

RESEARCH

# Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a metropolis of China

Yunjun Zhang<sup>1†</sup>, Yuying Li<sup>1†</sup>, Lu Wang<sup>2</sup>, Mingyuan Li<sup>3</sup> and Xiaohua Zhou<sup>1,2,4\*</sup>

\*Correspondence:

azhou@math.pku.edu.cn

<sup>1</sup>Department of Biostatistics,  
School of Public Health, Peking  
University, Xueyuan Road, 100191  
Beijing, China

Full list of author information is  
available at the end of the article

<sup>†</sup>Equal contributor

## Abstract

**Background:** COVID-19 caused rapid mass infection worldwide. Understanding its transmission characteristics including heterogeneity is of vital importance for prediction and intervention of future epidemics. In addition, transmission heterogeneity usually evokes super spreading events (SSEs) where certain individuals infect large numbers of secondary cases. Till now, studies of transmission heterogeneity of COVID-19 and its underlying reason are far from reaching an agreement.

**Methods:** We collected information of all infected cases between January 21 and February 26, 2020 from official public sources in Tianjin, a metropolis of China. Utilizing a heterogeneous transmission model based on branching process along with a negative binomial offspring distribution, we estimated the reproductive number  $R$  and the dispersion parameter  $k$  which characterized the transmission potential and heterogeneity, respectively. Furthermore, we studied the SSE in Tianjin outbreak and evaluated the effect of control measures undertaken by local government based on the heterogeneous model.

**Results:** A total of 135 confirmed cases (including 34 imported cases and 101 local infections) in Tianjin by February 26th 2020 entered the study. We grouped them into 43 transmission chains with the largest chain of 45 cases and the longest chain of 4 generations. The estimated reproduction number  $R$  was at 0.67 (95%CI: 0.54~0.84), and the dispersion parameter  $k$  was at 0.25 (95% CI: 0.13~0.88). A super spreader causing six infections in Tianjin, was identified. In addition, our simulation results showed that the outbreak in Tianjin would have caused 165 infections and sustained for 7.56 generations on average if no control measures had been taken by local government since January 28th.

**Conclusions:** Our analysis suggested that the transmission of COVID-19 was subcritical but with significant heterogeneity and may incur SSE. More efforts are needed to verify the transmission heterogeneity of COVID-19 in other populations and its contributing factors, which is important for developing targeted measures to curb the pandemic.

**Keywords:** COVID-19; Super spreading; Transmission heterogeneity

## 1 Background

- 2 In December 2019, many cases of viral pneumonia-like disease similar to severe acute
- 3 respiratory syndrome (SARS) were detected in Wuhan, China. Later they were con-

4   firmed to be caused by a novel coronavirus, provisionally called 2019 novel coro-  
5   navirus (2019-nCoV). On January 30th 2020, World Health Organization (WHO)  
6   declared the 2019-nCoV outbreak as a global health emergency of international con-  
7   cern [1]. On February 11th 2020, WHO named the diseases CoronaVirus Disease  
8   2019 (COVID-19) and announced "severe acute respiratory syndrome coronavirus 2  
9   (SARS-CoV-2)" as the name of the new virus. On March 11th 2020, WHO declared  
10   COVID-19 a pandemic. As of April 18th 2020, more than 2,160,207 cases were con-  
11   firmed globally and the National Health Commission (NHC) of China reported a  
12   total of 82,735 cases of COVID-19 in mainland China, including 77,062 recoveries  
13   and 4,632 deaths.

14    The dynamics of an infectious disease outbreak depends on both the potential and  
15   the heterogeneity of disease transmission. The transmission potential of an infectious  
16   pathogen is usually represented by its reproduction number (denoted as  $R$ ) which  
17   is the average number of secondary cases caused by a typical infectious individual.  
18   Estimates of  $R > 1$  (i.e., supercritical outbreak), such as for the early outbreak  
19   of the severe acute respiratory syndrome coronavirus (SARS-CoV) in China 2003  
20   [2], indicate the great risk for an infection pathogen to generate a major outbreak;  
21   estimates of  $R < 1$  (i.e., subcritical outbreak), such as for the Middle East res-  
22   piratory syndrome coronavirus (MERS-CoV) in South Korea 2018 [3], imply that  
23   the outbreak is slowing down with declining trend of incidence. Though different in  
24   transmission potential, both SARS and MERS shared the same feature of high level  
25   of heterogeneity which involved the uneven transmission patterns with a number of  
26   super-spreading events (SSEs), where some individuals spread to a disproportion-  
27   ate number of individuals, as compared to most individuals who infected a few or  
28   none [2, 4]. SSEs of SARS and MERS were responsible for triggering the initial out-  
29   breaks in several large cities such as Beijing, Hongkong and Singapore, and hence  
30   sustaining the spread of disease worldwide. Besides, SSEs were also documented

31 in many other infectious disease [5]. Detection of transmission heterogeneity may  
32 direct prevention efforts and reduce future infections [6, 7, 8].

