



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Reconstruction of the Transmission Chain of COVID-19 Outbreak in Beijing's Xinfadi Market, China

Tianyi Luo^{1,2}, Jiaojiao Wang¹, Quanyi Wang^{3,4}, Xiaoli Wang^{3,4}, Pengfei Zhao¹, Daniel Dajun Zeng¹, Qingpeng Zhang⁵, Zhidong Cao^{1,*}

¹ Institute of Automation Chinese Academy of Sciences, Beijing, 100190, China

² University of Chinese Academy of Sciences, Beijing, 100049, China

³ Beijing Research Center for Preventive Medicine, Beijing Center for Disease Prevention and Control, Beijing, China

⁴ Beijing Municipal Key Laboratory of Clinical Epidemiology, Beijing, China

⁵ School of Data Science, City University of Hong Kong, Hong Kong SAR, China

ARTICLE INFO

Article history:

Received 7 September 2021

Revised 9 January 2022

Accepted 16 January 2022

ABSTRACT

Objectives: The aim of the study was to reconstruct the complete transmission chain of the COVID-19 outbreak in Beijing's Xinfadi Market using data from epidemiological investigations, which contributes to reflecting transmission dynamics and transmission risk factors.

Methods: We set up a transmission model, and the model parameters are estimated from the survey data via Markov chain Monte Carlo sampling. Bayesian data augmentation approaches are used to account for uncertainty in the source of infection, unobserved onset, and infection dates.

Results: The rate of transmission of COVID-19 within households is 9.2%. Older people are more susceptible to infection. The accuracy of our reconstructed transmission chain was 67.26%. In the gathering place of this outbreak, the Beef and Mutton Trading Hall of Xinfadi market, most of the transmission occurs within 20 m, only 19.61% of the transmission occurs over a wider area (>20 m), with an overall average transmission distance of 13.00 m. The deepest transmission generation is 9. In this outbreak, there were 2 abnormally high transmission events.

Conclusions: The statistical method of reconstruction of transmission trees from incomplete epidemic data provides a valuable tool to help understand the complex transmission factors and provides a practical guideline for investigating the characteristics of the development of epidemics and the formulation of control measures.

© 2022 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Introduction

On June 11, the first locally confirmed case of COVID-19 in Beijing, China, was found in Xinfadi Market after 56 consecutive days with no new local confirmed cases. To assess the extent of infection, Beijing Center for Disease Prevention and Control implemented a screening campaign of SARS-CoV-2 infection over the city. Between June 15 and July 10, a total of more than 10 million citizens and 5342 environmental samples were screened. Eventually, 368 Quantitative Real-time Polymerase Chain Reaction (qRT-PCR) positive cases were confirmed (Pang et al., 2020).

The Beijing Center for Disease Control and Prevention conducted an epidemiological investigation on 368 cases and obtained valuable data such as incidence data, contact tracking data, and spatial data of patients. Many studies have been carried out on the basis of this dataset. Some studies have focused on investigating the effect of interventions deployed in Beijing after the outbreak in Xinfadi agricultural products (XFD) market (Cui et al., 2021; Wang et al., 2021; Wei, Guan, Zhao, Shen, & Chen, 2020). Such studies usually use modified susceptible–exposed–infectious–recovered (SEIR) transmission models, depending on the specific scenario. Han (Han et al., 2021) used spatial autocorrelation analysis and Spearman correlation analysis to research the spatial clustering characteristics of the COVID-19 pandemic and the impact of environmental factors in Beijing. In addition, many biological and medical scholars (Chen, Shi, Zhang, Wang, & Sun, 2021; Pang et al., 2020) analyzed the epidemiological characteristics and

* Corresponding author. Zhidong Cao, No. 95, Zhongguancun East Road, Haidian District, Beijing, China. Phone number: +86-10-82544716; fax: 010-82544799.
E-mail address: zhidong.cao@ia.ac.cn (Z. Cao).

clinical manifestations of this outbreak. However, the available transmission link data in the investigation are limited, which cannot restore the complete transmission process. The complete transmission tree offers many potential benefits. In particular, the complete transmission tree can 1) lead to an improved understanding of transmission dynamics and 2) generate valuable intelligence on key epidemiological parameters and risk factors for transmission, which paves the way for more targeted and cost-effective interventions.

Obtaining and reconstructing transmission link data are challenging. First, and most important, transmission dynamics are usually not observed. There is great uncertainty in the course of infection, which is impossible to directly measure an individual's exposure to a potential source of infection (Salje et al., 2016). The great uncertainty in the source of infection has made available link data scarce. Second, many vital points in the course of a patient's illness depend on the patient's recollection (Cauchemez et al., 2011). The recall uncertainty led to the deviation of the data. Third, factors that affect the risk of infection are multiple and complex. They often intertwine features of individuals, for example, age, behavior, mobility, the places they visited, and their social network. This complex intertwinement complicates the reconstruction of the propagation chain modeling.

