Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2

Kaiyuan Sun[1#*], Wei Wang[2#], Lidong Gao[3#], Yan Wang[2], Kaiwei Luo[3], Lingshuang Ren[2], Zhifei Zhan[3], Xinghui Chen[2], Shanlu Zhao[3], Yiwei Huang[3], Qianlai Sun[3], Ziyan Liu[3], Maria Litvinova[4], Alessandro Vespignani[5, 4], Marco Ajelli[6], Cécile Viboud[1†], Hongjie Yu[2†*]

[1]Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA.

[2]School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China.

[3]Hunan Provincial Center for Disease Control and Prevention, Changsha, China.

[4]ISI Foundation, Turin, Italy.

[5]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA USA.

[6]Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA.

*Correspondence to: kaiyuan.sun@nih.gov (K.S.); yhj@fudan.edu.cn (H.Y.)
# These authors contributed equally to this work.
† These are co-senior authors of this work.

Abstract

A long-standing question in infectious disease dynamics is the role of transmission heterogeneities, particularly those driven by demography, behavior and interventions. Here we characterize transmission risk between 1,178 SARS-CoV-2 infected individuals and their 15,648 close contacts based on detailed contact tracing data from Hunan, China. We find that 80% of secondary transmissions can be traced back to 14% of SARS-CoV-2 infections, indicating substantial transmission heterogeneities. Regression analysis suggests a marked gradient of transmission risk scales positively with the duration of exposure and the closeness of social interactions, after adjusted for demographic and clinical factors. Population-level physical distancing measures confine transmission to families and households; while case isolation and contact quarantine reduce transmission in all settings. Adjusted for interventions, the reconstructed infectiousness profile of a typical SARS-CoV-2 infection peaks just before symptom presentation, with ~50% of transmission occurring in the pre-symptomatic phase. Modelling results indicate that achieving SARS-CoV-2 control would require the synergistic efforts of case isolation, contact quarantine, and population-level physical distancing measures, owing to the particular transmission kinetics of this virus.

Introduction

While the age dependency in clinical severity of COVID-19 has been well documented (*1–5*), there is limited information on how transmission risk vary with age, clinical presentation, and contact types (*6–12*). Individual-based interventions such as case isolation, contact tracing and quarantine have been shown to accelerate case detection and interrupt transmission chains (*13*). However, these interventions are typically implemented in conjunction with population-level physical distancing measures, and their effects on contact patterns and transmission risk remain difficult to separate (*14–24*). A better understanding of the factors driving SARS-CoV-2 transmission is key to achieve epidemic control while minimizing societal cost, particularly as countries relax physical distancing measures.

Hunan, a province in China adjacent to Hubei where the COVID-19 pandemic began, experienced sustained SARS-CoV-2 transmission in late January and early February 2020, but the outbreak was swiftly suppressed thereafter. As in many other provinces in China, epidemic control was achieved by a combination of individual-based interventions targeting cases and their contacts and population-level physical distancing measures. In this study, we reconstruct transmission chains for all identified SARS-CoV-2 infections in Hunan, as of April 3, 2020, based on granular epidemiological information collected through extensive surveillance and contact tracing efforts. We identify the demographic, clinical and behavioral factors that drive transmission heterogeneities and evaluate how interventions modulate the topology of the transmission network. Further, we reconstruct the infectiousness profile of SARS-CoV-2 over the course of a typical infection and estimate the feasibility of epidemic control by individual and population-based interventions.

We analyze detailed epidemiological records for 1,178 SARS-CoV-2 infected individuals and their 15,648 close contacts, representing 19,227 separate exposure events, compiled by the Hunan Provincial Center for Disease Control and Prevention. Cases were identified between January 16 and April 2, 2020; index cases were captured by passive surveillance and laboratory confirmed by RT-PCR. Individuals who were close contacts of the index cases were followed for at least 2 weeks after the last exposure to the infected individual. Prior to February 7, 2020, contacts were tested if they developed symptoms during the quarantine period. After February 7, 2020, RT-PCR testing was required for all contacts, and specimens were collected at least once from each contact during quarantine, regardless of symptoms. Upon positive RT-PCR test results, infected individuals were isolated in dedicated hospitals, regardless of their clinical severity, while their contacts were quarantined in medical observation facilities.

The dataset includes 210 epidemiological clusters representing 831 cases, with additional 347 sporadic cases (29%) unlinked to any cluster (detailed in Materials & Methods). For each cluster, we stochastically reconstruct transmission chains and estimate the timing of infection most compatible with each patient's exposure history. We analyze an ensemble of 100 reconstructed transmission chains to account for uncertainties in exposure histories (Fig. 1 visualizes one realization of the transmission chains). We observe between 0 and 4 generations of transmission, with the largest cluster involving 20 SARS-CoV-2-infected individuals. The number of secondary infections ranges from 0 to 10, with a distribution of secondary infections best characterized by a negative binomial distribution with mean $\mu = 0.40$ (95% CI, 0.39 to 0.41) and variance $\mu(1 + \mu/k) = 0.95$ (95% CI, 0.92 to 0.97), where $k = 0.30$ (95% CI, 0.29 to 0.30) is the dispersion parameter (Fig. 1). This suggests 80% of secondary infection can be traced back to 14%

of SARS-CoV-2 infected individuals, indicating substantial transmission heterogeneities at the individual-level.

Results

Characterizing SARS-CoV-2 transmission heterogeneities at the individual level

To dissect the individual transmission heterogeneities and identify predictors of transmission, we analyze the infection risk among a subset of 14,622 individuals who were close contacts of 870 SARS-CoV-2 patients. This dataset excludes primary cases whose infected contacts report a travel history to Wuhan. The dataset represents 74% of all SARS-CoV-2 cases in the Hunan epidemic, for whom contacts have been carefully monitored, capturing 17,750 independent exposure events.

We start by characterizing variation in transmission risk across the diverse set of 17,750 exposures. We focus on quantifying how the duration, timing and type of contact impact transmission risk, accounting for other factors including age, sex, clinical presentation, travel history of the primary case, as well as age and sex of the contacts. Exposures are grouped into 5 categories based on the type of contact settings, namely: household, extended family, social, community, and healthcare (Table S2), with the duration of exposure approximated by the time interval between the initial and final dates of exposure. We also stratify exposures by the date of occurrence, with January 25, 2020 marking the beginning of enhanced physical distancing measures in Hunan (based on Baidu Qianxi mobility index (*25*), Fig. S1A insert). To address putative variation in infectiousness over the course of infection, we distinguish exposures based on whether the exposure window contains the time of symptom onset of the primary case, a period associated with high viral shedding. We use a mixed effects multiple logistic regression model (GLMM-logit) to quantify the effects of these factors on transmission (see Fig. S1A for regression results, and Table S3 for a detailed definition of all risk factors and summary statistics).

We find a marked gradient of transmission risk scales positively with closeness of social interactions (Fig. S1A): household contacts pose the highest risk of transmission (see also (*12*)), followed by contacts in the extended family, social and community settings. Contacts in the healthcare setting have the lowest risk, suggesting that adequate protective measures were adopted by patients and healthcare staff in Hunan, China. Interestingly, the impact of physical distancing differs by transmission setting (Table 1): enhanced physical distancing measures elevates the risk of transmission in the household, likely due to increased contact frequency at home as a result of physical confinement during the "lockdown". In contrast, reduced within-city mobility is associated with a reduction in transmission risk per contact opportunity in the community and social settings, possibly caused by adoption of prudent behaviors such as mask wearing, hand washing and coughing/sneezing etiquette. We also find that longer exposure window is a significant risk factor, with one additional day of exposure increasing the transmission risk by 10% (95% CI, 5% to 15%). Transmission risk is higher around the time of symptom presentation of the primary case (Table 1). And while susceptibility to SARS-CoV-2 gradually increases with age, we find no statistical support for age differences in infectivity (Fig. S1A), in agreement with previous findings (*12*).

For each of the 17,750 contact exposure events, we estimate the probability of transmission using the point estimate of the baseline odds and odds ratios from the GLMM-logit regression (Fig. S1A). In Fig. 2A, we plot the distribution of transmission probabilities in the household, extended family, social, and community settings separately. The average "per-contact" transmission probability is highest in the household (7.1%, 95% CI, 1.2% to 19.3%), followed by

family (1.7%, 95% CI, 0.4% to 5.7%) and social settings (0.9% with 95% CI, 0.3% to 2.7%), and lowest in the community (0.4% with 95% CI, 0.1% to 1.1%). The gradient of transmission probabilities across settings is the joint effects of increasing duration of exposure with closeness of social interactions (Fig. 2B), superimposed on setting-specific risk differences (Fig. S1A). It is worth noting that these "per-contact" transmission probabilities were evaluated in a situation of intense interventions measures and high population awareness of the disease, and thus, they may be not generalizable elsewhere.

The number of contacts is also a key driver of individual transmission potential and varies by transmission setting. Fig. 2C presents the contact degree distribution, defined as the number of unique contacts per individual. We find that the distributions of individual contact degree are over-dispersed with dispersion parameter $0 < k < 1$ across all settings. Furthermore, household ($k = 0.72$) and extended family ($k = 0.64$) contacts are less dispersed than social ($k = 0.19$) and community ($k = 0.14$) contacts, suggesting that contact heterogeneities are inversely correlated with the closeness of social interactions. Fig. 2D visualizes the age-specific contact patterns between the primary cases and their contacts, demonstrating diverse mixing patterns across settings. Specifically, household contacts present the canonical "three-bands" pattern with the diagonal representing age-assortative interactions and the two off-diagonals representing inter-generational mixing (*26*, *27*). Other settings display more diffusive mixing patterns by age.

Next, we summarize the overall transmission potential of an individual by calculating the cumulative contact rate (CCR) of the primary case. The CCR captures how contact opportunities vary with demography, temporal variation in the infectiousness profile, an individual's contact degree, and interventions. (See Section 4.2 in Materials and Methods for detailed definition). Through regression analysis, we focus on how the overall transmission opportunity of an infected individual is affected by different intervention measures across transmission settings. After adjusting for age, sex, clinical presentation, and travel history to Wuhan, we find that physical distancing measures increase CCRs in the household and extended family and decrease CCRs in social and community settings (Fig. 2E). In contrast, faster case isolation (measured as the time between isolation, or pre-symptomatic quarantine, and symptom onset) universally reduces CCRs, decreasing transmission opportunities across all settings (Fig. 2E).

Characterizing the natural history of SARS-CoV-2 infection

We have characterized the SARS-CoV-2 transmission risk factors and have shown that individual and population-based interventions have a differential impact on contact patterns and transmission potential. Next, we use our probabilistic reconstruction of infector-infectee pairs to further dissect transmission kinetics and project the impact of interventions on SARS-CoV-2 dynamics and control. Based on the reconstructed transmission chains, we estimate a median serial interval of 5.7 days, with an inter-quartile range (IQR) of 2.8 to 8.7 days, which represents the time interval between symptom onset of an infector and his/her infectee (Fig. S2B). The median generation interval, defined as the interval between the infection times of an infector and his/her infectee, is 5.3 days, with an IQR of 3.1 to 8.7 days (Fig. S2A). We estimate that 63.2% (95% CI, 59.6% to 66.4%) of all transmission events occur before symptom onset, which is comparable with findings from other studies (*6–8*, *10–13*, *19*, *28*). However, these estimates are impacted by the intensity of interventions, as we will show later. In Hunan, interventions including case isolation, contact tracing, and close-contact quarantine were in place throughout the epidemic.

Case isolation and contact quarantine are meant to prevent potentially infectious individuals from contacting susceptible individuals, effectively shortening the infectious period. As a result, we would expect right censoring of the generation and serial interval distributions (*29*). Symptomatic cases represent 86.5% of all SARS-CoV-2 infections in our data; among these patients, we observe longer generation intervals for cases isolated later in the course of their infection (Fig. 3A). The median generation interval increases from 4.1 days (IQR, 1.9 to 7.2 days) for cases isolated 2 day since symptom onset, to 7.0 days (IQR, 3.6 to 11.1 days) for those isolated more than 6 days after symptom onset ($p<0.001$, Mann-Whitney U test). We observe similar tends for the serial interval distributions (Fig. 3B). The median serial interval increases from 1.7 days (IQR, -1.5 to 4.7 days) for cases isolated less than 2 day after symptom onset, to 7.3 days (IQR, 3.4 to10.9 days) for those isolated more than 6 days after symptom onset ($p<0.001$, Mann-Whitney U test).

Faster case isolation restricts transmission to the earlier stages of infection, thus inflating the contribution of pre-symptomatic transmission (Fig. 3C). The proportion of pre-symptomatic transmission is estimated at 86.6% (95% CI, 80.8% to 92.3%) if cases are isolated within 2 day of symptom onset, while this proportion decreases to 47.5% (95% CI, 41.4% to 53.3%) if cases are isolated more than 6 days after symptom onset ($p<0.001$, Mann-Whitney U test).

To adjust for censoring due to case isolation and reconstruct the infectiousness profile of a SARS-CoV-2 infection in the absence of intervention, we characterize the changes in the speed of case isolation over time in Hunan. Fig. S4 shows the distributions of time from symptom onset to isolation during three different phases of epidemic control, coinciding with major changes in COVID-19 case definition (*Phase I*: before Jan. 27th; *Phase II*: Jan. 27th – Feb. 4th; *Phase III*: after Feb. 4th, Fig. S3) (*30*). In *Phase I*, 78% of cases were detected through passive surveillance; as a result, most cases were isolated after symptom onset (median time from onset to isolation 5.4 days, IQR (2.7, 8.2) days, Fig. S4A). In contrast, in *Phase III*, 66% of cases were detected through active contact tracing, shortening the median time from onset to isolation to -0.1 days with IQR (-2.9, 1.7) days, Fig. S4C. *Phase II* is intermediate. We use mathematical models (detailed in Materials and Methods) to dynamically adjust the serial interval distribution for censoring and apply the same approach to the time interval between a primary case's symptom onset and onward transmission (Fig. S6A-B). These censoring-adjusted distributions can be rescaled by the basic reproduction number R_0 to reflect the risk of transmission of a typical SARS-CoV-2 case since the time of infection or since symptom onset (Fig 3D-E). Assuming no interventions were in place, we estimated that infectiousness peaks near the time of symptoms onset (Fig. S6B), consistent with our regression estimates that transmission risk is higher if the onset of the primary case occurred within the window of exposure (Table 1).

