Quantitative Biology 2021, 9(1): 93–99
https://doi.org/10.15302/J-QB-020-0232

LETTER

Forecasting the development of the COVID-19 epidemic by nowcasting: when did things start to get better?

Yuehui Zhang[1], Lin Chen[2], Qili Shi[3], Zhongguang Luo[2], Libing Shen[4]

[1] Spine Center, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China
[2] Department of Digestive Diseases of Huashan Hospital, Fudan University, Shanghai 200040, China
[3] Stem Cell and Regenerative Medicine Laboratory, Ningbo No. 2 Hospital, Ningbo 315010, China
[4] Institute of Neuroscience, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
* Correspondence: luozg8@126.com, shenlibing@ion.ac.cn

Received June 16, 2020; Revised August 8, 2020; Accepted August 31, 2020

***Background*: Now the coronavirus disease 2019 (COVID-19) epidemic becomes a global phenomenon and its development concerns billions of peoples' lives. The development of the COVID-19 epidemic in China could be used as a reference for the other countries' control strategy.**
***Methods*: We used a classical susceptible-infected-recovered (SIR) model to forecast the development of the COVID-19 epidemic in China by nowcasting. The linear regression analyses were employed to predict the COVID-19 epidemic's inflexion point. Finally, we used a susceptible-exposed-infected-recovered (SEIR) model to simulate the development of the COVID-19 epidemic in China throughout 2020.**
***Results*: Our nowcasts show that the COVID-19 transmission rate started to slow down on January 30. The linear regression analyses further show that the inflexion point of this epidemic would arrive between February 17 and 18. The final SEIR model simulation forecasted that the COVID-19 epidemic would probably infect about 82,000 people and last throughout 2020 in China. We also applied our method to USA's and global COVID-19 data and the nowcasts show that the development of COVID-19 pandemic is not optimistic in the rest of 2020.**
***Conclusion*: The COVID-19 epidemic's scale in China is much smaller than the previous estimations. After implemented strict control and prevention measures, such as city lockdown, it took a week to slow down the COVID-19 transmission and about four weeks to really mitigate the COVID-19 prevalence in China.**

Keywords: forecast; COVID-19 epidemic; development

Author summary: Now the coronavirus disease 2019 (COVID-19) is a global pandemic and the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a threat to the health of billions of peoples. China imposed a strict prevention and control measure to fight it. This study shows that the effectiveness of China's measure emerged after one week of its implementation and it took about four weeks to really mitigate the COVID-19 transmission in China. Our results might have a certain reference value for the policy makers all over the world.

INTRODUCTION

The coronavirus disease 2019 (COVID-19) is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This virus is clustered together with the SARS-CoVs in bat and regarded as a SARS-like virus [1]. The disease so far claims at least 3000 lives in China. Japan, Thailand, and U.S. have already reported the COVID-19 infection cases [2]. Mysteriously, the SARS-CoV-2 antibodies were detected among the individuals enrolled in a prospective lung cancer screening trail in Italy as early as Sep 2019 [3], suggesting a possible multiple geographical onset of COVID-19 spread even before its identification in China [4]. In order to contain its devastating power, Chinese government has mobilized huge man power and poured vast resource to combat this virus. Unfortunately, the COVID-19's shadow has already overcast more than 100 countries. The development of this epidemic may not be on everyone's tongue, but is always on his mind. Therefore, it is almost compelling to develop a method to predict where this epidemic is going in real time.

Here we present a mathematic strategy to forecast the development of the COVID-19 epidemic in China by nowcasting. First, we provided the COVID-19 epidemic nowcasts for the next day based on a susceptible-infected-recovered (SIR) model. By doing so, we concluded that the COVID-19 transmission rate in China started to slow down after Jan 30. Second, after the infection nowcasts were no longer accurate, we used linear regression to forecast the transmission inflexion point of the COVID-19 epidemic, which shows that the inflexion point of this epidemic would arrive between Feb 17 and 18. Finally, we used the susceptible-exposed-infected-recovered (SEIR) model to simulate the possible development of this epidemic in China throughout 2020 in accordance with our predictions. By now, the COVID-19 epidemic has exacerbated into a global pandemic. We also applied our method to USA's and global COVID-19 data and the results propose that there is no any sign of imminent amelioration for this pandemic.

