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

Shixiong Hu[1#], Wei Wang[2#], Yan Wang[2#], Maria Litvinova[3#], Kaiwei Luo[1], Lingshuang Ren[2], Qianlai Sun[1], Xinghui Chen[2], Ge Zeng[1], Jing Li[2], Lu Liang[4], 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[5], Alessandro Vespignani[3,6], Cécile Viboud[5], Lidong Gao[1†], Marco Ajelli[7†], Hongjie Yu[2†]

Author Affiliations:

1. Hunan Provincial Center for Disease Control and Prevention, Changsha, China (S. Hu, K. Luo, Q. Sun, G. Zeng, Z. Deng, H. Yang, Z. Liu, L. Gao)
2. School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China (W. Wang, Y. Wang, L. Ren, X. Chen, J. Li, W. Zheng, M. Li, J. Guo, K. Wang, X. Chen, H. Yan, H. Shi, Z. Chen, Y. Zhou, H. Yu)
3. ISI Foundation, Turin, Italy (M. Litvinova, A. Vespignani)
4. West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China (L. Liang)
5. Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA (K. Sun, C. Viboud)
6. Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA USA (A. Vespignani)
7. Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA (M. Ajelli)

[#] These authors are joint first authors and contributed equally to this work.

[†] These authors are joint senior authors and contributed equally to this work.

Corresponding Authors:

Lidong Gao, MSc, Hunan Provincial Center for Disease Control and Prevention, No. 450 Middle Furong Road, Kaifu District, 410005, Changsha, China (810173358@qq.com)

Marco Ajelli, PhD, Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, 47405 IN, USA (marco.ajelli@gmail.com)

Hongjie Yu, PhD, School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, No. 138 Yixueyuan Road, Xuhui District, 200032, Shanghai, China (yhj@fudan.edu.cn)

Word count (abstract): 378

Word count (main text): 3504

Running head: Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission

Abstract

Background

Several parameters driving the transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) remain unclear, including age-specific differences in infectivity and susceptibility, and the contribution of inapparent infections to transmission. Robust estimates of key time-to-event distributions remain scarce as well.

Methods

We collected individual records for 1,178 SARS-CoV-2 infected individuals and their 15,648 contacts identified by contact tracing and monitoring over the period from January 13 to April 02, 2020 in Hunan Province, China. We provide descriptive statistics of the characteristics of cases and their close contacts; we fitted distributions to time-to-key-events distributions and infectiousness profile over time; and we used generalized linear mixed model to estimate risk factors for susceptibility and transmissibility of SARS-CoV-2.

Results

We estimated the mean serial interval at 5.5 days (95%CI -5.0, 19.9) and the mean generation time at 5.5 days (95%CI 1.7, 11.6). The infectiousness was estimated to peak 1.8 days before symptom onset, with 95% of transmission events occurring between 7.6 days before and 7.3 days after the date of symptom onset. The proportion of pre-symptomatic transmission was estimated to be 62.5%. We estimated that at least 3.5% of cases were generated asymptomatic individuals. SARS-CoV-2 transmissibility was not significantly different between working-age adults (15-59 years old) and other age groups (0-14 years old: p-value=0.16; 60 years and over: p-value=0.33), whilst susceptibility to SARS-CoV-2 infection was estimated to increase with age (p-value=0.03). In addition, transmission risk was higher for household contacts (p-value<0.001), decreased for higher generations within a cluster (second generation: odds ratio=0.13, p-value<0.001; generations 3-4: odds ratio=0.05, p-value<0.001, relative to generation 1), and decreased for infectors with a larger number of contacts (p-value=0.04).

Interpretation

Our findings warn of the possible relevant contribution of children to SARS-CoV-2 transmission. When lockdown interventions are in place, we found that odds of transmission are highest in the household setting but, with the relaxation of interventions, other settings (including schools) could bear a higher risk of transmission. Moreover, the estimated relevant fraction of pre-symptomatic and asymptomatic transmission highlight the importance of large-scale testing, contact tracing activities, and the use of personnel protective equipment during the COVID-19 pandemic.

Key words: transmissibility, risk factors, contact tracing, coronavirus

Introduction

The outbreak of coronavirus disease 2019 (COVID-19) started in December 2009 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 18.4 million SARS-CoV-2 infected individuals confirmed worldwide as of August 06, 2020 [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 thus far to disentangle. A related question is the probability of asymptomatic transmission from young individuals. 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 households, homeless shelters, and nursing facilities, supports the role of pre-symptomatic and asymptomatic transmission [10,12-16]. 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, especially as countries relax costly lockdown policies and move towards case-based interventions (e.g., case isolation, quarantine of contacts, contact tracing). 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 [1,15,17]. It is important to update these estimates using large-scale and harmonized epidemiological datasets.

In this study, we analyze 1,178 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-April 02, 2020. This comprehensive and detailed dataset compiled by the Hunan Provincial CDC sheds light on SARS-CoV-2 transmission patterns, risk factors, and the distribution of key time-to-event parameters.

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) [18]. 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 (Appendix 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). Confirmed COVID-19 cases were categorized by clinical severity, including mild, moderate, severe and critical illnesses (as defined in Appendix, Tab. S1).

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 her/his 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 [18]. 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.

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 (Appendix p3-4).

Close contacts, sporadic cases, and clusters

Close contacts were defined as individuals who had close-proximity interactions (within 1 meter) 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., a work colleague or classmate), 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).

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 (Appendix p3). Epidemiologically linked cases were classified according to the generation time 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) who did not belong to any of the reported clusters.

We define pre-symptomatic transmission as a direct transmission event that takes place before the date of symptom onset of the infector, while asymptomatic transmission is a transmission event from a person who never developed symptoms.

Ethical approval statement

This study was approved by the ethic committee of the Hunan CDC with a waiver of informed consent due to a public health outbreak investigation (IRB No. 2020005).

Role of the funding source

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Statistical analysis

We provide descriptive statistics of the characteristics of cases and their close contacts, including demographic factors and exposures (Appendix p5-p7). 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 her/his secondary cases), the generation time (i.e., the time interval between infection of the primary case and of her/his secondary cases), and the infectiousness profile (i.e., the daily distribution of the probability of transmission since the date of symptom onset; see [15,19] and Appendix p7-p10 for methods). We also estimated the interval from symptom onset to the sampling date of first PCR by using a maximum likelihood estimator and fitting three distributions (Weibull, gamma, and lognormal) (Appendix p10). 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 (Appendix p11). Additionally, generalized linear mixed-effects model, GLMM, for binary data with logit link were 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 8,159 individual records of contacts who were exposed to locally transmitted cases (see appendix p11-12). 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 3.5.0.

Results

Sample description

Between January 23, 2020 and April 02, 2020, 1,019 symptomatic cases and 159 asymptomatic subjects were reported and screened for inclusion (Fig. S1 and Tab. 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 1,178 SARS-CoV-2 infections, we identified 831 epidemiologically linked cases in 210 clusters. Of these clusters,

499 SARS-CoV-2 infections in 123 clusters had a clear epidemiological link to a previous SARS-CoV-2 infected individual. From 15,648 close contacts, 6,412 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 9,236 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 Fig. S2. 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 (Tab. 1). Cases aged 0-19 years presented milder or no clinical symptoms, while patients aged 40 years and older had more severe illness (P<0.001).

Time-to-key-event distributions

We analyzed 268 locally-acquired confirmed cases belonging to 114 clusters, with information on both the date(s) of exposure and transmission generation in the cluster. We found that the best fitting distribution of incubation period was a Weibull distribution with a mean of 6.4 days (95% CI: 0.7, 16.6 days) (Tab. S3). We performed a sensitivity analysis excluding cases having only exposure end date (17 individuals) and we obtained similar estimates (Appendix, Tab. S3). Symptom onset dates were available for 245 transmission pairs; the resulting serial interval was estimated to have a mean of 5.5 days (95%CI: -5.0, 19.9 days) and a median of 4.8 days, based on a fitted 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. The mean time interval from symptom onset to the sampling date of first PCR was estimated to be 4.7 days (95% CI: -2.9, 14.7 days) using the best fitting gamma distribution, based on 531 PCR positive individuals. The generation time was estimated to be 5.5 days (95% CI: 1.7, 11.6 days). The estimated distributions of the incubation period and of the generation time show stark similarities (Fig. 2B).

Pre-symptomatic transmission

Infectiousness was estimated to peak 1.8 days before symptom onset (Fig. 2A). We estimated the proportion of pre-symptomatic transmission (area under the curve, Fig. 2A) at 62.5%, with 95% of transmission events occurring between -7.6 days and 7.3 days of the date of symptom onset, under the intensive contact tracing and isolation strategy undertaken by the Hunan Province. From the analysis of the transmission chains reconstructed by field investigations, 43 pre-symptomatic transmission events were recorded in 23 clusters. A subset of those clusters is shown in Fig. 3A.

Asymptomatic transmission

From the analysis of contact tracing records, we identified 8 clusters with evidence of asymptomatic transmission. There were 11 asymptomatic infectors (5 primary and 6 secondary infections) associated with 15 of 25 local transmission events (10 secondary and 5 tertiary, Fig. 3B).

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 (Fig. S4). The results suggest that people aged 15-59 years generated a larger mean number of cases than younger (0-14 years old) and older (60+ years old) individuals. Moreover, individuals over 60 years were infected more often, suggesting increased susceptibility. Next, to account for the possible effect of multiple confounding factors on the probability of transmission, we performed a multivariate regression analysis. 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 (Tab. 2). Infectiousness was not significantly different between working-age adults (15-59 years old) and other age groups (0-14 years old: p-value=0.162; 60 years and over: p-value=0.332]); in contrast, susceptibility to SARS-CoV-2 infection increased with age (p-value=0.028, Model 2 in Tab. 2). Further, household contacts were associated with a significantly larger risk of SARS-CoV-2 infection than other types of contact. The GLMM model suggests two other statistically significant risk factors: the generation in the transmission chain and the number of contacts identified for an infector (Tab. 2). In particular, the transmission risk in the first generation was significantly higher than the later generations, possibly due to improved case isolation and contacts quarantine that deplete the number of susceptible individuals in the cluster. 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/contacts and clinical severity of an infector were not statistically significant, did not modify the estimated odds ratios for the other variables, and did not improve the fit of the model (Tab. S6, Tab. S7, and Fig. S5). In particular, this analysis indicates no statistically significant difference in transmissibility between symptomatic and asymptomatic individuals.

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 infectiousness by age, while susceptibility to SARS-CoV-2 infection increased with age. We provide evidence of both pre-symptomatic and asymptomatic SARS-CoV-2 transmission, with the former potentially accounting for up to 62.5% of all transmission events in this dataset. In addition, we estimate that SARS-CoV-2 transmission in households is responsible for most of 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 of the serial interval and incubation period were obtained from a limited number of infector-infectee pairs or from different data sources, thus suffering of large uncertainty [20,21]. This may explain the large variability of the estimates, ranging from 4.0 days to 7.8 days for the serial interval [1,15,21-24] and from 4.8 days to 8.0 days for the incubation period [1,23,25-29]. Our estimates fall within these intervals. Unlike the serial interval and the incubation period, only a few studies [30,31] provide estimates of the generation time, which is hard to directly infer from field investigations, as it requires knowledge of the infection date of both the infector and her/his infectees. Here, following an approach similar to He, et al [15], we estimate the mean generation time at 5.5 days, in general agreement with Ferretti, et al [31].

Previous studies show a relatively high proportion of pre-symptomatic transmission, but estimates vary significantly, ranging between 13-62% [15,31,32]. Our estimate (62.5%) is on the high end of the range found in the literature. This is 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 analysis suggests a key role of interventions (e.g., contact tracing and case isolation) in decreasing the risk of infection, as the risk

of infection decreased with the number of the generations in the transmission chain.

We found evidence of asymptomatic transmission in several clusters, with 15 secondary cases linked to asymptomatic infectors. Other studies provide evidence of asymptomatic infection [12,32,33], but do not quantify its contribution to transmission. In our study, we cannot provide a point estimate, as a fraction of asymptomatic infections may have been missed despite extensive PCR testing performed by the Hunan CDC. In fact, testing of contacts only focused on symptomatic individuals before February 7, 2020, while all contacts were tested afterwards. Therefore, our finding cannot be used to quantitatively estimate the percentage of asymptomatic infections. However, we can provide a lower bound of the share of transmission events associated with asymptomatic infectors. We estimated this share to be at least 3.5% (15/432) of transmission events, in agreement with Chen, et al (4.5% (6/132), p=0.602) [34] and Liu, et al (2.6% (24/914), p=0.402) [35].

In agreement with previous studies, we found that the risk of infection from a household member is larger than those resulting from other contacts [10,36]. 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 [37], we found that SARS-CoV-2 transmissibility decreased with the number of contacts, although the effect is small. Further studies are needed to explain this connection.

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 [23], 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 infecting for first-generation infectors was significantly higher than the ones for the later generation infectors. Together with a small number of infectious children in the first generation, this contributed to observing a lower total number of infections generated by children (see Tab. S9). However, when accounting for all confounding factors, including the generation, we found no statistical evidence of differential transmissibility by age group (Tab. 2). Moreover, comparing with the mean of number of reported contacts (range: 18.2-22.3) before COVID-19 epidemic in China and other countries [9,38-40], the number of individual contacts of each infectious child in contact tracing was considerably lower (mean: 7.72) during the outbreak in Hunan. This implies that caution should be applied when evaluating policies that increase the number of contacts among children, such as re-opening of 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]).

