

Transmission characteristics of the COVID-19 outbreak in China: a study driven by data

Meili Li

Donghua University

Pian Chen

Donghua University

Qianqian Yuan

Donghua University

Baojun Song

"Montclair State University"

Junling Ma (✉ junlingm@uvic.ca)

Universidad Victoria <https://orcid.org/0000-0002-0197-2317>

Research article

Keywords: Asymptomatic transmission, Reproduction number, Incubation period, Exponential growth rate

Posted Date: June 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-25738/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

The COVID-19 outbreak has been a serious public health threat worldwide. Individually documented case descriptions of COVID-19 are published by Chinese provinces (excluding Hubei). We use these descriptions to study the transmission characteristics in China, and how they are influenced by public awareness and control measures.

Methods

Dates for infection, symptom onset, quarantine, hospitalization and diagnosis, and sources of infection are tabulated from published cases descriptions. We use MCMC to estimate the exponential growth rate of cases infected in Hubei, and parametrize the distributions for the incubation period, and periods from symptom onset to hospitalization and diagnosis. Graph of infection is constructed and used to estimate the reproduction number in each generation.

Results

The median incubation period is 6.2 days overall, and 7.3% patients have an incubation period longer than the recommended 14-day quarantine period. The median period from symptom onset to hospitalization is 4.4 days before the lockdown of Wuhan city, and 2.1 days after the lockdown. The median period from symptom onset to diagnosis is 8.8 days before the lockdown, and 4.8 days after the lockdown. The number of cases in Hubei doubled every 3 days before the lockdown. In other provinces, reproduction number decreased from 1.51 in the first-generation patients to less than 0.21 in later generation patients.

Conclusions

We parametrized incubation period, and the periods from symptom onset to hospitalization and diagnosis. Some (possibly a majority of) patients were infected before the symptom onset of their sources of infection, causing a low success rate of official (not self) quarantines. Social distancing and self-quarantine greatly reduced the reproduction number, and shorted the periods from symptom onset to hospitalization and diagnosis.

Background

The COVID-19 outbreaks around the world have become a major global public health threat [1–3]. Many estimates of the scope and magnitude of the outbreak have been quickly published [4–7] using both estimations from epidemic curves and individual level data of a single region. For example, a crucial

parameter of an outbreak is the reproduction number, which is the average number of secondary infections caused by an average infectious individual. The daily number of new infections will decline if it is less than unity. The reproduction number of a patient in a fully susceptible population is called the basic reproduction number. Not surprisingly, this parameter received much attention, and has been estimated to be between 1.32 and 6.47 [8–15]. The incubation period can only be estimated reliably from individual level data and has been estimated to be between 5 and 9 days [9, 11, 16–19]. But both approaches have limitations: many disease parameters cannot be reliably estimated from epidemic curves only [20], while most of the studies using individual data have limited sample sizes.

Each province except Hubei officially published the daily update of case descriptions of COVID-19. We use these individually documented case descriptions to conduct our research. The documented cases and their data sources are tabulated in Supplementary Material S1. Some of these descriptions are well documented with important epidemiological information including both the source of infection and dates such as contacting the source of infection, symptom onset, quarantine, hospitalization and diagnosis.

This is a much more complete dataset than ones used in previous studies. We aim to use it to give a more precise parametrization of key epidemic parameters such as the incubation period, and the periods from symptom onset to hospitalization and diagnosis. We will also establish a graph of transmission and use it to estimate the reproduction number in each generation of transmission. We will study the changes in reproduction numbers and in the periods from symptom onset to hospitalization and diagnosis to estimate the effect of social distancing and self-quarantine measures that has been strictly implemented in China.

Case descriptions from many provinces also include the information whether a patient hospitalized in these provinces had visited (or is a resident of) Hubei province before symptom onset. We use this information to estimate the exponential growth rate of the Hubei outbreak before strict control measures are implemented in Wuhan city.

Our study will help to understand why the Chinese outbreaks in their provinces except the epicenter can be controlled so effectively. It will also help mathematical modelers to calibrate their epidemic models and make more precise predictions in the scenarios of their interest.

Methods

Data Our research is based on 3547 individually documented case descriptions published by provinces of China (excluding Hubei province that are publicly available. We have not participated in original data collection from patients. Each case description includes some or all of the dates of contact, symptom onset, quarantine, hospitalization and diagnosis. Furthermore, 331 descriptions contain source of infection. For the latest possible date of being infected, we also consider the constraints that the date cannot be later than the date of symptom onset and cannot be later than the hospitalization date of the source of infection. Supplementary Materials S1 contains the tabulated data and the data sources for

each case description. The numbers of case descriptions that are valid for the estimations are listed in Table 1.

Parametrization of period distributions The incubation period distribution is estimated from the case descriptions with dates of infection and symptom onset. Similarly, the periods from symptom onset to hospitalization and diagnosis are estimated from the case descriptions with the corresponding date information. The periods are parametrized as continuous variables with a gamma, a log-normal, or a Weibull distribution. The observations are given in either days or ranges of days, which are converted into continuous time periods. We then compute the probability that the period falls into the observed ranges. We can then construct the following log-likelihood function

[Please see the supplementary files section to view the equation.]

Reproduction Number We construct the graph of transmissions from the 643 descriptions with source of infection, where nodes represent patients and arcs represent transmissions. Some of the sources may have multiple possible sources. We assume that all sources have equal probability to be the true source; this probability is labeled as the weight of the arcs. The number of secondary infections of a patient is the sum of the weights on its outgoing arcs. We then compute the average of secondary infections over the patients.

We label the individuals by generations of infection, to study how reproduction number changes with generation. However, due to the uncertainty in the source of infection of some patients, the generations of these patients may be probabilistic, with probabilities computed from normalizing the weights of their incoming arcs.

Exponential growth rate in Hubei We tabulate the daily number of patients for patients who either visited Hubei or came from Hubei, with symptom onsets on or before the lockdown of Wuhan on January 23. We assume that these case counts are proportional to the cases in Hubei, and are independently Poisson distributed with a mean growing exponentially, μe^{rt} , and construct the following log-likelihood function

[Please see the supplementary files section to view the equation.]

where n_t is the observed number of patients who either visited or is a resident of Hubei showing symptom on day t . We then use MCMC to estimate the exponential growth rate r .