33 The transmission potential of COVID-19 has been studied based on mathematical  
34 models, yielding consistent evidence for the high level of reproduction number  $R$   
35 (1.95~2.2) within a completely susceptible population [9, 10, 11]. However, evidence  
36 for the transmission heterogeneity of COVID-19 and the related SSEs was limited  
37 and conflicting. So far, two studies found no heterogeneity in Singapore [12] and  
38 in a southern city of China [13], but one study confirmed heterogeneity based on  
39 data from 46 countries [14]. In addition, several possible SSEs have been reported  
40 in China[15, 16] which involved large numbers of infections. Hence, it would be  
41 essential to further explore the transmission heterogeneity and SSE in COVID-19  
42 pandemic and its relevant factors

43 In this study, we investigated the transmission characteristics of COVID-19 using  
44 the epidemiological data of all the confirmed cases in a metropolis in Northern  
45 China. We analysed the transmission chains and then estimated the reproduction  
46 number and the transmission heterogeneity using a branching model along with  
47 a negative binomial offspring distribution [17]. In addition, we identified the SSE  
48 according to the 99-percentile criterion in [2] and assessed the effect of control  
49 measures employed by local government.

## 50 **Method**

### 51 **Data Collection**

52 Tianjin is a municipality and a coastal metropolis in Northern China, with a popu-  
53 lation of 15.66 million and an area of 11,966 square kilometers. Since the first case  
54 of COVID-19 in Tianjin was confirmed on January 21st, a total of 135 cases were  
55 confirmed by the time of data collection (February 26th, 2020) of the current study.  
56 From 6:00 on February 22 to 18:00 on February 26, the city had no new confirmed  
57 cases for 108 consecutive hours. The municipal government publicized the epidemi-

58 ological information of the cases every day. Diagnosis of COVID-19 were based on  
59 the protocol issued by China CDC [18].

60 From the official websites of the Municipal Health Commissions [19], we retrieved  
61 data of the 135 confirmed COVID-19 cases in Tianjin, including demographic char-  
62 acteristics, epidemiological characteristics, i.e., travel history and contact history  
63 with confirmed/suspicious cases. According to these information, the infection rela-  
64 tionship within the scope of Tianjin can be specified. Each case was given a unique  
65 number according to its sequential order as reported.

66 We adopted the definition in [17] to define a transmission chain as a group of  
67 cases connected by an unbroken series of local transmission events. We grouped all  
68 the confirmed cases in Tianjin into transmission chains and for each transmission  
69 chain, we identified the primary case, calculated the chain size (i.e., the total number  
70 cases including the primary case) and reconstructed the transmission history (i.e.,  
71 who infected whom). According to the extent of resolution of the reconstructed  
72 information, three types of transmission chains were further identified. The first  
73 type was the *simple transmission chain* in which the transmission history could  
74 be clearly recovered. The second type was the *ordinary transmission chain* for  
75 which the transmission history was not clear but the single primary case could be  
76 clearly identified and the chain size was also clear. The third type was the *complex*  
77 *transmission chain* for which the chain size was clear but the primary case could  
78 not be clearly identified. More than one individual in the chain exhibited similar  
79 behavior/clinical characteristics, so we have to regard them as the primary cases.

80 As COVID-19 was first reported in Wuhan, a city in the middle of China, a  
81 primary case was defined if the individual had a history of travel to or residence in  
82 Wuhan within one month, had direct contact with an individual who had confirmed  
83 infection outside Tianjin, having fever outside Tianjin, or had the earliest onset of  
84 symptoms in the transmission chain.

85 Analytical approach

86 *Inference of Transmission Characteristics*

87 To quantify the transmission potential and heterogeneity for the COVID-19 out-  
88 break in Tianjin, we adopted a stochastic model based on branching process to  
89 characterize both the distribution of secondary cases (i.e., a negative binomial dis-  
90 tribution) and the resulting distribution of transmission chain sizes under the same  
91 parameterization of reproduction number  $R$  and dispersion parameter  $k$  (lower value  
92 indicating higher heterogeneity) for subcritical epidemic [17].

93 We first fitted the stochastic model to the retrieved information of the sim-  
94 ple/ordinary transmission chains. Then, in handling a complex transmission chain  
95 with two primary cases, we regarded it could be separated into two ordinary trans-  
96 mission chains, each of which was led by a primary case. The difficulty lay in that  
97 the exact size of each ordinary chain was unclear. We dealt with this ambiguity in  
98 two ways: one was the combinatorial method in [17] by allowing for all the possi-  
99 ble combinations and treated the sum as an overall probability; the other was to  
100 adopt the expectation-maximization (EM) algorithm by treating the sizes of two  
101 separated chains as latent variables. The EM algorithm also estimated the latent  
102 chain size, which was informative for dividing the complex transmission chain into  
103 constituent chains (Section B, Appendix 1).

104 To identify the possible SSEs in Tianjin outbreak, we adopted the definition in  
105 [2] to define a super-spreader as any infected individual causing more infections  
106 than would occur in 99% of infectious histories in a homogeneous population. The  
107 transmission in a homogeneous situation was modelled with a *Poisson* distribution  
108 which is the special case of negative binomial distribution without heterogeneity.

109 *Assessment of Control Measures*

110 Based on the stochastic model accounting for transmission heterogeneity, we as-  
111 sessed the control measures imposed by the Tianjin government. We compared the

112 transmission characteristics (i.e.,  $R$  and  $k$ ) for the periods of before and after the  
113 control measures taking effect. Then based on the estimate of transmission char-  
114 acteristics before control, we simulated the expected distribution of outbreaks that  
115 might occur in Tianjin if no control measures were taken. Each simulated outbreak  
116 was initiated with 43 infectious individuals (as the number of transmission chains  
117 in Tianjin), and was propagated with the branching process model on the basis of  
118 inferred characteristics.