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 her/his contacts, we could not always accurately reconstruct the entire transmission chain and avoid recall bias in individual records. Moreover, we cannot rule out the possibility of indirect exposures (e.g., contaminated surfaces), which may affect the identification of epidemiological links. Second, our sample size did not allow us to distinguish between different time periods of the pandemic in Hunan, while controlling for all the other covariates. Changes in population awareness and reactive behavioral response to the outbreak may affect the estimates provided in this study.

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 that transmission can occur up to 7 days before symptoms onset lends support to personal precautions such as mask wearing. 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 in several occasions [5,41-43]; time will tell whether schools can become a major foci of transmission in the coming months.

Author contributions

S. Hu, W. Wang, Y. Wang, L. Gao, 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. Gao, M. Ajelli, and H. Yu were responsible for its conception and design. K. Luo, L. Ren, Q. Sun, X. Chen, G. Zeng, J. Li, L. Liang, Z. Deng, W. Zheng, M. Li, H. Yang, J. Guo, K. Wang, X. Chen, Z. Liu, H. Yan, H. Shi, Z. Chen, and Y. Zhou were responsible for the acquisition, analysis, or interpretation of data. S. Hu, W. Wang, M. Litvinova, M. Ajelli and H. Yu drafted the manuscript. K. Sun, A. Vespignani, C. Viboud, L. Gao, M. Ajelli, H. Yu made critical revision of the manuscript for important intellectual content. W. Wang, Y. Wang, M. Litvinova, and M. Ajelli did the data analysis. K. Luo, Q. Sun, G. Zeng, Z. Deng, H. Yang, Z. Liu, and K. Sun provided administrative, technical, or material support.

Conflicts of Interest Disclosures

Hongjie Yu 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.

Funding

National Science Fund for Distinguished Young Scholars (No. 81525023), National Science and Technology Major Project of China (No. 2017ZX10103009-005, No. 2018ZX10713001-007, No. 2018ZX10201001-010), and Hunan Provincial Innovative Construction Special Fund: Emergency response to COVID-19 outbreak (No. 2020SK3012).

Reference

1. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine* 2020.
2. World Health Organization. Coronavirus disease (COVID-19) pandemic. 2020. https://www.who.int/emergencies/diseases/novel-%ADcoronavirus-%AD2019.
3. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 170. 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200805-covid-19-sitrep-198.pdf?sfvrsn=f99d1754_2 (accessed August 06 2020).
4. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. *The Lancet Public Health* 2020; **5**(5): e243-e4.
5. United Nations Educational SaCO. Education: From disruption to recovery. 2020. https://en.unesco.org/covid19/educationresponse (accessed Jul. 9 2020).
6. Sinha IP, Harwood R, Semple MG, et al. COVID-19 infection in children. *The Lancet Respiratory Medicine* 2020; **8**(5): 446-7.
7. Poletti P, Tirani M, Cereda D, et al. Probability of symptoms and critical disease after SARS-CoV-2 infection. *arXiv e-prints*, 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200608471P (accessed June 01, 2020).
8. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet*.
9. Zhang J, Litvinova M, Liang Y, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* 2020; **368**(6498): 1481-6.
10. Jing QL, Liu MJ, Zhang ZB, et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *Lancet Infect Dis* 2020.
11. Wu JT, Leung K, Bushman M, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 2020; **26**(4): 506-10.
12. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection: A Narrative Review. *Ann Intern Med* 2020.
13. Baggett TP, Keyes H, Sporn N, Gaeta JM. Prevalence of SARS-CoV-2 Infection in Residents of a Large Homeless Shelter in Boston. *JAMA* 2020.
14. Arons MM, Hatfield KM, Reddy SC, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *N Engl J Med* 2020; **382**(22): 2081-90.
15. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* 2020.
16. Cheng HY, Jian SW, Liu DP, Ng TC, Huang WT, Lin HH. Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. *JAMA internal medicine* 2020.
17. Cereda D, Tirani M, Rovida F, et al. The early phase of the COVID-19 outbreak in Lombardy, Italy. *arXiv e-prints*, 2020. https://ui.adsabs.harvard.edu/abs/2020arXiv200309320C (accessed March 01, 2020).
18. National Health Commission of the People's Republic of China. Diagnosis and treatment guideline on pneumonia infection with 2019 novel coronavirus (6th trial edn). 2020.
19. Ashcroft P, Huisman JS, Lehtinen S, et al. COVID-19 infectivity pro1le correction. *ArXiv e-prints* 2020.
20. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*.
21. Du Z, Xu X, Wu Y, Wang L, Cowling B, Meyers LA. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerging Infectious Disease journal* 2020; **26**(6): 1341.
22. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis* 2020; **93**: 284-6.
23. Bi Q, Wu Y, Mei S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their

close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases* 2020.
24. Ali ST, Wang L, Lau EHY, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* 2020.
25. Zhang J, Litvinova M, Wang W, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect Dis* 2020; **20**(7): 793-802.
26. Linton NM, Kobayashi T, Yang Y, 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* 2020; **9**(2).
27. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill* 2020; **25**(5).
28. Tian S, Hu N, Lou J, et al. Characteristics of COVID-19 infection in Beijing. *J Infect* 2020; **80**(4): 401-6.
29. Xu T, Chen C, Zhu Z, et al. Clinical features and dynamics of viral load in imported and non-imported patients with COVID-19. *Int J Infect Dis* 2020; **94**: 68-71.
30. Ganyani T, Kremer C, Chen D, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveill* 2020; **25**(17).
31. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 2020; **368**(6491).
32. Kimball A, Hatfield KM, Arons M, 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* 2020; **69**(13): 377-81.
33. Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *New England Journal of Medicine* 2020.
34. Chen Yi WA, Yi Bo, Ding Keqin, Wang Haibo, Wang Jianmei, Shi Hongbo, Wang Sijia, Xu Guozhang. The epidemiological characteristics of infection in close contacts of COVID-19 in Ningbo city. *Chinese Journal of Epidemiology* 2020; **41**(00).
35. 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* 2020.
36. Li W, Zhang B, Lu J, et al. The characteristics of household transmission of COVID-19. *Clin Infect Dis* 2020.
37. Cauchemez S, Donnelly CA, Reed C, et al. Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *N Engl J Med* 2009; **361**(27): 2619-27.
38. Johnstone-Robertson SP, Mark D, Morrow C, et al. Social Mixing Patterns Within a South African Township Community: Implications for Respiratory Disease Transmission and Control. *American Journal of Epidemiology* 2011; **174**(11): 1246–55.
39. Ibuka Y, Ohkusa Y, Sugawara T, et al. Social contacts, vaccination decisions and influenza in Japan. *J Epidemiol Community Health* 2016; **70**(2): 162-7.
40. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying Age-Related Rates of Social Contact Using Diaries in a Rural Coastal Population of Kenya. *PLoS One* 2014; **9**(8): e104786.
41. U.S.News. Hundreds of South Korean Schools to Close After Reopening. 2020.
42. Yahoo News. Primary school forced to close and pupils and staff told to quarantine after catering team member gets COVID-19. 2020.
43. Stein-Zamir C, Abramson N, Shoob H, et al. A large COVID-19 outbreak in a high school 10 days after schools' reopening, Israel, May 2020. *Euro Surveill* 2020; **25**(29): 2001352.

Tables

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

Characteristics	Symptomatic cases (n=1,019)	Asymptomatic subjects (n=159)	Close contacts of cases with SARS-CoV-2 infections (n=15,648)[a,b]
Age, years			
Median (interquartile range, IQR)	45 (34-55)	36 (19-52)	40 (27-52)
0-9	25 (2.5)	19 (11.9)	1,124 (7.2)
10-19	23 (2.3)	22 (13.8)	1,167 (7.5)
20-29	119 (11.7)	19 (11.9)	2,019 (12.9)
30-39	213 (20.9)	29 (18.2)	2,880 (18.4)
40-49	242 (23.7)	27 (17.0)	2,892 (18.5)
50-59	202 (19.8)	20 (12.6)	2,604 (16.6)
60-69	120 (11.8)	16 (10.1)	1,234 (7.9)
70-79	55 (5.4)	6 (3.8)	686 (4.4)
≥80	20 (2.0)	1 (0.6)	278 (1.8)
Missing	0 (0)	0 (0)	764 (4.9)
Sex			
Male	526 (51.6)	75 (47.2)	7,984 (51)
Female	493 (48.4)	84 (52.8)	7,397 (47.3)
Missing	0 (0)	0 (0)	267 (1.7)
Exposure history[b]			
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	-	-	2,771 (17.7)
Relative contacts	-	-	7,284 (46.5)
Social contacts	-	-	4,550 (29.1)
Other close contacts	-	-	5,709 (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)
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. 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 1,019 symptomatic COVID-19 cases and 159 subjects with asymptomatic SARS-CoV-2 infections.

b. Percentages may not total 100 because of one individual associated with multiple observed exposures and contacts.

c. 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.

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	Model 1*		Model 2*	
			OR (95%CI)	P-value	OR (95%CI)	P-value
Intercept	-	-	0.27 (0.11, 0.65)	0.004	0.02 (0, 0.21)	0.002
Age of infectors						
0-14 y	25	193	0.25 (0.04, 1.75)	0.162	-	-
15-64 y	355	6,833	Reference	-	-	-
65+ y	81	1,133	0.64 (0.26, 1.59)	0.332	-	-
Log-transformed age	461	8,159	-	-	1.62 (0.91, 2.90)	0.104
Age of contacts						
0-14 y	22	936	0.58 (0.34, 0.98)	0.042	-	-
15-64 y	154	6,411	Reference	-	-	-
65+ y	33	812	1.65 (1.03, 2.65)	0.038	-	-
Log-transformed age	209	8,159	-	-	1.26 (1.02, 1.55)	0.028
Type of contact						
Household contacts	-	1,021	Reference	-	Reference	-
Relative contacts	-	3,084	0.11 (0.07, 0.17)	<0.001	0.11 (0.07, 0.18)	<0.001
Social contacts	-	2,227	0.06 (0.03, 0.11)	<0.001	0.06 (0.03, 0.11)	<0.001
Other contacts	-	1,827	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.13)	<0.001
Generation of COVID-19 transmission						
G1	-	2,121	Reference	-	Reference	-
G2	-	2,987	0.13 (0.06, 0.31)	<0.001	0.12 (0.05, 0.28)	<0.001
G3-4	-	965	0.05 (0.02, 0.19)	<0.001	0.05 (0.02, 0.18)	<0.001
Multiple exposure[a]	-	598	0.10 (0.03, 0.41)	0.001	0.10 (0.03, 0.40)	0.001
Unknown	-	1,488	0.03 (0.01, 0.10)	<0.001	0.03 (0.01, 0.10)	<0.001
Infector's contacts						
Total number of contacts	-	8,159	0.99 (0.97, 1.00)	0.028	0.99 (0.97, 1.00)	0.036
Gender of infectors						
Female	239	4,067	Reference	-	Reference	-
Male	222	4,092	1.76 (0.97, 3.21)	0.063	1.77 (0.97, 3.21)	0.061
Gender of contacts						
Female	424	4,017	Reference	-	Reference	-
Male	427	4,142	1.02 (0.74, 1.40)	0.907	1.01 (0.73, 1.38)	0.973

a: Contacts who were exposed to multiple cases of different generations of COVID-19 transmissions.

* Age of infectors and contacts were considered either as categorical (model 1) or continuous log-transformed (model 2) variables.

Figure Legends

Figure. 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 (date of symptom onset) by source of infection. (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. (F) Geographical distribution of sporadic and clustered individuals with SARS-CoV-2 infections. (E) Geographical distribution of SARS-CoV-2 symptomatic and asymptomatic infected individuals by source of infection. (F) Geographical distribution of close contacts of SARS-CoV-2 symptomatic and asymptomatic infected individuals. (G) Age distribution of SARS-CoV-2 symptomatic and asymptomatic infected individuals.

Figure. 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. (B) Estimated distribution of the incubation period by Weibull distributions and of the generation time by gamma distributions.

Figure. 3. Timing of transmission events and SARS-CoV-2 infected individuals in clusters showing evidence of pre-symptomatic and asymptomatic transmission. (A) Transmission chain in 8 randomly selected clusters with pre-symptomatic SARS-CoV-2 transmission. (B) Transmission chain in all the clusters showing evidence of asymptomatic SARS-CoV-2 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 diagnosis to the first RT-PCR positive for asymptomatic subjects. Timeline of events (right panels).

Supporting information

Contents

Definitions of cases of SARS-CoV-2 infection 2
Specimen collection and laboratory testing 5
Overview of COVID-19 epidemics in Hunan Province, China 5
Characteristics of clusters of COVID-19 7
Incubation period 7
Serial interval 8
Infectiousness profile over time 9
Generation time 10
Interval from symptom onset to the sampling date of first PCR 10
Total and mean number of infections by age of infector and of infectee 11
Quantifying the impact of potential drivers on the susceptibility and infectivity of SARS-CoV-2 13
References 26

Definitions of cases of SARS-CoV-2 infection

Suspected COVID-19 cases

A suspected COVID-19 case is defined as a person who meets three clinical criteria OR two clinical criteria and one of epidemiological criteria:

a) Clinical criteria: i) acute respiratory illness; ii) radiographic evidence of COVID-19 viral pneumonia; iii) normal or decreased white blood cells count in the early stage of the disease and normal or decreased lymphocyte count.

b) Epidemiological criteria: i) history of travel to or residence in Wuhan or domestic location reporting community transmission or countries/territories/areas/overseas reporting widespread SARS-CoV-2 transmission during the 14 days prior to symptom onset; ii) contact with any confirmed cases during the 14 days prior to symptom onset; iii) cluster of contact with COVID-19 patients (nucleic acids amplification test positive) within 14 days before symptom onset or to individuals with fever and/or symptoms of respiratory infection within 14 days.

