

## Research



**Cite this article:** Wang Z *et al.* 2022

Assessing the asymptomatic proportion of SARS-CoV-2 infection with age in China before mass vaccination. *J. R. Soc. Interface* **19**: 20220498.

<https://doi.org/10.1098/rsif.2022.0498>

Received: 8 July 2022

Accepted: 22 September 2022

### Subject Category:

Life Sciences—Mathematics interface

### Subject Areas:

biomathematics, computational biology

### Keywords:

SARS-CoV-2, asymptomatic infection, age-stratified compartment model

### Authors for correspondence:

Huaiyu Tian

e-mail: [tianhuaiyu@gmail.com](mailto:tianhuaiyu@gmail.com)

Christopher Dye

e-mail: [christopher.dye@biology.ox.ac.uk](mailto:christopher.dye@biology.ox.ac.uk)

†These authors contributed equally to the study.

‡These authors jointly supervised this work.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6238061>.

# Assessing the asymptomatic proportion of SARS-CoV-2 infection with age in China before mass vaccination

Zengmiao Wang<sup>1,†</sup>, Peiyi Wu<sup>1,†</sup>, Jingyuan Wang<sup>2,3,†</sup>, José Lourenço<sup>4</sup>, Bingying Li<sup>1</sup>, Benjamin Rader<sup>5,6</sup>, Marko Laine<sup>7</sup>, Hui Miao<sup>8</sup>, Ligu Wang<sup>9</sup>, Hongbin Song<sup>9</sup>, Nita Bharti<sup>10</sup>, John S. Brownstein<sup>5,12</sup>, Ottar N. Bjornstad<sup>10,11,‡</sup>, Christopher Dye<sup>13,‡</sup> and Huaiyu Tian<sup>1,‡</sup>

<sup>1</sup>State Key Laboratory of Remote Sensing Science, Center for Global Change and Public Health, College of Global Change and Earth System Science, Beijing Normal University, Beijing, People's Republic of China

<sup>2</sup>School of Computer Science and Engineering, Beihang University, Beijing, People's Republic of China

<sup>3</sup>Peng Cheng Laboratory, Shenzhen, People's Republic of China

<sup>4</sup>Biosystems and Integrative Sciences Institute, University of Lisbon, Lisbon, Portugal

<sup>5</sup>Computational Epidemiology Lab, Boston Children's Hospital, Boston, MA, USA

<sup>6</sup>Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

<sup>7</sup>Meteorological Research Unit, Finnish Meteorological Institute, Helsinki, Finland

<sup>8</sup>Department of Statistics, College of Art and Science, Ohio State University, Columbus, OH, USA

<sup>9</sup>Center of Disease Control and Prevention, PLA, Beijing, People's Republic of China

<sup>10</sup>Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park, PA, USA

<sup>11</sup>Department of Entomology, College of Agricultural Sciences, Pennsylvania State University, University Park, PA, USA

<sup>12</sup>Harvard Medical School, Harvard University, Boston, MA, USA

<sup>13</sup>Department of Biology, University of Oxford, Oxford, UK

ZW, 0000-0001-6954-6069; JL, 0000-0002-9318-2581; BR, 0000-0002-6095-0193; ML, 0000-0002-5914-6747; NB, 0000-0003-1940-7794; JSB, 0000-0001-8568-5317; HT, 0000-0002-4466-0858

Some asymptomatic individuals carrying SARS-CoV-2 can transmit the virus and contribute to outbreaks of COVID-19. Here, we use detailed surveillance data gathered during COVID-19 resurgences in six cities of China at the beginning of 2021 to investigate the relationship between asymptomatic proportion and age. Epidemiological data obtained before mass vaccination provide valuable insights into the nature of pathogenicity of SARS-CoV-2. The data were collected by multiple rounds of city-wide PCR testing with contact tracing, where each patient was monitored for symptoms through the whole course of infection. The clinical endpoint (asymptomatic or symptomatic) for each patient was recorded (the pre-symptomatic patients were classified as symptomatic). We find that the proportion of infections that are asymptomatic declines with age (coefficient =  $-0.006$ , 95% CI:  $-0.008$  to  $-0.003$ ,  $p < 0.01$ ), falling from 42% (95% CI: 6–78%) in age group 0–9 years to 11% (95% CI: 0–25%) in age group greater than 60 years. Using an age-stratified compartment model, we show that this age-dependent asymptomatic pattern, together with the distribution of cases by age, can explain most of the reported variation in asymptomatic proportions among cities. Our analysis suggests that SARS-CoV-2 surveillance strategies should take account of the variation in asymptomatic proportion with age.

## 1. Introduction

The proportion of asymptomatic infections shows vast differences across the world, varying from 1.21% to 91.88% [1–3]. Once age-stratified, the fraction of asymptomatic cases also show substantial variation in age [4–6]. Due to the potential

threat of asymptomatic transmission in the community [7–10], numerous studies have been performed to decipher the underlying mechanism [11–21] and quantify the major drivers of heterogeneity in asymptomatic infection [22,23].

The main obstacles to understanding the drivers of asymptomatic transmission include assessment of asymptomatic status (for example, some studies define asymptomatic cases only based on the symptom status when testing [24,25]), testing bias [23,26] and sampling bias (for example, testing is more likely to be completed in hospitalized individuals [27,28]). Because passive surveillance is often limited to symptomatic cases, a substantial proportion of asymptomatic cases are expected to be underreported [29,30]. Asymptomatic transmission poses a great challenge to local governments and health systems in terms of control policy [22,29,31–35]. Therefore, it is necessary to identify the factors driving such heterogeneity in asymptomatic proportions to design more efficiently surveillance measures that include such subclass of the infected.

To address this heterogeneity, intensive, active surveillance of SARS-CoV-2 infection is needed. Detailed surveillance data collected during COVID-19 resurgences in six cities of China in 2021 provide a valuable opportunity to study this issue. Multiple rounds of population-level PCR testing in each city accelerated the speed of SARS-CoV-2 infection identification (the cases have a short time interval between infection and detection). Besides, it also improved the detection rate among the currently infected irrespective of clinical outcome. The clinical endpoint (asymptomatic or symptomatic) for each patient was obtained by detailed surveillance across the whole course of infection. The pre-symptomatic individuals are classified as symptomatic because they will develop symptoms during the course of infection. More importantly, epidemiological data prior to mass vaccination provide valuable insight for the nature of pathogenicity of SARS-CoV-2. Using this unique dataset, we explored the potential factors that may contribute to observed variations in the proportion of asymptomatic infections, and reported thereof. Our analysis has implications for understanding the role of asymptomatic individuals in COVID-19 transmission and can help guide future disease control policies that depend on the frequency of asymptomatic cases as a hidden subgroup of infections.