Depending on the data types, there are 2 most common approaches to infer the propagation chain. The "pairwise approach" is based on onset time and genetic data, which builds a disease transmission model and incorporates a genetic model that describes Pairwise genetic distance between putative transmission pairs (Jombart et al., 2014; Lau, Marion, Streftaris, & Gibson, 2015; Worby et al., 2016). The "phylogenetic approach" uses genetic data to infer the unobserved history of coalescent events between sampled pathogen genomes in the form of a phylogenetic tree. It infers transmission trees consistent with this phylogeny using epidemiological data (Klinkenberg, Backer, Didelot, Colijn, & Wallinga, 2017). But gene sequence data do not always provide spread information of the epidemics. Genetic diversity across most outbreaks is low, and a significant portion of genetic sequences is expected to be identical. The informativeness of genetic sequence data is also limited by complex evolutionary behavior. Gradually, the patient's attributes and behavioral data were used, including symptom onset time, contact tracking data, spatial data, and location data of visits (Campbell, Cori, Ferguson, & Jombart, 2019; Cauchemez et al., 2011).

We take the outbreak of novel coronavirus in Beijing in June 2020 as the case study. We analyzed detailed data describing the second outbreak caused by Beijing Xinfadi. We built a propagation model and used the reliable Bayesian data-augmentation statistical techniques (Cauchemez et al., 2006; Salje et al., 2016) to account for the uncertainty of infection sources, the date of unobserved onset, and infection. Then, we reconstructed the chain of propagation of the outbreak and assessed the influence of spatial distance, family relationship, visiting relationship, and so on, on transmission risk. The accuracy of the propagation chain is also verified.

Methods

Data Collection

The outbreak investigation was performed by the Beijing Center for Diseases Prevention and Control. The research team investigated the cluster of infections caused by the XFD market on June 11, 2020, and July 12, 2020. A total of 368 cases of COVID-19 infection have been reported in this outbreak. There were 335 confirmed cases and 33 asymptomatic cases. There are 14 trading halls in XFD market. The beef and mutton trading hall (referred to as B1 Hall) has been identified as the virus spread from XFD market

(Pang et al., 2020). In addition, the basic personal information and the onset information of each case were also collected. Personal information included sex, age, native place, and address. The information on the occurrence included the type of infection, the dates of onset, the date of diagnosis, the type of exposure to the XFD, and the information of the staff's booth in the XFD. Asymptomatic infections were defined as individuals who have not developed any symptoms but test positive for SARS-CoV-2 by nucleic acid tests.

Statistical Inference

Assuming that individuals with a positive nucleic acid test are infected with SARS-CoV-2, we built a statistical model (Salje et al., 2016) to ascertain risk factors for transmission. In particular, the model was used to estimate the role that the social relation, location, sex, age, and exposed type had on transmission dynamics.

We built a transmission model for the force of infection exerted on individual i at time t (Salje et al., 2016),

$$\lambda_i(t) = \sum_{j:t_j < t}^N \lambda_{j \rightarrow i}(t|x_j, x_i) \tag{1}$$

where $\lambda_{j \rightarrow i}(t|x_j, x_i)$ is the instantaneous hazard of transmission from individual j to individual i at time t :

$$\lambda_{j \rightarrow i}(t|x_j, x_i) = \beta(x_j, x_i) f(t - t_j|x_j, x_i) \tag{2}$$

$\beta(x_j, x_i)$ represents the transmission rate between individuals j and i , where i and j are in the same aggregation relationship. The transmission rate was estimated for 5 types of pairs of individuals.

The set of pairs of individuals was partitioned in 5 types of interaction: "household" (i.e., individuals from the same household or same stall in the XFD market), "colleagues" (i.e., individuals who work for the same company), "the B1 Hall" (i.e., individuals of different stalls in the B1 Hall of the XFD market), "exposed" (i.e., individuals have been exposed together in the same place), and "others" (i.e., pairs of individuals without any of the aforementioned relationships). A hierarchy was set up so that each pair had one and only one type of interaction: household > colleagues > the B1 Hall > exposed > others. For example, a pair of 2 colleagues from the same household was defined as a household pair.

$$\beta(x_j, x_i) = \begin{cases} \beta_h \cdot \beta_{sex}(x_i) \cdot \beta_{age}(x_i), & \text{for household interaction} \\ \beta_c \cdot \beta_{sex}(x_i) \cdot \beta_{age}(x_i), & \text{for colleagues interaction} \\ \beta_b \cdot g(x_i, x_j) \cdot \beta_{sex}(x_i) \cdot \beta_{age}(x_i), & \text{for the B1 Hall interaction} \\ \beta_e \cdot \beta_{sex}(x_i) \cdot \beta_{age}(x_i), & \text{for the exposed interaction} \\ \beta_d \cdot \beta_{sex}(x_i) \cdot \beta_{age}(x_i), & \text{for others interaction} \end{cases} \tag{3}$$

where β_{sex} characterizes the role of sex on risk of infection (male is the reference group), β_{age} characterizes the role of age on risk of infection (individuals' age smaller than 60 years are the reference group), and β_{expose} characterizes the role of expose type on the risk of infection. $g(x_i, x_j)$ characterizes the transmission kernel for individuals in the different aggregation relationships and is a function of the distance. We used an exponential distribution with parameter of a to characterize the transmission kernel.

$f(t - t_j|x_j, x_i)$ represents the infectivity of individual j over time and can be approximated by the generation time distribution (the time between 2 successive infections). The generation time follows Weibull distribution (Shape parameter = 2.826, scale parameter = 5.665) (Ferretti et al., 2020). Details about the transmission model are given in **Supplementary Material**.

