XINN, Volume ■ ■

Supplemental Information

Modeling the Epidemic Trend of the 2019 Novel Coronavirus Outbreak in China

Mingwang Shen, Zhihang Peng, Yanni Xiao, and Lei Zhang

Supplementary Appendix:

This is a supplementary document describing mathematical modelling details presented in the main text and parameters estimation.

1. Model formulation

We proposed a dynamic compartmental model to describe the transmission of COVID-19 in Hubei province, China. The population is divided into six compartments (**Figure S1**): susceptible individuals (*S*), asymptomatic (but infectious) individuals during the incubation period (*E*), undiagnosed infectious individuals with symptoms (*I*), diagnosed individuals with isolation and treatment (*T*), recovered (*R*) and dead (*D*) individuals. The total population size is denoted by *N*, where *N*=*S*+*E*+*I*+*T*+*R*.

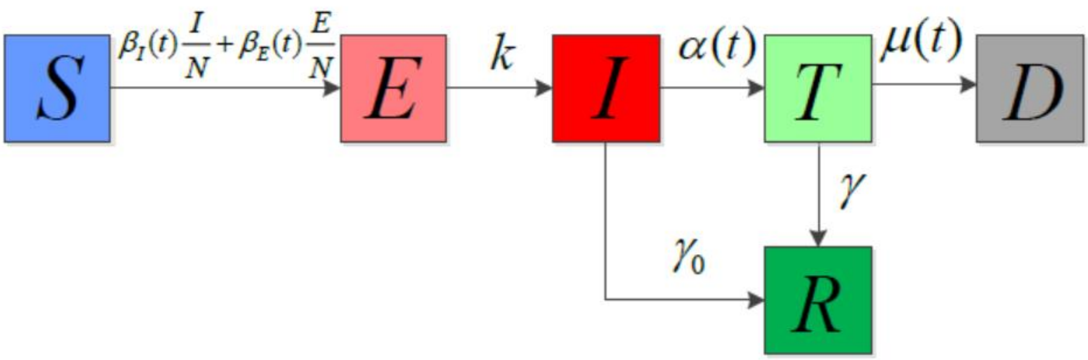

Figure S1. A schematic flow diagram of the transmission of COVID-19.

Susceptible individuals become infected by contacts with latent (*E*) and undiagnosed infectious individuals with symptoms (*I*) in the public space (e.g. public transportations, supermarkets, offices, etc) and household (home or other private space). The overall rate of infection (total new infetions) is given by the sum of rates of infections via these routes. That is,

$$\Lambda_{total} = \Lambda_{pub} + \Lambda_{pri}. \quad (1)$$

For each of route of transmission,

(a) public contacts

$$\Lambda_{pub} = \beta_E^{pub}(t)\frac{(S-S_f)E}{N-N_f} + \beta_I^{pub}(t)\frac{(S-S_f)I}{N-N_f}, \quad (2)$$

(b) household contacts

$$\Lambda_{pri} = \beta_E^{pri}(t)\frac{S_f E}{N_f} + \beta_I^{pri}(t)\frac{S_f I}{N_f}. \quad (3)$$

where

$$\beta_I^{pub}(t) = \beta m_1(t)(1-\theta p_1(t)), \beta_E^{pub}(t) = (1-\varepsilon)\beta_I^{pub}(t),$$
$$\beta_I^{pri}(t) = \beta m_2(t)(1-\theta p_2), \beta_E^{pri}(t) = (1-\varepsilon)\beta_I^{pri}(t). \quad (4)$$

Here β denotes the probability of transmission per contact with the infectious individuals with symptoms. We assumed that for contacts with the latent individuals this probability is lower, i.e. $(1-\varepsilon)\beta$ where $0 \leq \varepsilon \leq 1$ denotes the reduction in per-act transmission probability. The parameters $m_1(t)$ and $m_2(t)$ represent the average number of daily person-to-person contacts in the public space and household, $p_1(t)$ and p_2 denote the proportion of mask usage in the public space and household, respectively, and θ is the effectiveness of face mask/respirators in infection prevention.

