Jersey (NJ). The red curve represents the mean of the number of presymptomatic infections, and the area between dotted lines is the 95% confidence interval. The green curve represents the mean of the number of reported cases, and the area between dotted lines is the 95% confidence interval. The blue curve represents the mean of the number of asymptomatic cases or mildly symptomatic cases, and the area between dotted lines is the 95% confidence interval. (b) Proportion of presymptomatic infections among active infections. The red curve represents the mean proportion of presymptomatic infections, and the area between dotted lines is the 95% confidence interval.

infections in other classes, peaking at 71.87% (95% CI: [70.35%–73.39%]) on March 22, 2020. Among all of the states, this proportion ranged from 66.47% (95% CI: [64.57%–68.37%]) in Tennessee to 82.05% (95% CI: [80.66%–83.33%]) in New York on that day. Nationwide, this proportion declined after the spring pandemic wave, reaching its minimal value of 51.88% (95% CI: [50.18%–53.69%]) on November 16, 2020 (Figure 3(a)). Among all of the states, this proportion ranged from 47.61% (95% CI: [45.87%–49.34%]) in Wisconsin to 74.54% (95% CI: [71.84%–77.24%]) in Vermont on that day (Figure 3(b)). On that day besides Vermont, New Hampshire had a significantly larger proportion of infections from the presymptomatic transmission than their surrounding states, i.e., 73.25% (95% CI: [70.70%–75.80%]). From March 24 to November 16, New York was the state with the largest decline in this proportion, as it decreased from 82.05% (95% CI: [80.66%–83.33%]) to 51.66% (95% CI: [50.00%–53.33%]), while the states of Vermont, Hawaii and Maine had only slight changes in this proportion. This result may be due to the differences in population mobility among these states.

To quantify the contribution of presymptomatic transmission to disease spread, we also studied the contribution of the presymptomatic transmission of COVID-19 to the effective reproduction number. According to the next generation matrix method [34], the $m\beta Z_2\frac{S}{N}$ part of the effective reproduction number $R_e = (\alpha\beta D + (1-\alpha)\mu\beta D + m\beta Z_2)\frac{S}{N}$ was completely caused by presymptomatic transmission. We used this to measure the contribution of presymptomatic transmission to the effective reproduction number. The effective reproductive number contributed by presymptomatic transmission in New Jersey and Illinois exhibited peaks as high as 7.07 (95% CI: [6.70–7.44]) and 5.17 (95% CI: [4.85–5.49]) respectively, exceeding the national peak of 4.45 (95% CI: [4.24–4.67]) which occured in the spring pandemic wave (Figure 4(a)). The contribution of presymptomatic transmission to the effective reproduction number for Texas and Florida also reached peaks of 4.40 (95% CI: [4.11–4.68]) and 4.71 (95% CI: [4.40–5.01]), respectively, and exhibited another small peak in July and June, respectively, after declining for several months. The effective reproduction number associated with presymptomatic infection exceeded 1 again, reaching 1.61 (95% CI: [1.52–1.70]) and 1.61 (95% CI: [1.51–1.70]) for Texas and Florida, respectively. This finding implies that presymptomatic transmission is one of the major factors leading to the summer pandemic wave.

We further measured the effective reproduction number contributed by the transmission of reported infections and asymptomatic or mildly symptomatic infections, and we measured the contribution proportion of different transmission routes to the effective reproduction number, particularly the presymptomatic transmission, transmission by reported infections and transmission by asymptomatic or mildly symptomatic infections respectively. We found that, during the study period, presymptomatic transmission contributed more than 50% in the early stage of the pandemic, as averaged over 200 independent simulations (Figure 5). Nationwide, the contribution proportion of presymptomatic transmission peaked in April at 66.28% (95% CI: [64.51%–68.05%]). Analyzing by state, California, Texas, Florida, Illinois and New Jersey also peaked in contributions in April. When the contribution proportion of presymptomatic transmission reached the national peak on April 9, 2020, this proportion ranged from 60.91% (95% CI: [59.07%–62.76%]) in New Mexico to 72.58% (95% CI: [70.78%–74.34%]) in Michigan. The contribution proportion declined with the development of this disease, but the minimum still reached 51.43% (95% CI: [49.70%–53.16%]) on November 14, 2020, which still played a significant role in the continuous spread of the pandemic. Analyzing by state, on that day, the contribution proportion ranged from 49.51% (95% CI:

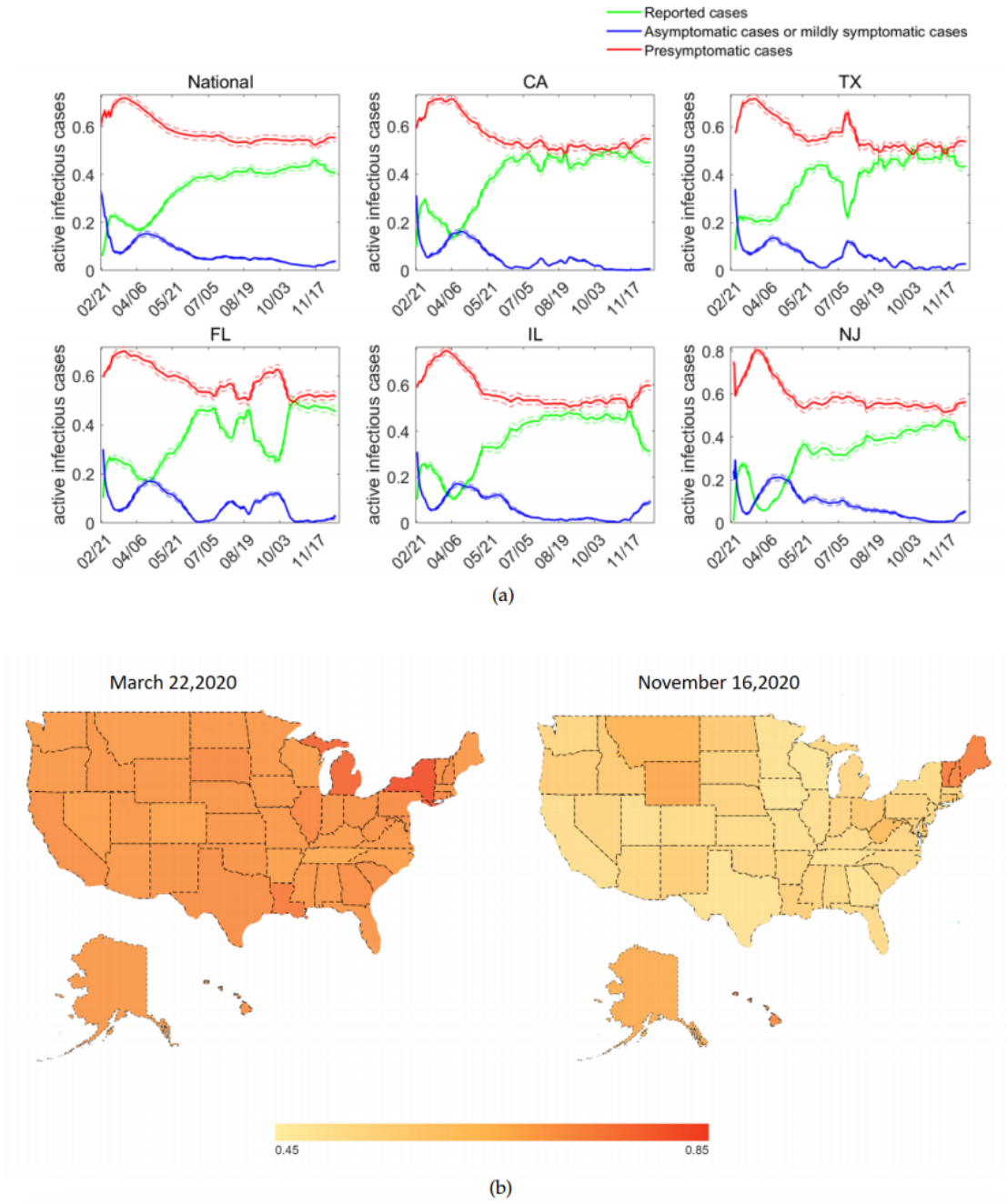

Figure 3. (a) Proportion of daily new infections from active infections for different classes. The red curves represent the mean proportions infected by presymptomatic infections, with the area between dotted lines being the 95% confidence interval. The green curves represent the mean proportions infected by reported cases, with the area between dotted lines being the 95% confidence interval. The blue curves represent the mean proportions infected by asymptomatic cases or mildly symptomatic cases, with the area between dotted lines being the 95% confidence interval. (b) Estimated spatial variation in the proportions of daily new infections from the presymptomatic transmission.

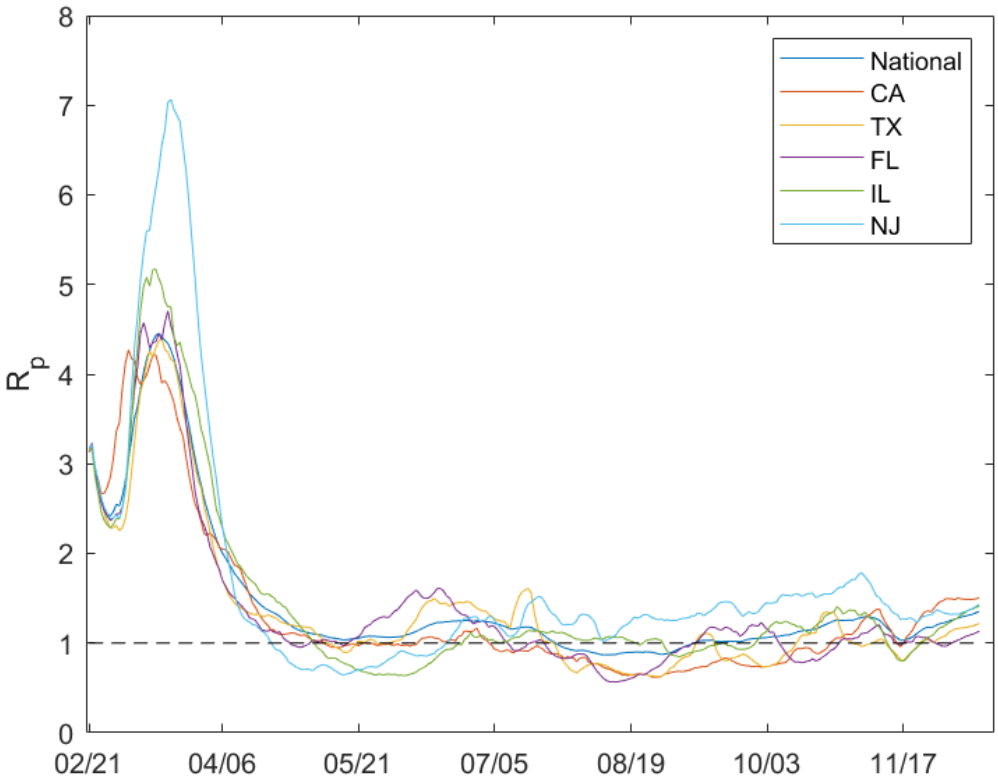

Figure 4. Effective reproduction number contributed by presymptomatic transmission, which is $R_p = (m\beta Z_2)\frac{S}{N}$.

[47.79%–51.23%]) in California to 69.20% (95% CI: [67.19%–71.20%]) in Vermont (Figure 6). The average contribution proportion of Vermont and Maine located in the northeast, was 56.17% which was significantly larger than their surrounding states. In 2020, the state with the largest decline in contribution proportion from presymptomatic transmission was New Jersey, which decreased from 71.46% (95% CI: [69.66%–73.26%]) to 50.42% (95% CI: [48.70%–52.15%]), and the state with the smallest decline was Washington D.C. which decreased by 0.77% (95% CI: [–0.79%–2.32%]).

3.4. Counterfactual simulation by controlling presymptomatic transmission

Our results show that presymptomatic transmission constituted a stable and crucial contributor to the COVID-19 disease spread throughout the entire study period. Measures to control presymptomatic transmission through non-pharmaceutical interventions, such as keeping social distance and restricting the opening of public places, would prevent and control the disease spread [19, 37–43]. To quantify the effect of controlling presymptomatic transmission on the COVID-19 disease spread, we performed a counterfactual simulation on the number of daily cases by multiplying the daily relative transmission rate parameters of presymptomatic infection by 0.5 times, 0.3 times and 0.1 times, respectively, while maintaining other parameters to be the same as their original values. Nationwide, the daily number of newly reported infections was significantly reduced

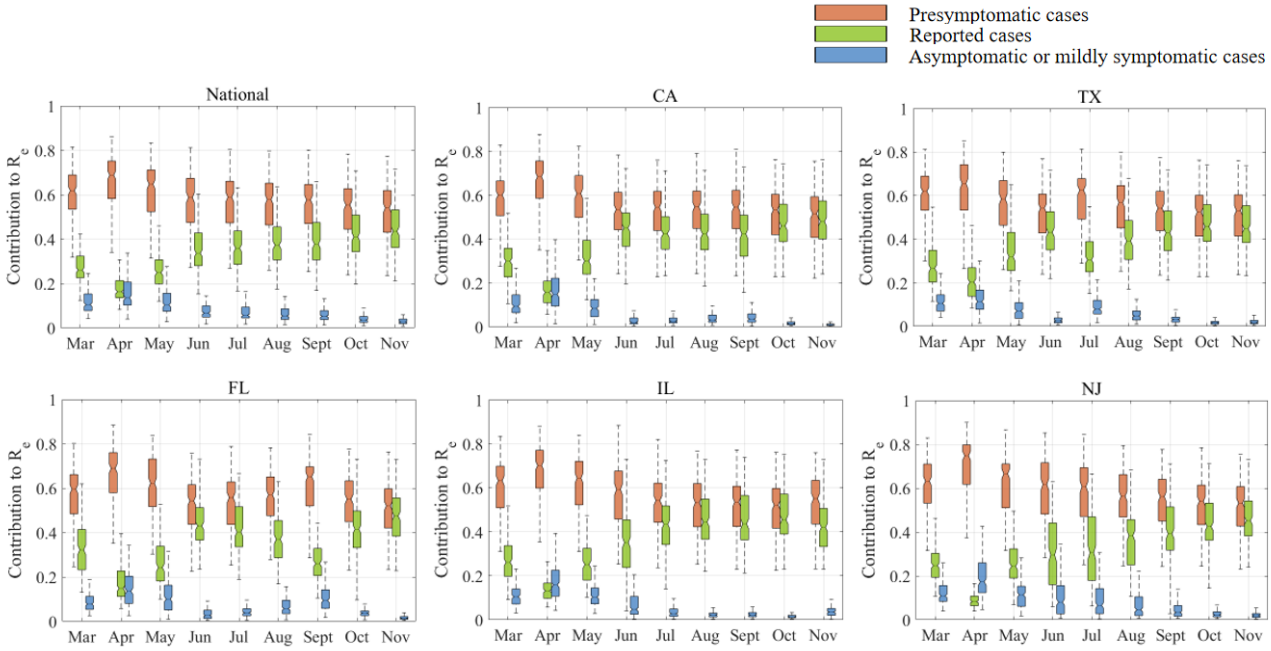

Figure 5. Contribution of different transmission routes to the effective reproduction number. Red boxes represent the contribution of presymptomatic transmission to the effective reproduction number, blue boxes represent the contribution of transmission by reported infections to the effective reproduction number and green boxes represent the contribution of transmission by asymptomatic or mildly symptomatic infections to the effective reproduction number. The boxes show the median and interquartile of 500 simulations, and whiskers show the maximum and minimum values.