Results

Asymptomatic transmission Among the 98 patients whose case description contains source of infection and dates of contacting the source, and whose source also has date of contacting their sources, 24 were infected before their sources of contacts showed major symptoms, while 5 were infected after their sources showed symptom. The other 69 patients were infected during a period that contains the date of symptom onset of their sources, and thus are inconclusive.

Incubation period Figure 1 shows the observed and best-fit incubation period. The best-fit distribution is a gamma distribution, which parameters are summarized in Table 2. Females has a slightly longer median incubation period; as shown in Figure 2, the probability mass function of the incubation period for males is larger in the first four days, smaller for days 9-12. The fraction of patients who have an incubation period longer than a given period is shown in Figure 3. For the recommended 14-day quarantine period, this fraction is 7.3% (5.8%, 9.0%).

Periods from symptom onset to hospitalization and diagnosis The observed and best-fit distributions for the periods from symptom onset to hospitalization and diagnosis are shown in Figure 4 and 5. The parameters are summarized in Tables 3 and 4. We also compared the periods for individuals who showed symptom before and after the lockdown of Wuhan on January 23, because the lockdown is an extreme control measure that greatly increased the awareness of the outbreak, which led to both behavior changes and the strengthen of control measures in other provinces. The results show that the periods reduced significantly after the lockdown. The successfully quarantined cases (quarantined before the date of symptom onset) are shown in the upper panel of Figure 4 as negative periods. Only 66 cases were successfully quarantined, with the first successful case quarantined on January 22, one day before the lockdown of Wuhan. These cases represent an overall 4.3% success rate, or 5.3% among patients who showed symptom after (and 0% for those before) the lockdown of Wuhan. Self-quarantines are not counted here because they may still lead to within-household transmissions.

Reproduction Number The graph of transmissions is shown in Figure 6, with patients color-coded by their provinces. The distribution of the generation of these patients and the reproduction number of patients in each generation are shown in Figure 7. Overall, the reproduction number of all patients is $R=0.65$, and that of patients who infected others is $R=1.63$. The reproduction number reduces from $R=1.51$ for the first-generation patients in provinces other than Hubei, to less than 0.21 in second and later generation patients. For those who infected others, the reproduction number remained between 1.42 and 1.68 within the first four generations before it decreases to 1.25 in the seventh generation.

Exponential growth rate in Hubei The daily counts of symptom onsets for patients who have visited (or are residents of) Hubei are shown in Figure 8. The counts decrease after January 23 due to the lockdown of Wuhan. The estimated exponential growth rate for these patients who showed symptom on or before January 23 is 0.23 (0.21, 0.26) per days. The best-fit exponential growth curve is also shown in Figure 8.

Discussions

More patients are observed to be infected before the symptom onset of their source of infections, than those infected after the symptom onset of their sources, while a majority are ambiguous due to their long duration of contact with their sources (e.g., in workplace or household settings). Thus, asymptomatic infections may be an important factor driving the outbreaks, and may greatly reduce the effectiveness of contact tracing and quarantine. This is also demonstrated by the 5.3% success rate for quarantine even after the lockdown of Wuhan.

The patients were hospitalized and diagnosed much faster after they showed symptoms. These changes are most likely due to both behaviors change driven by increased awareness and faster public health responses.

Social distancing prevented a majority of patients from infecting others. However, for the patients who did, social distancing did not reduce their secondary infections drastically. It is very likely that a majority of these transmission occur in households, and social distancing is not effective in reducing household transmissions. Instead, it greatly reduced the probability that a patient infects another household. Thus, for mathematical modeling of disease dynamics of social distancing, considering household structures may be important.

Conclusions

In the data that we tabulated suggest that asymptomatic transmission may occur more frequently than asymptomatic transmission in Chinese provinces other than Hubei, where contact tracing and social distancing are strongly implemented. Thus, under strong control measures, asymptomatic transmission may be a crucial factor driving the outbreaks.

We parameterized the incubation period from the data, which suggests that 7.3% of patients have an incubation period longer than the recommended 14-day quarantine period. These patients may be especially risky for quarantine failures and subsequently cause community transmissions. Males on average have a slightly shorter incubation period, i.e., we expect slightly more males to show symptoms within the first 4 days, and slightly more females to show symptoms in 9–12 days.

The lockdown of Wuhan, the social distancing, and the self-quarantine measures had a major impact on the control of the outbreaks in Chinese provinces other than Hubei. The times from symptom onset to hospitalization and diagnosis were greatly reduced after the lockdown. In addition, the reproduction number of patients was reduced by 82.6% after the first generation.

Due to strictly implemented contact tracing and social distancing measures in other provinces than Hubei, patients who have visited (or are residents of) Hubei and are hospitalized in other provinces form a reliable sample of the Hubei outbreak for the period up to the lockdown of Wuhan. We estimated that the number of patients doubled every 3 days before strict control measures are implemented.

The case descriptions published by Chinese provinces contain sources of infections and dates of contacting the sources, quarantine, symptom onset, hospitalization, and diagnosis, in addition to gender and age of the patients. They are valuable source of information for the COVID-19 outbreak in China. This is the first effort that tabulates and aggregates the data, even though the data sources are publicly available. Our effort allows more detailed analysis of the Chinese outbreak in the future.

Abbreviations

MCMC
Markov Chain Monte Carlo
DIC
Deviance Information Criterion

Declarations

Ethics approval and consent to participate

Not applicable. The case descriptions are published by Chinese provinces, and are publicly available. We have not participated in original data collection from patients.

Consent for publication

Not applicable.

Availability of data and materials

Our tabulated data in Supplementary material S1 are publicly available. S1 contains the data sources that we used to tabulate the data. The data sources are published by Chinese provinces, and are publicly available.

Competing interests

The authors declare that they have no competing interests.

Funding

This research is supported by National Natural Science Foundation of China (No. 11771075) (ML), State Scholarship Fund of China (CSC No. 201906635011) (ML), and a Fundamental Research Grant for Chinese Universities (ML) in data collection, analysis and interpretation of data and writing the manuscript, and a Natural Sciences and Engineering Research Council Canada Discovery Grant (JM) in the design of the study, analysis and interpretation of data and writing the manuscript.

