ARTICLE

Check for updates

Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China

Shixiong Hu[1,11], Wei Wang[2,11], Yan Wang[2,11], Maria Litvinova [3,4,11], Kaiwei Luo[1], Lingshuang Ren[2], Qianlai Sun[1], Xinghui Chen[2], Ge Zeng[1], Jing Li[2], Lu Liang[5], Zhihong Deng[1], Wen Zheng[2], Mei Li[2], Hao Yang[1], Jinxin Guo[2], Kai Wang[2], Xinhua Chen[2], Ziyan Liu[1], Han Yan[2], Huilin Shi[2], Zhiyuan Chen[2], Yonghong Zhou[2], Kaiyuan Sun[6], Alessandro Vespignani[4,7], Cécile Viboud[6], Lidong Gao[1,12✉], Marco Ajelli [3,7,12✉] & Hongjie Yu [2,8,9,10,12✉]

Several mechanisms driving SARS-CoV-2 transmission remain unclear. Based on individual records of 1178 potential SARS-CoV-2 infectors and their 15,648 contacts in Hunan, China, we estimated key transmission parameters. The mean generation time was estimated to be 5.7 (median: 5.5, IQR: 4.5, 6.8) days, with infectiousness peaking 1.8 days before symptom onset, with 95% of transmission events occurring between 8.8 days before and 9.5 days after symptom onset. Most transmission events occurred during the pre-symptomatic phase (59.2%). SARS-CoV-2 susceptibility to infection increases with age, while transmissibility is not significantly different between age groups and between symptomatic and asymptomatic individuals. Contacts in households and exposure to first-generation cases are associated with higher odds of transmission. Our findings support the hypothesis that children can effectively transmit SARS-CoV-2 and highlight how pre-symptomatic and asymptomatic transmission can hinder control efforts.

[1] Hunan Workstation for Emerging Infectious Disease Control and Prevention, Chinese Academy of Medical Sciences, Hunan Provincial Center for Disease Control and Prevention, Changsha, China. [2] School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China. [3] Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA. [4] ISI Foundation, Turin, Italy. [5] West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China. [6] Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA. [7] Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA, USA. [8] Department of Infectious Diseases, Huashan Hospital, Fudan University, Shanghai, China. [9] Shanghai Key Laboratory of Infectious Diseases and Biosafety Emergency Response, Shanghai, China. [10] Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, China. [11] These authors contributed equally: Shixiong Hu, Wei Wang, Yan Wang, Maria Litvinova. [12] These authors jointly supervised this work: Lidong Gao, Marco Ajelli, Hongjie Yu. ✉email: 810173358@qq.com; marco.ajelli@gmail.com; yhj@fudan.edu.cn

The outbreak of coronavirus disease 2019 (COVID-19) was first detected in December 2019 in Wuhan, China[1]. The outbreak, caused by the SARS-CoV-2 virus, quickly spread globally, leading WHO to declare a pandemic on March 11, 2020[2]. Despite more than 94.1 million SARS-CoV-2 infected individuals confirmed worldwide as of January 19, 2021[3], there are still many unknowns in the epidemiology and natural history of COVID-19.

A key question under debate is whether the infectivity of individuals with, and susceptibility to, SARS-CoV-2 infection differs by age. In particular, the role of children in SARS-CoV-2 transmission has yet to be fully understood. Schools were closed in the early months of the pandemic in most countries[4,5], so that the low proportion of cases notified in young individuals[6] could be attributed to a low probability of developing symptoms[7,8], a low susceptibility to infection[9–11], and/or few contact opportunities relative to other age groups. The importance of each of these factors has been difficult to disentangle thus far. A related question is the probability of asymptomatic transmission. In fact, it is often argued that the COVID-19 pandemic has been difficult to tackle because of the importance of pre-symptomatic and asymptomatic transmission. Evidence from confined settings such as households, homeless shelters, and nursing facilities, supports the role of pre-symptomatic and asymptomatic transmission[1,10,12–15]. Yet, a quantification of the contribution of asymptomatic and pre-symptomatic transmission in large populations is still lacking.

A full understanding of SARS-CoV-2 transmission patterns and risk factors is crucial to plan targeted COVID-19 responses, at least until a considerable fraction of the population is immune to the infection either through natural exposure or vaccination. So far, case-based interventions (e.g., case isolation, quarantine of contacts, contact tracing) have been the backbone of response strategies in most countries, at times in combinations with partial or full lockdowns. To define the temporal characteristics of the response strategies (e.g., duration of the quarantine and isolation period, definition of contacts to be traced) it is crucial to understand the age profile of infectiousness and to have robust estimates of key time-to-event distributions such as the generation time. These distributions were estimated in the early days of the pandemic based on the very first few clusters of cases and are thus subject to high uncertainty and variability between different studies[1,16]. It is important to update these estimates using large-scale and harmonized epidemiological datasets.

In this study, we analyze 1178 SARS-CoV-2 infected individuals and their 15,648 contacts identified by contact tracing operations carried out in the Hunan Province of China over the period from January 13 to April 02, 2020. This comprehensive and detailed dataset compiled by the Hunan Provincial CDC sheds light on SARS-CoV-2 transmission risk factors, and the distribution of key time-to-event parameters.

Results

Sample description. Between January 23, 2020 and April 02, 2020, 1019 symptomatic cases (i.e., PCR positive subjects who showed symptoms—see the "Methods" section for the detailed definition) and 159 asymptomatic subjects (i.e., PCR positive subjects who did not show symptoms—see "Methods" section for the detailed definition) were reported and screened for inclusion (Supplementary Fig. 1 and Table 1). Through active contacts tracing, a total of 15,648 close contacts were identified, of whom 471 contacts were positive for SARS-CoV-2 infection. Among 1178 SARS-CoV-2 infections, we identified an infector for a total of 432 transmission events, 831 epidemiologically linked cases (including index cases) in 210 clusters. Of these clusters, 499 SARS-CoV-2 infections in 123 clusters had a clear epidemiological

link to an individual previously infected with SARS-CoV-2. From 15,648 close contacts, 6412 were identified by forward contact tracing and resulted in the identification of 285 symptomatic cases and 63 asymptomatic SARS-CoV-2 positive subjects. The remaining 9236 close contacts were identified through backward contact tracing. The distribution of the cases and close contacts in time and space is presented in Fig. 1 and Supplementary Figs. 2 and 3. Overall, the median age of symptomatic cases and asymptomatic subjects, and their close contacts were 45 (IQR: 34–55), 36 (IQR: 19–52), and 40 (IQR: 27–52) years, respectively (Table 1). Cases aged 0–14 years presented milder or no clinical symptoms, while patients aged 15 years and older had more severe illness ($P < 0.001$) (Supplementary Fig. 4D).