Evaluating the impact of individual and population-based interventions on SARS-CoV-2 transmission

Next, we use the estimated infectiousness profile of SARS-CoV-2 (Fig. 3D-E) to evaluate the impact of case isolation on transmission. We first set a baseline reproduction number R_0 for SARS-CoV-2, in the absence of control. Results from a recent study (*30*) suggest that the observed initial growth rate in Wuhan was 0.15 day^{-1} (95% CI, 0.14 to 0.17), although the growth rate could be substantially lower (0.08 day^{-1}) if accounting for changes in case definition. Conservatively, we consider the upper value of the growth rate at 0.15 day^{-1} together with our generation interval distribution adjusted for censoring (Fig. S6A), to estimate R_0. We obtain a baseline reproduction number $R_0 = 2.23$ (95% CI, 2.13 to 2.43), using the renewal equation framework (*31*). This

represents a typical scenario of unmitigated SARS-CoV-2 transmissibility in an urban setting. The reconstructed infectiousness profile in the absence of control is shown in solid red lines in Fig. 3D-E, with respect to time of infection and symptom onset respectively. Notably, SARS-CoV-2 infectiousness peaks slightly before symptom onset (-0.1 days on average), with 86% of the overall infectiousness concentrated within ± 5 days of symptom onset and 52% of the overall infectiousness in the pre-symptomatic phase (Fig. 3E).

Next, we evaluate the impact of case isolation on transmission by considering three different intervention scenarios mimicking the speed of isolation in the three phases of the Hunan epidemic control. We first assume that 100% of infections are detected and isolated and that isolation is fully protective (i.e., there is no onward transmission after the patient has been isolated/quarantined). The infectiousness profiles of the three intervention scenarios are shown in dashed lines in Fig. 3D-E. We find that the basic reproduction number decreases in all intervention scenarios, but the projected decrease is not sufficient to interrupt transmission (Fig. 3D, $R_0^E = 1.77$ for *Phase I*, $R_0^E = 1.54$ for *Phase II*, and $R_0^E = 1.10$ for *Phase III*).

We further relax the assumption of 100% case detection and isolation and relate changes in the basic reproduction number to the efficacy of surveillance and compliance with case isolation and contact quarantine (measured as the fraction of total infections isolated) as well as the speed of isolation (delay from symptom onset to isolation, phase diagram in Fig. 3F). Dashed lines in Fig. 3F illustrate 30%, 40% and 50% of reduction in R_0. To reduce the R_0 in half (the minimum amount of transmission reduction required to achieve control for a baseline $R_0 \sim 2$), 100% of infections would need to be isolated even if individuals are isolated as early as the day of symptom onset. In practice, epidemic control is unrealistic to achieve if case isolation and quarantine of close contacts are the only measures in place.

Individual-based interventions are unlikely to be the sole mode of SARS-CoV-2 control in the months ahead. Layering additional physical distancing measures (e.g. through increased teleworking, reduced operation in the service industry, or broader adoption of face mask wearing), could provide substantial relief on the burden of case isolation and contact quarantine. The synergistic effects of these interventions are illustrated in Fig. 3G. We find that a 30% reduction in transmission from population-level measures would require a 70% case detection rate to achieve epidemic control, assuming that cases can be promptly isolated on average upon symptom presentation. Of note, a 30% reduction in transmission could be achieved in various ways and does not necessarily require physical distancing measures. It could also encompass the benefits of residual population-level immunity from the first wave of COVID-19, especially in hard-hit regions (*32*, *33*). As a sensitivity analysis, we further consider a more optimistic scenario with a lower baseline $R_0 = 1.59$, corresponding to an epidemic growth rate of 0.08 day^{-1} (95% CI, 0.06 to 0.10) in Wuhan (*30*), which is adjusted for reporting changes. As expected, control is much easier to achieve in this scenario: if detected SARS-CoV-2 infections are effectively isolated on average 2 days after symptom onset, a 25% population-level reduction in transmission coupled with a 43% infection isolation rate is able to achieve control (Fig. 3H).

Discussion

To our knowledge, our study is the most comprehensive analysis of contact tracing data so far. Detailed information on 1,178 SARS-CoV-2 infected individuals along with their 15,648 contacts has allowed us to dissect the behavioral and clinical drivers of SARS-CoV-2 transmission; to evaluate how transmission opportunities are modulated by individual and population-level

interventions, and to characterize the typical infectiousness profile of a case. Informed by this understanding, particularly the importance of pre-symptomatic transmission, we have evaluated the plausibility of SARS-CoV-2 control through individual and population-based interventions.

Contacts in healthcare settings pose the lowest risk of transmission in Hunan, suggesting that adequate protective measures against SARS-CoV-2 were taken in hospitals and medical observation centers (Table 1). The risk of transmission scales positively with the closeness of social interactions, with a lower per-contact risk estimated for community exposures (including contacts in the public transportation system, food and entertainment venues), intermediate risk for social and extended family settings, and highest risk in the household. The transmission risk associated with household exposures is further elevated when intense physical distancing is enforced, and for contacts that last longer. These lines of evidence support that SARS-CoV-2 transmission is facilitated by close proximity, confined settings, and high frequency of contacts. We cannot evaluate the relative risks of transmission in other settings such as schools, workplaces, conferences, prisons, or factories, as no contacts in these settings were reported in the Hunan dataset.

Regression analysis indicates a higher risk of transmission when an individual is exposed to a SARS-CoV-2 patient around the time of symptom onset, in line with our reconstructed infectiousness profile that peaks just before symptom onset. These epidemiological findings are in agreement with viral shedding studies (*6*, *34*, *35*). We estimate that overall in Hunan, ~63% of all transmission events were from pre-symptomatic individuals, in line with estimates from other modeling studies (*6*, *7*, *10*, *12*, *36*). However, this proportion is inflated by case isolation and contact quarantine measures, with right-censoring affecting transmission primarily in the symptomatic phase. We estimate that the relative contribution of pre-symptomatic transmission drops to ~52% in an uncontrolled scenario where case-based interventions are absent.

Case isolation reduces the "effective" infectious period of SARS-CoV-2 infected individuals by blocking contacts with susceptible individuals. We observe that faster isolation significantly reduces CCRs across settings (Fig. 2E). We also observe shorter serial and generation intervals and a larger fraction of pre-symptomatic transmission when individuals are isolated faster (Fig. 3A-C). In contrast, population-level physical distancing measures have differential impacts on CCRs, decreasing CCRs in social and community settings, while increasing CCRs in the household and family. As a result, strict physical distancing confines the epidemic mostly to families and households (see also Fig. S7). The precise impact of physical distancing on transmission is difficult to separate from that of individual-based interventions. However, our analysis suggests that physical distancing changes the topology of the transmission network by affecting the number and duration of interactions. Interestingly, the topological structure of the household contact network is highly clustered (*37*), and high clustering is expected to hinder epidemic spread (*38*, *39*). Thus, these higher-order topological changes could contribute to reducing transmission beyond the effects expected from an overall reduction in CCRs. Observationally, the effectiveness of physical distancing measures on reducing COVID-19 transmission has been demonstrated in China (*16*, *40*) and elsewhere (*41*).

We have explored the feasibility of SARS-CoV-2 epidemic control against two important metrics related to case isolation and contact quarantine: the speed of isolation and the infection isolation proportion (Fig. 3F). For a baseline transmission scenario compatible with the initial growth phase of the epidemic in Wuhan, we find that epidemic control solely relying on case isolation and quarantine of close contact is difficult to achieve. Layering case isolation and

quarantine of close contact with moderate physical distancing makes control more likely over a range of plausible parameters - a situation that could be further improved by residual immunity from the first wave of SARS-CoV-2 circulation (*32*, *33*). Successful implementation of contact tracing requires a low-level of active infections in the community, as the number of contacts to be monitored is several folds the number of infections (~13 contacts were being traced per SARS-CoV-2 infected individual in Hunan). The timing of easing of lockdown measures should align with the capacities of testing and contact tracing efforts, relative to the number of active infections in the community. Technology-based approaches could also facilitate intense contact tracing efforts (*7*, *42*).

It is important to point out several caveats. Our study is likely underpowered to assess the transmission potential of asymptomatic infections given the relatively small fraction of these infections in our data (13.5% overall and 22.1% of infections captured through contact tracing). There is no statistical support for decreased transmission from asymptomatic individuals (Fig. S1A), although we observe a positive, but non-significant trend in transmission risk scaling with disease severity. There is conflicting evidence from viral shedding studies; viral load appears independent of clinical severity in some studies (*6*, *23*, *35*, *43*) while others suggest faster viral clearance in asymptomatic individuals (*44*). Before February 7 in Hunan, a fraction of contacts were only tested upon symptom presentation, which may affect our estimates on age-specific susceptibility, as younger individuals are less likely to develop symptom (*45*). The rate of asymptomatic infections and their impact on transmission have profound implications on the feasibility of control through individual-based interventions. Careful serological studies combined with virologic testing in households and other controlled settings will be needed to fully resolve the role of asymptomatic infections and viral shedding on transmission.

In conclusion, detailed contact tracing data illuminate important heterogeneities in SARS-CoV-2 transmission driven by biological and behavioral factors and modulated by the impact of interventions. Crucially, and in contrast to SARS-CoV-1, the ability of SARS-CoV-2 to transmit during the host's pre-symptomatic phase makes it particularly difficult to achieve epidemic control (*46*). Our risk factor estimates can provide useful evidence to guide the design of more targeted and sustainable mitigation strategies, while our reconstructed transmission kinetics will help calibrate further modeling efforts. Moving forward, it will be particularly important to intensify collection and analysis of rich contact tracing data to monitor how transmission risk changes over time with growing population immunity, waxing and waning of interventions, and reactive changes in human behavior and contact opportunities.

References

1. W. Guan, Z. Ni, Y. Hu, W. Liang, C. Ou, J. He, L. Liu, H. Shan, C. Lei, D. S. C. Hui, B. Du, L. Li, G. Zeng, K.-Y. Yuen, R. Chen, C. Tang, T. Wang, P. Chen, J. Xiang, S. Li, J. Wang, Z. Liang, Y. Peng, L. Wei, Y. Liu, Y. Hu, P. Peng, J. Wang, J. Liu, Z. Chen, G. Li, Z. Zheng, S. Qiu, J. Luo, C. Ye, S. Zhu, N. Zhong, Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).

2. J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, G. M. Leung, Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020).

3. G. Grasselli, A. Zangrillo, A. Zanella, M. Antonelli, L. Cabrini, A. Castelli, D. Cereda, A. Coluccello, G. Foti, R. Fumagalli, G. Iotti, N. Latronico, L. Lorini, S. Merler, G. Natalini, A. Piatti, M. V. Ranieri, A. M. Scandroglio, E. Storti, M. Cecconi, A. Pesenti, Baseline Characteristics and Outcomes of 1591 Patients Infected with SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA - J. Am. Med. Assoc.* **323**, 1574–1581 (2020).

4. R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).

5. K. Sun, J. Chen, C. Viboud, Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digit. Heal.* **2**, e201–e208 (2020).

6. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).

7. L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, C. Fraser, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science (80-.).* **368** (2020), doi:10.1126/science.abb6936.

8. A. Kimball, K. M. Hatfield, M. Arons, A. James, J. Taylor, K. Spicer, A. C. Bardossy, L. P. Oakley, S. Tanwar, Z. Chisty, J. M. Bell, M. Methner, J. Harney, J. R. Jacobs, C. M. Carlson, H. P. McLaughlin, N. Stone, S. Clark, C. Brostrom-Smith, L. C. Page, M. Kay, J. Lewis, D. Russell, B. Hiatt, J. Gant, J. S. Duchin, T. A. Clark, M. A. Honein, S. C. Reddy, J. A. Jernigan, A. Baer, L. M. Barnard, E. Benoliel, M. S. Fagalde, J. Ferro, H. G. Smith, E. Gonzales, N. Hatley, G. Hatt, M. Hope, M. Huntington-Frazier, V. Kawakami, J. L. Lenahan, M. D. Lukoff, E. B. Maier, S. McKeirnan, P. Montgomery, J. L. Morgan, L. A. Mummert, S. Pogosjans, F. X. Riedo, L. Schwarcz, D. Smith, S. Stearns, K. J. Sykes, H. Whitney, H. Ali, M. Banks, A. Balajee, E. J. Chow, B. Cooper, D. W. Currie, J. Dyal, J. Healy, M. Hughes, T. M. McMichael, L. Nolen, C. Olson, A. K. Rao, K. Schmit, N. G. Schwartz, F. Tobolowsky, R. Zacks, S. Zane, Asymptomatic and presymptomatic SARS-COV-2 infections in residents of a long-term care skilled nursing facility - King County, Washington, March 2020. *Morb. Mortal. Wkly. Rep.* **69**, 377–381 (2020).

9. W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, V. J. Lee, Presymptomatic transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. *Morb. Mortal. Wkly. Rep.* **69**, 411–415 (2020).

10. Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, L. A. Meyers, Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg. Infect. Dis.* **26**, 1341–1343 (2020).

11. E. Lavezzo, E. Franchin, C. Ciavarella, G. Cuomo-Dannenburg, L. Barzon, C. Del Vecchio, L. Rossi, R. Manganelli, A. Loregian, N. Navarin, D. Abate, M. Sciro, S. Merigliano, E. Decanale, M. C. Vanuzzo, F. Saluzzo, F. Onelia, M. Pacenti, S. Parisi, G. Carretta, D. Donato, L. Flor, S. Cocchio, G. Masi, A. Sperduti, L. Cattarino, R. Salvador, K. A. M. Gaythorpe, I. C. L. C.-19 R. Team, A. R. Brazzale, S. Toppo, M. Trevisan, V. Baldo, C. A. Donnelly, N. M. Ferguson, I. Dorigatti, A. Crisanti, Suppression of COVID-19 outbreak in the municipality of Vo, Italy. *medRxiv* (2020), doi:10.1101/2020.04.17.20053157.

12. S. Hu, W. Wang, Y. Wang, M. Litvinova, 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, Y. Zhou, K. Sun, A. Vespignani, C. Viboud, L. Gao, M. Ajelli, H. Yu, Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. *medRxiv* (2020), doi:10.1101/2020.07.23.20160317.

13. Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, T. Feng, 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.* (2020), doi:10.1016/S1473-3099(20)30287-5.

14. A. Pan, L. Liu, C. Wang, H. Guo, X. Hao, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, S. Wei, T. Wu, Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA*. **323**, 1915 (2020).

15. B. J. Cowling, S. T. Ali, T. W. Y. Ng, T. K. Tsang, J. C. M. Li, M. W. Fong, Q. Liao, M. Y. Kwan, S. L. Lee, S. S. Chiu, J. T. Wu, P. Wu, G. M. Leung, Impact assessment of non-pharmaceutical interventions

against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Heal.* **5**, e279–e288 (2020).

