Supplementary methods

Calculating the arrival rate of imported cases

The number of exposed cases at time t is simplified as

$$\begin{aligned}
I_{i_exposed}(s) &= \int_0^\infty \beta I_i(s-a)(1-F(a))da \\
&= \int_0^\infty \beta I_0 e^{s-a}\left(1-\left(1-e^{-a\left(\frac{1}{\tau}+\frac{1}{T_g}\right)}\right)\right)da \\
&= \beta I_0 e^{\Lambda s}\int_0^\infty e^{-a\left(\frac{1}{\tau}+\frac{1}{T_g}+\Lambda\right)} \\
&= \beta I_0 e^{\Lambda s}\left[\frac{e^{-a\left(\frac{1}{\tau}+\frac{1}{T_g}+\Lambda\right)}}{\left(\frac{1}{\tau}+\frac{1}{T_g}+\Lambda\right)}\right]_0^\infty
\end{aligned} \tag{S1}$$

$$I_{i\_exposed}(t) = \left(\frac{\beta}{\Lambda+\frac{1}{\tau}+\frac{1}{T_g}}\right) I_0 e^{\Lambda t} \tag{S2}$$

Therefore, the rate of imported cases is

$$Imp^+ = (1-c)M_{ji}I_{i\_exposed} = \left(\frac{\beta}{\Lambda+\frac{1}{\tau}+\frac{1}{T_g}}\right)(1-c)M_{ji}I_0 e^{\Lambda t} \tag{S3}$$

Calculating number of imported and secondary cases

We have the following formulas

$$\frac{dImp}{dt} = Imp^+ - \left(\frac{1}{T_g}+\frac{1}{T_{qr}}\right) Imp \tag{S4}$$

$$\frac{dSec}{dt} = \beta\frac{S_i}{N}Imp - \left(\frac{1}{T_g}+\frac{1}{T_{qr}}\right) Sec \tag{S5}$$

Now, to solve the Eq.(S4) for Imp, let $Imp = y$. Thus,

$$\frac{dy}{dt} + \left(\frac{1}{T_g}+\frac{1}{T_{qr}}\right) y = \left(\frac{\beta}{\Lambda+\frac{1}{\tau}+\frac{1}{T_g}}\right)(1-c)M_{ji}I_0 e^{\Lambda t} \tag{S6}$$

Again, let, $p = \left(\frac{1}{T_g} + \frac{1}{T_{qr}}\right)$, $Q(t) = \left(\frac{\beta}{\Lambda + \frac{1}{\tau} + \frac{1}{T_g}}\right)(1-c)M_{ji}I_0 e^{\Lambda t}$ and $u(t) = e^{\int p dt}$. Therefore, we have, $u\frac{dy}{dt} + upy = uQ(t)$, then by using the method of integration by parts we get, $uy = \int uQ(t)dt$ and eventually, $e^{pt}y = \int e^{pt}Q(t)dt$. Finally, we have,

$$
\begin{aligned}
e^{\left(\frac{1}{T_g}+\frac{1}{T_{qr}}\right)t} y &= \int e^{\left(\frac{1}{T_g}+\frac{1}{T_{qr}}\right)t} \left(\frac{\beta}{\Lambda + \frac{1}{\tau} + \frac{1}{T_g}}\right)(1-c)M_{ji}I_0 e^{\Lambda t} dt \\
\Rightarrow e^{\left(\frac{1}{T_g}+\frac{1}{T_{qr}}\right)t} y &= \left(\frac{\beta}{\Lambda + \frac{1}{\tau} + \frac{1}{T_g}}\right)(1-c)M_{ji}I_0 \frac{e^{\left(\Lambda + \frac{1}{T_g}+\frac{1}{T_{qr}}\right)t}}{\left(\Lambda + \frac{1}{T_g} + \frac{1}{T_{qr}}\right)}
\end{aligned}
$$

Hence,

$$
Imp(t) = \left(\frac{1}{\Lambda + \frac{1}{T_g} + \frac{1}{T_{qr}}}\right)\left(\frac{\beta}{\Lambda + \frac{1}{\tau} + \frac{1}{T_g}}\right)(1-c)M_{ji}I_0 e^{\Lambda t} \tag{S7}
$$

