medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Estimating the daily trend in the size of COVID-19 infected population in Wuhan

Qiushi Lin[a], Taojun Hu[b], Xiao – Hua Zhou[a,c,d*]

a Beijing International Center for Mathematical Research, Peking University, Beijing, China, 100871
b School of Mathematical Sciences, Peking University, Beijing, China, 100871
c Department of Biostatistics, School of Public Health, Peking University, Beijing, China, 100871
d Center for Statistical Science, Peking University, Beijing, China, 100871

* Corresponding author: azhou@math.pku.edu.cn

Abstract

Since December 2019, an outbreak of pneumonia cause by a novel coronavirus (COVID-19) has occurred in Wuhan, Hubei Province, China. Cases have been exported to other places in China, as well as over 20 countries around the world. We provide estimates of the daily trend in the size of the epidemic in Wuhan based on detailed information of 10,940 confirmed cases outside Hubei Province.

Background

As of February 9, 2020, the National Health Commission (NHC) of China had confirmed a total of 40,171 cases of novel coronavirus pneumonia (NCP) in mainland China, including 6,484 severe cases, 908 deaths, and 32,81 recoveries. An additional total of 23,589 suspected cases were reported. Wuhan, the epicenter of the NCP outbreak, had confirmed 16,902 cases. The NHC had received confirmed reports including a total of 36 in Hong Kong Special Administrative Region, China, 10 in Macau Special Administrative Region, China, and 18 in Taiwan, China. [1] More than 300 cases had been detected outside China.

Although huge amount of medical resources and personnel has already been supporting Hubei Province, the admission capacity of hospitals is still not enough for all the NCP patients. Wuhan will transform hotels, venues, training centers and college dorms into places for quarantine and treatment of NCP patients, and 13 mobile cabin hospitals will be built to provide over 10,000 beds. [2] Therefore, a careful and precise understanding of the potential number of cases in Wuhan is crucial for the prevention and control of NCP outbreak. You et al. (2020) proposed a model to estimate the total number of cases of NCP in Wuhan. [3] However, their model can only give an estimate of the cumulative number of cases until a certain date. In this article, a more elaborate method for simultaneously estimating the number of cases in Wuhan every day is established using the individual information of confirmed cases outside Hubei Province.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Results

We estimate the number of cases that should be reported in Wuhan by January 11, 2020, is 4,090 (95% confidence interval [CI]: 3,975 – 4,206) and 56,833 (95% CI: 55,242 – 58,449) by February 9, 2020. Figure 1 shows how the estimated number of cases in Wuhan increased over time, together with the 95% confidence bands. As shown in Figure 2, the reporting rate had grown rapidly from 1.41% (95% CI: 1.37% - 1.45%) on January 11, 2020, to 29.74% (95% CI: 28.92% - 30.60%) on February 9, 2020. The date of first infection is estimated as November 30, 2020.