16. J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, H. Yu, Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science (80-.).* **368**, 1481–1486 (2020).

17. S. W. Park, K. Sun, C. Viboud, B. T. Grenfell, J. Dushoff, Potential roles of social distancing in mitigating the spread of coronavirus disease 2019 (COVID-19) in South Korea. *medRxiv* (2020), doi:10.1101/2020.03.27.20045815.

18. A. Aleta, D. Martin-Corral, A. P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini, S. Merler, A. Pentland, A. Vespignani, E. Moro, Y. Moreno, D. Mart, A. Pastore, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, Modeling the impact of social distancing testing contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. *medRxiv* (2020), doi:10.1101/2020.05.06.20092841.

19. J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi, X. Chen, Q. Wu, Y. Liang, X. Wang, J. Yang, K. Sun, I. M. Longini, M. E. Halloran, P. Wu, B. J. Cowling, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, H. Yu, 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).

20. K. Leung, J. T. Wu, D. Liu, G. M. Leung, First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet*. **395**, 1382–1393 (2020).

21. M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. Elizabeth Halloran, I. M. Longini, A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science (80-.).* **368**, 395–400 (2020).

22. H. Tian, Y. Liu, Y. Li, C. H. Wu, B. Chen, M. U. G. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, B. Wang, P. Yang, Y. Cui, Y. Song, P. Zheng, Q. Wang, O. N. Bjornstad, R. Yang, B. T. Grenfell, O. G. Pybus, C. Dye, An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science (80-.).* **368**, 638–642 (2020).

23. D. Cereda, M. Tirani, F. Rovida, V. Demicheli, M. Ajelli, P. Poletti, F. Trentini, G. Guzzetta, V. Marziano, A. Barone, M. Magoni, S. Deandrea, G. Diurno, M. Lombardo, M. Faccini, A. Pan, R. Bruno, E. Pariani, G. Grasselli, A. Piatti, M. Gramegna, F. Baldanti, A. Melegaro, S. Merler, The early phase of the COVID-19 outbreak in Lombardy, Italy. *arXiv* (2020) (available at http://arxiv.org/abs/2003.09320).

24. A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, N. Davies, A. Gimma, K. van Zandvoort, H. Gibbs, J. Hellewell, C. I. Jarvis, S. Clifford, B. J. Quilty, N. I. Bosse, S. Abbott, P. Klepac, S. Flasche, Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 553–558 (2020).

25. Baidu.com, Baidu Qianxi: Within city mobility index. *https://qianxi.baidu.com/2020/* (2020).

26. J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, W. J. Edmunds, Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med.* **5**, e74 (2008).

27. M. Litvinova, Q. H. Liu, E. S. Kulikov, M. Ajelli, Reactive school closure weakens the network of social interactions and reduces the spread of influenza. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 13174–13181 (2019).

28. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).

29. S. T. Ali, L. Wang, E. H. Y. Lau, X.-K. Xu, Z. Du, Y. Wu, G. M. Leung, B. J. Cowling, Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* (2020), doi:10.1126/science.abc9004.

30. T. K. Tsang, P. Wu, Y. Lin, E. H. Y. Lau, G. M. Leung, B. J. Cowling, Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *Lancet Public Heal.* **5**, e289–e296 (2020).

31. J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B Biol. Sci.* **274**, 599–604 (2007).

32. E. S. Rosenberg, J. M. Tesoriero, E. M. Rosenthal, R. Chung, M. A. Barranco, L. M. Styer, M. M. Parker, S.-Y. John Leung, J. E. Morne, D. Greene, D. R. Holtgrave, D. Hoefer, J. Kumar, T. Udo, B. Hutton, H. A. Zucker, Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. *Ann. Epidemiol.* (2020), doi:10.1016/j.annepidem.2020.06.004.

33. S. Stringhini, A. Wisniak, G. Piumatti, A. S. Azman, S. A. Lauer, H. Baysson, D. De Ridder, D. Petrovic, S. Schrempft, K. Marcus, S. Yerly, I. Arm Vernez, O. Keiser, S. Hurst, K. M. Posfay-Barbe, D. Trono, D. Pittet, L. Gétaz, F. Chappuis, I. Eckerle, N. Vuilleumier, B. Meyer, A. Flahault, L. Kaiser, I. Guessous, Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet* (2020), doi:10.1016/s0140-6736(20)31304-0.

34. R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmann, K. Zwirglmaier, C. Drosten, C. Wendtner, Virological assessment of hospitalized patients with COVID-2019. *Nature*. **581**, 465–469 (2020).

35. K. K. W. To, O. T. Y. Tsang, W. S. Leung, A. R. Tam, T. C. Wu, D. C. Lung, C. C. Y. Yip, J. P. Cai, J. M. C. Chan, T. S. H. Chik, D. P. L. Lau, C. Y. C. Choi, L. L. Chen, W. M. Chan, K. H. Chan, J. D. Ip, A. C. K. Ng, R. W. S. Poon, C. T. Luo, V. C. C. Cheng, J. F. W. Chan, I. F. N. Hung, Z. Chen, H. Chen, K. Y. Yuen, Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect. Dis.* **20**, 565–574 (2020).

36. Y. Liu, S. Funk, S. Flasche, The contribution of pre-symptomatic infection to the transmission dynamics of COVID-2019. *Wellcome Open Res.* (2020), doi:10.12688/wellcomeopenres.15788.1.

37. Q. H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, A. Vespignani, Measurability of the epidemic reproduction number in data-driven contact networks. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12680–12685 (2018).

38. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).

39. E. M. Volz, J. C. Miller, A. Galvani, L. Meyers, Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput. Biol.* **7**, e1002042 (2011).

40. N. M. Ferguson, K. E. C. Ainslie, C. E. Walters, H. Fu, S. Bhatia, H. Wang, X. Xi, M. Baguelin, S. Bhatt, A. Boonyasiri, O. Boyd, L. Cattarino, C. Ciavarella, Z. Cucunuba, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, S. L. van Elsland, R. FitzJohn, K. Gaythorpe, A. C. Ghani, W. Green, A. Hamlet, W. Hinsley, N. Imai, D. Jorgensen, E. Knock, D. Laydon, G. Nedjati-Gilani, L. C. Okell, I. Siveroni, H. A. Thompson, H. J. T. Unwin, R. Verity, M. Vollmer, P. G. T. Walker, Y. Wang, O. J. Watson, C. Whittaker, P. Winskill, C. A. Donnelly, S. Riley, Evidence of initial success for China exiting COVID-19 social distancing policy after achieving containment. *Wellcome Open Res.* **5**, 81 (2020).

41. S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, M. Monod, P. N. Perez-Guzman, N. Schmit, L. Cilloni, K. E. C. Ainslie, M. Baguelin, A. Boonyasiri, O. Boyd, L. Cattarino, L. V. Cooper, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, B. Djaafara, I. Dorigatti, S. L. van Elsland, R. G. FitzJohn, K. A. M. Gaythorpe, L. Geidelberg, N. C. Grassly, W. D. Green, T. Hallett, A. Hamlet, W. Hinsley, B. Jeffrey, E. Knock, D. J. Laydon, G. Nedjati-Gilani, P. Nouvellet, K. V. Parag, I. Siveroni, H. A. Thompson, R. Verity, E. Volz, C. E. Walters, H. Wang, Y. Wang, O. J. Watson, P. Winskill, X. Xi, P. G. Walker, A. C. Ghani, C. A. Donnelly, S. M. Riley, M. A. C. Vollmer, N. M. Ferguson, L. C. Okell, S. Bhatt, Estimating the effects of non-pharmaceutical interventions on

COVID-19 in Europe. *Nature* (2020), doi:10.1038/s41586-020-2405-7.

42. J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, S. Flasche, B. J. Quilty, N. Davies, Y. Liu, S. Clifford, P. Klepac, M. Jit, C. Diamond, H. Gibbs, K. van Zandvoort, S. Funk, R. M. Eggo, Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Heal.* **8**, e488–e496 (2020).

43. Q.-X. Long, X.-J. Tang, Q.-L. Shi, Q. Li, H.-J. Deng, J. Yuan, J.-L. Hu, W. Xu, Y. Zhang, F.-J. Lv, K. Su, F. Zhang, J. Gong, B. Wu, X.-M. Liu, J.-J. Li, J.-F. Qiu, J. Chen, A.-L. Huang, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* (2020), doi:10.1038/s41591-020-0965-6.

44. N. V. V. Chau, V. T. Lam, N. T. Dung, L. M. Yen, N. N. Q. Minh, L. M. Hung, N. M. Ngoc, N. T. Dung, D. N. H. Man, L. A. Nguyet, L. T. H. Nhat, L. N. T. Nhu, N. T. H. Ny, N. T. T. Hong, E. Kestelyn, N. T. P. Dung, N. T. Phong, T. C. Xuan, T. T. Hien, T. N. H. Tu, R. B. Geskus, T. T. Thanh, N. T. Truong, N. T. Binh, T. C. Thuong, G. Thwaites, L. Van Tan, The natural history and transmission potential of asymptomatic SARS-CoV-2 infection. *medRxiv* (2020), doi:10.1101/2020.04.27.20082347.

45. P. Poletti, M. Tirani, D. Cereda, F. Trentini, G. Guzzetta, G. Sabatino, V. Marziano, A. Castrofino, F. Grosso, G. Del Castillo, R. Piccarreta, A. L. C.-19 T. Force, A. Andreassi, A. Melegaro, M. Gramegna, M. Ajelli, S. Merler, Probability of symptoms and critical disease after SARS-CoV-2 infection. *arXiv* (2020) (available at http://arxiv.org/abs/2006.08471).

46. C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, Factors that make an infectious disease outbreak controllable. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6146–6151 (2004).

47. Diagnosis and treatment guideline on pneumonia infection with 2019 novel coronavirus (6th edition). *Natl. Heal. Comm. People's Repub. China* (2020), (available at http://www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df351d7da8aefc2.shtml).

48. B. W. Silverman, *Density estimation for statistics and data analysis* (CRC press, 1986), vol. 26.

49. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*. **17**, 261–272 (2020).

50. S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).

51. D. Bates, M. Mächler, B. M. Bolker, S. C. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

52. W. N. Venables, B. D. Ripley, *Modern Applied Statistics with S* (Springer, New York, Fourth., 2002; http://www.stats.ox.ac.uk/pub/MASS4).

Acknowledgements

The authors acknowledge Dr Christophe Fraser from the University of Oxford, Dr David Spiro from Fogarty International Center, National Institutes of Health, and Dr Peter Kilmarx from

Fogarty International Center, National Institutes of Health for their helpful comments on the manuscript. This article does not necessarily represent the views of the NIH or the U.S. government. **Funding:** Hongjie Yu acknowledges financial support from the National Science Fund for Distinguished Young Scholars (number 81525023) and the National Science and Technology Major Project of China (numbers 2018ZX10201001–010, 2018ZX10713001–007, and 2017ZX10103009–005). Lidong Gao acknowledges financial support from Hunan Provincial Innovative Construction Special Fund: Emergency response to COVID-19 outbreak (No. 2020SK3012). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. **Author contributions:** C.V. and H.Y. are joint senior authors. K.S. C.V. and H.Y. designed the experiments. L.G., W.W, and Y.W. collected data. K.S., W.W., L.G., and Y.W. analyzed the data. K.S., W.W., L.G., Y.W., K.L., L.R., Z.Z., X.C., S.Z., Y.H., Q.S., Z.L., M.L., A.V., M.A., C.V., and H.Y. interpreted the results. K.S., W.W., L.G., Y.W., M.L., A.V., M.A., C.V. and H.Y. wrote the manuscript. **Competing interests:** Hongjie Yu has received research funding from Sanofi Pasteur, GlaxoSmithKline, Yichang HEC Changjiang Pharmaceutical Company, and Shanghai Roche Pharmaceutical Company. All other authors declare no competing interests. **Ethics statement:** The study was approved by the Institutional review board from Hunan Provincial Center for Disease Control and Prevention (IRB#2020005). Data were deidentified, and informed consent was waived. All necessary patient/participant consent has been obtained and the appropriate institutional forms have been archived. **Data and materials availability:** the datasets generated during and/or analyzed during the current study are not publicly available due to privacy concerns but are available from the corresponding authors on reasonable request.

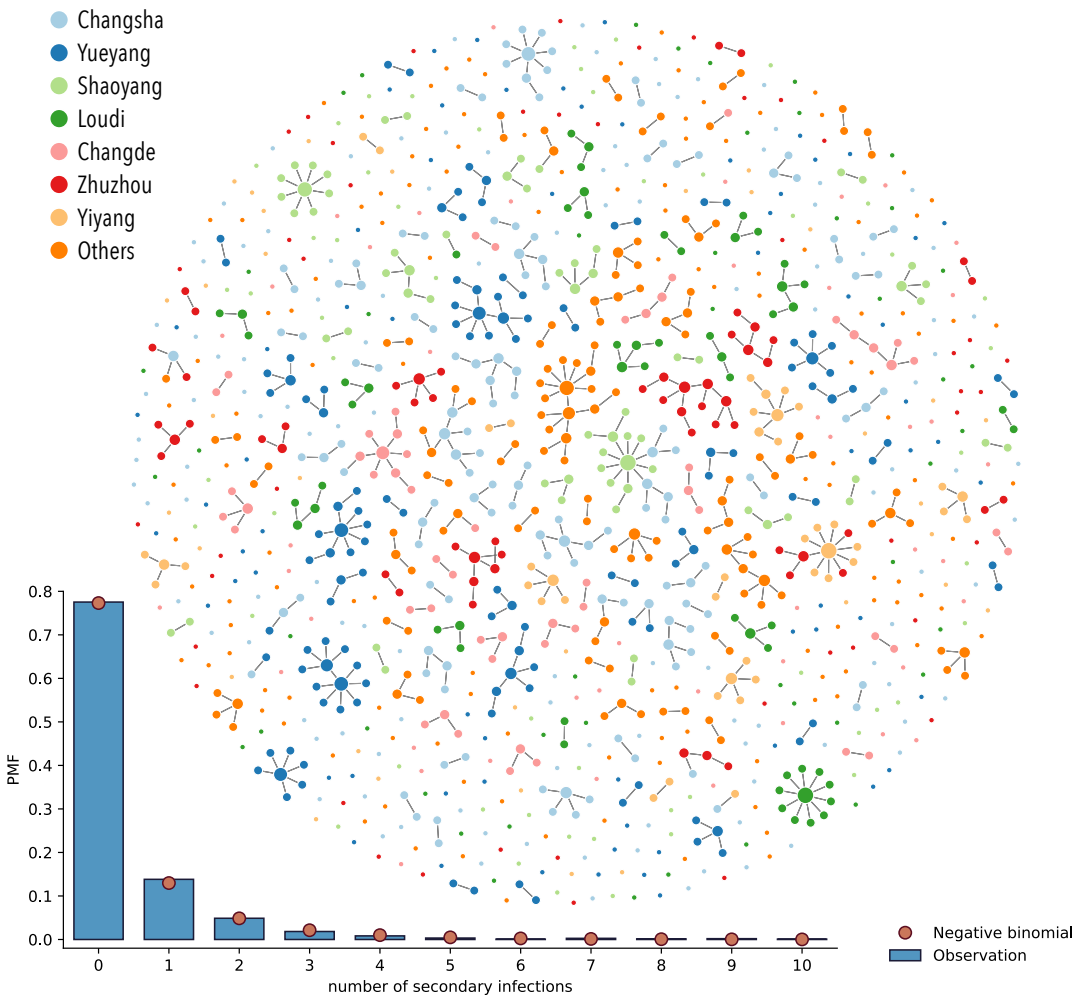

Fig. 1. SARS-CoV-2 transmission chains. Top: Realization of the reconstructed transmission chains among 1,178 SARS-CoV-2 infected individuals in Hunan province. Each node in the network represents a patient infected with SARS-CoV-2 and each link represents an infector-infectee relationship. Colors of the node denote the reporting prefecture of infected individuals. Bottom: The bar plot shows the distribution of the number of secondary infections based on 100 stochastic samples of the reconstructed transmission chains. Red dots represent fit from a negative binomial distribution.

