

On the study of the dynamic model of COVID-19 in Wuhan

Chong You

Peking University

Xin Gai

Peking University

Yuan Zhang (✉ zhangyuan@math.pku.edu.cn)

Peking University

Xiao-Hua Zhou (✉ azhou@math.pku.edu.cn)

Peking University

Research Article

Keywords: COVID-19, transmission dynamics, ascertainment rate

Posted Date: November 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-96422/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

On the study of the dynamic model of COVID-19 in Wuhan

Chong You¹⁺, Xin Gai²⁺, Yuan Zhang^{3,4*}, Xiao-Hua Zhou^{1,2,4*}

1. Beijing International Center for Mathematical Research, Peking University, Beijing, China, 100871

2. Department of Biostatistics, School of Public Health, Peking University, Beijing, China, 100871

3. School of Mathematical Sciences, Peking University, Beijing, China, 100871

4. Center for Statistical Sciences, Peking University, Beijing, China, 100871

⁺Joint first author * Joint corresponding author

email: azhou@math.pku.edu.cn, zhangyuan@math.pku.edu.cn

Understanding the transmission mechanism and effects of interventions is critical to the prevention and control of the COVID-19 pandemic. A recent study by Hao et al (2020) [1] provided an interesting perspective on the transmission dynamics of COVID-19 in Wuhan and inferred that 87% of the infections before 8 March 2020 were not laboratory-confirmed. In this paper, we clarify the definitions of the model compartments and raise questions in regard to the underlying homogenous assumption within compartments and settings of the parameters in the dynamic model by Hao et al (2020), and furthermore offer a modified model to resolve these potential limitations. Compared with the model in Hao et al (2020), the active virus carriers were predicted to persist for a longer period in our model which is well consistent with the active virus carriers detected in Wuhan in mid-May.

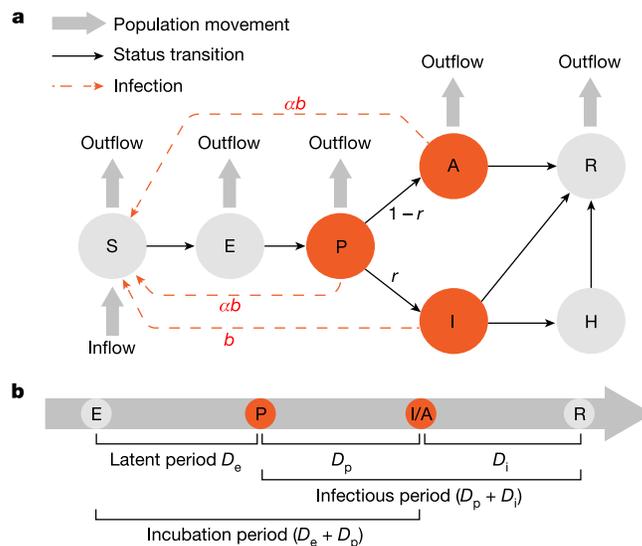


Figure 1: Illustration of the SAPHIRE model by Hao et al. (2020).

While Hao et al. (2020) brought useful insight into the spread of COVID-19 pandemic in Wuhan, it is a challenge to accurately understand, interpret and utilize the model they proposed due to the vagueness in the definitions of compartments and inconsistency in the settings of parameters. Here we rationalize the use of the model by clarifying the definitions of the model compartments as follows. The proposed SAPHIRE model by Hao et al (2020) included seven compartments susceptible (S), exposed (E), presymptomatic infectious (P), ascertained infectious (I), unascertained infectious (A), isolation in hospital (H) and removed (R) as illustrated in Figure 1. For ease of understanding, we suggest understanding R as the loss of transmissibility pathologically in order 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 transited from P to I or A depending on whether they would be laboratory-confirmed *in the future*, and r is the ratio that a patient would be laboratory-confirmed, namely, the ascertainment rate. Note that for a case to be laboratory-confirmed, the patient must be both symptomatic and tested positive by RT-PCR, 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 were included in the model for simplicity. The individuals in A would then lose their transmissibility pathologically and got into R . In the meantime, 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 tested positive by RT-PCR), or got isolated in hospital (H , namely lost their transmissibility physically) and then lost their transmissibility pathologically (R) eventually. The parameters r (ascertainment rate) and b (transmission rate) vary across five time periods on the basis of key events (“Chunyun”) and containment interventions. It is worth noting that, 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 laboratory-confirmed cases who reported their date of symptoms onset was on time t .

Based on such interpretation of the SAPHIRE model, four major concerns are to be raised.

(1) The initial ascertainment rate r was estimated based on the assumption of perfect ascertainment 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 in Hao et al (2020), see the justification in Appendix A.

(2) The individuals in A can be very different including asymptomatic and mild cases, as well as severe cases as evidence by deaths of clinically confirmed cases reported in [2], it is hence not rational to assign a same transmission rate to all individual 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). In fact, 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 . On the other hand, when medical resources were replenished and strong

screening and public awareness campaign were implemented, the remaining unascertained cases should be mostly asymptomatic or mildly symptomatic, and hence the transmission rate would be closer to that of P . See why this issue can **NOT** be easily resolved in Appendix A.

(3) As mentioned in Hao et al (2020) the clinically diagnosed cases were excluded in the model, however, there was indeed a significant amount of cases in A who were not laboratory-confirmed but clinically confirmed and isolated in (cabin) hospital in Wuhan during February 2020 and hence lost their transmissibility before they actually got into R namely lost their transmissibility pathologically which implies that clinically confirmed cases in A would have a faster rate to get into R than other cases in A [3]. Though only the data of laboratory-confirmed cases was used in Hao et al (2020), this does not mean the isolation due to clinical diagnosis can be simply ignored in the model.

(4) The pre-determined symptomatic infectious period $D_i = 2.9$ Days is highly questionable. The symptomatic infectious period D_i is the mean time from symptom onset to loss of transmissibility pathologically in our understanding, and the value was calculated based on the claim that 44% of secondary cases were infected during the index cases' presymptomatic stage by He et al (2020) [4]. Regardless of whether such claim is correct (a matters arising to that study was published), we have to notice that this 44% of presymptomatic spread was estimated based on the confirmed cases with isolation measure outside Wuhan, which is certainly not appropriate to be used to estimated 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 Table S1 in Appendix A for detail. Hence a more decent choice of D_i is essential.

