Biosafety and Health 7 (2025) 395–402

Contents lists available at ScienceDirect

Biosafety and Health

journal homepage: www.elsevier.com/locate/bsheal

Original Research

Attribution of super-spreaders to the COVID-19 outbreak

Rui Wang, Jianping Huang [*], Xinbo Lian, Yingjie Zhao, Shujuan Hu, Beidou Zhang

Collaborative Innovation Center for West Ecological Safety (CIWES), College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China

A R T I C L E I N F O

Article history:
Received 28 June 2025
Revised 29 September 2025
Accepted 30 September 2025
Available online 5 October 2025

Keywords:
Coronavirus disease 2019 (COVID-19)
Superspreading events (SSEs)
Transmission chain
Reproduction number
Dispersion parameter

A B S T R A C T

The explosive growth during the early stages and the sustained transmission in the later phases of the coronavirus disease 2019 (COVID-19) pandemic may be closely linked to superspreading events (SSEs), yet in-depth research into their specific mechanisms and quantitative effects remains limited. This study, based on data from 4,519 COVID-19 cases across eight regions in China, reconstructed transmission chains and quantified key parameters such as the basic reproduction number (R_0) and dispersion parameter (k), revealing a high degree of heterogeneity in COVID-19 transmission. The results showed that the majority of COVID-19 cases were mild, with female cases in some regions being significantly older than males. Epidemic curves were highly similar in geographically proximal areas, with the longest transmission chain reaching nine generations. The transmission parameters revealed a serial interval of 1.27–4.71 days, R_0 ranging from 0.87 to 2.65, and k values between 0.50–2.04, demonstrating that super-spreaders serve as critical drivers of epidemic spread. We found that 1.35 % of cases identified as super-spreaders directly responsible for 40.09 % of secondary cases. Occupationally, students and catering staff were identified as high-risk groups for super-spreading. Geographically, household or community transmission served as the main driver of SSEs in six regions, while school-based transmission dominated in one region. These findings provide crucial scientific evidence for advancing our understanding of COVID-19 transmission dynamics and informing precision prevention strategies.

1. Introduction

From the 2003 severe acute respiratory syndromes outbreak to coronavirus disease 2019 (COVID-19) in late 2019, fast-spreading diseases have caused major global health crises, economic problems, and social instability. Most disease models assume infections spread evenly—where every infected individual is equally likely to infect others [1]. However, studies conducted since 2003 reveal a different pattern [2–4]. In actual outbreaks, disease transmission isn't evenly distributed—a small number of individuals infect significantly more people than others. Sometimes this uneven spread leads to "superspreading events (SSEs)" [5]. In these cases, just a few infected people (called "super-spreaders") end up passing the disease to many others. While these people make up only a small fraction of all cases, they cause most of the new infections. A super-spreader is typically defined as an infected individual who transmits the virus to an unusually high number of secondary cases, significantly exceeding the average transmission number expected based on the basic reproduction number (R_0) [6]. For instance, an estimated 20 % of COVID-19 cases may have been responsible for up to 80 % of transmissions [7,8]. SSEs are strongly associated with explosive early-stage epidemic growth and sustained transmission in later phases, presenting significant challenges to containment efforts [9]. Therefore, identifying and controlling super-spreaders and SSEs is critical for effective epidemic response, as it enables the disruption of transmission chains and mitigation of the outbreak's impact.

The transmission dynamics of COVID-19 can be quantitatively characterized by two key epidemiological parameters: the R_0 and the dispersion parameter (k). R_0 refers to the average number of secondary infections generated by a typical primary infection case in a fully susceptible population, without any external interventions, over the course of its infectious period [10]. It is a theoretical value that measures the inherent transmission capacity of a pathogen and is primarily used for risk assessment in the early stages of an outbreak. R_t (also known as the time-varying reproduction number) refers to the average number of secondary infections that a typical infection case is expected to generate during its remaining infectious period at time t, taking into account the existing immunity in the population (due to prior infections or vaccinations) and the various interventions being implemented (such as social distancing, mask mandates, etc.) [11]. It is a dynamically changing indicator that reflects the real-time development of the outbreak and

---

* Corresponding author: Collaborative Innovation Center for West Ecological Safety (CIWES), College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China.
E-mail address: hjp@lzu.edu.cn (J. Huang).

https://doi.org/10.1016/j.bsheal.2025.09.010
2590-0536/

HIGHLIGHTS

Scientific question

Superspreading events are considered to be closely associated with the explosive growth in early stages and sustained transmission in later phases of coronavirus disease 2019 (COVID-19) pandemic. However, their specific mechanisms and quantitative effects remain to be elucidated.

Evidence before this study

Previous studies have highlighted significant transmission heterogeneity in COVID-19, with super-spreaders substantially amplifying outbreaks. However, systematic analysis of super-spreader profiles and local transmission chains in mainland China remains limited.

New findings

Our study found that 1.35 % of cases were identified as super-spreaders directly responsible for 40.09 % of secondary cases. Occupationally, students and catering staff were identified as high-risk groups for super-spreading. Geographically, household or community transmission served as the main driver of superspreading events in six regions. Regional variations were observed in key transmission parameters: basic reproduction number R_0 (0.87–2.65), dispersion coefficient k (0.50–2.04), and serial interval (1.27–4.71 days), demonstrating spatial heterogeneity in epidemic transmission dynamics.

Significance of the study

By systematically identifying super-spreaders as pivotal outbreak drivers and their associated risk factors, this study provides critical evidence for developing precision public health strategies tailored to local transmission dynamics.

the effectiveness of control measures. In the early stages of an outbreak, when population immunity is close to zero and no effective interventions are implemented, the R_t can be approximately equal to the R_0. An $R_0 > 1$ indicates that the outbreak will grow, while an $R_0 < 1$ suggests that the outbreak will gradually subside. The variation in the number of secondary cases generated by each primary case is characterized using a negative binomial distribution. Within this framework, the k is a value that needs to be estimated from data. A lower k value indicates greater heterogeneity in transmission (i.e., a higher degree of overdispersion), meaning that a larger proportion of transmission events are driven by a small number of individuals [5,12,13]. A 2020 study in China's Hong Kong Special Administrative Region estimated an overall reproductive number R of 0.58 and a k of 0.43, highlighting substantial variation in individual infectiousness [13]. Similarly, research from Singapore in 2020 reported an even lower k value (0.11), further underscoring the pivotal role of SSEs in driving transmission [14,15]. Data from multi-center studies globally indicate that when R_0 is maintained at 2–3, k values generally fall within a low range of 0.10 [15]. Recent research based on genomic data from multiple European countries also supports this finding, estimating k values between 0.1 and 0.5 [16]. These findings collectively suggest that COVID-19 prevention and control strategies need to consider not only the average transmission rate but also the impact of transmission heterogeneity, particularly the early identification and intervention of SSEs.

