

## Methods and Applications

# Determining the Covertness of COVID-19 — Wuhan, China, 2020

Chong You<sup>1,&</sup>; Xin Gai<sup>2,&</sup>; Yuan Zhang<sup>3,4,#</sup>; Xiaohua Zhou<sup>1,2,4,#</sup>

## ABSTRACT

**Introduction:** The coronavirus disease 2019 (COVID-19) pandemic has been going on for over a year and has reemerged in several regions. Therefore, understanding the covertness of COVID-19 is critical to more precisely estimating the pandemic size, especially the population of hidden carriers (those with very mild or no symptoms).

**Methods:** A stochastic dynamic model was proposed to capture the transmission mechanism of COVID-19 and to depict the covertness of COVID-19. The proposed model captured unique features of COVID-19, changes in the diagnosis criteria, and escalating containment measures.

**Results:** The model estimated that, for the epidemic in Wuhan, 79.8% (76.7%–82.7%) of the spread was caused by hidden carriers. The overall lab-confirmation rate in Wuhan up until March 8, 2020 was 0.17 (0.15–0.19). The diagnostic rate among patients with significant symptoms went up to 0.82 on March 8, 2020 from 0.43 on January 1, 2020 with escalating containment measures and nationwide medical supports. The probability of resurgence could be as high as 0.72 if containment measures were lifted after zero new reported (lab-confirmed or clinically confirmed) cases in a consecutive period of 14 days. This probability went down to 0.18 and 0.01 for measures lifted after 30 and 60 days, respectively.

**Discussion:** Consistent with the cases detected in Wuhan in mid-May, 2020, this study suggests that much of the COVID-19 pandemic is underreported and highly covert, which suggests that strict measures must be enforced continuously to contain the spread of the pandemic.

## INTRODUCTION

The outbreak of coronavirus disease 2019 (COVID-19) has been going on for over a year and has been deemed as a once-in-a-century health crisis (*1*). Many countries that believed they had gone through the

worst are now again grappling with new outbreaks (*1*). A major driving force in the persistence of COVID-19 is transmission caused by hidden carriers with very mild or no symptoms who are unaware of their infection (*2*). Due to the covertness of COVID-19 and overburdened medical resources, low diagnostic rates have been observed globally. The number of infected cases estimated by seroprevalence in various regions of the US were 6 to 24 times higher than the reported number (*3*). Therefore, understanding the covertness of COVID-19 is critical to getting a more precise picture of the pandemic size, especially the population of hidden carriers, and accordingly, making public health decisions such as the timing to lift containment measures.

Previous studies have unraveled the covert features of COVID-19, including incubation period, proportion and transmissibility of asymptomatic infections (*4–7*), and overall reporting rates (*7–8*). Hao et al. (2020) provided an important perspective on the transmission dynamics of COVID-19 in Wuhan. The key finding inferred that 87% of the infections before March 8, 2020 were not lab-confirmed (note that a lab-confirmed case is defined as a case who is symptomatic and tested positive for COVID-19; clinically-confirmed cases and detected asymptomatic cases were excluded) (*8*). However, due to limitations such as violation of the homogeneity assumption in compartment dynamic models, flawed inference of transmissibility parameters, and the inability to distinguish the covert nature of COVID-19 from external factors (i.e., overburdened medical resources), the study failed to capture some important information of covertness which led to an overoptimistic estimate on the all-clearance date in Wuhan (see [Supplementary Material Appendix A](#) for details, available in <http://weekly.chinacdc.cn/>).

In this paper, inspired by the SAPHIRE model in Hao et al. (2020), an improved dynamic model was proposed to depict the covertness of COVID-19 to gain a better understanding of the diagnostic rate and the probability of resurgence under different policies.

## METHODS

The proposed model included 6 compartments: susceptible ( $S$ ), exposed ( $E$ ), presymptomatic infectious ( $P$ ), infectious with significant symptoms ( $I$ ), infectious without significant symptoms ( $A$ ), and removed ( $R$ ). The individuals in  $I$  had significant symptoms that would be diagnosable but might not be lab-confirmed. For example, when the medical system was seriously overburdened, even a severe symptomatic case might not be diagnosed or hospitalized promptly before the viral load had dropped to the level that the infection could no longer be detected by tests. Meanwhile, compartment  $A$  included hidden carriers who were unlikely to be detected. Therefore, compared to the original SAPHIRE model, individuals in  $A$  or  $I$  would now show better homogeneity, which was a required underlying assumption in compartmental dynamic models. Individuals in  $A$  could only transit to  $R$  by losing transmissibility pathologically, while patients in  $I$  might reach  $R$  either by losing their transmissibility pathologically or by isolation upon lab-confirmation or clinical diagnosis. The transmission rate was set to vary over different time periods based on the local social events, medical resources, and implementation of containment measures. The covertness of COVID-19 can be better understood by dividing the diagnostic rate (hereafter referred to as lab-confirmation rate) into the following two components:

- 1) the natural characteristics of COVID-19 on covertness, i.e. the fraction of those infectious virus carriers without significant symptoms throughout their course of disease;

- 2) factors related to medical resources and containment measures, which could be replenished and help improve the diagnostic rate in the later stages of the outbreak.

In the proposed model, these components were estimated separately. In addition, in reality, transmissibility decreased towards the end of the infectious period (9), hence, the assumption of a constant transmission rate throughout compartment  $I$  and  $A$  might potentially lead to an overestimation of effective reproduction number  $R_e$  in the early stages. The proposed model had also taken these issues into consideration (see Appendix B for the solution details and ODE system description, available in <http://weekly.chinacdc.cn/>).