## 2. Methods

### 2.1. The age-dependent asymptomatic proportion and the total asymptomatic proportion

Based on the observed data, the asymptomatic proportion is dependent on age. Let  $p_1, \dots, p_7$  represent the asymptomatic proportion of detected cases in age groups 0–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60+, respectively. So, the total asymptomatic proportion ( $p$ ) is calculated as follows:

$$p = \frac{\sum_{i=1}^G p_i N_i}{\sum_{i=1}^G N_i}, \quad \text{where } N_i \text{ is the number of cases in age group } i.$$

The total asymptomatic proportion is a combination of the age-dependent asymptomatic proportion and the SARS-CoV-2 case distribution in age.

### 2.2. Age-stratified compartment model fit to detected SARS-CoV-2 data

Given the stochasticity and heterogeneity in the transmission of SARS-CoV-2 [36,37], we developed an age-stratified discrete

stochastic compartment model that incorporated asymptomatic and symptomatic cases. Specifically, we consider susceptible ( $S_i$ ), latent ( $L_i$ ), pre-symptomatic ( $L_{Pi}$ ), infectious asymptomatic ( $I_{Ai}$ ), infectious symptomatic ( $I_{Si}$ ), removed symptomatic ( $R_{Si}$ ) and removed asymptomatic ( $R_{Ai}$ ) individuals for age group  $i$ ,  $i = 1, \dots, G$ . The latent compartment represents the individuals who are exposed to the virus but not contagious. The pre-symptomatic compartment represents the individuals who spread the virus before they develop symptoms, which is a feature of SARS-CoV-2. Note that control measures (population-level PCR tests and contact tracing) were implemented, so infected individuals will be identified through population-level testing and contact tracing. Therefore, they will be quarantined and exit the transmission chain. Isolation is the dominant factor in removed status. We assume that a constant proportion ( $\theta$ ) of infected people exit from compartments  $L_i$ ,  $L_{Pi}$ ,  $I_{Ai}$  and  $I_{Si}$ . Individuals receiving PCR tests will stay in the  $S$  compartment if the test results are negative. Individuals exiting from  $L_i$  compartment may develop symptoms with proportion of  $1 - p_i$ , where  $p_i$  is the proportion of  $L_i$  that becomes  $I_{Ai}$  in age group  $i$ . We did not consider the Ct value from the PCR tests and assumed that the PCR tests will be positive if the individual is infected. The model structure and the parameters involved are shown in electronic supplementary material, figure S1. The full set of equations representing the transmission for age group  $i$  is given by

$$\begin{aligned} S_i(t+1) &= S_i(t) - \beta \mu_i S_i(t) \sum_{j=1}^G \frac{C_{ij} \{I_{Sj}(t) + rI_{Aj}(t) + kL_{Pj}(t)\}}{N_j}, \\ L_i(t+1) &= L_i(t) + \beta \mu_i S_i(t) \sum_{j=1}^G \frac{C_{ij} \{I_{Sj}(t) + rI_{Aj}(t) + kL_{Pj}(t)\}}{N_j} \\ &\quad - \theta L_i(t) - \sigma(1 - \theta) p_i L_i(t) - \sigma(1 - \theta)(1 - p_i) L_i(t), \\ I_{Ai}(t+1) &= I_{Ai}(t) + \sigma(1 - \theta) p_i L_i(t) - \theta I_{Ai}(t), \\ L_{Pi}(t+1) &= L_{Pi}(t) + \sigma(1 - \theta)(1 - p_i) L_i(t) - \theta L_{Pi}(t) \\ &\quad - \sigma_{\text{pre}}(1 - \theta) L_{Pi}(t), \\ I_{Si}(t+1) &= I_{Si}(t) + \sigma_{\text{pre}}(1 - \theta) L_{Pi}(t) - \theta I_{Si}(t). \end{aligned}$$

In this study,  $\sigma$  is the rate from  $L_i$  to  $L_{Pi}$  or  $I_{Ai}$ .  $\sigma_{\text{pre}}$  is the rate from  $L_{Pi}$  to  $I_{Si}$ . Therefore,  $1/\sigma$  is the latent period for asymptomatic cases and pre-symptomatic cases.  $1/\sigma_{\text{pre}}$  is the pre-symptomatic period for the symptomatic. Let  $\beta$  represent the transmission probability for the effective contact with infectious symptomatic individuals. We take the transmission probability for the effective contact with infectious asymptomatic individuals to be  $r\beta$  and the transmission probability for the effective contact with pre-symptomatic individuals to be  $k\beta$ .  $\mu_i$  is the susceptibility to infection for age-group  $i$  [38].  $C_{ij}$  is the number of contacts between age group  $i$  and age group  $j$  per day.  $G$  is the number of age groups. The values and priors for the parameters in the model are listed in electronic supplementary material, table S1. The prior information for  $p_i$  was set according to the observation in each city.

Previous studies have shown that the contacts among different age groups changed markedly before and during the Chinese lockdown [32,39]. Considering this fact, two contact matrices describing average contacts among age groups before and during lockdown (electronic supplementary material, table S2) were incorporated into the model. For each city, the baseline contact matrix is used before lockdown date and the lockdown contact matrix is used after lockdown date. Although population-level testing and contact tracing were implemented in each city, the effectiveness probably differed among cities. To account for this, we used the location-specific proportion of infection detection by contact tracing and population-level testing (represented by parameter  $\theta$  in electronic supplementary material, figure S1). A pre-symptomatic stage was also included because a COVID-19 case can become infectious before developing symptoms. We have two sets of observations: the daily new asymptomatic infections and the daily new

symptomatic infections given by case onset. These observations correspond to the following equations:

$$R_S(t) = \sum_{i=1}^C R_{Si}(t), \quad \text{where } R_{Si}(t) = \theta I_{Si}(t) + \theta L_{Pi}(t) + \theta(1 - p_i)L_i(t),$$

$$R_A(t) = \sum_{i=1}^C R_{Ai}(t), \quad \text{where } R_{Ai}(t) = \theta I_{Ai}(t) + \theta p_i L_i(t).$$

In the stochastic model, we assume that the daily asymptomatic infections and the daily new symptomatic infections at time  $t$  followed the Poisson distribution with mean of  $R_S(t)$  and  $R_A(t)$ , respectively. We fitted  $R_S(t)$  and  $R_A(t)$  to the two sets of observations with 3-day rolling mean. Model fitting was performed using the Metropolis–Hastings Markov chain Monte Carlo algorithm with the MATLAB (v. R2020a) toolbox `mcmcstat` [40] (<https://github.com/mjlaine/mcmcstat>). One hundred thousand iterations were set for burn-in. After that, another 100 000 iterations were performed. For Shijiazhuang and Harbin, the onset date for the first case was a few days before the date of confirmation (i.e. the date non-pharmaceutical interventions (NPIs) were launched), and  $\theta$  was set to be zero until the confirmation date for the first case. The fitting results are shown in electronic supplementary material, table S3.

