Reconstruction of Transmission Pairs for novel Coronavirus Disease 2019 (COVID-19) in mainland China: Estimation of Super-spreading Events, Serial Interval, and Hazard of Infection

Xiao-Ke Xu¹⁺, PhD, Xiao-Fan Liu²⁺, PhD, Ye Wu ³⁺, PhD, Sheikh Taslim Ali⁵⁺, PhD, Zhanwei Du⁶⁺, PhD, Paolo Bosetti⁸, PhD, Eric H.Y. Lau⁵, PhD, Benjamin J. Cowling⁵, PhD, Lin Wang⁷⁺, PhD

¹College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China
²Web Mining Lab, Department of Media and Communication, City University of Hong Kong, Hong Kong Special Administrative Region, China
³Computational Communication Research Center, Beijing Normal University, Zhuhai 519087, China
⁴School of Journalism and Communication, Beijing Normal University, Beijing 100875, China
⁵WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China
⁶Department of Integrative Biology, University of Texas at Austin, Austin, Texas 78705, USA
⁷Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK
⁸Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, UMR2000, CNRS, Paris 75015, France

⁺These authors contributed equally to this study.

Corresponding author: Lin Wang (lw660@cam.ac.uk), Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, United Kingdom. Tel: +33 632115679

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com.
Summary): A unique COVID-19 line-list database comprising 1,407 transmission pairs that formed 643 clusters in mainland China outside Hubei province was reconstructed to estimate super-spreading events, serial intervals, and hazard of infection for household versus non-household settings.
Abstract

**Background.** Knowledge on the epidemiological features and transmission patterns of COVID-19 is accumulating. Detailed line-list data with household settings can advance the understanding of COVID-19 transmission dynamics.

**Methods.** A unique database with detailed demographic characteristics, travel history, social relationships, and epidemiological timelines for 1,407 transmission pairs that formed 643 transmission clusters in mainland China was reconstructed from 9,120 COVID-19 confirmed cases reported during January 15 - February 29, 2020. Statistical model fittings were used to identify the super-spreaders and estimate serial interval distributions. Age and gender-stratified hazard of infection were estimated for household versus non-household transmissions.

**Results.** There were 34 primary cases identified as super-spreaders, with 5 super-spreadening events occurred within households. Mean and standard deviation of serial intervals were estimated as 5.0 (95% CrI: 4.4, 5.5) and 5.2 (95% CrI: 4.9, 5.7) days for household transmissions and 5.2 (95% CrI: 4.6, 5.8) and 5.3 (95% CrI: 4.9, 5.7) days for non-household transmissions, respectively. Hazard of being infected outside of households is higher for age between 18 and 64 years, whereas hazard of being infected within households is higher for young and old people.

**Conclusions.** Non-negligible frequency of super-spreadening events, short serial intervals, and a higher risk of being infected outside of households for male people of working age indicate a significant barrier to the identification and management of COVID-19 cases, which requires enhanced non-pharmaceutical interventions to mitigate this pandemic.

**Keywords:** COVID-19; transmission; super-spreadening event; serial interval; hazard of infection
Introduction

In December 2019, a novel coronavirus disease (COVID-19) emerged in Wuhan of Hubei Province, China. The World Health Organization (WHO) announced a public health emergency of international significance on 30 January 2020 and classified the threat as a global pandemic on 11 March 2020. More than 6 million confirmed cases and 371,166 deaths have been reported from more than 200 countries and territories as of 2 June 2020.

On 23 January 2020, China raised the national emergency response to the highest level, which triggered an unprecedented travel ban starting from the lockdown of Wuhan on 23 January, and 14 cities in Hubei Province on 24 January, and more than 30 provinces thereafter. Despite this unprecedented intervention, we estimated that COVID-19 cases had been introduced into 130 (95% CrI: 190, 369) cities in mainland China prior to the lockdown of Wuhan on 23 January 2020. Similar findings on the rapid geographic expansion of COVID-19 have also been reported in several recent studies. Starting from the last week of January 2020, more than 260 Chinese cities have implemented intensive social distancing and confinement policies, which brought the epidemic under control. However, the epidemic has still caused more than 10,000 confirmed cases in China outside Hubei Province within a month.