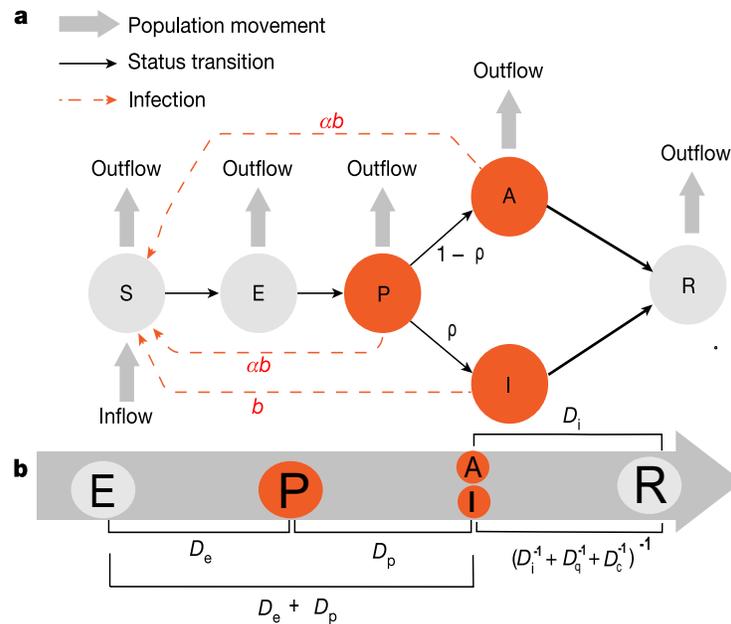


Figure 2: Illustration of the modified SAPHIRE model. The modified model includes seven compartments: susceptible (S), exposed (E), presymptomatic infectious (P), ascertainable infectious (I), unascertainable infectious (A) and removed (R). **A.** Relationship between different compartments. Two parameters of interest are ρ (ascertainable rate) and b (transmission rate). **B.** Schematic disease course of COVID-19. In this model, the unascertainable compartment A includes asymptomatic and mild cases whose symptoms were not significant enough to be detected, while the ascertainable compartment I includes symptomatic patients whose symptom were significant enough to be possibly ascertained.

To solve the aforementioned limitations in the current model, we propose a modified version of the SAPHIRE model. In our modified model, S , E and P together with dynamics and parameters associated with them remain unchanged. We redefine I as “**ascertainable infectious**” than “ascertained infectious”, that is, unlike in the SAPHIRE model, individuals I are not guaranteed to be ascertained but are those ones with symptom significant enough that could be possibly ascertained, for example a severe symptomatic case might not get laboratory-confirmed if his/her RT-PCR test was negative due to the prolonged waiting time. In the meanwhile, A is modified to include patients with no/mild symptoms exclusively who were **NOT** ascertainable. With such modification, individuals in A or I would be more homogeneous which is a required underlying assumption in any compartmental dynamic model. Furthermore, R now stands for removed for any reason which is in turn a combination of R and H in the original model, see Figure 2 for illustration. Note that patients in A can only transit to R by losing transmissibility pathologically while patients in I may reach R by either losing their transmissibility pathologically, or isolation upon laboratory-confirmation (tested positive by RT-PCR), or isolation upon clinical diagnosis. Thus, 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 laboratory-confirmed diagnosis; and D_c is the duration from illness onset to clinical diagnosis to be set as a step function which equals to infinity and 10 days before and after 2 February. Thus, the alternative model described above can be described by the following ODE system:

$$\frac{dS}{dt} = n - \frac{bS(\alpha P + \alpha A + I)}{N} - \frac{nS}{N} \quad (1.)$$

$$\frac{dE}{dt} = \frac{bS(\alpha P + \alpha A + I)}{N} - \frac{E}{D_e} - \frac{nE}{N} \quad (2.)$$

$$\frac{dP}{dt} = \frac{E}{D_e} - \frac{P}{D_p} - \frac{nP}{N} \quad (3.)$$

$$\frac{dA}{dt} = \frac{(1 - \rho)P}{D_p} - \frac{A}{D_i} - \frac{nA}{N} \quad (4.)$$

$$\frac{dI}{dt} = \frac{\rho P}{D_p} - \frac{I}{D_i} - \frac{I}{D_q} - \frac{I}{D_c} \quad (5.)$$

where b is the unknown transmission rate for ascertainable cases and varies over five time periods as in Hao et al (2020); α is the ratio of the transmission rate of presymptomatic/unascertainable cases to that of ascertainable cases

and is prefixed; ρ is the unknown fraction of infections with significant symptoms; D_e , D_p , D_i , D_q and D_c are the latent period, presymptomatic infectious period, symptomatic 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 A is reasonable to be a constant over time and equal to the one for P , and in addition, the ascertainment 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. We refer readers to Appendix B and C for estimation method, choices of initial values, parameter settings and sensitive analysis for our modified model.

Using the same estimation method as in Hao et al (2020), our model fit the observed data well. The effective reproduction R_e was 5.24 (5.08 - 5.39) and 4.57 (4.44 - 4.70) respectively in the first two periods namely 1-9 January (before Chunyun) and 10-22 January (Chunyun), then dropped dramatically to 1.19 (1.13 - 1.24), 0.41 (0.39 - 0.43) and 0.2 (0.17 - 0.23) in the later three periods with escalating containment measures, see Fig. 3B for more details. Note that all CI's without further specifications are 95% CI's throughout this paper. Compare with reproduction numbers in other published studies, our estimate in the first period is in the range but on the high side, it is possibly because most of the reproduction numbers were estimated for the period after 9 January namely after our first period, and in addition, the early data were that complete which might lead to an overestimation in reproduction number [5,6,7]. The estimated cumulative number of infections up until 8 March was 182,433 (158,964 - 208,763) by fitting data from all 5 periods, this number increased to 189,352 (164,974 - 216,793) if the trend of the fourth period was assumed, 406,004 (348,208 - 472,443) if the trend of the third period was assumed or 11,837,055 (10,996,111 - 12,643,132) if the trend of the second period was assumed. These represent a 3.7%, 55.1% and 98.46% 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 can be hence regarded as the number of infection in/from Wuhan.