## RESULTS

The same data from January 1 to February 29 from Wuhan in Hao et al. (2020) were used here for comparison. A total of 5 time periods were identified as in Hao et al. (2020) in Wuhan for the transmission rate to vary. Based on the stochastic dynamic model, the effective reproduction  $R_e$  in Wuhan was 4.49 (4.01–5.00) and 4.10 (3.71–4.52), respectively, in the first 2 periods of January 1–9 (before *Chunyun*, the period of intense travel preceding the Spring Festival) and January 10–22 (*Chunyun*), then dropped dramatically to 1.00 (0.92–1.07), 0.43 (0.40–0.46), and 0.27 (0.22–0.33) in the later 3 periods (Figure 1B). These implied that the pandemic had been preliminarily controlled since the third period. It was estimated that 79.8% (76.7%–82.7%) of the spread of the disease was caused by hidden carriers (namely population in  $A$  and  $P$ ). The estimated cumulative number of infections up until March 8 was 194,302 (170,190–220,691) and the overall lab-confirmation rate was 0.17 (0.15–0.19), which was of a similar order as in Hao et al (2020).

However, such a low lab-confirmation rate was mostly caused by the high proportion of asymptomatic or very mild cases which was estimated to be 0.74 (0.71–0.77), while the diagnostic rate among patients with significant symptoms (namely patients in  $I$ ) went up to 0.82 on March 8 from 0.43 on January 1 because of the escalating containment measures and nationwide medical support (Figure 1C). This estimation indicated that most of those with significant symptoms in Wuhan would be promptly diagnosed in the later stage of the outbreak. The clearance of all active infections would occur on June 3 (May 15 to July 5) assuming the trend remained unchanged as in the fifth period. These estimates were significantly more consistent with the confirmed cases reported in Wuhan in May, and the hidden carriers detected in subsequent city-wide testing (10). In regard to the decision making of continuous surveillance and interventions, the model found that if all measures were lifted after zero new reported (lab-confirmed or clinically-confirmed) cases in a consecutive period of 14 days, the probability of resurgence was still as high as 0.72 (Figure 1D). This probability went down to 0.18 and 0.01 if measures were lifted after zero new confirmed case in a consecutive period of 30 and 60 days, respectively. See Appendix C for more details, available in <http://weekly.chinacdc.cn/>.

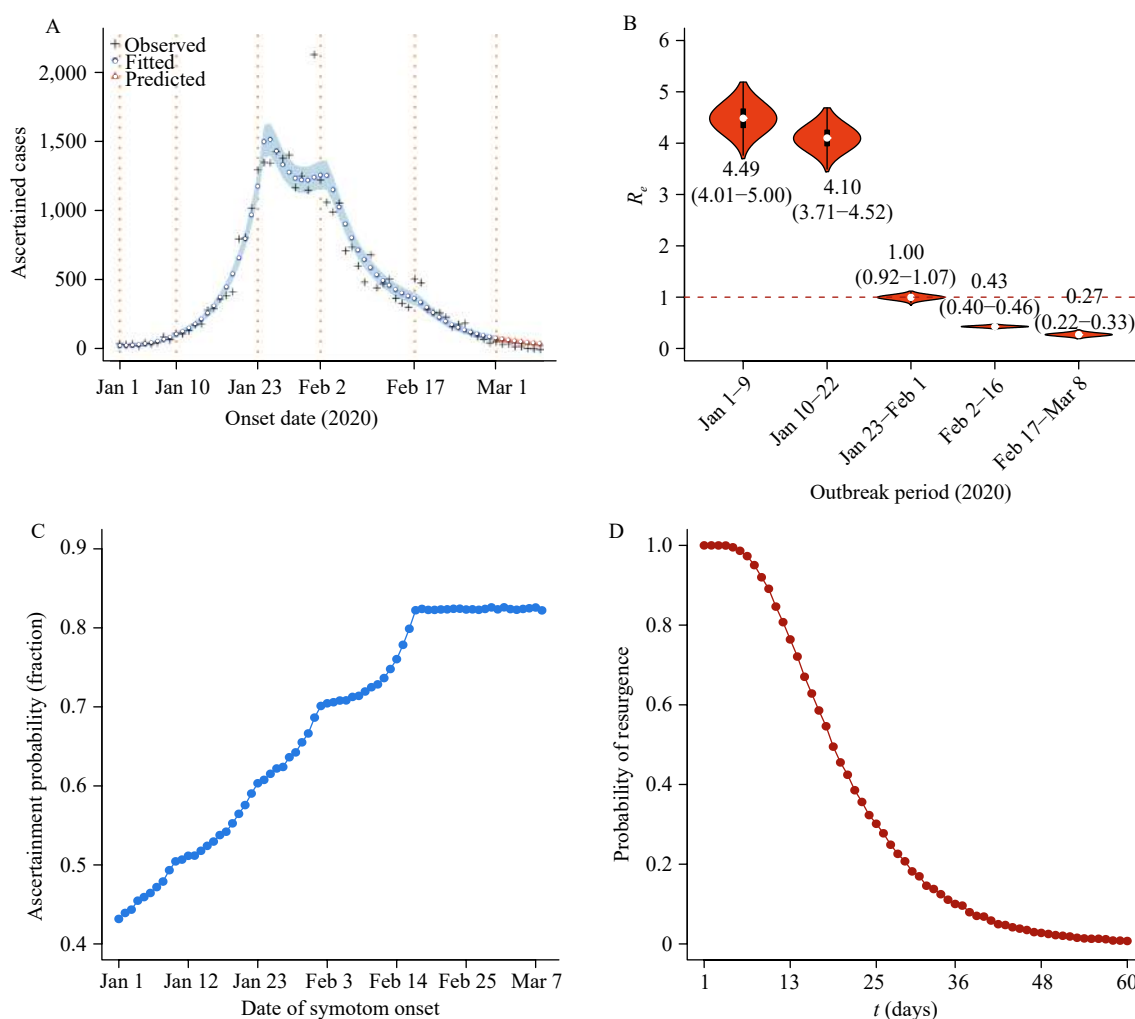


FIGURE 1. Results of primary analysis: epidemic trends, effective reproduction numbers, ascertainment probabilities, and probabilities of resurgences. (A) Fitting and prediction using parameters from the fifth period (February 17–29), the shaded area is the 95% credible intervals, and the colored points are the mean values based on 20,000 MCMC samples. (B) Distribution of  $R_e$  estimates from 20,000 MCMC samples. (C) Diagnostic rate among patients with significant symptoms. (D) Probability of resurgence if control measures were lifted after zero new lab-confirmed cases for  $t$  days consecutively.