Time-to-key-event distributions. To estimate the duration of the incubation period, we analyzed 268 confirmed cases with locally acquired infections belonging to 114 clusters, with information on both the date(s) of exposure and generation in the transmission chain of the cluster. We found that the best fitting distribution of the incubation period was a Weibull distribution with mean 6.4 days (median: 5.7, IQR: 3.2–8.8) (Supplementary Table 3). We performed two sensitivity analyses, one excluding cases with only an exposure end date (17 individuals) and another one where we inferred the earliest exposure date for 7 of those 17 individuals. The results of both sensitivity analyses are consistent with the main analysis (Supplementary Table 3).

Symptom onset dates were available for 245 transmission pairs; the resulting serial interval had an estimated mean of 5.5 days (median: 4.8, IQR: 0.8–9.4), based on fitting a gamma distribution. By considering only pairs with a single identified infector, we find that 14.0% (31/221) of the empirical serial intervals were negative, which means that the symptom onset date of the infectee precedes the symptom onset date of their potential infector. To assess whether the serial interval changed over the course of the epidemic, we divided the outbreak in Hunan into two periods: the first one running from the detection of the first case up to January 23 and the second one from January 24 (date of the activation of the Level 1 Emergency Response) to April 2 (date of the detection of the last confirmed case). The mean serial interval decreased from 7.0 (median: 6.6, IQR: 3.5–10.1) days in the first period, to 4.1 (median: 3.2, IQR: −1.1, 8.4) days in the second period.

The mean generation time was estimated to be 5.7 days (median: 5.5, IQR: 4.5–6.8). The difference between the estimated mean incubation period and mean generation time is less than one day (Fig. 2B).

The mean time interval from symptom onset to the date of collection of the sample for PCR testing was estimated to be 4.7 days (median: 4.2, IQR: 1.4–7.4) using the best fitting gamma distribution, based on 531 confirmed cases. The mean time interval from symptom onset to laboratory confirmation was estimated to be 6.4 (median: 6.0, IQR: 3.3–9.1) days, based on 952 confirmed cases.

Pre-symptomatic transmission. Infectiousness was estimated to peak 1.8 days before symptom onset (Fig. 2a). It is important to stress that our estimate provides a measure of the probability of infecting contacts at any time after the time of exposure to the infector. As such, this is an empirical measure of the transmissibility over time of infectors, which accounts for human behavior (contacts) and performed interventions (e.g., case isolation, precautionary behaviors). We estimated the proportion of pre-symptomatic transmission (area under the curve, Fig. 2a) at 59.2%, with 95% of transmission events occurring between −8.8 and 9.5 days since the date of symptom onset. The proportion of

Table 1 Characteristics of symptomatic cases, asymptomatic subjects, and their close contacts in Hunan Province, China.

Characteristics	Symptomatic cases[a] (n = 1019)	Asymptomatic subjects[a] (n = 159)	Close contacts of cases with SARS-CoV-2 infections (n = 15,648)[b,c]
Age, years			
Median (interquartile range, IQR)	45 (34–55)	36 (19–52)	40 (27–52)
0–14	33 (3.2)	28 (17.6)	1706 (10.9)
15–64	849 (83.3)	119 (74.8)	11,662 (74.5)
≥65	137 (13.4)	12 (7.5)	1516 (9.7)
Missing	0 (0)	0 (0)	764 (4.9)
Sex			
Male	526 (51.6)	75 (47.2)	7984 (51)
Female	493 (48.4)	84 (52.8)	7397 (47.3)
Missing	0 (0)	0 (0)	267 (1.7)
Exposure history[c]			
Residence in or travel history from Hubei Province	439 (43.1)	31 (19.5)	–
Contact with other confirmed cases or person with acute respiratory infections	366 (35.9)	90 (56.6)	–
Household contacts	–	–	2771 (17.7)
Relative contacts	–	–	7284 (46.5)
Social contacts	–	–	4550 (29.1)
Other close contacts[d]	–	–	5709 (36.5)
Contact with person traveled to Hubei Province	616 (60.5)	71 (44.7)	–
Exposure not determined	296 (29.0)	48 (30.2)	–
Clinical severity			
Asymptomatic subjects	–	159 (100)	104 (0.7)
Symptomatic subjects	1019 (100)	–	367 (2.4)
Mild patients	299 (29.3)	–	153 (1.0)
Moderate patients	570 (55.9)	–	174 (1.1)
Severe patients	119 (11.7)	–	31 (0.2)
Critical patients	31 (3.0)	–	9 (0.1)

Note: Data are presented as no. (%) of cases/contacts unless otherwise indicated.
[a]All symptomatic cases and asymptomatic subjects were PCR confirmed (see "Methods" section for the detailed definitions).
[b]A total of 471 cases of SARS-CoV-2 infections were identified among 15,648 close contacts in Hunan province, which were also included in 1019 symptomatic COVID-19 cases and 159 subjects with asymptomatic SARS-CoV-2 infections.
[c]Percentages may not total 100 as one individual may be associated with multiple observed exposures and contacts.
[d]Other close contacts refer to caregivers and patients in the same ward, persons in the same transportation vehicle, and those providing service for the case in public places.

pre-symptomatic transmission (area under the curve) increased from 50.8% for the period from January 5 to January 23, to 76.7% for the period from January 24 to April 2, when intensive contact tracing and isolation strategy were undertaken by the Hunan Province authorities. From the analysis of the transmission chains reconstructed through field investigations, 43 pre-symptomatic transmission events were recorded in 23 clusters. A subset of those clusters is shown in Fig. 3.

Asymptomatic transmission. From the analysis of contact tracing records, we identified 8 clusters (25 local transmission events) with evidence of asymptomatic transmission. There were 11 asymptomatic infectors (5 primary and 6 secondary infectors) associated with 15 transmission events (10 secondary and 5 tertiary infections, Supplementary Fig. 5). No asymptomatic individual was a cluster index case (i.e., did not trigger a contact tracing investigation), although 5 of them were primary infectors.

SARS-CoV-2 risk factors. We first explored differences in the age of SARS-CoV-2 infectors and infectees through the construction of age-specific transmission matrices (Supplementary Fig. 6). The results suggest that people aged 15–64 years generated a larger mean number of cases than younger (0–14 years old) and older (65+ years old) individuals. Moreover, individuals aged 65 years and older were infected more often. Note that these three age groups were chosen to represent three key segments of the

population, namely (i) younger than working-age (students and preschoolers), (ii) working-age population, and (iii) retiring-age individuals. We have also examined the proportion of transmission events associated with asymptomatic infectors. In our sample, we had 432 transmission events with an identified infector, 15 of which (3.5%) were associated with asymptomatic infectors. However, the share increases to 8.5% (10/118) if we focus on transmission events occurring after February 7, 2020 and to 60% (15/25) if we consider only clusters with at least one asymptomatic transmission event.