Author's contributions

JM designed the study. ML, PC and QY tabulated the data. All authors contributed equally in analysis. ML wrote the first draft, authors contributed equally in editing. All authors have read and approved the manuscript.

Acknowledgments

ML's work is carried out during the period of working as a visiting scholar at Montclair State University.

References

1. National Health Commission of the People's Republic of China. Briefing of the epidemic, 2020. http://www.nhc.gov.cn/xcs/xxgzbd/gzbd_index.shtml.
2. World Health Organization. Coronavirus disease (COVID-2019) situation reports, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>.
3. World Health Organization. Novel coronavirus, 2020. https://www.who.int/csr/don/archive/disease/novel_coronavirus/en/.
4. T.-M. Chen, J. Rui, Q.-P. Wang, et al. A mathematical model for simulation the phase-based transmissibility of a novel coronavirus. *Infect. Dis. Poverty*, 9:24, 2020.
5. J. Li. A robust stochastic method of estimating the transmission potential of 2019-nCoV. 2020. <https://arxiv.org/abs/2002.03828>.
6. H. Tian, Y. Liu, Y. Li, et al. An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china. *Science*, 2020. doi: 10.1126/science.abb6105.
7. H. Wang, Z. Wang, Y. Dong, et al. Estimating the number of 2019 novel coronavirus cases in Chinese Mainland. 2020. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3529449.
8. Z. Du, X. Xu, Y. Wu, et al. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg. Infect. Dis.*, 2020. <https://doi.org/10.3201/eid2606.200357>.
9. Q. Li, X. Guan, P. Wu, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*, 382:1199-1207, 2020.
10. T. Liu, J. Hu, J. Xiao, et al. Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China. 2020. <https://www.biorxiv.org/content/10.1101/2020.01.25.919787v2>.
11. S. Ma, J. Zhang, M. Zeng, et al. Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. 2020. <https://www.medrxiv.org/content/medrxiv/early/2020/03/24/2020.03.21.20040329.full.pdf>.
12. J. M. Read, J. R. E. Bridgen, D. A. T. Cummings, et al. 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. 2020. <https://www.medrxiv.org/content/10.1101/2020.01.23.20018549v2>.
13. E. Shim, A. Tariq, W. Choi, et al. Transmission potential and severity of COVID-19 in South Korea. *International Journal of Infectious Diseases*, 93:339-344, 2020.
14. B. Tang, X. Wang, Q. Li, et al. Estimation of transmission risk of 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine*, 9:462, 2020.

15. J. T. Wu, K. Leung, and G. M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, 395:689-697, 2020.
16. J. A. Backer, D. Klinkenberg, and J. Wallinga. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill*, 25(5):2000062, 2020.
17. C. Leung. The difference in the incubation period of 2019 novel coronavirus (SARS-CoV-2) infection between travelers to {Hubei} and nontravelers: The need for a longer quarantine period. *Infection Control & Hospital Epidemiology*. 2020. doi: 10.1017/ice.2020.81
18. N. M. Linton, T. Kobayashi, Y. Yang, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J. Clin. Med.*, 9:538, 2020.
19. L. Tindale, M. Coombe, J. E. Stockdale, et al. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. 2020.
<https://www.medrxiv.org/content/10.1101/2020.03.03.20029983v1>.
20. N. Tuncer and T. T. Le. Structural and practical identifiability analysis of outbreak models. *Mathematical Biosciences*, 299:1-18, 2018.

Tables

Table 1. The number of valid cases descriptions (i.e., containing information necessary to compute the parameters) that are available and used in parameterization.

Parameter	Valid case descriptions
Incubation period	646 (316 female, 330 males)
Period from onset to hospitalization	1452 (300 with onset before and 1152 after Jan 23)
Period from onset to diagnosis	2747 (550 with onset before and 2181 after Jan 23)
Reproduction number	643

Table 2. The best-fit distribution of the incubation period and their parameters (with 95% confidence intervals in parentheses). The ranges containing the period for 95% patients and the medians are in days.

Patients	Distribution	Median	Shape	Scale	Range
All	Gamma	6.2 (5.8, 6.6)	2.65 (2.36, 3.04)	2.67 (2.32, 2.97)	1.2 - 17.3
Males	Gamma	5.9 (5.4, 6.4)	2.45 (2.05, 2.96)	2.80 (2.29, 3.35)	0.9 - 17.9
Females	Gamma	6.5 (6.0, 7.1)	2.91 (2.46, 3.52)	2.54 (2.08, 2.95)	1.7 - 19.2

Table 3. The parameters of the best-fit distribution of the period from symptom onset to hospitalization, for all patients, those whose symptom onset is before the lockdown of Wuhan on January 23, and those on or after, with 95% confidence intervals in parentheses. The ranges containing the period for 95% patients and the medians are in days.

Patient	Distribution	Median	Shape	Scale	Range
All	Gamma	2.5 (2.3, 2.6)	0.85 (0.78, 0.92)	4.53 (4.12, 5.00)	0 - 12.7
Before Jan 23	Weibull	4.4 (4.0, 4.9)	1.26 (1.14, 1.38)	5.94 (5.40, 6.52)	0 - 13.4
After Jan 23	Gamma	2.1 (1.9, 2.3)	0.77 (0.70, 0.85)	4.39 (3.91, 4.94)	0 - 11.0

Table 4. The parameters of the best-fit distribution of the period from symptom onset to diagnosis, for all patients, those whose symptom onset is before the lockdown of Wuhan on January 23, and those on or after, with 95% confidence intervals in parentheses. The ranges containing the period for 95% patients and the medians are in days.