Note: (C) the y-axis is the probability that a patient with significant symptom can be ascertained.

Abbreviation: MCMC=Markov chain Monte Carlo.

## DISCUSSION

In this study, the model estimation is consistent with the number of cases detected in Wuhan in mid-May. In comparison with reproduction numbers in other published studies, the estimate in the first period is in the range but on the higher side (11–13). It is possibly due to most of the other published reproduction numbers being estimated for the period after first period in this study (January 9) and that earlier data were not as complete, which might lead to an overestimation in reproduction number. Furthermore, the results suggested that COVID-19 was highly covert, the spread of the disease in Wuhan

was mostly caused by hidden carriers, and the probability of resurgence was high even if the measures were retained for 14 consecutive days after reaching 0 new reported cases, which may explain the resurgence in new infections over the past few months in other countries. As a result, continuous and sometimes even painstaking endeavors have to be made in order to contain the spread of the pandemic. Large-scale testing is encouraged towards the end of a significant outbreak to identify and quarantine hidden carriers before a city or nation can be safely reopened. In particular, the model implies that it takes more than 4 months in Wuhan from a strict lockdown to the final clearance of all active infections.

The model in this study can be extended to fit

pandemic data outside Wuhan, though modifications are needed with respect to specific countries or regions. For example, in the United States, statistics on daily reported cases are publicly available, but 1) no distinction is made between symptomatic and asymptomatic cases, 2) the dates of symptoms onset are mostly unavailable, 3) most reported cases are only required to self-isolate, which means that even if a case is reported, there is still a chance of infecting others. Moreover, we also need to track the changes of non-pharmaceutical interventions within the region of interest, to which (14) can be a helpful reference. To accommodate such differences, an additional compartment is needed for reported cases.

To conclude, the proposed model reflects the unique features of COVID-19, the changes in the diagnostic criteria, and the escalating containment measures, and hence the corresponding model estimates offer a better understanding of the diagnostic rate and the probability of resurgence under different policies. COVID-19 is highly covert, 74% (71%–77%) of the virus carriers had no/mild symptoms, 80% (77%–83%) of the spread of the disease was caused by those hidden carriers, and as a result, the probability of resurgence is high.

This study shares some limitations in Hao et al. (2020), for example, the assumption of homogeneous transmission rate within the population while ignoring heterogeneity between groups by sex, age, geographical region, socioeconomic status. Moreover, the population movement in this study is modelled under the same relatively simple setting as in Hao et al. (2020). This is acceptable for the case of Wuhan due to the travel restriction since January 23. More sophisticated modelling on travel flows is needed in order to generalize this model to other regions. Finally, the recently reported SARS-CoV-2 mutants may pose potential challenge to the generalization of the proposed model. Even when control measures remain unchanged, the emergence and spread of new COVID-19 variants may change the transmission rate and thus make the epidemic trend deviate from our prediction. These will be explored in future research when more relevant epidemiological data are available.

**Conflicts of interest:** The Authors declare that there is no conflict of interest.

**Funding:** National Natural Science Foundation of China grant 8204100362 and The Bill & Melinda Gates Foundation (INV-016826).

**Acknowledgements:** Authors of Hao et al. (2020).

doi: 10.46234/ccdcw2021.048

\* Corresponding authors: Yuan Zhang, zhangyuan@math.pku.edu.cn;

Xiaohua Zhou, azhou@math.pku.edu.cn.

<sup>1</sup> Beijing International Center for Mathematical Research, Peking University, Beijing, China; <sup>2</sup> Department of Biostatistics, School of Public Health, Peking University, Beijing, China; <sup>3</sup> School of Mathematical Sciences, Peking University, Beijing, China; <sup>4</sup> Center for Statistical Sciences, Peking University, Beijing, China.

<sup>✉</sup> Joint first authors.

Copyright © 2021 by Chinese Center for Disease Control and Prevention. All content is distributed under a Creative Commons Attribution Non Commercial License 4.0 (CC BY-NC).

Submitted: January 15, 2021

Accepted: February 08, 2021

## REFERENCES

1. WHO. COVID-19 Emergency Committee highlights need for response efforts over long term. <https://www.who.int/news/item/01-08-2020-covid-19-emergency-committee-highlights-need-for-response-efforts-over-long-term>. [2020-8-1].
2. WHO. Virtual press conference on COVID-19 in the western Pacific, Remarks by Dr Takeshi Kasai. <https://www.who.int/westernpacific/news/speeches/detail/virtual-press-conference-on-covid-19-in-the-western-pacific>. [2020-8-18].
3. Havers FP, Reed C, Lim T, Montgomery JM, Klena JD, Hall AJ, et al. Seroprevalence of antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23–May 12, 2020. *JAMA Intern Med* 2020;180(12):1576–86. <http://dx.doi.org/10.1001/jamainternmed.2020.4130>.
4. Hu ZL, Song C, Xu CJ, Jin GF, Chen YL, Xu X, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci* 2020;63(5):706–11. <http://dx.doi.org/10.1007/s11427-020-1661-4>.
5. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung SM, Hayashi K, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int J Infect Dis* 2020;94:154–5. <http://dx.doi.org/10.1016/j.ijid.2020.03.020>.
6. Zhang Y, You C, Cai ZH, Sun JR, Hu WJ, Zhou XH. Prediction of the COVID-19 outbreak in China based on a new stochastic dynamic model. *Sci Rep* 2020;10(1):21522. <http://dx.doi.org/10.1038/s41598-020-76630-0>.
7. Li RY, Pei S, Chen B, Song YM, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 2020;368(6490):489–93. <http://dx.doi.org/10.1126/science.abb3221>.
8. Hao XJ, Cheng SS, Wu DG, Wu TC, Lin XH, Wang CL. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* 2020;584(7821):420–4. <http://dx.doi.org/10.1038/s41586-020-2554-8>.
9. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* 2020;581(7809):465–9. <http://dx.doi.org/10.1038/s41586-020-2196-x>.
10. City-wide screening results (5 days) released in Wuhan, 58 new asymptomatic cases detected. [2020-5-20]. (In Chinese).
11. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med* 2020;27(2):taaa021. <http://dx.doi.org/10.1093/jtm/taaa021>.
12. Sanche S, Lin YT, Xu CG, Romero-Severson E, Hengartner N, Ke RA, et al. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg Infect Dis* 2020;26(7):1470–7. <http://dx.doi.org/10.3201/eid2607.200282>.
13. Zhao SL, Chen H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quant Biol* 2020;8(1):11–9. <http://dx.doi.org/10.1007/s40484-020-0199-0>.
14. Hale T, Angrist N, Cameron-Blake E, Hallas L, Kira B, Majumdar S, et al. Oxford COVID-19 government response tracker. Blavatnik school of government. <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>. [2020-06-22].