Based on the same data used in Hao et al (2020), we estimated the cumulative number of laboratory-confirmed cases to be 32,755 (30,801 - 34,758) by 8 March which was close enough to the reported number of 32,583. This was equivalent to an overall ascertainment rate of 0.18 (0.16 - 0.20) given the estimated total infections of 182,433 (158,964 - 208,763). Despite the low overall ascertainment rate, the ascertainment rate in I went up to 0.822 on 8 March from 0.432 1 January, see Fig. 5 for more detail. The number of daily active infections (including presymptomatic, ascertainable and unascertainable infections, namely cases in P , I and A) peaked at 64,402 (54,900 - 75,158) on 1 February and dropped to 6,140 (5,032 - 7,402) on 8 March and 30 (15 - 48) on 14 May (Fig. 3F). Note that a 10-day city-wide screening was implemented in Wuhan from 14 May due to the new cases confirmed on 9 May after 35 consecutive days with no new confirmed cases [8]. If the trend remained unchanged as in the fifth period, the

number of ascertainable infections (I) would first become zero on 29 April (Apr 16 to May 12), and the clearance of all infections (namely $E + P + I + A = 0$) would occur on 30 Jun (7 Jun to 19 Jul) which was much later than 21 April (8 April to 12 May) estimated in Hao et al (2020). Compared with the estimates in Hao et al (2020), our estimate on A is much more heavily tailed (see Fig. 3F). Considering a few cases detected in Wuhan in mid-May, our 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) [9].

In regard to continuous surveillance and interventions, based on the modified model we found that if control measures were lifted 14 days after the first day of zero new ascertainable cases (ΔI), the probability of resurgence, defined as the number of active ascertainable cases greater than 100, could be almost as large as one, and the surge was predicted to occur on Day 20 (14 - 25) after lifting controls (Fig. 4B/C blue line). If controls were lifted after zero new ascertainable cases in a consecutive period of 14 days, the probability of resurgence was still as high as 0.90, with possible resurgence delayed to Day 37 (27 - 52) after lifting controls (Fig. 4B/C red line). This probability went down to 0.39 and 0.04 if controls were lifted after zero new ascertainable cases in a consecutive period of 30 and 60 days respectively. Compared with the results in Hao et al (2020), our estimates on the probability of resurgence are much higher which suggest continuous efforts in interventions is essential to contain the spread of the pandemic.

To conclude, our modified model improved the original model in Hao et al (2020) by (1) redefining A and I so that populations in each state were more homogenous; (2) taking the isolation measure on clinical diagnosed cases into account; and (3) correcting the unreasonable pre-determined symptomatic infectious period $D_i = 2.9$ Days in the model. We believe that the modified model provided a better prediction based on the fact that there were still active cases found in mid-May while in Hao et al (2020) the clearance of all active infections was predicted to occur on 21 April (8 April to 12 May). Our results suggest that control measures cannot be easily lifted while continuous efforts are needed to contain the spread of the pandemic; or a universal PT-PCR screening is essential to detect hidden cases before lifting control measure.

This modified model shares a same limitations pointed out in Hao et al (2020), that is the assumption of homogeneous transmission rate within the population and ignores heterogeneity between groups by sex, age, geographical region, socioeconomic status and most importantly the heterogeneity in disease courses which were also pointed out by Dr. Chaolong Wang through a direct communication. In reality, it is reasonable to believe that the transmissibility decays towards the end of infectious period, hence the assumption of a constant transmission rate b throughout compartment I might potentially lead to an overestimate of effective reproduction number R_e in the early stages. Fortunately, this limitation can be address in the following model modification. Split I into I_{early} (early stage of symptomatic

infectious period with a higher transmissibility), and I_{late} (late stage of symptomatic infectious period with a low transmissibility). The transition dynamics of these new states are as follows:

- (1) Let $I_{early} \rightarrow I_{late}$ with a rate of $1/3$ which corresponds to the setting in Hao et al (2020) and $I_{late} \rightarrow R$ with a rate of $1/7$. Thus, the expectation of the symptomatic infectious period would be $3+7=10$ days which is consistent with our choice of $D_i = 10$ days.
- (2) Patients in I_{early} had a transmission rate of b .
- (3) Patients in I_{late} had a transmission rate of βb , where $\beta \in [0,1]$ is the reduction factor of transmissibility in late stage, and is an unknown parameter to be inferred in the model.

Theoretically, this modification would grant us even better compartment homogeneity, and no additional technical difficulties were expected in inferring such a modified model with the same MCMC algorithm. Moreover, since the mean symptomatic infectious period remains the same more realistic value of 10 days, it is reasonable to expect the modified could still present the heavy tail phenomena as in the current study. However, in terms of coding, since such modification could cause substantial changes to the R-code of the original paper, we decide to first present theoretical argument here and postpone a full numerical report in the future work.

Acknowledgement

We thank authors of Hao et al (2020) and the journal for making the data and code publicly available, which made it possible for us to use and modify the code on the basis of the one in Hao et al (2020).

Author contribution statements

Chong You and Xin Gai contributed equally to the writing, model building and coding. Xiao-Hua Zhou and Yuan Zhang contributed equally to the overall design of the study and writing.

Conflict of interest statement

The Authors declare that there is no conflict of interest.

Reference

- [1] Hao, X., Cheng, S., Wu, D. et al. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature*, 584, 420–424 (2020).
- [2] Health Commission of Hubei Province. Situation report on 12 February 2020 (in Chinese). Available at http://wjw.hubei.gov.cn/bmdt/tzgg/202002/t20200213_2025582.shtml (2020).

- [3] 10,000 new "clinical diagnosed cases" confirmed in a day in Hubei (in Chinese). Available at <https://news.sina.cn/gn/2020-02-13/detail-iimxyqvz2457971.d.html> (2020).
- [4] He, X., Lau, E.H.Y., Wu, P. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26, 672–675 (2020).
- [5] Liu, Y., Gayle, A.A., Wilder-Smith, A., Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 27(2). (2020)
- [6] Sanche, S., Lin, Y., Xu, C. *et al.* High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases*, 26(7):1470-1477. (2020).
- [7] Zhao, S., Chen, H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quantitative Biology*, 8, 11–19. (2020).
- [8] Wuhan ended its 35 days of zero new cases (in Chinese). Available at <https://news.ifeng.com/c/7wNzkK9Br6a> (2020).
- [9] City-wide screening results (5 days) released in Wuhan, 58 new asymptomatic cases detected. Available at http://www.sanyarb.com.cn/content/2020-05/20/content_495360.htm (2020).
- [10] Hu, Z., Song, C., Xu, C. *et al.* Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Science China Life Sciences*. (2020).
- [11] Nishiura, H., Kobayashi, T., Suzuki, A., *et al.* Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International Journal of Infectious Disease*. (2020).
- [12] Zhang, Y., You, C., Cai, Z. *et al.* Prediction of the COVID-19 outbreak in China based on a new stochastic dynamic model. *medRxiv*. (2020).