Patient	Distribution	Median	Shape	Scale	Range
All	Gamma	5.5 (5.4, 5.7)	2.27 (2.15, 2.39)	2.84 (2.68, 3.00)	1.0 - 17.4
Before Jan 23	Gamma	8.8 (8.5, 9.2)	4.53 (4.03, 4.06)	2.11 (1.87, 2.38)	3.0 - 19.8
After Jan 23	Gamma	4.8 (4.6, 4.9)	2.21 (2.11, 2.29)	2.56(2.50, 2.68)	0.8 - 15.5

Figures

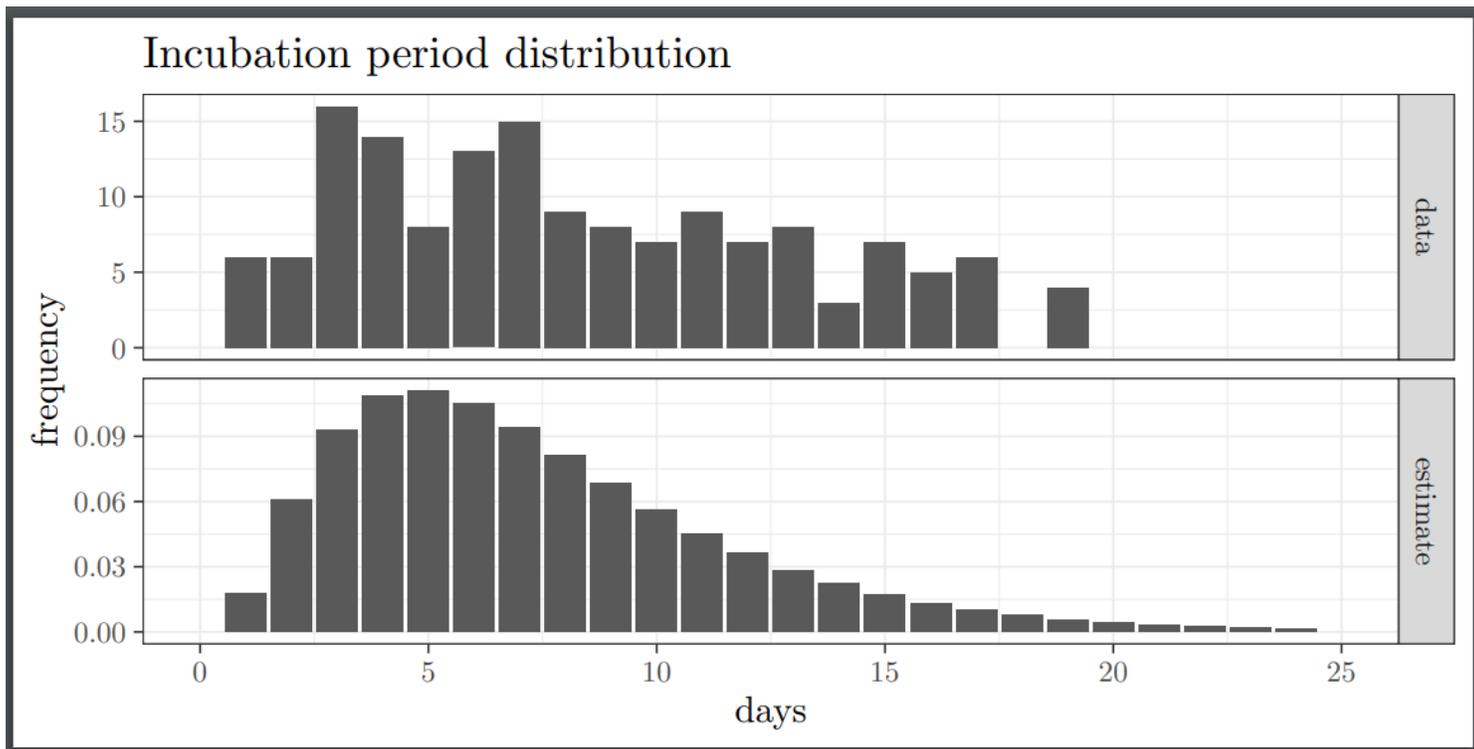


Figure 1

The observed and best-fit incubation period distribution. The data panel shows 157 patients whose date of being infected is not ambiguous (i.e., given as a single date). The best-fit distribution is computed from all 646 patients. The estimate panel shows the discretized best-fit distribution, computed from the average of 10 thousand MCMC samples.