Similarly, by following the same approach, solving Eq. (S5) for Sec we get,

$$
Sec(t) = \left(\frac{1}{\Lambda + \frac{1}{T_g} + \frac{1}{T_{qr}}}\right)\left(\frac{\beta}{\Lambda + \frac{1}{T_g} + \frac{1}{T_{qr}}}\right)\left(\frac{\beta}{\Lambda + \frac{1}{\tau} + \frac{1}{T_g}}\right)(1-c)M_{ji}I_0 e^{\Lambda t} \tag{S8}
$$

where τ is the incubation period, Λ is the growth rate and T_{qr} is the time to quarantine. Because of $\frac{dI_i}{dt} \approx \Lambda I_i$, we have the infected number at i as a function of time t, that is $I_i(t) = I_0 e^{\Lambda t}$. If we set $\beta = \frac{R_0}{T_g}$, we obtained the infectious disease spreading control function given k transmission waves before the community transmission:

$$
\begin{aligned}
\eta(k) &= \left(\frac{R_0}{T_g}\right)^k \left(\frac{1}{\Lambda + \frac{1}{T_g} + \frac{1}{T_{qr}}}\right)^k \left(\frac{1}{\Lambda + \frac{1}{\tau} + \frac{1}{T_g}}\right) \\
&= \left(\frac{R_0 T_{qr}}{R_0 T_{qr} + T_g}\right)^k \left(\frac{1}{\Lambda + \frac{1}{\tau} + \frac{1}{T_g}}\right)
\end{aligned} \tag{S9}
$$

The formula was simplified after we replaced $\Lambda + \frac{1}{T_g}$ by $\frac{R0}{T_g}$ where, $k = 0, 1, \ldots$. Under this notation, secondary infection cases can simply be calculated as $\eta(k=2)$ multiplied by the $(1-c)M_{ji}I_0 e^{\Lambda t}$:

$$
Imp(t) = \eta(k=1)(1-c)M_{ji}I_0 e^{\Lambda t} \tag{S10}
$$

$$Sec(t) = \eta(k=2)(1-c)M_{ji}I_0e^{\Lambda t} \quad \text{(S11)}$$

Finally, the cumulative number of imported, $CI_{Imp}(t)$, and secondary, $CI_{Sec}(t)$, cases were calculated by integrating $\beta(1-c)M_{ji}\eta(0)I_i$ and $\beta(1-c)M_{ji}\eta(1)I_i$ respectively over t.

Supplementary figures and tables

Outbreak spreads with low transmissibility

Since R_0 of COVID-19 was not fully known yet, we predicted the outbreak spreads using different R_0 settings. In most cases, the effective reproduction number R may decrease after many control measures are conducted [1]. Therefore, in addition to $R_0 = 2.92$, we made the same calculation under a low transmission setting with $R_0 = 1.4$, which was the lower bound of World Health Organization's estimate [2] and a medium transmission setting with $R_0 = 1.68$, which was near the lower bound of R_0 estimated by the MRC Centre for Global Infectious Disease Analysis at Imperial College London [3], to evaluate the arrival times of outbreak emergence among the top 10 visiting cities. Under these scenarios, following the suggestion from a recent study [4], we considered the initial infected number as 1000 on 31 December.

When $R_0 = 1.4$, the cumulative secondary infected number was slowly linearly increasing and the top visiting city had 6.8 cases on 28 January, which was below the critical threshold line $\nu = 8$ (Figure S1). For each of the top 10 cities, the arrival time at which the cumulative number of secondary cases larger than the critical threshold was determined (Table S1). Beijing, Shanghai, Guangzhou had the shortest required time periods. However, the arrival times of outbreak emergence in the above cities were between 31-34 days, corresponding to the end of January and the early of February, which were about more than 10 days later comparing to the actual reported data (Table S1). The mean arrival time of the top 10 visiting cities was 39 days, which was 21 days later than $R_0 = 2.92$. With $R_0 = 1.68$, the mean arrival time was 26.3 days. Given the reporting delay was about 10 days, we found that $R_0 = 2.92$ with incubation period 5.2 days gave a better prediction compared to other low R_0 or long incubation period settings.