Estimation

If the date of infection and date of symptom onset were fully observed for each case, it would be relatively straightforward to

perform likelihood-based inference. However, missing data made the estimation of transmission parameters challenging: (1) dates of infection were unobserved, and (2) 15 symptomatic cases and 33 asymptomatic cases had no observed onset date.

In this study, we used a Bayesian data augmentation framework to tackle the missing data problem (Cauchemez et al., 2011; Cauchemez, Carrat, Viboud, Valleron, & Boelle, 2004; Cauchemez et al., 2006). In the past, this approach has been successfully used to deal with similar problems (Cauchemez et al., 2004; Cori, Boëlle, Thomas, Leung, & Valleron, 2009; Walker et al., 2010). The dataset was “augmented” with missing dates of infection and a few missing dates of symptom onset. If we schematically denote y the observed data, z the augmented data, and θ the parameter vector, the joint posterior distribution of augmented data and model parameters is proportional to

$$P(z, \theta|y) \propto P(y|z)P(z|\theta)P(\theta) \tag{4}$$

This equation shows the hierarchical structure of our Bayesian model. On the righthand side of the equation, $P(y|z)$ is referred to as the “observation model,” which ensures that the augmented data are consistent with observed data. In agreement with a range of studies on COVID-19 pneumonia, the observation model relies on the assumption that the incubation period has a mean of 5 days and a variance of 2 days² (Li et al., 2020; Wang, Horby, Hayden, & Gao, 2020). $P(z|\theta)$ represents the “transmission model,” which describes the latent transmission process and is characterized by Eq. 2. $P(\theta)$ means the prior distribution of the parameters. Please see **Supplementary Material** for more details.

Prior Distributions

For all parameters except for the transmission kernel parameter, we used a lognormal prior distribution with a log(mean) equal to 0 and a log(variance) equal to 1. For the transmission kernel parameter, we used an exponential prior distribution with parameter of 0.0001.

MCMC Sampling Scheme

At every iteration of the Markov Chain Monte Carlo (MCMC) sampling scheme, we undertook the following:

- 1) Metropolis-Hastings update for the parameters in the model. At every iteration, all parameters were updated once. Metropolis-Hastings updates were performed on a log scale with the step size adjusted to achieve an acceptance probability between 20% and 30%.
- 2) Independent sampler for the days of infection. At each iteration, the day of infection was updated with an independent sampler for 50 randomly selected cases. Candidate values for the length of the incubation period were drawn from the incubation period distribution.
- 3) Independent sampler for the unobserved days of onset. For cases with unobserved dates of onset, the augmented date of onset was updated with an independent sampler.

Results

Outbreak Investigation

In June 2021, one of the largest cluster outbreaks of COVID-19 broke out in Beijing, the capital of China. In total, 368 cases of COVID-19 infection were reported in Beijing at 24:00 on June 11, 2020. On the basis of exposure surveys, 79 contacts have been identified for the 368 cases. Among all the cases, 73.9% had direct exposure history (including employees, staff, and visitors), and

Table 1
Parameter estimates

Parameter	Value	95% CI
β_h	9.184%	(8.479%, 9.889%)
β_c	3.166%	(2.956%, 3.376%)
β_b	0.228	(0.190, 0.265)
β_e	0.129%	(0.125%, 0.132%)
β_d	0.082%	(0.078%, 0.085%)
a	0.095	(0.053, 0.138)
ρ_{age}	0.695	(0.676, 0.714)
ρ_{sex}	0.913	(0.892, 0.935)

25.6% were indirectly related cases (having contact history with direct exposure cases or polluted environment) in XFD. Fig. 1A shows the number of cases by date of confirmation for the people with and without XFD exposed history. Most 169 XFD employees worked in the B1 Hall (119 cases, 70.4%). Fig. 1B shows the spatial plane of B1 Hall and the diagnosis. The B1 Hall is the main exposure site. Meanwhile, the resting place of XFD is also at high risk of exposure. Fig. 1C shows the number of cases by date of symptom onset and sex among B1 Hall.

Transmission tree

Transmission model was established, and the transmission risk factors were determined. All individuals with positive nucleic acid tests were included in the analysis as case studies. Data enhancement techniques were used to combine the uncertainty of the onset and unobserved infection’s date. Table 1 shows all the estimated parameters. We estimated the transmission probability between families (the same family and the same stall are considered family relationships), colleagues, different stalls in the B1 Hall, individuals with shared exposure history, and individuals with no apparent relationship. An exponential distribution kernel was used to characterize the distance between different booths in the B1 Hall (i.e., pairs of individuals operating at different booths). We found that there was a 9.2% probability of transmission between family members and the same booth in the B1 Hall (95% CI: 8.5–9.9%) (Fig. 2A), but in the B1 Hall, the probability of transmission from 15 m away was 0.5% (95% CI: 0.2–0.5%), with a probability of 0.02% (95% CI: 0.1–0.2%) at a distance of 50 m (Fig. 2B), indicating that transmission is highly concentrated. Women and men are similarly likely to be infected. Women were less likely to be infected by 0.91 factor (95% CI: 0.89–0.94) (Fig. 2C). The risk of infection in children and adults was of factor 0.70 as in the older individuals (95% CI: 0.68–0.71) (Fig. 2C).