Clinical severity of COVID-19 confirmed cases

We categorized confirmed COVID-19 cases according to their clinical severity, i.e., mild, moderate, severe, and critical case-patients. The details are presented in Tab. S1.

Table S1. *Definitions of clinical severity of COVID-19 cases.* [1]

Clinical severity	Definition[1]
Mild	Patients with mild symptoms, and no radiographic evidence of pneumonia
Moderate	Patients with fever, respiratory symptoms, and radiographic evidence

	of pneumonia
Severe	Patients had any of the following: a. respiratory distress, breathing rate ≥30 beats/min; or b. finger oxygen saturation ≤93% during resting state; or c. PaO_2/FiO_2 ≤300mmHg (1mmHg = 0.133kPa). Patients whose pulmonary imaging have obvious progress of lesions (>50%) within 24~48 hours are managed by severe case.
Critical	Patients had any of the following: a. respiratory failure and requires mechanical ventilation; or b. shock; or c. with other organ failures that requires ICU admission.

Epidemiologically-linked COVID-19 cases

An individual with an epidemiologic link is a SARS-CoV-2 infected individual who has either been exposed to a symptomatic or an asymptomatic individual, or had the same exposure as the SARS-CoV-2 infected individuals. Generally, epidemiologically-linked cases include, but are not limited to SARS-CoV-2 infected individuals' 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., a work colleague or classmate), and other close contacts (i.e., caregivers and patients in the same ward, persons sharing a vehicle, and those providing a service for the case in public places) who have been close-proximity interactions (within 1 meter) with index case-patient and have acquired SARS-CoV-2 infections.

The flowchart describing the selection criteria of the analyzed subjects is shown in Fig. S1.

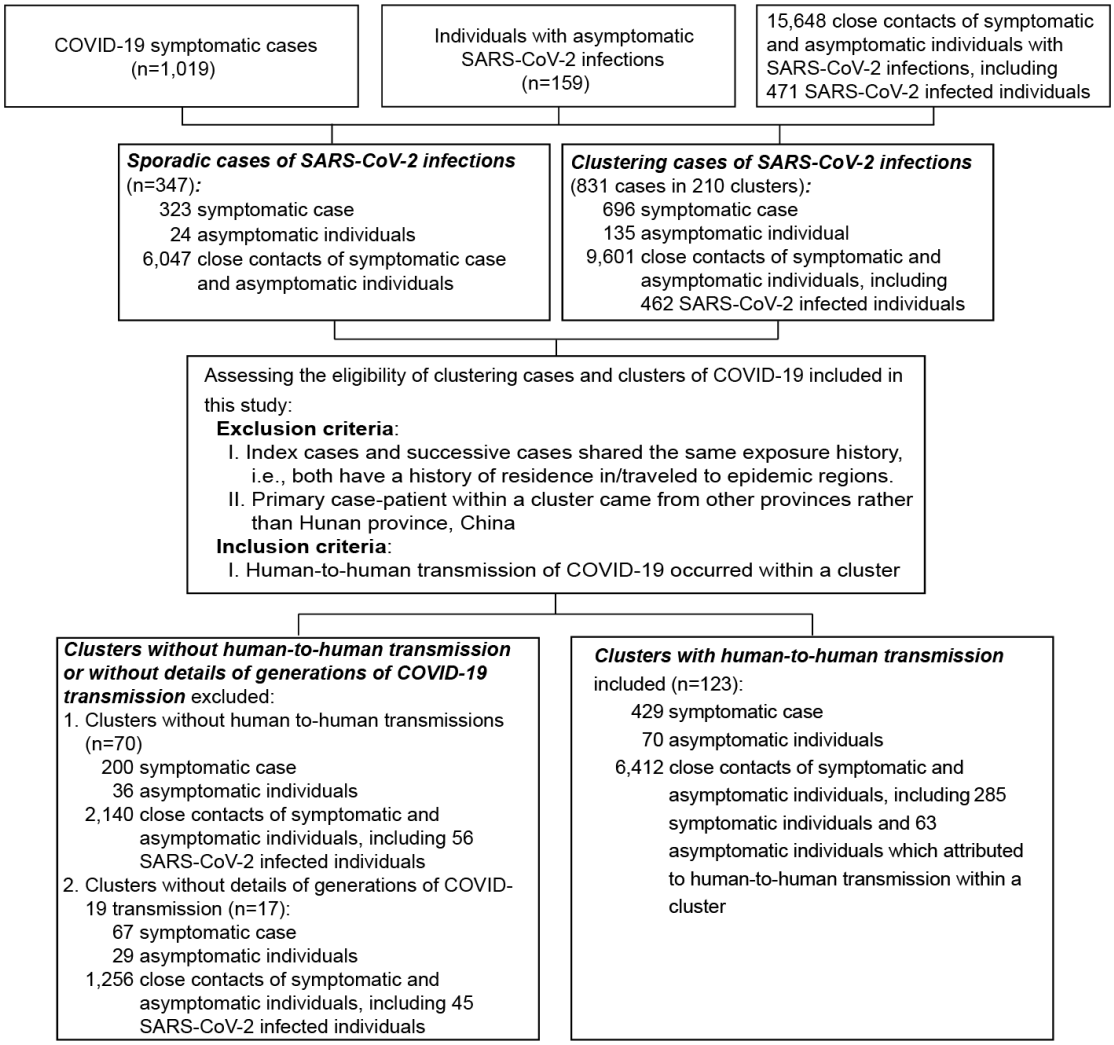


Figure S1. *Flowchart describing the COVID-19 symptomatic cases, individual with asymptomatic infections, close contacts, and clusters with human-to-human transmission.*

Specimen collection and laboratory testing

Since January 27, the designated hospitals and local Centers for Diseases Prevention and Controls (CDC) were approved to conduct real-time RT-PCR assay for diagnosis of COVID-19 using uniform laboratory testing procedures established by the World Health Organization (WHO). Total RNA was extracted using automated Nucleic Acid Extraction System 9600E (Xi'an TianLong Science and Technology Co., Ltd., Xi'an, China). Real-time reverse transcription polymerase chain reaction (RT-PCR) assay for SARS-CoV-2 was performed using a SARS-CoV-2 ORF1ab/N gene detection kit (Biogerm Medical Biotechnology Co.,Ltd, Shanghai, China), a product based on the recommendation of the National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention. The open reading frame 1ab gene (ORF1ab) and nucleocapsid gene (N) were amplified and tested. Results were reported positive when both the ORF1ab gene and N gene were positive. Specimens tested as Ct-value of >=35 and <39.2 were retested for confirmation, a retest Ct-value of >=39.2 was treated as positive, otherwise negative.

Overview of COVID-19 epidemics in Hunan Province, China

Overall, the dynamics of the epidemic in Hunan followed an exponential growth before January 23, 2020, and a decrease in the number of cases after February 1, 2020 (Fig. S2). Age descriptive statistics by other covariates are shown in Fig. S3.

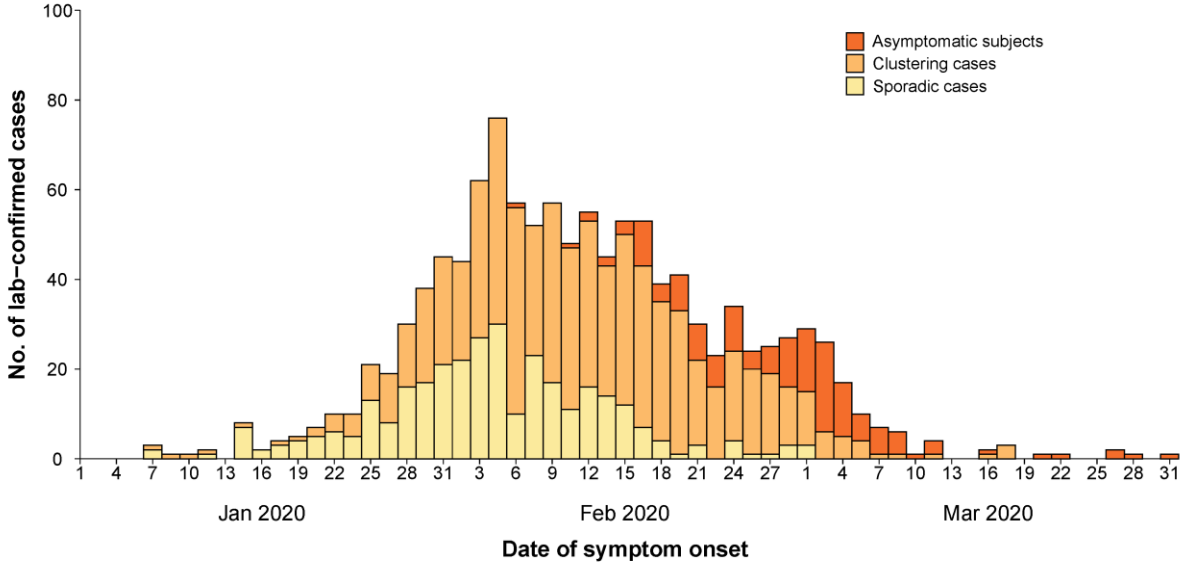


Figure S2. *Epidemic curve of sporadic COVID-19 cases, individuals with asymptomatic SARS-COV-2 infections, and clustering COVID-19 cases in Hunan, China.*

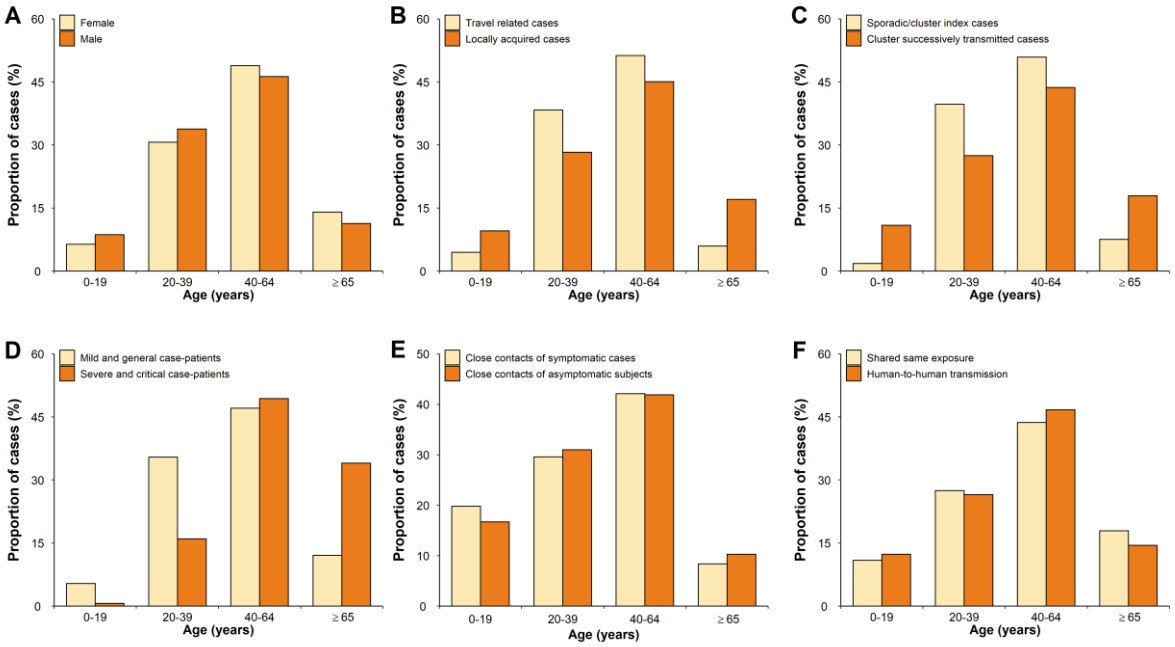


Figure S3. *Age distribution of COVID-19 symptomatic cases, asymptomatic subjects as well as their close contacts. (A) Age and sex distribution of COVID-19 symptomatic cases. (B) Age distributions of travel related COVID-19 cases and locally acquired COVID-19 cases. (C)*

Age distributions of sporadic cases, index cases and successively transmitted cases in clusters. (D) Age distribution by clinical severity. (E) Age distribution of contacts. (F) Age distributions of COVID-19 cases by type of exposure.

Characteristics of clusters of COVID-19

Cluster size was defined as the total number of *COVID-19 symptomatic cases* and asymptomatic individuals in a cluster. We characterized 123 clusters with clear evidence of human-to-human transmission, which includes 499 of the COVID-19 cases presented in Tab. S2. Cluster size distribution was bimodal, with most clusters were between 2 and 4 cases (94/123, corresponding to 76.4%). The largest cluster included 20 cases. The median cluster size was 3 (Tab. S2).

Table S2. *Characteristics of cluster with human-to-human transmission COVID-19 identified in Hunan province, China.*

Cluster size[a]	Number of clusters of a given size (n=123)	Total number of cases (n=499)	Index case (n=142)	Case with single exposure*			Cases with multiple exposures* (n=51)
				Secondary case (n=234)	Third-generation case (n=59)	Fourth-generation case (n=13)	
2 cases	45 (36.6)	90 (18.0)	45 (31.7)	45 (19.2)	0 (0)	0 (0)	0 (0)
3 cases	26 (21.1)	78 (15.6)	34 (23.9)	38 (16.2)	0 (0)	0 (0)	6 (11.8)
4 cases	23 (18.7)	92 (18.4)	29 (20.4)	42 (17.9)	8 (13.6)	0 (0)	13 (25.5)
5 cases	10 (8.1)	50 (10.0)	11 (7.7)	27 (11.5)	6 (10.2)	1 (7.7)	5 (9.8)
>5 cases	19 (15.4)	189 (37.9)	23 (16.2)	82 (35.0)	45 (76.3)	12 (92.3)	27 (52.9)

a. Index cases were included when calculating the clusters size for each cluster.

b. Number of clustering cases at a given size of group.