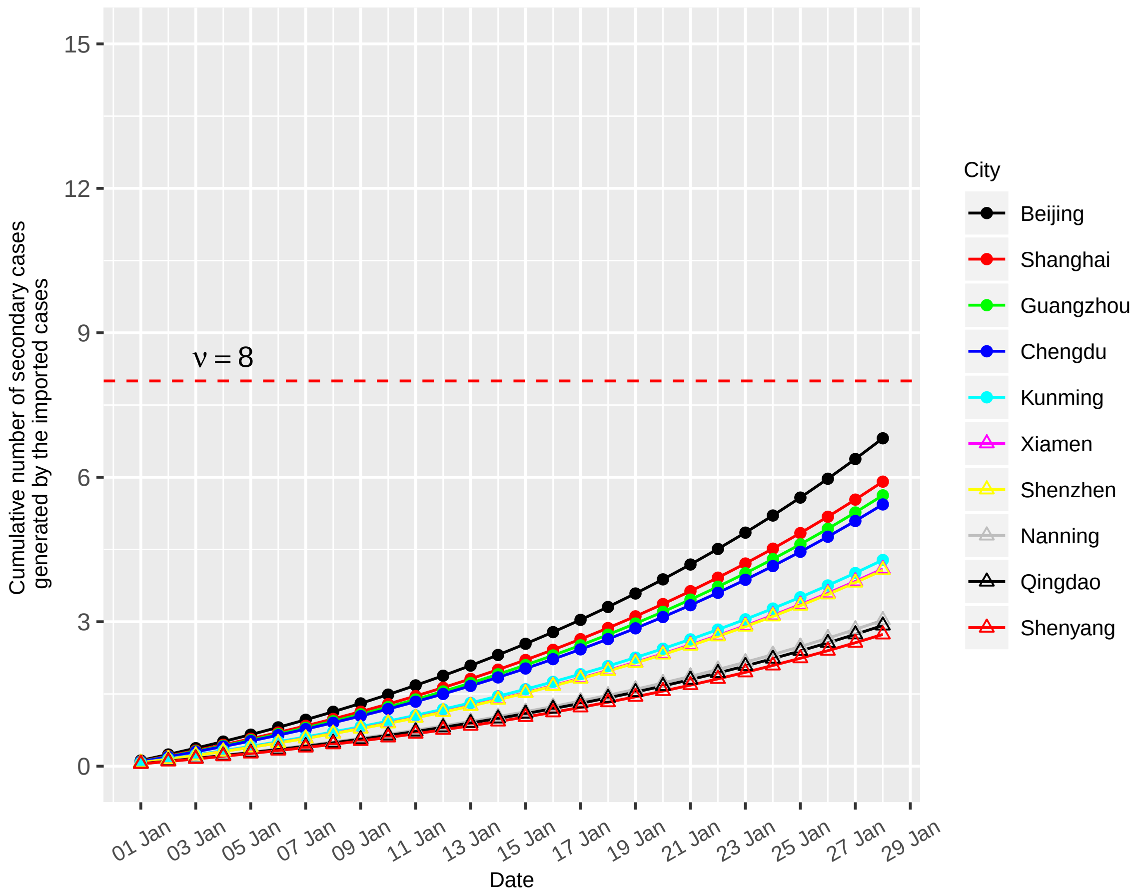

Figure S1. Cumulative number of secondary cases using a low setting of transmissibility in all top 10 visiting cities. Under the basic reproduction number $R_0 = 1.4$ and incubation time $\tau = 5.2$ days were used, all cities had a lower number of cumulative secondary cases than the critical threshold number $\nu = 8$. The top most visiting city Beijing had the secondary cases 7 persons by 28 January whereas the 10^{th} city had only 3.

