Full-spectrum dynamics of the coronavirus disease outbreak in Wuhan, China: a modeling study of 32,583 laboratory-confirmed cases

Short title: Transmission dynamics of COVID-19 in Wuhan

Xingjie Hao[1,2,#], Shanshan Cheng[1,2,#], Degang Wu[1,2], Tangchun Wu[1,3,*], Xihong Lin[4,*], Chaolong Wang[1,2,*]

[1] Ministry of Education Key Laboratory of Environment and Health, and State Key Laboratory of Environmental Health (Incubating), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China;
[2] Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China;
[3] Department of Occupational and Environmental Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
[4] Department of Biostatistics, Harvard T.H. Chan School of Public Health, and Department of Statistics, Harvard University, Boston, MA, USA;

[#] Co-first authors.
[*] Corresponding authors.

Emails: chaolong@hust.edu.cn (C.W.); xlin@hsph.harvard.edu (X.L.); wut@tjmu.edu.cn (T.W.)

Disclaimer: *An early version of this manuscript was part of a preprint entitled "Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China", which was posted on medRxiv on March 6, 2020. (https://www.medrxiv.org/content/10.1101/2020.03.03.20030593v1)*

ABSTRACT

Vigorous non-pharmaceutical interventions have largely suppressed the COVID-19 outbreak in Wuhan, China. We developed a susceptible-exposed-infectious-recovered model to study the transmission dynamics and evaluate the impact of interventions using 32,583 laboratory-confirmed cases from December 8, 2019 till March 8, 2020, accounting for time-varying ascertainment rates, transmission rates, and population movements. The effective reproductive number R_0 dropped from 3.89 (95% credible interval: 3.79-4.00) before intervention to 0.14 (0.11-0.28) after full-scale multi-pronged interventions. By projection, the interventions reduced the total infections in Wuhan by 96.5% till March 8. Furthermore, we estimated that 79% (lower bound: 60%) of the total infections were unascertained, potentially including asymptomatic and mild-symptomatic cases. The probability of resurgence was 0.22 and 0.10 based on models with 79% and 60% infections unascertained, respectively, assuming interventions were lifted after a 14-day period of no new ascertained infections. These results provide important implications for continuing surveillance and interventions to eventually contain the outbreak.

The novel coronavirus disease (COVID-19) caused by SARS-CoV-2 was detected in Wuhan, China, in December 2019.[1] Many early cases were connected to the Huanan Seafood Market, which was shut down on January 1, 2020 to stop potential zoonotic infection. Nevertheless, the high population density of Wuhan together with the increased social activities before the Chinese New Year catalyzed the outbreak in January, 2020. The massive human movement during the holiday travel season *Chunyun*, which started on January 10, further expedited spreading of the outbreak. Shortly after the confirmation of human-to-human transmission, the Chinese authorities implemented the unprecedented city lockdown of Wuhan on January 23 to contain the geographic spread, followed by a series of non-pharmaceutical interventions to reduce virus transmission, including suspension of all intra- and inter-city transportation, cancelation of public events, social distancing, and home quarantine of mild-symptomatic patients. From February 2, strict stay-at-home policy for all residents and centralized quarantine of all patients and close contacts were implemented to stop family and community transmission. Furthermore, a city-wide door-to-door universal symptom survey was carried out during February 17-19 by designated community workers to identify previously undetected symptomatic cases. Details of the interventions were described in Pan et al.[2] These drastic interventions, together with the improved medical resources and healthcare manpower from all over the country, have effectively bent the epidemic curve and reduced the attack rate in Wuhan, shedding light on the global efforts to control the COVID-19 outbreak.[2]

Several recent studies have reported a nonnegligible proportion of asymptomatic cases with transmissibility,[3-9] which were not considered by previous epidemiological studies.[2,10] Furthermore, the number of officially reported cases was much smaller than those estimated by earlier modeling-based studies using international cases exported from Wuhan prior to the city lockdown,[11-13] implying a substantial number of unascertained cases. Using reported confirmed cases from 375 cities in China, a recent modeling study concluded that substantial unascertained cases had facilitated the rapid

spreading of COVID-19.[14] In addition, accounting for the unascertained cases has refined the estimation of case fatality risk of COVID-19, leading to a better understanding of the clinical severity of the disease.[15] Modeling both ascertained and unascertained cases can facilitate interpretation of transmission dynamics and epidemic trajectories.

Based on comprehensive epidemiological data from Wuhan,[2] we developed an extended susceptible-exposed-infectious-recovered (SEIR) model to delineate the full spectrum of COVID-19 outbreak in the epicenter, accounting for the unascertained cases, population movement, and different intervention strengths across time periods (**Fig. 1**). Specifically, we compartmentalized the population into S susceptible, E latent, I ascertained infectious, A unascertained infectious, H hospitalized, and R removed individuals. Compared with the classic SEIR model, we explicitly modeled population movement[11] and introduced two additional compartments A and H to model that only ascertained cases would seek for medical care and thus be quarantined by hospitalization. We chose to model from January 1, 2020 to avoid complications from potential zoonotic infection from the Huanan Seafood Market. We divided the outbreak into five time periods based on key events and interventions: January 1 to 9 (before *Chunyun*), January 10 to 22 (*Chunyun*), January 23 to February 1 (city lockdown), February 2 to 16 (centralized quarantine), and February 17 to March 8 (community screening). We assumed a constant population size of 10 million with equal numbers of daily inbound and outbound travelers (500,000 before *Chunyun*, 800,000 during *Chunyun*, and 0 after city lockdown).[11] Furthermore, we assumed transmission rate and ascertainment rate did not change in the first two periods, because few interventions were implemented before January 23, while they were allowed to vary in later periods to reflect different intervention strengths. We used Markov Chain Monte Carlo (MCMC) to estimate these parameters by assuming the daily incidence following a Poisson distribution, while the other parameters were set based on previous epidemiological investigations[16,17] or from our data (**Methods**). We assumed the same

transmissibility in unascertained and ascertained cases, because unascertained cases were likely asymptomatic or mild-symptomatic, who were reported to have similar viral load in upper respiratory specimens as those from symptomatic patients.[18,19]

Intuitively, due to the force of infection contributed by the unascertained cases, in the absence of interventions, the epidemic curve would grow faster than the expected curve under a model with no unascertained cases. Thus, growth rate of the epidemic curve provides information to estimate the ascertainment rate. We first simulated epidemic curves with two periods to test the performance of our method (**Methods**). As shown in **Figs. S1-S2**, our method could accurately estimate the ascertainment and transmission rates when the model was correctly specified and was relatively robust to misspecification of parameters and initial states. It is worth noting that while estimation of transmission rate was a bit sensitive to misspecification of the length of latent or infectious period, estimation of the effective reproductive number R_0 appeared to be more robust. For example, misspecification of the infectious period by 25% lower would lead to ~20% overestimation in the transmission rate but only ~10% underestimation in R_0, both for the first period (**Fig. S2**).