## Supplementary Material

### Appendix A: Concerns on the SAPHIRE model proposed by Hao et al. (2020) (1)

The SAPHIRE model included 7 compartments susceptible ( $S$ ), exposed ( $E$ ), presymptomatic infectious ( $P$ ), ascertained infectious ( $I$ ), unascertained infectious ( $A$ ), isolation in hospital ( $H$ ), and removed ( $R$ ). For ease of understanding,  $R$  was suggested understanding as the loss of transmissibility pathologically to be distinguished from  $H$ . An individual in  $S$  would be infected by individuals in  $P$ ,  $I$ , or  $A$  with different transmissibility to get into  $E$  and then  $P$  after a latent period. At the time point of symptoms onset, an individual transitioning from  $P$  to  $I$  or  $A$  depending on whether they would be lab-confirmed in the future, and the ascertainment rate ( $r$ ) is the ratio that a patient would be lab-confirmed. For a case to be lab-confirmed, the patient must be both symptomatic and testing positive, which means individuals in  $I$  must be symptomatic, while those who were in  $A$  could be asymptomatic and their symptoms onset stage was just a hypothetical one which was included in the model for simplicity. The individuals in  $A$  would then lose their transmissibility pathologically and transitioned into  $R$ . Meanwhile, individuals in  $I$  would either lose their transmissibility pathologically ( $R$ ) before they got confirmed and isolated in hospital, which implies that a patient can be no longer infectious but still test positive or become isolated in hospital ( $H$ , lost their transmissibility physically) and eventually then lost their transmissibility pathologically ( $R$ ). The parameters  $r$  and transmission rate ( $b$ ) vary across five time periods based on key events (e.g., *Chunyun*, the period of intense travel preceding the Spring Festival) and containment interventions. Different from most of other dynamic models fitting number of confirmed diagnosis at time  $t$ , the numbers of individuals in all compartment in this model were not directly observable except in  $I$  where  $I(t)$  is the number of lab-confirmed cases who reported their date of symptoms onset was on time  $t$ .

In consideration of the SAPHIRE model described as above, the following four aspects can be potentially improved:

1) The initial ascertainment rate  $r$  was estimated based on the assumption of perfect ascertainments in Singapore ignoring asymptomatic individuals which certainly gave an over-conservative estimate of  $r$  under the current model as mentioned in Hao et al. (2020). In addition,  $r$  should be a continuous function rather than a step function over the five time periods, see the justification in [Appendix D](#).

2) The individuals in  $A$  could be very different including asymptomatic, mild cases, and severe cases as evidence by deaths of clinically confirmed cases reported in (2), it is hence not optimal to assign a same transmission rate to all individuals in  $A$  (note that the proposed transmission rate in  $A$  was identical to that of the presymptomatic infectious period  $P$  and was  $\alpha = 55\%$  of that in  $I$ ). At the beginning of the pandemic, the medical resources were overburdened. It was likely to have a larger fraction of patients with severe symptoms in  $A$  and thus the transmission rate would be close to that of  $I$ . When medical resources were replenished and strong screening and public awareness campaign were implemented, the remaining unascertained cases should mostly be asymptomatic or mildly symptomatic, and the transmission rate would be closer to that of  $P$ . See why this issue can not be easily resolved in [Appendix D](#).

3) Though only the data of lab-confirmed cases were used in the SAPHIRE model, the isolation due to clinical diagnosis cannot be simply ignored in the model. As mentioned in Hao et al. (2020) the clinically diagnosed cases were excluded in the model. However, there were a significant number of cases in  $A$  of the SAPHIRE model who were not lab-confirmed but clinically confirmed and isolated in *fangcang* (square-cabin) hospitals in Wuhan during February 2020 and thus lost their transmissibility before they actually got into  $R$ , which implies that clinically-confirmed cases in  $A$  would have a faster rate to get into  $R$  than other cases in  $A$  (3).

4) The pre-determined symptomatic infectious period of  $D_i = 2.9$  Days is questionable. The symptomatic infectious period  $D_i$  was the mean time from symptom onset to loss of transmissibility pathologically, and the value was calculated based on the claim that 44% of secondary cases were infected during the index cases' pre-symptomatic stage by He et al. (2020) (4). Regardless of whether such claim is correct (a matter arising paper to He et al. (2020) was published), this 44% of pre-symptomatic spread was estimated based on the confirmed cases with isolation measures outside Wuhan, which is certainly not appropriate to be used to estimate mean time from

symptom onset to loss of transmissibility pathologically. Furthermore, another defect in the calculation of  $D_i$  is the inconsistency in the study of Hao et al. (2020), where a constant infectiousness was assumed across the presymptomatic and symptomatic phases of ascertained cases in estimating  $D_i$  while in the meantime  $\alpha = 0.55$  was used as the ratio of transmission rate of cases in  $P$  (presymptomatic) to that of in  $I$  (symptomatic). It is important to note that unlike other pre-determined parameters in the model, the value of  $D_i$  is quite crucial to the model estimates of interest, see [Supplementary Table S1](#) in [Appendix D](#) for detail. Hence a more accurate choice of  $D_i$  is essential. It is worth to note that the clearance of all active infections in Hao et al. (2020) was predicted to occur on April 21 (April 8 to May 12), which was not consistent with the active cases detected in Wuhan in mid-May.