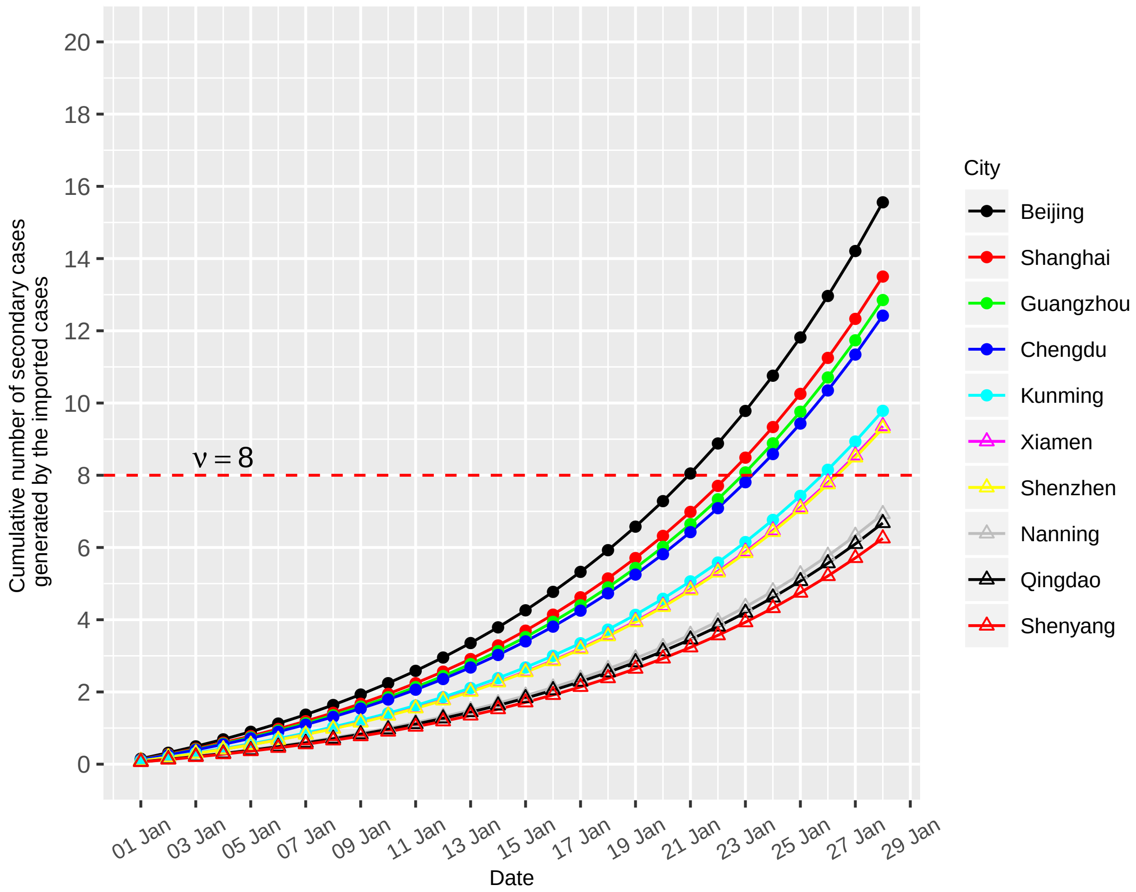

Figure S2. Cumulative number of secondary cases using a medium setting of transmissibility in all top 10 visiting cities. Under the basic reproduction number $R_0 = 1.68$ and incubation time $\tau = 5.2$ days, seven out of top 10 visiting cities had the number of cumulative secondary cases higher than the threshold by 28 January. Beijing had a total of 15.5 secondary cases.

