Data Fixing by Data Fitting: Estimating the Unreported Cases During the Early COVID -19 Outbreak in Hubei, China

Kamlesh Sarkar Xiang-Sheng Wang

Department of Mathematics,University of Louisiana at Lafayette

Objectives

- Address the sudden spike in reported cases due to the change in case definition on February 13, 2020.
- Derive a formula from the Kermack-McKendrick model to estimate and correct unreported cases.
- Use the new formula to fit inconsistent COVID-19 data and improve parameter estimation.
- Quantify hidden cases in Hubei (60,856)(95% CI: [33513, 91206]) and Wuhan (29,374) (95% CI: [18205, 40665]).
- Identify peak infection dates—February 6, 2020, for Hubei and February 8, 2020, for Wuhan.
- Estimate key parameters, including the basic reproduction number (R_0) for Hubei (2.334) (95% CI: [2.053, 2.711]) and Wuhan (2.189) (95%CI: [1.992, 2.448]) .

Methods

We adopt an epidemic model proposed by Kermack and McKendric [2]:

$$\begin{aligned} S'(t) &= -\beta S(t)I(t)/[S(t)+I(t)], \\ I'(t) &= \beta S(t)I(t)/[S(t)+I(t)] - \gamma I(t), \\ R'(t) &= \gamma I(t), \end{aligned} \tag{1}$$

where β is the disease transmission rate, and γ is the removal rate of infected individual. To simplify our analysis, we make the following assumptions:

(A1) Among all the cumulative infected cases at the end of the outbreak, $S(t)$ represents susceptible individual at time t.

(A2) The infected individuals, denoted by $I(t)$, are infectious even during the incubation period with no symptoms.

(A3) The removed individuals $R(t)$ include all infected individuals who are no longer infectious due to recovery, death, or quarantine.

The solution to model (1) has a closed from, and is given by

$$\begin{aligned} S(t) &= N\left[1+(R_0-1)e^{(\beta-\gamma)(t-t_p)}\right]^{-R_0/(R_0-1)}, \\ I(t) &= N(R_0-1)e^{(\beta-\gamma)(t-t_p)}\left[1+(R_0-1)e^{(\beta-\gamma)(t-t_p)}\right]^{-R_0/(R_0-1)}, \\ R(t) &= N - N\left[1+(R_0-1)e^{(\beta-\gamma)(t-t_p)}\right]^{-1/(R_0-1)}, \end{aligned} \tag{2}$$

where N is the final outbreak size, t_p is the peak time when $I(t)$ reaches its maximum, and $R_0 = \frac{\beta}{\gamma}$ is the basic reproduction number.

we introduce a new parameter H to capture the hiding information from the data. To sum up, we will fit the reported cumulative case numbers by the following piecewise defined function.

$$C(t) = \begin{cases} K_1 - K_1\left[1+(R_0-1)e^{(\beta-\gamma)(t-t_p)}\right]^{-1/(R_0-1)}, & t < t_c, \\ K_2 - K_2\left[1+(R_0-1)e^{(\beta-\gamma)(t-t_p)}\right]^{-1/(R_0-1)} - H, & t > t_c, \end{cases} \tag{3}$$

where $K_1 = p_1 N$, $K_2 = p_2 N$, and H is the hiding case number from the reported data. Note that the final reported cumulative case number is $C(\infty) = K_2 - H$.

Data

We use the reported cumulative case number from the websites of the Health Commission of Hubei Province [1] and the National Health Commission of the Peoples Republic of China [4]. Whenever there is a correction on a reported date, we update the corresponding value in the previous reported date. After correction, the reported cumulative case numbers of Hubei province from January 21, 2020 to February 21, 2020 are

270	375	444	549	729	1052	1423	2714
3554	4586	5806	7153	9074	11177	13522	16678
19665	22112	24953	27013	29631	31728	33366	47163
51986	54406	56249	58182	59989	61682	62457	63088

Table 1. Cumulative COVID-19 case numbers in Hubei Province.

Notice that the number jumps abruptly from 33366 to 47163.