## 119 **Results**

### 120 **Characteristics of COVID-19 Cases**

121 Among the 135 patients included in the study, 72 (53.3%) were males and 63 (46.7%)  
122 females, with an average (standard deviation) age of 47.8 (18.3) years and 50.1 (15.2)  
123 years, respectively. As shown by Fig.1 A, of the total 135 cases, 34 (25.2%) cases  
124 were imported cases. Among the 101 cases of local transmission, the majority (55,  
125 40.7%) were infected in household, and 35 (25.9%) cases were infected in public  
126 places including a shopping mall and working places, and 11 (8.1%) were unclear  
127 in the source of infection. When we examined the chronological development of the  
128 infection in Tianjin, we found that the imported cases and the household infec-  
129 tions dominated the confirmed cases of the early and the later stage of outbreak  
130 respectively. The transition happened around February 3rd.

### 131 **Reconstructed Transmission Chains**

132 The 135 cases were grouped into 43 transmission chains including 36 simple chains  
133 (47 cases, average size 1.3), 5 ordinary chains (78 cases, average size 15.6), and 2  
134 complex chains (10 cases, average size 5). (Table 1, Figure 2). Detail information  
135 of three type of transmission chains shown in Section A of additional file 1.

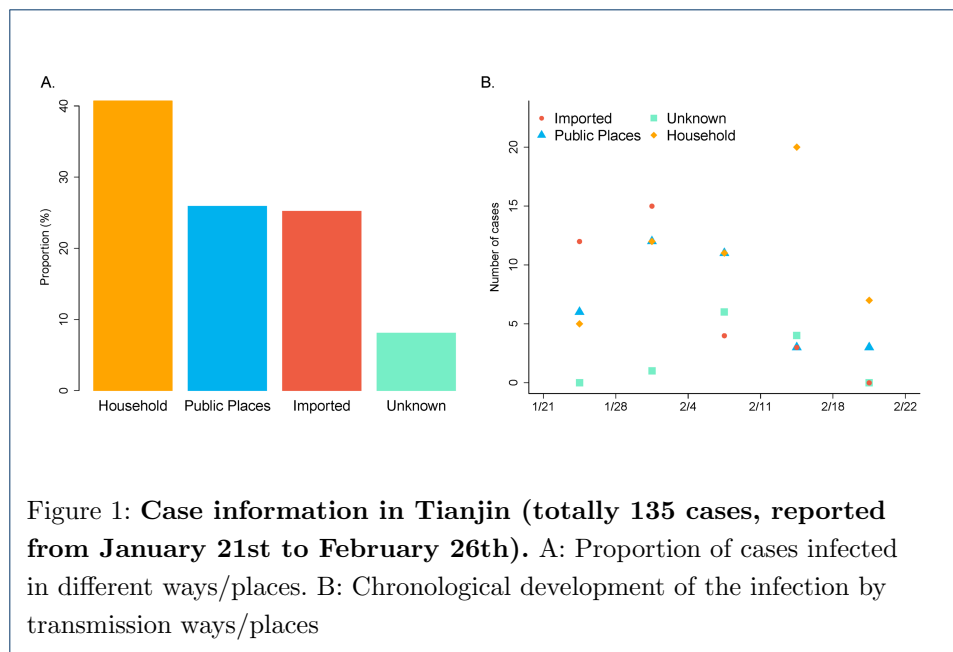


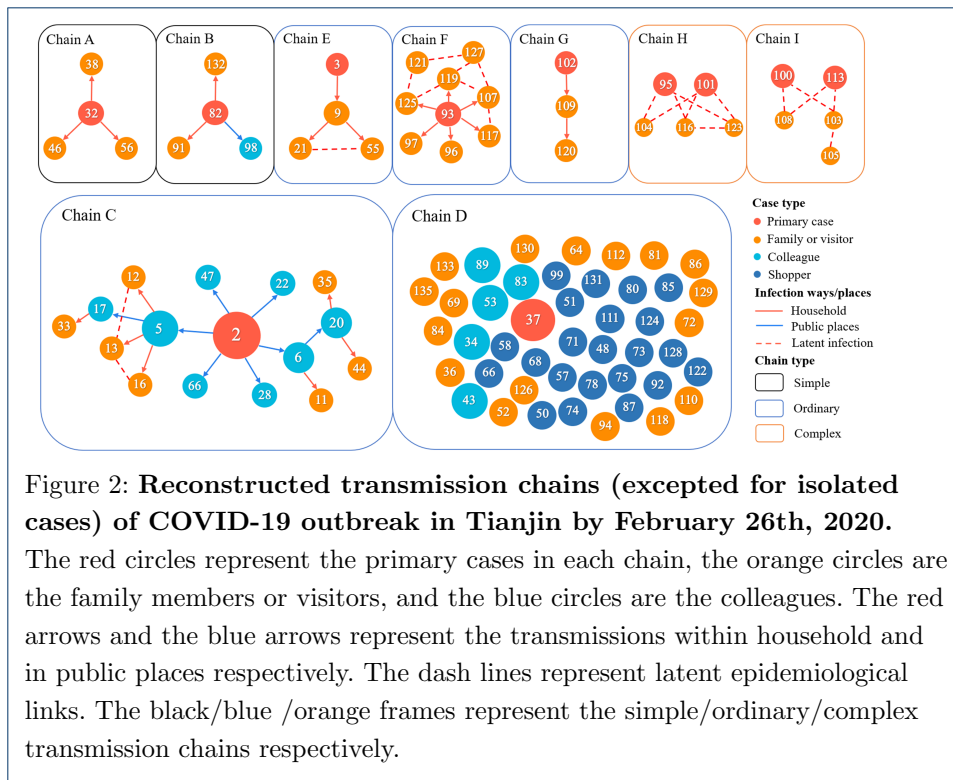
Table 1: Three types of COVID-19 transmission chains in Tianjin.