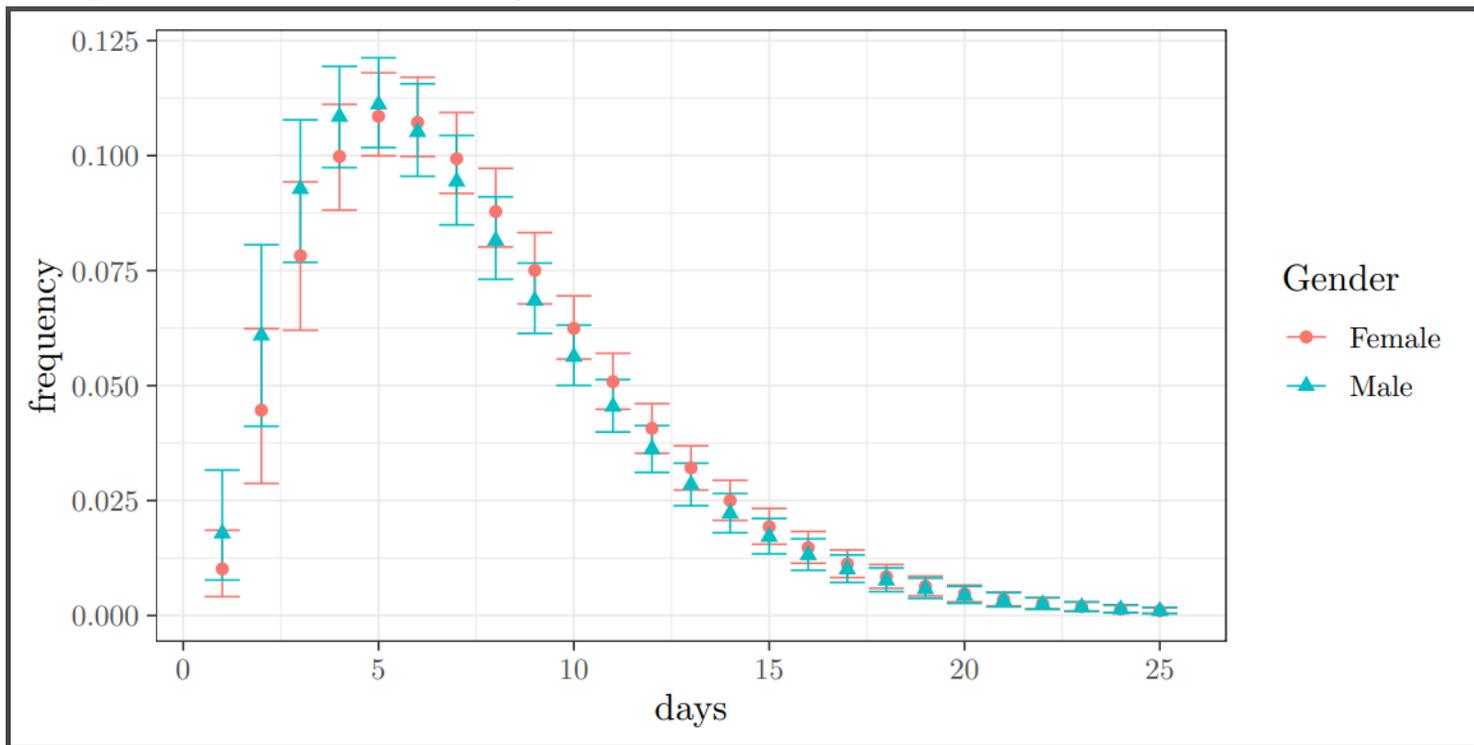


Figure 2

The best-fit incubation period distributions (dots) and the 95% confidence intervals (bars) by gender.

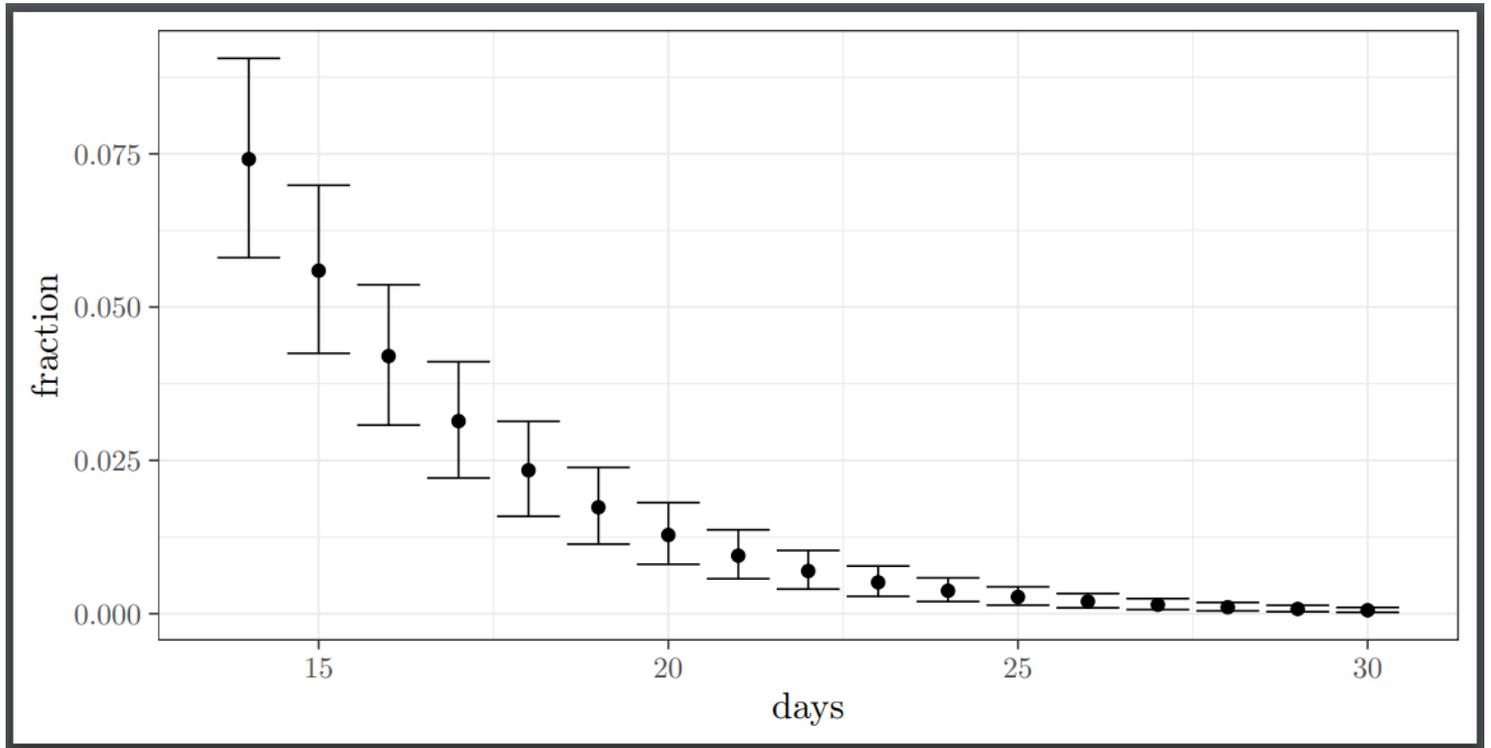


Figure 3

The fraction of patients with an incubation period longer than n days for $14 \leq n \leq 30$.