Also, the Wuhan cumulative case numbers from January 17, 2020 to February 21, 2020 are

45	62	121	198	258	363	425	495	572
618	698	1590	1905	2261	2639	3215	4109	5142
6384	8351	10117	11618	13603	14981	16902	18454	19558
32081	35991	37914	39462	41152	42752	44412	45027	45346

Table 2. Cumulative COVID-19 case numbers in Wuhan City.

Parameter Estimation

Report date	K_1	K_2	H	$C(\infty)$	R_0	β	γ	Peak time t_p
02/15/2020	41236	136679	67791	68888	2.394	0.528	0.22	02/06/2020
02/16/2020	38786	142666	79145	63521	1.997	0.594	0.297	02/06/2020
02/17/2020	38009	155787	93240	62547	1.869	0.629	0.337	02/06/2020
02/18/2020	39017	141742	77734	64008	2.036	0.585	0.287	02/06/2020
02/19/2020	40978	123531	57045	66486	2.356	0.533	0.226	02/06/2020
02/20/2020	41076	122818	56217	66601	2.372	0.531	0.224	02/06/2020
02/21/2020	40811	124939	58630	66309	2.328	0.536	0.230	02/06/2020

Table 3. The parameter estimation for Hubei cumulative data with different last report dates.

Report date	K_1	K_2	H	$C(\infty)$	R_0	β	γ	Peak time t_p
02/15/2020	21652	140829	100268	40561	1.254	1.303	1.039	02/08/2020
02/16/2020	21603	145774	105119	40655	1.253	1.310	1.045	02/08/2020
02/17/2020	22813	107525	64673	42852	1.503	0.813	0.541	02/08/2020
02/18/2020	24464	86501	41087	45414	1.830	0.617	0.337	02/08/2020
02/19/2020	26842	74109	25431	48678	2.288	0.513	0.224	02/08/2020
02/20/2020	27043	73453	24519	48934	2.326	0.507	0.218	02/08/2020
02/21/2020	26432	75872	27684	48188	2.207	0.525	0.238	02/08/2020

Table 4. The parameter estimation for Wuhan cumulative data with different last report dates.

Data Fitting

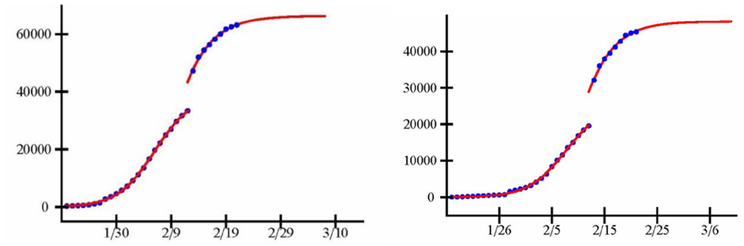

Figure 1. Reported data (dots) and fitted model (curve) for the cumulative cases of COVID-19 outbreak in Hubei province (left panel) and Wuhan city (right panel), respectively.

Conclusion and Discussion

- A simple and biologically meaningful formula defined by equation (3) was proposed to estimate key epidemic parameters.
- This formula not only fits the inconsistent data well but also corrects it by capturing the unreported case numbers hidden within the data through data fitting.
- The estimated basic reproduction number matches prior findings [3] despite using a different method and data type.
- The epidemic peak occurred at the beginning of February 2020, and the first epidemic wave would approach its end in March 2020.
- For more detail of this work, can be found at [5].

Acknowledgement

I would like to thank **Xiang-Sheng Wang** for his support and guidance throughout my research. I also extend my sincere thanks to the organizers of the MATH for All Conference at NOLA for giving me the opportunity to present my work.

References

[1] Health Commission of Hubei Province (2020). News. `http://www.hubei.gov.cn/fbjd/dtyw/`. [Online; accessed 21-February-2020].

[2] Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721.

[3] 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. (2020). Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England journal of medicine*, 382(13):1199–1207.

[4] National Health Commission of the People's Republic of China (2020). News. `http://www.nhc.gov.cn/yjb/s7860/new_list.shtml`. [Online; accessed 21-February-2020].

[5] Sarkar, K. and Wang, X. (2024). Data fixing by data fitting: Estimating the unreported cases during the early covid-19 outbreak in hubei, china. *Journal of Basic & Applied Sciences*, 20:92–97.