RESULTS AND DISCUSSION

At the very beginning, we did not know that SARS-CoV-2 has an incubation period. Thus, we started with a susceptible-infected-recovered (SIR) model with 20,000,000 possible susceptibles (twice the population size of Wuhan city) and 27 initial infected cases (news reported that there were 27 confirmed infected cases on Dec 31, 2019) [5]. The susceptible population was set as large as possible, because we did not how many people this disease was going to affect on Jan 24, 2020.

Using the data from World Health Organization's COVID-19 situation reports and the SIR model, we nowcasted the numbers of infections and recoveries on the next day from Jan 23, 2020. We did not use the SEIR model for nowcasting, because the number of exposed cases was unknown and would not be reported. Through changing the transmission rate (β) and recovery rate (γ) in our SIR model, we first calibrated the numbers in our model to the reported data on the previous day as close as we could. For example, the reported infected and recovered cases were 830 and 34 on Jan 23, respectively. By changing β and γ, the infection and recovery numbers on Jan 23 were set to be 836.96 and 37.44 in our SIR model. The model predicted that the infection and recovery numbers were 1020.23 and 43.76 on Jan 24, while the reported infected and recovered cases were 1287 and 38 on the same day (Tables 1 and 2). We only provided a nowcast for the next day, because the transmission of the COVID-19 epidemic in China was strongly influenced by the government's control, *e.g.*, the lockdown of Wuhan city and the restricted traffics in the other provinces, which make its development a stochastic process and long-term prediction inaccurate.

Comparing our nowcasts to the reported data, we can see the general trend of the COVID-19 epidemic's development in China (Fig. 1A and B). Figure 1A shows that before Jan 30, 2020, our model infection predictions are always lower than the reported infection numbers, but higher than the reported ones after it, except on Feb 12, when the clinically diagnosed cases were started to be added to the total number of infected cases.

Table 1 The reported infected and recovered cases and the calibrated number of infected and recovered cases in our SIR model

Date	The reported infected cases	The calibrated number of infected cases	The reported recovered cases	The calibrated number of recovered cases
2020/01/23	830	876.96	34	37.44
2020/01/24	1287	1289.74	38	41.27
2020/01/25	1975	1991.59	49	54.38
2020/01/26	2744	2764.4	51	56.87

Table 2 The reported infected and recovered cases and the predicted infected and recovered on the next day

Date	The reported infected cases	The predicted infected cases	The reported recovered cases	The predicted recovered cases
2020/01/24	1287	1020.23	38	43.76
2020/01/25	1975	1525.83	49	48.99
2020/01/26	2744	2365.41	51	64.73
2020/01/27	4515	3326.47	60	68.55