Incubation period

We estimated the time from infection to symptom onset (i.e., the incubation period) based on

information about the likely exposure of confirmed COVID-19 cases. Exposure was ascertained for cases without a history of travel to Wuhan/Hubei. The exposure information was provided in the form of a time interval bounded by the dates of the first and last possible exposure. If the exposure start date of the case was missing or before that of the first infector, it was replaced by the exposure start date of the first infector. For the rest cases without dates of first exposure, they were imputed by the random numbers generated from a gamma distribution that best fitted the data of time intervals between the first and last exposure. We also performed a sensitivity analysis excluding these cases. We estimated the distribution of interval-censored exposure data by using maximum likelihood and compared three distributions (Weibull, gamma, and lognormal). The goodness of fit was assessed using Akaike information criterion (AIC). Results are presented in Tab. S3.

Table S3. *Estimates of the incubation period based on the analysis of 114 clusters and 268 cases.*

Distribution	Parameters [mean (SD)]	Mean (days)	95%CI (days)	AIC
Gamma	shape = 2.08(0.21), rate = 0.33 (0.04)	6.3	0.8 – 17.7	782.6
Weibull	shape = 1.58 (0.09), scale=7.11(0.33)	6.4	0.7 – 16.6	775.5
Lognormal	meanlog = 1.57(0.06), sdlog = 0.82(0.04)	6.7	1.0 – 23.9	839.7
Sensitivity analysis				
gamma	shape = 2.05(0.22), rate = 0.33 (0.04)	6.1	0.7 – 17.1	733.6
Weibull	shape = 1.57 (0.09), scale=6.85(0.33)	6.2	0.6 – 16.1	726.8
lognormal	meanlog = 1.53(0.06), sdlog = 0.83(0.05)	6.4	0.9 – 23.3	788.4

Serial interval

We analyzed clusters of COVID-19 cases with known epidemiological links and no travel history to Wuhan/Hubei to estimate the interval between onset of symptoms in primary (index)

cases and the onset of symptoms in secondary cases generated by these primary cases (i.e., the serial intervals). For cases with several possible infectors, a time interval bounded by the symptom-onset dates of the first and last possible infectors was provided as the symptom onset interval of primary cases. We fitted a gamma distribution of interval-censored data by maximum likelihood to estimate the serial interval distribution, using dates of symptom onsets for consecutive generations of cases within clusters. Results are shown in the main text.

Infectiousness profile over time

Following the approach similar to He, et al [2], and accounting for the correction proposed by Ashcroft, et al [3], the infectiousness profile (i.e., transmission probability from primary cases to a secondary case) was inferred using the serial intervals from confirmed transmission pairs combined with the incubation period distribution fitted in our analysis. Assuming that the infectiousness profile $\beta_c(t_I - t_{S1})$ follows a gamma distribution with a time shift c to allow for start of infectiousness (t_I) c days prior to the date of symptom onset (t_{S1}). The serial intervals distribution $f(t_{S2} - t_{S1})$ would be the convolution between the infectiousness profile and incubation period distribution $g(t_{S2} - t_I)$, where t_{S2} is the date when secondary case shows symptoms. The parameter vector θ, which includes shape and scale of the gamma distribution and the time shift c, were estimated using maximum likelihood based on the convolution of serial interval and incubation period. Allowing for the start of infectiousness to be around symptom onset and taking into account the window of symptom onset (t_{S1l}, t_{S1u}), the likelihood function was given by

$$L(t_{S1u}, t_{S1l,} t_{S2} | \theta) = \int_{t_{S1l,}}^{t_{S1u}} \int_{-\infty}^{t_{S2}} \beta_c(t_I - t_{S1}) g(t_{S2} - t_I) d_{t_I} d_{t_{S1}}$$

The results of the estimation are presented in the main text.

Generation time

Generation time - that is the time interval between infection of the primary case (t_{I1}) and infection of the secondary cases (t_{I2}) generated by such primary case - was inferred using the data of incubation period combined with infectiousness profile estimated in our analysis. We considered that infected cases would show symptoms at certain time (t_S) before or after onset of infectiousness. Assuming that the distribution of generation time follows a gamma distribution $\varphi(t_{I2} - t_{I1})$, the observed distribution of incubation period $g(t_S - t_{I1})$ can be inferred as the convolution between the infectiousness profile $\beta_c(t_{I2} - t_S)$ and the generation time distribution. We constructed a likelihood function based on the convolution, which was fitted to the observed incubation period, with t_{I1} provided in the form of a time interval bounded by the dates of the first and last possible exposure (t_{E1}, t_{E2}), given by

$$L(t_{E1}, t_{E2}, t_S|\alpha, \beta) = \int_{t_{E1}}^{t_{E2}} \int_{t_{I1}}^{+\infty} \varphi(t_{I2} - t_{I1})\beta_c(t_{I2} - t_S) d_{t_S} d_{t_{I1}}$$

Shape parameter (α) and rate parameter (β) of the gamma distribution of generation time were estimated using maximum likelihood and presented in the main text.

Interval from symptom onset to the sampling date of first PCR

We estimated the time interval from symptom onset to the sampling date of first PCR based on information of PCR positive individuals. Individuals with information on both the date of symptom onset and the date of first PCR sampling were included. We estimated the distribution of the interval by using maximum likelihood and compared three distributions (Weibull, gamma, and lognormal). The goodness of fit was assessed using Akaike information criterion (AIC).

As described above, the infectiousness profile peaked before the day of symptom onset. This may be driven by the control measures like isolation of infectors. We estimated the distribution of interval from symptom onset to the sampling date of first PCR to evaluate the timing of identification and isolation of infectious individuals. Results are presented in the main text.

Total and mean number of infections by age of infector and of infectee

From 254 certain transmission pairs, we estimate the total (Fig. S4A) and mean (Fig. S4B) number of infections by age. These matrices are descriptive and do not account for confounding factors other than age. Therefore, they cannot be used to estimate susceptibility and infectivity by age group. For example, the lower mean number of infections generated by children (0-14 years old) with respect to adults is the joint effect of several factors. According to our regression analysis, one of these factors is the generation of infection. Infected individuals in generation one have much higher odds of transmitting the infection, probably due to the case isolation and quarantine of close contacts that increase with the generation. Coupled with the low proportion of children in the first generation as compared to adults (we remind that the schools were closed during the entire study period and close community management policies were in place), this may have contributed to lower number of infections generated by children. The summary tables by age and generation are reported in Tab. S4 and Tab. S5.

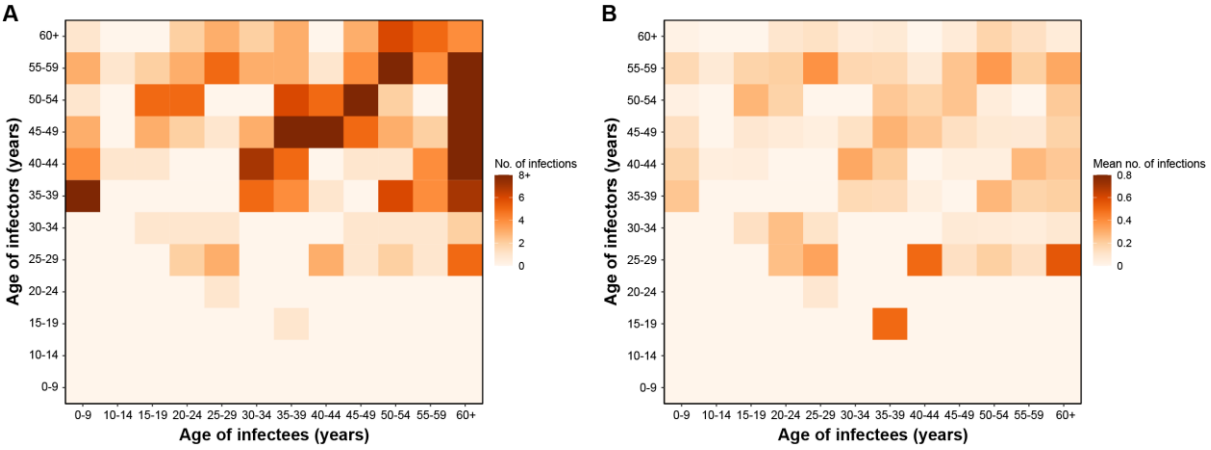


Figure S4. *Number of infections by age of infector and of infectee. Each cell in the matrices refers to the total number of infections (A) and the mean number of infections (B) caused by an infector of a given age.*

Table S4. *Summary of contact tracing data by age of infectors and generation of transmission.*

	No. of infectors	Number of infected contacts (independent of contact's age)	Number of contacts (independent of contact's age)	% of infected among contacts (N of infected contacts/ N of contacts)	Average number of infector's contacts
Age of infectors					
0-14 y	25	2	193	1.04	7.72
G1	1	0	1	0	-
G2	7	1	59	1.69	-
G3-4	6	0	52	0	-
Multiple exposure	3	1	25	4.00	-
Unknown	8	0	56	0	-
15-64 y	355	188	6,833	3.00	19.25
G1	59	84	1,951	4.00	-
G2	149	70	2,324	3.00	-
G3-4	50	17	839	2.03	-
Multiple exposure	30	8	551	1.45	-
Unknown	67	9	1,168	1.00	-
65+ y	81	19	1,133	2.00	13.99

G1	6	7	169	4.14	-
G2	46	10	604	1.66	-
G3-4	7	2	74	2.70	-
Multiple exposure	4	0	22	0	-
Unknown	18	0	264	0	-

Table S5. *Summary of contact tracing data by age of contacts and generation of transmission.*

	Number of contacts	Number of infected contacts	Total number of infectors (independent of infector's age)	% of infected among contacts (N of infected contacts/ N of contacts)	Average number of infector's contacts
Age of contacts					
0-14 y	936	22	309	2.35	3.03
G1	174	7	50	4.02	-
G2	349	8	131	2.29	-
G3-4	131	5	46	3.82	-
Multiple exposure	82	1	26	1.22	-
Unknown	200	1	56	0.50	-
15-64 y	6,411	154	446	2.00	14.37
G1	1,723	65	66	4.00	-
G2	2,363	60	198	3.00	-
G3-4	718	13	61	1.81	-
Multiple exposure	466	8	33	1.72	-
Unknown	1,141	8	88	1.00	-
65+ y	812	33	279	4.06	2.91
G1	224	19	50	8.48	-
G2	275	13	114	4.73	-
G3-4	116	1	41	0.86	-
Multiple exposure	50	0	19	0	-
Unknown	147	0	55	0	-

Quantifying the impact of potential drivers on the susceptibility and infectivity of SARS-CoV-2

We analyzed the odds ratio of SARS-CoV-2 transmission given the characteristics of the infectors and their contacts. To consider the clustering effect of an infector and a cluster, mixed effect logit models (i.e., generalized linear mixed-effect model, GLMM, for binary

data with the logit link) were used to explore potential drivers of the susceptibility and infectivity of SARS-CoV-2 virus. The specifications of the GLMM models are defined as follows:

$$g(u_i) = \alpha + \beta_1 Age_infector_i + \beta_2 Age_contact_i + \beta_3 Contact_type_i + \beta_4 Generation_infector_i + \beta_5 Exposure_level_i + \beta_6 Gender_infector_i + \beta_7 Gender_contact_i + \beta_8 Case_type_i + u_0 + u_1$$

Where:

- g is a logit link function;
- α is the intersect
- $Age_infector_i$ is the fixed effects of the age group of the infector in the successful (1) or unsuccessful (0) transmission event i;
- $Age_contact_i$ is the age group of the contact (potential infectee) in the successful/unsuccessful transmission event i;
- $Contact_type_i$ is the type of contact occurred in the successful/unsuccessful transmission event i;
- $Generation_infector_i$ is the generation of the successful/unsuccessful transmission event i;
- $Exposure_level_i$ is the number of close contacts of the infector involved in the successful/unsuccessful transmission event i;
- $Gender_infector_i$ is the gender of the infector in the successful/unsuccessful transmission event i;
- $Gender_contact_i$ is the gender of contact in the successful/unsuccessful transmission event i;

- $Case_type_i$ is the clinical severity of the infector (i.e., asymptomatic subjects, mild case, moderate case, severe and critical cases of SARS-CoV-2 virus) involved in the successful/unsuccessful transmission event i;
- u_0 and u_1 are random effects attributed to an infector and a cluster, respectively. $u_i = E[Y|(u_0, u_1)]$ is the mean of the response variable Y_i of a given value of the random effects.

The random effects of the GLMM are presented in Table S6. The results for fixed effects, including 3 age groups for infector's and infectee's age, are presented in the Table S7 and Figure S6. To evaluate the disaggregated effects of age, we also used transformed (log) continuous age variables (i.e., age of infectors and contacts) (Tab. S8). The goodness-of-fit evaluation was based on the estimates provided in the Table S9. Model diagnostic measures and residuals plots (Fig. S5) were evaluated by DHARMa residual diagnostics for hierarchical models [4].