## Appendix B: The modified SAPHIRE model

Inspired the SAPHIRE model in Hao et al. (2020), an improved dynamic model was proposed. The proposed model included 6 compartments: susceptible ( $S$ ), exposed ( $E$ ), presymptomatic infectious ( $P$ ), infectious with significant symptoms ( $I$ ), infectious without significant symptom ( $A$ ), and removed ( $R$ ). Note that patients in  $A$  could only transit to  $R$  by losing transmissibility pathologically while patients in  $I$  may reach  $R$  by either losing their transmissibility pathologically, isolating upon lab-confirmation, or isolating upon clinical diagnosis. Therefore, the transition rate from  $I$  to  $R$  is given by  $D_i^{-1} + D_q^{-1} + D_c^{-1}$ , where  $D_i$  and  $D_q$  are the period of the symptomatic infectious period and duration from illness onset to lab-confirmation; and  $D_c$  is the duration from illness onset to clinical diagnosis which varies at different time period (here let  $D_c$  equals to infinity and 10 days before and after February 2 in Wuhan).

Furthermore, it has been found that the transmissibility decays towards the end of infectious period, hence the assumption of a constant transmission rate  $b$  throughout compartment  $I$  and  $A$  might potentially lead to an overestimate of the effective reproduction number  $R_e$  in the early stages. A preliminary solution to this problem is to split  $I$  into  $IE$  (early stage of symptomatic infectious period with a higher transmissibility) and  $IL$  (late stage of symptomatic infectious period with a low transmissibility) and split  $A$  into  $AE$  (early stage of asymptomatic/mild infectious period with a higher transmissibility) and  $AL$  (late stage of asymptomatic/mild infectious period with a low transmissibility). The transition dynamics of these states are as follows:

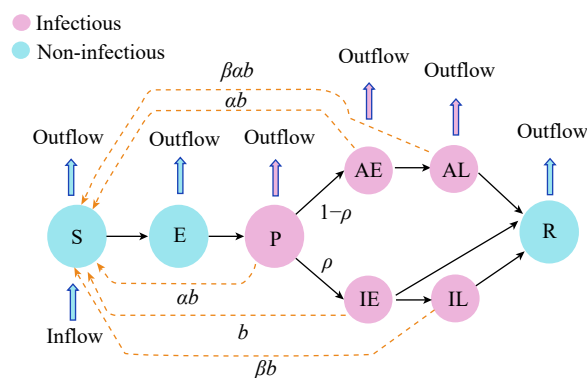
(1) Let  $IE \rightarrow IL$  and  $AE \rightarrow AL$  with a rate of  $1/D_i' = 1/3$  which corresponds to the setting in the SAPHIRE model, while  $IL \rightarrow R$  and  $AL \rightarrow R$  with a rate of  $1/D_i' = 1/7$ . Thus, the expectation of the symptomatic infectious period would be  $3 + 7 = 10$  days which is consistent with the choice of  $D_i = 10$  days.

(2) Individuals in  $IE$  have a transmission rate of  $b$  and individuals in  $AE$  have a transmission rate of  $\alpha b$ .

(3) Individuals in  $IL$  have a transmission rate of  $\beta_i b$  and individuals in  $AL$  have a transmission rate of  $\alpha \beta_i b$ , where  $\beta_i, \beta_a \in [0, 1]$  are the reduction factors of transmissibility in late stage and are unknown parameters to be inferred in the model.

Such modifications would grant us even better compartment homogeneity. See [Supplementary Figure S1](#) for model illustration.

The ODE system of the above model is as follow:



SUPPLEMENTARY FIGURE S1. Transmission dynamics of COVID-19 in Wuhan for modified SAPHIRE model .

$$\begin{aligned}
\frac{dS}{dt} &= n - \frac{bS(\alpha P + \alpha A_e + \alpha \beta_a A_l + I_e + \beta_i I_l)}{N} - \frac{nS}{N} \\
\frac{dE}{dt} &= \frac{bS(\alpha P + \alpha A_e + \alpha \beta_a A_l + I_e + \beta_i I_l)}{N} - \frac{E}{D_e} - \frac{nE}{N} \\
\frac{dP}{dt} &= \frac{E}{D_e} - \frac{P}{D_p} - \frac{nP}{N} \\
\frac{dA_e}{dt} &= \frac{(1-\rho)P}{D_p} - \frac{A_e}{D_i^e} - \frac{nA_e}{N} \\
\frac{dA_l}{dt} &= \frac{A_e}{D_i^e} - \frac{A_l}{D_i^l} - \frac{nA_l}{N} \\
\frac{dI_e}{dt} &= \frac{\rho P}{D_p} - \frac{I_e}{D_i^e} - \frac{I_e}{D_q} - \frac{I_e}{D_c} \\
\frac{dI_l}{dt} &= \frac{I_e}{D_i^e} - \frac{I_l}{D_i^l} - \frac{I_l}{D_q} - \frac{I_l}{D_c}
\end{aligned}$$