We then applied our model to fit the daily incidences in Wuhan from January 1 to February 29, assuming the ascertainment rate at the early outbreak was about 0.5, and used the fitted model to predict the trend from March 1 to 8. As shown in **Fig. 2A**, our model fit the observed data well, except for the outlier on February 1, which might be due to approximate-date records of many patients admitted to the mobile cabin hospitals set up after February 1. The slight overprediction for the last three days (March 6 to 8) was likely due to the delay in laboratory confirmation. The transmission rate decreased from 1.77 (95% credible intervals [CrI]: 1.73-1.81) in the first two periods to 0.57 (0.55-0.59), 0.21 (0.20-0.22), and 0.08 (0.06-0.10) in the later three periods, respectively (**Table S3**), which could be translated into R_0 of 3.93 (3.83-4.04), 3.89 (3.79-4.00), 1.23 (1.18-1.28), 0.44 (0.42-0.46) and 0.14 (0.11-0.18) for the five periods, respectively

(**Fig. 2B, Table S4**). We estimated the number of total infections, which included both ascertained and unascertained cases, till March 8 to be 163,328 (136,308-193,140) if the trend of the fourth period was assumed (**Fig. 2C**), or 514,457 (411,833-631,018) if the trend of the third period was assumed (**Fig. 2D**), or 4,514,872 (4,434,347-4,572,588) if the trend of the second period was assumed (**Fig. 2E**), in comparison to the estimated total infections of 157,894 (131,673-186,720) by fitting data from all five periods (**Fig. 2A**). These numbers were translated to 3.3%, 69.3%, and 96.5% reduction of infections due to the interventions in different periods.

Strikingly, we estimated low ascertainment rates across periods, which were 0.18 (0.16-0.20) for the first two periods, and 0.22 (0.19-0.25), 0.17 (0.15-0.19), and 0.29 (0.25-0.35) for the other three periods, respectively (**Table S3**). Even with the universal community symptom screening implemented on February 17 to 19, the ascertainment rate was only increased to 0.29. Based on the fitted model using data from January 1 to February 29, we projected the cumulative number of ascertained cases to be 32,588 (28,920-36,473) by March 8, close to the actual reported number of 32,583. This was equivalent to an overall ascertainment rate of 0.21 (0.18-0.24) given the estimated total infections of 157,894 (131,673-186,720). The model also projected that the number of daily active infections in Wuhan, including both ascertained and unascertained, peaked at 14,393 (11,735-17,284) on February 3 and dropped afterwards to 161 (109-222) on March 8 (**Fig. 2F**). If the trend remained unchanged, the number of ascertained infections would first become zero on March 24 (95% CrI: March 17 to March 30), while the clearance of all infections would occur on April 18 (April 6 to May 8), 2020 (**Table S5**).

The large fraction of unascertained cases has important implications for continuing surveillance and interventions.[20] Based on stochastic simulations, we estimated the probability of resurgence after lifting all controls, assuming the transmission rate, ascertainment rate, and daily population movement were resumed to values of the first

period (**Methods**). Because of the latent period and unascertained cases, the source of infection would not be completely cleared shortly after the first day of zero ascertained cases. We found that if control measures were lifted 14 days after the first day of zero ascertained cases, despite sparse new cases might be ascertained during the observation period, there would still be as high as 0.90 probability of resurgence, and the surge is predicted to occur on day 27 (95% CrI: 17-44) after lifting controls (**Fig. 3**). If we imposed a more stringent criterion of lifting controls after observing no ascertained cases in a consecutive period of 14 days, the probability of resurgence would drop to 0.22, with possible resurgence delayed to day 33 (95% CrI: 22-54) after lifting controls (**Fig. 3**). These results highlighted the risk of ignoring unascertained cases in switching intervention strategies, despite using an over-simplified model without considering other factors such as imported cases, changes in temperature and humidity, and a stepwise lifting strategy that is currently adopted by Wuhan and other cities in China.

We performed a series of sensitivity analyses to test the robustness of our results by considering the outlier data point on February 1, varying lengths of latent and infectious periods, lower transmissibility for unascertained cases, and different initial states (**Tables S3-S5**, **Figs. S4-S11**). Our major findings of remarkable decrease in R_0 after interventions and the existence of substantial unascertained cases was robust in all sensitivity analyses. Consistent with simulation results, the estimated ascertainment rates increased with decreasing initial numbers of latent and unascertained cases (**Figs. S9-S11**). If we assumed an extreme scenario with no unascertained cases in the early outbreak (model S8; **Fig. S11**), the estimated ascertainment rate would be 0.40 (0.35-0.45) overall and 0.55 (0.47-0.64) for the last period, which would represent an upper bound of the ascertainment rate. In this model, because of the higher ascertainment rate compared to the main analysis, we estimated a lower probability of resurgence of 0.10 when lifting controls after 14 days of no ascertained cases, and a longer time to resurgence, occurring on day 37 (95% CrI: 24-58) after lifting controls (**Fig. 3**). We also tested a simplified model assuming complete ascertainment anytime, but this simplified

model performed significantly worse than the full model, especially in fitting the rapid growth before interventions as expected (**Fig. S12**).

To evaluate our model assumptions, we compared with a published modeling study, which was based on independent data of early cases exported from Wuhan to other countries.[11] Wu et al.[11] estimated that 75,815 (37,304-130,330) individuals had been infected in Greater Wuhan as of January 25, 2020, while we estimated a total number of 40,883 (34,468-47,799) infections by the same day. The discrepancy was mainly due to different assumptions of population size, which was 19 million for the Greater Wuhan Area including surrounding cities by Wu et al.[11] versus 10 million for the Wuhan city in our analysis. After accounting for the difference in population size used, the estimates of infection prevalence were indeed highly consistent (0.41% in Wu et al. versus 0.37% in our analysis). The remarkable consistency supported the validity of key assumptions in our main analysis, including the initial states. For example, if we assumed no unascertained cases initially (model S8; **Fig. S11**), while the model still fit the epidemic curve well, the estimated total infections would be 21,393 (17,383-25,911) by January 25. This was equivalent to an infection prevalence of 0.21% (0.17-0.26%), much lower than the estimate based on Wu et al.[11]