To further explore how the probability of SARS-COV-2 infections changes with a change in each covariate, the average marginal effects of age of infector and contacts, type of contact between infector and contact were estimated across all contacts, holding the effect of other covariate constant (Fig. S7).

In addition, to explore possible non-linearity in the connection of age and of the number of contacts with SARS-CoV-2 transmission, we used generalized additive mixed models (GAMM). We used the same specifications as in the GLMM models presented in the main

text. The summary of the results of the GAMM models is shown in Fig. S8 and in Tab. S10. The obtained results suggest that the risk of SARS-CoV-2 transmission monotonically increase with the age of contacts and with the number of infector's contacts. This is consistent with the patterns that have been shown in GLMM models.

Table S6. *Random effect of infectors and clusters in generalized linear mixed model with categorized age of infectors and contacts.*

	Infector		Cluster	
	Variance	Standard deviation	Variance	Standard deviation
Step 1-1	6.14	2.48	0.27	0.52
Step 1-2	3.88	1.97	0.75	0.87
Step 1-3	3.06	1.75	0.99	1.00
Step 1-4	2.98	1.73	0.96	0.98
Step 1-5[a]	2.98	1.73	0.96	0.98
Step 1-6	2.93	1.71	0.97	0.98
Step 2-1	5.98	2.45	0.34	0.58
Step 2-2	3.73	1.93	0.84	0.92
Step 2-3	3.01	1.74	1.06	1.03
Step 2-4	2.94	1.71	1.01	1.00
Step 2-5[b]	2.94	1.71	1.01	1.00
Step 2-6	2.88	1.70	1.05	1.02

[a] model 1

[b] model 2

Table S7. *Stepwise regression analysis of factors associated with the probability of acquiring SARS-CoV-2 infections in generalized linear mixed models with categorized age of infectors and contacts.*

Characteristics	No. of contact	Step 1-1		Step 1-2		Step 1-3		Step 1-4		Step 1-5[a]		Step 1-6	
		OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value
Intercept	-	0.03 (0.01, 0.06)	<0.001	0.22 (0.11, 0.44)	<0.001	0.37 (0.17, 0.83)	0.015	0.27 (0.11, 0.65)	0.003	0.27 (0.11, 0.65)	0.004	0.26 (0.1, 0.71)	0.009
Age of infectors													
0-14 y	193	0.17 (0.02, 1.35)	0.093	0.28 (0.04, 2.18)	0.226	0.24 (0.03, 1.72)	0.155	0.25 (0.04, 1.75)	0.161	0.25 (0.04, 1.75)	0.162	0.28 (0.04, 2.04)	0.211
15-64 y	6833	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-
65+ y	1133	0.46 (0.17, 1.25)	0.128	0.65 (0.25, 1.71)	0.382	0.63 (0.25, 1.58)	0.326	0.64 (0.26, 1.59)	0.332	0.64 (0.26, 1.59)	0.332	0.56 (0.22, 1.43)	0.226
Age of contacts													
0-14 y	936	0.55 (0.32, 0.93)	0.026	0.58 (0.34, 0.98)	0.043	0.58 (0.34, 0.98)	0.041	0.58 (0.34, 0.98)	0.042	0.58 (0.34, 0.98)	0.042	0.58 (0.34, 0.98)	0.042
15-64 y	6411	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-
65+ y	812	1.66 (1.03, 2.67)	0.038	1.66 (1.03, 2.68)	0.036	1.64 (1.02, 2.62)	0.041	1.65 (1.03, 2.65)	0.038	1.65 (1.03, 2.65)	0.038	1.64 (1.02, 2.63)	0.04
Type of contact													
Household contacts	1021	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-
Relative contacts	3084	0.11 (0.07, 0.17)	<0.001	0.10 (0.07, 0.16)	<0.001	0.11 (0.07, 0.17)	<0.001	0.11 (0.07, 0.17)	<0.001	0.11 (0.07, 0.17)	<0.001	0.11 (0.07, 0.17)	<0.001
Social contacts	2227	0.06 (0.03, 0.11)	<0.001	0.05 (0.03, 0.10	<0.001	0.06 (0.03, 0.11)	<0.001	0.06 (0.03, 0.11)	<0.001	0.06 (0.03, 0.11)	<0.001	0.06 (0.03, 0.11)	<0.001
Other contacts	1827	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.12)	<0.001	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.13)	<0.001
Generation of COVID-19 transmission													
G1	2121	-	-	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-
G2	2987	-	-	0.15 (0.06, 0.36)	<0.001	0.13 (0.05, 0.29)	<0.001	0.13 (0.06, 0.31)	<0.001	0.13 (0.06, 0.31)	<0.001	0.15 (0.06, 0.35)	<0.001
G3-4	965	-	-	0.07 (0.02, 0.27)	<0.001	0.05 (0.02, 0.19)	<0.001	0.05 (0.02, 0.19)	<0.001	0.05 (0.02, 0.19)	<0.001	0.06 (0.02, 0.21)	<0.001
Multiple exposure[b]	598	-	-	0.11 (0.03, 0.47)	0.003	0.09 (0.02, 0.35)	<0.001	0.10 (0.03, 0.41)	0.001	0.10 (0.03, 0.41)	0.001	0.12 (0.03, 0.51)	0.004
Unknown	1488	-	-	0.03 (0.01, 0.11)	<0.001	0.03 (0.01, 0.10)	<0.001	0.03 (0.01, 0.1)	<0.001	0.03 (0.01, 0.10)	<0.001	0.03 (0.01, 0.12)	<0.001
Levels of exposure to an infector													
Total number of contacts	8159	-	-	-	-	0.99 (0.97, 1.00)	0.028	0.99 (0.97, 1.00)	0.028	0.99 (0.97, 1.00)	0.028	0.98 (0.97, 1)	0.018
Gender of infectors													
Female	4067	-	-	-	-	-	-	Reference	-	Reference	-	Reference	-

Male	4092	-	-	-	-	-	-	1.76 (0.97, 3.21)	0.064	1.76 (0.97, 3.21)	0.063	1.67 (0.91, 3.06)	0.096
Gender of contacts								-	-				
Female	4017	-	-	-	-	-	-	-	-	Reference	-	Reference	-
Male	4142	-	-	-	-	-	-	-	-	1.02 (0.74, 1.40)	0.907	1.02 (0.74, 1.40)	0.911
Clinical severity of infectors													
Asymptomatic subjects	898	-	-	-	-	-	-	-	-	-	-	0.72 (0.24, 2.16)	0.556
Mild cases	2848	-	-	-	-	-	-	-	-	-	-	Reference	-
Moderate case	2798	-	-	-	-	-	-	-	-	-	-	0.99 (0.48, 2.06)	0.979
Severe and critical case	1615	-	-	-	-	-	-	-	-	-	-	1.50 (0.58, 3.87)	0.403

[a] model 1

[b] Contacts who were exposed to multiple cases with different generations of COVID-19 transmissions.

Table S8. *Stepwise regression analysis of factors associated with the probability of acquiring SARS-CoV-2 infections in generalized linear mixed models with log-transformed age of infectors and contacts.*

Characteristics	No. of contact	Step 2-1		Step 2-2		Step 2-3		Step 2-4		Step 2-5[a]		Step 2-6	
		OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value
Intercept	-	0 (0, 0.01)	<0.001	0.01 (0, 0.19)	0.001	0.02 (0, 0.29)	0.003	0.02 (0, 0.21)	0.002	0.02 (0, 0.21)	0.002	0.02 (0, 0.28)	0.003
Age of infectors													
Log-transformed age	8159	1.92 (0.99, 3.73)	0.054	1.62 (0.87, 3.01)	0.128	1.61 (0.90, 2.90)	0.108	1.62 (0.91, 2.89)	0.103	1.62 (0.91, 2.90)	0.104	1.53 (0.85, 2.75)	0.160
Age of contacts													
Log-transformed age	8159	1.28 (1.04, 1.58)	0.019	1.26 (1.03, 1.56)	0.028	1.26 (1.02, 1.55)	0.029	1.26 (1.02, 1.55)	0.028	1.26 (1.02, 1.55)	0.028	1.26 (1.02, 1.55)	0.029
Type of contact													
Household contacts	1021	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-
Relative contacts	3084	0.11 (0.07, 0.18)	<0.001	0.11 (0.07, 0.17)	<0.001	0.11 (0.07, 0.18)	<0.001	0.11 (0.07, 0.18)	<0.001	0.11 (0.07, 0.18)	<0.001	0.11 (0.07, 0.18)	<0.001
Social contacts	2227	0.06 (0.03, 0.11)	<0.001	0.05 (0.03, 0.09)	<0.001	0.06 (0.03, 0.11)	<0.001	0.06 (0.03, 0.11)	<0.001	0.06 (0.03, 0.11)	<0.001	0.06 (0.03, 0.11)	<0.001
Other contacts	1827	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.12)	<0.001	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.13)	<0.001	0.07 (0.04, 0.13)	<0.001

Generation of COVID-19 transmission														
G1	2121	-	-	Reference	-	Reference	-	Reference	-	Reference	-	Reference	-
G2	2987	-	-	0.14 (0.06, 0.33)	<0.001	0.12 (0.05, 0.27)	<0.001	0.12 (0.05, 0.28)	<0.001	0.12 (0.05, 0.28)	<0.001	0.13 (0.06, 0.30)	<0.001
G3-4	965	-	-	0.07 (0.02, 0.25)	<0.001	0.05 (0.01, 0.18)	<0.001	0.05 (0.02, 0.18)	<0.001	0.05 (0.02, 0.18)	<0.001	0.05 (0.02, 0.19)	<0.001
Multiple exposure[b]	598	-	-	0.11 (0.03, 0.44)	0.002	0.08 (0.02, 0.34)	<0.001	0.10 (0.03, 0.40)	0.001	0.10 (0.03, 0.40)	0.001	0.11 (0.03, 0.45)	0.002
Unknown	1488	-	-	0.03 (0.01, 0.11)	<0.001	0.03 (0.01, 0.10)	<0.001	0.03 (0.01, 0.10)	<0.001	0.03 (0.01, 0.10)	<0.001	0.03 (0.01, 0.11)	<0.001
Levels of exposure to an infector													
Total number of contacts	8159	-	-	-	-	0.99 (0.97, 1.00)	0.037	0.99 (0.97, 1.00)	0.036	0.99 (0.97, 1.00)	0.036	0.99 (0.97, 1.00)	0.028
Gender of infectors													
Female	4067	-	-	-	-	-	-	Reference	-	Reference	-	Reference	-
Male	4092	-	-	-	-	-	-	1.77 (0.97, 3.21)	0.061	1.77 (0.97, 3.21)	0.061	1.71 (0.94, 3.13)	0.079
Gender of contacts								-	-				
Female	4017	-	-	-	-	-	-	-	-	Reference	-	Reference	-
Male	4142	-	-	-	-	-	-	-	-	1.01 (0.73, 1.38)	0.973	1.01 (0.73, 1.39)	0.968
Clinical severity of infectors													
Asymptomatic subjects	898	-	-	-	-	-	-	-	-	-	-	0.74 (0.25, 2.23)	0.592
Mild cases	2848	-	-	-	-	-	-	-	-	-	-	Reference	-
Moderate case	2798	-	-	-	-	-	-	-	-	-	-	1.00 (0.48, 2.08)	0.999
Severe and critical case	1615	-	-	-	-	-	-	-	-	-	-	1.23 (0.49, 3.12)	0.661

[a] model 2

[b] Contacts who were exposed to multiple cases with different generations of COVID-19 transmissions.

Table S9. *Assessing the fit of generalized linear mixed model with categorized age of infectors and contacts.*

Model	AIC (ΔAIC)	BIC	loglink	deviance	χ^2	P-value[a]
Step 1-1	1577.5	1647.6	-778.8	1557.5	-	-
Step 1-2	1541.9 (35.6)	1640.0	-756.9	1513.9	43.612	<0.001
Step 1-3	1539.5 (2.4)	1644.6	-754.7	1509.5	4.433	0.035

Step 1-4	1538.0 (1.5)	1650.1	-753.0	1506.0	3.456	0.063
Step 1-5[b]	1540.0 (-2.0)	1659.1	-753.0	1506.0	0.014	0.907
Step 1-6	1544.6 (-4.6)	1684.7	-752.3	1504.6	1.412	0.703
Step 2-1	1579.1	1635.2	-781.6	1563.1	-	-
Step 2-2	1541.9 (37.2)	1626.0	-759.0	1517.9	45.188	<0.001
Step 2-3	1539.9 (2.0)	1631.0	-757.0	1513.9	4.024	0.045
Step 2-4	1538.4 (1.5)	1636.5	-755.2	1510.4	3.529	0.060
Step 2-5[c]	1540.4 (-2.0)	1645.5	-755.2	1510.4	0.001	0.974
Step 2-6	1545.7 (-5.3)	1671.9	-754.9	1509.7	0.642	0.887

[a]ANOVA analyses were performed.