### 2.3. The SARS-CoV-2 case data for six cities of China

We collected data on SARS-CoV-2 infections from six cities in China (Shijiazhuang and Xingtai from Hebei province, Changchun and Tonghua from Jilin province, Harbin and Suihua from Heilongjiang province). The population size for the six cities is approximately 9.8 million, 8.0 million, 8.6 million, 2.1 million, 9.5 million and 5.2 million, respectively. To detect all ongoing infections as early as possible, the cities launched population-level PCR tests with more than 48 million tests in total (the actual number of tests should be higher than this since not all the information about the number of PCR tests for each round of population-level testing could be collected for this study); for example, Shijiazhuang performed three rounds of testing and Xingtai conducted as many as 10 rounds (electronic supplementary material, table S4).

For each city, the detailed surveillance data were collected from local Centers for Disease Control and Prevention, including the gender, age and onset date. For each patient, the clinical endpoint (asymptomatic or symptomatic) was also collected over a multi-week period.

### 2.4. The mobility data

In China, human movements were anonymously collected at the city level with mobile phone data, through location-based services (LBS) employed by the popular Baidu applications. We used relative volume of inflows movement for each city from the migration flows database (<http://qianxi.baidu.com/>). The study period for Shijiazhuang, Xingtai, Changchun, Tonghua, Harbin and Suihua was 23 December 2020–12 February 2021, 1 January 2021–24 January 2021, 10 January 2021–4 February 2021, 15 January 2021–9 February 2021, 6 January 2021–8 February 2021 and 9 January 2021–5 February 2021, respectively.

### 2.5. Age-dependent contact pattern

There are no specific contact matrices for the six cities in our study. To perform the analysis, the contact matrices for Wuhan, Shanghai, Changcha and Shenzhen before and after lockdown were downloaded from the studies of Zhang *et al.* [32,39]. The number in each cell of contact matrices before lockdown were averaged across four cities. There are 14 age groups in the original contact matrix. Considering that our data may be not able to support such detailed age groups, the averaged contact matrix was merged according to the age groups in our study (i.e. seven age groups). This is the baseline contact matrix. The same procedure

was applied to the contact matrices after lockdown. The lockdown contact matrix was also obtained. The cities of Shijiazhuang, Xingtai and Tonghua performed lockdown intervention, while other cities did not. To facilitate the comparison among these cities, we took the date with the lowest inner-city mobility as the lockdown date (electronic supplementary material, figure S2). Before this date, the baseline contact matrix (describing the contact rate among the different age groups) was used. After this date, the lockdown contact matrix was used. The date for lockdown and the date with the lowest inner-city mobility were closed (electronic supplementary material, table S5). Here, we used the same contact matrices (i.e. the baseline contact matrix and the lockdown contact matrix; electronic supplementary material, table S2) for each city in our study by assuming similar living habits across cities in China.

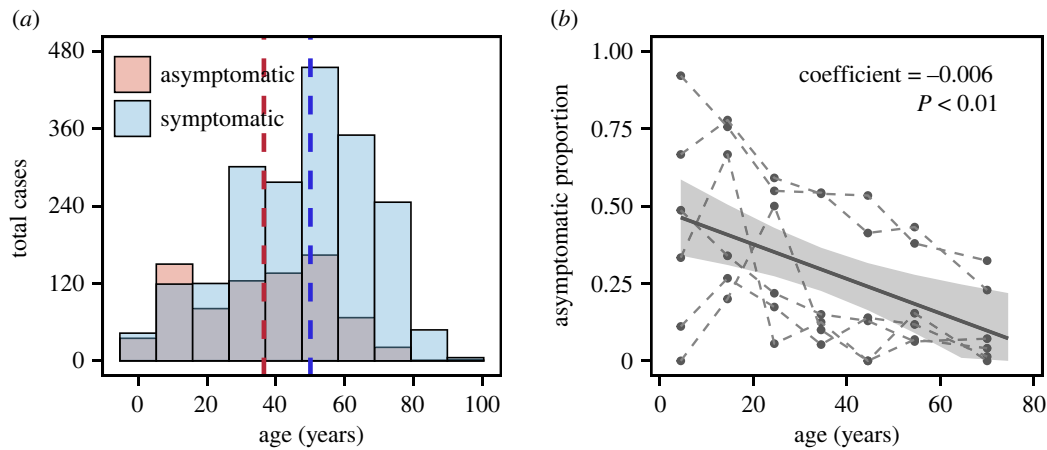
## 3. Results

### 3.1. The proportion of asymptomatic infections from multiple rounds of population-level PCR tests with patient follow-up

In order to understand heterogeneity in proportions of asymptomatic infections under demographic influence, we used detailed surveillance data gathered during COVID-19 resurgences in six cities of China at the beginning of 2021 to address this question. Previous studies show that local demography may play a role in the observed geographical and age-related heterogeneity of asymptomatic proportions (electronic supplementary material, figures S3–S5 and table S6). In this study, six cities in China (Shijiazhuang and Xingtai from Hebei province, Changchun and Tonghua from Jilin province, Harbin and Suihua from Heilongjiang province) were geographically close and had similar COVID-19 resurgences temporally, while manifesting heterogeneous asymptomatic proportions. All cities launched multi-round population-level PCR testing (applied to all the individuals in a city; electronic supplementary material, table S4) combined with detailed contact tracing and surveillance, making it a quick identification of SARS-CoV-2 infection regardless of whether displaying symptoms. Contact tracing was triggered if an individual was positive by PCR test, irrespective of the symptom status. So, the primary cases (being symptomatic or asymptomatic) from contact tracing were included in the analysis.

With detailed surveillance data including information on age distribution and clinical endpoint of each infection, we first looked into the statistical property of proportion of asymptomatic SARS-CoV-2 infection among different age groups. In total, there were 2744 PCR-confirmed SARS-CoV-2 infections with the highest number of 1040 in Shijiazhuang and the lowest number of 80 in Xingtai (electronic supplementary material, table S4). Among all SARS-CoV-2 infections, there were 1964 symptomatic infections with a median age of 50 and 780 asymptomatic infections with a median age of 36, indicating younger age groups are at risk of developing infectious but asymptomatic (figure 1a). The age distribution of infections varied by location: the largest proportion of positive cases was in individuals aged greater than 60 in Shijiazhuang, Changchun, Tonghua and Suihua, while the largest proportion was in age group of 20–29 for Xingtai and in age group of 40–49 for Harbin (figure 2).