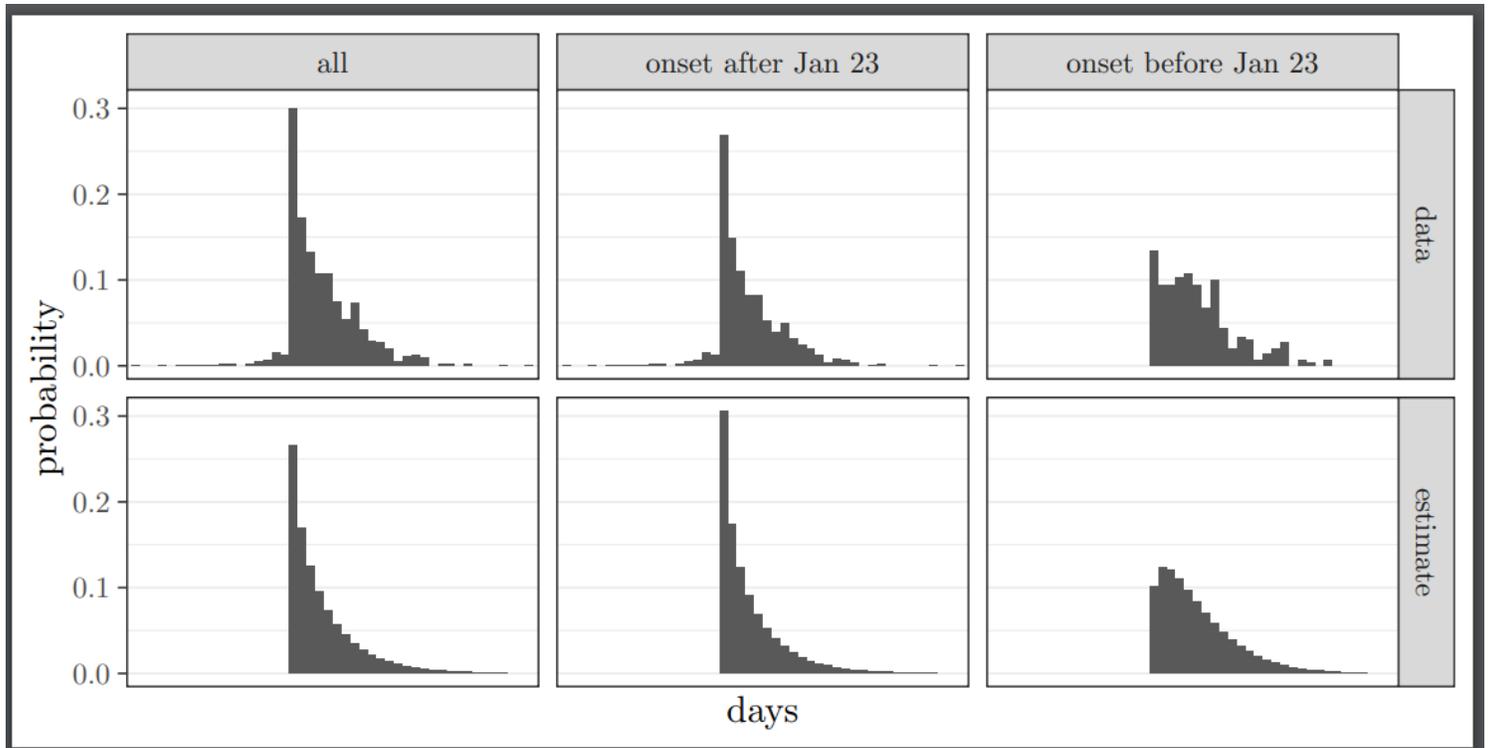


Figure 4

The observed periods from symptom onset to quarantine or hospitalization, and best-fit distribution for the period from symptom onset to hospitalization. The negative periods in data represent the 66 patients who were officially quarantined before they show symptoms. Their date of hospitalization is the same as their date of symptom onset. The estimate panel shows the discretized best-fit distribution, computed from the average of 10 thousand MCMC samples.

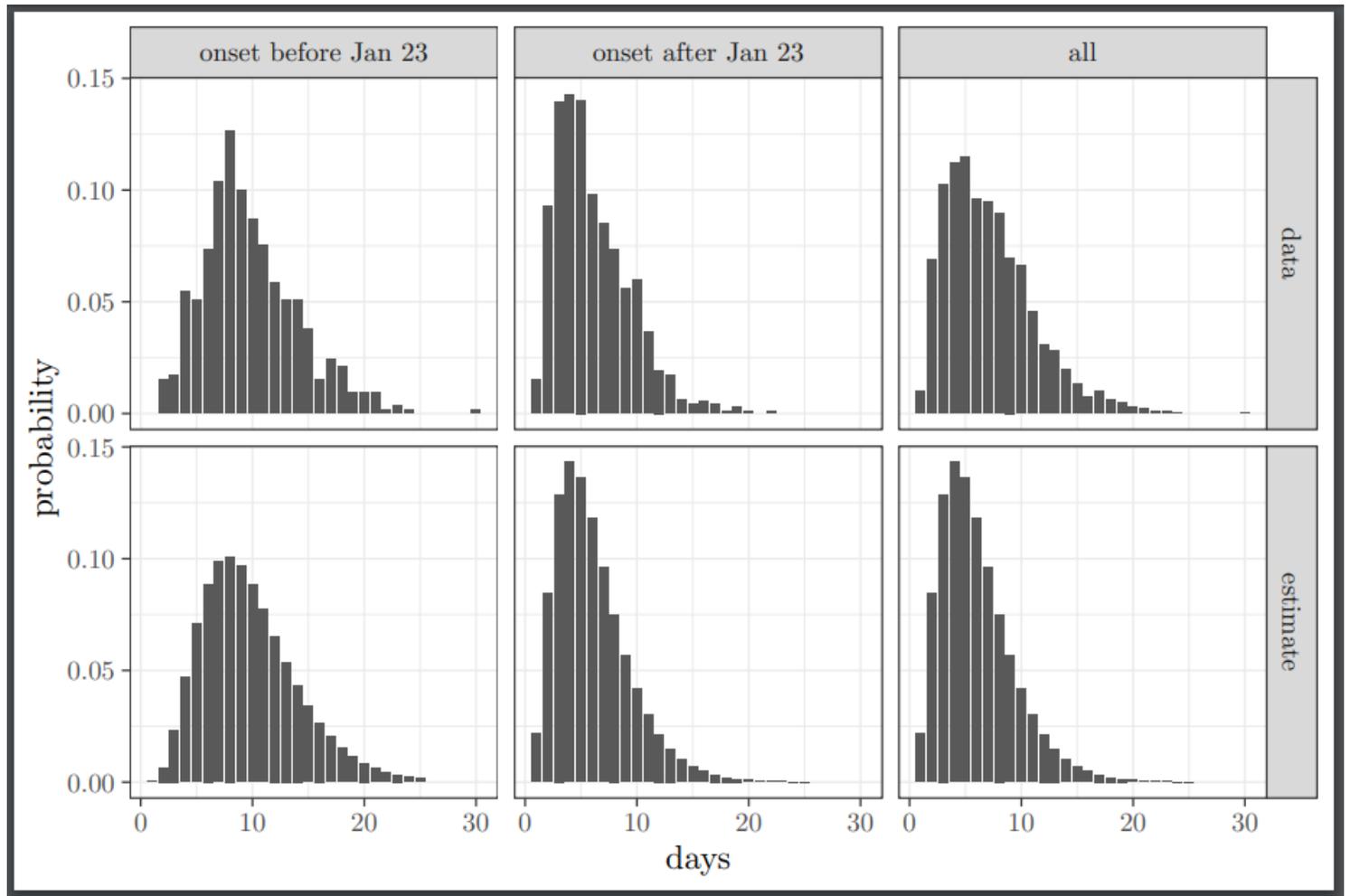


Figure 5

The observed and best-fit distributions for the period from symptom onset to diagnosis, for all patients, those who showed symptoms before the lockdown of Wuhan city on January 23, and those after. The estimate panel shows the discretized best-fit distribution, computed from the average of 10 thousand MCMC samples.