It is important to stress that these estimates do not account for several confounding factors (e.g., all index cases are symptomatic, children are more likely to be in later generations of transmission, see Supplementary Tables 6 and 7). To account for the possible effect of multiple confounding factors, we thus performed a multivariate regression analysis (GLMM). We found that the age of the contact, the contact setting, and the generation of the infector in a cluster were important risk factors for transmission (Table 2). Infectiousness was not significantly different between working-age adults (15–64 years old) and other age groups (0–14 years old: p-value = 0.210; 65 years and over: p-value = 0.306); in contrast, susceptibility to SARS-CoV-2 infection increased with age (p-value = 0.028, Model 2 in Table 2). We found no statistically significant difference in transmissibility between symptomatic and asymptomatic individuals. Furthermore, household contacts were associated with a significantly larger risk of SARS-CoV-2 infection than other types of contact. The GLMM
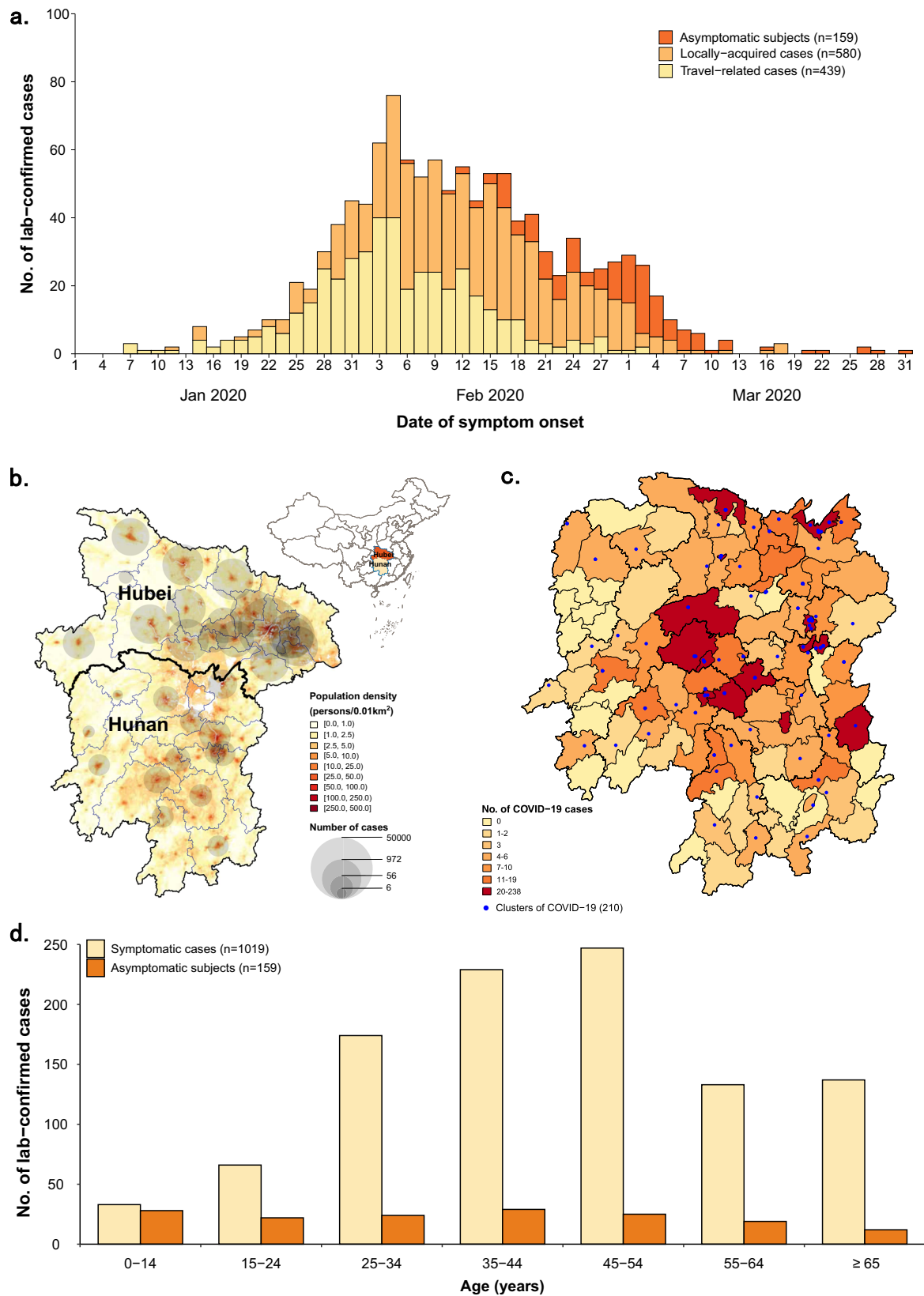
Fig. 1 Temporal, geographical, and age distribution of SARS-CoV-2 infected individuals stratified by the presence of symptoms and source of infection in Hunan Province, China. a Daily number of new SARS-CoV-2 infected individuals by date of symptom onset by source of infection. For asymptomatic individuals we used date of the first RT-PCR positive result. **b** Geographical distribution of SARS-CoV-2 infected individuals in Hunan Province, and geo-locations of Hubei and Hunan provinces. **c** Geographical distribution of SARS-CoV-2 clusters. **d** Age distribution of SARS-CoV-2 symptomatic and asymptomatic infected individuals.
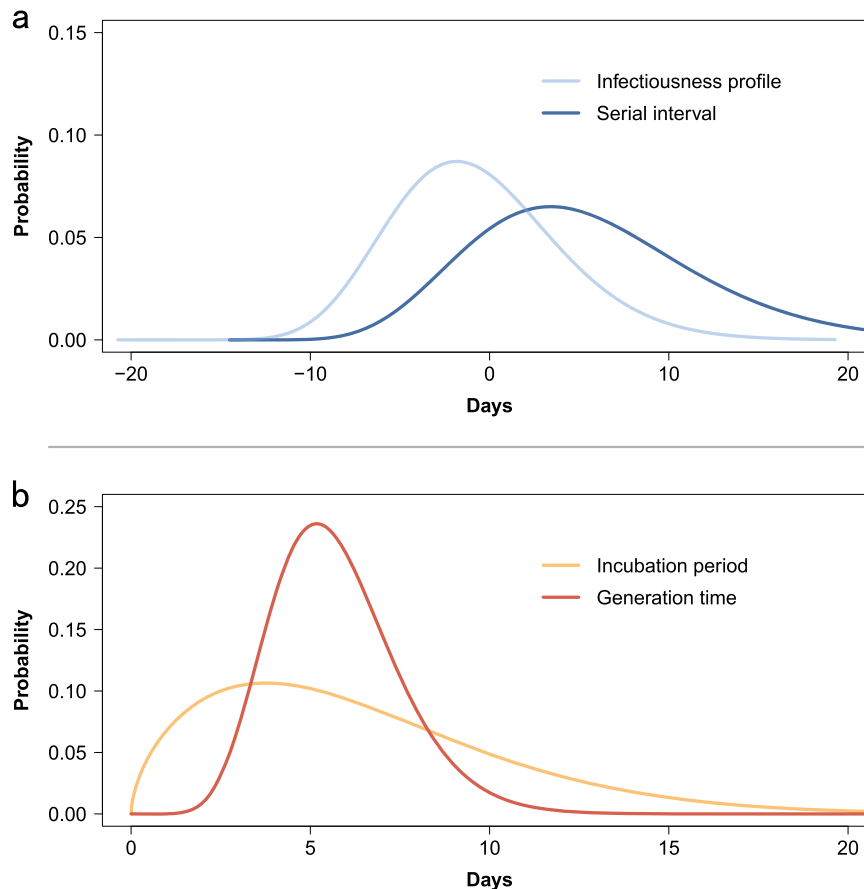
Fig. 2 Quantifying the serial interval, infectiousness profile, incubation period, and generation time fitted by gamma or Weibull distributions.
a Estimated distribution of the serial interval and of the infectiousness profile by gamma distributions based on 245 transmission pairs. **b** Estimated distribution of the incubation period by Weibull distributions and of the generation time by gamma distributions based on 268 confirmed cases with locally acquired infection belonging to 114 clusters.