Our finding of a large fraction of unascertained cases, despite the strong surveillance in Wuhan, indicated the existence of many asymptomatic or mild-symptomatic but infectious cases, highlighting a key challenge to the COVID-19 epidemic control.[21] There is accumulating evidence on the existence of many asymptomatic cases. For example, asymptomatic cases were estimated to account for 18% of the infections onboard the Diamond Princess Cruise ship[7] and 31% of the infected Japanese evacuated from Wuhan.[8] In addition, it was recently reported that 29 of the 33 (88%) infected pregnant women were asymptomatic by universal screening of 210 women admitted for delivery between March 22 and April 4 in New York City.[9] Several reports also highlighted the difficulty in detecting COVID-19 cases: about two thirds of the cases

exported from mainland China remained undetected worldwide,[22] and the detection capacity varied from 11% in low surveillance countries to 40% in high surveillance countries.[23,24] By modeling the epidemics in other cities, it was also estimated that the ascertainment rate of infected individuals was about 24.4% in China (excluding Hubei province)[13] and 14% in Wuhan prior to travel ban.[14] Consistent with these studies, our extensive analyses of the most comprehensive epidemic data from Wuhan also indicated an ascertainment rate between 13% and 40% (**Table S3**). The large fraction of unascertained cases would lead to about 25 days delay between the first occurrence of no ascertained cases and the clearance of all infections (**Table S5**), imposing a high risk of resurgence after lifting controls (**Fig. 3**). Therefore, understanding the proportion of unascertained cases and the asymptomatic transmissibility will be critical for prioritization of the surveillance and control measures.[20,25] Currently, Wuhan is implementing a strategy to normalize and restore societal activities gradually while maintaining strong disease surveillance. The experience and outcome of Wuhan will be valuable to other countries who will eventually face the same issue.

We noted that our R_0 estimate of 3.93 (3.83-4.04) before interventions was at the higher end of the range of R_0 (1.40-6.49 with a median of 2.79) reported by previous studies using different data sources, time periods, and statistical methods.[13,14,16,26-28] Several plausible reasons might explain the discrepancy, including potential impacts of unascertained cases, more complete case records in our analysis, and different starting date of the model. If we assumed no unascertained cases (**Fig. S12**), the estimated R_0 would be 2.89 (2.88-2.90) without interventions, aligned well with the early epidemiological investigations.[16,26] We further considered a model starting from the first COVID-19 case reported in Wuhan (**Fig. S13**), from which we estimated an R_0 of 3.58 (3.49-3.69) before January 23, 2020, similar to the value of 3.15 reported by a recent study.[27] Nevertheless, this reproductive number was still much higher than the earlier estimates and those for SARS and MERS,[29,30] featuring another challenge to control the spread of COVID-19.

Taken together, our modeling study delineated the full-spectrum dynamics of the COVID-19 outbreak in Wuhan, and highlighted two key features of the outbreak: a high proportion of asymptomatic or mild-symptomatic cases, and high transmissibility. These two features synergistically propel the global pandemic of COVID-19, imposing grand challenges to control the outbreak. Nevertheless, lessons from Wuhan have demonstrated that vigorous and multifaceted containment efforts can considerably control the size of the outbreak, as evidenced by the remarkable decrease of R_0 from 3.89 to 0.14 and an estimated 96.5% reduction of infections till March 8. These are important information for other countries combatting the outbreak.

Some limitations of our study should be noted. First, we need field investigations and serologic studies to confirm our estimate of the ascertainment rate, and the generalizability to other places is unknown. This may depend on the detection capacity in different locations.[24] Second, due to the delay in laboratory tests, we might have missed some cases and therefore underestimated the ascertainment rate, especially for the last period. Third, our model assumed a homogeneous population without considering heterogeneity by sex, age, geographic regions and socioeconomic status. Finally, we could not evaluate the impact of individual interventions by the epidemic curve from a single city, because many interventions were applied simultaneously. Future work by modeling transmission between different groups and joint analysis with data from other cities will lead to deeper insights on the effectiveness of different control strategies.[27,31]

Codes availability

R codes are available at http://chaolongwang.github.io/codes_covid19.zip.

Acknowledgements: We thank all staff at the national, provincial and municipal Center for Disease Control and Prevention for providing the data. This study was partly supported by the Fundamental Research Funds for the Central Universities (2019kfyXMBZ015), the 111 Project (X.H., S.C., D.W., C.W., T.W.). X.L. is supported by Harvard University.

Author contributions: TW, XL and CW designed the study. XH, SC, XL and CW developed statistical methods. XH, SC, and DW performed data analysis. CW wrote the first draft of the manuscript. All authors reviewed and edited the manuscript.

Competing interests: The authors declare no competing interests.

REFERENCES

1 World Health Organization. Coronavirus disease 2019 (COVID-19) situation report-76. (2020).

2 Pan, A. *et al.* Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA*, doi:10.1001/jama.2020.6130 (2020).

3 Bai, Y. *et al.* Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA*, doi:10.1001/jama.2020.2565 (2020).

4 Hoehl, S. *et al.* Evidence of SARS-CoV-2 Infection in Returning Travelers from Wuhan, China. *N Engl J Med* **382**, 1278-1280, doi:10.1056/NEJMc2001899 (2020).

5 Tong, Z. D. *et al.* Potential Presymptomatic Transmission of SARS-CoV-2, Zhejiang Province, China, 2020. *Emerg Infect Dis* **26**, doi:10.3201/eid2605.200198 (2020).

6 Hu, Z. *et al.* Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci*, doi:10.1007/s11427-020-1661-4 (2020).

7 Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill* **25**, doi:10.2807/1560-7917.ES.2020.25.10.2000180 (2020).

8 Nishiura, H. *et al.* Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int J Infect Dis*, doi:10.1016/j.ijid.2020.03.020 (2020).

9 Sutton, D., Fuchs, K., D'Alton, M. & Goffman, D. Universal Screening for SARS-CoV-2 in Women Admitted for Delivery. *N Engl J Med*, doi:10.1056/10.1056/NEJMc2009316 (2020).

10 Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi* **41**, 145-151, doi:10.3760/cma.j.issn.0254-6450.2020.02.003 (2020).

11 Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet (London, England)* **395**, 689-697, doi:10.1016/S0140-6736(20)30260-9 (2020).

12 Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30144-4 (2020).

13 Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, doi:10.1126/science.aba9757 (2020).

14 Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, doi:10.1126/science.abb3221 (2020).

15 Wu, J. T. *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine*, doi:10.1038/s41591-020-0822-7 (2020).

16 Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* **382**, 1199-1207, doi:10.1056/NEJMoa2001316 (2020).

17 Guan, W. J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*, doi:10.1056/NEJMoa2002032 (2020).

18 Zou, L. *et al.* SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients.

N Engl J Med **382**, 1177-1179, doi:10.1056/NEJMc2001737 (2020).

19 Cereda, D. *et al.* The early phase of the COVID-19 outbreak in Lombardy, Italy. *arXiv* (2020).

20 Lipsitch, M., Swerdlow, D. L. & Finelli, L. Defining the Epidemiology of Covid-19 - Studies Needed. *N Engl J Med* **382**, 1194-1196, doi:10.1056/NEJMp2002125 (2020).