Estimation of the population size and number of susceptibles for each route of transmission

For household contacts, the overall population size (N_f) is estimated as the total number of households members that are at risk of COVID-19 infection, whereas the number of susceptible households members (S_f) is the difference between N_f and the number of infected individuals in these households. We assumed that the number of the households at risk of infection is the same as the number of individuals infected in public space because the probability of two or more household members being infected at the same time but at different public venues is very small. Hence, the entry of N_f is $r\Lambda_{pub}$, where r is the average number of household members in a Chinese family and Λ_{pub} is the number of individuals infected in public space as shown in eq. (2). We assumed that the infected family members become recovered after the mean period $1/\xi$. Further, the entry of susceptible household members S_f is $(r-1)\Lambda_{pub} - \Lambda_{pri}$, where Λ_{pri} denotes the number of infected household members through household transmission as shown in eq. (3).

For public contacts, the overall population size is the number of residents (N) in Hubei province minus the overall population size (N_f) in household contacts, whereas the number of susceptibles is the number of individuals free of COVID-19 infection (S) minus the susceptibles (S_f) in household contacts.

Modelling disease progression

Individuals in the incubation period (E) progress to the infectious compartment (infected but undiagnosed) at a rate k. Infectious individuals are assumed to be diagnosed at the rate $\alpha(t)$ and then isolated and treated. We also assumed very strict isolation so isolated individuals could not further infect others. Infected individuals are assumed to recover naturally at the rate γ_0. Treated individuals recover at the rate γ or die due to the disease at the rate $\mu(t)$. The model is described by the following system of ordinary differential equations:

$$\begin{cases} \frac{dS}{dt} = -\Lambda_{pub} - \Lambda_{pri}, \\ \frac{dN_f}{dt} = r\Lambda_{pub} - \xi N_f, \\ \frac{dS_f}{dt} = (r-1)\Lambda_{pub} - \Lambda_{pri} - \xi S_f, \\ \frac{dE}{dt} = \Lambda_{pub} + \Lambda_{pri} - kE, \\ \frac{dI}{dt} = kE - \alpha(t)I - \gamma_0 I, \\ \frac{dT}{dt} = \alpha(t)I - (\gamma + \mu(t))T, \\ \frac{dR}{dt} = \gamma_0 I + \gamma T, \\ \frac{dD}{dt} = \mu(t)T. \end{cases} \qquad (5)$$

The cumulative number of deaths is tracked by the last equation of D in eq. (5) and the cumulative number of confirmed cases C (note that C is not an epidemiological state) is governed by the equation

$$\frac{dC}{dt} = \alpha(t)I. \qquad (6)$$

2. Data and parameter estimation

We collected the data on the number of daily and cumulative confirmed cases and deaths from 15th January 2020 (as the starting date for the epidemic model, i.e. t=0, because after which the data were reported regularly) to 30th August 2020 from National Health Commission [1] and the Health Commission of Hubei Province [2]. The mean incubation time for COVID-19 is about 5.2 days ($1/k$=5.2) [3]. The mean time from symptoms to natural recovery is approximately ten days ($1/\gamma_0$=10) [4]. The mean number of members in a Chinese household is four (r=4) [5]. The average recovery period for treated individuals is 2 weeks ($1/\gamma$=14) [6]. We assumed that the duration of recovery for infected family members is 4 weeks ($1/\xi = 28$). The probability of transmission per contact with latent individuals is assumed to be half ($1 - \varepsilon$=0.5 [0.1-0.9]) of that with infectious individuals [6].