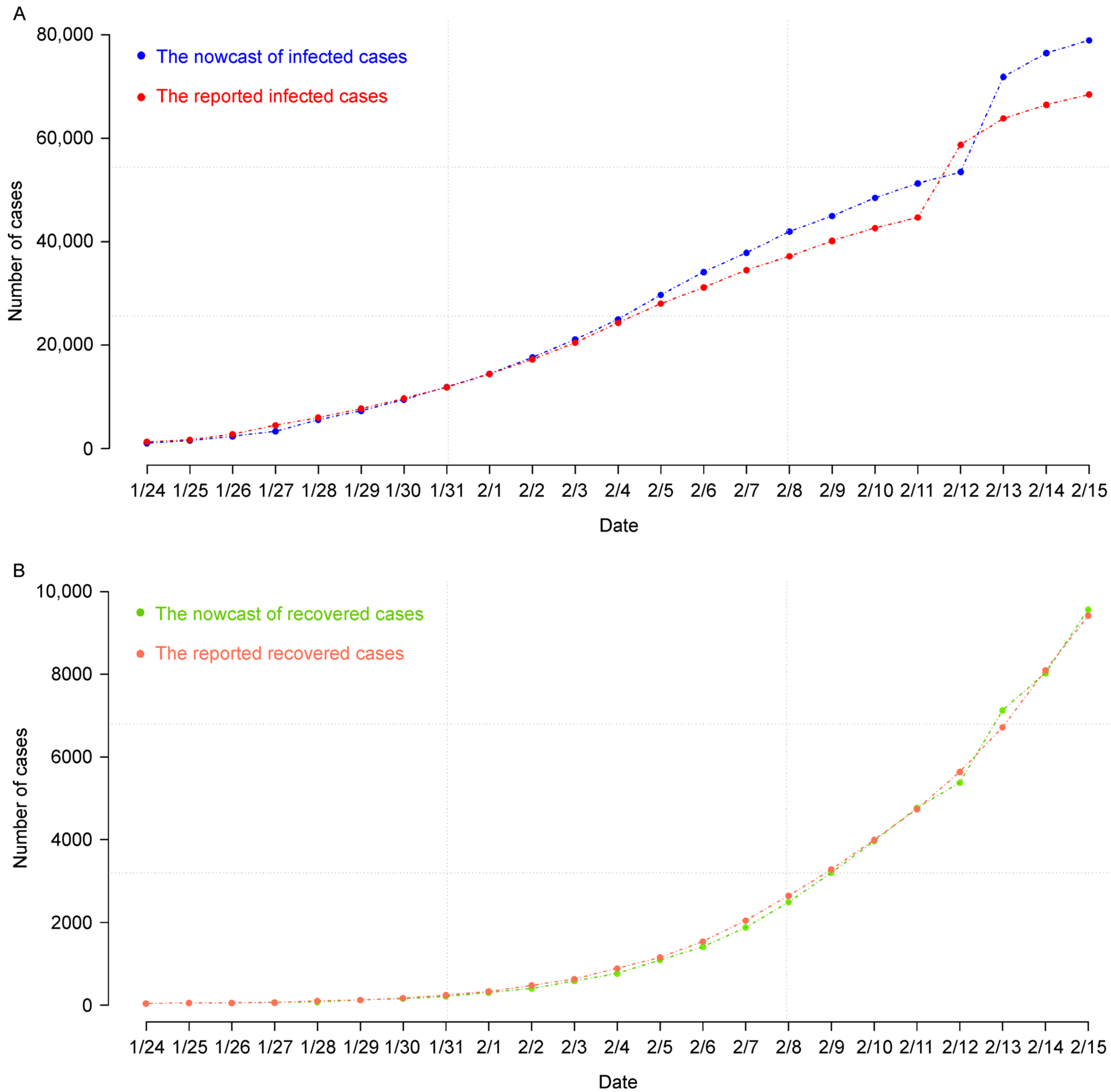

Figure 1. Comparison of our nowcasts with the reported numbers in China. (A) The nowcasts of infected cases vs. the reported infected cases from Jan 24 to Feb 15, 2020. (B) The nowcasts of recovered cases vs. the reported recovered cases from Jan 24 to Feb 15, 2020.

Thus, we concluded that the COVID-19 transmission rate started to slow down after Jan 30. Figure 1B shows that our model recovery predictions are always higher than the reported recovery numbers between Jan 29 and Feb 10. Together, these results propose that the things are starting to get better from the last day of January, 2020.

After Feb 1, our infection nowcasts have become more and more inaccurate. When the increased infection cases reached the highest number of 14,108 on Feb 12 (including 1820 lab-confirmed cases and 12,288 clinically diagnosed cases) and began to sharply decline after it, we used the different mathematic tool to forecast the inflexion point of the COVID-19 transmission in China. Two linear regression formulas were fitted with the daily increased infection and recovery numbers and the two fitted lines are crossed between Feb 17 and 18 (Fig. 2A). Therefore, we infer that the inflexion point of the COVID-19 epidemic in China would appear around Feb 18.

