Transmission characteristics of the COVID-19 outbreak in China: a study driven by data Supplementary Material S2

Meili Li[1] Pian Chen[1] Qianqian Yuan[1] Baojun Song[2]
Junling Ma[3,*]

[1]College of Science, Donghua University, Shanghai, China 201620

[2]Department of Applied Mathematics and Statistics,
Montclair State University, Montclair, NJ, USA 07043

[3]Department of Mathematics and Statistics,
University of Victoria, Victoria, BC, Canada V8W 2Y2

[*]corresponding: junlingm@uvic.ca

April 3, 2020

Case Descriptions

Our research is based on the individually documented case descriptions from each province of China (excluding those reported by Hubei province). Here are five exemplary case descriptions out of the total 3547 cases that we collected.

- Case 34 in Tianjin (Patient 230 in Supplementary Material S1): Female, age 52. Lives in Baodi District of Tianjin. A sales representative in Baodi Shopping Center. A coworker of hers visited Beijing on January 16 and 17. During the visit, the coworker contacted a patient with high fever in Beijing, then showed fever on January 25, positive on PCR test and is on waiting to be diagnosed until January 31 (See case 37 below). Case 34 showed fever on January 27, visited Baodi Haibin Hospital on the same day, and has been diagnosed to be the 34th case in Tianjin.

- Case 37 in Tianjin (Patient 233 in S1): Female, age 43. Lives in Baodi District of Tianjin. A coworker of Case 34. Has not recently visited Wuhan. Visited Beijing on January 16 and 17, and contacted a patient with high fever during the visit. Showed fever on January 25. Took medication and the symptom was relieved. Stopped the medication on January 30, and showed fever on the same day. Visited Baodi People's Hospital on January 31, positive on PCR test, and has been diagnosed to be the 37th case in Tianjin.

- Case 53 in Tianjin (Patient 249 in S1): Female, age 46. A coworker of Case 34 and 37 in Tianjin. Worked in Baodi Shopping Center from January 1 to January 23, 2020. Showed fever and coughs on January 31, and visited Baodi People's Hospital on February 1. Diagnosed to be the 53th case in Tianjin.

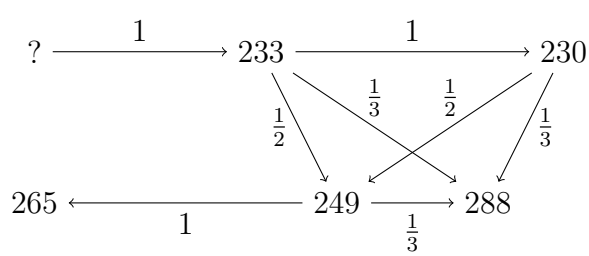
- Case 69 in Tianjin (Patient 265 in S1): Female, age 76. Lives in Baodi District of Tianjin. A relative of Case 53. She did not visit Baodi Shopping Center. Showed fever on February 3, visited Wuqing People's Hospital on February 4, and diagnosed on the same day to be the 69th case in Tianjin.

- Case 92 in Tianjin (Patient 288 in S1): Female, age 63. Lives in Baodi District of Tianjin. Has not recently visited Wuhan. Visited Baodi Shopping Center with other family members on January 23. Showed fever, clogged nose, and vomiting on January 30. Visited Baidi People's Hospital on February 7, diagnosed on February 10 to be the 92nd case in Tianjin.

Dates The dates are discrete variables. However, the incubation period and the periods from symptom onset to hospitalization and diagnosis are commonly considered continuous variables in mathematical modeling. We convert the dates into intervals of continuous time. To do so, we first establish an initial time $t_0 = 0$, to be 0:00AM on December 1, 2019. For example, the symptom onset of Patient 233 (Case 37 of Tianjin) is January 25, which corresponds to the interval $(55, 56)$ days. Her dates of being infected correspond to the range $(46, 48)$ days.

Periods Each period is a time difference between two dates. For example, the incubation period is the time difference between the date of being infected and the symptom onset. If the earliest date corresponds to an interval (a_1, a_2) and the latest date corresponds to an interval (b_1, b_2), where $a_2 < b_1$, then the period between the two dates has a range $(b_1 - a_2, b_2 - a_1)$. For example, the incubation period of Patient 233 has a range $(7, 10)$.

Constructing the contact graph The examples above correspond to the following contact graph.



Since Patient 249 can be infected by either 230 or 233, each of its incoming arcs weighs 1/2. Similarly, each of the incoming arcs of 288 weighs 1/3. Thus, the number of secondary infections of 233 is 11/6.

Estimation of the distributions

To estimate the incubation period and the periods from symptom onset to isolation and diagnosis, we assume that each distribution is either gamma, log-normally, or Weibull distributed (an exponential distribution is also considered for the period from symptom onset to hospitalization).

Likelihood Since each period has a range, the likelihood of observing a period is the probability that the period is within the range. This is applied to both the incubation period and the periods from symptom onset to hospitalization and diagnosis. Let $F(t;\vec{\theta})$ be the cumulative distribution function (CDF) of some period, where $\vec{\theta}$ is the set of parameters for the distribution, e.g., for a gamma distribution, $\vec{\theta} = (s, \mu)$ where s is the shape parameter and μ is the scale parameter. It is customarily assumed that the period for each individual is independently identically distributed. If the range for a specific period of Patient n is (t_n, τ_n), then the likelihood function is

$$L(\vec{\theta}) = \prod_n F(\tau_n, \vec{\theta}) - F(t_n, \vec{\theta})\,. \tag{1}$$

The log-likelihood function is thus

$$\ell(\vec{\theta}) = \ln L(\vec{\theta}) = \sum_n \ln \left[F(\tau_n, \vec{\theta}) - F(t_n, \vec{\theta}) \right]\,. \tag{2}$$

Model selection With the log-likelihood function, we then use Markov Chain Monte Carlo (MCMC) to find the posterior distribution for the distribution parameter $\vec{\theta}$, using a Gibbs sampler through the R2jags package of the R language. The prior distributions are taken as wide uniform distributions. The distribution with the smallest Deviance Information Criterion (DIC) is selected as the best-fit model [1], with the results summarized in Tables 1-3.

References

[1] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2013.

Patients	Gamma	Log-normal	Weibull
All	**1420.5**	1439.7	1424.2
Males	**711.6**	722.5	713.0
Females	**709.8**	717.0	713.1

Table 1: The DIC for the fitting results of the incubation periods. The smallest DIC values are printed in bold.

Patients	Exponential	Gamma	Log-normal	Weibull
All	7249.9	**7234.8**	7413.6	7243.3
Onset before Jan 23	1625.4	1612.5	1658.6	**1607.3**
Onset after Jan 23	5568.4	**5541.5**	5666.8	5550.7

Table 2: The DIC for the fitting results of the period from symptom onset to hospitalization. The smallest DIC values are printed in bold.

Patients	Gamma	Log-normal	Weibull
All	**14914.8**	15130.2	14917.2
Onset before Jan 23	**3221.3**	3254.8	3231.6
Onset after Jan 23	**11271.3**	11411.3	11294.2

Table 3: The DIC for the fitting results of the period from symptom onset to diagnosis. The smallest DIC values are printed in bold.