After updating the infection time and correcting the onset time of the model, the incubation period of 100 experimental results was statistically analyzed, and the incubation period distribution obtained is shown in Fig. 2D. The incubation period of 5 days was the most common, accounting for 18.2% of the total cases. The incubation period of 50.2% of cases was less than 5 days, and the incubation period of 20.5% of cases was more than 7 days.

According to the transmission risk between each pair, we reconstructed 100 transmission trees with the highest risk of infection according to the data. The constructed propagation tree links were compared with the identified links obtained through the survey. Of the 79 known links, an average of 52 links of the reconstructed propagation trees matched. The average accuracy was 67.26% (Fig. 3A). Analysis of these trees showed that family transmission accounted for 23.56% of all transmission events (Fig. 3C). Transmission between colleagues outside XFD accounted for 12.17% of all transmission events (95% CI: 18–25%), 91.29% of spread events were related to XFD, and 18.62% of the cases occurred in different booths in the XFD B1 Hall (Fig. 3C). In the B1 Hall of the XFD, most of the transmission occurs within 20 m. Only 19.61% of the trans-

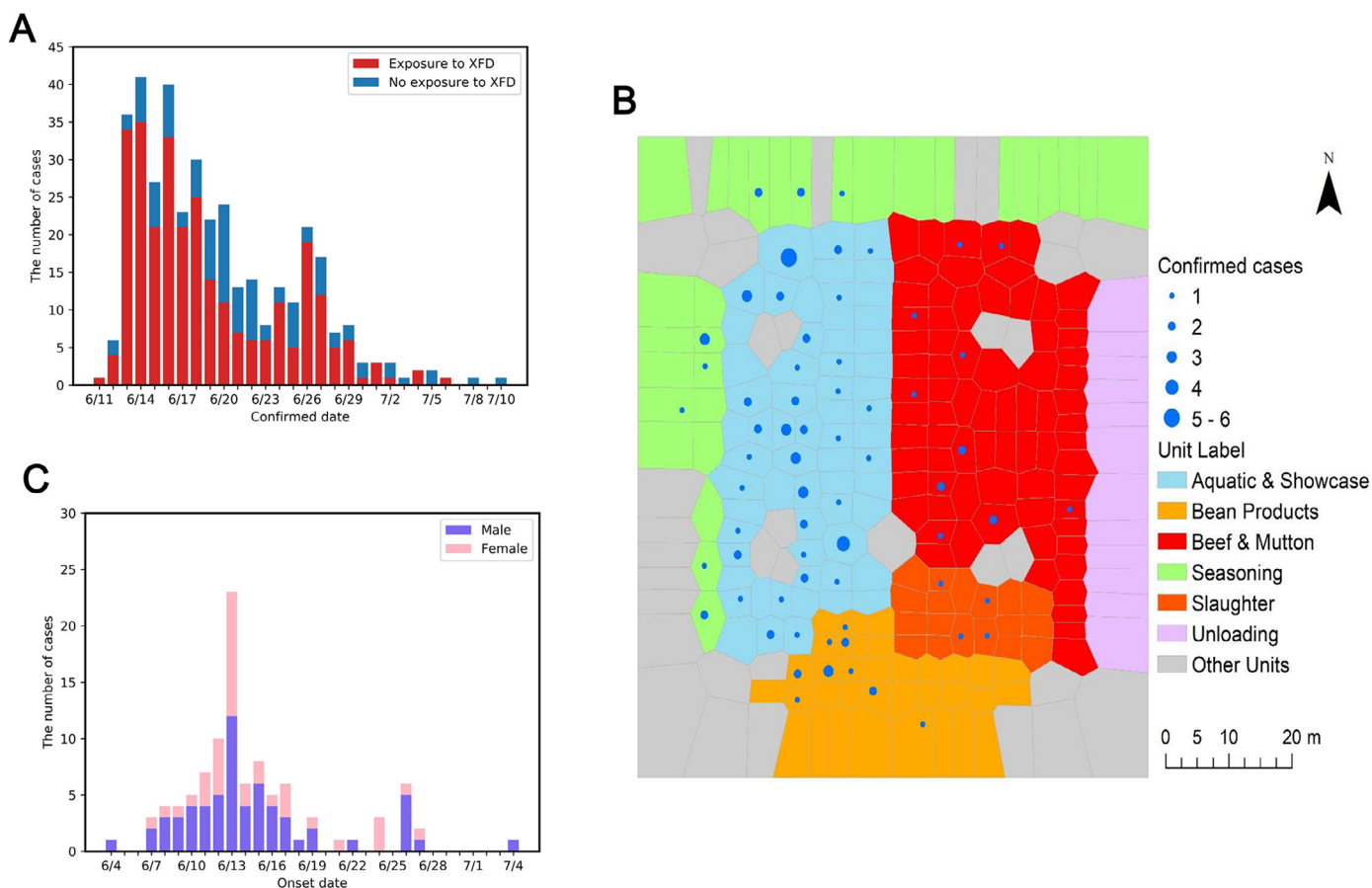


Fig. 1. Epidemiological data collected. (A) Number of cases by date of confirmation for different types of individuals. (B–C) Survey of Xinfadi agricultural products (XFD) market B1 Hall with (B) plane figure of B1 Hall and number of confirmed cases. (C) Number of cases by date of symptom onset and sex among B1 Hall.

mission occurs over a wider area (>20 m) (Fig. 3B), with an overall average transmission distance of 13.00 m.