Current research on COVID-19 transmission heavily relies on the susceptible-exposed-infectious-removed (SEIR) model. However, it is important to note that the classical SEIR framework assumes no transmission during the exposed (E) phase, which may not fully capture the transmission dynamics of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), as evidence suggests the occurrence of presymptomatic transmission. Consequently, some studies employ modified SEIR models (allowing transmission in the late stages of the exposed period) or, alternatively, use the susceptible-infectious-removed (SIR) model as a simplified approximation to simulate overall transmission trends. Nevertheless, these models typically assume homogeneous mixing within the population and fixed parameters, inevitably overlooking critical mechanistic details such as individual contact history and the complex topology of transmission chains. Most critically, their core framework fails to reflect the extreme heterogeneity in individual-level infectiousness-particularly SSEs-which is a defining characteristic of SARS-CoV-2 transmission [17]. The systematic contact tracing strategy implemented in China's COVID-19 epidemic prevention and control efforts has provided unique data support for transmission dynamics research. By constructing a comprehensive transmission chain database, we can more accurately quantify key transmission parameters. This study, based on epidemiological investigation data from COVID-19 outbreaks in eight regions of China in 2022, systematically reconstructed the transmission chain and achieved precise estimates of parameters such as the R_0 and the k. The research not only deeply revealed the spatiotemporal distribution characteristics of COVID-19 transmission heterogeneity but also clearly pointed out the crucial determinant role of super-spreaders in epidemic outbreaks and transmission. By quantifying transmission dynamics parameters, we have provided empirical evidence for optimizing prevention and control strategies for major infectious diseases.

2. Materials and methods

2.1. Study regions

This study selected eight representative regions with COVID-19 outbreaks as research objects. These regions all established complete epidemiological investigation and contact tracing systems during the COVID-19 epidemic in 2022, providing a reliable data foundation for reconstructing transmission chains. The study regions included: 1) Tianjin Municipality (Tianjin); 2) Anyang of Henan Province (Anyang); 3) Hangzhou of Zhejiang Province (Hangzhou); 4) Shaanxi Province (Shaanxi), given that the outbreaks in Xi'an City, Hanzhong City, and Baoji City within the province shared a common source and belonged to the same transmission chain, they were analyzed together; 5) Haining City in Zhejiang Province (Haining); 6) Taiyuan City in Shanxi Province (Taiyuan); 7) Beijing Municipality (Beijing); 8) Zhengzhou City in Henan Province (Zhengzhou).

2.2. Data collection

The COVID-19 epidemic data used in this study were sourced from epidemic reports published on the official websites of the National Health Commission [18] and the Health Commissions of corresponding provinces and cities [19–24]. The cumulative number of cases and the time ranges for each study region are as follows: Tianjin (430 cases, January 8 to February 7, 2022), Anyang (468 cases, January 8 to January 29, 2022), Hangzhou (114 cases, January 26 to February 3, 2022), Shaanxi (398 cases, March 5 to March 28, 2022), Haining (172 cases, April 3 to April 13, 2022), Taiyuan (351 cases, April 3 to April 27, 2022), Beijing (2231 cases, April 22 to June 29, 2022), and Zhengzhou (355 cases, May 1 to May 22, 2022).

2.3. Statistical analysis

Detailed definitions of epidemic curve patterns [25], transmission chain classifications [26], and super-spreader [26–28] criteria are provided in the Supplementary Data.

Data were organized and statistically analyzed using R software (version 4.2.3). The calculations of the R_t and k values utilized the EpiEstim package (version 2.2.4), MASS package (version 7.3.58.1), and boot package (version 1.3.31). First, we estimated R_t using the parametric method in the EpiEstim package. The estimation was performed with a 7-day sliding window, parameterized using the mean and standard deviation of the serial interval (SI). Based on the theoretical principle that $R_t \approx R_0$ holds during the initial outbreak phase when population immunity is negligible and intervention measures are not yet implemented, we estimated R_0 by calculating the mean value of R_t from the first three time points in the reconstructed transmission dynamics timeline. Second, the k value was estimated using a negative binomial regression model, assessing transmission heterogeneity by comparing the actual case distribution with the fitted negative binomial distribution curve. For both R_0 and k values, 95 % confidence intervals (CI) were calculated using the bootstrap method (with R = 1,000 replications).

3. Results

3.1. Characteristics of COVID-19 cases

The demographic and infection information of the cases is shown in Fig. 1A, with detailed statistical analysis results provided in Tables S1–S2. Among the 1,933 cases with available gender information (from six regions), there were 884 males (45.73 %) and 1,049 females (54.27 %), with no statistically significant difference in gender distribution ($t = -0.637$, $P = 0.538$). The age range of cases was 1 to 95 years. Comparisons of gender and age between regions revealed that the average age of female cases in Anyang and Taiyuan was signif-