The proportion of facial mask usage in the public space was drastically increased during the outbreak in Wuhan [7][8]. We assumed a logistic growth for this percentage (**Figure S2a**), i.e. $p_1(t) = p_{ini} + \frac{\bar{p}-p_{ini}}{1+\exp(-0.5(t-t_{ini}))}$ where $t_{ini} = 11$ is the time when the metropolitan-wide quarantine (23rd Janurary) was initiated with additional three days delay to show the effect of quarantine, p_{ini} is the base proportion of facial mask usage in the public space before the outbreak, and $\bar{p}$ is the maximum proportion of facial mask usage in the public space during the outbreak. An observational survey [9] in Beijing showed that about 10% of people wear facial mask in routine life in winter, so we chose $p_{ini} = 10\%$ as the base value of the proportion of facial mask usage in the absence of COVID-19. The data in [7] showed that 97.6% of customers wore facial masks in shopping places in Wuhan during quarantine (23rd January-10th February 2020). Another online survey [10] showed that 97.3%-99.3% of people wore facial mask in the public space. Thus, we chose $\bar{p} = 97.6\%$ (97.3%-99.3%). The coverage ratio of facial mask in private space p_2 is set to zero [11]. The effectiveness of mask in preventing infection (θ) is chosen to be 85% (66%-93%) based on a recent meta-analysis on the effectiveness of face masks for COVID-19 [12].

The data in [13] show that the number of daily contacts had reduced by 80% during the outbreak in Wuhan and Shanghai, so we assumed that the average number of daily contacts in the public space $m_1(t)$ (**Figure S2d**) is reduced by 80% in the base case, described by a decreasing logistic function $m_1(t) = m_{ini} + \frac{0.2m_{ini}-m_{ini}}{1+\exp(-0.5(t-t_{ini}))}$, where m_{ini} is the background daily contact number in the public space before the quarantine and will be estimated by model fitting. Home confinement led to three times longer 'stay-at-home' duration than the pre-quarantine level [5]. We assumed that the average number of daily contacts in a household (**Figure S2e**) increased from 4 to 12, described by an increasing logistic function $m_2(t) = 4 + \frac{12-4}{1+\exp(-0.5(t-t_{ini}))}$.

The diagnosis rate increases gradually as more health resources become available. We described it by an increasing logistic function (**Figure S2b**) $\alpha(t) = \alpha_{ini} + \frac{\bar{\alpha}-\alpha_{ini}}{1+\exp(-0.5(t-t_{\alpha}))}$, where $\alpha_{ini} = 1/7$ is the initial diagnosis rate (i.e. the mean time from symptoms onset to diagnosis is 7 days) [14], $\bar{\alpha} = 1/3$ is the maximal diagnosis rate during the epidemic (i.e. the mean time from symptoms onset to diagnosis is shortened to 3 days) [6], and t_{α} is the time when the diagnosis rate is half of the initial and maximal diagnosis rate (to be estimated by data fitting).

The disease-induced death rate of treated individuals $\mu(t)$ (**Figure S2c**) in Hubei Province decreases over time as more health resources become available [6][15]. We used an decreasing logistic function $\mu(t) = \mu_{ini} + \frac{\bar{\mu}-\mu_{ini}}{1+\exp(-\mu_0(t-t_{\mu}))}$, where μ_{ini} and $\bar{\mu}$ $(< \mu_{ini})$ are the initial and minimal disease-induced death rate, μ_0 is the decrease rate of death rate, and t_{μ} is the time when the death rate is half of the initial and minimal death rate. All the four parameters will be estimated by data fitting.

The total population size in Hubei Province is 59,170,000 based on the China Population and Employment Statistics Yearbook in 2019. The initial values of the disease states in 15th Jan 2020 in Hubei Province are given by I(0)=41, T(0)=0, R(0)=0, D(0)=2, N(0)=59,170,000, $N_f(0) = rI(0) = 164$, and $S_f(0) = (r-1)I(0) = 123$. We left E(0) to be estimated by the fitting. The above parameter values are shown in **Table S1**.