model suggests two other statistically significant risk factors: the generation in the transmission chain and the number of contacts identified for an infector (Table 2). In particular, the transmission risk in the first generation was significantly higher than later generations, possibly due to improved case isolation and contacts quarantine that deplete the number of susceptible individuals in the cluster. The same results were observed when accounted for the time period of the epidemic (Supplementary Tables 10 and 11). We also found a slight but significant decrease in transmission risk from cases who reported more contacts. The inclusion of other potential risk factors, such as the gender of infectors and the gender of the contacts were not statistically significant, did not modify the estimated odds ratios for the other variables, and did not improve the fit of the model (Supplementary Tables 9 and 10, and Supplementary Fig. 7).

Discussion
This analysis of SARS-CoV-2 transmission patterns and risk factors in Hunan, China, is based on the largest contact tracing dataset considered thus far. We found no difference in transmissibility between symptomatic and asymptomatic individuals and between age groups, while susceptibility to SARS-CoV-2 infection increased with age. We provide evidence of both presymptomatic and asymptomatic SARS-CoV-2 transmission, with the former potentially accounting for up to 59.2% of all transmission events in this dataset. In addition, we estimate that SARS-CoV-2 transmission in households is responsible for most

secondary and tertiary infections. Further, within a cluster, individuals who were exposed to primary cases experienced a significantly higher risk of SARS-CoV-2 infection than those exposed to later cases.

The exposure history data used in this study were collected from in-depth epidemiological investigations, allowing us to provide robust estimation of several key time-to-event distributions. Previous estimates suffered of large uncertainty, ranging from 3.0 to 7.8 days for the serial interval[1,17–22] and from 4.8 to 8.0 days for the incubation period[1,23–28]. We note that short estimates of the serial interval such as the one obtained for Brazil[21] tend to be skewed as secondary cases tend to recall more recent contacts, which is especially true when a major epidemic is unfolding[29]. This appears not to be the case in Hunan where the exponential growth phase of the outbreak lasted only about two weeks[23] and the effort heavily relied on forward contact tracing. Still, our estimates fall within these intervals. Moreover, in agreement with Ali et al.[30], we found that the mean serial interval shortened over time, reflecting increased timeliness of case isolation that truncates successful onward transmission. Unlike the serial interval and the incubation period, only a few studies[31,32] provide estimates of the generation time, as it is hard to directly infer it from field investigations given that it requires information on the infection dates of both the infector and their infectees. In this work, we estimate the mean generation time at 5.7 days (median: 5.5, SD: 1.8), in general agreement with Ferretti et al.[32] (median: 5.0 days; SD: 1.9 days). Solid estimates of the generation time

Fig. 3 Timing of transmission events and SARS-CoV-2 infected subjects in randomly selected clusters showing evidence of pre-symptomatic transmission. Square symbols indicate symptomatic cases and circular symbols indicate asymptomatic subjects. Age, sex, and generation in a cluster are shown for each SARS-CoV-2 infected individual (left panels), with information on date of illness onset for symptomatic cases and date of the first RT-PCR positive result for asymptomatic subjects and for symptomatic cases without date of illness onset (symbol "*"). Timeline of events (right panels).

are key as, in conjunction with epidemic growth rate, they can be used to estimate the reproduction number of an epidemic[33,34]. In the absence of such data, many studies so far have relied on the distribution of the serial interval as an approximation of the generation time[1,35]. However, individual variability in the duration of the incubation period is expected to widen the distribution of the serial interval with respect to that of the generation time. This is highlighted by the IQR of the two distributions estimated here, namely the mean of the serial interval was estimated at 5.5 days (IQR: 0.9–9.4) and that of the generation time at 5.7 days (IQR: 4.5–6.8).

Previous studies show a relatively high proportion of pre-symptomatic transmission, but estimates vary significantly, ranging between 13 and 62%[1,32,36]. Our estimate (59.2%) nears the high end of the range found in the literature. This may be due to two main factors. First, the fraction of pre-symptomatic transmission heavily depends on the intensity of contact tracing and isolation strategy (e.g., whether cases are promptly isolated in dedicated facilities at the time of symptom onset or are isolated at home). Second, the depth of the contact tracing investigation may determine the rate of ascertainment of index cases. Our study highlighted that the contact tracing system in Hunan has

Table 2 Estimating the association of demographic and behavioral factors with the risk of acquiring and transmitting SARS-CoV-2.