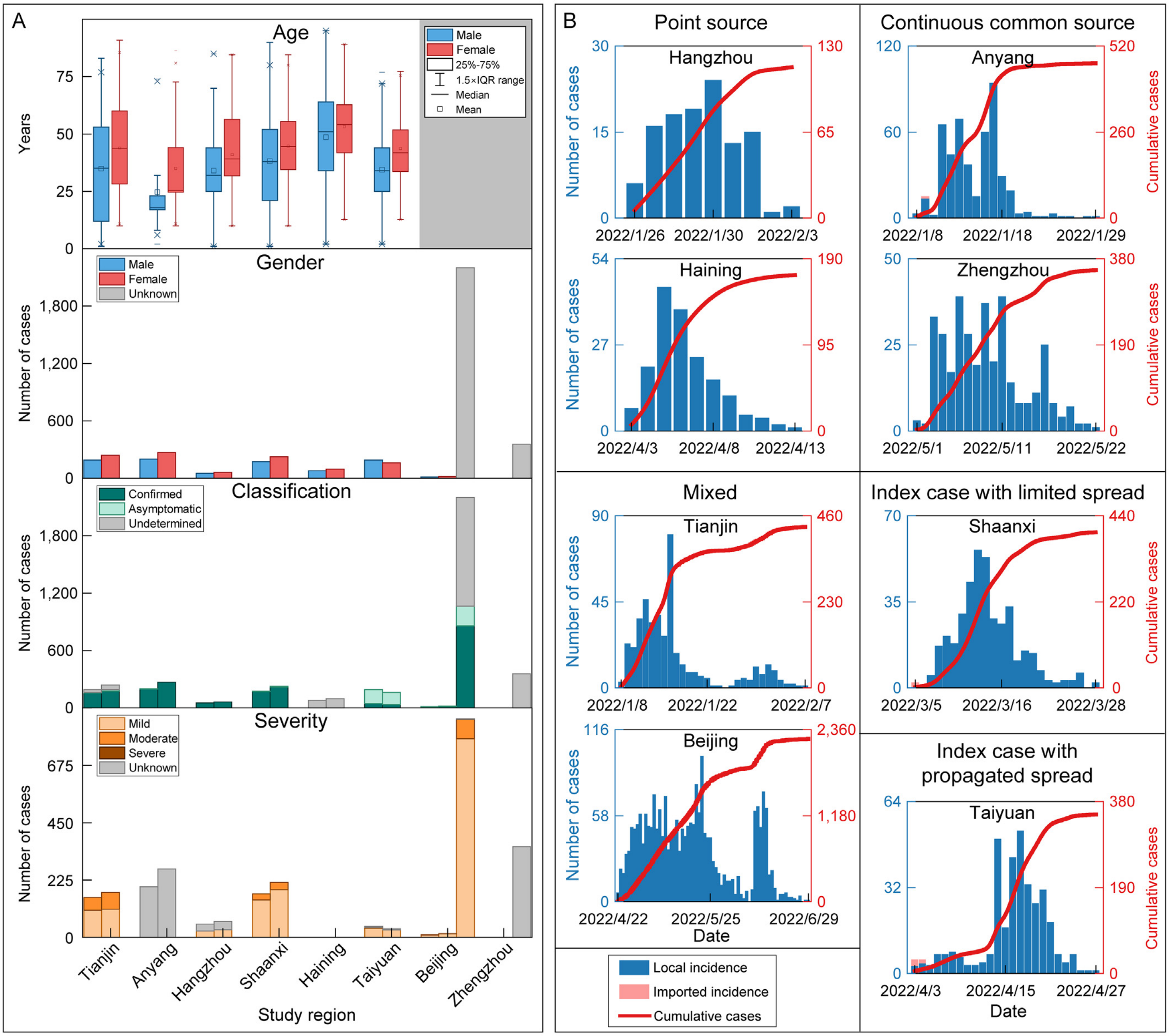

Fig. 1. COVID-19 in eight regions in China. A) Demographic and infection information of infected cases. Gray areas and gray bars indicate that relevant information is unavailable. B) Epidemic curves of infected cases. Abbreviations: COVID-19, coronavirus disease 2019; IQR, interquartile range.

icantly higher than that of males ($P = 0.017$ and $P = 0.046$, respectively), while there were no statistically significant differences in gender-age comparisons in other regions ($P > 0.05$). Among the 2,779 cases with clear clinical classifications, confirmed cases accounted for 81.29 % (2,259/2,779), with mild cases comprising 65.25 % (1,474/2,259) of the confirmed cases. Comparative analyses of classification case numbers and severity levels by gender across six regions found no statistical differences ($\chi^2 = 4.659$, $P = 0.097$; $\chi^2 = 2.003$, $P = 0.367$).

Based on the reporting dates, we conducted a stratified analysis of local and imported cases (Fig. 1B), revealing distinct transmission patterns across regions: Hangzhou and Haining exhibited typical point-source exposure patterns, with epidemic durations of 9–11 days. The epidemic curves showed a unimodal shape characterized by rapid rise and decline, suggesting transient SSEs initiated by a single infection source. Anyang and Zhengzhou demonstrated continuous common-source exposure, with epidemic curves maintaining plateau phases for 7 and 9 days respectively, indicating prolonged SSEs. Tianjin and Beijing displayed mixed exposure patterns, with bimodal curves separated by 15–20 day intervals, reflecting initial common-source transmission followed by multigenerational SSEs. Shaanxi and Taiyuan exhibited index case-initiated transmission chains, with Taiyuan showing progressively shortened interpeak intervals, suggesting cumulative effects of multigenerational SSEs.

3.2. Reconstruction of transmission chains

To gain a deeper understanding and quantify the heterogeneous characteristics presented by these transmission patterns, we constructed detailed and comprehensive transmission chains for the COVID-19 outbreaks in these eight regions using exhaustive and valid exposure history information. These chains were classified into simple transmission chains (Fig. S1) and ordinary transmission chains (Fig. 2), with detailed chain characteristics outlined in Table S3.