Across the six cities, the proportion of asymptomatic infections declined with age (figure 1b; coefficient =  $-0.006$



**Figure 1.** Pattern of SARS-CoV-2 infections in six cities in China before mass vaccination. (a) The total asymptomatic infection (light red) and symptomatic infection (light blue) distribution from six cities (Shijiazhuang, Xingtai, Changchun, Tonghua, Harbin and Suihua) in age groups. The red dashed line represents the average of age in asymptomatic SARS-CoV-2 infections. The blue dashed line represents the average of age in symptomatic SARS-CoV-2 infections. (b) The grey dots connected by dashed line represent asymptomatic proportion from six Chinese cities in our study. A linear regression was built (asymptomatic proportion = intercept + coefficient  $\times$  age). The grey line represents the linear regression using pooled asymptomatic proportion under each age group across cities with 95% CI in light grey shadow (coefficient =  $-0.006$ ,  $p < 0.01$ ).

(95% CI:  $-0.008$  to  $-0.003$ ) and  $p < 0.01$  from linear regression). To consider the heterogeneity in each city, a meta-analysis was applied to the coefficients of age from each city. The asymptomatic proportion was still significantly associated with age (figure 3). The total or ‘crude’ asymptomatic proportion (i.e. the asymptomatic SARS-CoV-2 infections divided by all the SARS-CoV-2 infections for each city; see Methods) varied considerably, from 8% in Tonghua to 18% in Shijiazhuang and 51% in Harbin (white bars with black borders in figure 4b). The bias in this asymptomatic proportion calculation should be very small for the following reasons: (i) the populations were tested multiple times during the COVID-19 flare-ups and (ii) the identified cases were isolated and under 14-day health monitoring to create a definite clinical outcome (asymptomatic or symptomatic) per individual. For each age group, there was heterogeneity in the probability of being asymptomatic across the cities (figure 1b). Identification of the underlying biological mechanism of this heterogeneity is beyond the scope of the current study, as it mainly focuses on the overall, cross-city asymptomatic proportion across all ages.

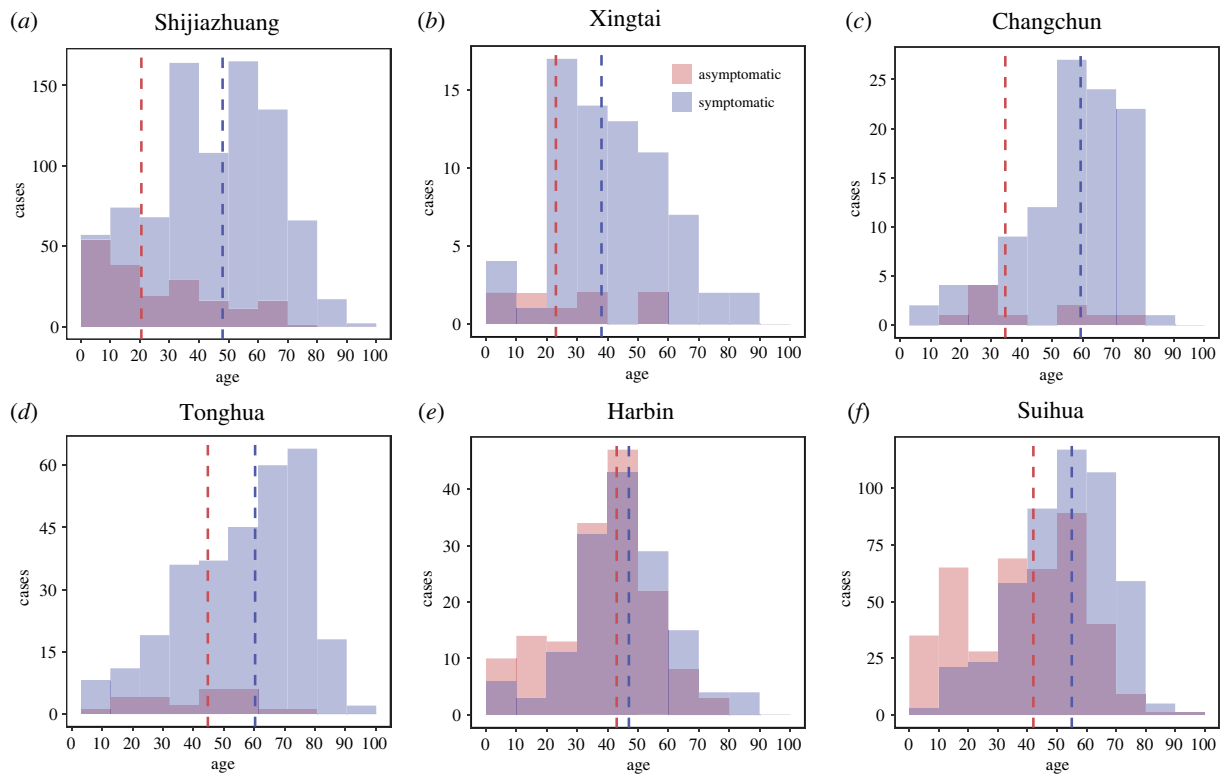
### 3.2. Age-dependent asymptomatic proportions and age distribution of cases contribute to the total asymptomatic proportion

To explore the mechanism underlying heterogeneity in overall, cross-city asymptomatic proportions, we developed an age-stratified compartment model that incorporated age-dependent asymptomatic proportions to reconstruct observed epidemic trajectories, including both asymptomatic and symptomatic cases (figure 4a; electronic supplementary material, figure S1). We hypothesized that cities report different total asymptomatic proportions of SARS-CoV-2 infections because of the interplay between the age-dependent asymptomatic proportion of cases and the age distribution of infections. The proportion of asymptomatic infections was assumed to decline with age, in