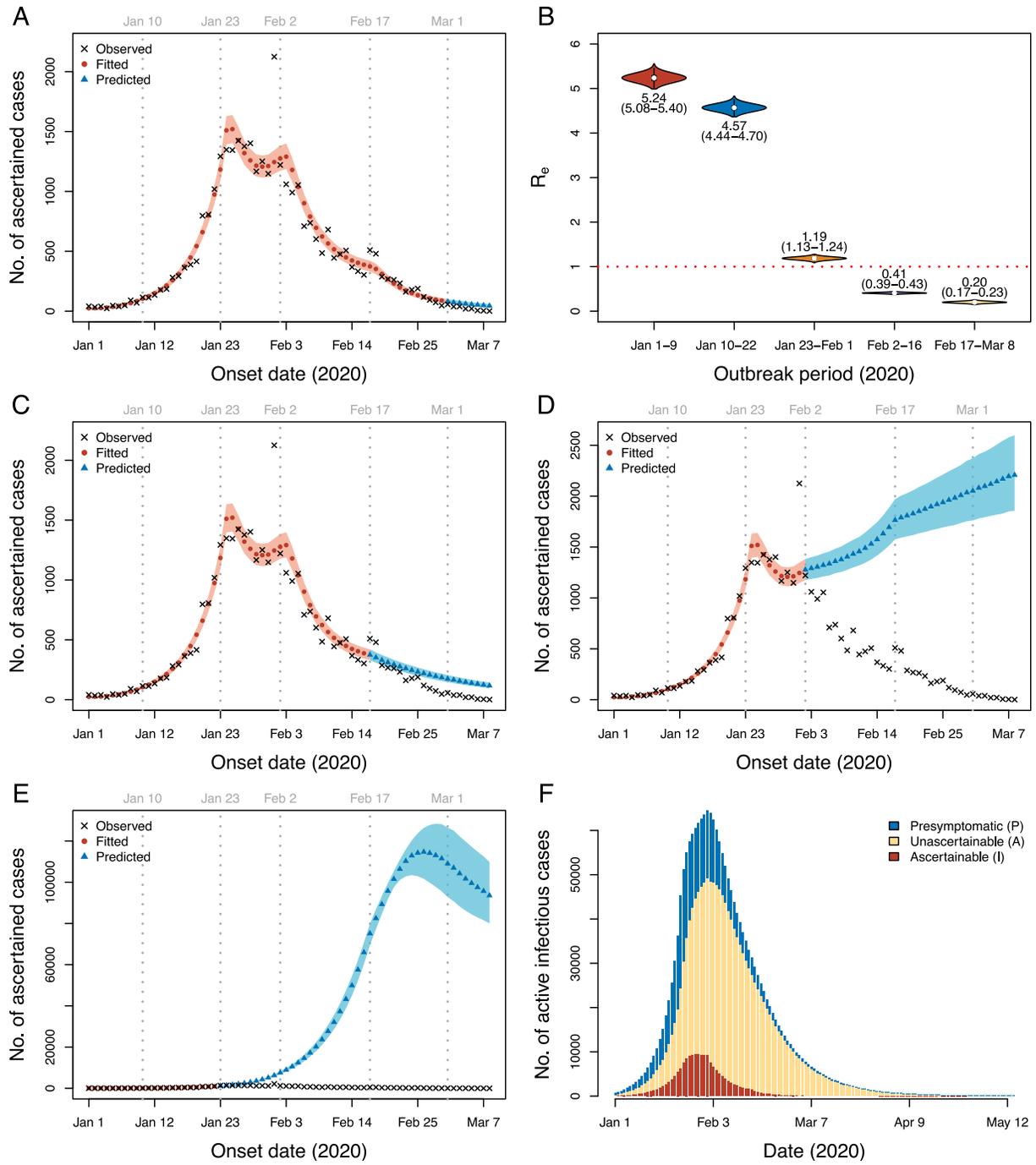


Figure 3: Modelling the COVID-19 epidemic in Wuhan. The same data from 1 January to 29 February in Hao et al (2020) were used to estimate model parameters. **A.** Fitting and prediction using parameters from the fifth period (17 February–29 February). **B.** Distribution of R_e estimates from 10,000 MCMC samples. **C.** Prediction using parameters estimated from the fourth period (2 February–16 February). **D.** Prediction using parameters estimated from the third period (23 January to 1 February). **E.** Prediction using parameters estimated from the second period (10 January to 22 January). The shaded areas in **A**, **C**–**E** are 95% credible intervals, and the colored points are the mean values based on 10,000 MCMC samples. **F.** Estimated number of active infectious cases in Wuhan.