In Tianjin, 128 of 430 cases had clear contact history, forming 26 chains (18 simple, 8 ordinary) with a max of three generations. Of these cases, 28.13 % (36/128) led to second-generation cases. One super-spreader (Tianjin Case 2) was identified in a school and directly infected 15 people. The other super-spreader (Tianjin Case 13) infected 19 people through a gathering. In Anyang, 464 of 468 cases formed 2 chains (1 simple, 1 ordinary) reaching six generations. Among these cases, 12.07 % (56/464) led to second-generation cases. Two super-spreaders (Anyang Case 7 and 48) together directly infected 278 people in a school environment. The other two super-spreaders (Anyang Case 8 and 151) directly infected 58 people through household or community transmission. In Hangzhou, 100 of 114 cases formed 11 chains (7 simple, 4 ordinary) up to five generations. Of these cases, 22.00 % (22/100) led to second-generation cases. One super-spreader (Hangzhou Case 1), an employee of a catering company, directly infected 19 people. The other super-spreader (Hangzhou Case 29), a salesperson, infected 29 people through market or community transmission. In Shaanxi, 334 of 398 cases formed 43 chains (31 simple, 12 ordinary) up to seven generations. Among these cases, 29.94 % (100/334) led to second-generation cases. Notably, super-spreaders (Shaanxi Case 10, 38, and 62, as restaurant staff, triggered household cluster transmission through occupational exposure, forming highly concentrated transmission clusters. In Haining, 20 of 172 cases formed 3 chains (1 simple, 2 ordinary) up to three generations, with no super-spreaders. Among these cases, 30.00 % (6/20) led to second-generation cases. In Taiyuan, 179 of 351 cases formed 35 chains (26 simple, 9 ordinary) up to 5 generations. Among these cases, 29.05 % (52/179) led to second-generation cases. Two super-spreaders (Taiyuan Case 2 and 130), couriers, directly infected 24 people through community transmission. The other super-spreader (Taiyuan Case 7) infected 10 people through various transmission routes. In Beijing, 431 of 2,231 cases formed 66 chains (54 simple, 12 ordinary) up to eight generations. Among these cases, 28.07 % (121/431) led to second-generation cases. Two super-spreaders (Beijing Case 5 and 8) together directly caused 48 secondary infections in a school environment. One super-spreader (Beijing Case 27) infected 12 people due to shared meals. The other four super-spreaders (Beijing Case 28, 82, 99, and 728) infected 112 people through various routes. In Zhengzhou, 346 of 355 cases formed 20 chains (9 simple, 11 ordinary) up to nine generations. Among these cases, 38.15 % (132/346) led to second-generation cases. Two super-spreaders (Zhengzhou Case 86 and 162) together directly infected 25 people in a school environment. The other three super-spreaders (Zhengzhou Case 44, 192, and 298) infected 38 people through household or community transmission.

Through systematic analysis of transmission chains, we identified significant heterogeneous characteristics in the spread of COVID-19. Ordinary transmission chains ($n = 59$) involved a total of 1,583 cases, with an average size of 26.83 cases per chain. Simple transmission chains ($n = 147$) involved 419 cases, with an average size of 2.85 cases per chain. A total of 27 super-spreaders were identified, whose transmission chains involved 1,728 cases, with an average size of 64 cases per chain.

3.3. R_0 and k values of COVID-19

To quantitatively assess the transmission dynamics of COVID-19, we systematically calculated three key epidemiological parameters based on transmission chain data from various regions: SI, R_t, and k (Fig. 3). The results showed that Anyang had the largest mean SI of 4.71 (95 % CI: 4.39–5.02) days, indicating a relatively slower transmission speed in Anyang with longer infectious chains between cases. Haining had the smallest mean SI of 1.27 (95 % CI: 1.00–1.55) days, suggesting a faster transmission speed in Haining with shorter infectious chains, leading to a rapid increase in case numbers within a short period. The mean SI for other regions fell between these two values. Shaanxi had the highest R_0 of 2.65 (95 % CI: 2.34–2.88), indicating a relatively higher transmission potential in Shaanxi, making the epidemic more prone to spread. Haining had the lowest R_0 of 0.87 (95 % CI: 0.69–1.01), implying that in the absence of external interventions, the epidemic might gradually die out as each infected individual, on average, infects fewer than one person. The R_0 for other regions was between these two values. The k values for Tianjin, Anyang, and Shaanxi were <1, with the negative binomial distribution exhibiting a longer tail. This indicates high heterogeneity in transmission in these three regions, where a small number of infected individuals triggered a large number of secondary infections, evidencing a pronounced “super-spreader” effect. The timing of transmission in these regions showed marked clustering, suggesting that outbreaks could occur rapidly within a short period, driven by a few key individuals. The k values for Hangzhou, Haining, Taiyuan, Beijing, and Zhengzhou ranged between 1 and 5, indicating moderate transmission heterogeneity in these areas. The spread of the epidemic in these regions was relatively dispersed but still exhibited some clustering. Although there was no obvious “super-spreader” effect like that in Tianjin, Anyang, and Shaanxi, the transmission was not entirely uniform.

3.4. Super-spreaders are key drivers of COVID-19 outbreaks and spread

Among the eight regions studied, confirmed COVID-19 cases primarily exhibited mild clinical manifestations. Cases in Taiyuan and Anyang showed gender and age differences, with female cases having a significantly higher average age than male cases. Spatial transmission patterns revealed that neighboring regions with close geographical proximity exhibited highly similar epidemic curves. The longest