To enhance public health preparedness and awareness, Chinese health authorities have publicly reported detailed records of confirmed COVID-19 cases since mid-January. This provides a unique resource for studying the transmission patterns, routes, and risk factors of COVID-19.
Methods

Data Collection

In mainland China, 27 provincial and 264 urban health commissions have publicly posted 9,120 confirmed case reports online during January 15 – February 29, 2020, which accounts for 72% of all cases confirmed in mainland China outside Hubei Province. We compiled a unique line-list database using these reports, which contains detailed information about the demographic feature, social relationship, travel history and key epidemiological timelines (e.g., dates of symptom onset, confirmation, and hospitalization). In contrast to several published COVID-19 data repositories \(^{11-16}\) which focus on describing information about individual cases, our database allows to reconstruct transmission pairs and clusters by inferring potential causal associations among different cases. The entire dataset of transmission pairs is available at our GitHub (https://github.com/linwangidd/covid19_transmissionPairs_China). See Supplementary Materials for more details.

Statistical Analysis

We reconstructed 1,407 transmission pairs using the epidemiological evidence among reported cases. The section “Reconstruction of transmission pairs” in Supplementary Materials specifies how we identified a pair or a group of confirmed cases using information about their close contacts, stratified transmission pairs into household and non-household settings using information about familial relationships, and determined the direction of transmission between infector and infectee using information about travel histories. For each transmission pair, we term the infector the primary case and the infectee the secondary case. We also consider connected chains of confirmed cases, in which we term the original case the index and the entire chain of cases, including the index, the transmission cluster (Figure 1a).

We categorized each transmission pair by the social relationship between primary and secondary cases (e.g., familial members of the same household, non-household relatives, colleagues, classmates, friends, and other face-to-face contacts). Considering that during the Spring Festival travel season (January 10 – February 18, 2020) several billion human movements can occur because of the tradition of Chinese New Year (to visit and...
live with their original families), we considered any transmission pair with immediate familial relationships (e.g., a person’s spouse, parents, and children) as a household transmission pair, and with other familial relationships (e.g., a person’s siblings with age older than 17) or close contacts with no familial information (e.g., classmates, colleagues) as a non-household transmission pair. The numbers of household (662) and non-household (745) transmission pairs are almost even.

Following Lloyd-Smith et al, we defined the threshold of observing super-spreading events (SSE) as the 99-percentile of the offspring distribution for the number of secondary cases caused by a primary case. Household and non-household transmissions were combined together for computing the offspring distribution. To estimate the threshold of observing SSE, we used a Poisson, exponential, and power-law distribution to fit the empirical offspring distribution via the Distribution Fitter App in Matlab R2020a. Since the power-law distribution gives the smallest Akaike information criterion (AIC) compared to the Poisson and exponential distributions (Supplementary Table 2), the threshold of observing SSE is set as 3.78, which indicates the occurrence of an SSE if 4 or more secondary cases were infected by a single primary case.

For each transmission pair with known symptom onset times for both primary and secondary cases, we computed the empirical serial interval as the number of days between the symptom onset dates of the primary case and of the secondary case. Due to the presence of negative-valued serial intervals and the skewness of the empirical distribution (Figure 1b), we estimated the serial interval distributions by fitting a normal distribution via the Markov Chain Monte Carlo (MCMC) method with Gibbs sampling and non-informative flat prior. We confirmed the convergence of MCMC chains via trace plot and diagnosis, and obtained the posterior estimates of parameters by running 100,000 iterations with a burn-in of 40,000 iterations and a thinning interval of 10. Fitting serial intervals data with a Gumbel or Logistic distribution gives similar estimates (Supplementary Tables 3 - 5).
We estimated the age-stratified hazard of infection $\gamma_{H/N}(a, b)$ for household versus non-household transmissions by the ratio between the probability $p_H(a, b)$ that a secondary case of age group $b$ was infected by a primary case of age group $a$ within the same household and the probability $p_N(a, b)$ that a secondary case of age group $b$ was infected by a primary case of age group $a$ outside of households, i.e.,

$$\gamma_{H/N}(a, b) = \frac{p_H(a, b)}{p_N(a, b)}.$$ 

If $\gamma_{H/N}(a, b) > 1$, then the infection within households has a higher risk than the infection outside of households for secondary cases of age group $b$ being infected by primary cases of age group $a$.

We estimated the gender-specific hazard of infection for household versus non-household transmissions by the ratio between the probability that a secondary case of gender $b$ was infected by a primary case of gender $a$ within the same household and the probability that a secondary case of gender $b$ was infected by a primary case of gender $a$ via non-household transmission.