Fig. 4A shows the complementary cumulative distribution function (CCDF) of the number of secondary infections per infected individual. Of the 368 cases, 48.4% developed the secondary case. Of those with secondary cases, 45.96% had infected only 1 person. Those with more than 5 secondary cases accounted for 3.77% of the total number of cases. The average number of cases with more than 16 secondary cases was only 2. According to the reconstructed transmission tree, the 2 patients were staff at the booth in the B1 Hall of XFD Market. In total, 24.5% of people contributed to 80% of transmission events. Fig. 4B shows the age and sex propagation matrix of this propagation event constructed by the propagation tree. In this outbreak, the male-to-male transmission propagation was 32.06%, and the female-to-female propagation was the lowest (17.28%). We divided 368 cases into 4 age groups: 0~20 years, 20~40 years, 40~60 years, and 60~80 years. As can be seen from the figure, this transmission mainly occurred between ages 20 and 60 years. Fig. 4C shows the cumulative number of infections over time. The average depth of the reconstructed propagation tree is 9.2 (Fig. 4D), that is, the longest propagation generation in this outbreak is 9.2. Fig. 4E is an example of a reconstructed transmission tree.

Discussion

The spread of epidemics is driven by individual social connections and complex interactions between behavior and environment. The reconstruction approach of transmission trees from incomplete epidemic data provides a valuable tool to help trace the

source of the epidemic and understand the complex transmission factors. It provides a theoretical reference for the investigation of the characteristics of the development of epidemics and the development of control measures. In this study, we used the secondary outbreak of novel coronavirus in Beijing, the capital of China, as a case study; built a model to trace the source of transmission; and reconstructed a complete transmission tree. We have shown that the combination of detailed epidemiological data and mathematical models enables us to gain insight into the detailed dynamics of the spread of disease in confined spaces and more broadly in cities. The individual characteristics (such as age) and the geographical space, particularly in densely populated disease outbreak areas, have an essential influence on infection risk. This finding underscores the importance of considering specific sites when assessing the spread of the epidemic.

The study illustrates challenges epidemiologists face in studying the spread of infectious diseases. During an outbreak investigation, the route of transmission or date of infection is often not recorded, and cases are usually incorrectly recorded. The proposed data enhancement strategies can correctly interpret these uncertainties in the reasoning framework and thus significantly improve our ability to analyze epidemic data robustly. In this study, the joint posterior distribution of model parameters and augmented data is studied by Markov chain Monte Carlo sampling from the perspective of Bayesian. The data enhancement technology under Bayesian framework has a hierarchical structure: (1) the observation level ensured that the augmented data were consistent with the observed data, (2) the transmission level described the underlying epidemic process, and (3) the prior level specified the distribution of the parameters. This approach is a general framework that applies not