21 Munster, V. J., Koopmans, M., van Doremalen, N., van Riel, D. & de Wit, E. A Novel Coronavirus Emerging in China - Key Questions for Impact Assessment. *N Engl J Med* **382**, 692-694, doi:10.1056/NEJMp2000929 (2020).

22 Bhatia, S. *et al.* Report 6: Relative sensitivity of international surveillance (2020).

23 De Salazar, P. M., Niehus, R., Taylor, A., Buckee, C. O. & Lipsitch, M. *medRxiv*, doi:10.1101/2020.02.04.20020495 (2020).

24 Niehus, R., De Salazar, P. M., Taylor, A. & Lipsitch, M. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. *medRxiv*, doi:10.1101/2020.02.13.20022707 (2020).

25 Morens, D. M., Daszak, P. & Taubenberger, J. K. Escaping Pandora's Box - Another Novel Coronavirus. *N Engl J Med* **382**, 1293-1295, doi:10.1056/NEJMp2002106 (2020).

26 Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklov, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med*, doi:10.1093/jtm/taaa021 (2020).

27 Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, doi:10.1126/science.abb6105 (2020).

28 Chen, T. M. *et al.* A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infect Dis Poverty* **9**, 24, doi:10.1186/s40249-020-00640-3 (2020).

29 Wang, Y., Wang, Y., Chen, Y. & Qin, Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *J Med Virol*, doi:10.1002/jmv.25748 (2020).

30 Lipsitch, M. *et al.* Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966-1970, doi:10.1126/science.1086616 (2003).

31 Prem, K. *et al.* The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health*, doi:10.1016/S2468-2667(20)30073-6 (2020).

32 Brooks, S. P. & Gelman, A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434-455 (1997).

33 Backer, J. A., Klinkenberg, D. & Wallinga, J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill* **25**, doi:10.2807/1560-7917.ES.2020.25.5.2000062 (2020).

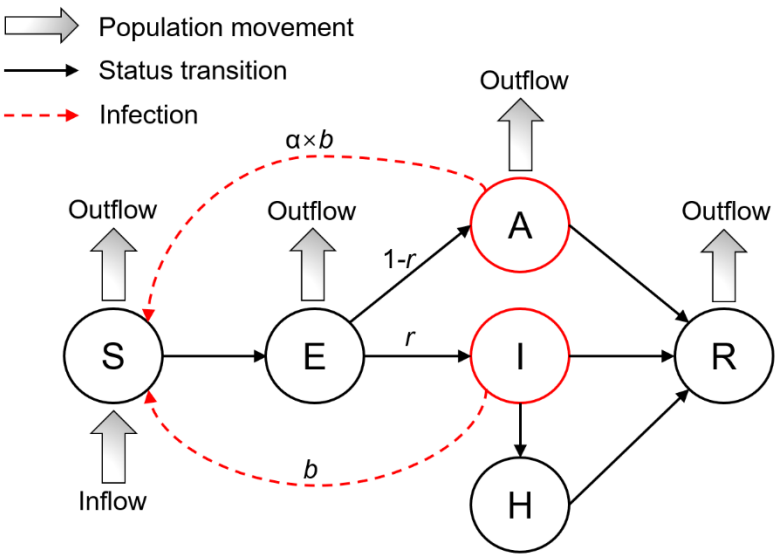

Fig. 1. Illustration of the extended SEIR model. The population in Wuhan were partitioned into six compartments: S (susceptible), E (latent), I (ascertained infections), A (unascertained infections), H (hospitalized), and R (removed). Two parameters of interests are *r* (ascertainment rate) and *b* (transmission rate), which are assumed to be varying across time periods. Details are described in the **Methods**.