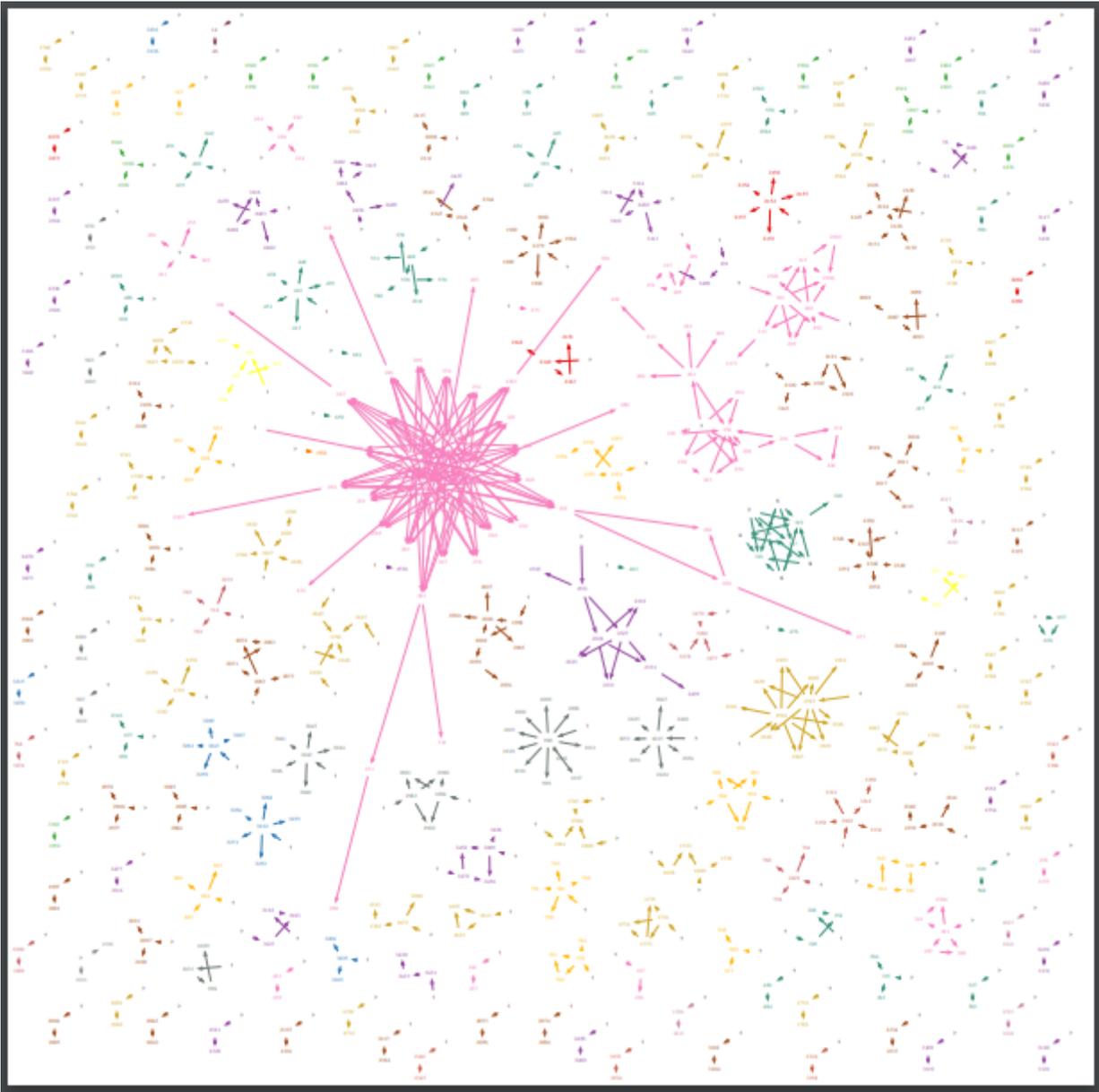


Figure 6

The graph of transmissions for 643 individually documented cases with a contact tracing information. The nodes represent patients, and arcs represent possible transmissions. The unknown sources are labeled as "?". The nodes and arcs are color-coded by province. Patient information, including their labels and provinces, are listed in Supplementary Material S1.