Chain Type	Amount of Chains	Total Number of Cases	Average Chain Size	Range of Chain Size
Simple transmission chain	36	47	1.3	1 - 4
Ordinary transmission chain	5	78	15.6	3 - 45
Complex transmission chain	2	10	5	5 - 5

136 Estimation of  $R$  and  $k$

137 Both the combinational method and the EM algorithm gave lower value of  $k$  with  
 138 the corresponding 95% confidence intervals (CIs) being lower than 1 (Table 2),  
 139 suggesting significant evidence for the transmission heterogeneity in the outbreak  
 140 at Tianjin. In addition, both combinatorial method and EM algorithm gave the same  
 141 estimate of reproductive number  $R = 0.67$  being lower than the critical value of 1,  
 142 indicating the local transmission would finally die out. Note that these estimate  
 143 of transmission potential was based on the data by February 26th, implying the  
 144 average trend over the same period.

145 When comparing the CIs from different methods, we observed that the EM algo-  
 146 rithm generated slightly narrower CI for both parameters than the combinatorial  
 147 method. In addition, the EM algorithm had advantage on the ability of providing



148 more information on the split of complex transmission chain since it also produced a  
 149 probabilistic insight of unknown partitions of the chain. Recall that the COVID-19  
 150 data in Tianjin contains two complex transmission chains of size 5 for each. For  
 151 these two complex chains, the EM algorithm tended to split it into one size-1 and  
 152 one size-4 transmission chain (posterior probability = 95.5%), instead of two chains  
 153 with size 2 and size 3 (posterior probability = 4.5%).

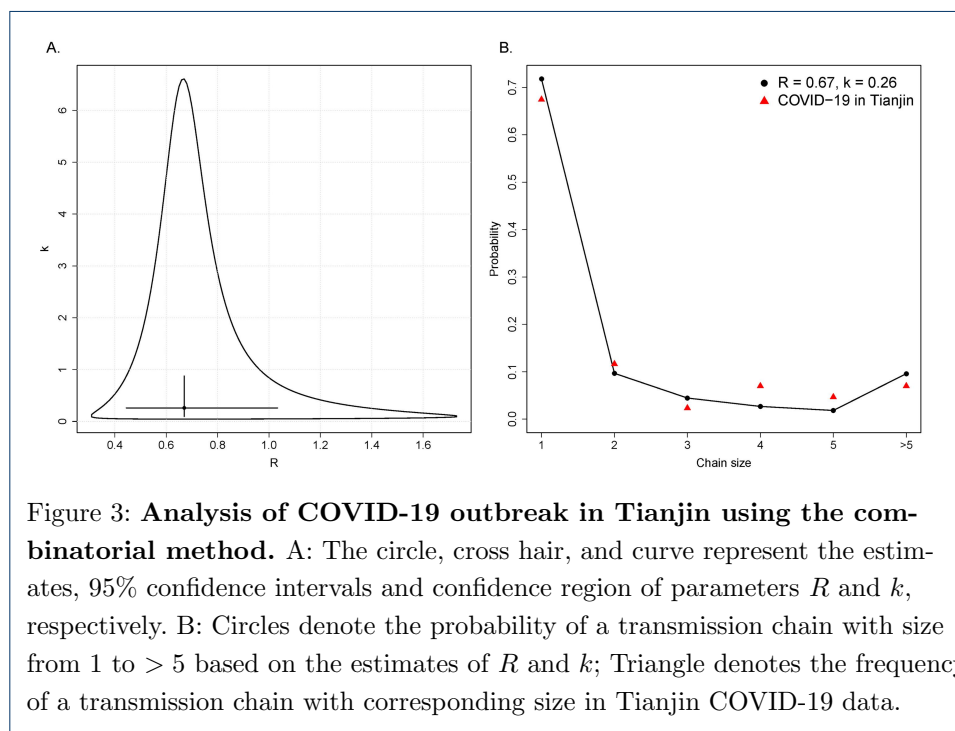
Table 2: Estimation and CI of the reproduction number  $R$  and dispersion parameter  $k$  based on the combinational method and the EM algorithm

	Combinatorial method (95% CI)	EM (95% CI)
$R$	0.67 (0.44, 1.03)	0.67 (0.54, 0.84)
$k$	0.26 (0.10, 0.88)	0.25 (0.13, 0.88)

154 Furthermore, the estimates along with the 95% CI and confidence regions of com-  
 155 binatorial method were plotted in Fig.3 A. The confidence region incorporated  
 156 the uncertainty in both parameters simultaneously, thus it undoubtedly gave much  
 157 wider range than the CI, especially for  $k$ . We also found (Fig.3 B) that the observed



158 chain size distribution was close to the expected distribution based on the estimated  
159 characteristics (i.e.,  $\hat{R} = 0.67$  and  $\hat{k} = 0.26$ ). Particularly, the probability of a chain  
160 with size over 5 was less than 10%, and the probability of a single infected case  
161 resulting in a chain of size 45 (the largest size of transmission chain in Tianjin) was  
162 about 0.8%. It indicated that the outbreak of a comparatively large chain was not  
163 likely to happen in Tianjin.