Most of the parameters outside Hubei are the same as in Hubei Province (**Table S1**) and we listed these different parameter values in **Table S2.** The initial values of the disease states in 20th Jan 2020 outside Hubei are given by I(0)=21, T(0)=0, R(0)=0, D(0)=0, N(0)= 1,400,050,000-59,170,000=1,340,880,000, $N_f(0) = rI(0) = 84$, and $S_f(0) = (r-1)I(0) = 63$.

We calibrated the model by the daily confirmed cases and deaths data from 15th January 2020 to 30th August 2020 by using a nonlinear least-squares method (**Figure S3**). The unknown parameters (**Table S1,S2**) were sampled within their ranges by the Latin hypercube sampling method and repeated 1000 times. For every simulation, we calculated the sum of square errors between the model output and data, and selected the top 10% with the least square errors to generate 95% confidence intervals. All analyses and simulations were performed in MATLAB R 2019b.

Based on these estimated parameter values, we used the model (Eq. (1)-(6)) to forecast the epidemic trend with and without social distancing, including daily confirmed cases and deaths over one year since the beginning of the epidemic (**Figure 1** in the main text).

Table S1. The values of parameters based on references or estimated by nonlinear least squares (NLS) methods in Hubei Province.

Parameter	Description	Range or 95% CI from NLS	Source
$1/k$	The mean incubation time (days)	5.2 (4.1-7.0)	[3]
$1/\gamma_0$	The mean time from symptoms onset to natural recovery (days)	10	[4]
$1/\gamma$	The mean time from diagnosis to recovery	14	[6]
$1/\xi$	The mean recovery period for infected family members (days)	28	[16]
r	The mean number of members in a family	4	[5]
E(0)	The initial value of latent individuals	499.9905 (443.1596-556.8214)	NLS
β	The per-act transmission probability in contact with infected individuals with symptoms	0.0550 (0.0539-0.0561)	NLS
ε	The reduction in per-act transmission probability if infection is in latency	50% (10-90%)	[4]
$p_1(t)$	The usage percentage of facial mask in the public space	$p_{ini} + \dfrac{\bar{p} - p_{ini}}{1 + \exp(-0.5(t - t_{ini}))}$	[7][8]
t_{ini}	Time when behaviour change begins (3 days after quarantine)	11	[7]
p_{ini}	Base percentage of facial mask usage in the public space before the epidemic	10%	[9]
$\bar{p}$	Percentage of facial mask usage in the public space during quarantine	97.6% (97.3-99.3%)	[7][10]
p_2	The usage percentage of mask in the private space	0%	[11]
θ	The effectiveness of mask in preventing infection	0.85 (0.66-0.93)	[12]
$m_1(t)$	The average number of daily contacts in the public space	$m_{ini} + \dfrac{0.2m_{ini} - m_{ini}}{1 + \exp(-0.5(t - t_{ini}))}$	[13]
m_{ini}	Base daily contact number in the public space	10.3676 (10.2540-10.4813)	NLS
$m_2(t)$	The average number of daily contacts in the households	$4 + \dfrac{12 - 4}{1 + \exp(-0.5(t - t_{ini}))}$	[5]

$\alpha(t)$	The diagnosis rate	$\alpha_{ini} + \frac{\bar{\alpha} - \alpha_{ini}}{1 + \exp(-0.5(t - t_\alpha))}$	[16]
α_{ini}	Initial diagnosis rate	1/7	[14]
$\bar{\alpha}$	Maximum diagnosis rate during the epidemic	1/3	[6]
t_α	The time when the diagnosis rate is half of the maximum diagnosis rate	24.6750 (20.1253-29.2247)	NLS
μ(t)	Disease-induced death rate	$\mu_{ini} - \frac{\bar{\mu} - \mu_{ini}}{1 + \exp(-\mu_0(t - t_\mu))}$	[6][15]
μ_{ini}	Initial disease-induced death rate	0.0120 (0.0119-0.0121)	NLS
$\bar{\mu}$	Final disease-induced death rate	0.8147 (0.7011-0.9283)× μ_{ini}	NLS
μ_0	Decrease rate in the disease-induced death rate	0.0100 (0.0089-0.0111)	NLS
t_μ	The time when the death rate is half of the initial and final death rate	60.0008 (54.3184-65.6832)	NLS

Table S2. The values of parameters based on references or estimated by nonlinear least squares (NLS) methods outside Hubei. The other parameters are the same as in Hubei Province as shown in **Table S1**.