[b] model 1

[c] model 2

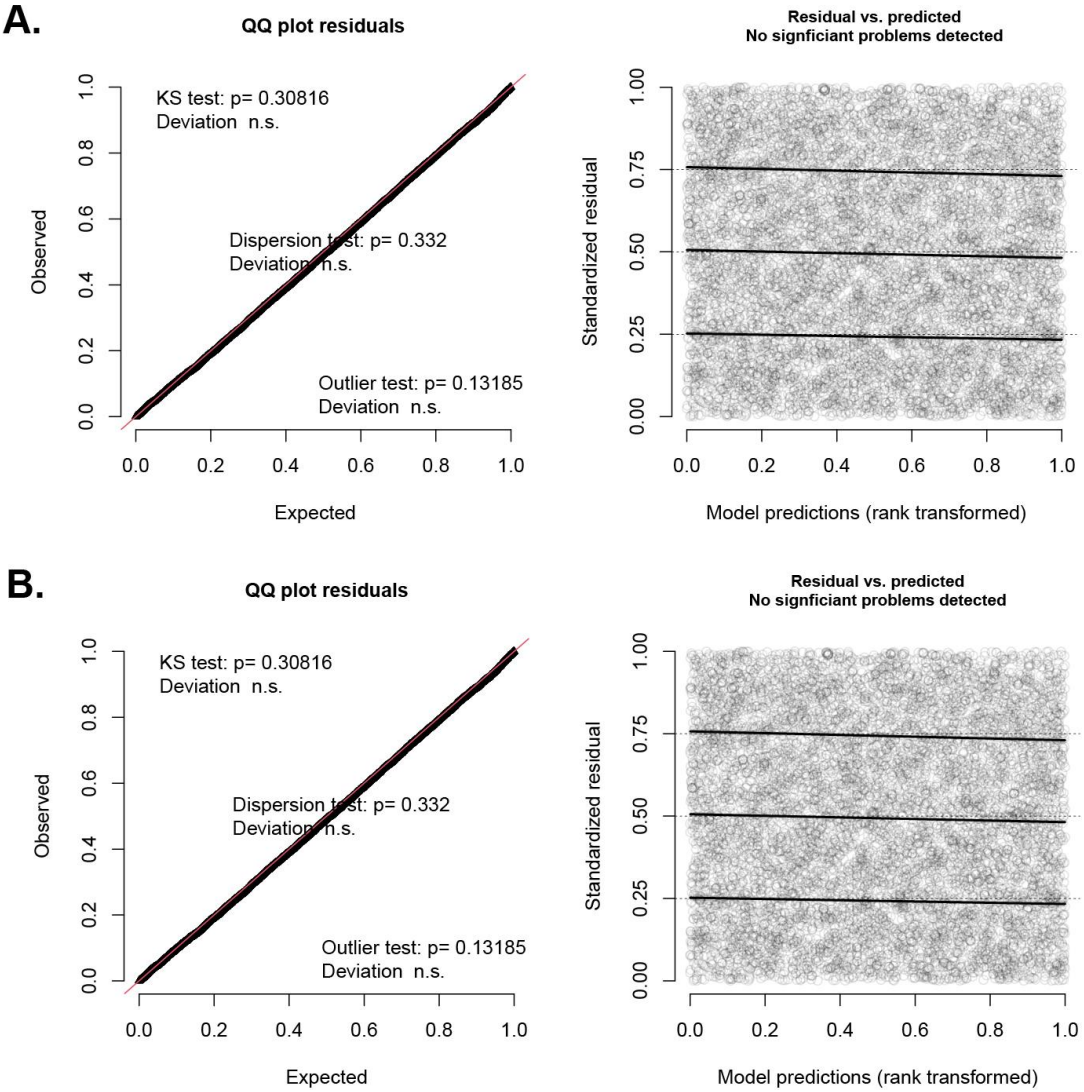


Figure S5*. DHARMa residual diagnostics for GLMM models. A: DHARMa residual diagnostics for model 1. B: DHARMa residual diagnostics for model 2.*

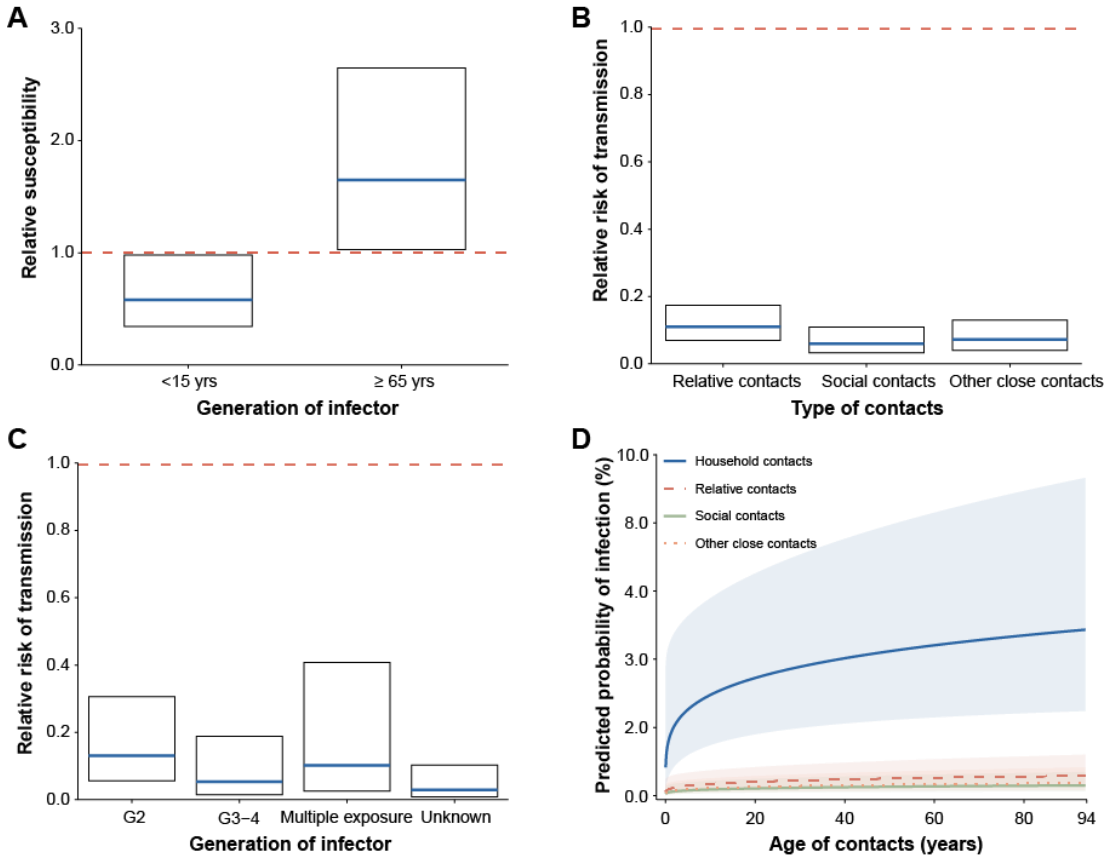


Figure S6. *Potential drivers of COVID-19 transmissions in Hunan Province, China. (A) The relative susceptibility of contacts who were younger than 15 years of age and of those who were older than 65 years of age (the reference group [red lines] was contacts who were 15-64 years of age); (B) The relative risk of SARS-CoV-2 infections among contacts with different type of exposures to an infector (the reference group [red lines] was contacts who were exposed to an infector in households); (C) The relative risk of SARS-CoV-2 infections among contacts with exposures to infectors with different generations of COVID-19 transmission (the reference group [red lines] was contacts who were exposed to an index case-patients); (D) The probability of SARS-CoV-2 infections at a given age of contacts in a specific setting.*

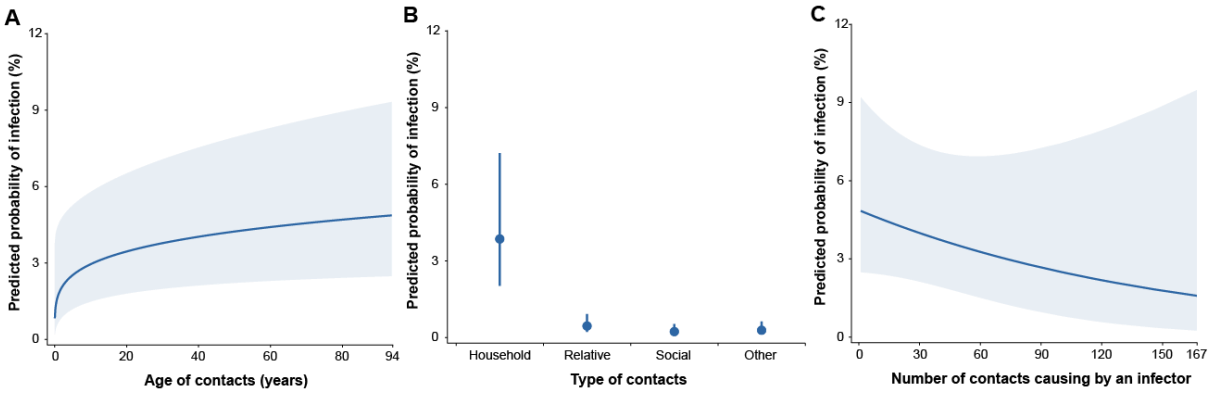


Figure S7. *The marginal effects of various covariates in GLMM model. The smoothed lines and shared areas represent the point estimates and 95% confidence interval for the probability of SARS-CoV-2 infections, respectively.*

Table S10. *Stepwise regression analysis of factors associated with the probability of acquiring SARS-CoV-2 infections in generalized additive mixed models with splines for age of infectors and contacts, as well as number of contacts.*

Characteristics	No. of contact	Step 4-1		Step 4-2	
		OR (95%CI)	P-value	OR (95%CI)	P-value
Parametric coefficients					
Intercept	-	0.25 (0.12, 0.51)	<0.001	0.23 (0.11, 0.47)	<0.001
Age of infectors					
0-14 y	193	0.33 (0.06, 1.66)	0.177	-	-
15-64 y	6,833	Reference	-	-	-
65+ y	1,133	0.60 (0.29, 1.23)	0.166	-	-
Age of contacts					
0-14 y	936	0.62 (0.38, 1.01)	0.055	-	-
15-64 y	6,411	Reference	-	-	-
65+ y	812	1.54 (0.99, 2.39)	0.055	-	-
Type of contact					
Household contacts	1,021	Reference	-	Reference	-
Relative contacts	3,084	0.14 (0.09, 0.21)	<0.001	0.14 (0.09, 0.21)	<0.001
Social contacts	2,227	0.09 (0.05, 0.15)	<0.001	0.08 (0.05, 0.14)	<0.001
Other contacts	1,827	0.10 (0.06, 0.17)	<0.001	0.10 (0.06, 0.17)	<0.001
Generation of COVID-19 transmission					
G1	2,121	Reference	-	Reference	-
G2	2,987	0.27 (0.15, 0.48)	<0.001	0.27 (0.15, 0.48)	<0.001
G3-4	965	0.17 (0.08, 0.38)	<0.001	0.18 (0.08, 0.40)	<0.001
Multiple exposure[b]	598	0.26 (0.10, 0.70)	0.007	0.28 (0.10, 0.75)	0.011
Unknown	1,488	0.07 (0.03, 0.18)	<0.001	0.08 (0.03, 0.20)	<0.001
Gender of infectors					
Female	4,067	Reference	-	Reference	-
Male	4,092	1.62 (1.02, 2.57)	0.041	1.65 (1.04, 2.62)	0.035
Gender of contacts					
Female	4,017	Reference	-	Reference	-

Male	4,142	1.01 (0.74, 1.36)	0.963	1.00 (0.74, 1.35)	0.988
Approximate significance of smooth terms					
Random effect					
Infector	-	-	<0.001	-	<0.001
Cluster	-	-	0.141	-	0.124
Age of infectors					
Spline for age	-	-	-	-	0.198
Age of contacts					
Spline for age	-	-	-	-	<0.001
Levels of exposure to an infector					
Spline for number of contacts	-	-	0.032	-	0.040