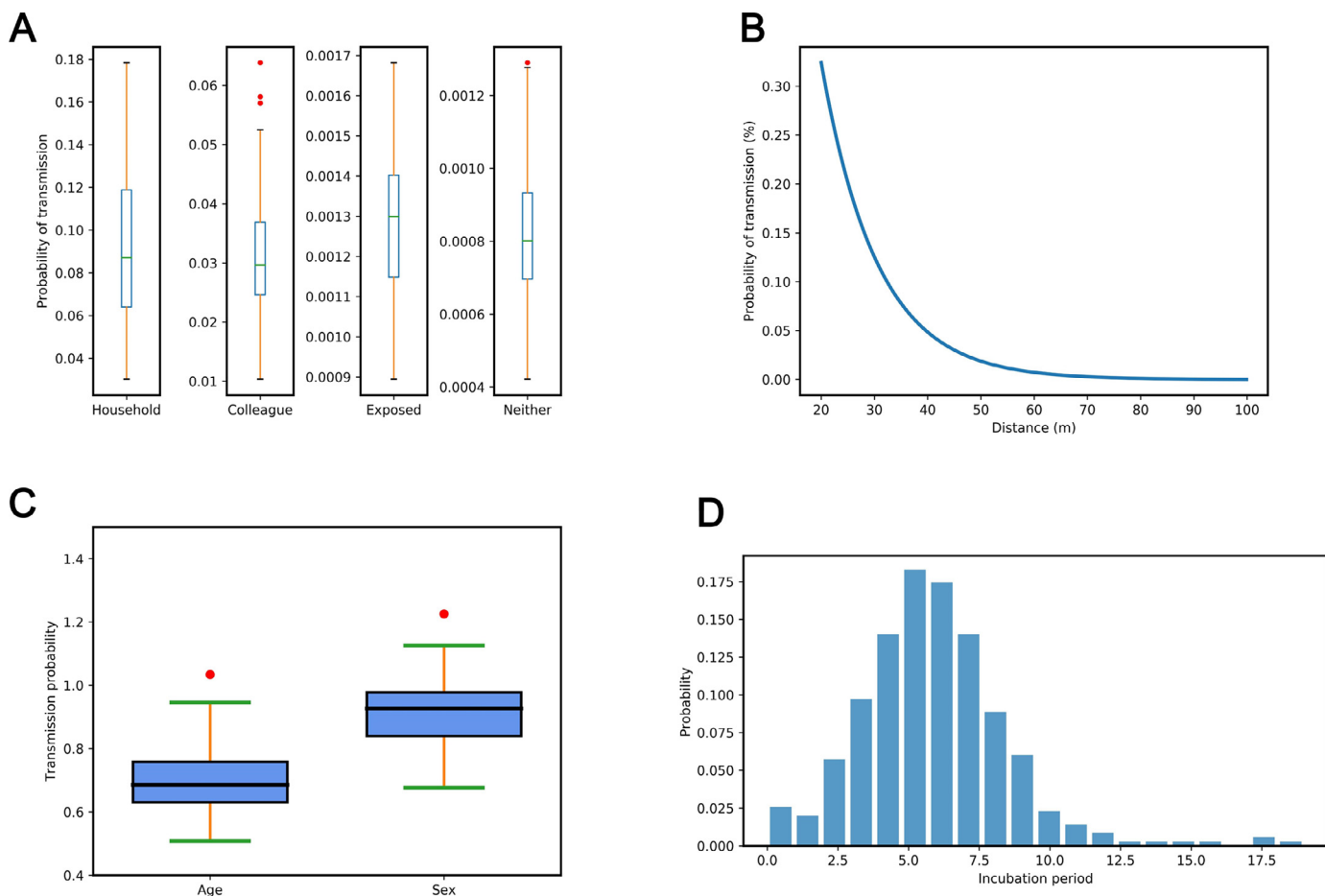


Fig. 2. Parameter estimates from transmission model. (A) Probability of transmission of the 2 individuals under different relationships (B) Transmission kernel. (C) Relative susceptibility for adults versus old people (those more than 60 years) (baseline) and females versus males (baseline). (D) The incubation period distribution estimated by the transmission model.

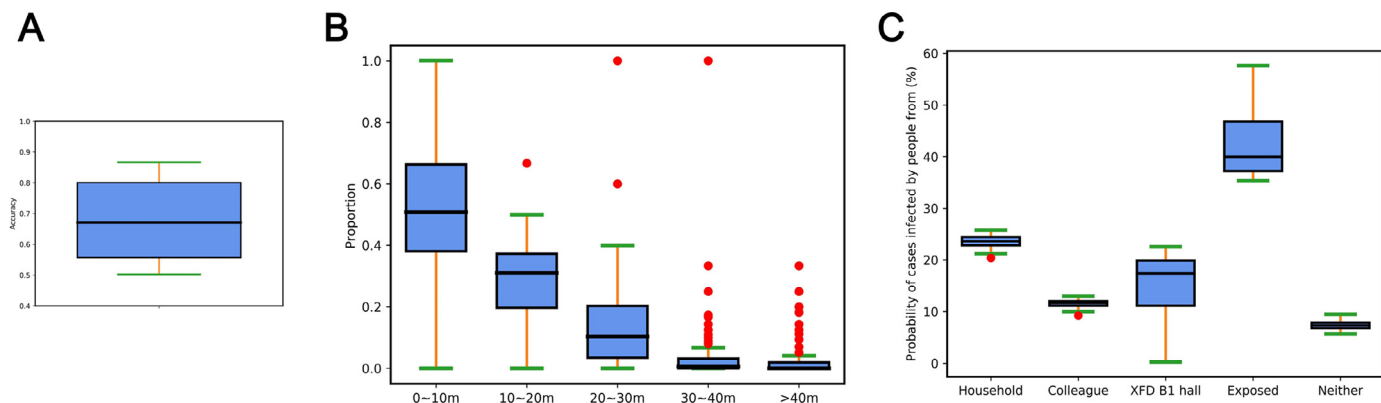


Fig. 3. In total, 100 transmission trees were reconstructed according to the experimental results. (A) Accuracy of the transmission link. (B) The proportion of transmission events at different distances in Xinfadi agricultural products (XFD) B1 Hall. (C) The proportion of cases infected by people from household, colleague, XFD B1 Hall, or exposed space or from the unknown relationship.

only to COVID-19 but also can be widely applied to extrapolate the transmission characteristics of other infectious diseases from missing data. The details of this approach can be designed for different scenarios. In our research, the collection of precise location data greatly assists in investigating the outbreak. This helps us consider the heterogeneity of individual spatial distances in the model framework rather than placing all individuals in a uniform space. Through the reconstruction of the transmission tree, we fully analyze the characteristics of this outbreak from the perspective of in-

dividuals. The inference of the complete transmission tree can improve our understanding of the transmission characteristics of the epidemic and grasp the risk factors of the disease. This provides insights into more targeted and cost-effective interventions.

The XFD outbreak can be broadly divided into 2 stages. The first stage was from May 28 to June 12. During this period, the epidemic started from the B1 Hall of the XFD Market and quickly spread among employees, visitors, and family members. Then the epidemic gradually spread to other regions of the XFD market and