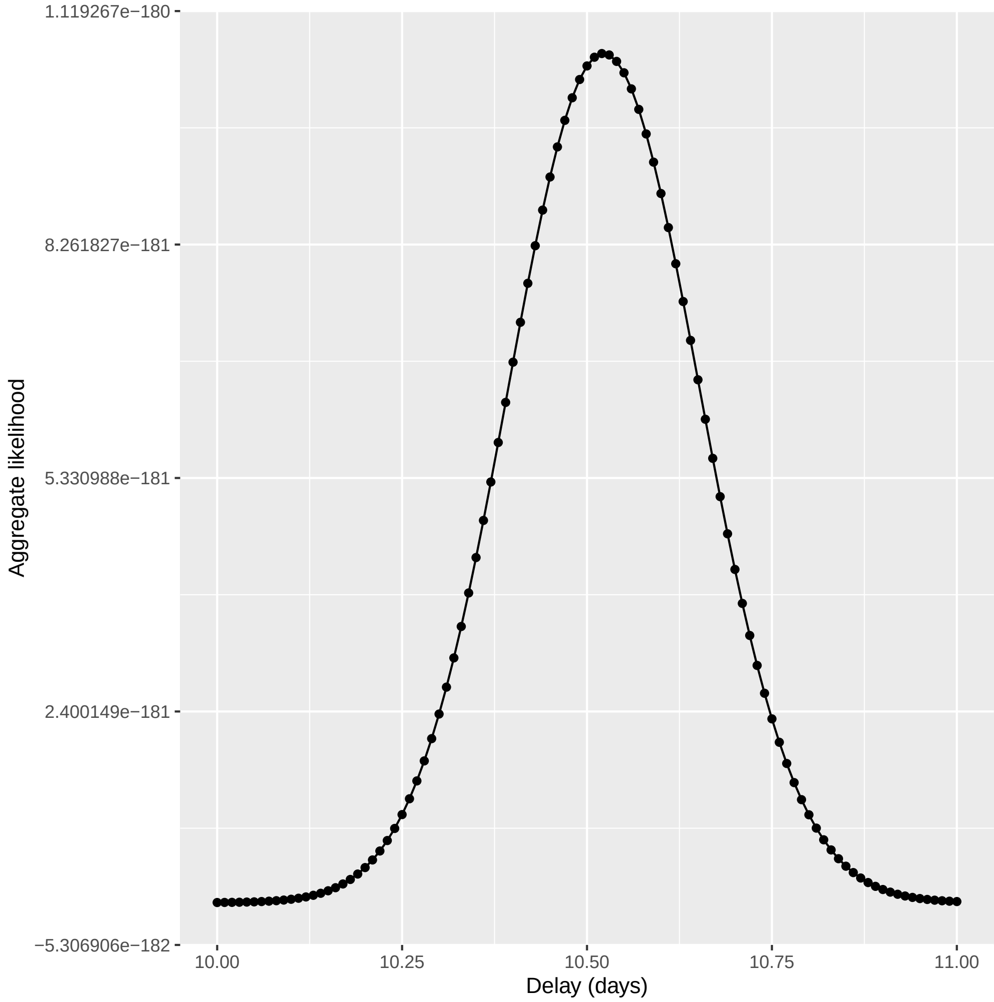

Figure S3. Maximum likelihood estimation of accumulated delay time (day) of disease detection for all top 10 visiting cities.