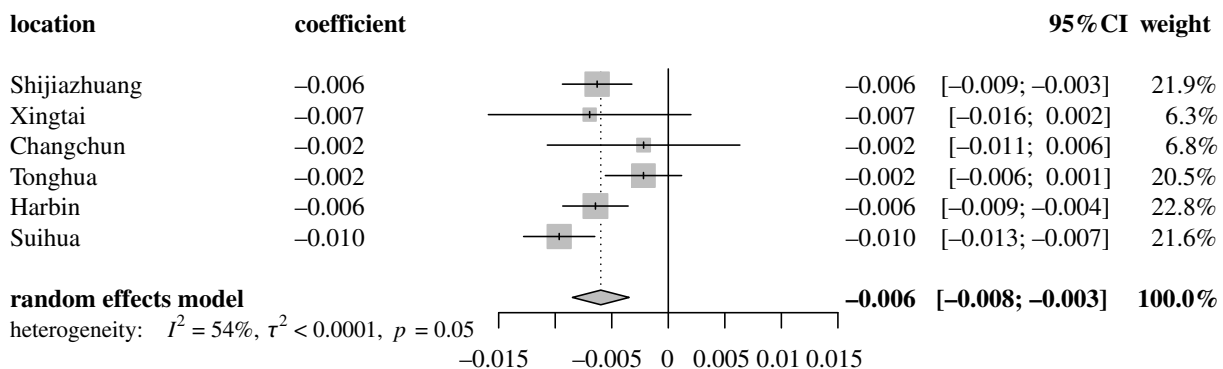
accordance to our findings and other reports [38]. The model accurately described the daily numbers of asymptomatic and symptomatic cases (electronic supplementary material, figure S6), the case distribution by age (figure 4b), the age-dependent proportion of asymptomatic cases (electronic supplementary material, figure S7), and the reported asymptomatic proportion of all cases (red bar in figure 4c) across all the cities (note that these were not involved in the loss function during fitting).

Based on the fit of the model, we next investigated the effect of the age-dependent asymptomatic proportions on total asymptomatic proportions. Our model suggested that if the proportion of asymptomatic cases had been the same across all age groups (i.e. age-independent), the total proportion of asymptomatic cases in each city would have been equal (yellow bar in figure 4c; electronic supplementary material, figure S8). However, this contradicted the observations and illustrated that the asymptomatic proportion depends on age, verifying that age-dependent asymptomatic proportions contribute to the total asymptomatic proportion.

Age distribution of cases might also have effects on the proportion of asymptomatic cases, and it will be affected by the interaction of initial infection seed among age groups, age structure and contact patterns of a city. We tested different combination of these factors by setting them to be either age-specific or homogeneous in age, but it had little effect on the total asymptomatic proportion (electronic supplementary material, figure S9). Then we considered a more extreme scenario where sustained transmission occurred only among the young population (contacts assumed only to take place among age groups (0–9) and (10–19)). In this case, the total asymptomatic proportion was predicted to be higher than observed (blue bar in figure 4b; electronic supplementary material, figure S10), because of the shifted age distribution of cases (electronic supplementary material, figure S11a–d). These results illustrate that transmission within specific age groups, such as school transmission and nursing home transmission, will greatly alter the total asymptomatic proportion.



**Figure 2.** The case distribution in age groups for symptomatic infection and asymptomatic infection. (a) Shijiazhuang, (b) Xingtai, (c) Changchun, (d) Tonghua, (e) Harbin and (f) Suihua. Symptomatic infection: light purple. Asymptomatic infection: light pink. The red dashed line represents the average of age in asymptomatic infections. The blue dashed line represents the average of age in symptomatic infections.



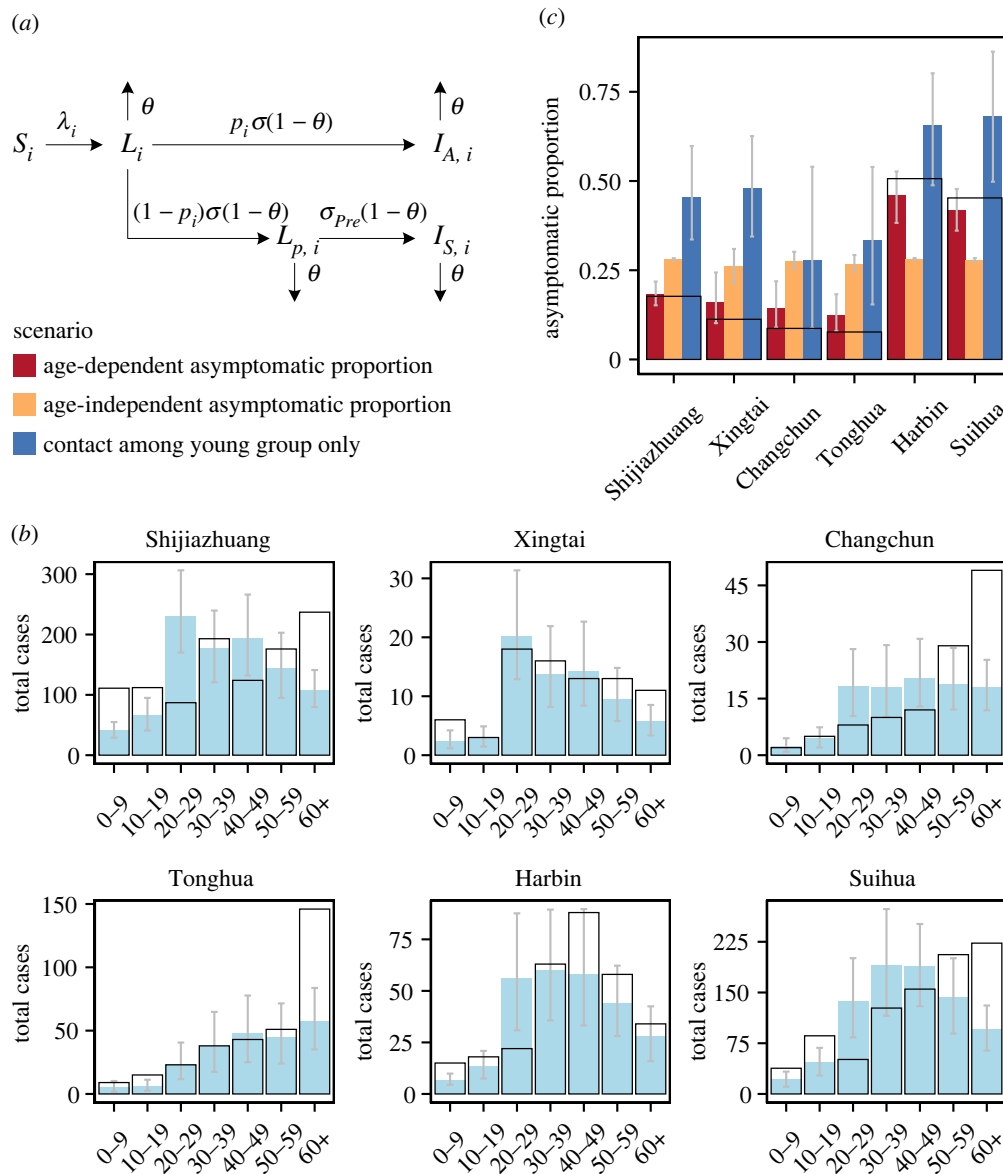
**Figure 3.** Pooled coefficient of age on the asymptomatic proportion in six Chinese cities. The coefficient was obtained from linear regression (asymptomatic proportion = intercept + coefficient  $\times$  age) built for each city in our study in figure 1b.