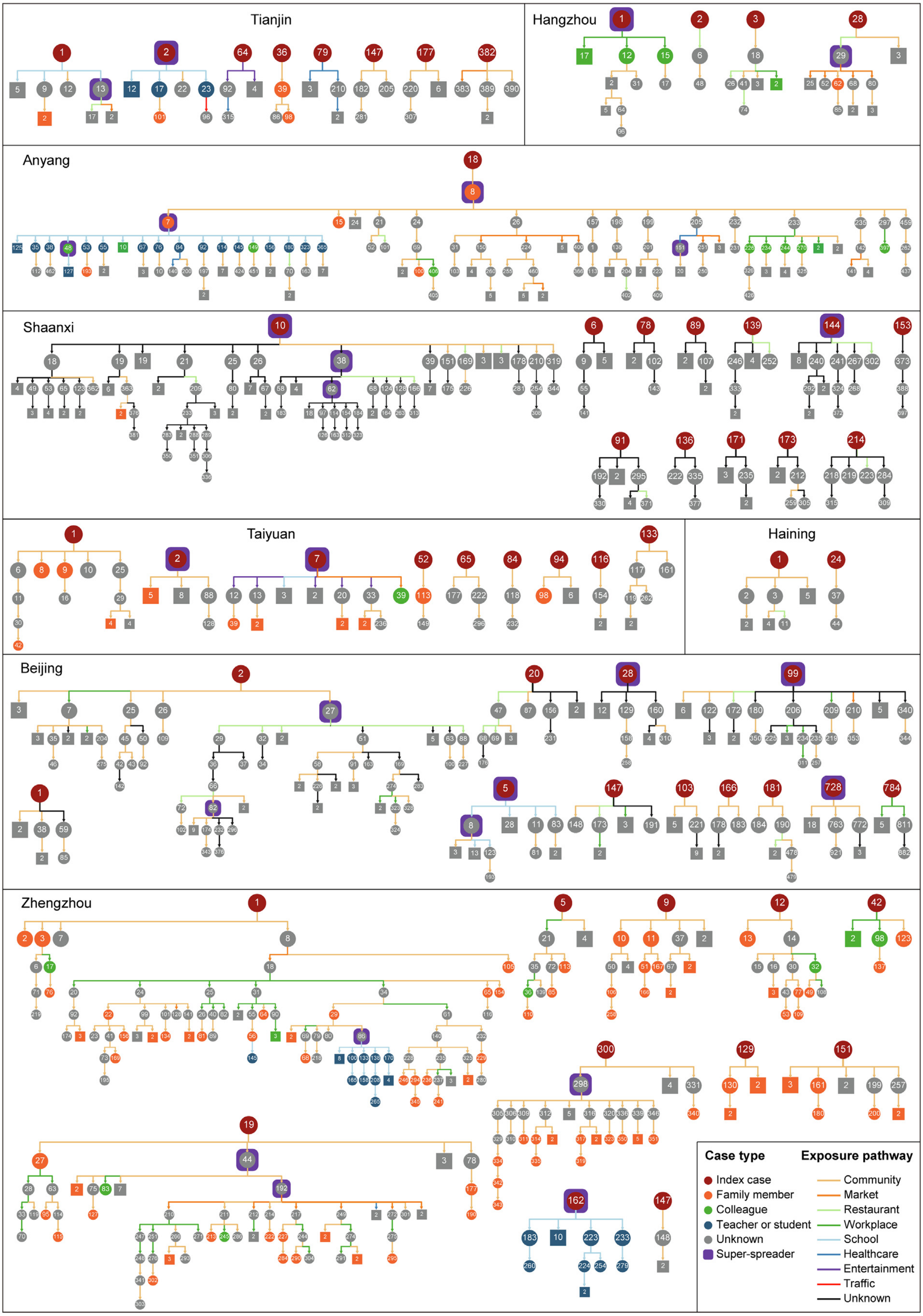

Fig. 2. Ordinary transmission chains of coronavirus disease 2019 (COVID-19) in eight regions. Nodes represent cases, numbers in circles indicate case report serial number, numbers in boxes indicate the number of cases, arrows indicate transmission relationships between cases, and purple squares represent super-spreaders.

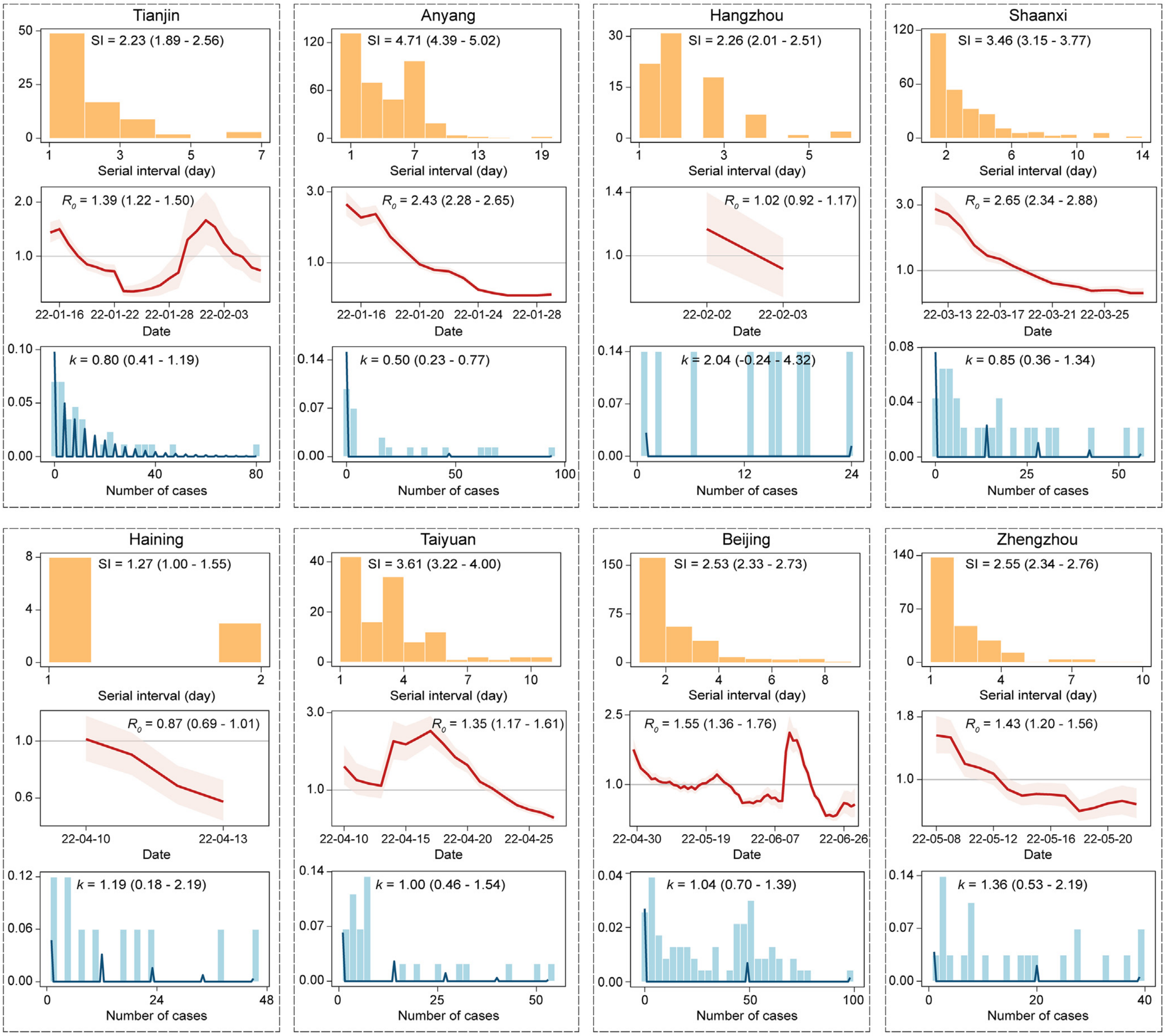

Fig. 3. Calculation results of SI, R_t, and k for eight regions. The yellow bars represent the frequency distribution of SI, the red line indicates the trend of R_t, the blue bars show the distribution density of case numbers, and the blue line represents the fitted curve of the negative binomial distribution. The data in parentheses represent the 95 % confidence interval. Abbreviation: SI, serial interval.