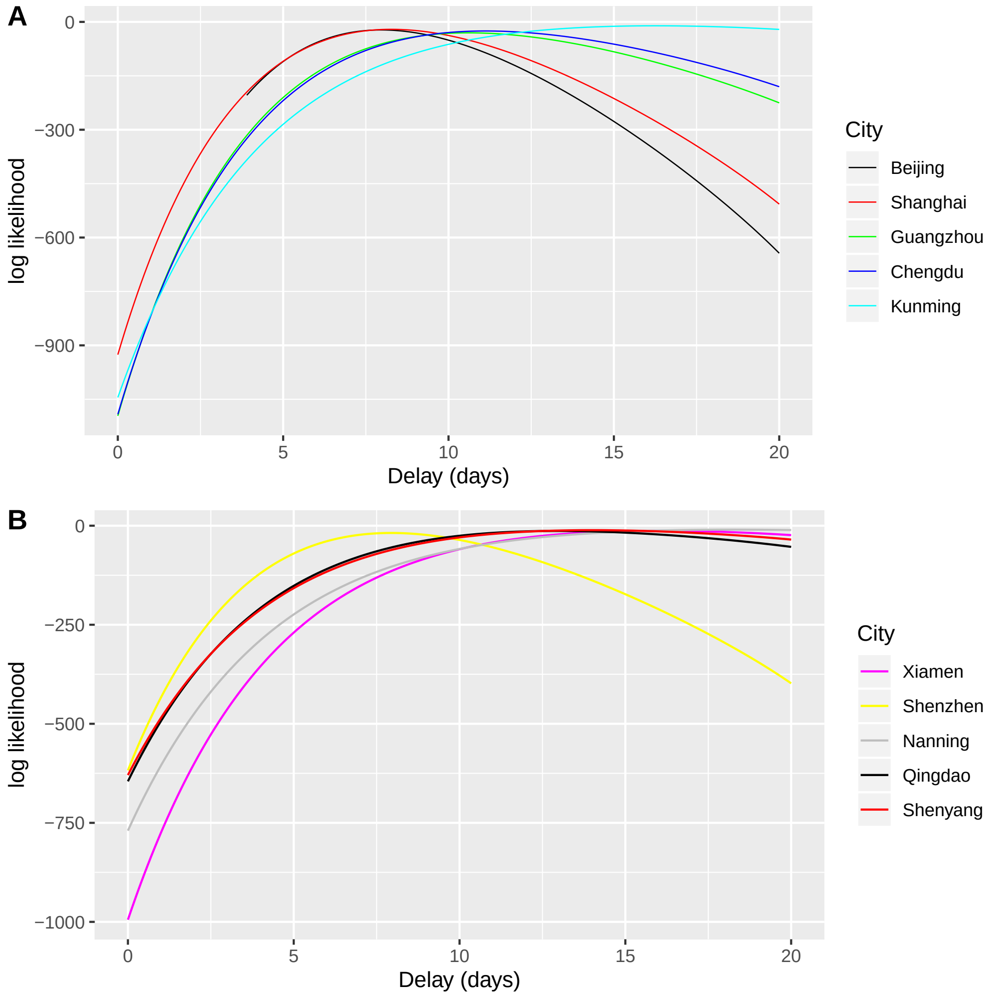

Figure S4. Parameter estimation of reporting delay time using maximum likelihood approach. A binomial distribution was assumed, with the total number of infections as the total population, and the probability of infection in each visiting city calculated from the migration matrix and the estimated numbers of imported and secondary cases. Given the observed cases between 22 to 28 January 2020, maximum likelihood estimate for reporting delay of infections, in days, for each of the top 10 visiting cities was obtained by maximizing the binomial loglikelihood function. (A) Top five visiting cities, (B) 2nd top five visiting cities.