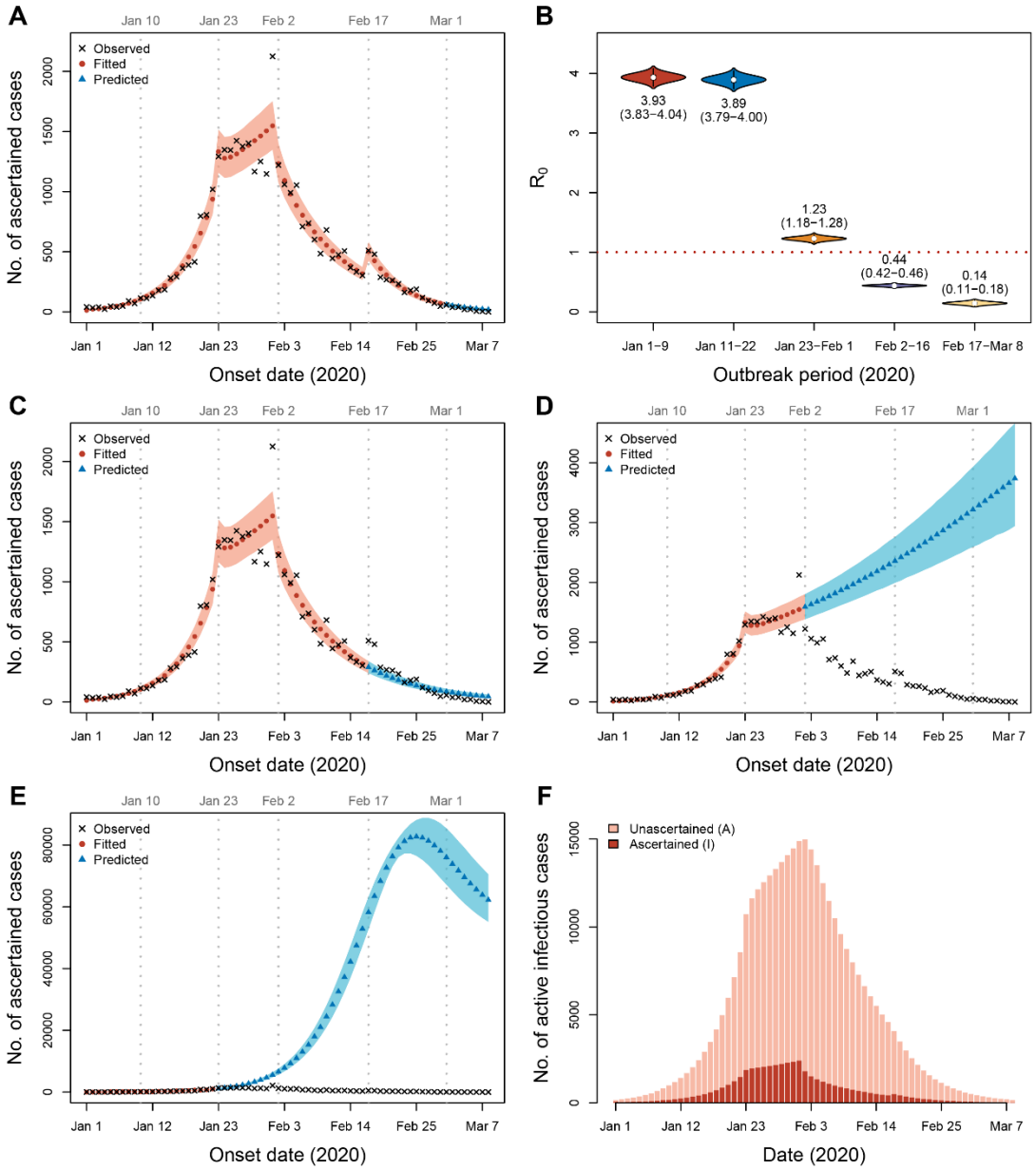

Fig. 2. Modeling the COVID-19 epidemic in Wuhan. Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

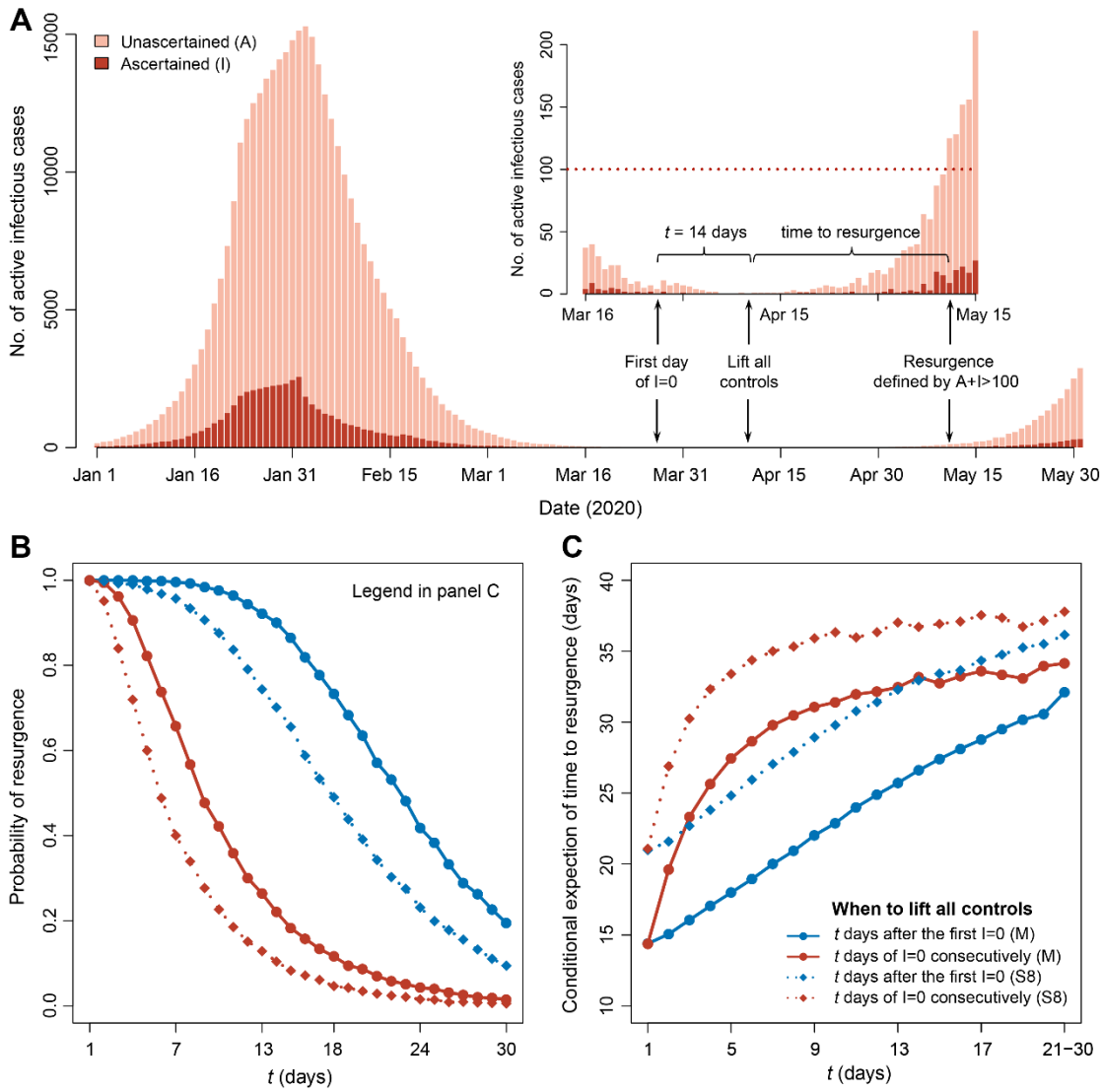

Fig. 3. Risk of resurgence after lifting controls. We considered the main model (M) and the sensitivity analysis model S8 (see **Methods**). In model M, we assumed the number of unascertained cases A was the same as the ascertained cases I during the early outbreak and thus had a 0.21 estimate of the overall ascertainment rate. In model S8, we assumed no unascertained cases during the early outbreak and thus had a 0.40 estimate of the overall ascertainment rate. For each model, we simulated epidemic curves based on 10,000 sets of parameter values from MCMC, assuming transmission rate b, ascertainment rate r, and population movement n were resumed to values before *Chunyun* after lifting controls. A resurgence was defined by when the number of active infections raised to over 100. (A) Illustration of a simulated curve under the main model with control measures lifted 14 days after the first day of no ascertained cases.

The inserted panel is a zoom-in plot from Mar 16 to May 15. (B) Probability of resurgence if control measures were lifted t days after the first observation of no ascertained cases, or after observing zero ascertained cases in a consecutive period of t days. (C) Expectation of time to resurgence conditional on the occurrence of resurgence. We grouped the last 10 days ($t = 21$ to 30) to calculate the expected time to resurgence because of their low probability of resurgence.

Methods

Data of COVID-19 cases in Wuhan

We used the same dataset as detailed in Pan et al.[2] Briefly, we extracted information of COVID-19 cases from December 8, 2019 till March 8, 2020 from the municipal Notifiable Disease Report System on March 9, 2020. We collected information of date of illness onset (the self-reported date of symptoms such as fever, cough, or other respiratory symptoms) and date of confirmed diagnosis (the laboratory confirmation date of SARS-CoV-2 in the bio-samples). For the consistency of case definition throughout the periods, we only included 32,583 laboratory-confirmed cases who were tested positive for SARS-CoV-2 by the real-time reverse-transcription-polymerase-chain-reaction (RT-PCR) assay or high-throughput sequencing of nasal and pharyngeal swab specimens.