transmission chain extended up to nine generations. Notably, 1.35 % (27/2,002) of cases identified as super-spreaders directly responsible for 40.09 % (720/1,796) of secondary cases. Across seven regions, a total of 27 infected individuals with superspreading characteristics were identified. Among the 15 super-spreaders with complete gender information, there were 7 males (46.67 %) and 8 females (53.33 %). In terms of occupational composition, students and catering staff were more likely to become super-spreaders. SSEs in six regions were primarily driven by household or community transmission, while one region was predominantly associated with school-based transmission. The range of SI across regions was 1.27–4.71 days, the R_0 ranged between 0.87–2.65, and the k spanned from 0.50 to 2.04. These findings indicate that super-spreaders are crucial drivers of COVID-19 outbreaks and spread. Their transmission characteristics are manifested as highly clustered spatio-temporal distribution, elevated risk associated with specific occupations and venues, and pronounced centrality within transmission networks (Fig. 4).

4. Discussion

This study systematically analyzed the transmission chain characteristics of 4,519 COVID-19 cases in eight regions of China, revealing the central role of super-spreaders in epidemic transmission and quantifying the heterogeneous features of COVID-19 transmission across multiple regions in China for the first time. The findings not only confirm the significant heterogeneity in COVID-19 transmission but also provide important scientific evidence for a deeper understanding of its transmission mechanisms and the formulation of precise prevention and control strategies.

Compared to the classic homogeneous transmission model proposed by Anderson [1,29], our study validated the significant heterogeneity (k = 0.50–2.04) in COVID-19 transmission through empirical data, supporting the super-spreading theory proposed by Lloyd-Smith et al. [5]. Compared to D.C. Adam et al.'s [13] study on the China's Hong Kong epidemic (k = 0.45), our study found lower k values in

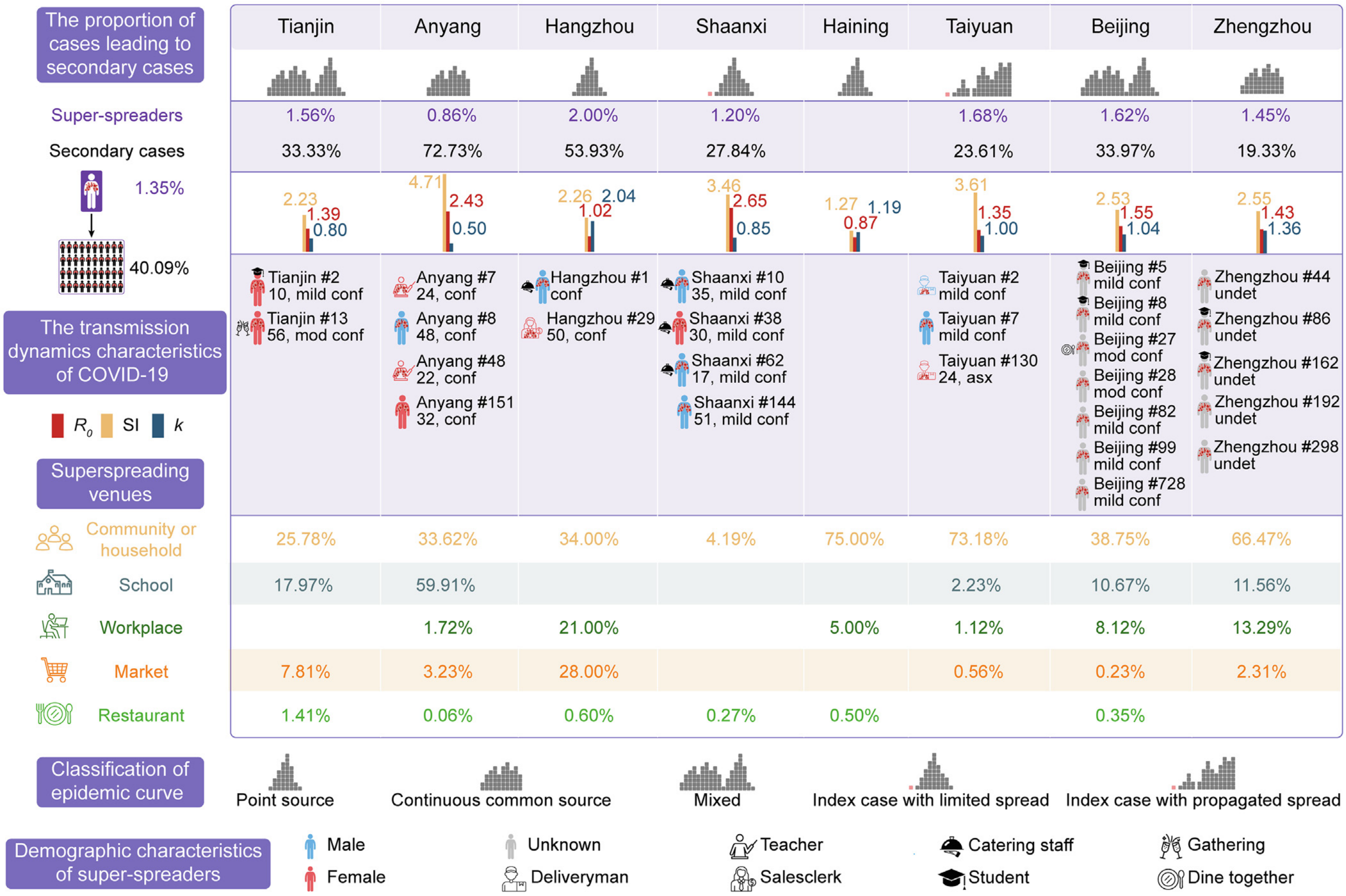

Fig. 4. Conceptual diagram of coronavirus disease 2019 (COVID-19) transmission dynamics and super-spreader characteristics in China. Abbreviations: conf, confirmed COVID-19; asx, asymptomatic COVID-19; undet, undetected COVID-19.