and there was a major decline for the pandemic waves in summer and winter (Figure 7). Specifically, if the relative transmission rate of presymptomatic infections is changed to half, three-tenths or one-tenth of the original value, as of December 11, the cumulative number of reported infections was reduced by 9,700,762 (95% CI: [8,830,504–10,860,945]), 13,043,612 (95% CI: [12,482,191–13,493,424]) or 14,297,830 (95% CI: [14,096,437–14,653,129]), respectively, which means that 60.79% (95% CI: [55.34%–68.06%]), 81.74% (95% CI: [78.22%–84.56%]) or 89.60% (95% CI: [88.34%–91.83%]) of the cumulative number of reported infections will be avoided, respectively. For the five representative states, the daily number of newly reported infections will also be significantly reduced although there will be some variation. If the relative transmission rate of presymptomatic infections is changed to half of the original value, as of December 11, the number of cumulative reported infections will be reduced by 90.13% (95% CI: [87.89%–91.64%]), 89.07% (95% CI: [85.37%–90.73%]), 62.98% (95% CI: [56.77%–70.41%]), 91.38% (95% CI: [90.14%–92.77%]) and 90.88% (95% CI: [87.96%–91.98%]) for California, Texas, New Jersey, Illinois and Florida, respectively. If the presymptomatic transmission is more strictly controlled and the relative transmission rate of presymptomatic infections is three tenths of the original value, the number of cumulative reported infections will be reduced by 95.71% (95% CI: [94.96%–96.31%]), 95.89% (95% CI: [94.05%–96.48%]), 83.67% (95% CI: [80.01%–86.36%]), 96.42% (95% CI:

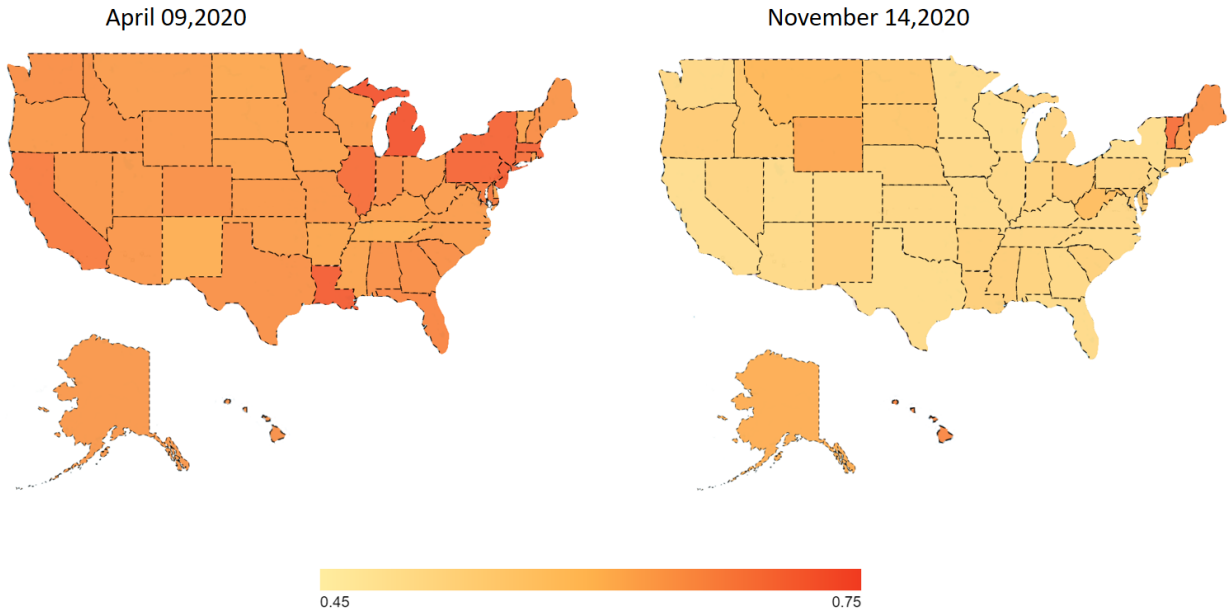

Figure 6. Estimated spatial variation in the proportions of contribution of different transmission routes to the effective reproduction number.

[95.62%–96.91%]) and 96.10% (95% CI: [95.57%–96.98%]) for California, Texas, New Jersey, Illinois and Florida, respectively. Further, if the relative transmission rate of presymptomatic infection is only one tenth of the original value, the number of cumulative reported infections will be reduced by 97.83% (95% CI: [97.58%–98.03%]), 98.11% (95% CI: [97.55%–98.40%]), 92.20% (95% CI: [90.79%–92.97%]), 98.11% (95% CI: [97.99%–98.32%]) and 98.27% (95% CI: [98.07%–98.62%]) for California, Texas, New Jersey, Illinois and Florida, respectively. The pronounced reduction in the number of daily infections suggests the importance of implementing non-pharmaceutical interventions aiming at controlling presymptomatic transmission.

4. Discussion

We developed a metapopulation SEPAIR network model that incorporates deterministic and stochastic spread based on population mobility data to investigate the dynamical characteristics of presymptomatic COVID-19 transmission in the USA in 2020. Unlike previous studies, our model inference system could capture the characteristics of the spatiotemporal dynamics of presymptomatic transmission during the ongoing disease spread in the USA. We found that the virus was highly infectious even in the presymptomatic stage during which its transmission rate reached 59.94% (95% CI: [58.08%–61.72%]) the transmission rate of reported infections. Second, hidden transmission in the presymptomatic stage had a significant impact on the ongoing disease transmission. The contagious incubation period was about 4.29 days (95% CI: [4.22–4.33]) .We found that the average proportion of infected individuals in the presymptomatic stage was consistently over 50% of all infected individuals. The contribution of presymptomatic transmission to daily new infections consistently exceeded 52% and the contribution of presymptomatic transmission to the effective