If the inflexion point did appear on Feb 18, we estimated that this disease would infect at least 80,000

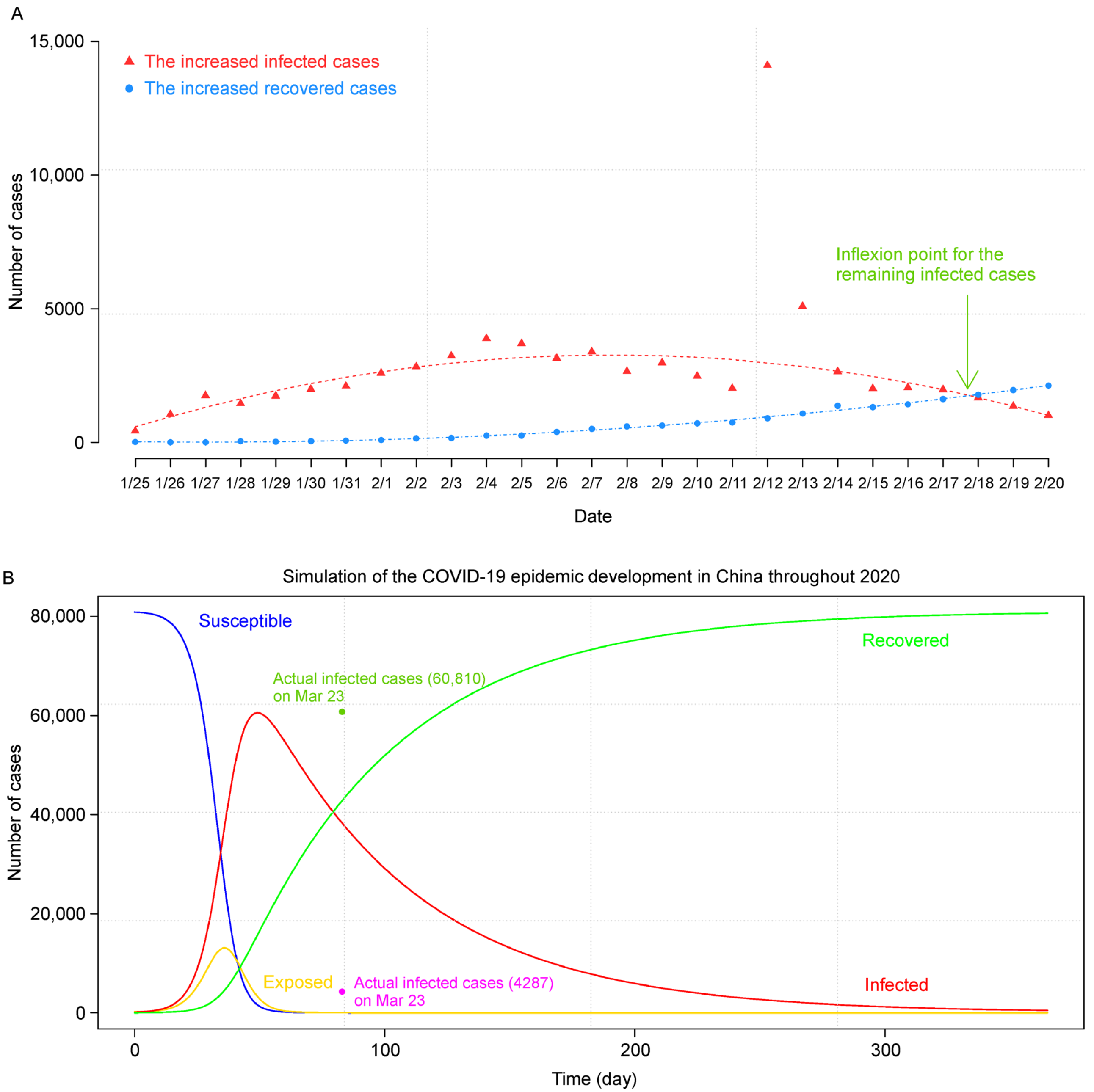

Figure 2. The forecast of the development of COVID-19 epidemic in China. (A) The forecast of the inflexion point of COVID-19 epidemic based on the increased infected and recovered cases. The daily increased infection and recovery numbers fitted with regression lines, $y_1 = 169.19 + 425.11x - 14.59x^2$ for infection (the number on Feb 12 treated as the outlier and not used for linear regression, $R^2 = 0.19$, p-value = 0.09) and $y_2 = 39.9684 - 22.3055x + 3.7608x^2 - 0.0029x^3$ for recovery ($R^2 = 0.99$, p-value < 2.2e−16). After Feb 15, the increased infection and recovery numbers were calculated with two fitted equations above. (B) SEIR model simulation of the COVID-19 epidemic transmission in China throughout 2020.