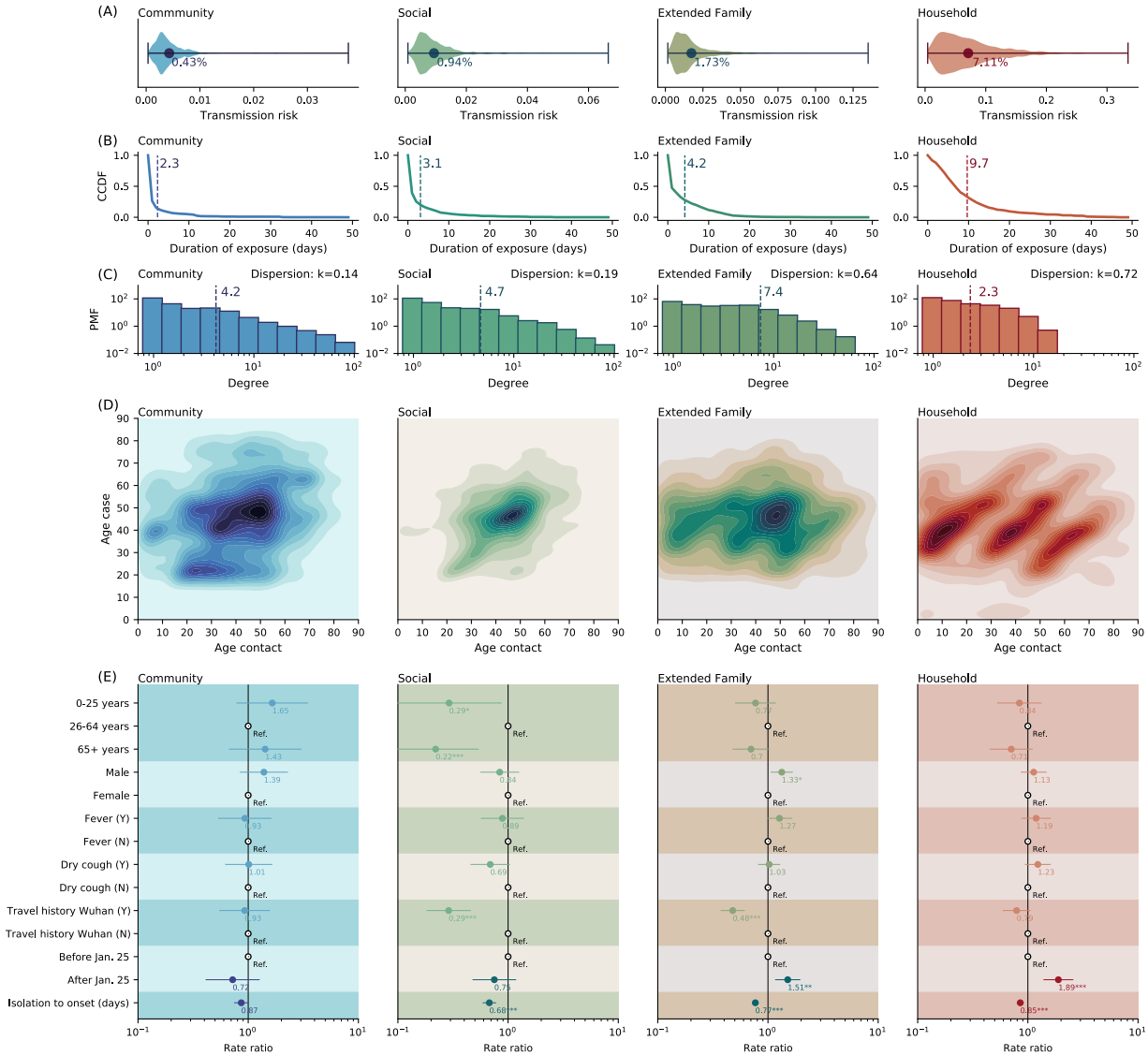

Fig. 2. Heterogeneity in contact rates of SARS-CoV-2 cases and impact of interventions, by category of contact. Columns from left to right represent community contacts (public transportation, food & entertainment), social contacts, extended family contacts, and household contacts. (A) Distribution of the transmission risk by setting, adjusted for all other covariates in Fig. S1. (B) Cumulative distribution function of the duration of exposure (i.e. the probability that exposure is longer or equal to a certain value). Dashed vertical lines indicate average values. (C) The distribution of the number of unique contacts (degree distribution) of the primary cases in each setting. The dashed vertical lines indicate average values. (D) Age distribution of SARS-CoV-2 case-contact pairs (contact matrices). (E) Rate ratios of negative binomial regression of the cumulative contact rates (CCRs) against predictors including the infector's age, sex, presence of fever/cough, Wuhan travel history, whether symptom onset occurred before social distancing was in place (before or after Jan. 25, 2020), and time from isolation to symptom onset. CCRs represent the sum of relevant contacts over a one-week window centered at the date of the primary case's symptom onset. Dots and lines indicate point estimates and 95% confidence interval of the rate ratios, numbers below the dots indicate the numerical value of the point estimates; Ref. stands for reference category; * indicates p-value<0.05, ** indicates p-value<0.01, *** indicates p-value<0.001.

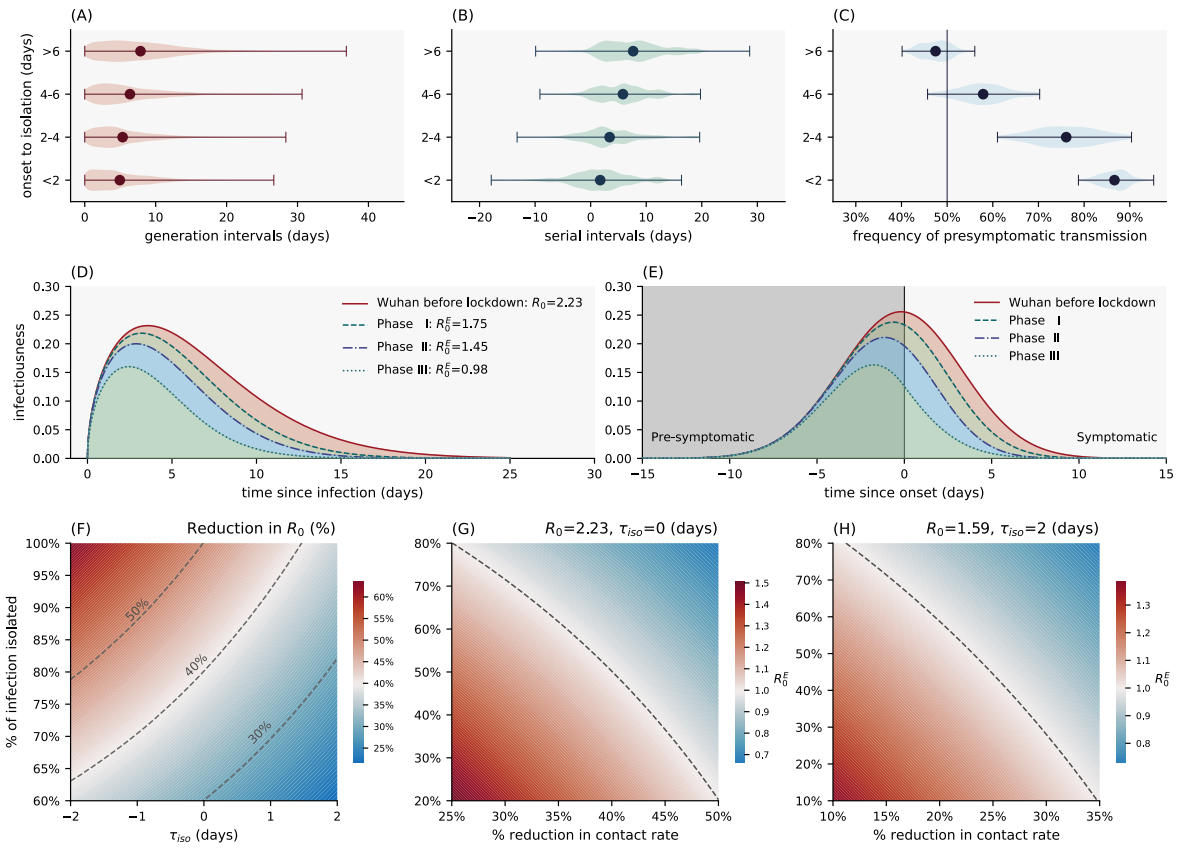

Fig. 3. (A) Violin plot of the generation interval distributions stratified by time from symptom onset to isolation/pre-symptomatic quarantine (B) Same as A but for the serial interval distributions (C) Same as A but for the fraction of pre-symptomatic transmission, among all transmission events, with vertical line indicating 50% of pre-symptomatic transmission. Dots represent the mean and whiskers represents minimum and maximum. (D) Estimated average transmission risk of a SARS-CoV-2 infected individual since time of infection under four intervention scenarios: the red solid line represents an uncontrolled epidemic scenario modelled after the early epidemic dynamics in Wuhan before lockdown; the dashed lines represent scenarios where quarantine and case isolation are in place and mimic Phase I, II, and III of epidemic control in Hunan. The shapes of these curves match that of the generation interval distributions in each scenario while the areas under the curve are equal to the ratio of the baseline/effective basic reproduction numbers (R_0/R_0^Es). (E) Same as in D but with time since symptom onset on the x-axis (colors are as in (D)). The vertical line represents symptom onset. (F) Reduction (percentage) in the basic reproduction number as a function of mean time from symptom onset (or from peak infectiousness for asymptomatic cases) to isolation τ_{iso} (x-axis) and fraction of SARS-CoV-2 infections being isolated (y-axis). The distribution of onset to isolation follows a normal distribution with mean τ_{iso} and standard deviation of 2 days. The dashed lines indicate 30%, 40% and 50% reductions in R0 under interventons. (G) Effective basic reproduction number as a function of population-level reduction in contact rates (i.e. through physical distancing, expressed as a percentage, x-axis) and isolation rate (fraction of total infections detected and further isolated). We assume baseline basic reproduction number $R_0 = 2.23$, and a normal distribution for the distribution from onset to isolation with mean of 0 days and standard deviation of 2 days. The dashed line represents the epidemic threshold $R^E = 1$. The blue area indicates region below the epidemic threshold (namely, controlled epidemic) and the red area indicates region above the epidemic threshold. (H) Same as in G but assuming $R_0 = 1.59$ (a more optimistic estimate of R_0 in Wuhan adjusted for reporting changes), and a normal distribution for the distribution from onset to isolation with mean of 2 days and standard deviation of 2 days.

Risk factors		Odds ratio	95% CI
Household contacts	Before 01/25/2020	2.17***	(1.38, 3.43)
	After 01/25/2020	3.72***	(2.38, 5.81)
Extended Family contacts	Before 01/25/2020	1.00	Reference
	After 01/25/2020	0.96	(0.61, 1.49)
Social contacts	Before 01/25/2020	0.63	(0.37, 1.07)
	After 01/25/2020	0.43*	(0.22, 0.86)
Community contacts	Before 01/25/2020	0.39**	(0.19, 0.78)
	After 01/25/2020	0.17**	(0.05, 0.62)
Healthcare contacts	Before 01/25/2020	0.18*	(0.04, 0.73)
	After 01/25/2020	0.07	(0.00, 1.07)
Duration of exposure (days)		1.10***	(1.05, 1.15)
Symptom onset within exposure window (Yes)		1.46*	(1.08, 1.97)

Table 1: SARS-CoV-2 transmission risk in Hunan by contact setting, duration of exposure, and whether the exposure window contains the date of symptom onset of the primary case – a period of intense viral shedding. Risk is further stratified by the date of implementation of social distancing interventions in Hunan, which is 01/25/2020. The regression model is adjusted for demographic characteristics of the cases and their contacts, clinical symptoms, and travel history. Details are provided in the Material and Methods, while the full results of the regression including additional risk factors are shown in Fig. S1. * indicates p-value<0.05, ** indicates p-value<0.01, *** indicates p-value<0.001.

Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2: Supplementary Materials

Kaiyuan Sun*, Wei Wang, Lidong Gao, Yan Wang, Kaiwei Luo, Lingshuang Ren, Zhifei Zhan, Xinghui Chen, Shanlu Zhao, Yiwei Huang, Qianlai Sun, Ziyan Liu, Maria Litvinova, Alessandro Vespignani, Marco Ajelli, Cécile Viboud, Hongjie Yu*.

Correspondence to: kaiyuan.sun@nih.gov (K.S.); yhj@fudan.edu.cn (H.Y.)