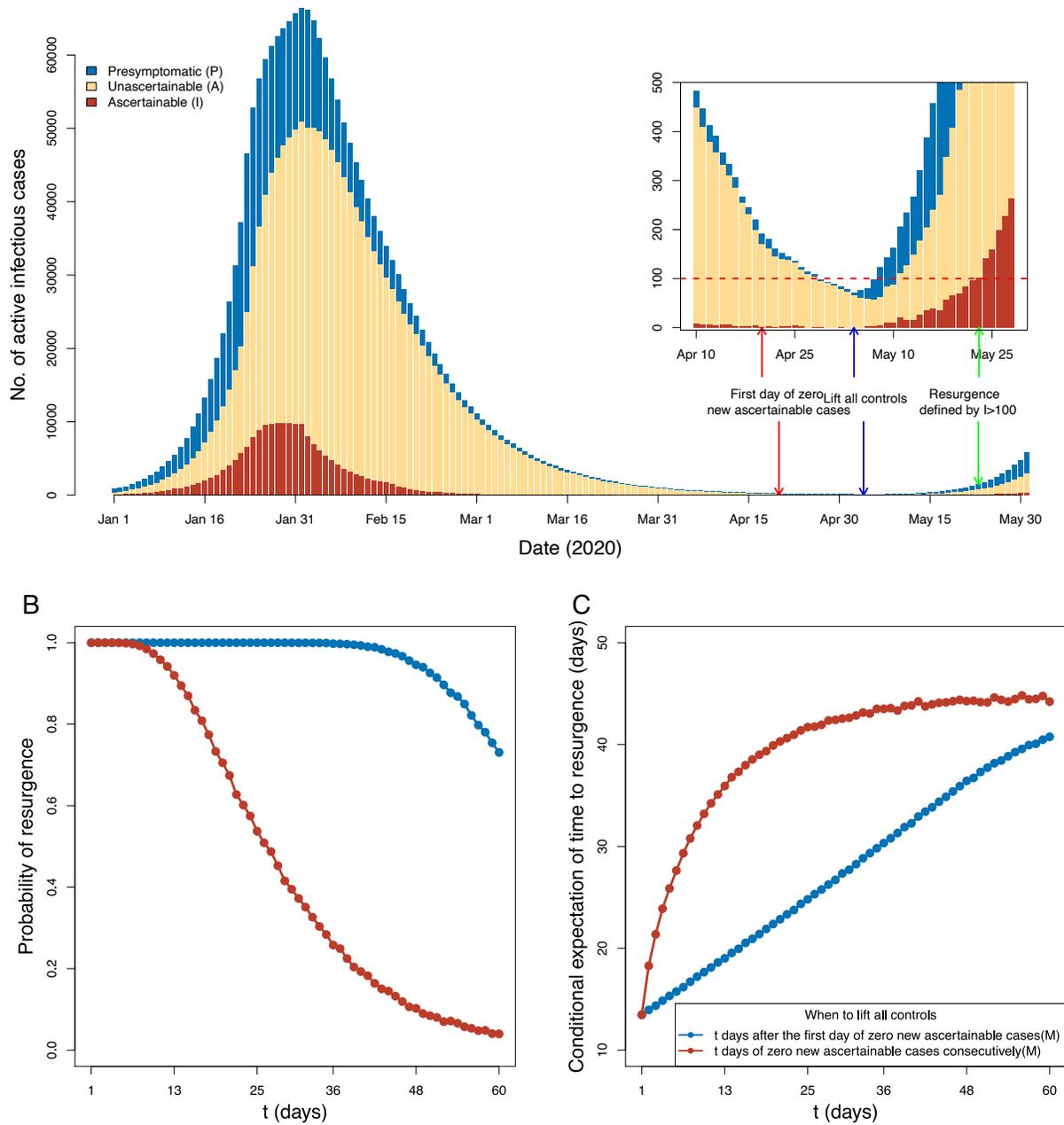


Figure 4: Risk of resurgence after lifting controls under the main model (M). The epidemic curves were simulated on the basis of 10,000 sets of parameters from MCMC, and set the transmission rate (b), ascertainable rate (ρ) and population movement (n) to their values in the first period after lifting controls as in Hao et al (2020). **A.** Illustration of a simulated curve under the main model, with control measures lifted 14 days after the first day of zero new ascertainable cases. The inset is an enlarged plot from 13 March to 28 May. **B.** Probability of resurgence if control measures were lifted t days after the first day of zero new ascertainable cases (in blue), or zero new ascertainable cases for t days consecutively (in red). **C.** Expectation of time to resurgence, conditional on the occurrence of resurgence.

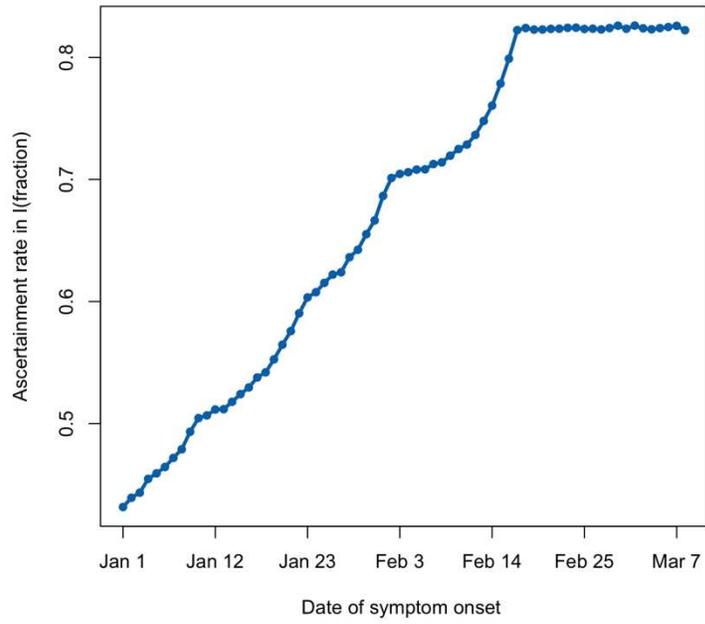


Figure 5: Ascertainment rate (fraction) in I over time. The ascertainment rate in I increased from 0.432 to 0.822 in 68 days indicating great efforts taken by the Wuhan government.