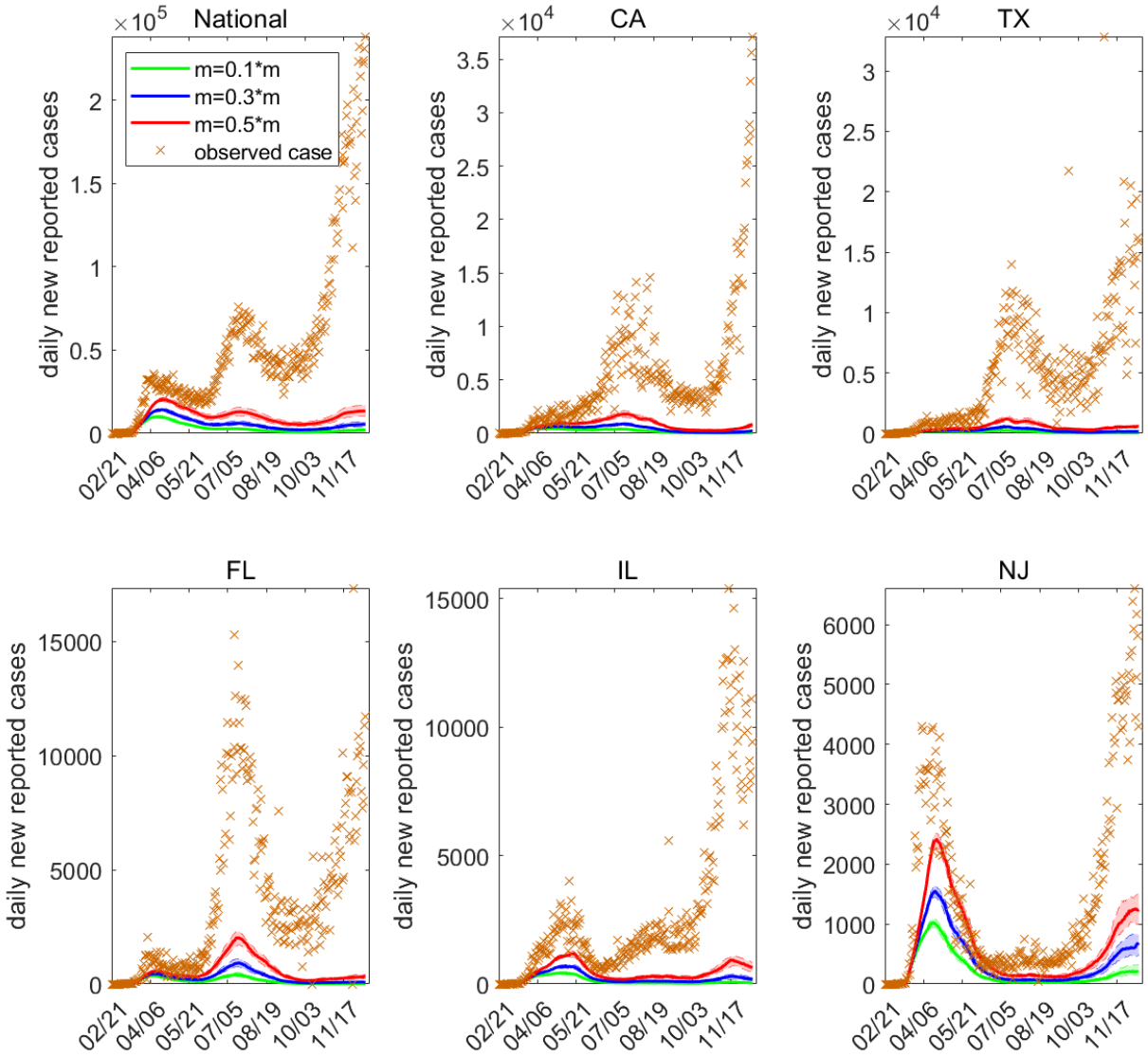

Figure 7. The counterfactual simulation model results obtained by controlling presymptomatic transmission. The brown cross represents the daily number of newly reported cases observed. The red curve represents the simulation of the daily number of newly reported cases by controlling the relative transmission rate of presymptomatic infections to 0.5 times the original value. The blue curve represents the simulation of the daily number of newly reported cases by controlling the relative transmission rate of presymptomatic infections to 0.3 times the original value. The green curve represents the simulation of the daily number of newly reported cases by controlling the relative transmission rate of presymptomatic infections to 0.1 times the original value.

reproduction number consistently exceeded 50% nationwide. Further, the current level of non-pharmacological interventions was insufficient to control the ongoing spread of the pandemic. If

the relative transmission rate of presymptomatic infection was reduced to be 0.5 times, 0.3 times or 0.1 times the original values in our model, respectively, 60.79% (95% CI: [55.34%–68.06%]), 81.74% (95% CI: [78.22%–84.56%]) and 89.60% (95% CI: [88.34%–91.83%]) of the reported infections would be avoided as of December 11. This result suggests that, in the formulation and implementation of prevention and control measures, it is necessary to increase the screening of people who do not show symptoms, and timely test and isolate presymptomatic cases [44]. It is therefore critical to take measures that can effectively reduce the risk of presymptomatic transmission, such as social distancing, wearing masks and vaccinations [45].

There are several advantages in the study. First, we leverage both the data of mobility among counties and the data of all counties to quantify presymptomatic transmission in our method. Our results are more reliable than those obtained by simply using the data of locations. Second, the data assimilation algorithm used in our study is efficient for high dimensional models. Therefore it can be applied in models characterizing the COVID-19 transmission throughout the United States of America.

There are also some limitations of this work. First, due to the lack of relevant data, in our model, we treat asymptomatic cases as a special case of mildly symptomatic cases and model it with the presymptomatic stage, which allowed us to roughly quantify the prevalence and contribution of presymptomatic transmission [23]. Second, characteristics such as age structure, gender, comorbidity and contact patterns are not considered in our model, even though they may play a significant role in disease transmission [46–48]. Lastly, we assume in our model that all reported cases are symptomatic and all symptomatic cases have been reported. However, this is a simplification for modeling and not always the case. Some reported cases may be asymptomatic if detected through contact tracing, and some symptomatic cases may not be reported.

Use of AI tools declaration

The authors declare that they have not used artificial intelligence tools in the creation of this article.

Conflict of interest

The authors declare that there is no conflict of interest.

Funding

W. Chen was funded by National Natural Science Foundation of China under grant 62388101, the National Key Research and Development Program of China under grant 2022YFF0902800, as well as Key Program of the Beijing Natural Science Foundation under grant Z180005. Q. Gao was funded by the National Key Research and Development Program of China under grant 2022YFF0902800. S. Pei was partly supported by funding from the National Institutes of Health grant R01AI163023, Centers for Disease Control and Prevention grants U01CK000592 and 75D30122C14289 and National Science Foundation grant DMS-2229605.

Availability of data

(1) Case information data set. This data set contains the number of daily confirmed cases and daily deaths in the 50 states of the USA and Washington, DC from February 21, 2020 to December 11, 2020. The data was taken from Johns Hopkins University, and it can be downloaded from https://datahub.io/core/covid-19. (2) A data set of commuting between states in the USA. It was taken from the commuter flow data of the five-year U.S. Census survey from 2011 to 2015 (https://www.census.gov/data/tables/2015/demo/metro-micro/commuting-flows-2015.html). This data set is available in [25], which recorded the number of commuters between counties. From this data set, we collected the state-level population commuting data for model calibration before the announcement of extensive control measures on March 15, 2020. (3) Resident travel data set of 51 states and regions in the USA. The data set was taken from the Foursquare Labs Inc laboratory in which the time, duration and location of users are determined by mobile device signals [27]. (4) Population data set. This data set contains the total population of each state in the USA and the age composition data of the population, which can be obtained from USAFacts (https://usafacts.org/) .