people, almost 1% of Wuhan's population. Then, we used the SEIR model to simulate what would happen if the inflexion point appeared on Feb 18 in accordance with our predictions above. After setting the susceptible population to 82,000, the transmission rate (β) to 0.33, the recovery rate (γ) to 0.016, and the incubation rate (σ) to 0.3, the simulation model forecasted that the infected cases would be 60,445 and the recovered cases would be 14,545 on Feb 18, 2020 (Fig. 2B). By this forecast, we would have COVID-19 as a company by the end of 2020, if the recovery rate were not improved (Fig. 2B). The actual reported infected cases and recovered cases were 57,805 and 14,376 on Feb 18, 2020. After Feb 18, the remaining COVID-19 patients in China declined day by day [6].

The mean incubation period for COVID-19 is around 3 days and the basic reproduction number (R_0) for COVID-19 is estimated to be 3.28 [7,8]. In our SEIR simulation model, the incubation rate (σ) is 0.3, which is quite close to the number above since the incubation period equals to $1/\sigma$. However, the R_0 in our model is 20.625, much larger than the other estimations ($R_0 = \beta/\gamma$). We set the transmission rate (β) and the recovery rate (γ) according to when the inflexion point would appear and how many infected and recovered cases would be predicted on Feb 18. Since our SEIR simulation is very close to real conditions, the extra-large R_0 is unlikely to be an artefact. The modeling of controlling COVID-19 in China shows that R_0 is not a stationary and fixed value [9]. According to three epidemic dynamic stages, Zhao and Chen estimate that R_0 ranges from 4.7 to 0.47 in Wuhan city [9]. The small R_0 at the third stage of COVID-19 epidemic in Wuhan demonstrates that strict quarantine measure is very effective. The extra-large R_0 in our SEIR simulation suggests that COVID-19 quickly exhausted the potential susceptible population and SARS-CoV-2 spread into every possible susceptible within a very short period of time in China. It actually proves the effectiveness of China's strict quarantine measure from another perspective.

On Mar 23, the number of new infected cases became zero and the number of remaining infected cases was 4287, which is much lower than our SEIR simulation's estimation (Fig. 2B). Moreover, the number of actual recovered cases was also higher than our SEIR simulation's estimation. The actual numbers of infected and recovered cases proposes that Chinese government does not only implement effective infection control and prevention measures but also actively engage in providing medical care for infected patients. The former decreases the COVID-19 transmission rate and the latter increases the recovery rate for the infected patient. That is why our results show that the scale of China's COVID-19 epidemic is much smaller than the previous estimations and China could contain the COVID-19 epidemic within two months. The other study also confirmed the effectiveness of Chinese government's COVID-19 response, especially its infection control and prevention measures [10].

By the end of March, the COVID-19 epidemic has developed into a global pandemic. The worst COVID-19 scenario is currently on the world's stage. We applied our method to USA's and global COVID-19 data between Jul 16 and Aug 2 to forecast the general trend of the COVID-19 pandemic in the rest of 2020. USA has the world's largest number of COVID-19 cases by now. Figure 3 shows our nowcasts of infected and recovered cases and the reported infected and recovered ones in USA and world. The reported infected cases are not significantly lower than the predicted ones in USA and world (Fig. 3A and C). In some days of July (*e.g.*, Jul 22), the reported infected cases were even higher than the predicted ones. Such trend can be also seen in the reported and predicted recovered cases (Fig 3B and D). Unlike what we observed in Fig. 1A, there is no clear division between the reported infected cases and the predicted infected ones in USA and world. This result proposes that there is no sign of imminent amelioration for this pandemic so far. Unfortunately, the COVID-19 pandemic is very like to last to next year.

DATA AND METHODS

The daily COVID-19 data of China were obtained from World Health Organization's situation reports (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/). The daily USA's and global COVID-19 data were retrieved from WHO coronavirus disease (COVID-19) dashboard (https://covid19.who.int/). The SIR and SEIR model construction and linear regressions were implemented in R package (version 3.6.2).