Figures

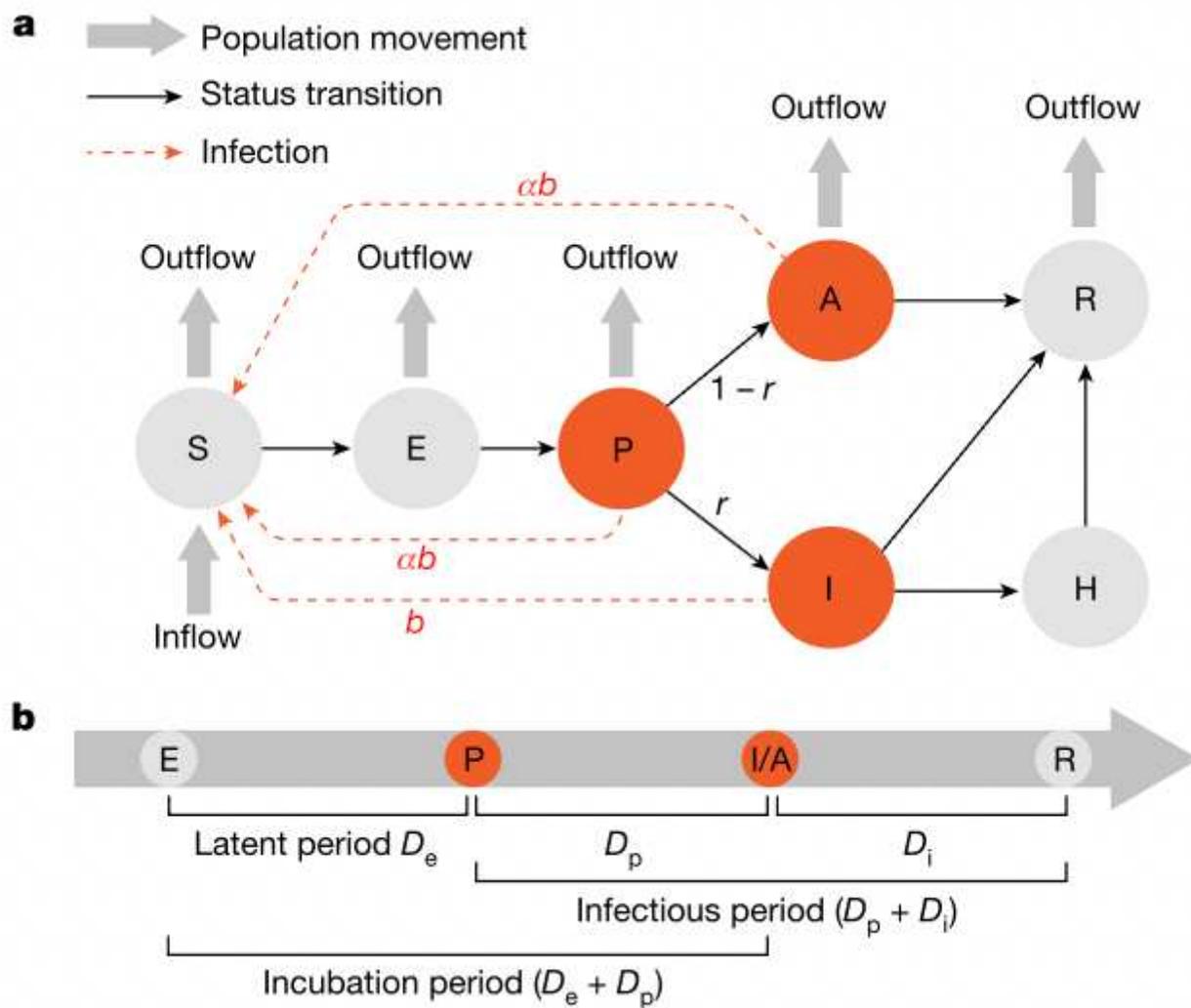


Figure 1

Illustration of the SAPHIRE model by Hao et al. (2020).

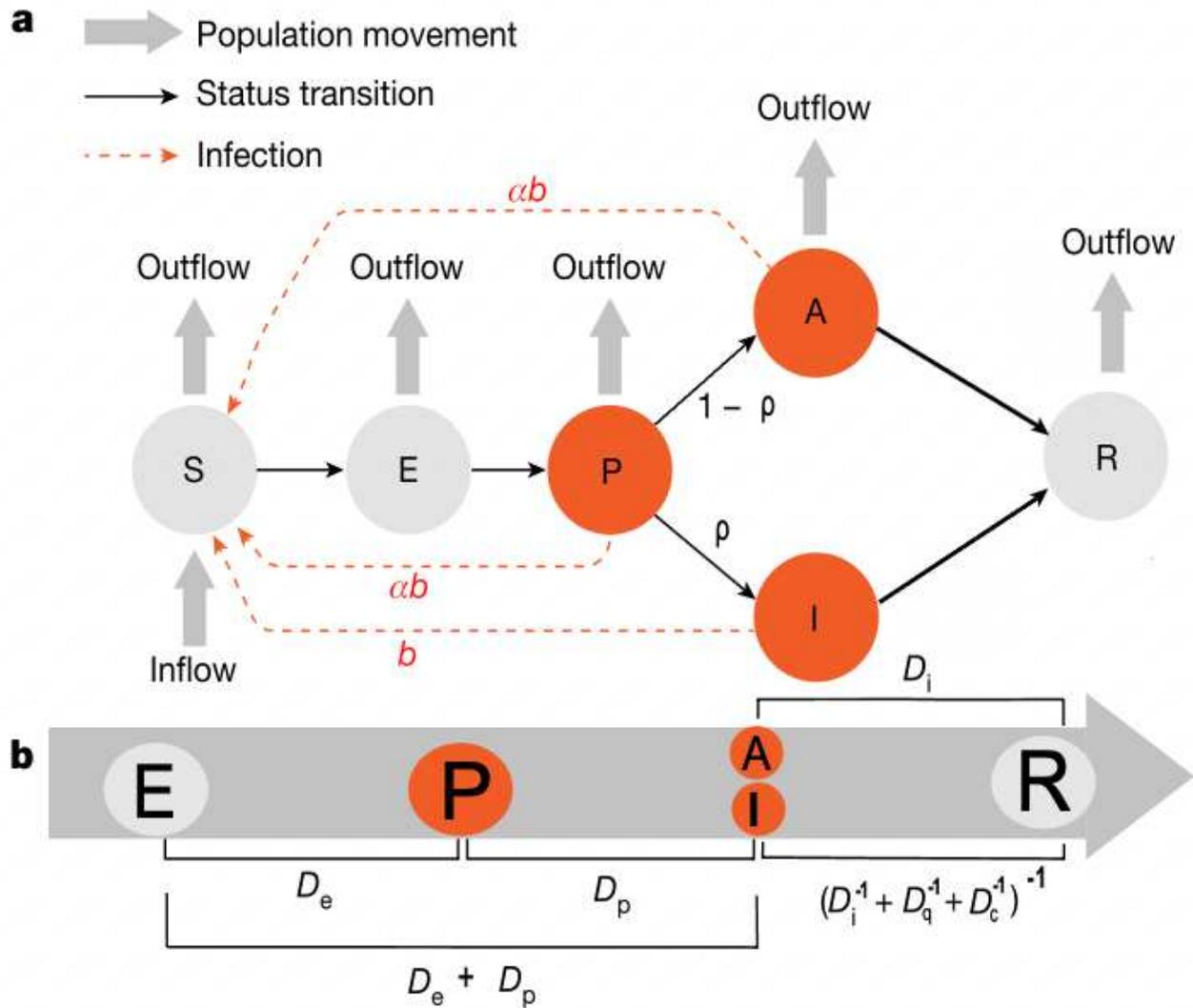


Figure 2

Illustration of the modified SAPHIRE model. The modified model includes seven compartments: susceptible (S), exposed (E), presymptomatic infectious (P), ascertainable infectious (I), unascertainable infectious (A) and removed (R). A. Relationship between different compartments. Two parameters of interest are ρ (ascertainable rate) and b (transmission rate). B, Schematic disease course of COVID-19. In this model, the unascertainable compartment A includes asymptomatic and mild cases whose symptoms were not significant enough to be detected, while the ascertainable compartment I includes symptomatic patients whose symptom were significant enough to be possibly ascertained.