The extended SEIR model

We extended the classic susceptible-exposed-infectious-recovered (SEIR) model to account for population movement, unascertained cases, and quarantine by hospitalization (**Fig. 1**). We chose to analyze from January 1, 2020, when the Huanan Seafood market was closed, and thus did not model the zoonotic force of infection. We assumed a constant population size $N = 10{,}000{,}000$ with equal daily inbound and outbound travelers n, where $n = 500{,}000$ for January 1-9, $800{,}000$ for January 10-22 due to *Chunyun*, and 0 after city quarantine from January 23.[11] We divided the population into S susceptible, E latent, I ascertained infectious, A unascertained infectious, H hospitalized, and R removed individuals. Dynamics of these six compartments across time t were described by the following set of ordinary differential equations:

$$\frac{\mathrm{dS}}{\mathrm{dt}} = n - \frac{bS(I+\alpha A)}{N} - \frac{nS}{N-I-H} \tag{1}$$

$$\frac{\mathrm{dE}}{\mathrm{dt}} = \frac{bS(I+\alpha A)}{N} - \frac{E}{D_e} - \frac{nE}{N-I-H} \tag{2}$$

$$\frac{\mathrm{dI}}{\mathrm{dt}} = \frac{rE}{D_e} - \frac{I}{D_q} - \frac{I}{D_i} \tag{3}$$

$$\frac{\mathrm{dA}}{\mathrm{dt}} = \frac{(1-r)E}{D_e} - \frac{A}{D_i} - \frac{nA}{N-I-H} \tag{4}$$

$$\frac{\mathrm{dH}}{\mathrm{dt}} = \frac{I}{D_q} - \frac{H}{D_h} \tag{5}$$

$$\frac{\mathrm{dR}}{\mathrm{dt}} = \frac{I+A}{D_i} + \frac{H}{D_h} - \frac{nR}{N-I-H} \tag{6}$$

where b was the transmission rate, defined as the number of individuals that an ascertained case can

infect per day; α was the ratio of the transmission rate of unascertained over ascertained cases; r was ascertainment rate; D_e and D_i were the latent and infectious periods; D_q was the duration from illness onset to hospitalization; and D_h was the hospitalization period. The effective reproductive number R_0 could be computed as

$$R_0 = (1-r)\,\alpha b D_i + rb\left(D_i^{-1} + D_q^{-1}\right)^{-1} \tag{7}$$

where D_i^{-1} and D_q^{-1} represented the rate of ascertained cases to recovery and hospitalization, respectively. Thus, $\left(D_i^{-1} + D_q^{-1}\right)^{-1}$ was the expected infectious period of ascertained cases after taking hospitalization into account.

Initial states and parameter settings

Initial states of the model and parameter settings of the main analysis were summarized in **Tables S1-S2**. We set $\alpha = 1$, assuming same transmissibility between unascertained and ascertained cases.[18,19] We set the latent period $D_e = 5.2$ days and the infectious period $D_i = 2.3$ days,[16] assuming the latent period equal to the incubation period and the infectious period equal to the serial interval minus the incubation period.[11] We set a long hospitalization period of $D_h = 30$ days. The duration from onset to hospitalization were estimated to be $D_q =$ 10.5, 7.5, 5, 3, and 1 days as half of the median difference between the onset and confirmed dates for each period, respectively. Our $D_q = 10.5$ days for the first period of January 1-9 matched well to the reported mean D_q of 9.1 days for 189 cases with onset during January 1-11.[16]

We assumed the ascertainment rate was 1/2 at the early outbreak, such that the number of initial unascertained cases $A(0)$ was the same as the number of initial ascertained cases $I(0)$, and the initial number of latent cases as twice of those ascertained cases with onset from January 1 to 5, 2020.

Estimation of parameters in the SEIR model

Considering the time-varying strength of control measures, we assumed $b = b_{12}$ and $r = r_{12}$ for the first two periods, $b = b_3$ and $r = r_3$ for period 3, $b = b_4$ and $r = r_4$ for period 4, and $b = b_5$ and $r = r_5$ for period 5. To estimate unknown parameters b_{12}, b_3, b_4, b_5 r_{12}, r_3, r_4, and r_5, we assumed the number of ascertained cases with illness onset on day d, denoted as x_d, followed a Poisson distribution with rate $\lambda_d = rE_{d-1}D_e^{-1}$, where E_{d-1} was the expected number of latent individuals on

day (d-1). We fit the observed data from January 1 to February 29 ($d = 1,2,\ldots,D$, and $D = 60$) and used the fitted model to predict the trend from March 1 to 8. Thus, the likelihood function was

$$L(b_{12}, b_3, b_4, r_{12}, r_3, r_4) = \prod_{d=1}^{D} \frac{e^{-\lambda_d} \lambda_d^{x_d}}{x_d!}. \quad (8)$$

We estimated b_{12}, b_3, b_4, b_5 r_{12}, r_3, r_4, and r_5 by Markov Chain Monte Carlo (MCMC) with Metropolis-Hastings algorithm and non-informative flat priors. We set a burn-in period of 200,000 iterations and continued to run 1,000,000 iterations with a sampling step size of 100 iterations. We repeated MCMC with three different sets of initial values and assessed the convergence by the trace plot and the multivariate Gelman-Rubin diagnostic (**Fig. S3**).[32] Estimates of parameters were presented as posterior means and 95% credible intervals (CrIs) from 10,000 MCMC samples. All the analyses were performed in R (version 3.6.2) and the Gelman-Rubin diagnostic was calculated using the *gelman.diag* function in the R package *coda*.

Stochastic simulations

We used stochastic simulations to obtain 95% CrI of fitted/predicted epidemic curve. Given a set of parameter values from MCMC, we performed the following multinomial random sampling:

$$(U_{S\to E}, U_{S\to O}, U_{S\to S}) \sim \text{Multinomial}(S_{t-1}; p_{S\to E}, p_{out}, 1 - p_{S\to E} - p_{out})$$

$$(U_{E\to I}, U_{E\to A}, U_{E\to O}, U_{E\to E}) \sim \text{Multinomial}(E_{t-1}; p_{E\to I}, p_{E\to A}, p_{out}, 1 - p_{E\to I} - p_{E\to A} - p_{out})$$

$$(U_{I\to H}, U_{I\to R}, U_{I\to I}) \sim \text{Multinomial}(I_{t-1}; p_{I\to H}, p_{I\to R}, 1 - p_{I\to H} - p_{I\to R})$$

$$(U_{A\to R}, U_{A\to O}, U_{A\to A}) \sim \text{Multinomial}(A_{t-1}; p_{A\to R}, p_{out}, 1 - p_{A\to R} - p_{out})$$