Materials and Methods

1. Data source

1.1 Epidemiological SARS-CoV-2 data

We collected data on 1,178 confirmed SARS-CoV-2 infections in Hunan Province, China, from January 16 to April 2, 2020, following the protocol for field epidemiological investigation developed by the National Health Commission of the People's Republic of China to identify potential COVID-19 cases (*1*). SARS-CoV-2 infections were identified by a combination of traffic entrance and community screening in high-risk populations who had a history of traveling to Wuhan City/Hubei Province, which captured travel-associated cases; passive surveillance in hospitals and outpatient practices, which captured symptomatic cases; and systematic monitoring of contacts of confirmed cases, which captured symptomatic and asymptomatic infections. All SARS-CoV-2 positive individuals in this database received positive laboratory confirmation of SARS-CoV-2 infection by RT-PCR test. Before February 7, 2020, contacts were tested if they developed symptoms during the quarantine period. After February 7, specimens were collected at least once from each contact during quarantine, regardless of symptoms. The information collected for each case includes age, sex, prefecture (of case being reported), clinical severity (asymptomatic, mild, moderate, severe, or critical, see Table S1 for definition), potential exposures (travel history to Wuhan or contact with confirmed SARS-CoV-2 infection), time windows of potential exposures, date of the start of isolation/pre-symptomatic quarantine, date of symptom onset (list of symptoms below), date of healthcare consultation, date of hospital admission and ICU admission (if applicable), and date of laboratory-confirmation. The list of symptoms observed and documented among all patients includes: *fever (57.7%), dry cough (36.4%), fatigue (23.9%), sputum (19.6%), headache (10.3%), muscle ache (8.6%), sore throat (7.8%), chills (7.6%), chest tightness (5.4%), diarrhea (5.2%), shortness of breath (5%), runny nose (4.2%), stuffy nose (4.2%), vomiting (2.2%), joint pain (2.0 %), nausea (1.9%), difficulty in breathing (1.4%), chest pain (1.3%), abdominal pain (0.5%), conjunctival hyperemia (0.3%)*. All epidemiological information and testing data were collected by the Hunan CDC staff or by trained local CDC personnel and entered in a systematic database.

Table S1. Definitions of clinical severity of SARS-CoV-2 infections

Clinical severity	Definition
Asymptomatic	SARS-CoV-2 positive individuals who do not show any symptoms throughout the course of infection.
Mild	Patients with mild symptoms and no radiographic evidence of pneumonia
Moderate	Patients with fever, or respiratory symptoms, and radiographic evidence of pneumonia
Severe	Patients who have 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 as severe case.
Critical	Patients who have any of the following: a. respiratory failure that requires mechanical ventilation; or b. shock; or c. other organ failures that requires ICU admission.

Dates of key events in the course of exposure and infection: For each SARS-CoV-2 positive individual in the database, information is compiled on the start/end date of exposure, along with the date of symptom onset (for symptomatic individuals) and laboratory confirmation. Biologically, the time of infection should occur before the onset of symptom or a positive RT-PCR test. Thus, we update the patient's end date of putative exposures in the database as the earliest of the reported exposure end date, date of symptom onset, or date of laboratory confirmation. If the start date of exposure is later than the date of symptom onset or positive RT-PCR test, it likely reflects recall error and we update the exposure start date as missing (1.9% of the records).

1.2 Contact tracing database

We collected data on 15,648 individuals in close contact with the 1,178 confirmed SARS-CoV-2 infections identified in Hunan Province, China, representing 19,227 unique exposure events following the national protocol (*1*). Information included age, and sex of the contacts, type of contacts (household, extended family, social, community, and healthcare, see Table S2 for definition), as well as the start and end dates of contact exposure. If the contact was confirmed with SARS-CoV-2 by RT-PCR, a unique identifier mapping the individual to the SARS-CoV-2 patient database was provided.

Table S2: Definition of contact types.

Contact Type	Definition
Household	A household member living with a SARS-CoV-2 infected individual.
Extended family	A family member not residing in the same household but who has been in close contact with the primary SARS-CoV-2 infected individual.
Social	Friends, coworkers and classmates who study, work or are in close contact with the primary infected individual.
Community	Staff who interact with SARS-CoV-2-infected individuals in restaurants, entertainment venues, or other service settings; passengers seated in close proximity to a SARS-CoV-2 infected individual.
Healthcare	Healthcare workers who provide diagnosis, treat or nurse a SARS-CoV-2 patient or other patients and caregivers in the same ward as a SARS-CoV-2 infected individual.

Any individual reporting encounters as described in Table S2 and occurring within <1m of a SARS-CoV-2 infected individual (irrespective of displaying symptom) was considered a close contact, at risk of SARS-CoV-2 infection. All records were extracted from the electronic database managed by Hunan Provincial Center for Disease Control and Prevention. All individual records were anonymized and de-identified before analysis.

1.3 Definition of a SARS-CoV-2 cluster

Based on the contact tracing database, we define a SARS-CoV-2 cluster as a group of two or more confirmed SARS-CoV-2 cases or asymptomatic infections with an epidemiological link, i.e. occurring in the same setting (e.g. home, work, community, healthcare, or other) and for which a direct contact between successive cases can be established within two weeks of symptom onset of the most recent case (alternatively, the date of RT-PCR test for asymptomatic infections). In total, there are 210 clusters recorded in the database, for a total of 831 SARS-COV-2 infections.

While clusters of cases are grouped together based on shared exposures, a subset of cases report additional exposures outside the cluster as possible causes of infection as well. As a result, there can be more than one index case within each cluster. In addition, for cases that only report exposures within the cluster, a unique infector cannot always be identified, given simultaneous SARS-CoV-2 exposures within the same cluster.

A sporadic case is defined as a laboratory-confirmed SARS-CoV-2 individual who does not belong to any of the reported clusters (i.e. a singleton who has no epidemiological link to other infections identified). In total, there are 347 sporadic cases recorded in the database.

Since the source and direction of transmission within a cluster cannot always be defined based on epidemiological grounds alone, we next turn to a modeling approach to probabilistically reconstruct possible infector-infectee transmission chains and further evaluate predictors of transmission.

2. Reconstruction of SARS-CoV-2 transmission chains

2.1 Sampling algorithm

For each cluster and each patient i in the cluster, the time of infection t_i^{inf} is stochastically sampled by randomly drawing from the incubation period distribution and subtracting this value from the reported time of symptom onset, i.e. $t_i^{inf} = t_i^{sym} - \tau_i^{incu}$, where τ_i^{incu} is the sampled incubation period and t_i^{sym} the date of symptom onset (*2*). The incubation period follows a Weibull distribution:

$$g_{incu}(\tau) = \frac{k}{\lambda}\left(\frac{\tau}{\lambda}\right)^{k-1} exp\left(-\left(\frac{\tau}{\lambda}\right)^{k}\right)$$

with shape parameter $k = 1.58$ and scale parameter $\lambda = 7.11$. The median incubation period is taken to be 5.56 days with IQR (3.14, 8.81) days (*2*).

The sampled time of infection t_i^{inf} must satisfy the following constrains:

- t_i^{inf} must fall within the start and end dates of the exposures identified by epidemiological investigation.
- For any infector-infectee pair, the time of infection of the infector $t_{infector}^{inf}$ must be earlier than the time of infection of the infectee $t_{infectee}^{inf}$, i.e. $t_{infector}^{inf} < t_{infectee}^{inf}$.

A SARS-CoV-2 infected individual may have multiple exposures (either through contacts with multiple SARS-CoV-2 infected individuals, or travel history to Wuhan in addition to contact with a SARS-CoV-2 individual). For an individual *i* who has multiple sources of exposure with a cluster, all other cases in contact with *i* are potential sources of infection, except for those whom *i* has infected. If the sampled infection time of infectee *i,* t_i^{inf}, satisfies the constraints of multiple exposures, we randomly choose one as the source of infection. If t_i^{inf} satisfies the constrains of none of the plausible exposures, we resample t_i^{inf} until individual *i* has one and only one valid source of infection. For individuals with missing onset dates (including all asymptomatic individuals), we set the time of infection as missing. The source of infection is then randomly chosen from all plausible exposures identified from epidemiological investigation.

We stochastically reconstruct 100 realizations of transmission chains to account for uncertainties in both the timing and source of exposures. 375 of the 831 (45%) SARS-CoV-2 infections do not have unique epidemiological link and their transmission routes may vary from one realization to another.

We remove all singletons from the reconstruction of transmission chains, since they are not epidemiologically linked to other cases, but we consider these singletons when we analyze the distribution of secondary cases and when we represent the transmission network in Fig. 1.

2.2 Distribution of the number of secondary infections among transmission chains

Next, we calculate the number of secondary infections for each of the 1,178 SARS-CoV-2 individuals based on the 100 reconstructed transmission chains among 831 cluster cases, and the 347 singletons. The distribution of secondary infections is shown in Fig. 1 We fit a negative binomial distribution to these data using package "pystan" version v2.19.1.1 with uniform prior. We estimated mean $\mu = 0.40$, 95% CI 0.39 to 0.41 and dispersion parameter $k = 0.30$, 95%CI 0.29 to 0.30.

3. Kinetics of SARS-CoV-2 transmission

3.1 Generation interval and serial interval distribution

The generation interval is defined as the time interval between the dates of infections in the infector and the infectee. We calculate the generation intervals of all the infector-infectee pairs based on 100 realization of the reconstructed transmission chains. The distribution of the generation interval is shown in Fig. S2A. The observed serial interval is defined as the time interval between dates of symptom onsets in the infector and the infectee. We calculate the serial interval of all the infector-infectee pairs based on 100 realizations of the reconstructed transmission chains with known dates of symptom onset. The distribution of the serial interval is shown in Fig. S2B.

3.2 Gauging the impact of case isolation on the distribution of the serial and generation intervals.

We select all infector-infectee pairs for which the infector has been isolated during the course of his/her infection, date of symptom onset is available, and times of infection have been estimated. We stratify the data by the infector's time interval between onset and isolation, τ_{iso}, with $\tau_{iso} \in \{(-\infty, 2), [2, 4), [4, 6), [6, +\infty)\ days\}$, and assess how the generation interval and serial interval distributions change with the speed of case isolation (Fig. 3A and Fig. 3B).

3.3 Speed of case isolation and relative contribution of pre-symptomatic transmission.

As cases are isolated earlier in the course of infection, we expect that the contribution of pre-symptomatic transmission will increase. This is because symptomatic transmission occurs after pre-symptomatic transmission and transmission will be blocked after effective isolation. In other words, isolated individuals remain infectious, but they can only effectively transmit before isolation, which is predominantly in their symptomatic phase. To validate the hypothesis that the contribution of pre-symptomatic transmission is affected by interventions, we first estimate the overall contribution of pre-symptomatic transmission among all reconstructed transmission chains. Let $tr_{i,j}^{k}$ represent each transmission event from an infector to infectee i, in realization j of the 100 sampled transmission chains; $k = 0$ indicates that infection in an infectee occurred before the time of symptom onset of his/her infector, denoting pre-symptomatic transmission, while $k = 1$ indicates that the time of infection occurred after the infector's symptom onset (i.e. post-symptomatic transmission). Thus, the overall fraction of pre-symptomatic transmission in realization j can be calculated using the following formula:

$$P_j^{pre} = \frac{\sum_i tr_{i,j}^{k}}{\sum_i \sum_k tr_{i,j}^{k}}$$

Mean and 95% CI of P^{pre} can be estimated over the 100 realizations of the reconstructed transmission chains. We further stratify P^{pre} by the time interval between an infector's symptom onset and isolation, considering four categories (days):

$$(-\infty, 0),\ [0,2),\ [2,4),\ [4,6),\ [6, +\infty)$$

The mean and variance (based on 100 realization of the sampled transmission chains) of P^{pre} for each category of the isolation intervals is shown in Fig. 3C.

3.3 Relative infectiousness profiles over time adjusted for case isolation.

In Hunan province, all COVID-19 cases regardless of clinical severity were managed under medical isolation in appointed hospitals; while contacts of SARS-CoV-2 infections were quarantined in designated medical observation centers. In Section 4, we estimate that the risk of transmission in the healthcare setting is the lowest among all contact settings, thus case isolation and contact quarantine are highly effective to block onward transmission after isolation/quarantine. As a result, the observed serial/generation intervals are shorter than they would be in the absence of case isolation and contact quarantine. The censoring effects are clearly demonstrated in Fig. 3A and Fig. 3B, where we observe that the median generation time drops from 7.0 days for $\tau_{iso} > 6\ (days)$ after symptom onset, to 4.1 days for $\tau_{iso} < 2\ (days)$.

Moreover, the speed of case isolation is not static over time. Fig. S4 shows the distribution of time from symptom onset to isolation in three different phases of epidemic control (*Phase I, II, and III*) defined by two major changes in COVID-19 case definition issued by National Health Commission on Jan. 27 and Feb. 4. The median time from symptom onset to isolation decreases from 5.4 days in *Phase I* to -0.1 days in *Phase III*, due to the expansion of "suspected" case definition (*3*) and strengthening of contact tracing effort (Fig. S3).

3.3.1 Generation interval adjusted for case isolation.

Estimating the generation interval distribution in the absence of interventions is important to understand the kinetics of SARS-CoV-2 transmission, as the shape of the generation interval distribution represents the population-average infectiousness profile since the time of infection. To minimize the potential error of flipping the directionality of infector-infectee relationship during contact tracing, we further limit our analysis to the infector-infectee pairs where the primary case had a travel history to Wuhan (and no other SARS-CoV-2 contact), while the secondary case did not have a travel history to Wuhan but was epidemiological linked to the primary case. To further reduce potential recall bias on the timing of symptom onset/exposure, we down-sample the outlier incubation periods. To do this in a statistically sound manner, we rely on the independence of the incubation periods of the infector and the infectee, and down-sample infector-infectee pairs whose joint likelihood of the observed incubation period pair is very low. Specifically, we first estimate the joint empirical distribution of the incubation periods of both the infector and infectee using the gaussian kernel density estimate (*4*) in the package "scipy" version v1.5.0 function "scipy.stats.gaussian_kde" (*5*). The joint likelihood of observing the incubation periods of a given infector-infectee pair based on the kernel density estimate is denoted as $p_{kde}(\tau_i^{incu}, \tau_j^{incu})$. The joint likelihood of the incubation period of the same infector-infectee pairs based on two independent draws from the Weibull distribution $g_{incu}(\tau) = \frac{k}{\lambda}\left(\frac{\tau}{\lambda}\right)^{k-1} exp\left(-\left(\frac{\tau}{\lambda}\right)^{k}\right)$ with shape parameter $k = 1.58$ and scale parameter $\lambda = 7.11$ (Section 2.1) is denoted as $p_E(\tau_i^{incu}, \tau_j^{incu})$. If $p_{kde}(\tau_i^{incu}, \tau_j^{incu}) > p_E(\tau_i^{incu}, \tau_j^{incu})$, it suggests the observed incubation periods are over-

represented relative to expectations, and vice versa. We introduce a down-sampling weight in accordance with the incubation period distribution as $w_{incu} = p_E(\tau_i^{incu}, \tau_j^{incu})/p_{kde}(\tau_i^{incu}, \tau_j^{incu})$.