The differential equations for susceptible-infected-recovered (SIR) model in this study are as follows:

$$\begin{cases} \dfrac{\mathrm{d}S}{\mathrm{d}t} = -\beta\dfrac{SI}{N} \\ \dfrac{\mathrm{d}I}{\mathrm{d}t} = \beta\dfrac{SI}{N} - \gamma I \\ \dfrac{\mathrm{d}R}{\mathrm{d}t} = \gamma I \end{cases}$$

where β is the transmission rate, γ is the recovery rate, S, I, and R represent the number of susceptible, infected, and recovered individuals, respectively; $N = S + I + R$.

The differential equations for SEIR model in this study are as follows:

$$\begin{cases} \dfrac{\mathrm{d}S}{\mathrm{d}t} = -\beta\dfrac{SI}{N} \\ \dfrac{\mathrm{d}E}{\mathrm{d}t} = \beta\dfrac{SI}{N} - \sigma E \\ \dfrac{\mathrm{d}I}{\mathrm{d}t} = \sigma E - \gamma I \\ \dfrac{\mathrm{d}R}{\mathrm{d}t} = \gamma I \end{cases}$$

where β is the transmission rate, γ is the recovery rate, σ is the rate at which an exposed person becomes infective, S, I, E, and R represent the number of susceptible, infected,