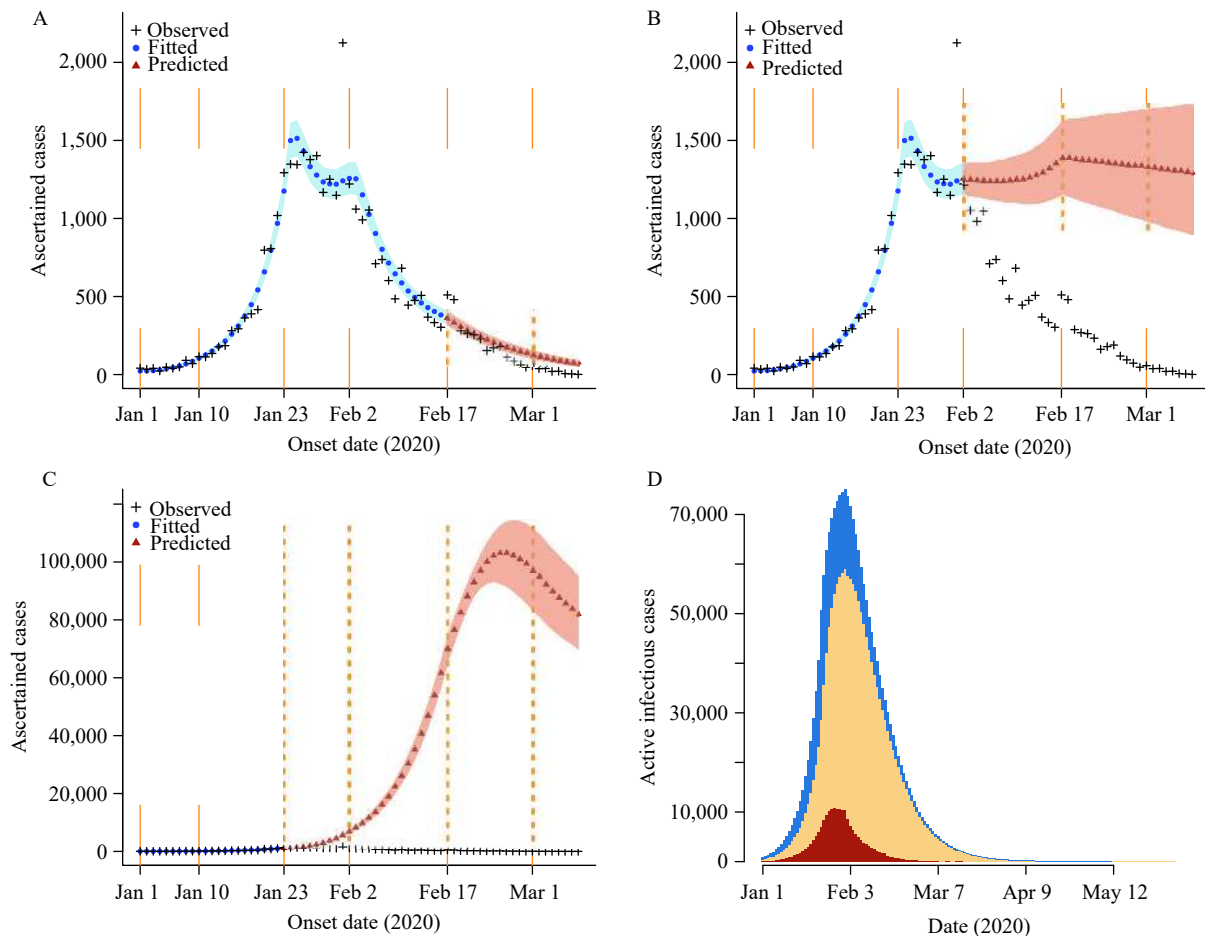
where  $b$  is the unknown transmission rate for significantly symptomatic cases in early stages and varies different time periods (set the five time periods as in Hao et al. (2020) for the demonstration in Wuhan);  $\alpha$  is the ratio of the transmission rate of presymptomatic/asymptomatic/mild cases to that of significantly symptomatic cases and is prefixed;  $\rho$  is the unknown fraction of infections with significant symptoms;  $\beta_i$  and  $\beta_a$  are defined as before, set  $\beta_i = \beta_a$  by assuming the trends of viral load are independent to symptoms;  $D_e$ ,  $D_p$ ,  $D_q$ , and  $D_c$  are the latent period, presymptomatic infectious period, duration from illness onset to isolation and duration from illness onset to clinical diagnosis respectively and are all predetermined. Under such setting, the transmission rate for  $AE$  is reasonable to be a constant over time and equal to the one for  $P$ . In addition, the lab-confirmation rate can be better presented as the function of the ratio between cases with insignificant (no/mild) and significant symptoms, and the time dependent ratio between the isolation/diagnosis and removal speed. [Appendix E and F](#) are for estimation method, choices of initial values, parameter settings, and sensitive analysis for the modified model.

## Appendix C: Detailed results

The similar estimation method in Hao et al. (2020) is used in this study. Note that all CIs without further specifications are 95% CIs throughout this paper. The estimated cumulative number of infections up until March 8 was 194,302 (170,190–220,691) by fitting data from all 5 periods, this number increased to 198,748 (173,856–226,051) if the trend of the fourth period was assumed, 355,907 (301,525–418,579) if the trend of the third period was assumed or 11,507,840 (10,688,446–12,336,104) if the trend of the second period was assumed, see [Supplementary Figure S2A–C](#) for details. These represented a 2.24%, 44.16%, and 96.91% reduction of infections by the measures taken in the fifth period, the fourth and the fifth periods combined, and the last three periods combined, respectively. Note that under the trend of second period the total number of infections exceeded the total population of Wuhan. It was because the population inflow and outflow in Wuhan was about 800,000 per day before lockdown, the estimated number of infections could therefore be regarded as the number of infection in/from Wuhan.

The number of daily active infections (including cases in  $P$ ,  $I$ , and  $A$ ) peaked at 75,093 (64,342–87,068) on February 1 and dropped to 6 (1–13) on May 14 ([Supplementary Figure S2D](#)). If the trend remained unchanged as in the fifth period, the number of significantly symptomatic infections ( $I$ ) would first become zero on April 12 (April 3 to April 23), and the clearance of all infections (namely  $E + P + I + A = 0$ ) would occur on June 4 (June 15 to July 5). Compared with the estimates in Hao et al. (2020), the estimate on  $A$  was much more heavily tailed implying the high covertness of COVID-19. Considering a few cases detected in Wuhan in mid-May, the estimate on the clearance of all active infections was more consistent with the official report in Wuhan than what was predicted in Hao et al. (2020) (5–6).

Regarding continuous surveillance and interventions, based on the modified model that if control measures were



SUPPLEMENTARY FIGURE S2. Results of secondary analysis: epidemic trends under hypothetic policies, and estimations of standing active virus carriers. (A) Prediction using parameters estimated from the fourth period (February 2–16). (B) Prediction using parameters estimated from the third period (January 23 to February 1). (C) Prediction using parameters estimated from the second period (January 10 to January 22). The shaded areas are 95% credible intervals, and the colored points are the mean values based on 20,000 MCMC samples. (D) Estimated number of pre-symptomatic infections ( $P$ , in blue), infections without significant symptoms ( $A$ , in yellow) and infections with significant symptoms ( $I$ , in red) in Wuhan.

lifted after zero new reported case in a consecutive period of 14 days, the probability of resurgence, defined as the number of active significantly symptomatic cases greater than 100, was still as high as 0.72, and the surge was predicted to occur on Day 41 (31–57) after lifting controls. Compared with the results in Hao et al. (2020), estimates on the probability of resurgence were much higher, which suggests continuous efforts in interventions is essential to contain the spread of the pandemic.