Parameter	Description	Range or 95% CI from NLS	Source
E(0)	The initial value of latent individuals	899.9925 (843.1606-956.8244)	NLS
β	The per-act transmission probability in contact with infected individuals with symptoms	0.0520 (0.0599-0.0531)	NLS
$m_1(t)$	The average number of daily contacts in the public space	$m_{ini}+\frac{0.2m_{ini}-m_{ini}}{1+\exp(-m_0(t-t_{ini}))}$	[13]
m_{ini}	Base daily contact number in the public space	9.9000 (9.7862-10.0138)	NLS
m_0	Decrease rate of daily contact number in the public space	0.6030 (0.5461-0.6599)	NLS
t_{ini}	Time when behaviour change begins (3 days after quarantine)	6.2000 (6.0861-6.3138)	NLS
$m_2(t)$	The average number of daily contacts in the households	$4+\frac{12-4}{1+\exp(-m_0(t-t_{ini}))}$	[5]
$p_1(t)$	The usage percentage of facial mask in the public space	$p_{ini}+\frac{\bar{p}-p_{ini}}{1+\exp(-m_0(t-t_{ini}))}$	[7][8]
$\alpha(t)$	The diagnosis rate	$\alpha_{ini}+\frac{\bar{\alpha}-\alpha_{ini}}{1+\exp(-\alpha_0(t-t_{\alpha}))}$	[16]
α_0	Increase rate of diagnosis rate	0.6000 (0.4862-0.7138)	NLS
t_{α}	The time when the diagnosis rate is half of the maximum diagnosis rate	6.2001 (6.0863-6.3139)	NLS
μ(t)	Disease-induced death rate	0.0015 (0.0013-0.0017)	NLS