Characteristics	No. of individuals	No. of contacts	No. of contacts per person	No. of secondary infections	Secondary infection attack rate (%, 95%CI)	Data-based OR (95%CI)	Model 1[a]		Model 2[a]	
							OR (95%CI)	P-value	OR (95%CI)	P-value
Age of infectors										
0–14 y	25	193	7.7	2	1.0 (0.1, 3.7)	0.36 (0.04, 1.16)	0.28 (0.04, 2.04)	0.210	–	–
15–64 y	355	6833	19.2	188	2.8 (2.4, 3.2)	Reference	Reference	–	–	–
65+ y	81	1133	14.0	19	1.7 (1.0, 2.6)	0.61 (0.42, 0.81)	0.62 (0.25, 1.55)	0.306	–	–
Log-transformed age	461	8159	17.7	–	–	–	–	–	1.57 (0.87, 2.81)	0.134
Age of contacts										
0–14 y	22	936	42.5	–	–	–	0.58 (0.34, 0.98)	0.041	–	–
15–64 y	154	6441	41.6	–	–	–	Reference	–	–	–
65+ y	33	812	24.6	–	–	–	1.65 (1.03, 2.64)	0.038	–	–
Log-transformed age	209	8159	39.0	–	–	–	–	–	1.26 (1.02, 1.55)	0.028
Type of contact										
Household contacts	–	1021	–	99	9.7 (8.0, 11.7)	Reference	Reference	–	Reference	–
Relative contacts	–	3084	–	56	1.8 (1.4, 2.4)	0.19 (0.17, 0.21)	0.11 (0.07, 0.17)	2.30e−21	0.11 (0.07, 0.18)	5.37e−21
Social contacts	–	2227	–	27	1.2 (0.8, 1.8)	0.12 (0.10, 0.15)	0.06 (0.03, 0.11)	3.95e−20	0.06 (0.03, 0.11)	1.70e−20
Other contacts	–	1827	–	27	1.5 (1.0, 2.1)	0.15 (0.12, 0.18)	0.07 (0.04, 0.13)	1.73e−18	0.07(0.04, 0.13)	9.75e−19
Generation of SARS-CoV-2 transmission										
G1	–	2121	–	91	4.3 (3.5, 5.2)	Reference	Reference	–	Reference	–
G2	–	2987	–	81	2.7 (2.2, 3.4)	0.63 (0.63, 0.65)	0.14 (0.06, 0.32)	3.81e−6	0.13 (0.05, 0.29)	1.25e−6
G3-4	–	965	–	19	2.0 (1.2, 3.1)	0.47 (0.34, 0.60)	0.05 (0.02, 0.19)	5.04e−6	0.05 (0.02, 0.18)	3.22e−6
Multiple exposure[b]	–	598	–	9	1.5 (0.7, 2.8)	0.35 (0.20, 0.54)	0.11 (0.03, 0.43)	0.002	0.11 (0.03, 0.42)	0.001
Unknown	–	1488	–	9	0.6 (0.3, 1.1)	0.14 (0.09, 0.21)	0.03 (0.01, 0.11)	8.94e−8	0.03 (0.01, 0.11)	7.45e−8
Levels of exposure to an infector										
Total number of contacts	–	8159	–	–	–	–	0.99 (0.97, 1.00)	0.022	0.99 (0.97, 1.00)	0.030
Gender of infectors										
Female	239	4067	17.0	87	2.1 (1.7, 2.6)	Reference	Reference	–	Reference	–
Male	222	4092	18.4	122	3.0 (2.5, 3.5)	1.43 (1.35, 1.47)	1.75 (0.96, 3.17)	0.067	1.75 (0.97, 3.18)	0.064
Gender of contacts										
Female	424	4017	9.5	–	–	–	Reference	–	Reference	–
Male	427	4142	9.7	–	–	–	1.02 (0.74, 1.41)	0.899	1.01 (0.73, 1.39)	0.964
Clinical severity of infectors										
Asymptomatic	80	898	11.2	15	1.7 (0.9, 2.7)	0.63(0.39, 0.87)	0.68 (0.25, 1.88)	0.460	0.72 (0.26, 1.98)	0.523
Symptomatic	381	7261	19.1	194	2.7 (2.3, 3.1)	Reference	Reference	–	Reference	–

Note that significance was tested using two-sided Wald test with $\alpha = 0.05$.
[a]Age of infectors and contacts were considered either as categorical (model 1) or continuous log-transformed (model 2) variables.
[b]Contacts who were exposed to multiple cases of different generations of SARS-CoV-2 transmission.

effectively identified a large number of pre-symptomatic transmission events and, at the same time, produced high-quality data that allowed us to advance the understanding of where and among whom SARS-CoV-2 transmission occurs. One of the main lessons learned by the successful history of epidemic containment in Hunan is the key role of rapid index case isolation, extensive identification of potentially exposed individuals and extensive search for the originating exposure events through contact tracing, which resulted in the decrease of the risk of infection; in fact, we found that the risk of infection significantly decreased with the number of the generations in the reconstructed transmission chains. As discussed in previous studies[15,30,31], the effectiveness of tracing and isolation/quarantine heavily depends on quick identification of cases. Here we estimated the mean time interval from symptom onset to PCR sample collection and to laboratory confirmation to be 4.7 and 6.4 days, respectively. However, contacts were quarantined preventively before the diagnosis was confirmed. Finally, it is important to note that, by definition, our analysis includes only contacts occurring up to 2 days prior to the symptom onset of the presumed infector (as per Chinese authorities' policy[37]). This may potentially lead to underestimation of pre-symptomatic transmission outside the household and skew the distribution of infectiousness. Future analyses of viral load data may provide further support to our estimates of the infectiousness profile over time.

We found evidence of asymptomatic transmission in several clusters, with 15 secondary cases (out of 432 transmission events) linked to asymptomatic infectors, similar to Chen et al. (6/132 events)[38] and Liu et al. (24/914 events)[39]. Other studies provide evidence of asymptomatic infection[12,36,40], but do not attempt to quantify its contribution to transmission. Our multivariate analysis shows no statistically significant difference in the transmissibility between symptomatic and asymptomatic individuals. This highlights that the low proportion of cases generated by asymptomatic individuals in this study (3.5%) can be partially explained by the lower probability of identification of asymptomatic index cases. In fact, by considering only clusters with at least one asymptomatic transmission event, the proportion of asymptomatic transmission increases to 60% (15/25). However, it is important to note our data cannot be used to estimate the probability of developing symptoms as a fraction of asymptomatic infections (e.g., entire clusters that consist only of asymptomatic subjects) may have been missed despite extensive PCR testing performed by the Hunan CDC. In fact, testing focused on symptomatic contacts before February 7, 2020, and was expanded to all contacts afterward. Therefore, our findings cannot be used to quantitatively estimate the percentage of infected individuals who develop symptoms.

In agreement with previous studies, we found that the risk of infection from a household member is larger than that resulting from other contacts[10,41]. This may be explained by the duration, type, and frequency of contacts between household members as well as the impact of interventions (such as household quarantine) on household contacts. Consistent with the transmissibility of H1N1pdm influenza during the 2009 pandemic in the US[42], we found that SARS-CoV-2 transmissibility decreased with the number of contacts, although the effect is small. Further cohort studies are needed to explain this connection, possibly recording number, type, and duration of contacts. It is important to stress that the observed significantly higher risk of infection in households calls for measures targeted at households, such as providing isolation shelters for mild cases that can remove SARS-CoV-2 infectors from households and thereby interrupt chains of within-household transmission[43,44]. Although we estimated a higher susceptibility to SARS-CoV-2 infection among the elderly, this finding has to be cautiously interpreted. This finding may

stem from a higher probability of infection detection among the elderly due to higher probability of developing symptoms and present severe illness[7].