### 3.3. The sensitivity analysis for the six cities in China

As the susceptibility to SARS-CoV-2 infection is expected to increase with age [38], we also investigated the effect of susceptibility to infection on the total asymptomatic proportion. By the simulation with homogeneous susceptibility among different age groups, we found that the total asymptomatic proportion was higher than the observations. This is due to more cases happening among the younger groups (electronic supplementary material, figure S11*e,f*). We also replaced the contact matrix with the contact matrices from Wuhan, Shanghai, Changsha and Shenzhen in China [32,39] and the total asymptomatic proportion did not show dramatic changes (electronic supplementary material, figure S12). The age-dependent seed, age-structure and social contact matrices were also evaluated. Multiple simulations were performed by changing the age-dependent seed, age-structure and social contact matrices of the population to be homogeneous

in age groups. The results showed that the predicted total asymptomatic proportion may be relatively unchanged among these combinations of models (electronic supplementary material, figures S13–S20).

To illustrate the robustness of our model, we tested the effects of different initial values for parameters and the estimations were robust to the different initial values (electronic supplementary material, table S7). Considering other NPIs implemented, we used the inner-city mobility as a proxy to test whether the fitting was improved or not by assuming that the reduction in mobility has an effect on the transmission probability. Specifically, transmission probability ( $\beta$ ) is scaled by  $\beta^* \text{Mobility} / \text{max\_Mobility}$ , where  $\text{max\_Mobility}$  is the maximum value of mobility during the outbreak. However, the model performance showed limited improvement by adding human mobility information (electronic supplementary material, table S8).



**Figure 4.** The estimated SARS-CoV-2 cases distribution in age groups in six cities and the total asymptomatic proportions under different scenarios. (a) Model diagram. The following compartments are considered: susceptible ( $S_i$ ), latent state ( $L_i$ ), pre-symptomatic ( $L_{p,i}$ ), infectious asymptomatic ( $I_{A,i}$ ), infectious symptomatic ( $I_{S,i}$ ) in age group  $i$ .  $\lambda_i$  is the force of infection in age group  $i$ ;  $1/\sigma$  is the latent period for asymptomatic and pre-symptomatic;  $p_i$  is the proportion of infections that manifest as asymptomatic cases in age group  $i$ ;  $1/\sigma_{pre}$  is the pre-symptomatic period for symptomatic;  $\theta$  is the proportion of infection detected by contact tracing and population-level testing (see Methods). Population-level testing is applied to the individuals in a city, i.e. the individuals in all the compartments. (b) Estimate of the developed model to the total number of SARS-CoV-2 cases in age groups for six cities. The total number of SARS-CoV-2 cases reported (black box) and the estimated total number of cases (light blue box) in each city and each age group. The grey line represents 95% CI. (c) The total asymptomatic proportions under different scenarios. Red: simulated, based on the observed age-dependent asymptomatic proportions. Yellow: simulated, with age-independent asymptomatic proportions. This proportion was calculated by the total asymptomatic cases divided by the total SARS-CoV-2 cases using the pooled data from six cities. Blue: simulated, with contacts among the 0-9 and 10-19 age groups only. The grey bar represents the 95% CI for each simulation. White bars with black borders represent the observed total asymptomatic proportion in each city.

## 4. Discussion

Our study provides insight into the observed variability in the proportion of asymptomatic SARS-CoV-2 infections. Based on previous studies [6,26,41-43] and on data from the present study before mass vaccination, the proportion of cases that is asymptomatic declines with age. Together, local age-dependent asymptomatic proportions along with the age distribution of cases lead to differences in the asymptomatic proportions observed in different populations. Cities with younger populations are more likely to have a high proportion of asymptomatic infections and are at a high risk of COVID-19 endemic due to undetected asymptomatic transmission.

A great deal of concern has been directed toward the asymptomatic proportion among infected vaccinated people, which has been reported to be higher than that among the unvaccinated. However, vaccines could not have contributed to heterogeneity in the asymptomatic proportion in the six cities of our study because the vaccine roll-out had not started at the time of data collection. It is also unlikely that exposure during the first COVID-19 wave contributed to the heterogeneity of asymptomatic proportions in this dataset because the total number of SARS-CoV-2 infections during the first wave in 2020 among these six cities was small: ranging from 7 in Tonghua to 198 in Harbin. Additionally, a retrospective serological survey conducted in six

provinces or municipalities (of Hubei province) from 10 April 2020 to 18 April 2020 suggested that less than 0.1% of the population carried antibodies against SARS-CoV-2 [44].

There are a few limitations in our study. Due to the lack of city-specific contact matrices, we used averaged contact matrices from four other cities in China to represent contact patterns in our research assuming similar living habits across cities in China. The sensitivity analyses showed very similar results using city-specific contact matrices. Although some studies have reported that COVID-19 vaccines lower the risk of developing symptomatic disease [45–47], we did not consider this effect of vaccines. Few studies have reported proportions of asymptomatic infections by age group or otherwise following mass vaccination campaigns. In addition, our conclusion about the heterogeneity of asymptomatic proportion may persist once vaccinated, since the effectiveness of vaccines also depends on the age. Though active surveillance was conducted in six cities and most of the infected individuals were detected by population-level testing and contact tracing, there is heterogeneity in the proportion of asymptomatic SARS-CoV-2 infection within each age group across the cities. The underlying mechanism could not be unraveled based on the current dataset, although it may be caused by small sample sizes, other biological variations in the infected population, natural variation in clinical manifestations for SARS-CoV-2 infections or testing procedures in practice between the six cities (despite standard protocols being used in all six cities). More studies would be warranted to investigate this question in the future.

In our study, most infections were identified in local communities and isolation centres. Frequent, multiple rounds of testing and contact tracing accelerated the early detection of cases and even before they showed symptoms. Routine PCR tests were also used for COVID-19 surveillance among high-risk groups (e.g. health workers) and there were no reported SARS-CoV-2 infections before the first documented cases for each city of our study. By all these means, we were able to identify asymptomatic cases. Nevertheless, the chance of finding asymptomatic and symptomatic cases could have been different. Even if this is true, it would not change our conclusion that age is a key determinant of the asymptomatic proportion. Analysing the SARS-CoV-2 infections from population-level testing would be a way to demonstrate the robustness of asymptomatic proportion calculation. However, our dataset is not a line-list (a separate line for each case) and how each individual infection was identified was unknown. This is a limitation of data collection in our study.

In summary, our results give a comprehensive profile of asymptomatic infection in six cities of China and explain factors that contribute to heterogeneity in the asymptomatic proportion. Our study provides some insights for policy related asymptomatic infection surveillance and control of future resurgences by taking age structure into consideration, especially in the era of vaccination and variants such as the Omicron era.