$$(U_{H\to R}, U_{H\to H}) \sim \text{Multinomial}(H_{t-1}; p_{H\to R}, 1 - p_{H\to R})$$

$$(U_{R\to O}, U_{R\to R}) \sim \text{Multinomial}(R_{t-1}; p_{out}, 1 - p_{out})$$

where $p_{S\to E} = b(I_{t-1} + \alpha A_{t-1})N^{-1}$, $p_{E\to I} = rD_e^{-1}$, $p_{E\to A} = (1-r)D_e^{-1}$, $p_{I\to H} = D_q^{-1}$, $p_{I\to R} = p_{A\to R} = D_i^{-1}$, $p_{H\to R} = D_h^{-1}$, and $p_{out} = n(N - I_{t-1} - H_{t-1})^{-1}$. The SEIR model described by Eqs. 1-6 is equivalent to the following stochastic dynamics:

$$S_t - S_{t-1} = n - U_{S\to E} - U_{S\to O} \quad (9)$$

$$E_t - E_{t-1} = U_{S\to E} - U_{E\to I} - U_{E\to A} - U_{E\to O} \quad (10)$$

$$I_t - I_{t-1} = U_{E\to I} - U_{I\to H} - U_{I\to R} \quad (11)$$

$$A_t - A_{t-1} = U_{E\to A} - U_{A\to R} - U_{A\to O} \quad (12)$$

$$H_t - H_{t-1} = U_{I\to H} - U_{H\to R} \quad (13)$$

$$R_t - R_{t-1} = U_{A\to R} + U_{I\to R} + U_{H\to R} - U_{R\to O}. \quad (14)$$

We repeated the stochastic simulations for all 10,000 sets of parameter values sampled by MCMC to construct the 95% CrI of the epidemic curve by the 2.5 and 97.5 percentiles at each time point.

Prediction of epidemic ending date and the risk of resurgence

Using the stochastic simulations described above, we predicted the first day of no new ascertained cases and the date of clearance of all active infections in Wuhan, assuming continuation of the same control measures as the last period (*i.e.*, same parameter values).

We also evaluated the risk of outbreak resurgence after lifting control measures. We considered lifting all controls (1) at t days after the first day of zero ascertained cases, or (2) after a consecutive period of t days with no ascertained infectious cases. After lifting controls, we set the transmission rate b, ascertainment rate r, and population movement n to be the same as the first period, and continued the stochastic simulation to stationary state. Time to resurgence was defined as the number of days from lifting controls to when the number of active infectious cases reached 100. We performed 10,000 simulations with 10,000 sets of parameter values sampled from MCMC (as described above). We calculated the probability of resurgence as the proportion of simulations in which a resurgence occurred, as well as the time to resurgence conditional on the occurrence of resurgence.

Simulation study for method validation

To validate the performance of our method in estimating parameters b and r, we performed two-period stochastic simulations (Eqs. 9-14) with transmission rate $b = 1.6$, ascertainment rate $r = 0.2$, daily population movement $n = 500{,}000$, and duration from illness onset to hospitalization $D_q = 10$ for the first period, and $b = 0.5$, $r = 0.2$, $n = 0$, and $D_q = 1$ for the second period. Lengths of both periods were set to 15 days, while the initial states and other parameters were set as the those in our main analysis (**Table S1**). We repeated stochastic simulations 25 times to generate 25 datasets. For each dataset, we applied our MCMC method to estimate b_1, b_2, r_1 and r_2, while setting all other parameters and initial values the same as the true values. We also tested the robustness to misspecification of the latent period D_e, the infectious period D_i, the initial number of latent cases $E(0)$, and the initial number of unascertained cases $A(0)$. In each test, we changed the specified value of a parameter (or initial state)

to be 25% lower or higher than its true value, while keeping all other parameters and initial states unchanged. We did not test the robustness to misspecification of other parameters or initial states because they either had little impacts (e.g., hospitalization period D_h) or were known with little uncertainty in the real data setting.

For the simulated datasets, we ran the MCMC method with 100,000 burn-in iterations and sampled parameter values from additional 100,000 iterations with a step size of 100 iterations. We took the mean across 1,000 MCMC samples as the final estimates and displayed results for 25 repeated simulations using boxplots.

Sensitivity analyses for the real data

We designed nine sensitivity analyses to test the robustness of our real data results. For each of the sensitivity analyses, we fixed parameters and initial states to be the same as the main analysis except for those mentioned below.

(S1) Adjust the reported incidences from January 29 to February 1 to their average. We suspected the spike of incidences on February 1 might be caused by approximate-date records among some patients admitted to the centralized quarantine after February 2. The actual illness onset dates for these patients were likely to be between January 29 and February 1.

(S2) Decrease the latent period D_e to be 4.0 days,[17] and accordingly adjust the initial number of latent cases $E(0)$ to be 276 as twice the number of reported incidences from January 1 to 4.

(S3) Increase the latent period D_e to 6.4 days [33], and accordingly adjust the initial number of latent cases $E(0)$ to be 450 as twice the number of reported incidences from January 1 to 6.

(S4) Double the infectious period D_i to 4.6 days.

(S5) Assume the unascertained cases have half transmissibility of the ascertained cases ($\alpha = 0.5$).

(S6) Assume the early ascertainment rate is 1/3 by setting $A(0) = 176$ and $E(0) = 555$.

(S7) Assume the early ascertainment rate is 2/3 by setting $A(0) = 44$ and $E(0) = 278$.

(S8) Assume the early ascertainment rate is 1 by setting $A(0) = 0$ and $E(0) = 185$.

(S9) Assume complete ascertainment across time by fixing $r_{12} = r_3 = r_4 = r_5 = 1$. We tested if the full model was significantly better than this simplified model using likelihood ratio test.

Supplementary Data

Table S1. Initial state of the SEIR model on December 31, 2019.

Variable	Meaning	Value	Note
$S(0)$	Number of susceptible individuals	9,999,427	$S = N - E - I - A - H - R$
$E(0)$	Number of latent cases	370	Twice of the number of cases with onset from Jan 1 to 5, 2020
$I(0)$	Number of ascertained infectious cases	88	Number of cases with onset before Jan 1, 2020 minus $H(0)$
$A(0)$	Number of unascertained infectious cases	88	Assuming $A(0) = I(0)$
$H(0)$	Number of hospitalized cases	27	Number of cases ascertained by Dec 31, 2019
$R(0)$	Number of removed individuals	0	Number of recovered patients by Dec 31, 2019

Table S2. Parameter settings for five periods in the main analysis.