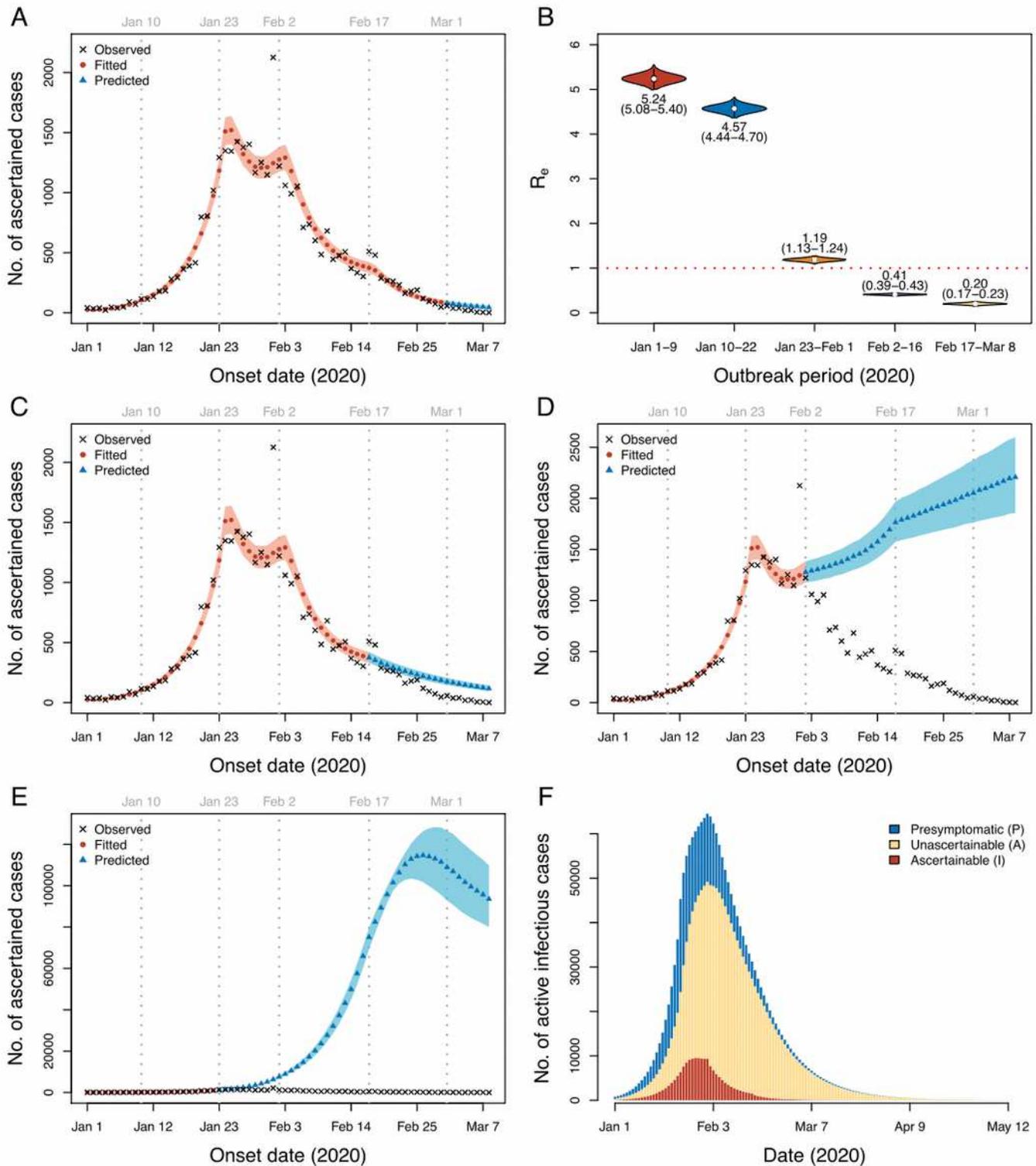


Figure 3

Modelling the COVID-19 epidemic in Wuhan. The same data from 1 January to 29 February in Hao et al (2020) were used to estimate model parameters. A. Fitting and prediction using parameters from the fifth period (17 February–29 February). B. Distribution of R_e estimates from 10,000 MCMC samples. C. Prediction using parameters estimated from the fourth period (2 February–16 February). D. Prediction using parameters estimated from the third period (23 January to 1 February). E. Prediction using

parameters estimated from the second period (10 January to 22 January). The shaded areas in A, C – E are 95% credible intervals, and the colored points are the mean values based on 10,000 MCMC samples. F. Estimated number of active infectious cases in Wuhan.