Despite the challenges of reporting a low number of infections among children and the complexity of establishing epidemiologic links between children and adults within households[19], we assessed the effects of infector and infectee demographics and other characteristics on SARS-CoV-2 susceptibility and infectivity. We found that the odds of infection was significantly higher for first-generation infectors than for later generation ones. Together with a small number of infectious children in the first generation, this contributed to a lower total number of infections generated by children (see Supplementary Table 11). However, when accounting for all confounding factors, including generation number, we found no statistical evidence of differential transmissibility by age group (Table 2). Interestingly, while younger individuals typically have more contacts than other age groups both in China[9,45,46] (range: 18.2–22.3 contacts per day) and elsewhere[9,47–51], the number of individual contacts reported by each infectious child in contact tracing data was considerably lower (mean: 7.7) during the outbreak in Hunan. Such a marked reduction in contacts was likely connected with the interventions in place (lockdown policy) and school closures (either for the New Year vacation and later as part of interventions). Therefore, caution should be applied when evaluating policies that increase the number of contacts among children, such as reopening schools or summer camps. In addition, our findings suggest that the risk of acquiring SARS-CoV-2 infection steadily increases with age (in agreement with Zhang et al.[9,11]). Nonetheless, it is important to remark that our estimates of the infectiousness by age groups are based on a small sample size of younger individuals. Further studies are needed to confirm our finding.

Our study is not without limitations. First, it suffers from the classic limitations of any epidemiological field investigation. Despite the longitudinal and in-depth investigation of each case and their contacts, we could not always accurately reconstruct the entire transmission chain and fully avoid recall bias in individual records. Also, the imperfect sensitivity of PCR testing should be taken into consideration, especially as it highly depends on the delay between the time of infection and specimen collection[52]. Unfortunately, we do not have a representative sample for the date of sample collection, thus we cannot correct for this factor, possibly leading to an underestimation of the number of SARS-CoV-2 infected asymptomatic individuals. Moreover, we cannot rule out the possibility of indirect exposures (e.g., contaminated surfaces), which may affect the identification of epidemiological links. High-resolution genomic and virologic surveillance data would be needed to decrease the uncertainty on the links identified by the epidemiological investigation and to better distinguish direct vs. community transmission[53,54]. Second, the duration of per-contact exposure was not reported in the dataset and we were thus unable to correct for this factor. This may contribute to explain the importance attributed to household contacts in our regression analysis and why individuals with more contacts have lower transmission risk per contact. Third, despite controlling known factors associated with transmissibility, we cannot exclude the possibility that there are other potential factors that may confound the estimated effect of current covariates.

In conclusion, the evidence of pre-symptomatic and asymptomatic SARS-CoV-2 transmission shown in this study underlines the key role of undetectable SARS-CoV-2 transmission that can hinder control efforts. Control measures should thus be tailored accordingly, especially contact tracing, testing, and isolation. Our findings show a high risk of acquiring SARS-CoV-2 from infected individuals not showing symptoms (either while pre-symptomatic or asymptomatic), thus supporting the

enhancement of personal precautions such as wearing a mask and improved hygiene practice. In addition, school reopening, and the consequent increase in the number of daily contacts among children and teenagers, is expected to increase the contribution of children to SARS-CoV-2 transmission. School outbreaks have already been reported on several occasions[5,55–57]; time will tell whether schools become a major source of transmission.

Methods

COVID-19 surveillance system, field epidemiological investigations, and contact tracing. In response to the COVID-19 outbreak, in late December 2019, the Chinese Center for Disease Control and Prevention (China CDC) launched a new surveillance system for COVID-19 cases. A description of the surveillance system is reported elsewhere[1]. On January 21, 2020, the first COVID-19 case was confirmed in Hunan Province. Since then, active field epidemiological investigations of suspected or confirmed SARS-CoV-2 infections as well as their contacts have been initiated.

The definition of suspected and confirmed COVID-19 cases (i.e., symptomatic individuals), as well as subjects with asymptomatic SARS-CoV-2 infections (i.e., asymptomatic subjects) was based on the New Coronavirus Pneumonia Prevention and Control Program published by the National Health Commission (NHC) of China and the World Health Organization (WHO)[58]. A suspected COVID-19 case was defined as a person who met one or more clinical criteria and had an epidemiological link to SARS-CoV-2 positive individuals or history of travel to/ from regions reporting widespread SARS-CoV-2 transmission (Supplementary Information, p2). A confirmed COVID-19 case was defined as a suspected case with positive real-time RT-PCR results, while an asymptomatic subject was defined as an individual with laboratory confirmation of SARS-CoV-2 infection, but without any clinical symptom (e.g., no fever or cough) within the quarantine/ observation period (i.e., 14 days). Confirmed COVID-19 cases were categorized by clinical severity, including mild, moderate, severe, and critical illnesses (as defined in Supplementary Table 1).

SARS-CoV-2 infected individuals were identified using a variety of measures. In particular, (i) to identify travel-associated cases, traffic entrance and community screening were performed in high-risk populations who had a history of traveling from/to Wuhan City/Hubei Province; (ii) to identify symptomatic cases, passive surveillance in hospitals and outpatient practices were monitored; (iii) to capture potential symptomatic cases and asymptomatic subjects, systematic tracing and monitoring of contacts of confirmed cases were performed. In particular, screening measures (i) and (iii) were used to identify SARS-CoV-2 infected asymptomatic individuals. Once a suspected or confirmed COVID-19 case was identified, a field epidemiology investigation was undertaken by the local CDC. Data were collected on demographic characteristics, clinical symptoms, and activity patterns starting 14 days before symptom onset and until confirmation or isolation in the hospital. All cases detected between January 16 and April 02, 2020 were interviewed using a standardized questionnaire. In addition, each individual with suspected or confirmed SARS-CoV-2 infection was asked to provide a list of locations she/he visited (e.g., workplace, health-care facilities) and their contacts. On the basis of this list, active contact tracing was then initiated by the investigation team. Screening interviews, checking of travel records based on public security cameras and traffic system, and digital health records were also collected to assess whether an individual met the definition of close contact. Once a close contact was identified and traced, she/he was quarantined at a designated place (e.g., hotel room) or at home and followed up for 14 days[58]. Close contacts were interviewed using a standardized form before they were quarantined. The form comprised basic demographic information (e.g., age and sex), and detailed a record of the timing, frequency, and type of exposures to the case(s) who triggered the investigation. An earlier version of the data from contact tracing operations containing only reduced descriptive information on contacts was used for the estimation of age-specific susceptibility in Zhang et al.[9].