Role of the Funding Source

The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of the French Agence Nationale de la Recherche, European Commission, European Research Council, National Institutes of Health, National Natural Science Foundation and National Social Science Foundation of China, and the Government of Hong Kong Special Administrative Region. The funders had no role in study design, data collection and analysis, preparation of the manuscript, or the decision to submit the manuscript for publication.

Results

We in total reconstructed 643 transmission clusters formed by 1,407 transmission pairs (Figure 1a). The size is smaller than 5 for 587 transmission clusters, whereas the size exceeds 20 for the largest cluster. We identified 34 primary cases as the super-spreaders, with 5 super-spreading events occurred within households. Stratification by household setting demonstrates that 356, 92, and 34 primary cases infected 1, 2, and $\geq 3$ secondary cases.
within households, respectively, and 276, 78, and 75 primary cases infected 1, 2, and \( \geq 3 \) secondary cases outside of households, respectively (Figure 1b).

Fitting a normal distribution to serial intervals data for household transmissions estimates the mean and standard deviation (sd) of serial intervals as 5.0 (95% CrI: 4.4, 5.5) and 5.2 (95% CrI: 4.9, 5.7) days, respectively. Given the posterior median estimates of the mean and sd, the median and interquartile range (IQR) of serial interval distribution are estimated to be 5.0 days and (1.5, 8.5) days for household transmissions. Fitting a normal distribution to serial intervals data for non-household transmissions estimates the mean and sd of serial intervals as 5.2 (95% CrI: 4.6, 5.8) and 5.3 (95% CrI: 4.9, 5.7) days, respectively. Given the posterior median estimates of the mean and sd, the median and IQR of the serial interval distribution is estimated to be 5.2 days and (1.6, 8.8) days for non-household transmissions. See Supplementary Table 3 for more results. Notably, 25 of 339 household- and 25 of 340 non-household transmission pairs reported negative-valued serial intervals, implying pre-symptomatic transmission.

We performed several sensitivity analysis on estimating serial interval distributions (Supplementary Tables 4 - 5), such as the stratification of transmission pairs by the location of primary cases (imported versus local), and estimating transmission pairs with more clear epidemiological evidence (e.g., primary case linked only to a single secondary case). Results of these sensitivity analysis are consistent with those estimated with all transmission pairs.

Hazard of being infected within households was higher for the age groups of young (<18) and elderly (>65) people, whereas the hazard of being infected outside of households was higher for age groups between 18 and 64 years (Table 1). Primary cases of elderly (>65) people were more prone to cause household infections. Hazard of infection between different genders was higher for households than non-household transmission (Table 2).
Discussion

We have built a line-list database with detailed demographic information, travel history, epidemiological timelines, and social relationships for 1,407 transmission pairs that formed 643 transmission clusters in mainland China outside Hubei Province. We identified 34 primary cases as super-spreaders. Majority of superspreading events were observed for non-household transmissions, which is consistent with a recent study on transmission settings of COVID-19 (e.g., hospitals, residential care, prisons, boarding schools, cruise ships). This indicates the importance of non-pharmaceutical interventions (e.g. isolation, quarantine, social distancing, and confinement) in mitigating the COVID-19 epidemic.

Household studies are helpful to identify risk factors for certain demographic groups. The analysis of the age-stratified and gender-specific hazard of infection suggests a higher risk of infection within households for age groups of young (<18), elderly (>65) and female people. The higher risk of being infected outside of households for male people of age between 18 and 64 years may indicate their role in driving household secondary infections, perhaps because these were travelers from Wuhan of working age.

We identified 50 transmission pairs (~3.5%) with secondary case reported symptom onset earlier than primary case (i.e., negative-valued serial intervals), which is consistent with recent clinical reports and epidemiological studies. We estimated that the mean serial interval is around 5 days for both household and non-household infections, which is considerably shorter than the mean serial interval estimated for SARS (e.g., 8.4 days) and MERS (e.g. 7.6 days).

Our findings have several limitations. First, the household sizes and primary cases with no secondary infections were not provided from the original public case reports. This may give rise to biased estimates if we estimate the household reproduction number and secondary attack rate from the raw data. Field surveys will be helpful to adjust such biases. Second, the information on nosocomial infections and public gathering settings was not available from original case reports, so that the observation of super-spreading events may be less common from our dataset. Third, caution is needed when attempting to generalize the age-stratified hazard of infection to other
demographic settings. For example, in our study (Table 1), children (<18) never acquired COVID-19 from other children at home may be more a reflection of the usual household composition in Chinese cities (single child living with parents) than the transmission characteristics of the virus. China had a lower proportion of household with multiple children\(^{33}\), which may reduce the risk of transmission between children especially during lock down and school closure.