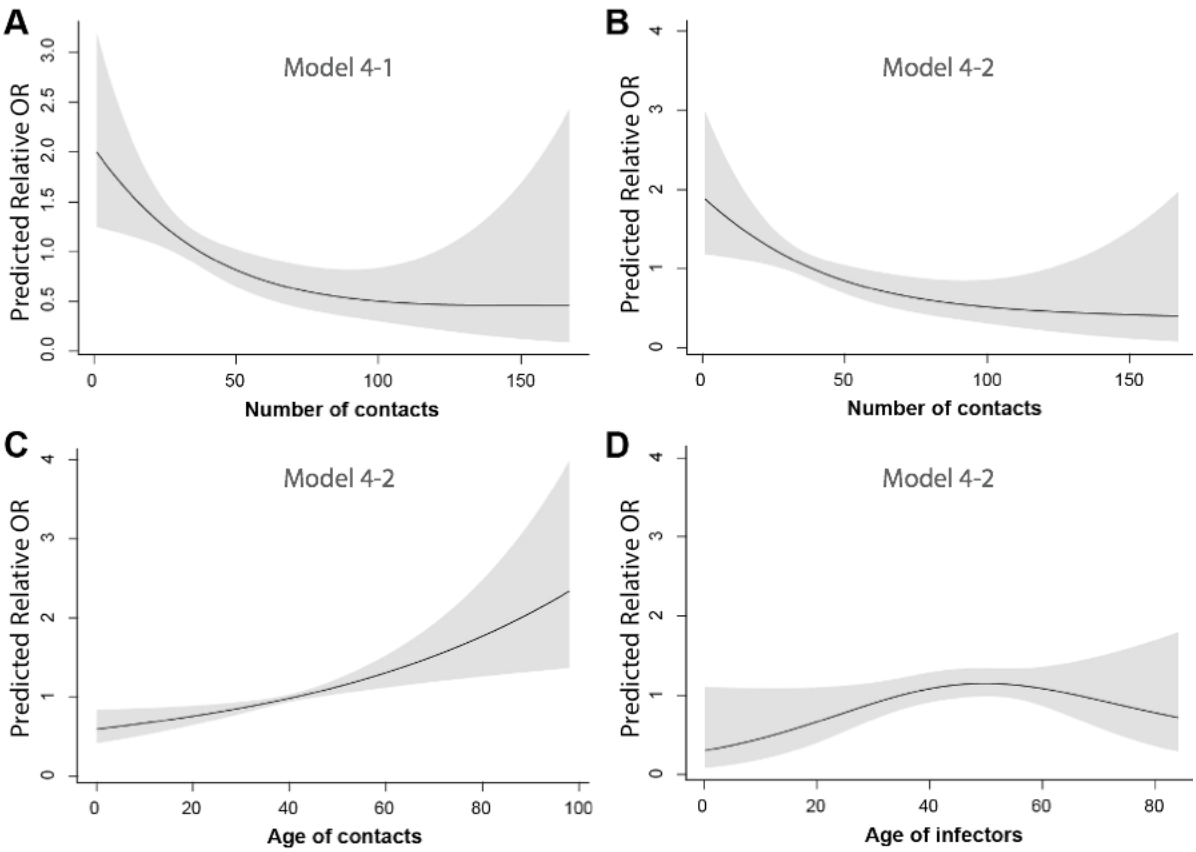
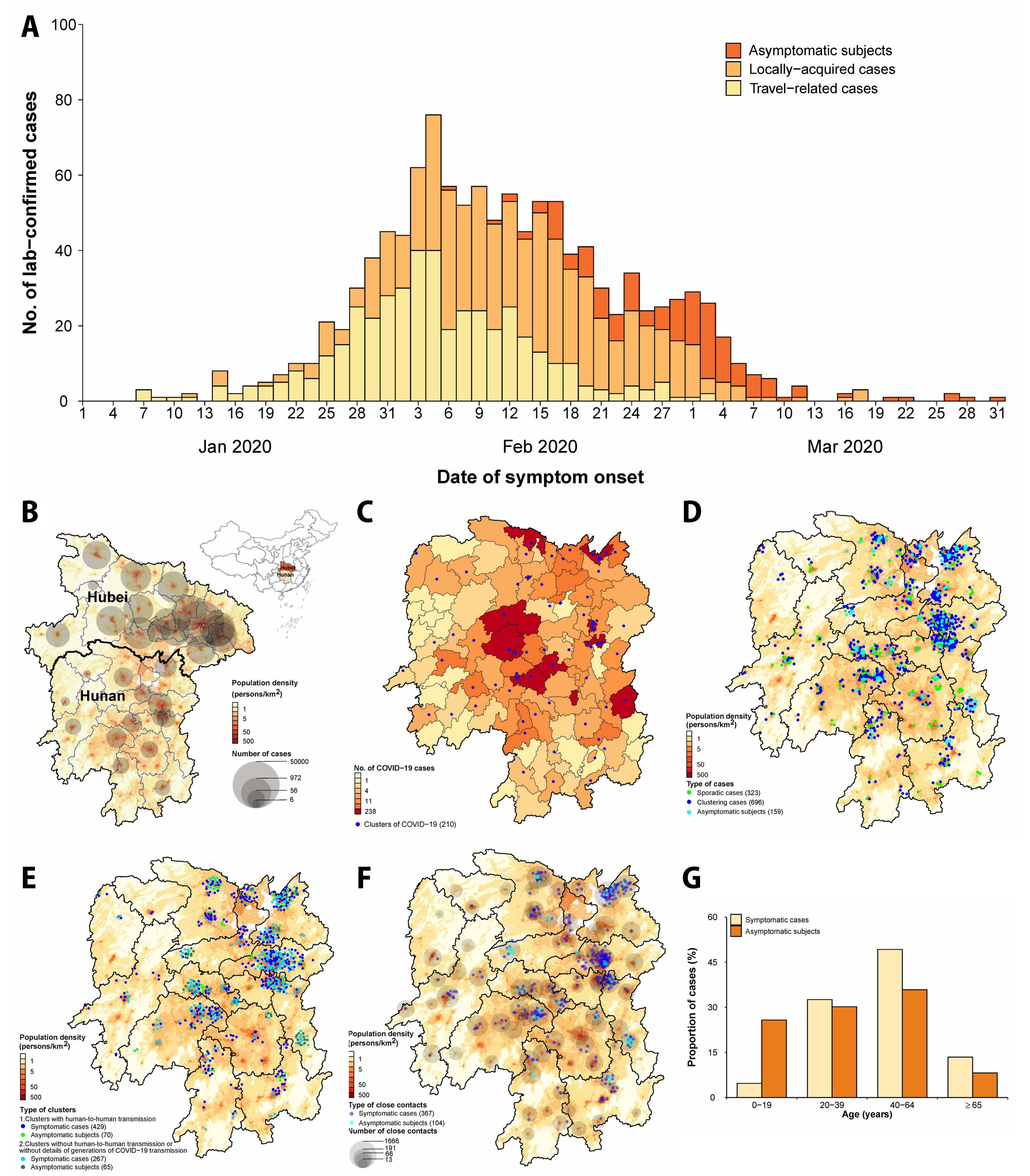
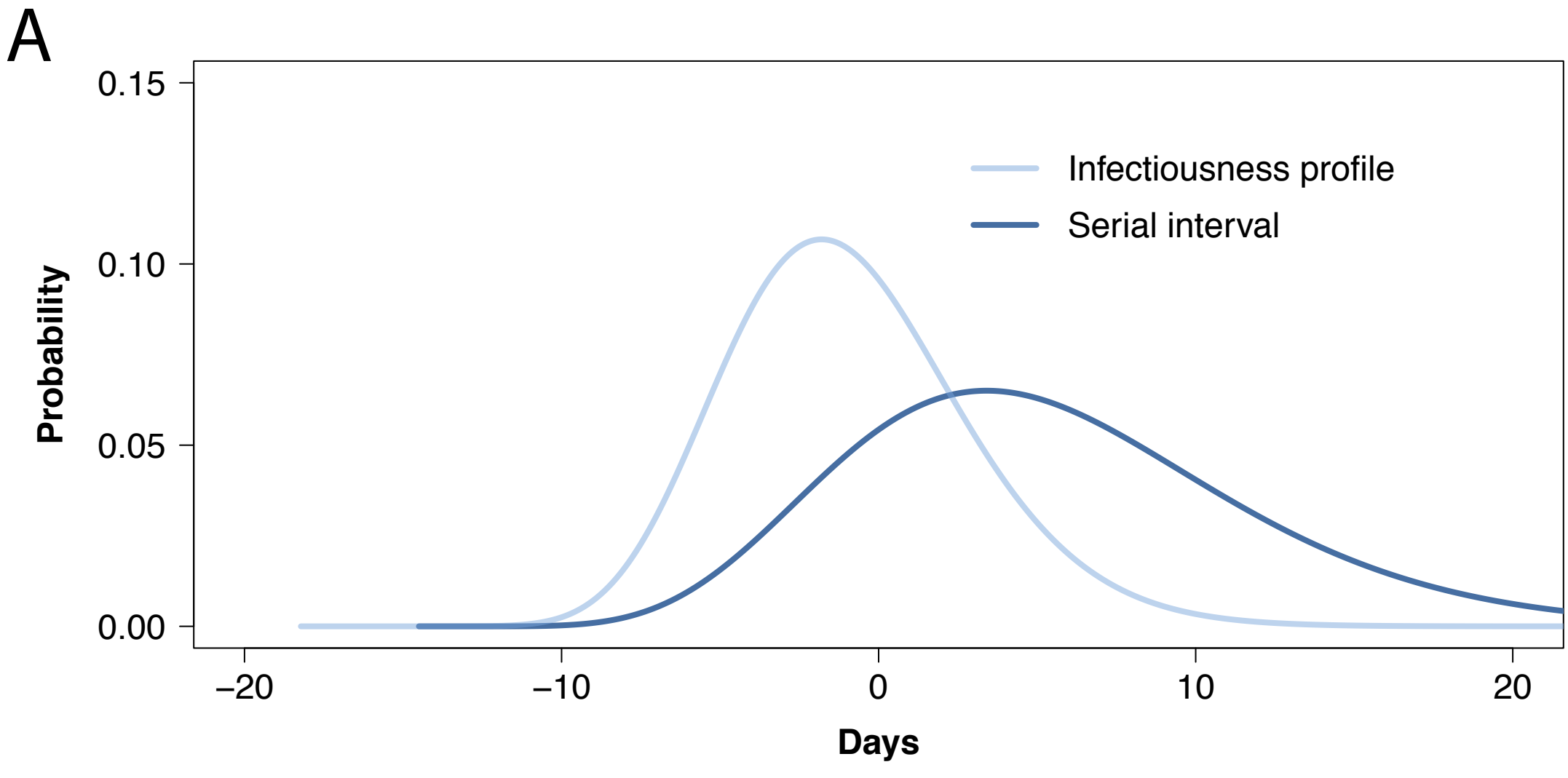


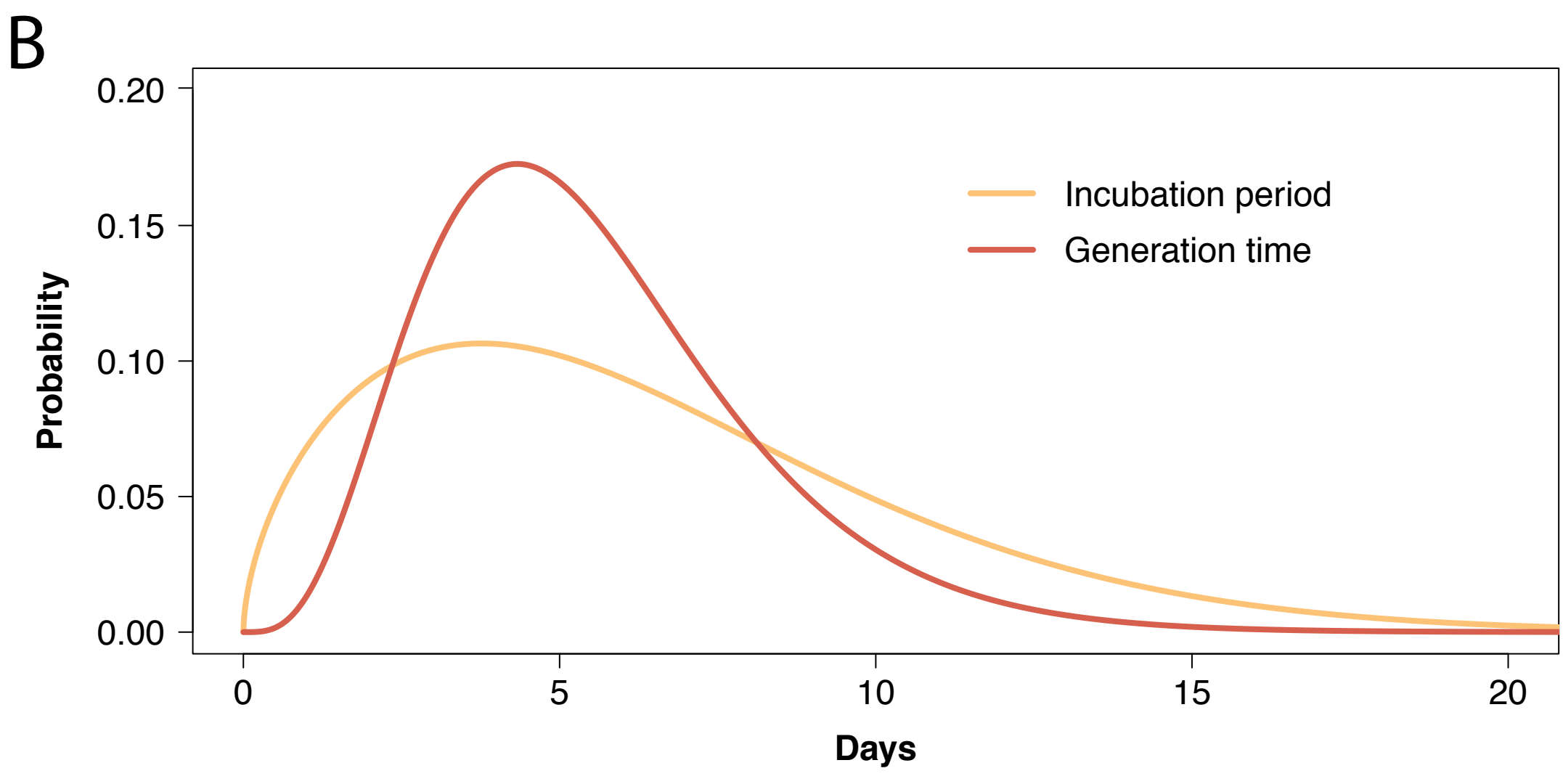
Figure S8. *GAMM-predicted non-linear 3-knot splines for age of infector, age of contacts, and the number of contacts. A. model 4-1; B to D. model 4-2. The shaded area delimits the 95% confidence intervals of the spline functions.*

References

1. National Health Commission of the People's Republic of China. Diagnosis and treatment guideline on pneumonia infection with 2019 novel coronavirus (6th trial edn). 2020.
2. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* 2020; **26**(5): 672-5.
3. Ashcroft P, Huisman JS, Lehtinen S, et al. COVID-19 infectivity pro1le correction. *ArXiv e-prints* 2020.
4. Florian Hartig. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. 2016.