References

1. *World Health Organization*, Coronavirus disease 2019 (COVID-19): Situation report, 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
2. *World Health Organization*, WHO Coronavirus (COVID-19) Dashboard, 2021. Available from: https://covid19.who.int 2021
3. M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, et al., First case of 2019 novel coronavirus in the United States, *N. Engl. J. Med.*, **382** 2020, 929–936. https://doi.org/10.1056/NEJMoa2001191
4. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, *Lancet Infect. Dis.*, **20** (2020), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1
5. A. L. Rasmussen, S. V. Popescu, SARS-CoV-2 transmission without symptoms, *Science*, **371** (2021) , 1206–1207. https://doi.org/10.1126/science.abf9569
6. L. Tian, X. Li, F. Qi, Q. Y. Tang, V. Tang, J. Liu, et al., Harnessing peak transmission around symptom onset for non-pharmaceutical intervention and containment of the COVID-19 pandemic, *Nat. Commun.*, **12** (2021), 1147. https://doi.org/10.1038/s41467-021-21385-z
7. D. P. Oran, E. J. Topol, Prevalence of asymptomatic SARS-CoV-2 infection, *Ann. Int. Med.*, **174** (2021), 286–287. https://doi.org/10.7326/L20-1285
8. S. W. Park, D. M. Cornforth, J. Dushoff, J. S. Weitz, The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak, *Epidemics*, **31** (2020), 100392. https://doi.org/10.1016/j.epidem.2020.100392
9. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020, *Eurosurveillance*, **25** (2020), 2000180.

10. J. F. W. Chan, S. Yuan, K. H. Kok, K. K. W. To, H. Chu, J. Yang, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster, *Lancet*, **395** (2020), 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9

11. L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, et al., Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing, *Science*, **368** (2020), eabb6936. https://doi.org/10.1126/science.abb6936

12. Z. D. Tong, A. Tang, K. F. Li, P. Li, H. L. Wang, J. P. Yi, et al., Potential presymptomatic transmission of SARS-CoV-2, Zhejiang province, China, 2020, *Emerg. Infect. Dis.*, **26** (2020), 1052. https://doi.org/10.3201%2Feid2605.200198

13. S. J. Krieg, J. J. Schnur, M. L. Miranda, M. E. Pfrender, N. V. Chawla, Symptomatic, presymptomatic, and asymptomatic transmission of SARS-CoV-2 in a university student population, August—November 2020, *Public Health Rep.*, **137** (2022), 1023–1030. https://doi.org/10.1177/00333549221110300

14. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, et al., Temporal dynamics in viral shedding and transmissibility of COVID-19, *Nat. Med.*, **26** (2020), 672–675. https://doi.org/10.1038/s41591-020-0869-5

15. D. Buitrago-Garcia, D. Egli-Gany, M. J. Counotte, S. Hossmann, H. Imeri, A. M. Ipekci, et al., Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis, *PLoS Med.*, **17** (2020), e1003346. https://doi.org/10.1371/journal.pmed.1003346

16. S. M. Moghadas, M. C. Fitzpatrick, P. Sah, A. Pandey, A. Shoukat, B. H. Singer, et al., The implications of silent transmission for the control of COVID-19 outbreaks, *Proc. Natl. Acad. Sci. USA*, **117** (2020, 17513–17515. https://doi.org/10.1073/pnas.2008373117

17. L. C. Tindale, J. E. Stockdale, M. Coombe, E. S. Garlock, W. Y. V. Lau, M. Saraswat, et al., Evidence for transmission of COVID-19 prior to symptom onset, *Elife*, **9** (2020), e57149. https://doi.org/10.7554/eLife.57149

18. Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, L. A. Meyers, Serial interval of COVID-19 among publicly reported confirmed cases, *Emerg. Infect. Dis.*, **26** (2020), 1341. https://doi.org/10.3201/eid2606.200357

19. S. A. McDonald, F. Miura, E. R. A. Vos, M. van Boven, H. E. de Melker, F. R. M. van der Klis, et al., Estimating the asymptomatic proportion of SARS-CoV-2 infection in the general population: Analysis of nationwide serosurvey data in the Netherlands, *Eur. J. Epidemiol.*, **36** (2021), 735–739. https://doi.org/10.1007/s10654-021-00768-y

20. W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, V. J. Lee, Presymptomatic transmission of SARS-CoV-2—Singapore, january 23—march 16, 2020, *Morb. Mortal. Wkly. Rep.*, **69** (2020), 411. https://doi.org/10.3724/SP.J.1123.2021.03024

21. C. Martínez, H. Serrano-Coll, Á. Faccini, V. Contreras, K. Galeano, Y. Botero, et al., SARS-CoV-2 in a tropical area of Colombia, a remarkable conversion of presymptomatic to symptomatic people impacts public health, *BMC Infect. Dis.*, **22** (2022), 1–5. https://doi.org/10.1186/s12879-022-07575-0

22. R. Subramanian, Q. He, M. Pascual, Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity, *Proc. Natl. Acad. Sci. USA*, **118** (2021), e2019716118. https://doi.org/10.1073/pnas.2019716118

23. X. Hao, S. Cheng, D. Wu,T. Wu, X. Lin, C. Wang, Reconstruction of the full transmission dynamics of COVID-19 in Wuhan, *Nature*, **584** (2020), 420–424. https://doi.org/10.1038/s41586-020-2554-8

24. R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, et al., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), *Science*, **368** (2020), 489–493. https://doi.org/10.1126/science.abb3221

25. M. Monod, A. Blenkinsop, X. Xi, D. Hebert, S. Bershan, S. Tietze, et al., Age groups that sustain resurging COVID-19 epidemics in the United States, *Science*, **371** (2021), eabe8372. https://doi.org/10.1126/science.abe8372

26. S. Pei, S. Kandula, W. Yang, J. Shaman, Forecasting the spatial transmission of influenza in the United States, *Proc. Natl. Acad. Sci. USA*, **115** (2018), 2752–2757. https://doi.org/10.1073/pnas.1708856115

27. S. Pei, S. Kandula, J. Shaman, Differential effects of intervention timing on COVID-19 spread in the United States, *Sci. Adv.*, **6** (2020), eabd6370. https://doi.org/10.1126/sciadv.abd6370

28. J. L. Anderson, An ensemble adjustment kalman filter for data assimilation, *Mon, Weather Rev.*, **129** (2001), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129¡2884:AEAKFF¿2.0.CO;2

29. J. Shaman, A. Karspeck, Forecasting seasonal outbreaks of influenza, *Proc. Natl. Acad. Sci. USA*, **109** (2012), 20425–20430. https://doi.org/10.1073/pnas.1208772109

30. S. Pei, T. K. Yamana, S. Kandula, M. Galanti, J. Shaman, Overall burden and characteristics of COVID-19 in the United States during 2020, preprint, medRxiv: 2021.02. https://doi.org/10.1101/2021.02.15.21251777

31. M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Process.*, **50** (2002), 174–188. https://doi.org/10.1109/78.978374

32. C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering, *Mon. Weather Rev.*, **136** (2008), 4629–4640. https://doi.org/10.1175/2008MWR2529.1

33. W. Yang, M. Lipsitch, J. Shaman, Inference of seasonal and pandemic influenza transmission dynamics, *Proc. Natl. Acad. Sci. USA*, **112** (2015), 2723–2728. https://doi.org/10.1073/pnas.1415012112

34. O. Diekmann, J. Heesterbeek, M. G. Roberts, The construction of next-generation matrices for compartmental epidemic models, *J. R. Soc. Interface*, **7** (2010), 873–885. https://doi.org/10.1098/rsif.2009.0386