some regions of China's mainland, indicating more significant transmission heterogeneity. This may reflect the influence of factors such as higher population density and differences in social contact patterns in cities of China's mainland. Furthermore, the cross-regional comparison in this study revealed spatial heterogeneity in transmission parameters. By combining R_0 and k values, a transmission risk classification can be proposed, providing a basis for formulating regionally differentiated prevention and control strategies. When $R_0 > 1.5$ and $k < 1.0$, it indicates high transmission potential and strong dependence on superspreading for COVID-19 (e.g., Anyang and Shaanxi in this study), which is considered high-risk. When $R_0 = 1.0$–1.5 and $k = 1.0$–5.0, it indicates moderate transmission potential and heterogeneity for COVID-19 (e.g., Taiyuan and Hangzhou), which is considered medium-risk. When $R_0 < 1.0$ or $k > 5.0$, it indicates a self-extinguishing trend or uniform transmission for COVID-19 (e.g., Haining), which is considered low-risk.

In the process of deeply exploring the characteristics of super-spreaders, this study revealed the significant status of students and catering staff as high-risk groups. Therefore, it is recommended to implement regular screening strategies for these specific populations to detect potential infections early. At the same time, communities or households, and schools were identified as high-risk environments for frequent SSEs, which is consistent with the conclusions of previous studies, emphasizing the urgency of strengthening prevention and control measures in these places. Considering the increased infection risk among co-habitants due to long-term close contact and social interaction [29,30], accurately identifying super-spreaders and their environments is invaluable for effectively deploying preventive measures and minimizing the spread of infections. For regions with different risk levels, this study proposes differentiated prevention and control strategies: In high-risk areas where epidemic transmission is mainly driven by super-spreaders, measures such as precise transmission chain interruption, targeted venue shutdown, and enhanced traceability capabilities should be taken to quickly control the spread of the epidemic; In medium-risk areas facing the complex situation of coexisting superspreading and community transmission, dynamically adjusting venue control strategies and strengthening supervision of key populations become key to controlling the epidemic; In low-risk areas, where the transmission pattern approaches a Poisson distribution, strategies such as routine monitoring and protection of key populations are sufficient to cope with the situation, but high vigilance must still be maintained to prevent a resurgence of the epidemic.

This study has several limitations that should be considered when interpreting the results. First, the epidemiological data were collected from outbreaks in eight regions of China during the year 2022. Consequently, our estimates of the R_t and k primarily reflect the transmission characteristics of the Omicron variants predominant at that time. The generalizability of our findings to earlier or future variants with different virological properties (e.g., transmissibility, incubation period) may be limited. Second, while the eight regions provide diverse case studies, they may not be fully representative of the entire country or other international contexts. Geographic, demographic, and behavioral differences could influence transmission patterns. Future studies incorporating data from a wider range of locations are needed to validate the universality of our conclusions. Third, and crucially, our analysis could not fully account for the heterogeneity of nonpharmaceutical interventions across regions and over time. Although a major strength of our dataset is its detailed transmission chains, information on the precise timing and intensity of localized public health measures (e.g., school closures, workplace restrictions, gathering size limits, mobility changes) was not systematically integrated into our quantitative models. We strongly recommend that future research merge detailed transmission chain data with granular policy data to explicitly test this hypothesis and disentangle the effects of virus biology from the effects of public health policy. Fourth, our findings on SSEs locations must be interpreted in the context of shifting public health policies during the study period. As interventions intensified-particularly with widespread lockdowns-transmission modes shifted from diverse public venues (e.g., early periods) toward household settings (e.g., later periods). This may bias our venue-specific observations, potentially overestimating household transmission and underestimating the role of public venues that were restricted. Despite these limitations, the high-resolution contact tracing data from which our transmission chains were reconstructed provides a unique and invaluable resource for quantifying heterogeneity, offering a more realistic perspective than traditional modeling approaches.

Acknowledgements

This study was supported by Self-Supporting Program of Guangzhou Laboratory (SRPG22-007), Major Project of Guangzhou National Laboratory (GZNL2024A01004); National Key Research and Development Program of China (2023YFC3503400); Gansu Province Intellectual Property Program (Oriented Organization) Project (22ZSCQD02).

Conflcit of interest statement

The authors declare that there are no conflicts of interest.

Author contributions

Rui Wang: Writing – original draft, Writing – review & editing, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Jianping Huang:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Xinbo Lian:** Methodology, Data curation. **Yingjie Zhao:** Writing – review & editing, Methodology, Data curation. **Shujuan Hu:** Writing – review & editing, Methodology, Funding acquisition. **Beidou Zhang:** Writing – review & editing, Funding acquisition.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bsheal.2025.09.010.

References

[1] R.M.M. R.M. Anderson, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, Oxford, 1991. https://doi.org/10.1093/oso/9780198545996.001.0001.

[2] W. Liang, Z. Zhu, J. Guo, Z. Liu, W. Zhou, D.P. Chin, A. Schuchat, Severe acute respiratory syndrome, Beijing, 2003, Emerg. Infect. Dis. 10 (1) (2004) 25–31, https://doi.org/10.3201/eid1001.030553.

[3] S.M. Poutanen, D.E. Low, B. Henry, S. Finkelstein, D. Rose, K. Green, R. Tellier, R. Draker, D. Adachi, M. Ayers, et al., Identification of severe acute respiratory syndrome in Canada, N. Engl. J. Med. 348 (20) (2003) 1995–2005, https://doi.org/10.1056/NEJMoă34.

[4] Z. Shen, F. Ning, W. Zhou, X. He, C. Lin, D.P. Chin, Z. Zhu, A. Schuchat, Superspreading SARS events, Beijing, 2003, Emerg. Infect. Dis. 10 (2) (2004) 256–260, https://doi.org/10.3201/eid1002.030732.

[5] J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, W.M. Getz, Superspreading and the effect of individual variation on disease emergence, Nature 438 (7066) (2005) 355–359, https://doi.org/10.1038/nature04153.

[6] A. Teicher, Super-spreaders: a historical review, Lancet Infect. Dis. 23 (10) (2023) e409–e417, https://doi.org/10.1016/S1473-3099(23)00183-4.