#### 164 Super Spreading Event in Tianjin

165 We found notable transmission heterogeneity ( $k = 0.25$ ) in the local spreading of  
166 COVID-19 in Tianjin, suggesting that there was likely to be SSEs [5]. As SSEs  
167 happened more likely at the early stage of the outbreak to trigger out the local  
168 spreading in the epidemic of SARS and MERS, we therefore had a reason to suspect  
169 the existence of SSEs at the early stage of outbreak in Tianjin.

170 Adopting the criterion in [2] to define SSE, we calculated the cut-off as the 99th  
171 percentile of the *Poisson* distribution with mean value of 2.2 (i.e., the reproduction  
172 number at the early), which was 6, suggesting that an infected individuals who

173 infected 6 or more secondary cases could be regarded as a super spreader of the  
174 COVID-19 outbreak in Tianjin. According to this criterion, the second case con-  
175 firmed in Tianjin was a super spreader. This primary case infected six colleagues  
176 within close contact before being confirmed on January 21st, 2020 (Fig.2). These  
177 six colleagues, as secondary cases in this transmission chain, successively infected  
178 other colleagues or their relatives.

### 179 Effect of Government Control Measures

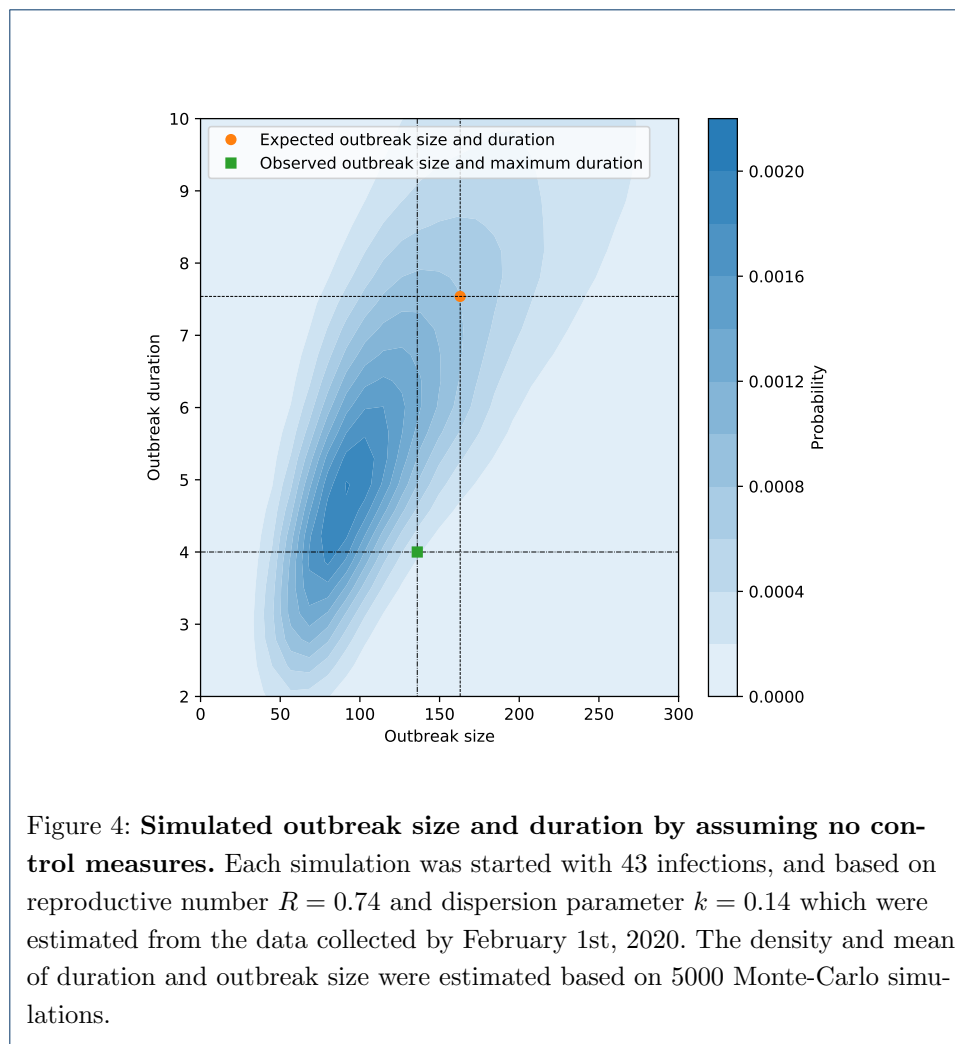
180 Tianjin municipal government deployed a series of policies to control the spread of  
181 COVID-19 from January 28th, 2020, including strong traffic restriction and quaran-  
182 tining individuals who had suspect contacts with confirmed cases[20]. By allowing  
183 for the median incubation period of 5.1 days [21], we assumed that the control  
184 measures had been taking effect since February 1st, 2020.

185 When we compared the transmission characteristics before and after the interven-  
186 tion taking effect (Table 3), we found that the reproduction number  $R$  decreased  
187 from 0.74 to 0.53, while  $k$  increased from 0.14 to 0.77, suggesting the decrease in  
188 both transmission potential and heterogeneity after taking control measures. How-  
189 ever, the overlapping CIs were likely to be caused by the small sample size.