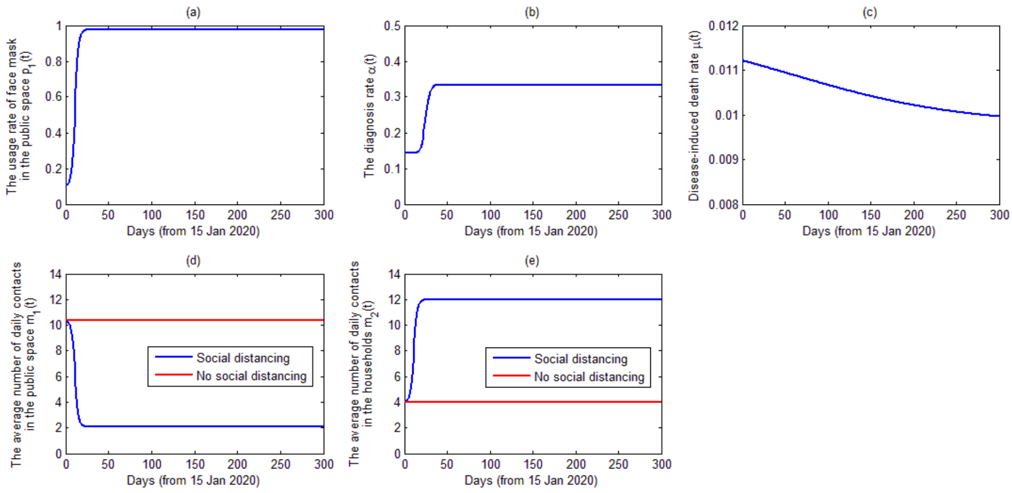

Figure S2. Time-dependent parameters in Hubei Province. (a) The percentage of facial mask use $p_1(t) = p_{ini} + \frac{\bar{p}-p_{ini}}{1+\exp(-0.5(t-t_{ini}))} = 10\% + \frac{97.6\%-10\%}{1+\exp(-0.5(t-11))}$ over time. (b) The diagnosed rate $\alpha(t) = \alpha_{ini} + \frac{\bar{\alpha}-\alpha_{ini}}{1+\exp(-\alpha_0(t-t_{ini}))} = \frac{1}{7} + \frac{\frac{1}{3}-\frac{1}{7}}{1+\exp(0.5(t-11))}$ over time. (c) The disease-induced death rate $\mu(t) = \mu_{ini} + \frac{\bar{\mu}-\mu_{ini}}{1+\exp\left(-\mu_0(t-t_\mu)\right)} = 0.0120 + \frac{0.8147\times0.0120-0.0120}{1+\exp\left(-0.0101(t-60.0008)\right)}$ over time. (d) The average number of daily contacts in the public space $m_1(t) = m_{ini} + \frac{0.2m_{ini}-m_{ini}}{1+\exp(-0.5(t-t_{ini}))} = 10.3676 + \frac{0.2\times10.3676-10.3676}{1+\exp(-0.5(t-11))}$ over time with (blue line) and without (red line) social distancing. (e) The average number of daily contacts in the public space $m_2(t) = 4 + \frac{12-4}{1+\exp(-0.5(t-11))}$ over time with (blue line) and without (red line) social distancing.

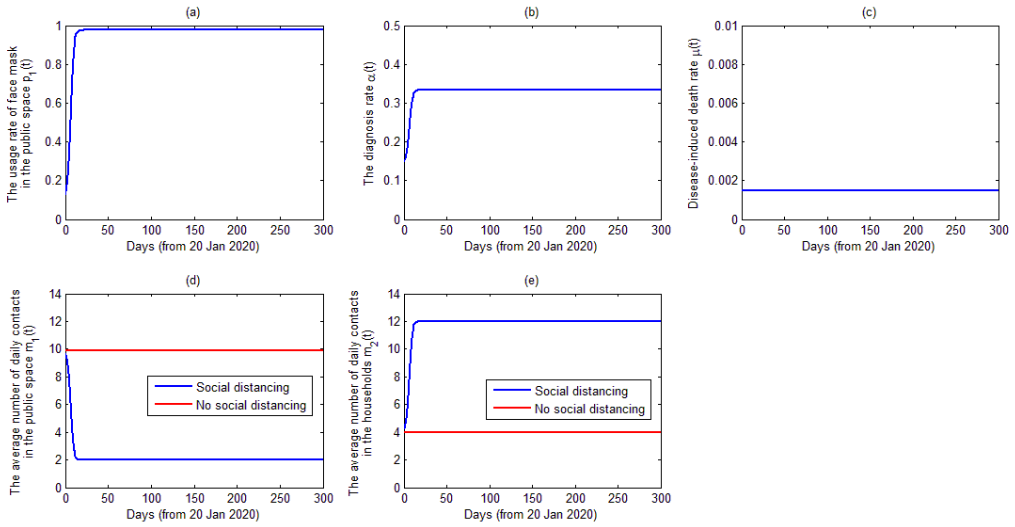

Figure S3. Time-dependent parameters outside Hubei. (a) The percentage of facial mask use $p_1(t) = p_{ini} + \frac{\bar{p}-p_{ini}}{1+\exp(-m_0(t-t_{ini}))} = 10\% + \frac{97.6\%-10\%}{1+\exp(-0.6030(t-6.2000))}$ over time. (b) The diagnosed rate $\alpha(t) = \alpha_{ini} + \frac{\bar{\alpha}-\alpha_{ini}}{1+\exp(-\alpha_0(t-t_\alpha))} = \frac{1}{7} + \frac{\frac{1}{3}-\frac{1}{7}}{1+\exp(0.6000(t-6.2001))}$ over time. (c) The disease-induced death rate $\mu(t) = 0.0015$ over time. (d) The average number of daily contacts in the public space $m_1(t) = m_{ini} + \frac{0.2m_{ini}-m_{ini}}{1+\exp(-m_0(t-t_{ini}))} = 9.9000 + \frac{0.2\times 9.9000-9.9000}{1+\exp(-0.6030(t-6.2000))}$ over time with (blue line) and without (red line) social distancing. (e) The average number of daily contacts in the public space $m_2(t) = 4 + \frac{12-4}{1+\exp(-0.6030(t-6.2000))}$ over time with (blue line) and without (red line) social distancing.