[7] K. Kupferschmidt, Why do some COVID-19 patients infect many others, whereas most don't spread the virus at all? https://www.science.org/content/article/why-do-some-covid-19-patients-infect-many-others-whereas-most-don-t-spread-virus-all, 2020 (accessed 19 May 2025).

[8] D. Mackenzie, Why coronavirus superspreaders may mean we avoid a deadly pandemic？ https://www.newscientist.com/article/2234388-why-coronavirus-superspreaders-may-mean-we-avoid-a-deadly-pandemic/, 2020 (Accessed 18 February 2025).

[9] T.R. Frieden, C.T. Lee, Identifying and interrupting superspreading events-implications for control of severe acute respiratory syndrome coronavirus 2, Emerg. Infect. Dis. 26 (6) (2020) 1059–1066, https://doi.org/10.3201/eid2606.200495.

[10] R. Breban, R. Vardavas, S. Blower, Theory versus data: how to calculate R_0?, PLoS One 2 (3) (2007) e282, https://doiorg/10.1371/journal.pone.0000282.

[11] A. Cori, N.M. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics, Am. J. Epidemiol. 178 (9) (2013) 1505–1512, https://doi.org/10.1093/aje/kwt133.

[12] S. Ryu, J.Y. Chun, S. Lee, D. Yoo, Y. Kim, S.T. Ali, B.C. Chun, Epidemiology and transmission dynamics of infectious diseases and control measures, Viruses 14 (11) (2022) 2510, https://doi.org/10.3390/v14112510.

[13] D.C. Adam, P. Wu, J.Y. Wong, E.H.Y. Lau, T.K. Tsang, S. Cauchemez, G.M. Leung, B.J. Cowling, Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong, Nat. Med. 26 (11) (2020) 1714–1719, https://doi.org/10.1038/s41591-020-1092-0.

[14] A. Tariq, Y. Lee, K. Roosa, S. Blumberg, P. Yan, S. Ma, G. Chowell, Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020, BMC Med. 18 (1) (2020) 166, https://doi.org/10.1186/s12916-020-01615-9.

[15] A. Endo, S. Abbott, A.J. Kucharski, S. Funk, Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China, Wellcome Open Res. 5 (2020) 67, https://doi.org/10.12688/wellcomeopenres.15842.3.

[16] E.B. Hodcroft, M.S. Wohlfender, R.A. Neher, J. Riou, C.L. Althaus, Estimating *Re* and overdispersion in secondary cases from the size of identical sequence clusters of SARS-CoV-2, PLoS Comput. Biol. 21 (4) (2025) e1012960, https://doi.org/10.1371/journal.pcbi.1012960.

[17] J. Huang, L. Zhang, X. Liu, Y. Wei, C. Liu, X. Lian, Z. Huang, J. Chou, X. Liu, X. Li, et al., Global prediction system for COVID-19 pandemic, Sci. Bull. 65 (22) (2020) 1884–1887, https://doi.org/10.1016/j.scib.2020.08.002.

[18] National Health Commission of the People's Republic of China, Daily briefing on novel coronavirus cases in China, http://en.nhc.gov.cn/DailyBriefing.html, 2023 (accessed 22 June 2023).

[19] Health Commission of Henan Province, Daily briefing on novel coronavirus cases in Henan province, https://wsjkw.henan.gov.cn/ztzl/xxgzbdfyyqfk/, 2023 (accessed 22 June 2023).

[20] Health Commission of Shaanxi Province, Daily briefing on novel coronavirus cases in Shaanxi province, http://sxwjw.shaanxi.gov.cn/sy/ztzl/fyfkzt/index.html, 2023 (accessed 22 June 2023).

[21] Health Commission of Beijing Municipal, Press releases, http://wjw.beijing.gov.cn/xwzx_20031/xwfb/index_20.html, 2023 (accessed 22 June 2023).

[22] Health Commission of Zhejiang Province, Daily briefing on novel coronavirus cases in Zhejiang province, https://wsjkw.zj.gov.cn/col/col1202101/index.html?uid=4978845&pageNum=51, 2023 (accessed 22 June 2023).

[23] Health Commission of Shanxi Province, Daily briefing on novel coronavirus cases in Shanxi province, http://wjw.shanxi.gov.cn/ztzl/xxgzbdfyyqfk/, 2023 (Accessed 22 June 2023).

[24] Tianjin Municipal Health Commission, Daily briefing on novel coronavirus cases in Tianjin city, https://wsjk.tj.gov.cn/ZTZL1/ZTZL750/YQFKZL9424/, 2023 (accessed 22 June 2023).

[25] Centers for Disease Control and Prevention, USA, Quick-learn lesson using an epi curve to determine mode of spread, https://www.cdc.gov/training/QuickLearns/epimode/index.html, 2025 (accessed 26 June 2025).

[26] J. Wallinga, P. Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures, Am. J. Epidemiol. 160 (6) (2004) 509–516, https://doi.org/10.1093/aje/kwh255.

[27] Q.J. Leclerc, N.M. Fuller, L.E. Knight, S. Funk, G.M. Knight, What settings have been linked to SARS-CoV-2 transmission clusters? Wellcome Open Res. 5 (2020) 83, https://doi.org/10.12688/wellcomeopenres.15889.2.

[28] J. Cai, W. Sun, J. Huang, M. Gamber, J. Wu, G. He, Indirect virus transmission in cluster of COVID-19 cases, Wenzhou, China, 2020, Emerg. Infect. Dis. 26 (6) (2020) 1343–1345, https://doi.org/10.3201/eid2606.200412.

[29] X. Zhao, Z. Shen, L. Sun, L. Cheng, M. Wang, X. Zhang, B. Xu, L. Tian, Y. Miao, X. Wu, et al., A network meta-analysis of secondary attack rates of COVID-19 in different contact environments, Epidemiol. Infect. 149 (2021) e219, https://doi.org/10.1017/S0950268821002223.

[30] C. Liu, J. Huang, S. Chen, D. Wang, L. Zhang, X. Liu, X. Lian, The impact of crowd gatherings on the spread of COVID-19, Environ. Res. 213 (2022) 113604, https://doi.org/10.1016/j.envres.2022.113604.