## Appendix D: Detailed explanation to issues in Hao et al. (2020)

The mean transmissibility in  $A$  defined by Hao et al. (2020) should be a time/configuration dependent function  $\alpha(t)$ , which were determined by averaging the transmissibility between patient with symptoms and those with no/mild symptom in  $A$ . Unfortunately, these two sub-populations of  $A$  were unidentifiable in the proposed model structure by Hao et al. (2020). One relatively easy potential fix of the issue was to replace  $\alpha A$  with  $\alpha_i A$  and allow  $\alpha_i$  changing over different time steps in the paper. However, according to the definition of  $A$ , patients that did not get diagnosed could only increase through symptom onset and decrease through the loss of infectivity. Thus, the evolution of the sub-population proportion within  $A$  must be continuous in time. Meanwhile, the changes in containment measures/lab-confirmed rates in different steps in this model could only lead to changes to the evolution rate but not the proportion itself. This fact made it inappropriate to treat  $\alpha$  as a step function. By similar



SUPPLEMENTARY TABLE S1. Sensitivity analysis on the choice of  $D_i$  the SAPHIRE model. With a longer infectious period, it is expected to obtain a higher reproduction number in the beginning of the epidemic and a longer time needed to clear all infections in the Wuhan.

$D_i$ (days)	Ascertainment rate	$R_e$ in Period 1	$R_e$ in Period 2	$R_e$ in Period 3	$R_e$ in Period 4	$R_e$ in Period 5	Date of the clearance of all infections
2.9	12.7% (10.2%–15.6%)	3.54 (3.41–3.68)	3.32 (3.20–3.45)	1.18 (1.11–1.25)	0.51 (0.47–0.54)	0.28 (0.23–0.32)	April 21 (April 8 to May 12)
5	12.2% (9.8%–15.1%)	4.16 (4.00–4.33)	3.80 (3.65–3.95)	1.37 (1.28–1.45)	0.49 (0.45–0.52)	0.21 (0.18–0.26)	May 7 (April 21 to May 30)
7	12.1% (9.8%–15.1%)	4.68 (4.48–4.88)	4.18 (4.00–4.35)	1.56 (1.45–1.67)	0.49 (0.45–0.53)	0.18 (0.15–0.22)	May 28 (May 9 to June 25)
9	11.9% (9.5%–14.7%)	5.12 (4.89–5.35)	4.48 (4.28–4.68)	1.76 (1.65–1.89)	0.51 (0.47–0.55)	0.16 (0.13–0.20)	June 21 (May 31 to July 19)
13	11.1% (8.9%–13.7%)	5.83 (5.56–6.10)	4.93 (4.71–5.14)	2.18 (2.03–2.34)	0.57 (0.52–0.61)	0.15 (0.12–0.19)	July 19 (July 18 to July 19)

argument, the setting of  $r$  in Hao et al. (2020) as a step function is also questionable.

## Appendix E: Estimation method

A major obstacle in parameter inference of the modified model is that now no compartment is directly observable. To be precise,  $I(d)$  is the inflow of  $I$  on day  $d$  satisfying a Poisson distribution with  $\lambda_d = \rho P_{d-1} D_p^{-1}$ , while the observed data only consisted of those who had symptoms onset on day  $d$  provided that they would be diagnosed in the future, namely, a sub-population in  $I$ . Therefore, the distribution of such sub-population was estimated under this model. Fortunately, by the thinning argument of a Poisson process, the size of the sub-population of interest satisfies Poisson distribution with  $\bar{\lambda}_d = r P_{d-1} D_p^{-1} q_d$ , where  $q_d$  is the probability that a patient with symptom onset on day  $d$  would be diagnosed in the future. Moreover given  $D_n$ , the duration from symptom onset to the time point that test turns to negative, and  $D_q$  are predetermined in Hao et al (2020), the exact value of  $q_d$  can be independently calculated using the following stochastic viewpoint of the dynamic model: consider  $N_1(t)$  a time-dependent Poisson Process/Poisson Point Measure, with intensity equals to  $D_q(d) = 21, 15, 10, 6, 3$  on different stages, and  $N_2(t)$  a time homogeneous Poisson Process with intensity  $D_n$  independent to  $N_1$ . For each  $d$ , define stopping times

$$\tau_{d,1} = \inf\{t \geq d, N_1(t) = N_1(t-) + 1\}$$

and

$$\tau_{d,2} = \inf\{t \geq d, N_2(t) = N_2(t-) + 1\}$$

be the first jumps times after  $d$ . Then  $q_d$  can be calculated can the probability that  $\tau_{d,1} < \tau_{d,2}$ . The specific values of  $q_d$  can either be calculated manually for each  $d$  or can be approximated numerically using frequencies obtained from multiple independent stochastic simulations. Since the precision of numerical simulation can be guaranteed by law of large numbers and large deviation theory, it was used here to approximate the values of  $q_d$ 's with  $10^5$  stochastic realizations for each  $d$ .

## Appendix F: Choices of initial values, parameter settings and sensitive analysis

Supplementary Table S2 provides a list of parameter settings in the modified model for all five periods. The initial values of  $I, A$  were estimated as follows:

(1) Let  $I^*(0)$  the number of symptoms onset cases during December 29 to 31 who would be lab-confirmed in the future;  $r_0 = 0.23$  be the initial lab-confirmed rate in Wuhan among symptomatic case which was calculated based on assuming complete diagnosis of early cases among symptomatic cases in Singapore; and  $\rho_0 = 0.7$  be the proportion of symptomatic patients (1,7-9).