35. J. L. Daly, B. Simonetti, K. Klein, K. E. Chen, M. K. Williamson, C. Antón-Plágaro, et al., Neuropilin-1 is a host factor for SARS-CoV-2 infection, *Science*, **370** (2020), 861–865. https://doi.org/10.1126/science.abd3072

36. S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, R. Ke, High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2, *Emerg. Inf. Dis.*, **26** (2020), 1470. https://doi.org/10.3201%2Feid2607.200282
37. S. Lai, N. W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J. R. Floyd, et al., Effect of non-pharmaceutical interventions to contain COVID-19 in China, *Nature*, **585** (2020), 410–413. https://doi.org/10.1038/s41586-020-2293-x
38. M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, *Science*, **368** (2020), 395–400. https://doi.org/10.1126/science.aba9757
39. H. Tian, Y. Liu, Y. Li, C. H. Wu, B. Chen, M. U. Kraemer, et al., An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China, *Science*, **368** (2020), 638–642. https://doi.org/10.1126/science.abb6105
40. B. F. Maier, D. Brockmann, Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China, *Science*, **368** (2020), 742–746. https://doi.org/10.1126/science.abb4557
41. J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, et al., Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China, *Science*, **368** (2020), 1481–1486. https://doi.org/10.1126/science.abb8001
42. M. U. Kraemer, C. H. Yang, B. Gutierrez, C. H. Wu, B. Klein, D. M. Pigott, et al., The effect of human mobility and control measures on the COVID-19 epidemic in China, *Science*, **368**, (2020), 493–497. https://doi.org/10.1126/science.abb4218
43. J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, V. Priesemann, Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions, *Science*, **369** (2020), eabb9789. https://doi.org/10.1126/science.abb9789
44. M. Du, Contact tracing as a measure to combat COVID-19 and other infectious diseases, *Am. J. Infect. Control*, **50** (2022), 638–644. https://doi.org/10.1016/j.ajic.2021.11.031
45. S. Liu, T. Yamamoto, Role of stay-at-home requests and travel restrictions in preventing the spread of COVID-19 in Japan, *Transp. Res. Part A*, **159** (2022), 1–16. https://doi.org/10.1016/j.tra.2022.03.009
46. N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, R. M. Eggo, Age-dependent effects in the transmission and control of COVID-19 epidemics, *Nat. Med.*, **26** (2020), 1205–1211. https://doi.org/10.1038/s41591-020-0962-9
47. A. James, M. J. Plank, R. N. Binny, A. Lustig, K. Hannah, S. Hendy, N. Steyn, A structured model for COVID-19 spread: Modelling age and healthcare inequities, *Math. Med. Biol.*, **38** (2021), 299–313. https://doi.org/10.1093/imammb/dqab006
48. E. A. Undurraga, G. Chowell, K. Mizumoto, COVID-19 case fatality risk by age and gender in a high testing setting in Latin America: Chile, March–August 2020, *Infect. Dis. Poverty*, **10** (2021), 1–11. https://doi.org/10.1186/s40249-020-00785-1

49. R. Verity, L. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, et al., Estimates of the severity of coronavirus disease 2019: A model-based analysis, *Lancet Infect. Dis.*, **20** (2020), 669–677. https://doi.org/10.1016/S1473-3099(20)30243-7

Appendix

A. SEPAIR model

The daytime metapopulation SEPAIR model is expressed as follows:

$$\begin{aligned} S_{ij}(t+dt_1) = S_{ij}(t) &- \frac{\beta_i S_{ij}(t)\sum_k I_{ki}(t)}{N_i^d(t)}dt_1 - \frac{\mu\beta_i S_{ij}(t)\sum_k A_{ik}(t)}{N_i^d(t)}dt_1 \\ &- \frac{m\beta_i S_{ij}(t)\sum_k P_{ik}(t)}{N_i^d(t)}dt_1 \\ &+ \theta dt_1 \frac{N_{ij} - I_{ij}(t)}{N_i^d(t)} \sum_{k\neq i} \frac{\bar{N}_{ik}\sum_l S_{kl}(t)}{N_k^d(t) - \sum_l I_{lk}(t)} \\ &- \theta dt_1 \frac{S_{ij}(t)}{N_i^d(t) - \sum_l I_{li}(t)} \sum_{k\neq i} \bar{N}_{ki} \end{aligned} \tag{A.1}$$

$$\begin{aligned} E_{ij}(t+dt_1) = E_{ij}(t) &+ \frac{\beta_i S_{ij}(t)\sum_k I_{ki}(t)}{N_i^d(t)}dt_1 + \frac{\mu\beta_i S_{ij}(t)\sum_k A_{ik}(t)}{N_i^d(t)}dt_1 \\ &+ \frac{m\beta_i S_{ij}(t)\sum_k P_{ik}(t)}{N_i^d(t)}dt_1 - \frac{E_{ij}(t)}{Z_1}dt_1 \\ &+ \theta dt_1 \frac{N_{ij} - I_{ij}(t)}{N_i^d(t)} \sum_{k\neq i} \frac{\bar{N}_{ik}\sum_l E_{kl}(t)}{N_k^d(t) - \sum_l I_{lk}(t)} \\ &- \theta dt_1 \frac{E_{ij}(t)}{N_i^d(t) - \sum_l I_{li}(t)} \sum_{k\neq i} \bar{N}_{ki} \end{aligned} \tag{A.2}$$

$$\begin{aligned} P_{ij}(t+dt_1) = P_{ij}(t) &+ \frac{E_{ij}(t)}{Z_1}dt_1 - \frac{P_{ij}(t)}{Z_2}dt_1 \\ &+ \theta dt_1 \frac{N_{ij} - I_{ij}(t)}{N_i^d(t)} \sum_{k\neq i} \frac{\bar{N}_{ik}\sum_l P_{kl}(t)}{N_{ij}^d(t) - \sum_l I_{lk}(t)} \\ &- \theta dt_1 \frac{P_{ij}(t)}{N_i^d(t) - \sum_l I_{li}(t)} \sum_{k\neq i} \bar{N}_{ki} \end{aligned} \tag{A.3}$$

$$I_{ij}(t+dt_1) = I_{ij}(t) + \alpha_i \frac{P_{ij}(t)}{Z_2}dt_1 - \frac{I_{ij}(t)}{D}dt_1 \tag{A.4}$$

$$A_{ij}(t+dt_1) = A_{ij}(t) + (1-\alpha_i)\frac{P_{ij}(t)}{Z_2}dt_1 - \frac{A_{ij}(t)}{D}dt_1$$

$$+\theta dt_1 \frac{N_{ij} - I_{ij}(t)}{N_i^d(t)} \sum_{k\neq i} \frac{\bar{N}_{ik} \sum_l A_{kl}(t)}{N_k^d(t) - \sum_l I_{lk}(t)}$$
$$-\theta dt_1 \frac{A_{ij}(t)}{N_i^d(t) - \sum_l I_{li}(t)} \sum_{k\neq i} \bar{N}_{ki} \tag{A.5}$$

$$N_i^d(t) = \sum_k N_{ik} \tag{A.6}$$

Similarly, the nighttime metapopulation SEPAIR model is expressed as follows:

$$S_{ij}(t+1) = S_{ij}(t+dt_1) - \frac{\beta_j S_{ij}(t+dt_1) \sum_k I_{kj}(t+dt_1)}{N_j^n} dt_2$$
$$-\frac{\mu\beta_j S_{ij}(t+dt_1) \sum_k A_{kj}(t+dt_1)}{N_j^n} dt_2$$
$$-\frac{m\beta_j S_{ij}(t+dt_1) \sum_k P_{kj}(t+dt_1)}{N_j^n} dt_2$$
$$+\theta dt_2 \frac{N_{ij}}{N_j^n} \sum_{k\neq j} \frac{\bar{N}_{jk} \sum_l S_{lk}(t+dt_1)}{N_k^n - \sum_l I_{lk}(t+dt_1)}$$
$$-\theta dt_2 \frac{S_{ij}(t+dt_1)}{N_j^n - \sum_k I_{kj}(t+dt_1)} \sum_{k\neq j} \bar{N}_{kj} \tag{A.7}$$

$$E_{ij}(t+1) = E_{ij}(t+dt_1) + \frac{\beta_j S_{ij}(t+dt_1) \sum_k I_{kj}(t+dt_1)}{N_j^n} dt_2$$
$$+\frac{\mu\beta_j S_{ij}(t+dt_1) \sum_k A_{kj}(t+dt_1)}{N_j^n} dt_2$$
$$+\frac{m\beta_j S_{ij}(t+dt_1) \sum_k P_{kj}(t+dt_1)}{N_j^n} dt_2 - \frac{E_{ij}(t+dt_1)}{Z_1} dt_2$$
$$+\theta dt_2 \frac{N_{ij}}{N_j^n} \sum_{k\neq j} \frac{\bar{N}_{jk} \sum_l E_{lk}(t+dt_1)}{N_k^n - \sum_l I_{lk}(t+dt_1)}$$
$$-\theta dt_2 \frac{E_{ij}(t+dt_1)}{N_j^n - \sum_k I_{kj}(t+dt_1)} \sum_{k\neq j} \bar{N}_{kj} \tag{A.8}$$

$$P_{ij}(t+1) = P_{ij}(t+dt_1) + \frac{E_{ij}(t+dt_1)}{Z_1} dt_2 - \frac{P_{ij}(t+dt_1)}{Z_2} dt_2$$
$$+\theta dt_2 \frac{N_{ij}}{N_j^n} \sum_{k\neq j} \frac{\bar{N}_{jk} \sum_l P_{lk}(t+dt_1)}{N_k^n - \sum_l I_{lk}(t+dt_1)}$$
$$-\theta dt_2 \frac{P_{ij}(t+dt_1)}{N_j^n - \sum_k I_{kj}(t+dt_1)} \sum_{k\neq j} \bar{N}_{kj} \tag{A.9}$$

$$I_{ij}(t+1) = I_{ij}(t+dt_1) + \alpha_j \frac{P_{ij}(t+dt_1)}{Z_2} dt_2 - \frac{I_{ij}(t+dt_1)}{D} dt_2 \tag{A.10}$$

$$\begin{aligned} A_{ij}(t+1) = A_{ij}(t+dt_1) + \left(1-\alpha_j\right) \frac{P_{ij}(t+dt_1)}{Z_2} dt_2 - \frac{A_{ij}(t+dt_1)}{D} dt_2 \\ +\theta dt_2 \frac{N_{ij}}{N_j^n} \sum_{k \neq j} \frac{\bar{N}_{jk} \sum_l A_{lk}(t+dt_1)}{N_k^n - \sum_l I_{lk}(t+dt_1)} \\ -\theta dt_2 \frac{A_{ij}(t+dt_1)}{N_j^n - \sum_k I_{kj}(t+dt_1)} \sum_{k \neq j} \bar{N}_{kj} \end{aligned} \tag{A.11}$$

$$N_i^n(t) = \sum_k N_{ki} \tag{A.12}$$

where $S_{ij}, E_{ij}, P_{ij}, I_{ij}, A_{ij}$ and N_{ij} are the numbers of susceptible individuals, exposed cases, presymptomatic cases, reported cases and asymptomatic or mildly symptomatic cases, as well as the total population living in state j and working in state i respectively. Z_1, D, μ, θ, m and Z_2 represent the latent period (non-infectious), the average duration of infection, the relative transmission rate of asymptomatic or mildly symptomatic cases, the adjustment parameter for population mobility, the relative transmission rate in the presymptomatic stage and the contagious incubation period in the presymptomatic stage (asymptomatic but contagious) respectively. dt_1 and dt_2 represent the duration of day and night, which are 8 hours and 16 hours, respectively. Since we assumed that the reported individuals are not allowed to travel to the state where they work, in the daytime, S_{ij} number of susceptible individuals who work in state i can only be infected by I_{ki} number of reported individuals who live in state i, A_{ik} number of asymptomatic or mildly symptomatic individuals who work in state i and P_{ik} number of presymptomatic individuals who work in state i and then become exposed individuals. The expected numbers for those exposed individuals are $\frac{\beta_i S_{ij}(t) \sum_k I_{ki}(t)}{N_i^d(t)}$, $\frac{\mu \beta_i S_{ij}(t) \sum_k A_{ik}(t)}{N_i^d(t)}$ and $\frac{m \beta_i S_{ij}(t) \sum_k P_{ik}(t)}{N_i^d(t)}$, respectively. Besides population mobility due to work commuting, we also consider random mobility. Random mobility has also been considered in the modeling of the COVID-19 transmission in previous studies [27]. Here we apply it the same way in our model. On the right-hand side of Eq (A.1), $\theta dt_1 \frac{N_{ij} - I_{ij}(t)}{N_i^d(t)} \sum_{k \neq i} \frac{\bar{N}_{ik} \sum_l S_{kl}(t)}{N_k^d(t) - \sum_l I_{lk}(t)}$ denotes the augmentation in number of susceptible people due to random mobility while $\theta dt_1 \frac{S_{ij}(t)}{N_i^d(t) - \sum_l I_{li}(t)} \sum_{k \neq i} \bar{N}_{ki}$ denotes the decrease in number of susceptible people due to random mobility.

B. Model calibration

The SEPAIR model generates the number of daily new reported cases and daily new deaths for each state. To map infections to deaths, we used an age-stratified infection fatality rate (IFR) and computed the IFR for each state as a weighted average by using demographic information on local age structure [27,49]. Data can be obtained from USAFacts (Https://usafacts.org/). To account for reporting delays, we mapped the simulated reported infections to reported cases by using a separate observational delay

model. In this delay model, we account for the time interval between a person transitioning from latent infection to symptom onset (i.e., $E \rightarrow I$) as well as for observational confirmation of that individual infection. To estimate this delay period, we applied a gamma distribution (a = 1.85, b = 3.24, mean = 6 days) for the confirmation of infection and a gamma distribution (a = 1.85, b = 11.35, mean = 21 days) to describe infections leading to death in the grid search method [27].