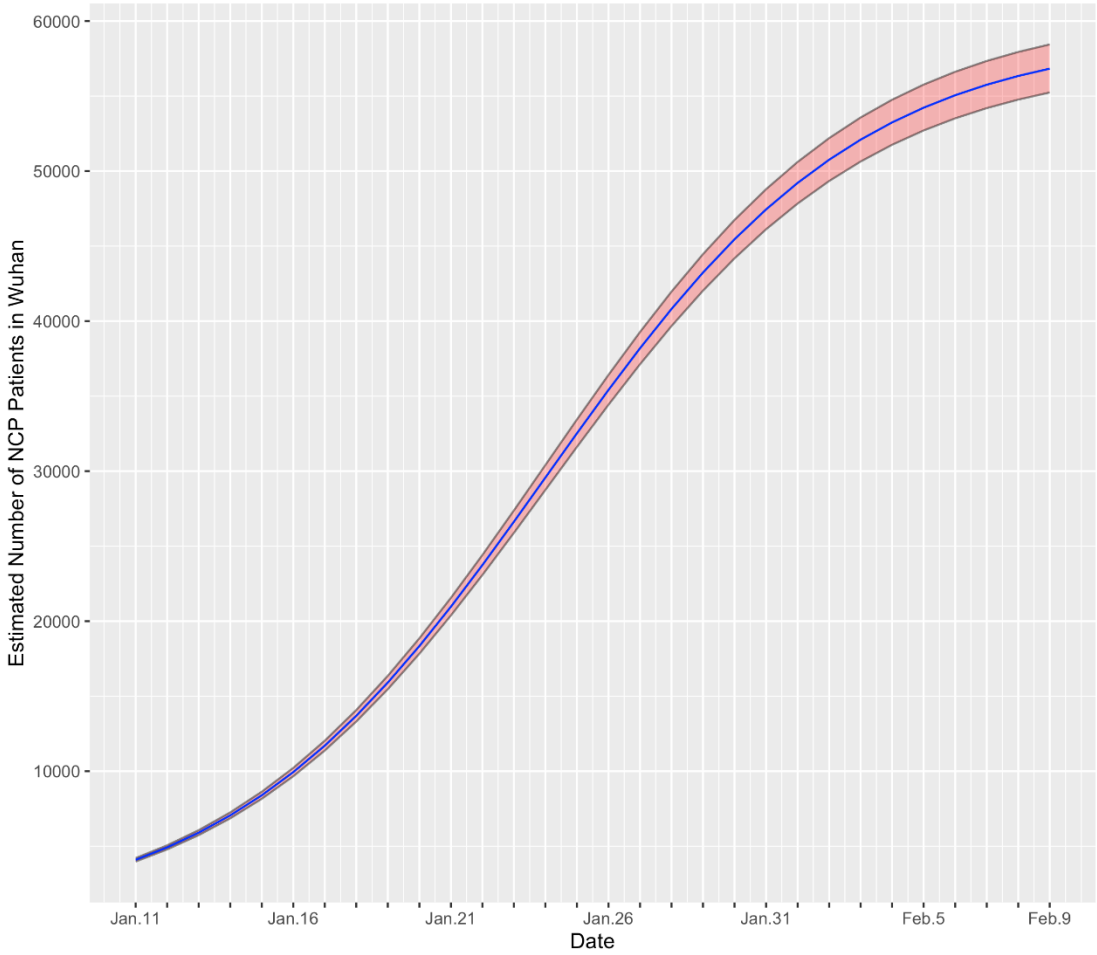

Figure 1. Estimate number of total cases in Wuhan.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