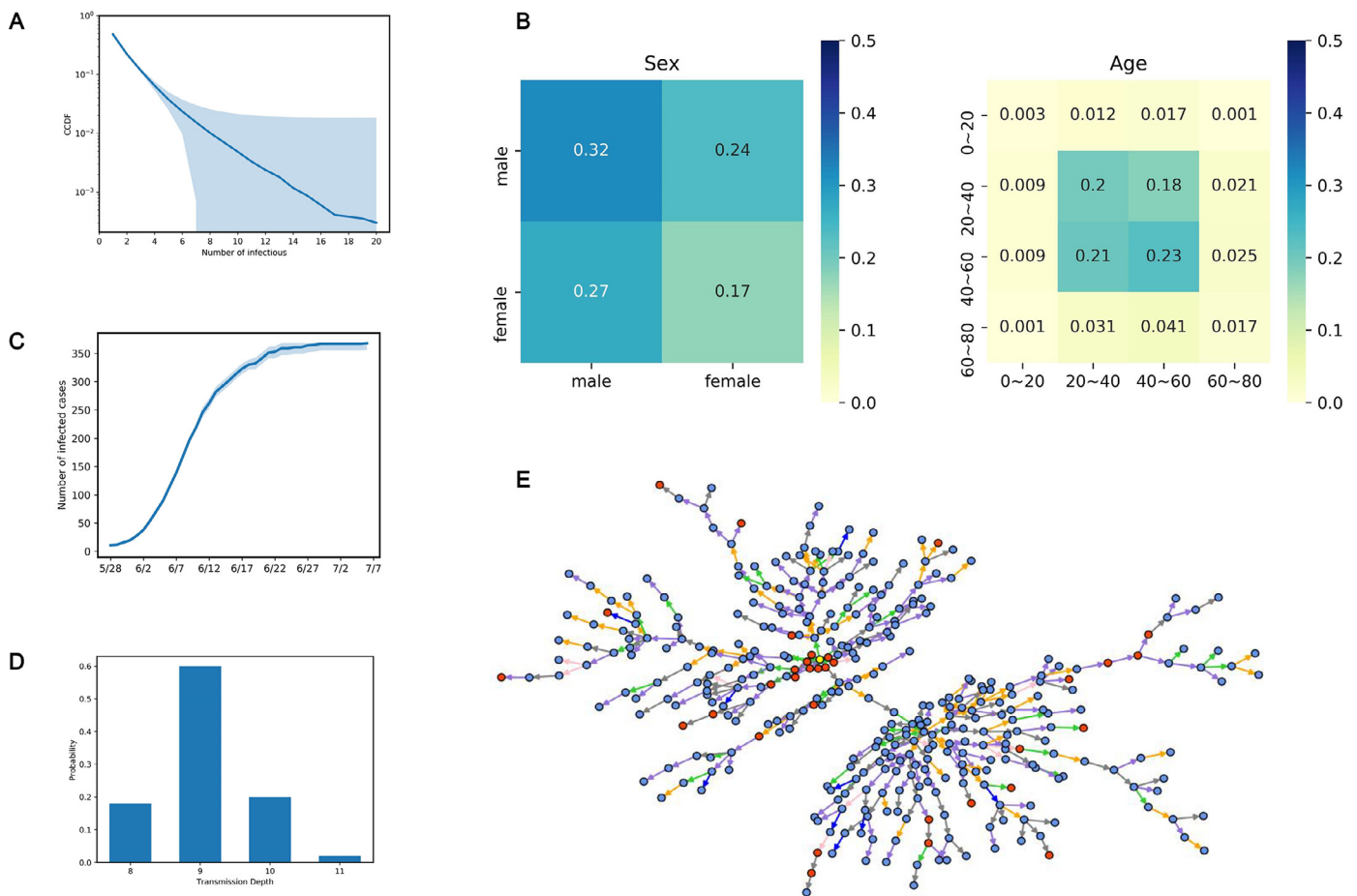


Fig. 4. Reconstruction of the transmission tree. (A) Complementary cumulative distribution function (CCDF) of the number of secondary infections per infected individual. (B) The proportion of transmission by sex and age group. (C) The cumulative number of infections. (D) The distribution of the depth of transmission tree. (E) An example of a reconstructed transmission tree.

spread to the population outside the market in a short time, giving to rise cases. The second stage was after June 12, when the first case was confirmed, and the XFD market was closed. The transmission rate in XFD market dropped significantly, whereas the spread from outside the market to the city was continuous. Timely isolation measures and blockade of the source have reduced the scale of the epidemic and brought it under effective control.

After constructing the entire transmission tree, the conclusion can be drawn that this outbreak is made up of multiple clustered propagation events. The spread of the virus is largely driven by the movement of people in XFD market. The virus then continues to spread in the households, companies, and other public places. The B1 Hall of the XFD market hall is where the outbreak originated and gathered. The transmission rate within this space is closely related to the distance between booths. When the booth distance exceeds 40 m, the transmission probability becomes very small. In this transmission event, only 2.83% of the transmission exceeded 40 m. Transmissions between booths of more than 40 m depend mainly on the movement of people. Therefore, during the epidemic, vendors should pay special attention to the prevention and control of nearby stalls and mobile visitors.

Household transmission is a typical cluster transmission, and the transmission probability between family members is the highest among the several relationships we set. Family transmission events accounted for approximately 23.67% of the transmission events, second only to the number of jointly exposed infectious events in XFD market. In family transmission, the most common is between ages 20 and 60 years. Older people are significantly

more susceptible than young people. Therefore, family prevention and control and priority protection measures for the older people should be paid more attention to.

Two abnormally high propagation events occurred during this propagation. One of the events source was an employee of the booth where the infection originated, and the other events source was also an employee of the B1 Hall, an asymptomatic infected person. These two cases infected up to 10% of the transmission events. According to the investigation and analysis, the asymptomatic patient had been infected as early as June 7, and the nucleic acid positive was detected only on June 26. This shows that asymptomatic patients still have a strong ability to infect in the asymptomatic stage. Because there are no significant symptoms, the early prevention and control of asymptomatic patients are difficult.