**Ethics.** The epidemiological and clinical data collection was exempt from Institutional Review Boards, because it was part of a public health investigation for COVID-19, issued by the National Health Commission of the People's Republic of China (<http://www.nhc.gov.cn/jkj/s3577/202003/4856d5b0458141fa9f376853224d41d7.shtm>).

**Data accessibility.** All datasets are publicly available. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. Code and data are available at the following repository: [https://github.com/WWMWaggie/COVID-19\\_Role\\_Of\\_Age](https://github.com/WWMWaggie/COVID-19_Role_Of_Age).

The data are provided in electronic supplementary material [48].

**Authors' contributions.** Z.W.: conceptualization, formal analysis, methodology, writing—original draft, writing—review and editing; P.W.: data curation, formal analysis, methodology, visualization, writing—review and editing; J.W.: data curation; J.L.: writing—review and editing; B.L.: data curation; B.R.: writing—review and editing; M.L.: writing—review and editing; H.M.: data curation; L.W.: data curation; H.S.: data curation; N.B.: writing—review and editing; J.S.B.: writing—review and editing; O.N.B.: conceptualization, supervision, writing—review and editing; C.D.: conceptualization, supervision, writing—review and editing; H.T.: conceptualization, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

**Conflict of interest declaration.** The authors declare no competing interests.

**Funding.** This study was supported by the Scientific and Technological Innovation 2030—Major Project of New Generation Artificial Intelligence (2021ZD0111201); National Natural Science Foundation of China (82073616, 82161148011, 72171013, 82204160); National Key Research and Development Program of China; Open Fund of State Key Laboratory of Remote Sensing Science (OFSLRSS202106); Beijing Science and Technology Planning Project (Z201100005420010); Beijing Science and Technology Planning Project; Beijing Advanced Innovation Program for Land Surface Science; Key Scientific and Technology Project of Inner Mongolia Autonomous Region (2021ZD0006); Fundamental Research Funds for the Central Universities (2021NTST17). The funders had no role in study design, data collection and analysis, the decision to publish, or in preparation of the manuscript.

**Acknowledgements.** We thank the thousands of Centers for Disease Control (CDC) staff and local health workers in China who collected data and continue to work to contain COVID-19 in China and elsewhere.

## References

- Oran DP, Topol EJ. 2021 The proportion of SARS-CoV-2 infections that are asymptomatic. *Ann. Intern. Med.* **174**, 655–662. (doi:10.7326/M20-6976)
- Gao Z, Xu Y, Sun C, Wang X, Guo Y, Qiu S, Ma K. 2021 A systematic review of asymptomatic infections with COVID-19. *J. Microbiol. Immunol. Infect.* **54**, 12–16. (doi:10.1016/j.jmii.2020.05.001)
- Nishiura H *et al.* 2020 Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int. J. Infect. Dis.* **94**, 154–155. (doi:10.1016/j.ijid.2020.03.020)
- He J, Guo Y, Mao R, Zhang J. 2021 Proportion of asymptomatic coronavirus disease 2019: a systematic review and meta-analysis. *J. Med. Virol.* **93**, 820–830. (doi:10.1002/jmv.26326)
- Poletti P *et al.* 2021 Association of age with likelihood of developing symptoms and critical disease among close contacts exposed to patients with confirmed SARS-CoV-2 infection in Italy. *JAMA Netw. Open* **4**, e211085–e211085. (doi:10.1001/jamanetworkopen.2021.1085)
- McDonald SA, Miura F, Vos ERA, van Boven M, de Melker HE, van der Klis FRM, van Binnendijk RS, den Hartog G, Wallinga J. 2021 Estimating the asymptomatic proportion of SARS-CoV-2 infection in the general population: analysis of nationwide serosurvey data in the Netherlands. *Eur. J. Epidemiol.* **1**, 3.
- Jones TC *et al.* 2021 Estimating infectiousness throughout SARS-CoV-2 infection course.