In sum, the notable threat of super-spreading events, short serial intervals, and a higher risk of being infected outside of households for adult male of working class (18–64 years old) indicate a significant barrier to the identification and management of COVID-19 cases, which requires enhanced non-pharmaceutical interventions to mitigate this pandemic.
Notes

Author contributions. XKX, XFL, YW, STA, and LW conceived the study. STA, ZD, EHYL, BJC, LW had roles in the study design. XKX, XFL, YW had roles in the data collection, data analysis and data interpretation. STA, ZD, PB, EHYL, BJC, LW had roles in the data interpretation, statistical modelling, and writing of the manuscript. All authors reviewed and approved the final version of the manuscript.

Acknowledgements and financial support. The authors thank all health workers and volunteers in China who collected, prepared, and shared data throughout this outbreak. They are in particular grateful to Simon Cauchemez, Henrik Salje, Lauren Ancel Meyers, Juliette Paireau, Qifang Bi, Bingyi Yang, and Lanfang Hu for comments and suggestions. This study was partially supported by the Investissement d’Avenir program, the Laboratoire d’Excellence Integrative Biology of Emerging Infectious Diseases program (ANR-10-LABX-62-IBEID), European Union V.E.O project, European Research Council (804744), National Institutes of Health (U01 GM087719), Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources (KF-2019-04-034), National Natural Science Foundation of China (61773091, 11875025, 11975025), Major Project of The National Social Science Fund of China (19ZDA324), and Health and Medical Research Fund, Food and Health Bureau, Government of the Hong Kong Special Administrative Region, China (COVID1910118).

Potential conflicts of interests. Prof. Benjamin J. Cowling reports honoraria from Sanofi Pasteur and Roche. The other authors report no other potential conflicts of interest.

Reproducible Research Statement:

Statistical code and data set: Available at: https://github.com/linwangidd/covid19_transmissionPairs_China
References

1. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). <https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)> (2020).

2. WHO Director-General’s opening remarks at the media briefing on COVID-19 – 11 March 2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (2020).

3. World Health Organization. Coronavirus disease 2019 (COVID-19), situation report - 133, 01 June 2020 (available at https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200601-covid-19-sitrep-133.pdf?sfvrsn=9a56f2ac_4).

4. Du, Z. et al. Risk for Transportation of 2019 Novel Coronavirus Disease from Wuhan to Other Cities in China. Emerg Infect Dis 26, doi:10.3201/eid2605.200146 (2020).

5. Wu, J. T., Leung, K. & Leung, G. M. 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, doi:10.1016/S0140-6736(20)30260-9 (2020).

6. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science, doi:10.1126/science.abb9757 (2020).

7. Tian, H. et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science, doi:10.1126/science.abf6105 (2020).

8. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science, doi:10.1126/science.abb4218 (2020).

9. Kucharski, A. J. et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis, doi:10.1016/S1473-3099(20)30144-4 (2020).

10. Lai, S. et al. Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. medRxiv, doi:10.1101/2020.03.03.20029843 (2020).

11. Xu, B., Kraemer, M. U. G. & Open, C. –D. C. G. Open access epidemiological data from the COVID-19 outbreak. Lancet Infect Dis, doi:10.1016/S1473-3099(20)30119-5 (2020).

12. Xu, B. et al. Epidemiological data from the COVID-19 outbreak, real-time case information. Sci Data 7, 106, doi:10.1038/s41597-020-0448-0 (2020).
Zhang, J. et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30230-9 (2020).

Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30120-1 (2020).

Sun, K., Chen, J. & Viboud, C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digital Health* **2**, PE201-208, doi:10.1016/S2589-7500(20)30026-1 (2020).

Lauer, S. A. et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*, doi:10.7326/M20-0504 (2020).

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

MathWorks. Working with Probability Distributions, <https://fr.mathworks.com/help/stats/working-with-probability-distributions.html> (2020).

Vink, M. A., Bootsma, M. C. & Wallinga, J. Serial intervals of respiratory infectious diseases: a systematic review and analysis. *Am J Epidemiol* **180**, 865-875, doi:10.1093/aje/kwu209 (2014).

Liu, Y., Eggo, R. M. & Kucharski, A. J. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet* **395**, e47, doi:10.1016/S0140-6736(20)30462-1 (2020).