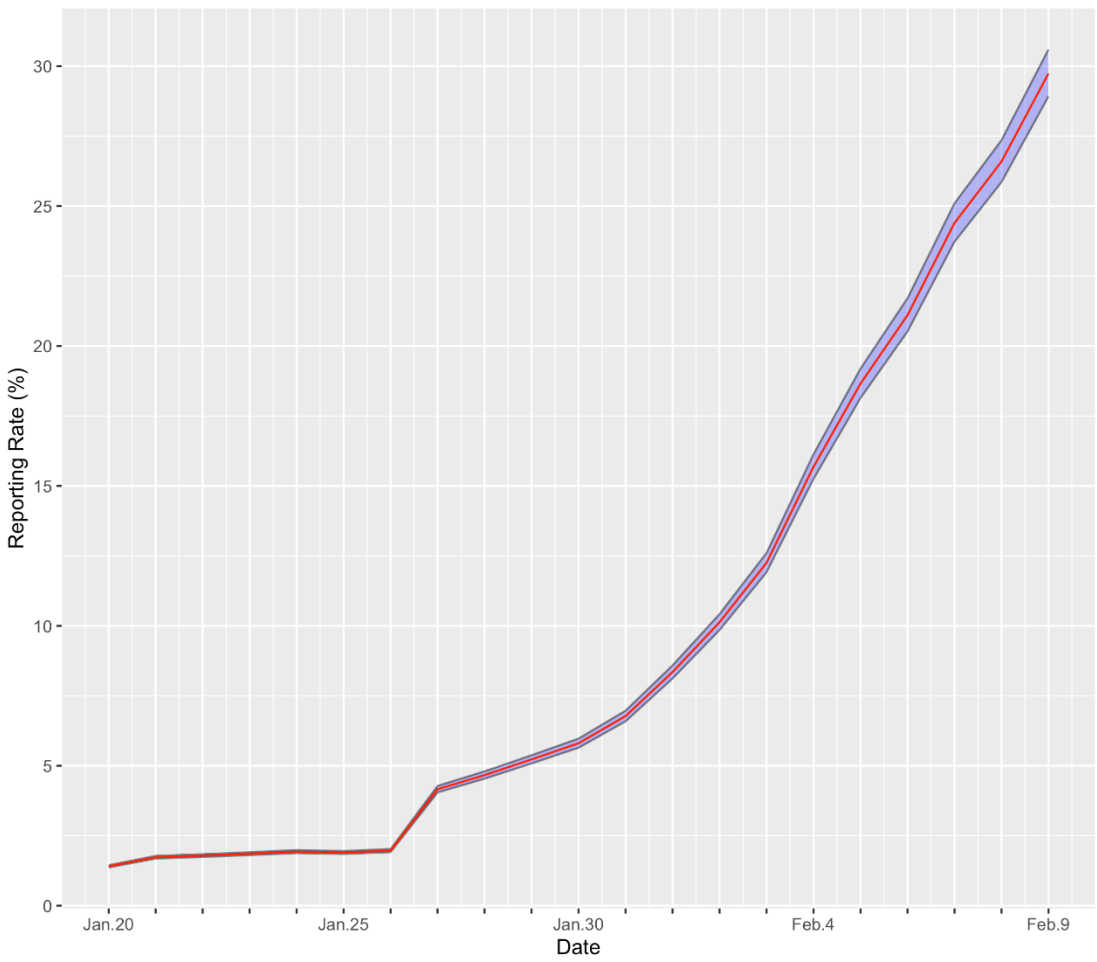

Figure 2. The ratio of reported number of cases to the estimated number.

Data Description

Data were retrieved from provincial and municipal health commissions in China as well as through ministries of health in other countries with detailed information of 10,940 confirmed cases outside Hubei Province, including region, gender, age, date of symptom onset, date of confirmation, history of travelling or residing in Wuhan, and date of leaving Wuhan. Among 7,499 patients with gender information, 3,508(46.78%) are female. The mean age of patients is 44.48, and the median age is 44. The youngest patient outside Hubei Province was confirmed only five days after birth, while the oldest is 97 years old.

Table 1. Demographical Characteristics of Patients with NCP outside Hubei Province.

Age group - years	Female (N=3,508)	Male (N=3,991)	No information (N=3,441)
0-20	97(3)*	149(4)	2
20-39	1,076(33)	1,348(36)	41

medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

40-59	1,424(43)	1,598(43)	38
60-79	630(19)	578(15)	39
⩾80	66(2)	60(2)	7
No information	214	257	3,279

* Number (%). The percentages do not take missing data into account.

We display the epidemiological data categorized by the date of confirmation in Table 2. Imported case means a patient that had been to Wuhan and detected outside Hubei on Day k. Local case means a confirmed case that had not been to Wuhan. Among the total of 10,940 cases, 6,828(62.41%) have such information. The number of imported cases reached its peak on January 29, 2020, and the fourth column of Table 2 shows that the proportion of imported cases declines over time. This might reflect the effect of containment measures taken in Hubei Province to control the NCP outbreak. [4] Meanwhile, the daily counts of local cases are over 300 from February 2, 2020, to February 7, 2020, which indicate that infections among local residents should be a major concern for authorities outside Hubei.

Table 2. Patient data categorized by the date of confirmation.

Day k	Imported Cases I_k (N=2,969)	Local Cases L_k (N=3,859)	Percentage $I_k/(I_k+L_k)$	Total Cases T_k (N=10,940)	Onset to detection (days)
⩽20-Jan*	23	1	96%	26	8.83
21-Jan	30	3	91%	45	8.71
22-Jan	51	6	89%	77	6.11
23-Jan	99	15	87%	184	5.37
24-Jan	138	15	90%	275	4.56
25-Jan	194	36	84%	358	4.15
26-Jan	198	49	80%	410	4.05
27-Jan	190	85	69%	483	5.00
28-Jan	283	104	73%	639	5.58
29-Jan	299	144	67%	723	5.43
30-Jan	291	214	58%	765	5.13
31-Jan	256	239	52%	803	5.05
1-Feb	160	252	39%	678	4.83

medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

2-Feb	159	310	34%	738	5.55
3-Feb	164	410	29%	896	5.76
4-Feb	119	336	26%	749	5.99
5-Feb	107	376	22%	749	6.10
6-Feb	83	380	18%	721	5.92
7-Feb	53	363	13%	617	6.17
8-Feb	42	286	13%	530	6.49
9-Feb	30	235	11%	474	6.66

* The count and average on the first row are taken over all cases confirmed by January 20, 2020.