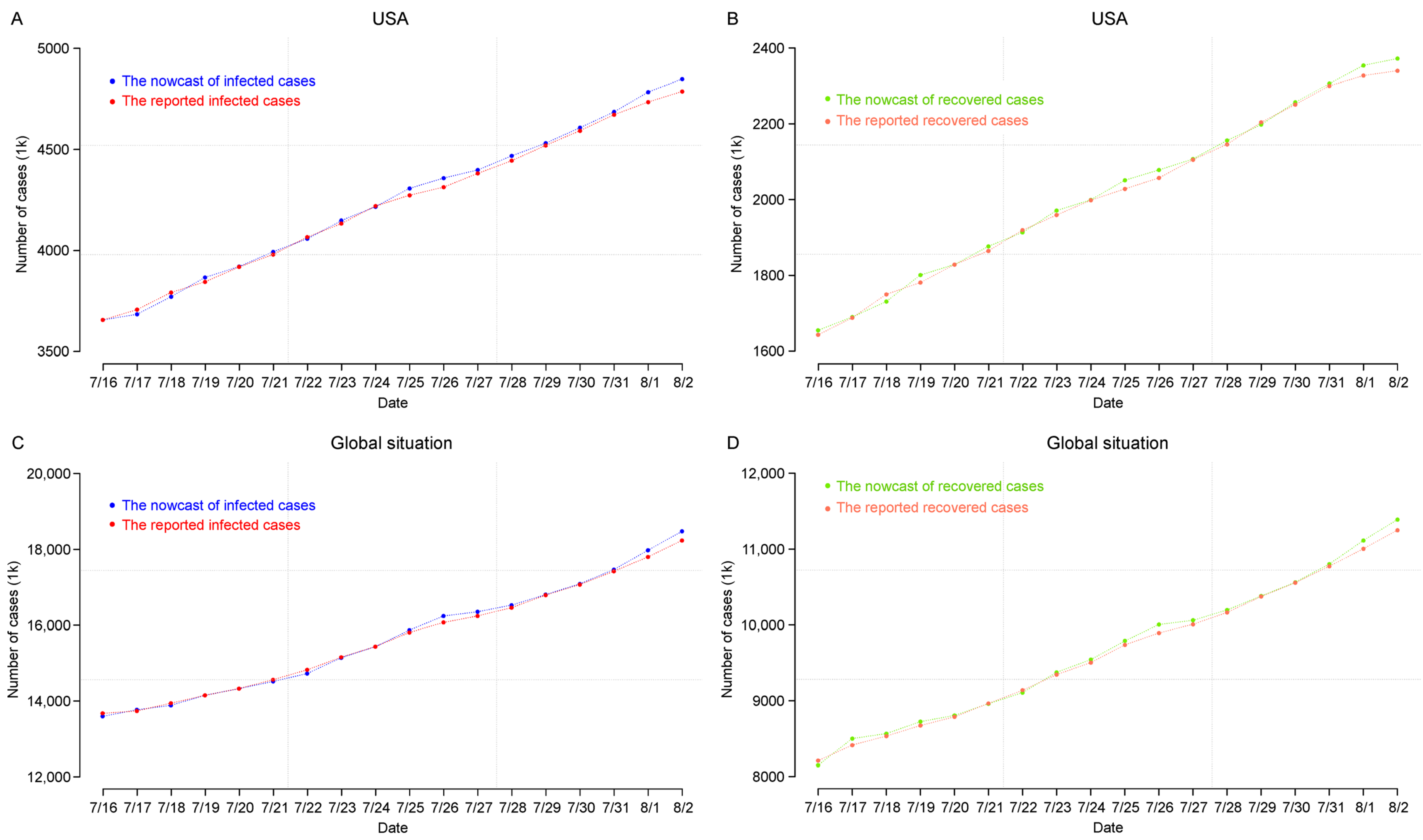

Figure 3. Comparison of our nowcasts with the reported numbers in USA and world. (A) The nowcasts of infected cases vs. the reported infected cases in USA from Jul 16 to Aug 2, 2020. (B) The nowcasts of recovered cases vs. the reported recovered cases in USA from Jul 16 to Aug 2, 2020. (C) The nowcasts of infected cases vs. the reported infected cases in world from Jul 16 to Aug 2, 2020. (D) The nowcasts of recovered cases vs. the reported recovered cases in world from Jul 16 to Aug 2, 2020. For the nowcasts in USA and world, we set the possible susceptible number as 160,000,000 and 3,850,000,000, respectively. Both numbers are the half population of USA and world. It is estimated that at least half population needs to be infected in a region for herd immunity to work.

exposed, and recovered individuals, respectively; $N = S + I + E + R$.

AUTHOR CONTRIBUTIONS

LBS devised this study and did modeling. YHZ collected and primarily processed data. YHZ, ZGL, LC and QLS analyzed the data and produced the figures. YHZ and ZGL wrote the first draft of this manuscript. ZGL and LBS revised the manuscript. All authors approved the submission of this work.

ACKNOWLEDGEMENTS

This study is supported by the National Natural Science Foundation of China (Nos. 81500503 and 81870456). We are grateful for Professor Liangsheng Zhang of Zhejiang University for his valuable advices on this work. LBS especially thanks his friends on WeChat for their zealous comments on this model, since it was first posted on Wechat Moments.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Yuehui Zhang, Lin Chen, Qili Shi, Zhongguang Luo and Libing Shen declare that they have no conflict of interests.

All procedures performed in studies were in accordance with the ethical standards of the institution or practice at which the studies were conducted, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

REFERENCES

1. Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W. and Hao, P. (2020) Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci. China Life Sci., 63, 457–460
2. WHO. Novel Coronavirus. Situation report–3 (2019-nCoV) 23 January 2020
3. Apolone, G., Montomoli, E., Manenti, A., Boeri, M., Sabia, F., Hyseni, I., Mazzini, L., Martinuzzi, D., Cantone, L., Milanese, G., *et al.* (2020) Unexpected detection of SARS-CoV-2 antibodies in the prepandemic period in Italy. Tumori, https://doi.org/10.1177/0300891620974755
4. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., *et al.* (2020) A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269
5. Vynnycky, E. and White, R. G. (2010) An Introduction to Infectious Disease Modelling, pp. 370. New York: Oxford University Press
6. Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., *et al.* (2020) Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J. Thorac. Dis., 12, 165–174
7. Liu, Y., Gayle, A. A., Wilder-Smith, A. and Rocklöv, J. (2020) The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Travel Med., 27, taaa021
8. Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S. M., Yuan, B., Kinoshita, R. and Nishiura, H. (2020) Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. J. Clin. Med., 9, 538
9. Zhao, S. and Chen, H. (2020) Modeling the epidemic dynamics and control of COVID-19 outbreak in China. Quant. Biol., 8, 11–19
10. Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., Wesolowski, A., Santillana, M., Zhang, C., Du, X., *et al.* (2020) Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature, 585, 410–413