To account for the "censoring" of generation interval distribution due to quarantine/case isolation, we first exclude generation intervals where transmission occurred after isolation of the infector (only 4.3% of the reconstructed transmission events, attesting to the effectiveness of isolation). We then divide the generation intervals into three groups based whether the date of symptom onset of the infectors fall within a given phase of epidemic control in Hunan. In Group 1 the illness onset of the infectors occurred before Jan. 27th (*Phase I*); in Group 2 the illness onset of the infector occurred between Jan. 27th and Feb. 4th (*Phase II*); in Group 3, the illness onset of the infector occurred after Feb. 4th (*Phase III*). For a given generation interval τ_{GI} of an infector-infectee pair in each group, we denote:

- The time of symptom onset of the infector as t_{onset}.
- The time of case isolation/quarantine of the infector as t_{iso}.
- The time of transmission from the infector to the infectee as $t_{inf.}$.
- The time interval between onset of the infector and transmission to the infectee $\tau_{oi} = t_{onset} - t_{inf.}$
- The time interval between infection times in the infector and infectee, i.e. the generation interval τ_{ii}
- The probability distribution from symptom onset to isolation as $P_i(\tau_{iso})$, where $i \in \{I, II, III\}$ denotes the different phases of epidemic control, determined by symptom onset in the infector t_{onset}. The functional form of $P_i(\tau_{iso})$ is shown in Fig. S4. The corresponding cumulative probability distribution is denoted as $C_{OI}^i(\tau_{iso})$.

The probability of this infection-infectee pair escaping the "censoring" due to quarantine and case isolation is $p_i^{esc.} = 1 - C_{OI}^i(\tau_{oi})$. For every n observations of the generation interval τ_{ii} under intervention $p_i(\tau_{iso})$ given τ_{oi}, there should be $m = \frac{n}{p_i^{esc.}}$ observations of τ_{GI} given τ_{oi} without intervention $p_i(\tau_{iso})$. Thus, we denote the sampling weight adjusted for case isolation as $w_{iso} = \frac{1}{p_i^{esc.}}$. The overall resampling weight of generation interval τ_{ii} between infector i and infectee j considering both incubation period distribution and censoring due to case isolation is given by $w_{sample}(i,j) = w_{incu} \times w_{iso} = \frac{p_E(\tau_i^{incu}, \tau_j^{incu})}{p_i^{esc.} \times p_{kde}(\tau_i^{incu}, \tau_j^{incu})}$. We resample from $\{\tau_{ii}(i,j)\}$ with sampling weights $w_{sample}(i,j)$ until we reach a sample size of $n = 10000$ to obtain the distribution of generation time $\{\tau_{ii}^{adj.}\}$ adjusted for censoring. The distribution of $\tau_{ii}^{adj.}$ reflects the generation interval that would have been observed in the absence of quarantine and case isolation/quarantine. We fit Weibull, gamma, and lognormal function to $\{\tau_{ii}^{adj.}\}$. The distribution of $\tau_{ii}^{adj.}$ is best described by the Weibull distribution:

$$g_{GI}^{adj.}(\tau) = \frac{k}{\lambda}\left(\frac{\tau}{\lambda}\right)^{k-1} exp\left(-\left(\frac{\tau}{\lambda}\right)^k\right)$$

with $k = 1.3$ and $\lambda = 9.61$ (Fig. S6A).

3.3.2 Distribution of time interval between symptom onset and transmission, adjusted for case isolation.

In contrast to the generation interval distribution, which characterize the relative infectiousness of a SARS-CoV-2 infection over time with respect to the time of infection, we now focus on the interval between symptom onset and transmission. This shifts the reference point of the infectiousness profile from the time of infection to the time of symptom onset. Namely the distribution of symptom onset to transmission adjusted for case isolation $\{\tau_{OT}^{adj.}\}$ represents the population-average relative infectiousness profile over time since the onset of symptom. Of note, since we observe substantial pre-symptomatic transmission for SARS-CoV-2, negative values of $\tau_{OT}^{adj.}$ are allowed.

Similarly to the previous section, we resample from $\{\tau_{OT}(i,j)\}$ with sampling weights $w_{sample}(i,j)$ until a sample of size $n = 10000$ is reached to obtain the distribution of symptom onset to transmission $\{\tau_{OT}^{adj.}\}$. The resampled distribution represents the infector's relative infectiousness (population average) with respect to the infector's symptom onset (Fig. S6B). The best-fit distribution is a normal distribution:

$$f_{OT}^{adj.}(\tau_{OT}^{adj.}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\tau_{OT}^{adj.} - \mu}{\sigma}\right)^2}$$

with mean $\mu = -0.22$ days and standard deviation $\sigma = 3.47$ days. After adjusting for case isolation, the fraction of transmission occurring during the pre-symptomatic phase of SARS-CoV-2 infection is 52%.

3.4 Estimating the basic reproduction number in Wuhan before lockdown

A recent study (*6*) estimated the initial growth rate of the epidemic in Wuhan at 0.15 day^{-1} 95% CI (95% CI, 0.14 to 0.17) ahead of the lockdown. The estimate is based on the daily rise in reported cases by onset date; adjustment for increased reporting due to a broadening case definition places the growth rate at 0.08 day^{-1} (*6*). The Euler–Lotka equation (*7*) describes the relationship between the basic reproduction number R_0, the epidemic growth rate r, and the generation interval distribution $g(\tau)$:

$$R_0 = \frac{1}{\int exp(-r * \tau) \times g(\tau) d\tau}$$

We assume that no effective intervention had been implemented in Wuhan by the time of the lockdown (Jan. 23). Using the generation time distribution adjusted for "censoring" due to quarantine and case isolation $g_{GI}^{adj.}(\tau)$ described in the previous section, we estimate the basic reproduction number in Wuhan during the exponential growth phase at $R_0^{Wuhan} = 2.23$, (95% CI, 2.13 to 2.43), based on the conservatively higher estimate of growth rate in this city (0.15 day^{-1} 95% CI (95% CI, 0.14 to 0.17). If we rescale the adjusted generation time distribution $g_{GI}^{adj.}(\tau)$ by a factor of R_0^{Wuhan}, the function

$$r_{GI}(\tau) = g_{GI}^{adj.}(\tau) \times R_0^{Wuhan}$$

represents the average risk of SARS-CoV-2 transmission to a secondary case at time τ since infection. The red line in Fig. 3D visualizes the functional form of $r_{GI}(\tau)$.

Similarly, if we rescale the adjusted distribution of symptom onset to transmission $f_{OT}^{adj.}(\tau)$ with R_0^{Wuhan}, the function

$$r_{OT}(\tau) = f_{OT}^{adj.}(\tau) \times R_0^{Wuhan}$$

represents the average risk of transmission to a secondary case at time τ since the symptom onset of the infector. The red line in Fig. 3E visualizes the functional form of $r_{OT}(\tau)$.

3.5 Evaluating the impact of case isolation and quarantine on SARS-CoV-2 transmission.

To evaluate the impact of quarantine and case isolation on the reduction of SARS-CoV-2 transmission at different phases of epidemic control, we denote the time intervals between a patient's time of infection to his/her time of isolation as τ_{inf}^{iso}. The corresponding probability distribution is $p_{ii}^{j}(\tau)$, where $j \in \{I, II, III\}$ denotes the phase of epidemic control. We denote the distribution of the incubation period τ_{incu} as $p_{incu}(\tau)$ and the distribution of symptom onset to isolation τ_{onset}^{iso} as $p_{oi}^{j}(\tau)$, for each phase j of epidemic control. We sample $\tau_{inf}^{iso} = \tau_{incu} + \tau_{onset}^{iso}$ numerically through independently sampling of τ_{incu} and τ_{onset}^{iso} and add them together. Fig. S5 shows the distribution of 10000 numerical sampling of τ_{inf}^{iso} at different phases of epidemic control. We fit the sampled distribution of τ_{inf}^{iso} to various probability distributions including normal, lognormal, gamma, Cauchy, logistic, and hyperbolic secant distribution. The top three fits are show in Fig. S5 and the best fit is selected based on the Akaike information criterion during each of the three phases of epidemic control. We denote cumulative density distribution of $p_{ii}^{j}(\tau)$ as

$$C_{ii}^{j}(\tau) = \int_{-\infty}^{\tau} P_{ii}^{j}(\tau') d\tau'$$

where $C_{ii}^{j}(\tau)$ gives the probability that transmission is blocked after time τ, where τ is the time since infection. The shaded areas in Fig. S5 visualize the probabilities $C_{ii}^{j}(\tau)$ for the best-fit distribution.

Assuming that all SARS-CoV-2 patients are subject to case isolation and quarantine efforts carried out in Hunan province, we can estimate the average risk of transmission $r_{GI}^{control(j)}(\tau)$ of an infected individual at time τ since his/her infection, during phase $j \in \{I,\ II,\ III\}$ of epidemic control as:

$$r_{GI}^{control(j)}(\tau) = r(\tau) \times \left(1 - C_{ii}^{j}(\tau)\right) = g_{GI}^{adj.}(\tau) \times R_0^{Wuhan} \times \left(1 - C_{ii}^{j}(\tau)\right)$$

The corresponding basic reproduction number assuming 100% SARS-CoV-2 infection detection rate is given by:

$$R_0^{j} = \int_{0}^{\infty} r_{GI}^{control(j)}(\tau)\, d\tau\, , j \in \{I, II, III\}$$

In Fig. 3D, we visualize the transmission profile with respect to infection time $r_{GI}^{control(j)}$ for all three phases of epidemic control (dashed lines) and shows the estimated values of the corresponding basic reproduction number R_0^{j}.

Similarly, following Section 3.3.1, $C_{oi}^{j}(\tau)$ gives the probability that transmission is blocked after time τ since symptom onset in the infector, for the 3 phases of epidemic control $j \in \{I,\ II,\ III\}$. We can estimate the average risk

of transmission $r_{OT}^{control(j)}(\tau)$ of an infected individual at time τ since his/her onset of symptom, during phase $j \in \{I,\ II,\ III\}$ of epidemic control as:

$$r_{OT}^{control(j)}(\tau) = r(\tau) \times \left(1 - C_{oi}^{j}(\tau)\right) = f_{OT}^{adj.}(\tau) \times R_0^{Wuhan} \times \left(1 - C_{oi}^{j}(\tau)\right)$$

In Fig. 3E, we visualize the transmission profile with respect to symptom onset time $r_{OT}^{control(j)}$ for all three phases of epidemic control (dashed lines).

3.6 Evaluating synergistic effects of individual-level and population-level interventions on SARS-CoV-2 transmission.

We start by characterizing the controllability of SARS-CoV-2 (measured as R_0 under control measures) as a function of infection isolation rate and the speed of case isolation/pre-symptomatic quarantine. In Fig. 3F, we plot the phase diagram of R_0 as a function of infection detection proportion (fraction of all SARS-CoV-2 infections detected) and the mean time from symptom onset to isolation/quarantine τ_{iso}. Contour lines indicates reductions in R0 from baseline non-intervention conditions. It is worth noting that we do not know the precise prevalence of truly asymptomatic infections as well as their role in transmission. Here we assume that asymptomatic cases have a similar shape of infectiousness profile over the course of infection as symptomatic cases, and a peak of infectiousness corresponding to the time of symptom onset in symptomatic cases, as shown in Fig. S6. The corresponding τ_{iso} for asymptomatic cases is measured as time from peak infectiousness to isolation. Here we assume that the distribution of symptom onset/peak infectiousness to isolation follows a normal distribution with mean τ_{iso} and standard deviation of 2 days.

We further consider the synergic effects of layering individual-based intervention (case isolation, contact tracing, and quarantine) with population-based interventions (ie, via physical distancing, measured as a reduction in effective contact rates). In Fig. 3G, we plot the phase diagram of R_0 as a function of the proportion of population-level contact reduction and infection isolation rate, with the average speed of isolation 0 days after symptom onset/peak infectiousness and standard deviation of 2 days. The blue area indicates the region below the epidemic threshold, where control is achieved, and the red area indicates region above the epidemic threshold.

Last, we consider a sensitivity analysis with a lower base $R_0 = 1.59$, using the growth rate of $r = 0.08$ observed in Wuhan data with adjustment for changes in reporting (Section 3.4). In Fig. 3F, we plot the phase diagram of R_0 as a function of % population-level contact reduction (i.e. through physical distancing) and isolation rate, assuming that SARS-CoV-2 infections are isolated 2 days after symptom onset/peak infectiousness on average with a standard deviation of 2 days. The blue area indicates the region below the epidemic threshold and the red area indicates region above the epidemic threshold.

4. Evaluating individual-level transmission heterogeneity of SARS-CoV-2

4.1 Regression analysis to evaluate the "per-exposure" risk of SARS-CoV-2 transmission as a function of demographical, epidemiological, clinical, and behavioral predictors.

In this section, we use a mixed effects multiple logistic regression model to evaluate the risk of SARS-CoV-2 transmission for each exposure reported in the contact tracing database. Each entry in the database represents a contact exposure between a SARS-CoV-2 infected individual and his/her contact. For individuals who were in contact with SARS-CoV-2 infected individual, the contact individual's age, sex, type of contact, the start/end dates of exposure, as well as the infection status (whether the exposed individuals was eventually infected with SARS-CoV-2) are carefully documented (Section 1.2). All SARS-CoV-2 infected individual (both primary cases and secondary infections via contact exposures) have unique identifiers that can be mapped to the SARS-CoV-2 patient line-list database, where additional information about the course of infection is also available (see Section 1.1 for detailed information). An individual in the contact tracing database can be exposed to multiple SARS-CoV-2 cases; further, an individual in the contact tracing database can be exposed to the same SARS-CoV-2 case through multiple independent exposures. All exposures are recorded independently.