The last column of Table 2 lists the mean time from symptom onset to confirmation for patients confirmed on Day k. The median duration of all cases is 5 days, and the mean is 5.54 days. In general, the detection period decreased in the first week after January 20, 2020, but increased since then. The improvements in detection speed and capacity might cause the initial decline, and the rise may be due to more thorough screening, leading to the detection of patients with mild symptoms who would otherwise not go to the hospitals. [5]

Assumptions

The proposed method relies on the following assumptions:

1. The daily probability of departing from Wuhan is a constant p over time.
2. There is a $d = d_1 + d_2$-day window between infection to detection, including a d_1-day incubation period and a d_2-day delay from symptom onset to detection.
3. Trip durations are long enough that a traveling patient infected in Wuhan will develop symptoms and be detected in other places rather than after returning to Wuhan.
4. All travelers leaving Wuhan, including transfer passengers, have the same risk of infection as local residents.
5. Traveling is independent of the exposure risk to COVID-19 or of infection status.
6. Patients are not able to travel d days after infection.
7. Recoveries are not considered in this method.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Some of the assumptions might not be realistic, but the data needed to account for them are not currently available. The following caveats are further noted:

a. Violation of Assumption 1 (e.g., the probability of leaving Wuhan is larger than p) would cause an overestimation of the number of cases in Wuhan.

b. Violation of Assumption 2 (e.g., the average time from infection to detection is longer than d days) would lead to an overestimation.

c. Violation of Assumption 4 (e.g., travelers have a lower risk of infection than residents in Wuhan) would cause an underestimation.

d. Violation of Assumption 5 (e.g., infected individuals are less likely to travel due to the health conditions) would cause an underestimation.

e. Given that the number of recoveries is relatively small compared to the size of COVID-19 infected population, Assumption 8 should not significantly influence the result.

Some of the violations are practiced in the Sensitivity Analysis.

Methods

Assume Day 1 is the date of infection for the very first case. Let N_t be the cumulative number of cases that should be confirmed in Wuhan by Day t, then the number of infections is N_{t-d} under Assumption 2. If $t \leq d$, there should be no confirmed cases. If $d < t \leq 2d$, then the number of imported cases x_t on Day t follows a binomial$(N_t,\ p)$ distribution. If $t > 2d$, under Assumption 6, x_t follows a binomial$(N_t - N_{t-d},\ p)$ distribution. Let X_t be the cumulative number of imported cases by Day t, then

$$X_t = \sum_{k=1}^{t} x_k \sim \text{Binomial}\left(\sum_{k=t-d+1}^{t} N_k, p\right), \quad t \geq 2d. \qquad (1)$$

The growth trend of the size N_t of infected population is determined by the following ordinary differential equation

medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

$$\frac{dN_t}{dt} = \frac{r}{K} N_t (K - N_t), \qquad r > 0, K > 0. \qquad (2)$$

This is the simplified version of the famous SIR model [6] in epidemiology, which does not take recoveries into account, such that equation (2) has an analytical solution

$$N_t = \frac{K}{1 + e^{-r(t-t_0)}} = K f_t, \qquad (3)$$

where $f_t = \frac{1}{1+e^{-r(t-t_0)}}$. Therefore, $X_t \sim \text{Binomial}(K \sum_{k=t-d+1}^{t} f_k, p)$. The parameter estimate $\widehat{K}$ is the maximizer of the likelihood function

$$l(K) = \binom{K \sum_{k=t-d+1}^{t} f_k}{X_t} p^{X_t} (1-p)^{K \sum_{k=t-d+1}^{t} f_k - X_t}. \qquad (4)$$

The lower and upper bound of the 95% confidence interval $[\widehat{K_l}, \widehat{K_u}]$ are values such that the cumulative distribution function $F(K) = \sum_{x=0}^{X_t} l(K)$ equals 0.975 and 0.025, respectively.

Determining the number of imported cases x_k plays a crucial role in the modeling procedure. Note that not all cases have clear records on the history of traveling or residing in Wuhan, we need to impute the missing values. Assume the proportion of imported cases in the patients with no information is the same as the observed proportion, then

$$x_k = T_k \times \frac{I_k}{I_k + L_k}, \qquad (5)$$

where the values of T_k, I_k, and L_k are listed in Table 2.