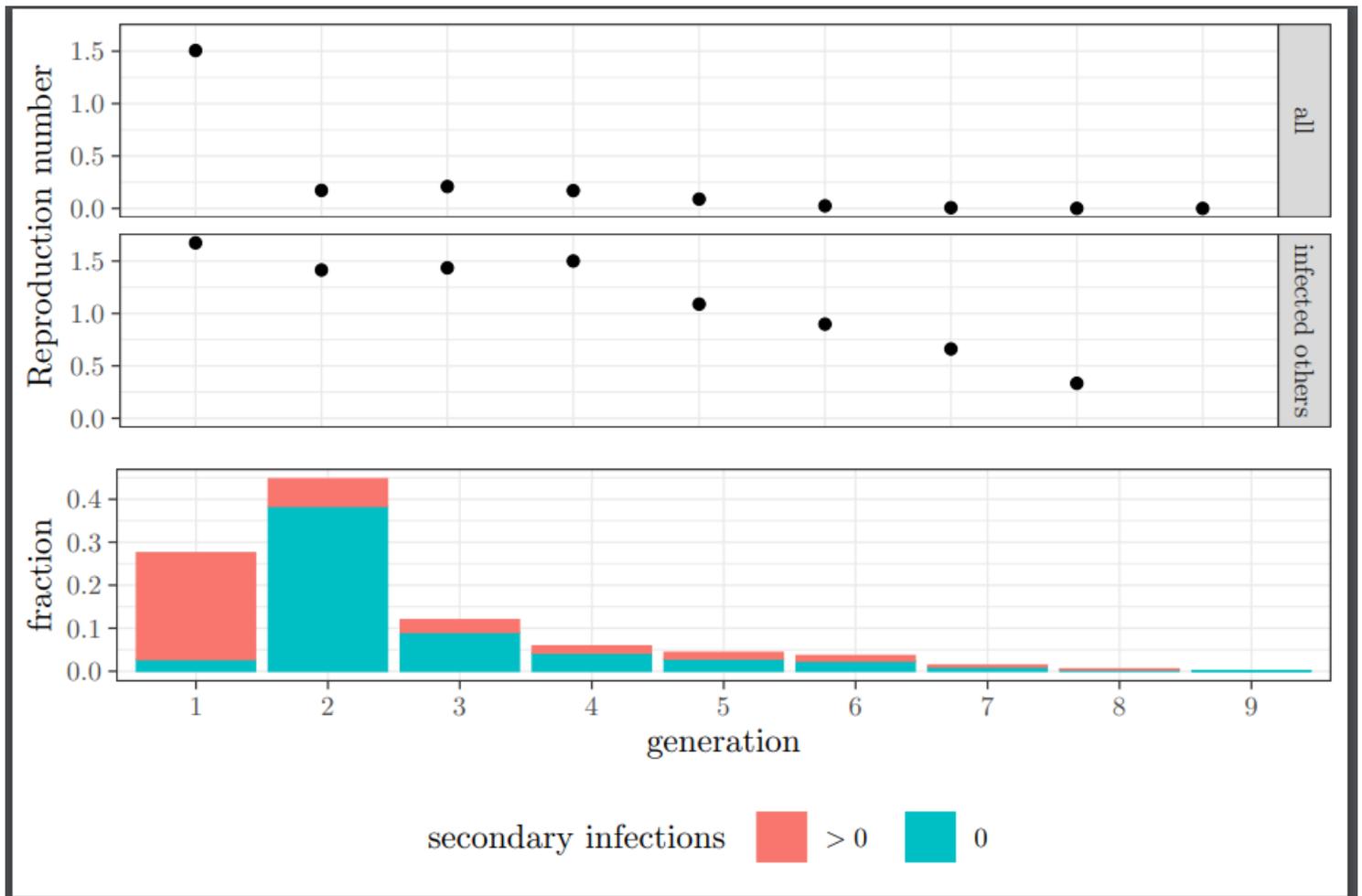


Figure 7

The reproduction number of patients in each generation. The upper panel includes all patients, the middle panel only includes patients who infected others. The lower panel shows the fractions of patients in each generation. The majority of the patients in the secondary generations and later did not infect others due to social distancing.

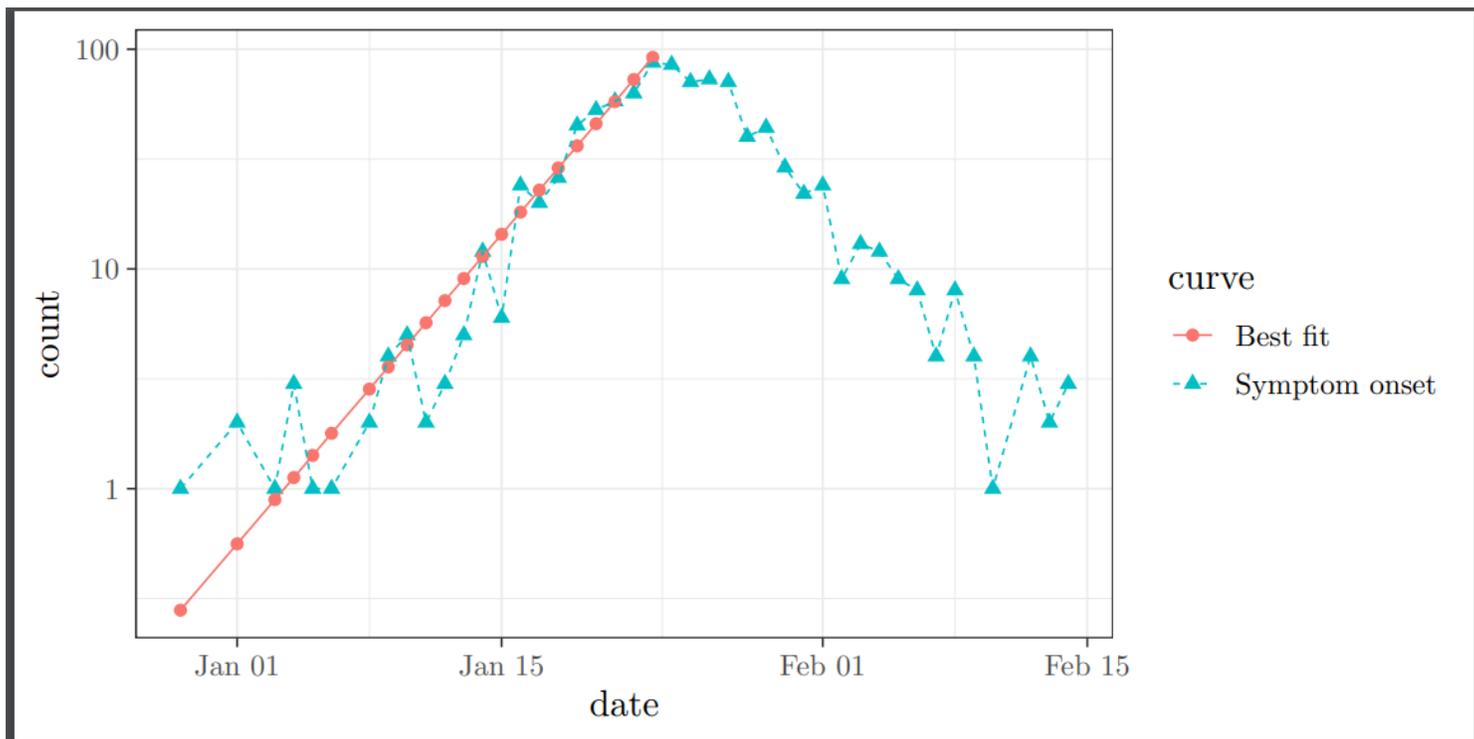


Figure 8

The daily counts of symptom onsets for patients who have either visited Hubei or are a resident of Hubei, and the best-fit exponential growth curve. The exponential growth rate is 0.23.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement1.xlsx](#)
- [supplement2.pdf](#)
- [MethodswithEquations.docx](#)