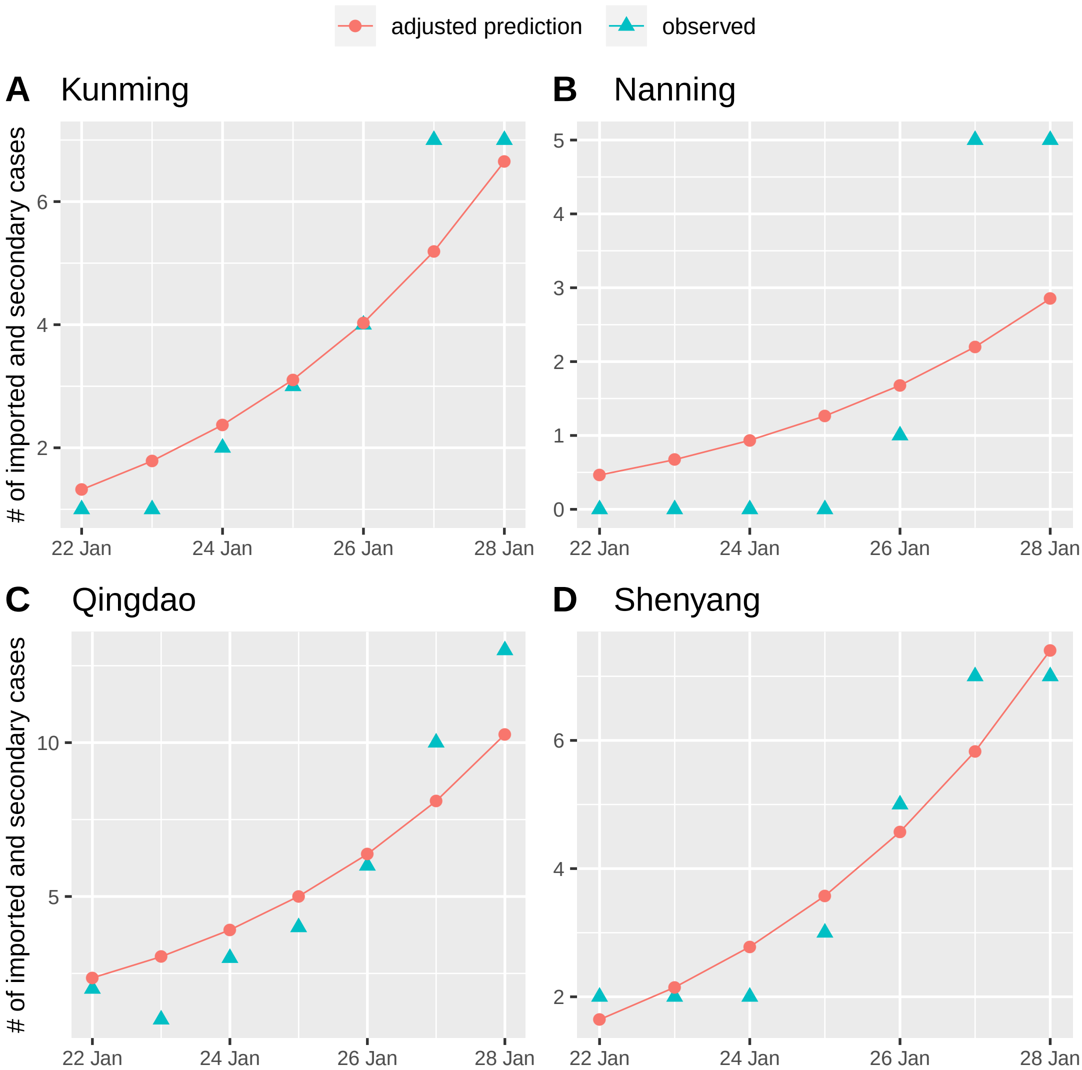

Figure S5. Observed number of confirmed cases and predicted number of imported and secondary cases. The predicted number of cases were adjusted by the reporting delay after using maximum likelihood estimation. The four cities that have the actual longer arrival times are listed.