In general, we reconstructed the complete transmission chain of the second outbreak of COVID-19 in Beijing through Bayesian data enhancement and unveiled the transmission characteristics of this epidemic through the analysis of the transmission chain. The conclusions from our study can guide the design of more targeted and sustainable mitigation strategies. Our reconstructed transmission models will help analyze risk factors for outbreaks and help calibrate future modeling efforts.

Funding sources

This work was supported in part by the National Natural Science Foundation of China [grant numbers 72025404, 71621002,

and 72074209], the New Generation Artificial Intelligence Development Plan of China (2015–2030) [grant number 2021ZD0111205], Beijing Natural Science Foundation [grant number L192012] and Beijing Nova Program [grant number Z201100006820085].

Conflict of interest statement

All authors have no conflict of interests to declare.

Ethical approval

The survey was discussed with the Beijing Center for Diseases Prevention and Control and the Institute of Automation Chinese Academy of Sciences, who reviewed the content and advised that it was exempt from the ethics committee review. This study was considered a continuation of the public health investigation associated with an emerging infectious disease.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ijid.2022.01.035](https://doi.org/10.1016/j.ijid.2022.01.035).

References

- Campbell F, Cori A, Ferguson N, Jombart T. Bayesian inference of transmission chains using timing of symptoms, pathogen genomes and contact data. *PLoS computational biology* 2019;15(3).
- Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, Group PHNW. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proceedings of the National Academy of Sciences* 2011;108(7):2825–30.
- Cauchemez S, Carrat F, Viboud C, Valleron A, Boelle P. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in medicine* 2004;23(22):3469–87.
- Cauchemez S, Temime L, Guillemot D, Varon E, Valleron A-J, Thomas G, Boëlle P-Y. Investigating heterogeneity in pneumococcal transmission: a Bayesian MCMC approach applied to a follow-up of schools. *Journal of the American Statistical Association* 2006;101(475):946–58.
- Chen H, Shi L, Zhang Y, Wang X, Sun G. The re-emergence from the COVID-19 epidemic of Beijing Xinfadi Market. *Medicine* 2021;100(31).
- Cori A, Boëlle P-Y, Thomas G, Leung GM, Valleron A-J. Temporal variability and social heterogeneity in disease transmission: the case of SARS in Hong Kong. *PLoS computational biology* 2009;5(8).
- Cui X, Zhao L, Zhou Y, Lin X, Ye R, Ma K, Shi H. Transmission dynamics and the effects of non-pharmaceutical interventions in the COVID-19 outbreak resurged in Beijing, China: a descriptive and modelling study. *BMJ open* 2021;11(9).
- Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Fraser C. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *science* 2020;6491:368.
- Han Y, Yang L, Jia K, Li J, Feng S, Chen W, Pereira P. Spatial distribution characteristics of the COVID-19 pandemic in Beijing and its relationship with environmental factors. *Science of The Total Environment* 2021;761.
- Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS computational biology* 2014;10(1).
- Klinkenberg D, Backer JA, Didelot X, Colijn C, Wallinga J. Simultaneous inference of phylogenetic and transmission trees in infectious disease outbreaks. *PLoS computational biology* 2017;13(5).
- Lau MS, Marion G, Streftaris G, Gibson G. A systematic Bayesian integration of epidemiological and genetic data. *PLoS computational biology* 2015;11(11).
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Wong JY. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England journal of medicine*; 2020.
- Pang X, Ren L, Wu S, Ma W, Yang J, Di L, Du S. Cold-chain food contamination as the possible origin of COVID-19 resurgence in Beijing. *National Science Review* 2020;7(12):1861–4.
- Salje H, Lessler J, Paul KK, Azman AS, Rahman MW, Rahman M, Cauchemez S. How social structures, space, and behaviors shape the spread of infectious diseases using chikungunya as a case study. *Proceedings of the National Academy of Sciences* 2016;113(47):13420–5.
- Walker PG, Cauchemez S, Métras R, Dung DH, Pfeiffer D, Ghani AC. A Bayesian approach to quantifying the effects of mass poultry vaccination upon the spatial and temporal dynamics of H5N1 in Northern Vietnam. *PLoS computational biology* 2010;6(2).
- Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *The lancet* 2020;395:470–3 (10223).
- Wang X, Lin X, Yang P, Wu Z-Y, Li G, McGoogan JM, Shi H-H. Coronavirus disease 2019 outbreak in Beijing's Xinfadi Market, China: a modeling study to inform future resurgence response. *Infectious diseases of poverty* 2021;10(1):1–10.
- Wei Y, Guan J, Zhao Y, Shen S, Chen F. Inference of start time of resurgent COVID-19 epidemic in Beijing with SEIR dynamics model and evaluation of control measure effect. *Zhonghua liu Xing Bing xue za zhi= Zhonghua Liuxingbingxue Zazhi* 2020;41 E077-E077.
- Worby CJ, O'Neill PD, Kyraios T, Robotham JV, De Angelis D, Cartwright EJ, Cooper BS. Reconstructing transmission trees for communicable diseases using densely sampled genetic data. *The annals of applied statistics* 2016;10(1):395.