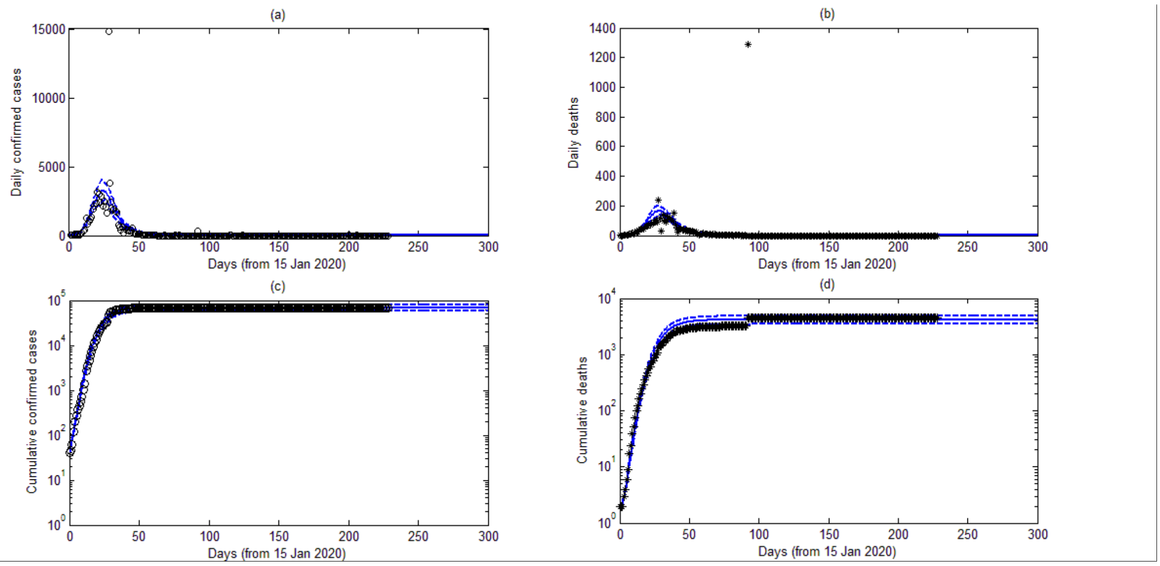

Figure S4. Model calibration by the number of daily and cumulative confirmed cases and deaths in Hubei Province. Dashed lines denote 95% confidence intervals.