190 Compared with the observation of 135 confirmed cases in total and the longest  
191 transmission chain with 4 generations (Fig.2), the simulation study showed that the  
192 local outbreak in Tianjin would have sustained for 7.56 generations and would have  
193 led to 165 infections on average if there were no control policies (Fig.4).

Table 3: Estimation of the reproductive number  $R$  and the dispersion parameter  $k$  for different periods

	<b>Before February 1st (95% CI)</b>	<b>After February 1st (95% CI)</b>
$R$	0.74 (0.39, 1.61)	0.53 (0.29, 0.96)
$k$	0.14 (0.04, 0.63)	0.77 (0.14, 31.47)



## 194 Discussion

195 Based on individual-level information of COVID-19 infection cases in the city of  
196 Tianjin, we discovered significant transmission heterogeneity ( $k = 0.25$ , 95%CI:  
197  $0.13 \sim 0.88$ ) and subcritical transmission potential ( $R = 0.67$ , 95%CI:  $0.54 \sim 0.84$ )  
198 and identified one super-spreader who infected 6 individuals, suggesting that the  
199 local outbreak in Tianjin was considerably heterogeneous even though it would  
200 finally die out. In addition, our numerical results successfully verify the effectiveness  
201 of government control deployed on January 28th in Tianjin.

202 Our finding of significant transmission heterogeneity of COVID-19 outbreak in  
203 Tianjin was inconsistent with the previous studies conducted in ShenZhen, China

204 ( $k = 0.58$ , 95%CI: 0.35~1.18) [13] and in Singapore ( $k = 0.4$ , 95%CI: 0.1~Inf)  
205 [12]. The absence of heterogeneity in these two studies could have been due to the  
206 failure of tracking some epidemiological links among cases. In addition, our result  
207 was consistent with another study conducted among 46 countries ( $k = 0.1$ , 95%CI:  
208 0.05~0.2) [14]. However, it should be noted that the large scale spreading of COVID-  
209 19 is likely to be heterogeneous because of many extrinsic factors, such as weather  
210 [22] and different control measures, may affect the transmission of pathogen, while  
211 our confirmation of transmission heterogeneity of COVID-19 in a local outbreak  
212 justified that some intrinsic properties of the pathogen might also lead to heteroge-  
213 neous transmission. Our finding is more meaningful for the development of targeted  
214 control measures.

215 Besides the super spreader in Tianjin as discussed in the results, some other  
216 SSEs were also identified based on our criterion in other cities of China, such as in  
217 Ningbo, Zhejiang [23], in Harbin, Heilongjiang [24], and in Hongkong [25]. With the  
218 tendency of pandemic of COVID-19, all the governments should strengthen controls  
219 to prevent SSEs.

220 Previous studies identified multiple underlying reasons for the emergence of SSEs  
221 and high transmission heterogeneity [5]. For example, SSEs of MERS-CoV and  
222 SARS-CoV were largely driven by the environmental and clinical factors, such as  
223 hospital transfer, substantial delay of diagnosis and fundamentally crowding of pop-  
224 ulation [4, 5]. The transmission heterogeneity and the identified SSE of COVID-19 in  
225 Tianjin, however, exhibited different features. The SSE happened in a working place,  
226 and there was no obvious delay of diagnosing the super-spreader (3 days from onset  
227 of symptom to diagnosis). In addition, the longest transmission chain (as shown in  
228 Fig.2) occurred after the initiation of Level one response of public health emergency  
229 by the Tianjin government. The public were then advised to avoid mass gathering  
230 and to stay at home. Obviously, the aforementioned reasons for the transmission

231 heterogeneity in MERS-CoV and SARS-CoV did not hold in the COVID-19 out-  
232 break. Therefore, we suspected the transmission heterogeneity of COVID-19 were  
233 driven by some other epidemiological characteristics, such as asymptomatic trans-  
234 mission and contagiousness during incubation period, i.e., transmission can occur  
235 before symptom onset. The onset of viral shedding prior to the onset of symptoms,  
236 or in cases that remain asymptomatic, is a classic factor that makes infectious dis-  
237 ease outbreaks difficult to control [26]. A recent reported transmission chain of large  
238 size in Harbin, China were triggered by asymptomatic infections [27], which justi-  
239 fied our speculation and highlighted the need for efficient measures such as rapid  
240 testing suspected cases to reduce the transmission from subclinical cases.

241 Additionally, our estimate of transmission potential ( $R < 1$ ) suggested that the  
242 spreading of COVID-19 would not cause large infection in Tianjin. This estimate  
243 was the average trend over the period of data collection (from January 21st to  
244 February 26th), and was considerable lower than that of other studies, especially  
245 at the early stage of COVID-19 outbreak [11]. The differences in the reproductive  
246 number reported from different studies are largely due to differences methods, dif-  
247 ferences in data sources and time periods used to estimate the reproductive number.  
248 In addition, as indicated with our evaluation based on simulation, the value of  $R < 1$   
249 indicated that the control measures imposed in succession by the local government  
250 had been considerable reduce the transmissibility of virus.