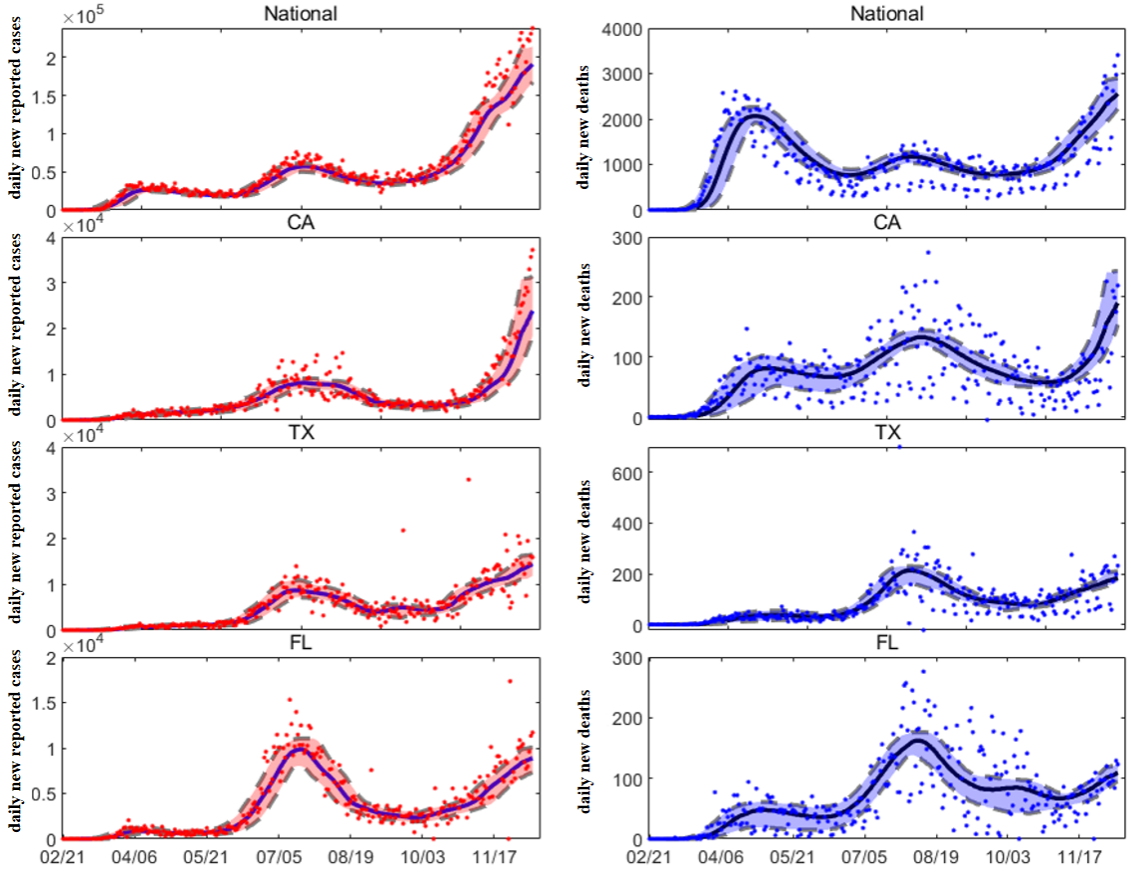

Figure B1. Model fitting for the numbers of daily new reported cases and daily new deaths in the USA, California, Texas and Florida. The red points represent the observed values of number of daily new reported cases, the blue points represent the observed values of number of daily new death cases and the curve and the filled area represent the median and the 95% confidence interval, respectively.

From Figure B1, we find that the SEPAIR model captures well the dynamics of the number of daily new reported cases and daily new deaths in the USA in 2020. There were three waves of the pandemic in the USA: the spring wave in April, the summer wave at the end of July and the more violent resurgence in winter. The number of reported cases in the USA briefly declined in the initial period due to a wide range of non-pharmaceutical interventions. However, there was a sustained recovery beginning in mid-2020, with new reported cases exceeding 100,000 in a single day on November 4, 2020. As of December 11, 2020, the six states with the largest numbers of reported cases in the USA were California (CA), Texas (TX), Florida (FL), Illinois (IL), New York (NY) and Ohio (OH); the six states with the most deaths were New York (NY), Texas (TX), California (CA),

Florida (FL), New Jersey (NJ) and Illinois (IL). Here, we present the simulation results for California, Texas, Florida, Illinois, New York and New Jersey in Figure B1, which shows the differences in the positions of the peaks of pandemic waves among these states. Specifically, California, Texas and Florida saw their first wave in the summer, and for New York and New Jersey, the first wave came in the spring. The number of deaths in the states that ushered in the peak of the diagnosis in the spring is very large, and it may be related to the shortage of medical resources in the early stage of the pandemic (Figure B2).

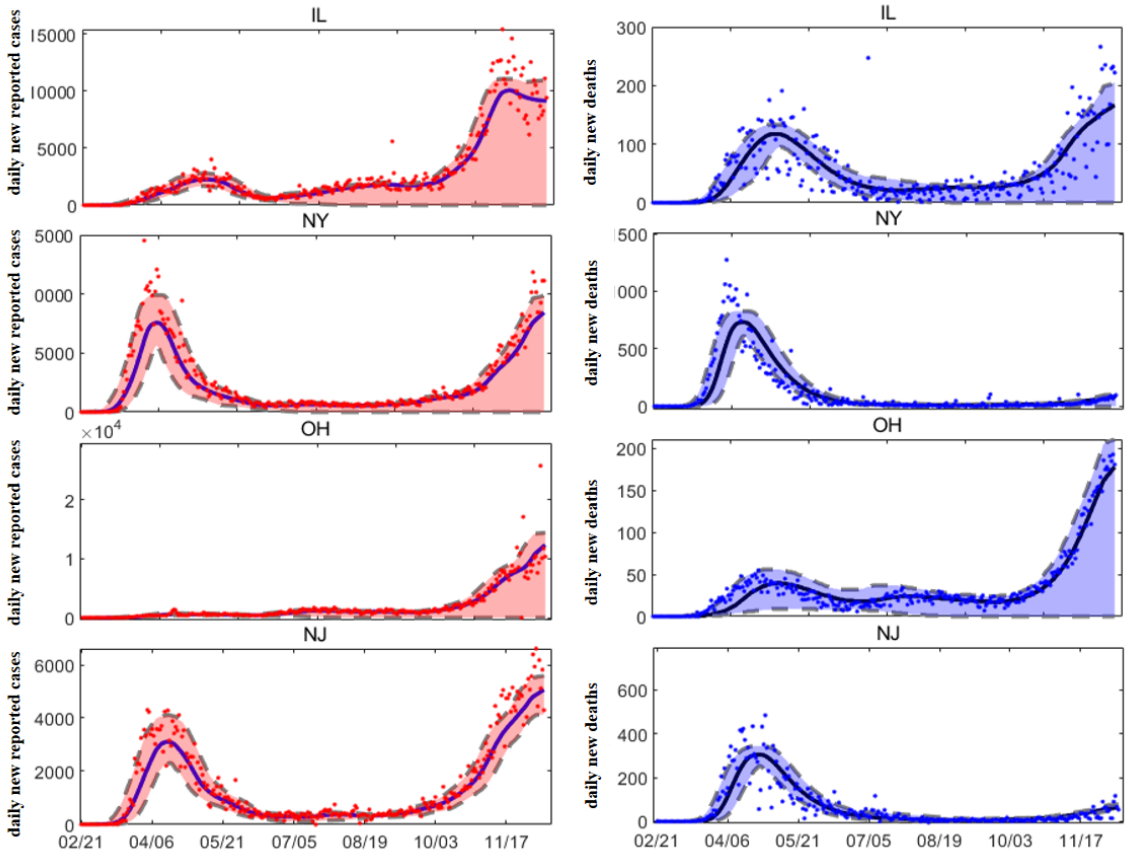

Figure B2. Model fitting for the number of daily new reported cases and daily new deaths in Illinois, New York, Ohio and New Jersey. The red points represent the observed values of number of daily new reported cases, the blue points represent the observed values of number of daily new death cases and the curve and the filled area represent the median and the 95% confidence interval, respectively.