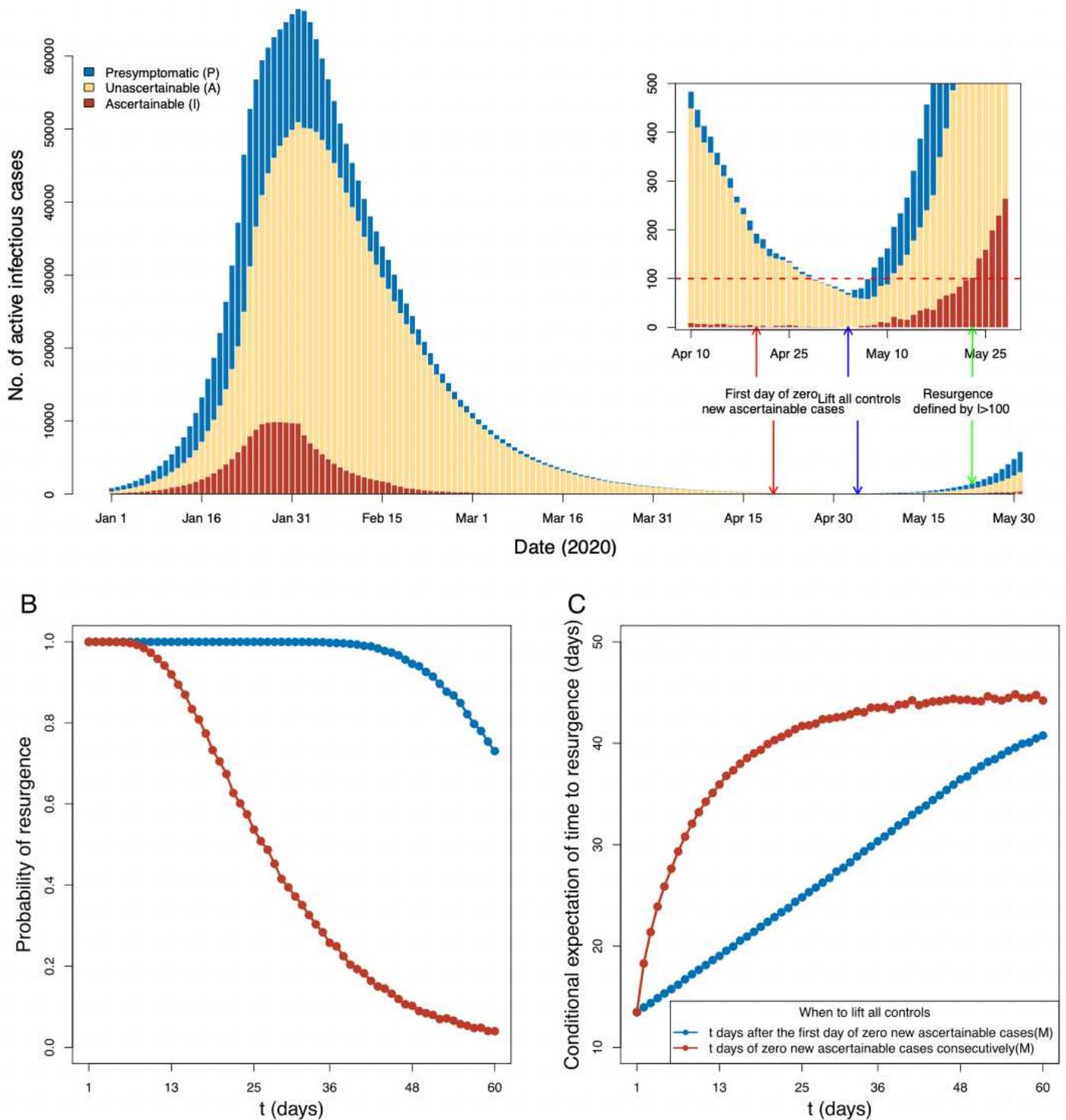


Figure 4

Risk of resurgence after lifting controls under the main model (M). The epidemic curves were simulated on the basis of 10,000 sets of parameters from MCMC, and set the transmission rate (b), ascertainable

rate (ρ) and population movement (n) to their values in the first period after lifting controls as in Hao et al (2020). A. Illustration of a simulated curve under the main model, with control measures lifted 14 days after the first day of zero new ascertainable cases. The inset is an enlarged plot from 13 March to 28 May. B. Probability of resurgence if control measures were lifted t days after the first day of zero new ascertainable cases (in blue), or zero new ascertainable cases for t days consecutively (in red). C. Expectation of time to resurgence, conditional on the occurrence of resurgence.

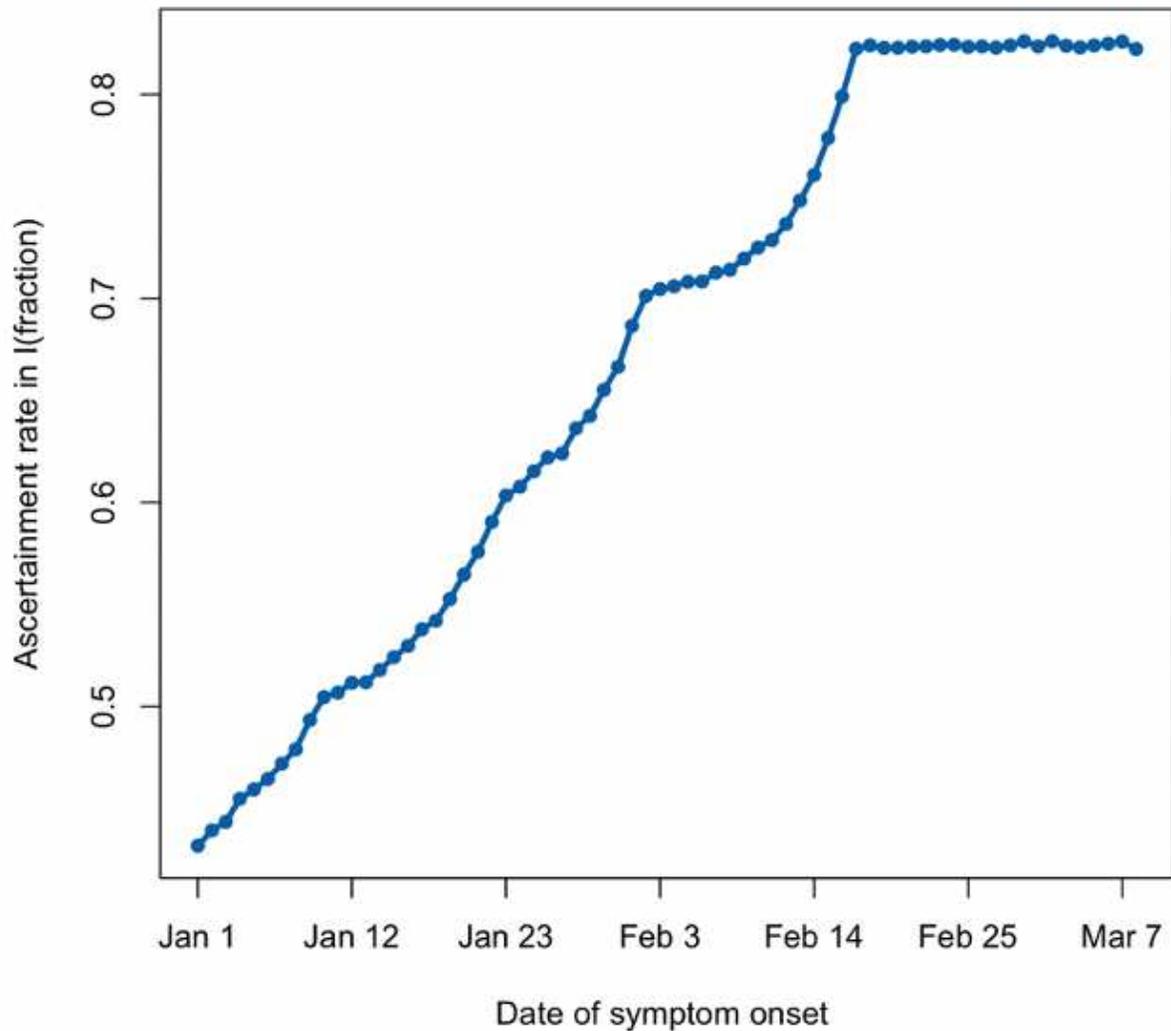


Figure 5

Ascertainment rate (fraction) in I over time. The ascertainment rate in I increased from 0.432 to 0.822 in 68 days indicating great efforts taken by the Wuhan government.

Supplementary Files

This is a list of supplementary files associated with this preprint. [Click to download.](#)

- [Appendices.docx](#)