251 Regarding the effect of control measure undertaking by the local government,  
252 there were changes in point estimates of  $R$  and  $k$  of before and after government  
253 control, but the difference was not statistically significant with overlapping confi-  
254 dence intervals. Even though, by considering the resulted decrease of 19% infection  
255 (from 165 cases to 135 cases) within 25 days, these control measures still had some  
256 practical implication. And the non-significant changes in parameter estimates might  
257 be due to the short period of observation.

258 Although the spread of COVID-19 is currently under control in China, it is still  
259 facing the risk and challenge brought by the resumption of work and imported cases  
260 from other countries. The verification of transmission heterogeneity and SSE in  
261 Tianjin outbreak of COVID-19 and the analysis of the underlying reasons remind  
262 the government to pay much attention to asymptomatic transmission and other  
263 factors that may lead to the transmission heterogeneity and SSEs. Furthermore,  
264 more efforts are also needed to explore the emergence of SSE and its contributing  
265 factors.

266 It is also worth mentioning the limitation of our study. Totally 34 imported cases  
267 were reported in Tianjin during the period of our study, but we identified 43 trans-  
268 mission chains, which meant some epidemiology links among cases were missing.  
269 The estimate of heterogeneity might be driven up if the confirmed cases were con-  
270 densed into fewer transmission chains. Meanwhile, we only analyzed COVID-19  
271 data from Tianjin (135 confirmed cases), a relatively small sample size compared  
272 with the total number in China (exceeds 80,000 confirmed cases), and our case data  
273 might be subject to bias or under-reporting, which could not destroy our conclusion  
274 of significant heterogeneity [28].

275 In conclusion, we proved that the transmission of COVID-19 is heterogeneous  
276 and identified the existence of one SSE in Tianjin. We also showed that the control  
277 measures undertaken by the local government in effect alleviated the outbreak in  
278 terms of infection size and duration. As a pandemic which is still spreading world-  
279 wide at a startling speed, the transmission characteristics of COVID-19 needs more  
280 exploration and investigation in a large scale.

#### 281 **Competing interests**

282 The authors declare that they have no competing interests.

#### 283 **Author's contributions**

284 Y.J.Z and X.H.Z designed and directed the project; Y.J.Z developed the theoretical framework; Y.Y.L, L.W, and

285 M.Y.L collected and analyzed the data; Y.J.Z Y.Y.L, L.W, and M.Y.L wrote the article.

286 **Acknowledgements**

287 The authors acknowledge support from the National Natural Science Foundation of China (grant number 82041023)  
288 and Zhejiang University special scientific research fund for COVID-19 prevention and control. The funders had no  
289 role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

290 **Author details**

291 <sup>1</sup>Department of Biostatistics, School of Public Health, Peking University, Xueyuan Road, 100191 Beijing, China.

292 <sup>2</sup>Beijing International Center for Mathematical Research, Peking University, Yiheyuan Road, 100871 Beijing, China.

293 <sup>3</sup>School of Mathematical Sciences, Peking University, Yiheyuan Road, 100871 Beijing, China. <sup>4</sup>Center for Statistical  
294 Science, Peking University, Yiheyuan Road, 100871 Beijing, China.

295 **References**

- 296 1. (WHO), W.H.O., et al.: Statement on the second meeting of the international health regulations (2005)  
297 emergency committee regarding the outbreak of novel coronavirus (2019-ncov). geneva: Who; 30 jan  
298 2020.[accessed 01 feb 2020]. Geneva, Switzerland (2005)
- 299 2. Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E., Getz, W.M.: Superspreading and the effect of individual  
300 variation on disease emergence. *Nature* **438**(7066), 355–359 (2005)
- 301 3. Kucharski, A., Althaus, C.: The role of superspreading in middle east respiratory syndrome coronavirus  
302 (mers-cov) transmission. *Euro surveillance* **20**(25), 21167 (2015)
- 303 4. Wong, G., Liu, W., Liu, Y., Zhou, B., Bi, Y., Gao, G.F.: Mers, sars, and ebola: The role of super-spreaders in  
304 infectious disease. *Cell Host & Microbe* **18**(4), 398–401 (2015)
- 305 5. Stein, R.A.: Super-spreaders in infectious diseases. *International Journal of Infectious Diseases* **15**(8), 510–513  
306 (2011)
- 307 6. Kucharski, A.J., Althaus, C.L.: The role of superspreading in middle east respiratory syndrome coronavirus  
308 (mers-cov) transmission. *Eurosurveillance* **20**(25), 21167 (2015)
- 309 7. Lau, M.S.Y., Dalziel, B.D., Funk, S., McClelland, A., Tiffany, A., Riley, S., Metcalf, C.J.E., Grenfell, B.T.:  
310 Spatial and temporal dynamics of superspreading events in the 2014–2015 west africa ebola epidemic.  
311 *Proceedings of the National Academy of Sciences of the United States of America* **114**(9), 2337–2342 (2017)
- 312 8. Faye, O., Boëlle, P.-Y., Heleze, E., Faye, O., Loucoubar, C., Magassouba, N., Soropogui, B., Keita, S., Gakou,  
313 T., Bah, E.H.I., Koivogui, L., Sall, A.A., Cauchemez, S.: Chains of transmission and control of ebola virus  
314 disease in conakry, guinea, in 2014: an observational study. *Lancet Infectious Diseases* **15**(3), 320–326 (2015)
- 315 9. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y., et al.:  
316 Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal*  
317 *of Medicine* (2020)
- 318 10. Riou, J., Althaus, C.L.: Pattern of early human-to-human transmission of wuhan 2019 novel coronavirus  
319 (2019-ncov), december 2019 to january 2020. *Eurosurveillance* **25**(4) (2020)
- 320 11. Liu, Y., Gayle, A.A., Wilder-Smith, A., Rocklöv, J.: The reproductive number of covid-19 is higher compared to  
321 sars coronavirus. *Journal of Travel Medicine* (2020)
- 322 12. Tariq, A., Lee, Y., Roosa, K., Blumberg, S., Yan, P., Ma, S., Chowell, G.: Real-time monitoring the transmission  
323 potential of covid-19 in singapore, february 2020. medRxiv (2020). doi:10.1101/2020.02.21.20026435.  
324 <https://www.medrxiv.org/content/early/2020/03/12/2020.02.21.20026435.full.pdf>
- 325 13. Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S.A., Zhang, T., et al.:  
326 Epidemiology and transmission of covid-19 in shenzhen china: Analysis of 391 cases and 1,286 of their close  
327 contacts. *MedRxiv* (2020)