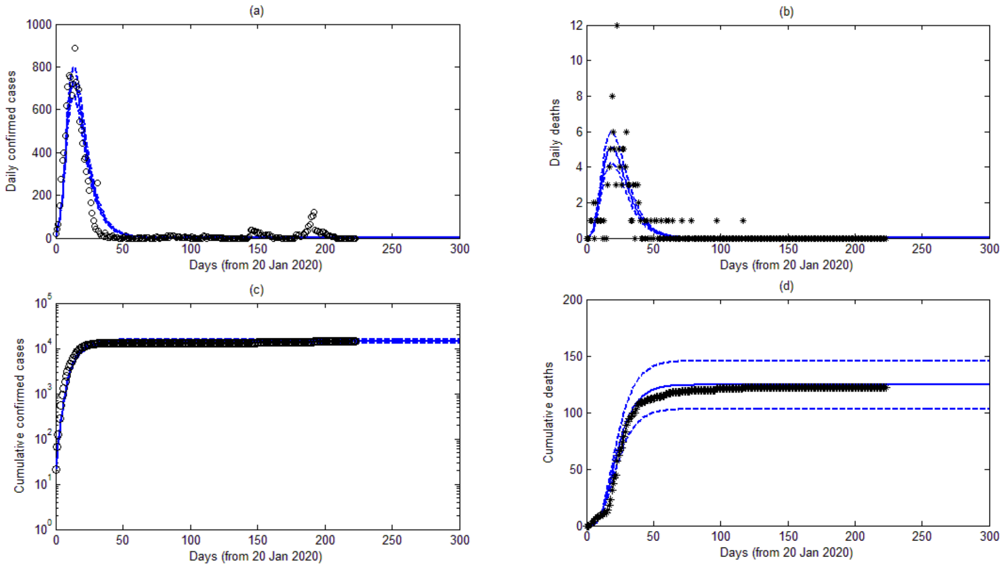

Figure S5. Model calibration by the number of daily and cumulative confirmed cases and deaths outside Hubei. Dashed lines denote 95% confidence intervals.

References

[1]. National Health Commission of China. Daily report of COVID-19. http://www.nhc.gov.cn/xcs/xxgzbd/gzbd_index.shtml [Accessed 31 Aug 2020]

[2]. Health Commission of Hubei Province. Daily report of COVID-19. http://wjw.hubei.gov.cn/ [Accessed 31 Aug 2020]

[3]. Li Q, Guan XH, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. 2020. N Engl J Med. DOI:10.1056/NEJMoa2001316

[4]. Wei Y, Lu Z, Du Z, et al. Fitting and forecasting the trend of COVID-19 by SEIR+CAQ dynamic model. Chin J Epidemiol. 2020; 41(4):470-475.

[5]. Chen B, Ai C, Ma L, et al. We should prepare for the transmission of COVID-19 in the long run: simulating the risk of the secondary outbreak after holiday. Journal of System Simulation. 2020; http://www.china-simulation.com/CN/Y2020/V0/I0/0

[6]. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf [Accessed 31 Aug 2020]

[7]. Kejixun. Liebao Mobile robot data: The facial mask useage rate reached 97.6% in Wuhan city during the epidemic. http://m.kejixun.com/article/200224/488396.shtml [Accessed 31 Aug 2020]

[8]. Huailiang Wu, Jian Huang, Casper J. P. Zhang, et al. Facemask shortage and the novel coronavirus (2019-nCoV) outbreak: Reflection on public health measures. https://doi.org/10.1101/2020.02.11.20020735 [Accessed 31 Aug 2020]

[9]. Mao M, Wang M, Zhang S. How many people wear facial masks in Beijing? A street mask observation practice was conducted here. https://www.thepaper.cn/newsDetail_forward_1592513 [Accessed 31 Aug 2020]

[10]. Gu J, Li J, Hao Y, et al. Bulletin of COVID-19 online survey. http://sph.sysu.edu.cn/article/1257 [Accessed 31 Aug 2020]

[11]. CCTV. How to protect yourself when returning? Whether it is necessary to wear facial masks at home? http://jiankang.cctv.com/2020/01/31/ARTIBRA7riu1PKMGS9BaqPkD200131.shtml [Accessed 31 Aug 2020]

[12]. Chu DK, Akl EA, Duda S, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet. 2020, 395(10242):1973-1987.

[13]. Zhang J, Litvinova M, Liang Y, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science. 368, 6498:1481-1486.

[14]. Huang CL, Wang YM, Li XW, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020; DOI:https://doi.org/10.1016/S0140-6736(20)30183-5

[15]. Yunpeng Ji, Zhongren Ma, Maikel P Peppelenbosch, et al. Potential association between COVID-19 mortality and healthcare resource availability. Lancet Glob Health 2020; https://doi.org/10.1016/S2214-109X(20)30068-1

[16]. Shen M, Peng Z, Guo Y, et al. Assessing the effects of metropolitan-wide quarantine on the spread of COVID-19 in public space and households. Int J Infect Dis. 2020, 96:503-505.