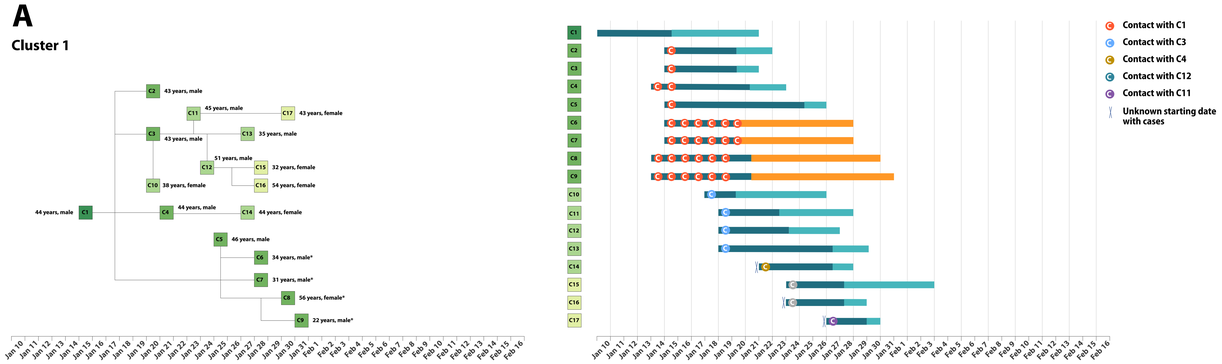

A
No. of lab−confirmed cases
Asymptomatic subjects
Locally−acquired cases
Travel−related cases
Jan 2020
Feb 2020
Mar 2020
Date of symptom onset
B
Hubei
Hunan
Population density (persons/km2)
Number of cases
C
No. of COVID−19 cases
Clusters of COVID−19 (210)
D
Population density (persons/km2)
Type of cases
Sporadic cases (323)
Clustering cases (696)
Asymptomatic subjects (159)
E
Population density (persons/km2)
Type of clusters
1.Clusters with human-to-human transmission
Symptomatic cases (429)
Asymptomatic subjects (70)
2.Clusters without human-to-human transmission or without details of generations of COVID−19 transmission
Symptomatic cases (267)
Asymptomatic subjects (65)
F
Population density (persons/km2)
Type of close contacts
Symptomatic cases (367)
Asymptomatic subjects (104)
Number of close contacts
G
Symptomatic cases
Asymptomatic subjects
Proportion of cases (%)
Age (years)
0−19
20−39
40−64
≥65


A
0.15
0.10
0.05
0.00
Probability
Infectiousness profile
Serial interval
−20
−10
0
10
20
Days

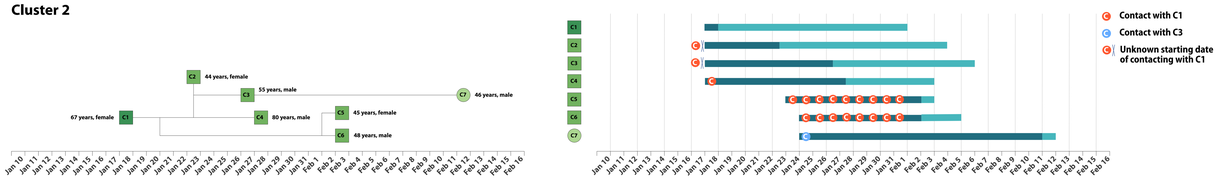

B
0.20
0.15
0.10
0.05
0.00
Probability
Incubation period
Generation time
0
5
10
15
20
Days

A

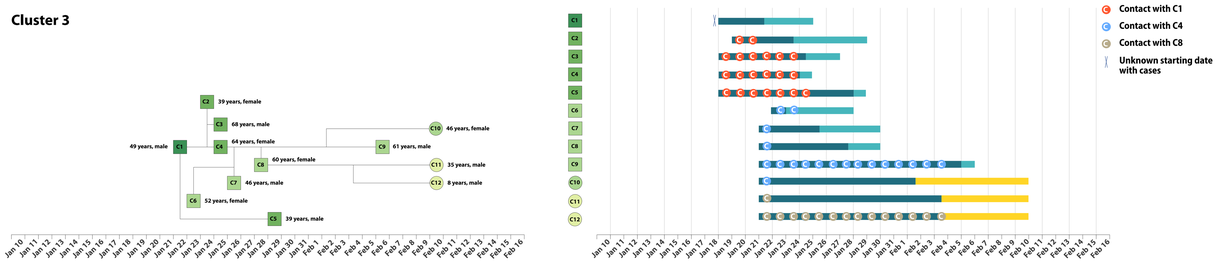
Cluster 1



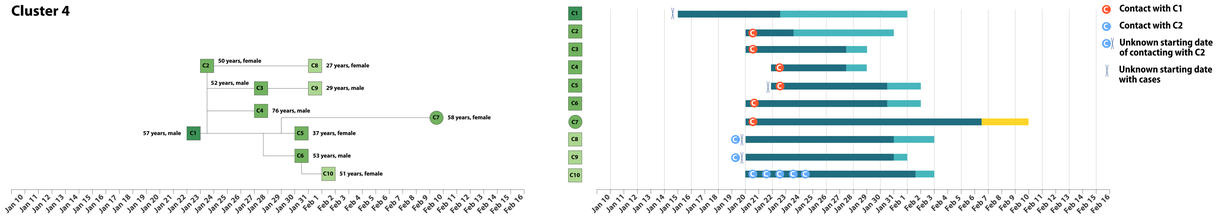
Cluster 2



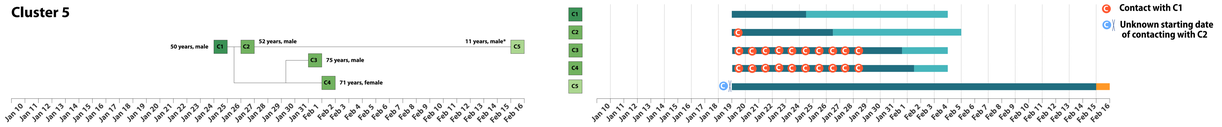
Cluster 3



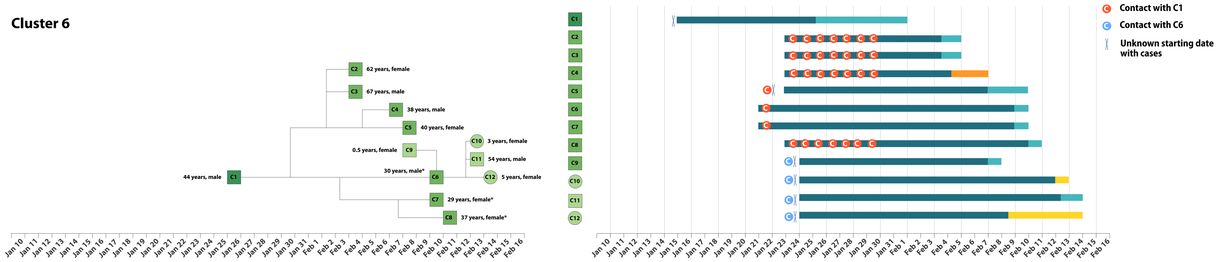
Cluster 4



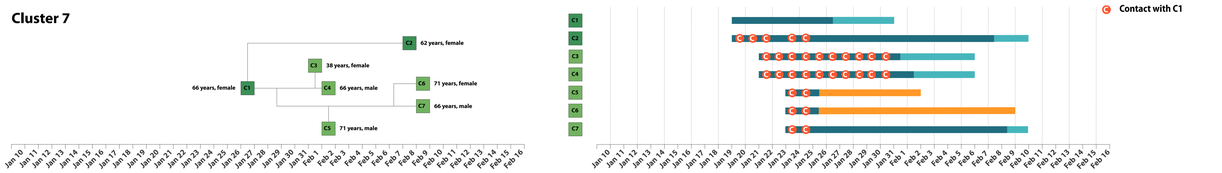
Cluster 5



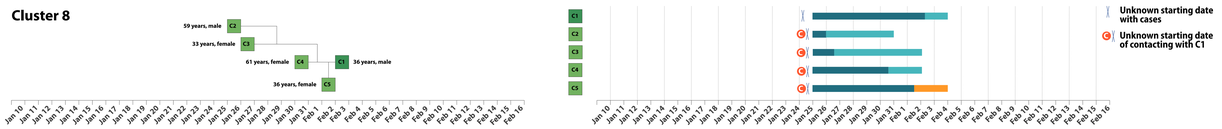
Cluster 6



Cluster 7

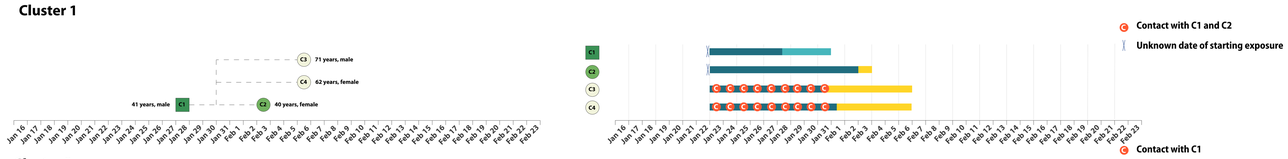


Cluster 8

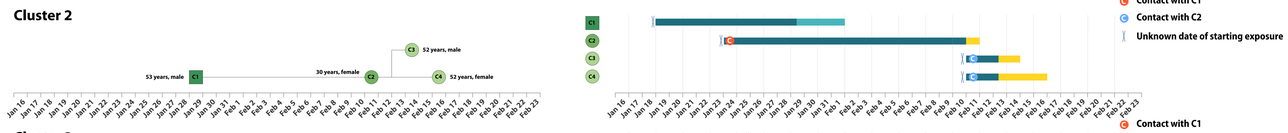


B

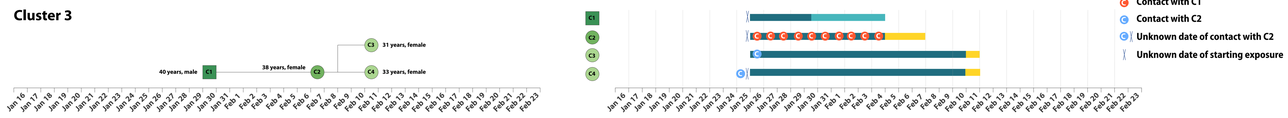
Cluster 1



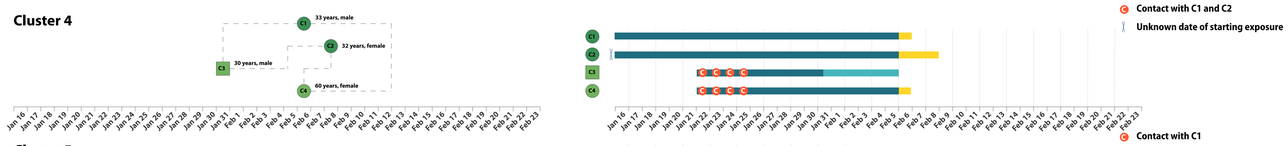
Cluster 2



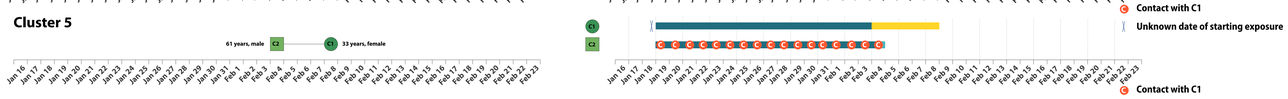
Cluster 3



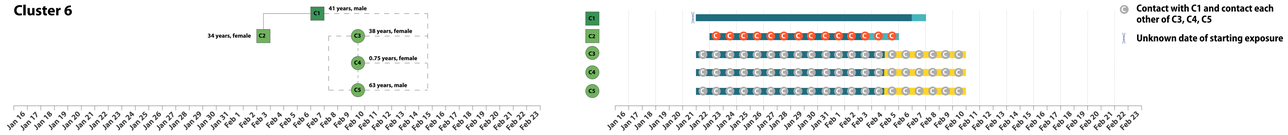
Cluster 4



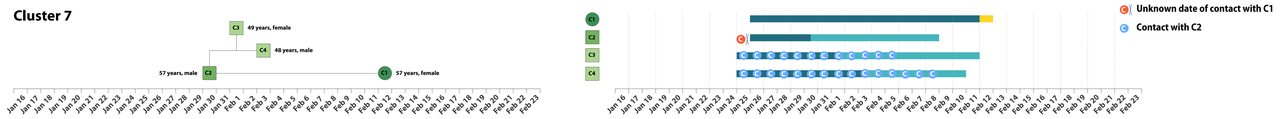
Cluster 5



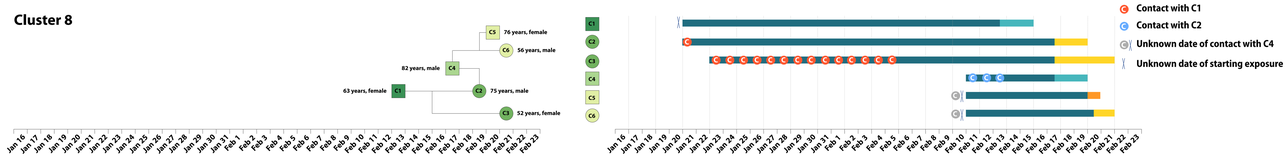
Cluster 6



Cluster 7



Cluster 8



- Period from contacts to symptom onset or isolation
- Period from symptom onset to isolation
- Period from isolation to symptom onset
- Period from isolation to date of laboratory confirmation

- Index cases
- Secondary cases
- Third-generation cases
- Fourth-generation cases
- Unknown

- Symptomatic cases
- Asymptomatic individuals
- Duration of date of onset between generations
- Case with multiple exposures