Specimen collection and laboratory testing. Upper respiratory specimens (nasopharyngeal and oropharyngeal swabs) were collected from all suspected cases as well as their close contacts. Before February 7, 2020 specimens were collected for testing from each close contact if she/he developed symptoms during quarantine period. After February 7, 2020, specimens were collected at least once during quarantine, regardless of symptoms. After January 27, the designated hospitals and local CDCs were approved to conduct real-time RT-PCR assay for diagnosis of COVID-19 using a standardized laboratory testing procedure according to the "Novel coronavirus pneumonia Diagnosis and Treatment Program" released by NHC of China. The assays were performed in laboratory equipped with BSL-2 facilities (Supplementary Information, p5).

Close contacts, sporadic cases, and clusters. Close contacts were defined as individuals who had close-proximity interactions (within 1 m) with clinically suspected and laboratory-confirmed SARS-CoV-2 cases, for the period from 2 days before, to 14 days after, the potential infector's symptom onset. For those exposed to asymptomatic subjects, the contact period was from 2 days before, to 14 days

after, a respiratory sample was taken for real-time RT-PCR testing. Close contacts included, but were not limited to, household contacts (i.e., household members regularly living with the case), relatives (i.e., family members who had close contacts with the case but did not live with the case), social contacts (i.e., work colleagues or classmates), and other close contacts (i.e., caregivers and patients in the same ward, persons sharing a vehicle, and those providing a service in public places, such as restaurants or movie theatres)[37].

A cluster of SARS-CoV-2 infections was defined as a group of two or more confirmed cases or asymptomatic subjects with an epidemiologic link (Supplementary Information, p3). Epidemiologically linked cases were classified according to the generation of SARS-CoV-2 transmission and the setting where exposure took place, with primary cases considered as first generation. A sporadic case was defined as a confirmed case of SARS-CoV-2 infection (either symptomatic or asymptomatic) that did not belong to any of the reported clusters.

We define pre-symptomatic transmission as a direct transmission event that took place before the date of symptom onset of the infector, while asymptomatic transmission is a transmission event from a person that did not develop symptoms within the quarantine/observation period.

Statistical analysis. We provide descriptive statistics of the characteristics of cases and their close contacts, including demographic factors and exposures (Supplementary Information, p5–7). We estimated the incubation period (i.e., the time delay from infection to illness onset), the serial interval (i.e., the time interval between the onset of symptoms in a primary case and in their secondary cases), the generation time (i.e., the time interval between infection of the primary case and of their secondary cases), and the infectiousness profile (i.e., the daily distribution of the probability of transmission since the date of symptom onset; see refs. [20,59] and Supplementary Information, p8–12 for methods). We also estimated the interval from symptom onset to the sampling date of first PCR and to laboratory confirmation by using a maximum likelihood estimator and fitting three distributions (Weibull, gamma, and lognormal) (Supplementary Information, p12–13). The goodness of fit was assessed using Akaike information criterion (AIC). We restrict the estimation of incubation period to 268 locally acquired infections with information on both the date(s) of exposure and generation of SARS-CoV-2 transmission in the cluster.

We rely on the contact tracing data to describe the age-specific contact matrices for SARS-CoV-2 infectors and their contacts (Supplementary Information, p17), and we present the number of contacts per person by demographical characteristics of SARS-CoV-2 infectors and their contacts. To focus on local transmission pairs with clearly epidemiological links, we excluded those travel-related cases and their successive cases (as people in the same cluster often share the same travel history, making hard to disentangle the transmission chain). The transmissibility of SARS-CoV-2 was measured by the secondary infection attack rate. In addition, generalized linear mixed-effects model, GLMM, for binary data with logit link was built to quantify the effects of potential drivers of susceptibility and infectivity of the SARS-CoV-2 virus (i.e., odds ratio and marginal effect), based on 8159 individual records of contacts who were exposed to locally generated cases (see Supplementary Information, p20–22). These risk factors include age and gender of infectors/contacts, type of contact, generation of SARS-CoV-2 transmission in a cluster, as well as the number of contacts of an infector. Statistical analyses were performed using the R software, version 3.6.3; the data were stored and maintained using Microsoft Office Excel 2019.

Ethical approval statement. The collection of specimens, epidemiological and clinical data for SARS-CoV-2 infected individuals and their close contacts is a part of a continuing public health investigation of an emerging outbreak, ruled in the Protocol on the Prevention and Control of COVID-19 by the National Health Commission of the People's Republic of China, which was exempt from ethical approval (i.e., institutional review board assessment). However, staff of the local CDCs also informed the subjects that the collection of specimens would not pose any risk to personal privacy and information security before collecting specimens and took the specimens after obtaining verbal informed consents of the subjects. This study was approved by the ethics committee of the Hunan CDC with a waiver of informed consent for using the records of SARS-CoV-2 infected individuals and their contacts for scientific research (IRB No. 2020005).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The original database containing confidential patient information cannot be made public. The data that support the findings of this study is available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

We have created a synthetic dataset to demonstrate the structure of the underlying data. The code, the synthetic dataset, and aggregated de-identified data to reproduce all the tables and figures in the main text and Supplementary Information are publicly available on GitHub at https://github.com/KristyWang/Cluster_Hunan.