- Science* **373**, eabi5273. (doi:10.1126/science.abi5273)
8. Al-Qahtani M, AlAli S, AbdulRahman AK, Salman Alsayyad A, Otoom S, Atkin SL. 2021 The prevalence of asymptomatic and symptomatic COVID-19 in a cohort of quarantined subjects. *Int. J. Infect. Dis.* **102**, 285–288. (doi:10.1016/j.ijid.2020.10.091)
  9. Fisman DN, Tuite AR. 2021 Asymptomatic infection is the pandemic's dark matter. *Proc. Natl Acad. Sci. USA* **118**, e2114054118. (doi:10.1073/pnas.2114054118)
  10. Sah P, Fitzpatrick MC, Zimmer CF, Abdollahi E, Juden-Kelly L, Moghadas SM, Singer BH, Galvani AP. 2021 Asymptomatic SARS-CoV-2 infection: a systematic review and meta-analysis. *Proc. Natl Acad. Sci. USA* **118**, e2109229118. (doi:10.1073/pnas.2109229118)
  11. Zou L *et al.* 2020 SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N. Engl. J. Med.* **382**, 1177–1179. (doi:10.1056/NEJMc2001737)
  12. Lavezzo E *et al.* 2020 Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* **584**, 425–429. (doi:10.1038/s41586-020-2488-1)
  13. Vespignani A *et al.* 2020 Modelling COVID-19. *Nat. Rev. Phys.* **2**, 279–281. (doi:10.1038/s42254-020-0178-4)
  14. Lee S *et al.* 2020 Clinical course and molecular viral shedding among asymptomatic and symptomatic patients with SARS-CoV-2 infection in a community treatment center in the Republic of Korea. *JAMA Intern. Med.* **180**, 1447–1452. (doi:10.1001/jamainternmed.2020.3862)
  15. Long QX *et al.* 2020 Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204. (doi:10.1038/s41591-020-0965-6)
  16. Yongchen Z, Shen H, Wang X, Shi X, Li Y, Yan J, Chen Y, Gu B. 2020 Different longitudinal patterns of nucleic acid and serology testing results based on disease severity of COVID-19 patients. *Emerg. Microbes Infect.* **9**, 833–836. (doi:10.1080/22221751.2020.1756699)
  17. He D, Zhao S, Lin Q, Zhuang Z, Cao P, Wang MH, Yang L. 2020 The relative transmissibility of asymptomatic COVID-19 infections among close contacts. *Int. J. Infect. Dis.* **94**, 145–147. (doi:10.1016/j.ijid.2020.04.034)
  18. Sayampanathan AA, Heng CS, Pin PH, Pang J, Leong TY, Lee VJ. 2021 Infectivity of asymptomatic versus symptomatic COVID-19. *Lancet* **397**, 93–94. (doi:10.1016/S0140-6736(20)32651-9)
  19. Tian H *et al.* 2020 An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642. (doi:10.1126/science.abb6105)
  20. Kraemer MUG *et al.* 2020 The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497. (doi:10.1126/science.abb4218)
  21. Rader B *et al.* 2020 Crowding and the shape of COVID-19 epidemics. *Nat. Med.* **26**, 1829–1834. (doi:10.1038/s41591-020-1104-0)
  22. Subramanian R, He Q, Pascual M. 2021 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**, e2019716118. (doi:10.1073/pnas.2019716118)
  23. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, Salanti G, Low N. 2020 Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. *PLoS Med.* **17**, e1003346. (doi:10.1371/journal.pmed.1003346)
  24. Pollock AM, Lancaster J. 2020 Asymptomatic transmission of COVID-19. *BMJ* **371**, m4851. (doi:10.1136/bmj.m4851)
  25. Meyerowitz EA, Richterman A, Bogoch II, Low N, Cevik M. 2020 Towards an accurate and systematic characterisation of persistently asymptomatic infection with SARS-CoV-2. *Lancet Infect. Dis.* **21**, e163–e169. (doi:10.1016/S1473-3099(20)30837-9)
  26. Kumar N, Shahul Hameed SK, Babu GR, Venkataswamy MM, Dinesh P, Kumar BGP, John DA, Desai A, Ravi V. 2021 Descriptive epidemiology of SARS-CoV-2 infection in Karnataka state, South India: transmission dynamics of symptomatic vs. asymptomatic infections. *EclinicalMedicine* **32**, 100717. (doi:10.1016/j.edinm.2020.100717)
  27. Zhou X, Li Y, Li T, Zhang W. 2020 Follow-up of asymptomatic patients with SARS-CoV-2 infection. *Clin. Microbiol. Infect.* **26**, 957–959. (doi:10.1016/j.cmi.2020.03.024)
  28. Kim SE *et al.* 2020 Viral kinetics of SARS-CoV-2 in asymptomatic carriers and presymptomatic patients. *Int. J. Infect. Dis.* **95**, 441–443. (doi:10.1016/j.ijid.2020.04.083)
  29. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. 2020 Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493. (doi:10.1126/science.abb3221)
  30. Wu SL *et al.* 2020 Substantial underestimation of SARS-CoV-2 infection in the United States. *Nat. Commun.* **11**, 1. (doi:10.1038/s41467-019-13993-7)
  31. Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C. 2020 Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584**, 420–424. (doi:10.1038/s41586-020-2554-8)
  32. Zhang J *et al.* 2020 Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486. (doi:10.1126/science.abb8001)
  33. Lai S *et al.* 2020 Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **585**, 410–413. (doi:10.1038/s41586-020-2293-x)
  34. Tian H, Bjornstad ON. 2020 Population serology for SARS-CoV-2 is essential to regional and global preparedness. *Lancet Microbe* **1**, e94. (doi:10.1016/S2666-5247(20)30055-0)
  35. Liu Y *et al.* 2021 Associations between changes in population mobility in response to the COVID-19 pandemic and socioeconomic factors at the city level in China and country level worldwide: a retrospective, observational study. *Lancet Digit. Health* **3**, e349–e359. (doi:10.1016/S25589-7500(21)00059-5)
  36. Althouse BM, Wenger EA, Miller JC, Scarpino S V, Allard A, Hébert-Dufresne L, Hu H. 2020 Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. (<https://arxiv.org/abs/2005.13689>)
  37. Lee EC, Wada NI, Grabowski MK, Gurley ES, Lessler J. 2020 The engines of SARS-CoV-2 spread. *Science* **370**, 406–407. (doi:10.1126/science.abd8755)
  38. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Eggo RM. 2020 Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211. (doi:10.1038/s41591-020-0962-9)
  39. Zhang J, Litvinova M, Liang Y, Zheng W, Shi H, Vespignani A, Viboud C, Ajelli M, Yu H. 2021 The impact of relaxing interventions on human contact patterns and SARS-CoV-2 transmission in China. *Sci. Adv.* **7**, eabe2584. (doi:10.1126/sciadv.abe2584)
  40. Laine M. 2020 MCMC toolbox for Matlab. Software. See <https://zenodo.org/record/4105825>.
  41. Maltezos HC *et al.* 2021 Transmission dynamics of SARS-CoV-2 within families with children in Greece: a study of 23 clusters. *J. Med. Virol.* **93**, 1414–1420. (doi:10.1002/jmv.26394)
  42. Al-Rifai RH, Acuna J, Al Hossany FI, Aden B, Al Memari SA, Al Mazrouei SK, Ahmed LA. 2021 Epidemiological characterization of symptomatic and asymptomatic COVID-19 cases and positivity in subsequent RT-PCR tests in the United Arab Emirates. *PLoS ONE* **16**, e0246903. (doi:10.1371/journal.pone.0246903)
  43. Liu P, Niu R, Chen J, Tang Y, Tang W, Xu L, Feng J. 2021 Epidemiological and clinical features in patients with coronavirus disease 2019 outside of Wuhan, China: special focus in asymptomatic patients. *PLoS Negl. Trop. Dis.* **15**, e0009248. (doi:10.1371/journal.pntd.0009248)
  44. Li Z *et al.* 2021 Antibody seroprevalence in the epicenter Wuhan, Hubei, and six selected provinces after containment of the first epidemic wave of COVID-19 in China. *Lancet Reg. Health West Pac.* **8**, 100094. (doi:10.1016/j.lanwpc.2021.100094)
  45. Jara A *et al.* 2021 Effectiveness of an inactivated SARS-CoV-2 vaccine in Chile. *N. Engl. J. Med.* **385**, 875–884. (doi:10.1056/NEJMoa2107715)
  46. Hyams C *et al.* 2021 Effectiveness of BNT162b2 and ChAdOx1 nCoV-19 COVID-19 vaccination at preventing hospitalisations in people aged at least 80 years: a test-negative, case-control study. *Lancet Infect. Dis.* **21**, 1539–1548. (doi:10.1016/S1473-3099(21)00330-3)
  47. Tanriover MD *et al.* 2021 Efficacy and safety of an inactivated whole-virion SARS-CoV-2 vaccine (CoronaVac): interim results of a double-blind, randomised, placebo-controlled, phase 3 trial in Turkey. *Lancet* **398**, 213–222. (doi:10.1016/S0140-6736(21)01429-X)
  48. Wang Z *et al.* 2022 Assessing the asymptomatic proportion of SARS-CoV-2 infection with age in China before mass vaccination. Figshare.