The daily probability of leaving Wuhan is estimated as the ratio of daily volume of travelers to the population of Wuhan City, 14 million. [10] Wuhan Municipal Transportation Management Bureau expected to send 15 million passengers during the 40-day Chinese New Year travel rush, 34% of whom would travel across 300 km. [7] This would imply, on average, the daily probability p of traveling from Wuhan to places outside Hubei would be 15*0.34/40/14=0.009. Li et al. estimated that the mean incubation period of 425 patients with NCP was 5.2 days (95% CI, 4.1 to 7.0). [8] The mean time from symptom onset to detection calculated from our data is 5.5 days, so we choose $d_1 = d_2 = 5$ days. January 29, 2020, has the maximum observed count of imported cases, so $N_t - N_{t-d}$ is maximized at $t =$ January

medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

29, 2020. It is clear from the analytical form (3) that $t_0 = t - \frac{d}{2} =$ January 24, 2020, which is shortly after the lockdown of Wuhan City. [4] Wu et al. estimated the epidemic doubling time as 6.4 days (95% CI 5.8–7.1) as of January 25, 2020. From this result, we estimate that $\frac{r}{2} = \frac{d \ln N_{t_0}}{dt} = \frac{\ln 2}{6.4} = 0.1$. Using these values for parameters $p, d, t_0,$ and $r,$ we can derive the maximum likelihood estimate $\widehat{K} = 59150,$ with 95% confidence interval [57494, 60831]. The estimate of the date of first infection is obtained by solving the equation $N_t = 1$.

Sensitivity Analysis

We explore the sensitivity of the estimate of total cases in to our assumptions and choices of parameters $p, d,$ and r. Note that $t_0 + \frac{d}{2} =$ January 29, 2020. Table 3 summaries the estimate the number of cases should be reported on January 11, 2020, and February 9, 2020 under baseline assumptions and alternative scenarios. Confidence intervals are omitted.

Table 3. Estimated case numbers on January 11and February 9, 2020, based on different choice of parameters.

	Baseline	Scenario 1	Scenario 2	Scenario 3
p	0.009	0.01	0.009	0.009
$\mathrm{d} = d_1 + d_2$	10	10	11	10
r	0.2	0.2	0.2	0.3
11-Jan	4,090	3,681	4,014	1,100
9-Feb	56,833	51,150	51,370	54,968

The currently reported number 16,902 on February 9, 2020, is substantially smaller than the lower bound of our most conservative scenario.

Conclusions

Measures taken by authorities to boost the admission capacity of hospitals in Wuhan are effective. The estimated reporting rate increases rapidly over the past few days. However, the reporting rate only reaches about 30% as of February 9, 2020, indicating that more thorough screening of all fever patients with a mild or moderate respiratory disease should be

medRxiv preprint doi: https://doi.org/10.1101/2020.02.12.20022277; this version posted February 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

conducted. More hospital beds and medical facilities should be built to accommodate and quarantine all the NCP patients and hence stop the spread of the virus in Wuhan.

References

[1] National Health Commission Update on February 10, 2020. National Health Commission of the People's Republic of China. http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm#NHCFeb10 [2020-2-10]

[2] Hubei ordered to admit all patients in hospitals. China Daily. https://www.chinadaily.com.cn/a/202002/09/WS5e3fba1ca3101282172760aa.html [2020-2-9]

[3] Chong You, Qiushi Lin, Xiao-hua Zhou. An Estimation of the Total Number of Cases of NCIP (2019-nCoV) — Wuhan, Hubei Province, 2019–2020[J]. China CDC Weekly, 2020, 2(6): 87-91.

[4] China declares lockdown in Wuhan on Thursday due to coronavirus outbreak. Tass. https://tass.com/world/1111981 [2020-1-23]

[5] Beijing to set up checkpoints in all residential communities. China Daily. https://www.chinadaily.com.cn/a/202002/10/WS5e415cb1a3101282172766c4.html [2020-2-1]

[6] Kermack WO, McKendrick AG. A Contribution to the Mathematical Theory of Epidemics". Proceedings of the Royal Society A, 1927, 115 (772): 700–721.

[7] Big data perspective: Wuhan in the Chinese New Year travel rush. Daily Economic News. https://m.nbd.com.cn/articles/2020-01-22/1402239.html [2020-1-22] (In Chinese).

[8] Qun Li, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med. 2020 January 29.

[9] Wu, Jianhong & Leung, Kathy & Leung, Gabriel. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet. 10.1016/S0140-6736(20)30260-9.

[10] 5 million-plus leave Wuhan: Mayor. China Daily. https://www.chinadaily.com.cn/a/202001/27/WS5e2dcd01a310128217273551.html [2020-1-27]