For each exposure in the contact-tracing database, the regression outcome is coded as 1 if the contact eventually becomes infected and 0 if not infected. For each exposure, a list of independent variables, their definitions, and corresponding values are shown in Table S3 (fixed effects in the mixed model):

Table S3: Fixed effect variables of the mixed effects multiple logistic regression model

Fixed effect	Definition	Category	Counts	%
Age (contact)	Age category of the contact. Age is categorized into three age categories: ***0-12 years***, ***13-25 years, 26-64***	***0-12 years***	1392/14662	9%
		13-25 years	1859/14662	13%

	years, ***65 years and older (+65 years)***; **26-*64 years*** is the reference category.	***26-64 years***	9323/14662	63%
		+65 years	1415/14662	10%
		NA	673/14662	5%
Sex (contact)	Sex of the contact (***male/female***). ***Female*** is the reference category.	***male***	7473/14662	51%
		female	6958/14662	47%
		NA	231/14662	2%
Age (case)	Age category of the case. Age is categorized into three age categories: ***0-12 years***, ***13-25 years, 26-64 years***, ***65 years and older (+65 years)***; **26-*64 years*** is the reference category. For main regression with data imputation (Fig. S1), we merge age brackets of ***0-12 years*** and ***13-25 years*** into ***0-25 years*** as only 3% of data in the ***0-12 years*** bracket. For regression of sensitivity analysis that removes missing data, we keep both ***0-12 years*** and ***13-25 years*** age brackets.	***0-12 years***	27/870	3%
		13-25 years	74/870	8%
		26-64 years	666/870	77%
		+65 years	103/870	12%
		NA	0/870	0%
Sex (case)	Sex of the case (***male/female***). ***Female*** is the reference category.	***male***	454/870	52%
		female	416/870	48%
		NA	0/870	0%
Clinical severity (case)	Clinical severity category of the case. Here we consider three categories: the first category represents ***asymptomatic*** cases; the second represents ***mild & moderate*** cases (reference category) and the third represents ***severe & critical*** cases. A definition of clinical severity is provided in Section 1.1, Table S1.	***asymptomatic***	108/870	12%
		mild	217/870	25%
		moderate	427/870	49%
		severe	94/870	11%
		critical	24/870	3%
		NA	0/870	0%
"Fever" (case)	If the SARS-CoV-2 case had ***"fever (Yes/No)"*** during the course of illness. Cases without ***"fever"*** are the reference class.	***Yes***	524/870	60%
		No	342/870	39%
		NA	4/870	1%
"Dry cough" (case)	If the SARS-CoV-2 case had ***"dry cough (Yes/No)"***. Cases without ***"dry cough"*** are the reference class.	***Yes***	314/870	36%
		No	552/870	63%
		NA	4/870	1%
Travel history Wuhan (case)	If the SARS-CoV-2 case had ***travel history to Wuhan (Yes/No)***: cases without ***travel history to Wuhan*** are the reference category.	***Yes***	356/870	41%
		No	459/870	53%
		NA	55/870	6%
Transmission settings	The type of interactions between a case and a contact: the 6 contact types are ***household***, ***extended family***, ***social***, ***community***, and ***healthcare*** contacts. Definition of the contact types are detailed in Section 1.2 Table S2. For exposures in each setting, we first define the exposure time as the midpoint between the start and end date of the exposure window. The median household exposure time is 01/25, 2020. We further divide the contacts in different settings into two categories: we denote household contact with exposure	***household (pre 01/25)***	964/17750	5%
		household (post 01/25)	924/17750	5%
		extended family (pre 01/25)	3141/17750	18%
		extended family (post 01/25)	2723/17750	15%

	time before 01/25/2020 as ***household (pre 01/25)***; we denote household contact with exposure time after 01/25/2020 as ***household (post 01/25)***. The reference class is ***household (pre 01/25)***. Similarly, we divide contacts in ***extended family***, ***social***, ***community***, and ***healthcare*** settings into: ***extended family (pre/post 01/25), social (pre/post 01/25), community (pre/post 01/25), healthcare (pre/post 01/25).***	***social (pre 01/25)***	2269/17750	13%
		social (post 01/25)	1626/17750	9%
		community (pre 01/25)	3328/17750	13%
		community (post 01/25)	822/17750	5%
		healthcare (pre 01/25)	740/17750	4%
		healthcare (post 01/25)	927/17750	5%
		NA	1313/17750	7%
Duration of exposure	Duration of exposure: defined as the time interval between the start and end date of exposure in days. The duration of exposure is a numeric variable.	***NA***	2045/17750	11%
Onset within exposure	Onset within exposure: defined as if the symptom onset of the primary case occurred within the exposure time window of the contact. (***Yes/No***), ***No*** is the reference class.	***Yes***	9837/17750	56%
		No	4118/17750	23%
		NA	3795/17750	21%

We also introduce random effects for each SARS-CoV-2 case, representing the individual-level infectiousness heterogeneity that is not explained by the independent variables representing fixed effects. These random effects also take into account the lack of independence of our observations.

A contact could report more than one SARS-CoV-2 exposure. If the contact eventually becomes infected, however, only one of the many exposures will be the actual source of infection. In this case, if we denote the number of exposures as n^{expo}, for each of the contact's n^{expo} exposure entries in the database with two different outcomes, either the contact became infected (1 as regression outcome) with regression weight $1/n^{expo}$ or the contact avoided infection from the same exposure (0 as regression outcome) with regression weight $(n^{expo} - 1)/n^{expo}$. We remove contacts who become infected but also have travel history to Wuhan (81/15646, <1%), as the infection could possibility originate from exposures in Wuhan in addition to exposure to local cases in Hunan.

A fraction of the regression variables has missing values in the contact-tracing database (see Table S3, column 3). We adopted the state-of-the-art "Multivariate Imputation by Chained Equations" algorithm (*8*) (implemented in R package "MICE" version 3.9.1 https://cran.r-project.org/web/packages/mice/index.html) to impute missing values in the database. All independent variables in Table S3 are used as predictors for data imputation. The number of multiple imputations is set as 10 with each imputation running 10 realizations. For each of the 5 realizations of imputed contact-tracing databases, we independently perform mixed effects multiple logistic regression of the risk of SARS-CoV-2 transmission with all exposures and variables described in Table S3 as covariates. The regression is performed using R package "lme4" (*9*) version v1.1-23 function "glmer" (https://cran.r-project.org/web/packages/lme4/index.html). The final odds-ratio estimates are pooled from the 5 independent regressions on 5 imputed databases using "MICE" package's "pool" function, based on Rubin's rule (*8*). The odds ratios of independent variables, their 95%CIs, and the baseline odds (intercept) are reported in Fig. S1A.

To examine the model's fit to the data, we explore (i) how well the model reproduces the age profiles of infector-infectee pairs and (ii) whether the model captures the amount of transmission that occurs in different settings (household, family, transportation, etc). We first randomly choose one of the five imputed contact-tracing databases. For each exposure entry in the imputed contact-tracing database, we calculate the model predicted risk of infection based on all fixed variables in the regression. We simulate the infection status of the contact according to the predicted risk by drawing from a binomial distribution. We repeat the process for all contacts, and further simulate 100 realizations of projected infection databases to gauge variability. Fig. S1C shows the observed age distribution of the infector-infectee pairs in the original data, and Fig. S1B visualizes the projected age distribution based on the regression model, averaged over 100 realizations. Violin plots in Fig. S1D show the relative fraction (with projection uncertainties) of transmission that is explained by each type of contacts, based on the model, while the dots in Fig.

S1D represents the empirical observations. We find that the model accurately captures the strong assertiveness of transmission in the 30-50 years age group, and the off diagonals that represent transmission between different generations. Further, the model reproduces the relative contribution of different types of contacts seen in the empirical data (Fig. S1D).

As a sensitivity analysis of imputing missing data (especially addressing the issue of imputing "onset within exposure" for SARS-CoV-2 infected individuals that are asymptomatic), we perform a GLMM-logit regression with entries of missing data removed. We've further break-up the age bracket of predictor "Age (case)" (Table S3) into ***0-12 years***, ***12-25 years***, ***26-64 years***, and ***65+ years***. We remove the predictor of "onset within exposure", however, for predictor "clinical severity (case)", we break down the category "mild & moderate", and "severe & critical" based on the whether onset of the primary case occurred within the exposure time window. "(-)" indicate symptom onset outside the exposure time window, while "(+)" indicate symptom onset within the exposure time window. The results of the regression are shown in Fig. S8.

4.2 Regression analysis evaluating predictors of individual contact patterns among SARS-CoV-2 cases and the impact of interventions

While Section 4.1 addresses predictors of "per-contact" transmission risk heterogeneity, in this section we aim to characterize variation in individual contact patterns of SARS-CoV-2 cases by type of contact. We are particularly interested in the impact of both individual-based and population-based intervention on contact rates. Intuitively, the overall transmission rate of an infectious individual can be interpreted as the sum of contact rates across contact categories weighted by the "per-contact" transmission risk. Thus, conditioning on all other predictors, higher contact rates would translate to higher transmission rates.

We use regression analysis to model the individual contact patterns of each symptomatic SARS-CoV-2 case, whose contacts are traced and documented in the contact-tracing database. We focus on symptomatic cases (the majority of our data) because we are particularly interested in contacts near the time of symptom onset, since we have previously shown that transmission risk is highest near symptom onset. We first define a time window $\tau_{symp.}$ of peak infectiousness as ± 5 days before and after each case's symptom onset $t_{symp.}$. This time window accounts for a majority (86%) of the total infection risk of a typical symptomatic SARS-CoV-2 infection (Fig. S6B). In addition, we consider the 4 main contact types separately: community, social, family, and household contacts. For each symptomatic SARS-CoV-2 case and contact type s, we denote the number of contacts on day i as k_i^s. Here each contact in k_i^s is weighted by the regression odds ratios of GLMM-logit, excluding effects from duration of exposure and if onset is within exposure time window. The cumulative daily contact rate $CCR_{\tau_{symp.}}^s$ within the time window $\tau_{symp.}$ for a given case is given by:

$$CCR_{\tau_{symp.}}^s = \sum_{i=t_{symp.}-5}^{t_{symp.}+5} k_i^s \times w(t_{symp.} - i)$$

Here $w(\tau)$ is the infectiousness profile with respect to symptom onset (Fig. S6B). Clearly, case isolation will impact an infected individual's contact rate, irrespective of whether the case is symptomatic. However here we restrict our analysis to symptomatic cases as the speed of case isolation and pre-symptomatic quarantine can be quantitatively measured as the time from isolation/pre-symptomatic quarantine to symptom onset.

To quantify the impact of socio-demographic factors and interventions on $CCR_{\tau_{symp.}}^s$, we consider a negative binomial regression with $CCR_{\tau_{symp.}}^s$ as the dependent variable and proxies of interventions intensities as independent variables in the regression. Specifically, we use a within-city mobility index as a proxy for the intensity of population-level social distancing, while we use time between isolation and symptom onset to measure the intensity of individual-level interventions (here, case isolation). We also include demographic and clinical predictors as independent variables to adjust for age and sex differences, as well as other changes in contact patterns. A full description of all regression variables is shown in Table S4:

Table S4: Variables of the negative binomial regression on cumulative contact rates.

Independent variable	Definition
Age (categorical)	Age category of the SARS-CoV-2 case. We consider three age categories: ***0-18 years***, ***19-64 years***, ***65 years and older***; ***19-64 years*** is the reference category.
Sex (Male/Female)	Sex of the contact (***male/female***). ***Female*** is the reference category.

Symptom fever (Y/N)	Whether the SARS-CoV-2 case had ***"fever"***. Cases without ***"fever"*** are the reference class.
Symptom dry cough (Y/N)	Whether the SARS-CoV-2 case had ***"dry cough"***. Cases without symptom ***"dry cough"***, i.e. Dry Cough (N), is the reference class.
Travel history to Wuhan (Y/N)	If the SARS-CoV-2 case reported a ***travel history to Wuhan***: cases without ***travel history to Wuhan*** are the reference category.
Physical distancing (Before/After Jan. 25)	Based on the within-city mobility index (Fig. S1A, insert) provided by Baidu Qianxi (*10*), we grouped the individual patients into categories depending on whether the patients symptom onsets occurred before and after January 25, 2020, corresponding to weak/strong physical distancing. Onsets occurred ***before Jan. 25*** (weak physical distancing) is the reference class.
Isolation to onset (days)	Time from case isolation to symptom onset. This is used as a proxy for individual-level intervention intensity. The larger the value, the earlier the case is being isolated. Positive values indicate isolation before symptom onset, negative values indicate isolation after symptom onset.

The regression is performed using the R package "MASS" (*11*) version 7.3-51.6 function "glm.nb" (https://cran.r-project.org/web/packages/MASS/index.html). The point estimates rate ratios along with their 95% CIs for each of the variables are presented in the bottom panels of Fig. 2E. We identify an effect of interventions on contact rates, along with clinical factors; these effects tend to be most intense in the social and transportation settings.