- 328 14. Endo, A., Abbott, S., Kucharski, A.J., Funk, S., *et al.*: Estimating the overdispersion in covid-19 transmission  
329 using outbreak sizes outside china. *Wellcome Open Research* **5**(67), 67 (2020)
- 330 15. Liu, Y., Eggo, R.M., Kucharski, A.J.: Secondary attack rate and superspreading events for sars-cov-2. *The*  
331 *Lancet* **395**(10227), 47 (2020)
- 332 16. Wu, W., Li, Y., Wei, Z., Zhou, P., Lyu, L., Zhang, G., Zhao, Y., He, H., Li, X., Gao, L., *et al.*: Investigation  
333 and analysis on characteristics of a cluster of covid-19 associated with exposure in a department store in tianjin.  
334 *Zhonghua liu Xing Bing xue za zhi= Zhonghua Liuxingbingxue Zazhi* **41**(4), 489–493 (2020)
- 335 17. Blumberg, S., Lloyd-Smith, J.O.: Inference of  $r$  and transmission heterogeneity from the size distribution of  
336 stuttering chains. *PLoS Computational Biology* **9**(5) (2013)
- 337 18. The Proposal of New Coronavirus Pneumonia Prevention(Sixth Edition).  
338 [http://www.chinacdc.cn/jkzt/crb/zl/szkb\\_11803/jszl\\_11815/202003/W020200309376009304000.pdf](http://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_11815/202003/W020200309376009304000.pdf)
- 339 19. Tianjin Municipal People's Government. <http://www.tj.gov.cn/xw/ztlz/tjsyqfk/yqtb/>
- 340 20. Tianjin: Come up with 18 Policies for Epidemic Prevention and Control.  
341 [http://www.tj.gov.cn/xw/spxw/202001/t20200129\\_3668285.html](http://www.tj.gov.cn/xw/spxw/202001/t20200129_3668285.html)
- 342 21. Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler,  
343 J.: The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases:  
344 Estimation and application. *Annals of Internal Medicine*
- 345 22. Neher, R.A., Dyrda, R., Druelle, V., Hodcroft, E.B., Albert, J.: Potential impact of seasonal forcing on a  
346 sars-cov-2 pandemic. *Swiss Medical Weekly* **150**(1112) (2020)
- 347 23. Health Commission of Ningbo. <http://wjw.ningbo.gov.cn/col/col142/index.html>
- 348 24. All 9 Family Members in New Year's Eve Dinners in Heilongjiang Were Confirmed.  
349 <http://hlj.people.com.cn/GB/n2/2020/0205/c220024-33767665.html>
- 350 25. Nine Members of Hongkong Family Feared Infected After Sharing Hotpot.  
351 [https://www.straitstimes.com/asia/east-asia/](https://www.straitstimes.com/asia/east-asia/coronavirus-nine-members-of-hong-kong-family-feared-infected-after-sharing-hotpot)  
352 [coronavirus-nine-members-of-hong-kong-family-feared-infected-after-sharing-hotpot](https://www.straitstimes.com/asia/east-asia/coronavirus-nine-members-of-hong-kong-family-feared-infected-after-sharing-hotpot)
- 353 26. Fraser, C., Riley, S., Anderson, R.M., Ferguson, N.M.: Factors that make an infectious disease outbreak  
354 controllable. *Proceedings of the National Academy of Sciences of the United States of America* **101**(16),  
355 6146–6151 (2004)
- 356 27. Health Commission of Heilongjiang Province.  
357 <http://yiqing.ljjk.org.cn/index/patients/newsinfo/id/1887.html>
- 358 28. Lloyd-Smith, J.O.: Maximum likelihood estimation of the negative binomial dispersion parameter for highly  
359 overdispersed data, with applications to infectious diseases. *PloS one* **2**(2) (2007)

#### 360 **Additional Files**

##### 361 Additional file 1

362 The additional file 1 is organized as follows. Section A describes detail information of three type of transmission  
363 chain. Section B presents the EM algorithm for the estimation of parameters  $R$  and  $k$ . Section C gives technical  
364 details for the construction of bootstrap confidence interval (CI) of the estimates from EM algorithm

##### 365 Additional file 2 — Data

366 The data information includes chain size, chain type, the Number of primary case, ID, onset of symptoms of primary  
367 case, confirmed date of primary case.