Leclerc QJ, F. N., Knight LE et al. What settings have been linked to SARS-CoV-2 transmission clusters? [version 1; peer review: 1 approved with reservations]. Wellcome Open Res 2020, 5:83 (https://doi.org/10.12688/wellcomeopenres.15889.1).

Cowling, B. J. et al. Impact assessment of non-pharmaceutical interventions against COVID-19 and influenza in Hong Kong: an observational study. *medRxiv*, doi:10.1101/2020.03.12.20034660 (2020).

Koo, J. R. et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30162-6 (2020).

Prem, K. et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health*, doi:10.1016/S2468-2667(20)30073-6 (2020).

Cowling, B. J. & Leung, G. M. Epidemiological research priorities for public health control of the ongoing global novel coronavirus (2019-nCoV) outbreak. *Euro Surveill* **25**, doi:10.2807/1560-7917.ES.2020.25.6.2000110 (2020).
Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 - Studies Needed. *N Engl J Med* **382**, 1194–1196, doi:10.1056/NEJMp2002125 (2020).

Bai, Y. *et al.* Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA*, doi:10.1001/jama.2020.2565 (2020).

Pan, X. *et al.* Asymptomatic cases in a family cluster with SARS-CoV-2 infection. *Lancet Infect Dis* **20**, 410–411, doi:10.1016/S1473-3099(20)30114-6 (2020).

Du, Z. *et al.* Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg Infect Dis* **26**, doi:10.3201/eid2606.200357 (2020).

He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* **26**, 672–675, doi:10.1038/s41591-020-0869-5 (2020).

Lipsitch, M. *et al.* Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966–1970, doi:10.1126/science.1086616 (2003).

Assiri, A. *et al.* Hospital outbreak of Middle East respiratory syndrome coronavirus. *N Engl J Med* **369**, 407–416, doi:10.1056/NEJMoal306742 (2013).

Nations., U. Household size and composition around the World 2017 : data
https://digitallibrary.un.org/record/3799696/files/household_size_and_composition_around_the_world_2017_data_booklet.pdf.
Table 1: Age-stratified hazard of infection for household versus non-household transmissions. Red (blue) shades indicate an increased (decreased) hazard of infection within households relative to outside of households, respectively.

| Primary cases | Secondary cases | 0~17 | 18~49 | 50~64 | 65+ | Total |
|---------------|-----------------|------|-------|-------|-----|-------|
| 0~17          |                 | 0.0  | 0.8   | 0.8   | 1.1 | 0.7   |
| 18~49         |                 | 6.3  | 0.7   | 0.9   | 2.0 | 1.1   |
| 50~64         |                 | 1.7  | 0.9   | 0.7   | 0.6 | 0.8   |
| 65+           |                 | 2.3  | 1.4   | 0.6   | 2.1 | 1.3   |
| Total         |                 | 3.5  | 0.8   | 0.8   | 1.4 | 1.0   |
Table 2: Gender-specific hazard of infection for household versus non-household transmissions. Red (blue) shades indicate an increased (decreased) hazard of infection within households relative to outside of households, respectively.

| Table 1b | Secondary cases |
|----------|-----------------|
|          | Male | Female | Total |
| Primary cases | Male | 0.6   | 1.6   | 1.0   |
|           | Female | 1.2   | 0.7   | 0.9   |
| Total     | 0.8 | 1.2   | 1.0   |
Figure Legends

Figure 1: COVID-19 transmission clusters.

(a) 643 transmission clusters, stratified by the size of cluster $n$. Red, green and blue nodes denote primary cases, household secondary cases and non-household secondary cases, respectively.

(b) Empirical offspring distribution for the number of secondary cases infected by each of the 809 primary cases.

(c) Empirical and estimated serial interval distributions for household or non-household transmissions. Pink and blue bars indicate the empirical distributions of serial interval data for household and non-household transmissions, respectively. Red curve indicates the estimated serial interval distributions for household transmissions, based on the posterior median estimates of the mean and standard deviation (sd) of normal distribution. Blue curve indicates the estimated serial interval distributions for non-household transmissions, based on the posterior median estimates of the mean and sd of normal distribution. Fitting results with alternative distributions (e.g., Gumbel or Logistic distribution) are available in Supplementary Tables 3 - 5.
Figure 1

n=2

n=3

n=5

n=4

n ≥ 6

Figure 1

Proportion

Number of secondary cases infected by each primary case

Density

Serial Interval (days)