Supplementary Figures

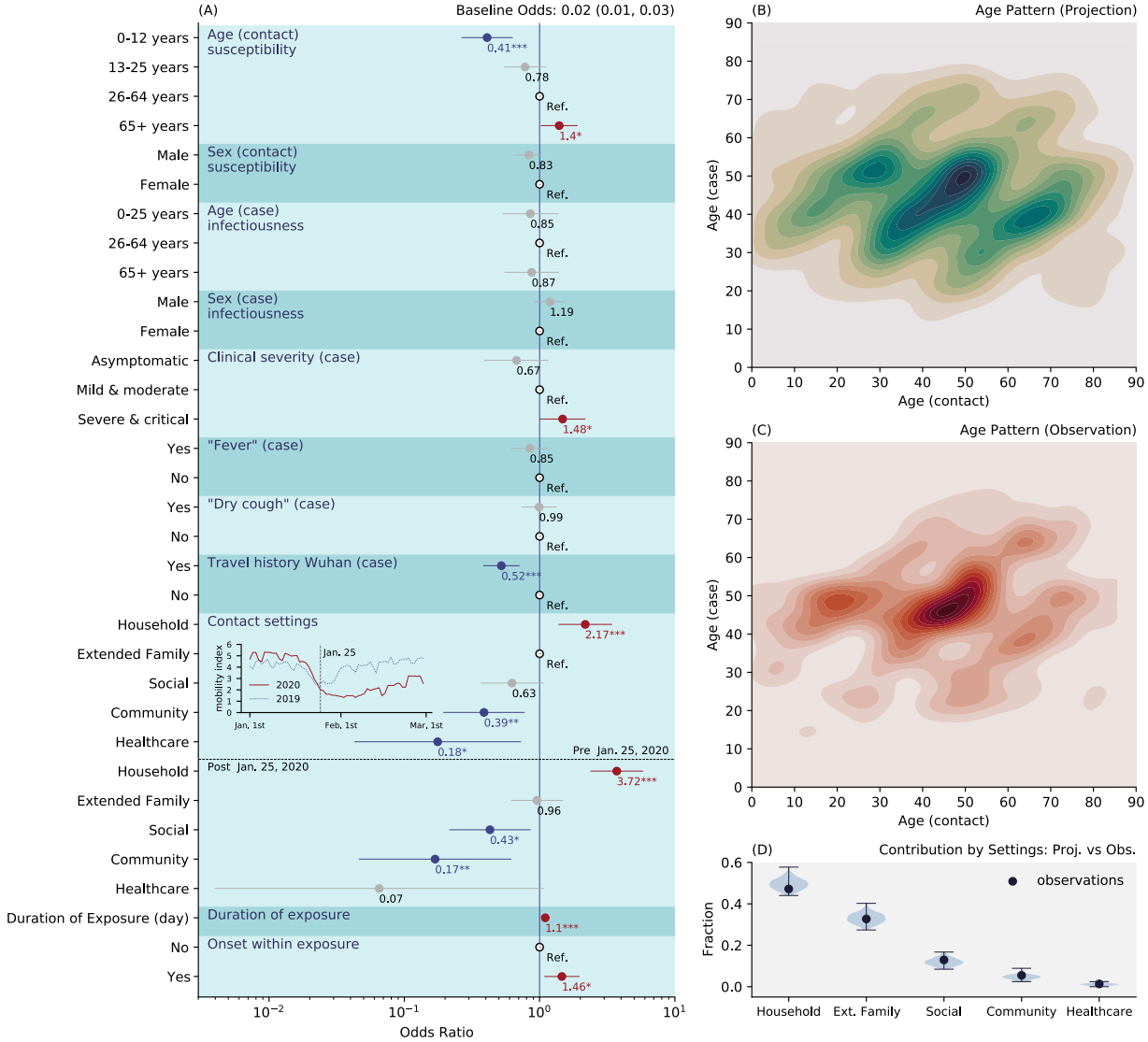

Fig. S1. (A) Individual predictors of transmission risk among close contacts of SARS-CoV-2 infected indivduals in Hunan. The predictors of the logistic regression as those indicated on the left (fixed effects) and we also include random effects for individual SARS-CoV-2 infections. Dots and lines indicate point estimates and 95% confidence interval of the odds ratio, numbers below the dots indicate the numerical value of the point estimates; "Ref." stands for reference category; * indicates p-value<0.05, ** indicates p-value<0.01, *** indicates p-value<0.001. Top inset indicates the within city mobility index in Changsha, Hunan for year 2020 and 2019, provide by Baidu Qianxi (*10*); dashed line indicates January 25, 2020. (B) Age distribution of projected infector-infectee pairs based on the regression model (average over 100 ensemble projections). (C) Age distribution of observed infector-infectee pairs. (D) The contribution of household, family, social, community, and healthcare contacts to transmission. Dots represent empricial observations and violin plots represents model estimates based on 100 ensemble projections.

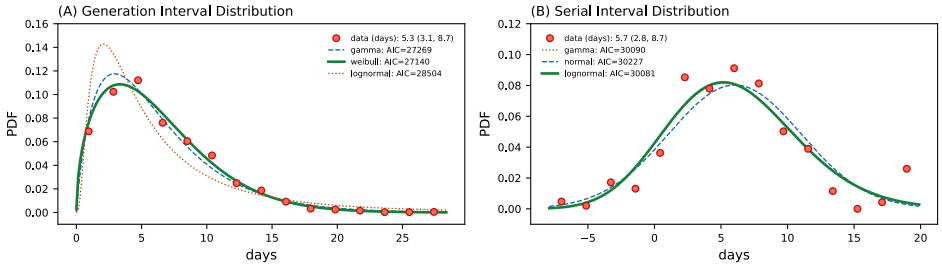

Fig. S2. (A) Distribution of generation interval (time interval between the infection of an infector and his/her infectee's). The distribution is calculated based on 100 realization of plausible transmission chains reconstructed based on the contact tracing database. The estimated generation intervals have a median of 5.3 days with IQR (3.1, 8.7). Grey bars represent data; solid lines represent the Weibull distribution fitted to the data (best fit based on AIC score); dashed lines represent lognormal and gamma distributions fitted to the data. (B) The distribution of serial interval (interval between the symptom onset time of an infector and his/her infectee's). As with the generation interval, the serial interval distribution is calculated over 100 realization of the plausible transmission chains reconstructed through the contact database. The estimated serial interval distribution has a median of 5.7 days with IQR (2.8, 8.7) days. Red dots represent data; solid lines represent the lognormal distribution fitted to the data (best fit based on AIC score); dashed lines represent normal and gamma distributions fitted to the data.

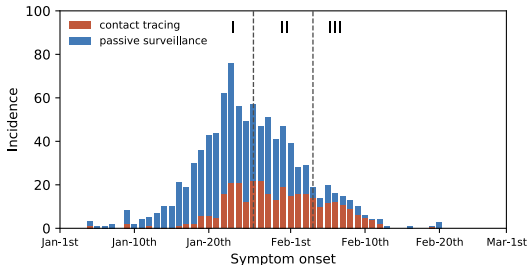

Fig. S3. Incidence of SARS-CoV-2 infections by onset date, for cases captured through contact tracing (red) or passive surveillance (blue). The dashes lines indicate the *Phase I, II, and III* of epidemic control.

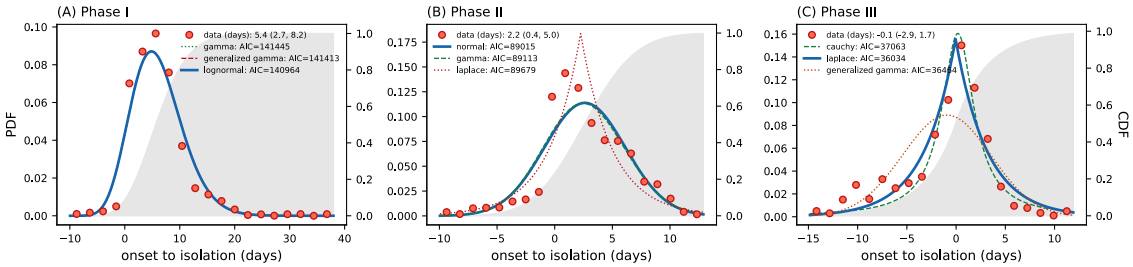

Fig. S4. Distribution of time from symptom onset to isolation in three different phases of epidemic control. Dots are data estimated from the transmission chains, lines are candidate distributions fitted to the data and solid lines distributions fitted best to the data based on Akaike information criterion (A) *Phase I* of epidemic control (before Jan. 27): time from onset to isolation has a median of 5.4 days with IQR (2.7, 8.2) days. (B) *Phase II* of epidemic control (Jan. 27 – Feb. 4): time from onset to isolation distribution has a median of 2.2 days with IQR (0.4, 5.0) days. (C)

Phase III of epidemic control (after Feb. 4): time from onset to isolation distribution has a median of -0.1 days with IQR (-2.9, 1.7) days.

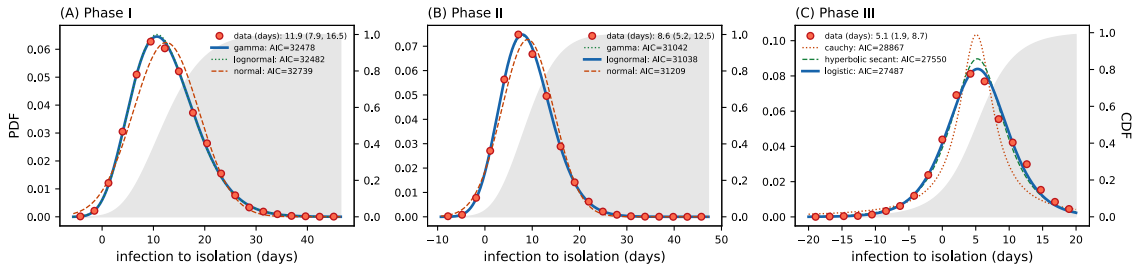

Fig. S5. Same as S3 but focusing on the time from infection to isolation. (A) Phase I of epidemic control (before Jan. 27): time from infection to isolation has a median of 11.9 days with IQR (7.9, 16.5) days (B) Phase II of epidemic control (Jan. 27 – Feb. 4): time from infection to isolation distribution has a median of 8.6 days with IQR (5.2, 12.5) days. (C) Phase III of epidemic control (after Feb. 4): time from infection to isolation distribution has a median of 5.1 days with IQR (1.9, 8.7) days.

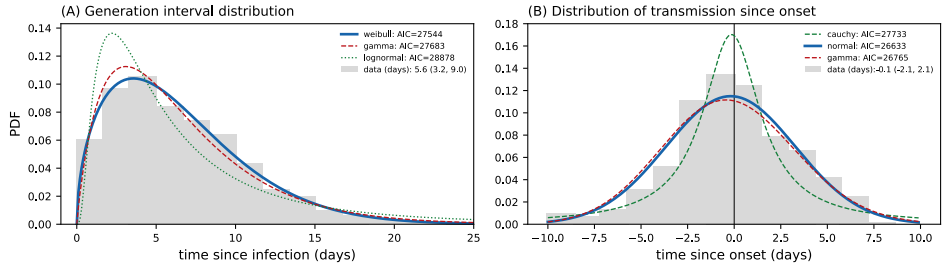

Fig. S6. (A) Distribution of generation interval $\tau_{\mathrm{GI}}^{\mathrm{adj.}}$ adjusted for censoring due to case isolation and quarantine. This represents the distribution that would have been observed in the absence of quarantine and case isolation. The distribution of $\tau_{\mathrm{GI}}^{\mathrm{adj.}}$ has a median of 5.6 days with IQR (3.2, 9.0) days. Grey bars represent data; solid lines represent the Weibull distribution fit to the data (best fit based on AIC score); dashed lines represent gamma and lognormal distributions fitted to the data. (B) Distribution of time from symptom onset to transmission $\tau_{OT}^{adj.}$; negative values represent pre-symptomatic transmission. Grey bars represent data; solid lines represent the normal distribution fitted to the data (best fit based on AIC score); dashed lines represent Cauchy and gamma distributions fitted to the data.

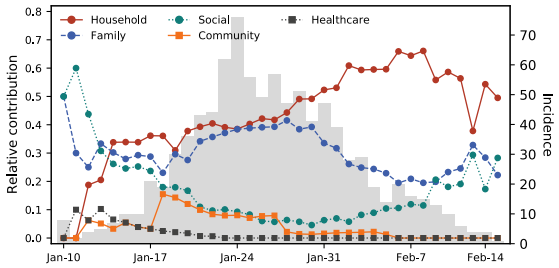

Fig. S7. Trends in the relative contribution of different types of contacts to SARS-CoV-2 transmission. Estimates are averaged over a 10-day moving window. The grey shade in the background indicates the time series of SARS-CoV-2 incidence in Hunan, China.

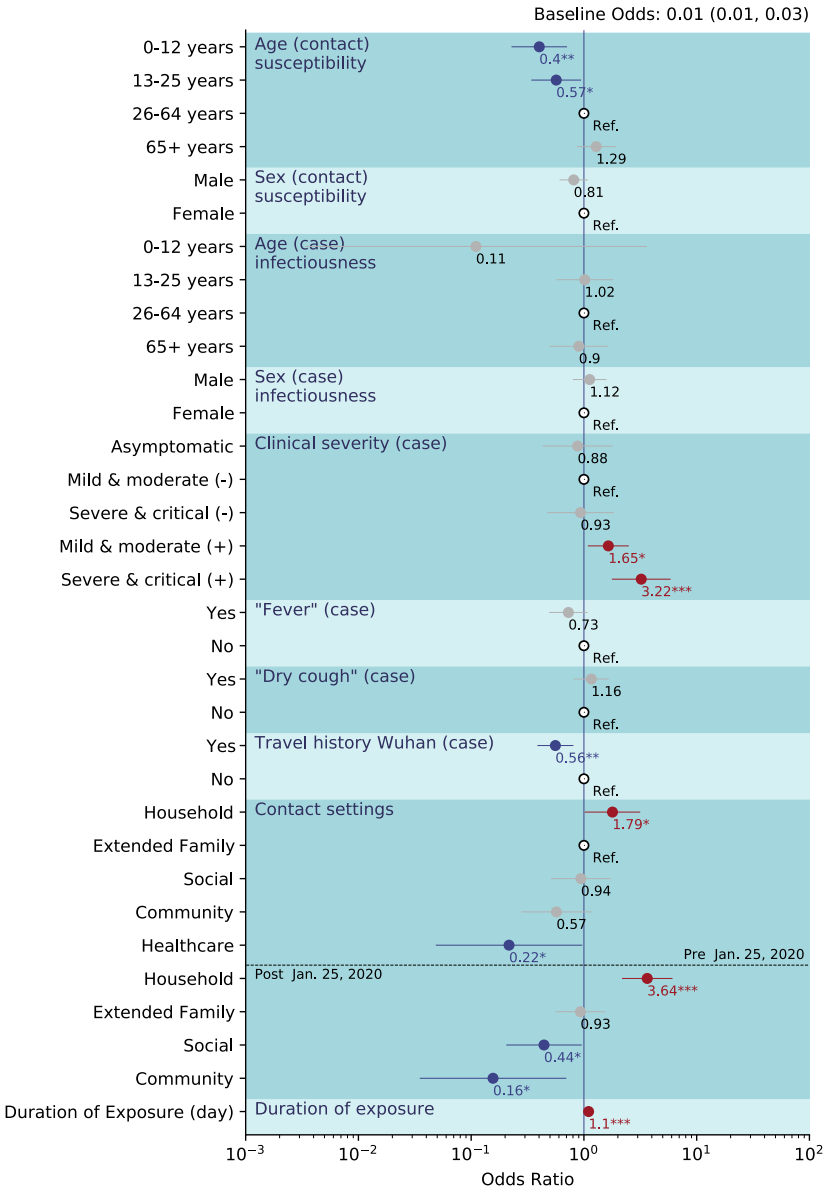

Fig. S8. A sensitivity analysis of GLMM-logit regression that removes missing values. The predictors of the logistic regression as those indicated on the left (fixed effects) and we also include random effects for individual SARS-CoV-2 infections. The "(-)" in "Mild & Moderate (-)" and "Sever & Critical (-)" indicate the SARS-CoV-2 infected individual's symptom onset occurred outside the exposure time window; The "(+)" in "Mild & Moderate (+)" and "Sever & Critical (+)" indicate the SARS-CoV-2 infected individual's symptom onset occurred within the exposure time window. Dots and lines indicate point estimates and 95% confidence interval of the odds ratio, numbers below the dots indicate the numerical value of the point estimates; "Ref." stands for reference category; * indicates p-value<0.05, ** indicates p-value<0.01, *** indicates p-value<0.001. Note that the regression results of odds ratio in healthcare setting after January 25 is not visualized due to very low point estimate (9.1×10^{-8}), with 0 of the 927 contact exposures in healthcare settings after January 25 led to secondary transmissions.