References

1. Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
2. World Health Organization. Coronavirus disease (COVID-19) pandemic. https://www.who.int/emergencies/diseases/novel-%ADcoronavirus-%AD2019 (2020).
3. World Health Organization. WHO coronavirus disease (COVID-19) dashboard. https://covid19.who.int (2021).
4. Van Lancker, W. & Parolin, Z. COVID-19, school closures, and child poverty: a social crisis in the making. *Lancet Public Health* **5**, e243–e244 (2020).
5. United Nations Educational, Scientific and Cultural Organization. Education: from disruption to recovery. https://en.unesco.org/covid19/educationresponse (2020).
6. Sinha, I. P. et al. COVID-19 infection in children. *Lancet Respir. Med.* **8**, 446–447 (2020).
7. Poletti, P. et al. 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 (2021).
8. Pollán, M. et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* **396**, 535–544 (2020).
9. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
10. Jing, Q. et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 1141–1150 (2020).
11. Wu, J. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020).
12. Oran, D. P. & Topol, E. J. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. *Ann. Intern. Med.* **173**, 362–367 (2020).
13. Baggett, T. P., Keyes, H., Sporn, N. & Gaeta, J. M. Prevalence of SARS-CoV-2 infection in residents of a large homeless shelter in Boston. *JAMA* **323**, 2191–2192 (2020).
14. Arons, M. M. et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N. Engl. J. Med.* **382**, 2081–2090 (2020).
15. Cheng, H. et al. High transmissibility of COVID-19 near symptom onset. Preprint at https://www.medrxiv.org/content/10.1101/2020.03.18.20034561v1 (2020).
16. Cereda, D. et al. The early phase of the COVID-19 outbreak in Lombardy, Italy. Preprint at https://arxiv.org/abs/2003.09320 (2020).
17. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286 (2020).
18. Du, Z. et al. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg. Infect. Dis.* **26**, 1341–1343 (2020).
19. Danis, K. et al. Cluster of coronavirus disease 2019 (COVID-19) in the French Alps, February 2020. *Clin. Infect. Dis.* **71**, 825–832 (2020).
20. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
21. Prete, C. A. et al. Serial interval distribution of SARS-CoV-2 infection in Brazil. *J. Travel Med.* taaa115, https://doi.org/10.1093/jtm/taaa115 (2020).
22. You, C. et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int. J. Hyg. Environ. Health* **228**, 113555 (2020).
23. Zhang, J. et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.* **20**, 793–802 (2020).
24. Linton, N. M. et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J. Clin. Med.* **9**, 538 (2020).
25. Backer, J. A., Klinkenberg, D. & Wallinga, J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eur. Surveill.* **25**, 2000062 (2020).
26. Bi, Q. et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919 (2020).
27. Tian, S. et al. Characteristics of COVID-19 infection in Beijing. *J. Infect.* **80**, 401–406 (2020).
28. Xu, T. et al. Clinical features and dynamics of viral load in imported and non-imported patients with COVID-19. *Int. J. Infect. Dis.* **94**, 68–71 (2020).
29. Park, S. W., Champredon, D. & Dushoff, J. Inferring generation-interval distributions from contact-tracing data. *J. R. Soc. Interface* **17**, 20190719 (2020).
30. Ali, S. T. et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* **369**, 1106–1109 (2020).
31. Ganyani, T. et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eur. Surveill.* **25**, 2000257 (2020).
32. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
33. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* **274**, 599–604 (2007).
34. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013).
35. Zhao, S. et al. Serial interval in determining the estimation of reproduction number of the novel coronavirus disease (COVID-19) during the early outbreak. *J. Travel Med.* **27**, taaa033 (2020).
36. Kimball, A. et al. Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility—King County, Washington, March 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 377–381 (2020).
37. National Health Commission of the People's Republic of China. The prevention and control of 2019 Novel Coronavirus Pneumonia (7th Edition). http://www.gov.cn/xinwen/2020-09/15/content_5543680.htm (2020).
38. Chen, Y. et al. The epidemiological characteristics of infection in close contacts of COVID-19 in Ningbo city. *Chin. J. Epidemiol.* **41**, 667–671 (2020).
39. Liu, Z., Chu, R., Gong, L., Su, B. & Wu, J. The assessment of transmission efficiency and latent infection period on asymptomatic carriers of SARS-CoV-2 infection. *Int. J. Infect. Dis.* **99**, 325–327 (2020).
40. Rothe, C. et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N. Engl. J. Med.* **382**, 970–971 (2020).
41. Li, W. et al. Characteristics of household transmission of COVID-19. *Clin. Infect. Dis.* **71**, 1943–1946 (2020).
42. Cauchemez, S. et al. Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *N. Engl. J. Med.* **361**, 2619–2627 (2009).
43. Chen, S. et al. Fangcang shelter hospitals: a novel concept for responding to public health emergencies. *Lancet* **395**, 1305–1314 (2020).
44. Dickens, B. L., Koo, J. R., Wilder-Smith, A. & Cook, A. R. Institutional, not home-based, isolation could contain the COVID-19 outbreak. *Lancet* **395**, 1541–1542 (2020).
45. Zhang, J. et al. Patterns of human social contact and contact with animals in Shanghai, China. *Sci. Rep.* **9**, 15141 (2019).
46. Guan, W. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
47. Johnstone-Robertson, S. P. et al. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. *Am. J. Epidemiol.* **174**, 1246–1255 (2011).
48. Ibuka, Y. et al. Social contacts, vaccination decisions and influenza in Japan. *J. Epidemiol. Community Health* **70**, 162–167 (2016).
49. Kiti, M. C. et al. Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya. *PLoS ONE* **9**, e104786 (2014).
50. Mossong, J. et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
51. Ajelli, M. & Litvinova, M. Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J. Theor. Biol.* **419**, 1–7 (2017).
52. Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D. & Lessler, J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann. Intern. Med.* **173**, 262–267 (2020).
53. Deng, X. et al. Case fatality risk of the first pandemic wave of novel coronavirus disease 2019 (COVID-19) in China. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciaa578 (2020).
54. Rockett, R. J. et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat. Med.* **26**, 1398–1404 (2020).
55. Lardieri, A. Hundreds of South Korean schools to close after reopening. U.S. News. https://www.usnews.com/news/world-report/articles/2020-05-29/hundreds-of-south-korean-schools-to-close-after-reopening (2020).
56. Manning, E. Primary school forced to close and pupils and staff told to quarantine after catering team member gets COVID-19. Yahoo News. https://uk.news.yahoo.com/coronavirus-primary-school-closes-caterer-075932511.html (2020).
57. Stein-Zamir, C. et al. A large COVID-19 outbreak in a high school 10 days after schools' reopening, Israel, May 2020. *Eur. Surveill.* **25**, 2001352 (2020).
58. National Health Commission of the People's Republic of China. Diagnosis and treatment guideline on pneumonia infection with 2019 novel coronavirus (6th trial edn). http://www.gov.cn/zhengce/zhengceku/2020-02/19/content_5480948.htm (2020).
59. Ashcroft, P. et al. COVID-19 infectivity profile correction. *Swiss Med. Wkly.* **150**, w20336 (2020).

Author contributions

S.H., W.W., Y.W., L.G., and H.Yu had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. L.G., M.A., and H.Yu were responsible for its conception and design. K.L., L.R., Q.S., X.C., G.Z., J.L., L.L., Z.D., W.Z., M.Li, H.Yang, J.G., K.W., X.C., Z.L., H.Yan, H.S., Z.C., and Y.Z. were responsible for the acquisition, analysis, or interpretation of data. S.H., W.W., M.Litvinova, M.A., and H.Yu drafted the manuscript. K.S., A.V., C.V., L.G., M.A., and H.Yu made critical revision of the manuscript for important intellectual content. W.W., Y.W., M.Litvinova, and M.A. did the data analysis. K.L., Q.S., G.Z., Z.D., H.Yang, Z.L., and K.S. provided administrative, technical, or material support. L.G., M.A., and H.Yu contributed equally to the supervision of this work.

Competing interests

A.V. reports a grant from Metabiota inc., M.A. has received research funding from Seqirus, and H.Y. has received research funding from Sanofi Pasteur, GlaxoSmithKline, Yichang HEC Changjiang Pharmaceutical Company, and Shanghai Roche Pharmaceutical Company. None of those research funding is related to COVID-19. All other authors report no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21710-6.

Correspondence and requests for materials should be addressed to L.G., M.A. or H.Y.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.