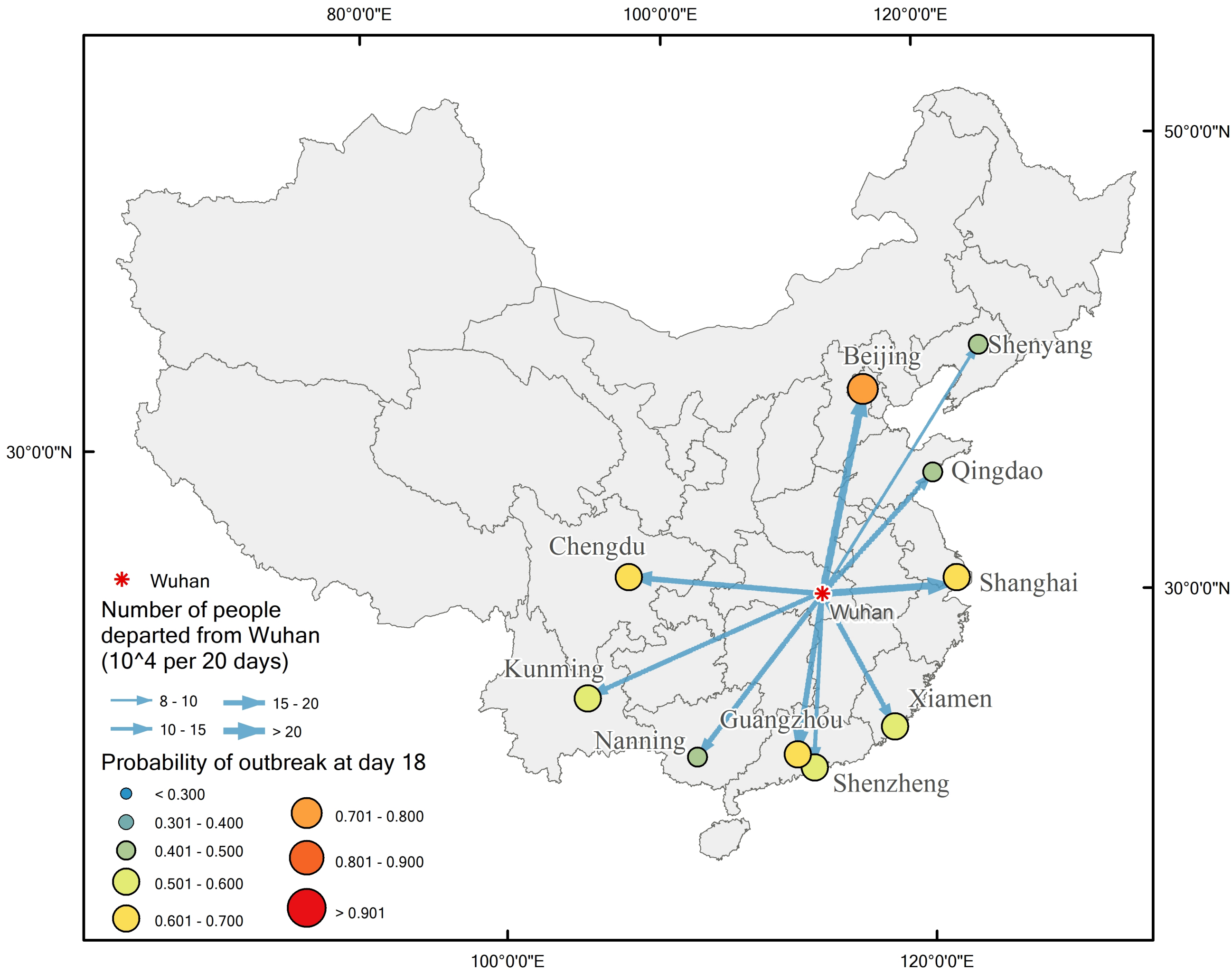

Figure S6. Risk map of COVID-19 for the top 10 visiting cities from Wuhan in Mainland China.

Table S1. Actual and predicted arrival time of outbreak emergence at top 10 connected cities in China. $R_0 = 1.4$ and $R_0 = 1.68$ with incubation time $\tau = 5.2$ days were used. CI_{Sec} is the cumulative number of secondary infected cases generated by the imported cases. The actual arrival time of outbreak is defined as the the date when the number of cumulative cases is larger than the threshold number 8 and the number of newly reported cases is larger than 5.

City	Actual arrival time (day)	$R_0 = 1.4$			$R_0 = 1.68$		
		Predicted arrival time (day)	CI_{Sec} at mean critical time (day 39)	Probability of outbreak at mean critical time (day 39)	Predicted arrival time (day)	CI_{Sec} at mean critical time (day 26.3)	Probability of outbreak at mean critical time (day 26.3)
Beijing	<23	31	13.2	0.715	21	13.0	0.709
Shanghai	<23	33	11.4	0.664	23	11.3	0.658
Guangzhou	25	34	10.9	0.646	23	10.7	0.640
Chengdu	26	35	10.5	0.633	24	10.3	0.627
Kunming	29	39	8.3	0.546	26	8.2	0.540
Xiamen	27	40	7.9	0.530	27	7.8	0.524
Shenzhen	25	40	7.9	0.528	27	7.7	0.522
Nanning	>30	45	5.9	0.428	30	5.8	0.423
Qingdao	>30	46	5.7	0.417	31	5.6	0.412
Shenyang	>30	47	5.3	0.396	31	5.2	0.391

Table S2. Maximum likelihood estimates of reporting delay (day) based on negative binomial distribution and binomial distribution. k represents over dispersion parameter in the negative binomial distribution and 95% CI is for 95% confidence interval.

City	Binomial model		Negative binomial model		
	Delay (days)	95% CI	Delay (days)	95% CI	k
Beijing	8	$7.5-8.5$	8.1	$7.4-8.7$	80
Shanghai	8.3	$7.7-8.9$	8.4	$7.6-9.1$	80
Guangzhou	10.7	$9.9-11.4$	11.4	$9.5-13.0$	4.37
Chengdu	11.2	$10.3-12.0$	11.7	$10.0-13.2$	5.92
Kunming	16.3	$14.8-17.9$	16.3	$14.8-18.0$	80
Xiamen	16	$14.5-17.7$	16.1	$14.5-17.7$	80
Shenzhen	8	$7.3-8.7$	8	$7.2-8.8$	80
Nanning	18.3	$16.1-19.9$	19.1	$15.4-20.0$	1.67
Qingdao	13	$11.7-14.4$	13	$11.6-14.5$	80
Shenyang	14.1	$12.6-15.7$	14.1	$12.5-15.7$	80

References

[1] Steven Riley, Christophe Fraser, Christl A Donnelly, Azra C Ghani, Laith J Abu-Raddad, Anthony J Hedley, Gabriel M Leung, Lai-Ming Ho, Tai-Hing Lam, Thuan Q Thach, et al. Transmission dynamics of the etiological agent of sars in hong kong: impact of public health interventions. *Science*, 300(5627):1961–1966, 2003.

[2] World Health Organization (WHO). Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). `https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulati-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-n` 2020.

[3] Natsuko Imai, Anne Cori, Ilaria Dorigatti, Marc Baguelin, Christl A Donnelly, Steven Riley, and Neil M Ferguson. Report 3: Transmissibility of 2019-nCoV. `https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf`, 2020.

[4] Jonathan M Read, Jessica Re Bridgen, Derek At Cummings, Antonia Ho, and Chris P Jewell. Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions. *medRxiv*, 2020.