(2) The initial population of symptoms onset patients in Wuhan namely  $A(0) + I(0)$  were hence given by  $I^*(0) r_0^{-1} \rho_0^{-1}$ .

(3) The ratio between  $A(0)$  and  $I(0)$ , it should be roughly the same as the unknown diagnosable ratio  $\rho$  in the

SUPPLEMENTARY TABLE S2. Parameter settings for five periods.

Parameter	Meaning	Jan 1–9	Jan 10–22	Jan 23–Feb 1	Feb 2–16	Feb 17–Mar 8
$b$	Transmission rate of significant cases	$b_{12}$	$b_{12}$	$b_3$	$b_4$	$b_5$
$\rho$	Fraction of infections with significant symptoms or diagnosable ratio	$\rho$	$\rho$	$\rho$	$\rho$	$\rho$
$\alpha$	Ratio of transmission rate of no/mild symptomatic patients to that of significantly symptomatic patients	0.55	0.55	0.55	0.55	0.55
$D_e$	Latent period	2.9	2.9	2.9	2.9	2.9
$D_p$	Presymptomatic infectious period	2.3	2.3	2.3	2.3	2.3
$D_q$	Duration from illness onset to lab-confirmation/isolation	21	15	10	6	3
$D_n$	Duration from symptom onset to negative test result	14	14	14	14	14
$D_c$	Duration from symptom onset to clinical-confirmed diagnosis/isolation	$\infty$	$\infty$	$\infty$	10	10
$N$	Population size	10,000,000	10,000,000	10,000,000	10,000,000	10,000,000
$n$	Daily inbound and outbound size	500,000	800,000	0	0	0

ODE system of the modified model. By “fix point iteration” method,  $\rho$  was estimated to be 0.26, hence  $I(0) = 0.26I^*(0)r_0^{-1}\rho_0^{-1}$  and  $A(0) = 0.74I^*(0)r_0^{-1}\rho_0^{-1}$ .

$$(4) AE(0) = A(0)D_i^e/D_i, AL(0) = A(0)D_i^l/D_i, IE(0) = I(0)D_i^e/D_i, IL(0) = I(0)D_i^l/D_i.$$

In the original paper,  $P_I(0)$  and  $E_I(0)$  stood for the numbers of lab-confirmed cases with onset during January 12, 2020 and during January 3 to 5, 2020. According to the same reasonings as for  $A(0) + I(0)$ , there were

SUPPLEMENTARY TABLE S3. Estimated transmission rates, overall lab-confirmation rate, and diagnosable ratio from the sensitivity analysis where  $D_n = 14$  is used in the main model.

	$r$	$b_{12}$	$b_3$	$b_4$	$b_5$	$\rho$	$\beta i = \beta a$
$D_n=10$	0.18(0.16, 0.21)	1.01(0.91, 1.10)	0.20(0.18, 0.23)	0.10(0.08, 0.13)	0.07(0.05, 0.1)	0.32(0.28, 0.37)	0.40(0.17, 0.66)
$D_n=14$	0.17(0.15, 0.19)	1.03(0.94, 1.13)	0.21(0.18, 0.24)	0.10(0.08, 0.12)	0.07(0.05, 0.09)	0.26(0.23, 0.29)	0.46(0.22, 0.71)
$D_n=21$	0.15(0.13, 0.17)	1.06(0.97, 1.15)	0.22(0.19, 0.24)	0.10(0.08, 0.12)	0.06(0.04, 0.09)	0.20(0.18, 0.23)	0.48(0.24, 0.73)

$E(0) = E_I(0)r_0^{-1}\rho_0^{-1}$  and  $P(0) = P_I(0)r_0^{-1}\rho_0^{-1}$ . The estimates of parameters of interest were given in Supplementary Table S3 with a sensitivity analysis on  $D_n$ . Note that it was reasonable to believe that the mean duration from symptom onset to negative test result is less than 21 days,  $D_n = 14$  was used in the main model (see upper left panel of Supplementary Figure S2 in (4) for reference). The estimates are relatively robust to the different choice of  $D_n$ .

## REFERENCES

- Hao XJ, Cheng SS, Wu DG, Wu TC, Lin XH, Wang CL. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* 2020;584: 420–4. <http://dx.doi.org/10.1038/s41586-020-2554-8>.
- Health Commission of Hubei Province. Situation report on 12 February 2020. [http://wjw.hubei.gov.cn/bmdt/tzgg/202002/t20200213\\_2025582.shtml](http://wjw.hubei.gov.cn/bmdt/tzgg/202002/t20200213_2025582.shtml). [2021-01-15]. (In Chinese).
- 10, 000 new “clinical diagnosed cases” confirmed in a day in Hubei. 2020. <https://news.sina.cn/gn/2020-02-13/detail-iimxyqvz2457971.d.html>. [2021-01-15]. (In Chinese).
- He X, Lau EHY, Wu P, Deng XL, Wang J, Hao XX, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* 2020;26: 672–5. <http://dx.doi.org/10.1038/s41591-020-0869-5>.
- Wuhan ended its 35 days of zero new cases. 2020. <https://news.ifeng.com/c/7wNzkK9Br6a>. [2021-01-15]. (In Chinese).
- City-wide screening results (5 days) released in Wuhan, 58 new asymptomatic cases detected. 2020. [http://www.sanyarb.com.cn/content/2020-05/20/content\\_495360.htm](http://www.sanyarb.com.cn/content/2020-05/20/content_495360.htm). [2021-01-15].
- Hu ZL, Song C, Xu CJ, Jin GF, Chen YL, Xu X, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci Chin Life Sci* 2020;63:706–11. <http://dx.doi.org/10.1007/s11427-020-1661-4>.
- Nishiura H, Kobayashi T, Suzuki A, Jung S, Hayashi K, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int J Infect Dis* 2020. <http://dx.doi.org/10.1016/j.ijid.2020.03.020>.
- Zhang Y, You C, Cai ZH, Sun JR, Hu WJ, Zhou XH. Prediction of the COVID-19 outbreak in China based on a new stochastic dynamic model. *Sci Rep* 2020. <http://dx.doi.org/10.1038/s41598-020-76630-0>.