Parameter	Meaning	Jan 1-9	Jan 10-22	Jan 23-Feb 1	Feb 2-16	Feb 17-Mar 8
b	Transmission rate of ascertained cases	b_{12}	b_{12}	b_3	b_4	b_5
r	Ascertainment rate	r_{12}	r_{12}	r_3	r_4	r_5
α	Ratio of transmission rate for unascertained cases compared to ascertained cases	1	1	1	1	1
D_e	Latent period	5.2	5.2	5.2	5.2	5.2
D_i	Infectious period	2.3	2.3	2.3	2.3	2.3
D_q	Duration from illness onset to hospitalization	10.5	7.5	5	3	1
D_h	Hospitalization period	30	30	30	30	30
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

Table S3. Estimated parameter values from the main and sensitivity analyses.

Analysis	b_{12}	b_3	b_4	b_5	r_{12}	r_3	r_4	r_5
Main	1.77 (1.73-1.81)	0.57 (0.55-0.59)	0.21 (0.20-0.22)	0.08 (0.06-0.10)	0.18 (0.16-0.20)	0.22 (0.19-0.25)	0.17 (0.15-0.19)	0.29 (0.25-0.35)
S1	1.77 (1.72-1.81)	0.53 (0.51-0.55)	0.21 (0.20-0.22)	0.08 (0.06-0.10)	0.18 (0.16-0.20)	0.23 (0.20-0.26)	0.19 (0.16-0.22)	0.33 (0.28-0.40)
S2	1.53 (1.49-1.56)	0.56 (0.54-0.58)	0.25 (0.24-0.26)	0.15 (0.13-0.16)	0.18 (0.16-0.20)	0.22 (0.19-0.25)	0.17 (0.15-0.20)	0.29 (0.25-0.35)
S3	2.02 (1.97-2.06)	0.59 (0.56-0.61)	0.16 (0.15-0.18)	0.01 (0.00-0.03)	0.18 (0.17-0.20)	0.22 (0.19-0.24)	0.17 (0.14-0.19)	0.29 (0.24-0.34)
S4	1.24 (1.21-1.28)	0.35 (0.34-0.37)	0.10 (0.09-0.11)	0.03 (0.02-0.04)	0.21 (0.19-0.23)	0.26 (0.23-0.29)	0.20 (0.17-0.23)	0.33 (0.28-0.39)
S5	3.07 (2.95-3.19)	1.00 (0.96-1.04)	0.38 (0.36-0.40)	0.14 (0.11-0.18)	0.16 (0.14-0.18)	0.19 (0.16-0.22)	0.14 (0.12-0.17)	0.25 (0.21-0.30)
S6	1.77 (1.72-1.81)	0.57 (0.54-0.59)	0.20 (0.19-0.21)	0.07 (0.06-0.09)	0.12 (0.11-0.13)	0.14 (0.12-0.16)	0.11 (0.10-0.13)	0.19 (0.16-0.23)
S7	1.78 (1.74-1.82)	0.58 (0.56-0.61)	0.21 (0.20-0.22)	0.08 (0.06-0.11)	0.24 (0.22-0.27)	0.29 (0.25-0.32)	0.22 (0.19-0.25)	0.38 (0.32-0.44)
S8	1.80 (1.76-1.84)	0.61 (0.59-0.63)	0.22 (0.21-0.24)	0.10 (0.08-0.13)	0.36 (0.32-0.39)	0.42 (0.36-0.47)	0.32 (0.28-0.37)	0.55 (0.47-0.64)
S9	1.53 (1.53-1.54)	0.74 (0.73-0.76)	0.34 (0.33-0.36)	0.65 (0.60-0.70)	-	-	-	-

The estimates were displayed as mean (95% CrI) based on 10,000 MCMC samples.

Table S4. Estimated R_0 for different periods from the main and sensitivity analyses.

Analysis	Jan 1-9	Jan 10-22	Jan 23-Feb 1	Feb 2-16	Feb 17-Mar 8
Main	3.93 (3.83-4.04)	3.89 (3.79-4.00)	1.23 (1.18-1.28)	0.44 (0.42-0.46)	0.14 (0.11-0.18)
S1	3.93 (3.82-4.04)	3.89 (3.77-4.00)	1.14 (1.09-1.18)	0.44 (0.42-0.46)	0.14 (0.11-0.18)
S2	3.40 (3.31-3.49)	3.37 (3.28-3.46)	1.20 (1.16-1.24)	0.53 (0.51-0.55)	0.27 (0.24-0.30)
S3	4.48 (4.37-4.61)	4.44 (4.32-4.56)	1.26 (1.21-1.32)	0.35 (0.32-0.38)	0.03 (0.00-0.06)
S4	5.35 (5.18-5.53)	5.26 (5.08-5.44)	1.43 (1.36-1.49)	0.40 (0.38-0.43)	0.10 (0.07-0.12)
S5	3.89 (3.78-4.00)	3.83 (3.71-3.94)	1.23 (1.18-1.28)	0.44 (0.42-0.47)	0.15 (0.11-0.18)
S6	3.97 (3.87-4.08)	3.94 (3.84-4.05)	1.24 (1.19-1.29)	0.44 (0.42-0.47)	0.15 (0.11-0.18)
S7	3.91 (3.81-4.02)	3.86 (3.75-3.97)	1.22 (1.18-1.27)	0.44 (0.42-0.46)	0.14 (0.11-0.18)
S8	3.88 (3.77-3.99)	3.79 (3.69-3.91)	1.21 (1.17-1.26)	0.45 (0.42-0.47)	0.14 (0.11-0.18)
S9	2.89 (2.88-2.90)	2.70 (2.69-2.71)	1.17 (1.15-1.20)	0.45 (0.43-0.46)	0.45 (0.42-0.49)

The estimates were displayed as mean (95% CrI) based on 10,000 MCMC samples.

Table S5. Prediction of the ending date of COVID-19 epidemic in Wuhan from the main and sensitivity analyses.

Analysis	First day of no ascertained infections	Clearance of all infections
Main	Mar 24 (Mar 17 to Mar 30)	Apr 18 (Apr 6 to May 8)
S1	Mar 24 (Mar 18 to Mar 30)	Apr 17 (Apr 5 to May 7)
S2	Mar 24 (Mar 18 to Mar 31)	Apr 17 (Apr 5 to May 7)
S3	Mar 24 (Mar 18 to Mar 30)	Apr 20 (Apr 7 to May 10)
S4	Mar 25 (Mar 19 to Apr 1)	Apr 24 (Apr 11 to May 15)
S5	Mar 24 (Mar 18 to Mar 30)	Apr 20 (Apr 7 to May 9)
S6	Mar 24 (Mar 17 to Mar 30)	Apr 21 (Apr 9 to May 10)
S7	Mar 24 (Mar 17 to Mar 30)	Apr 16 (Apr 4 to May 7)
S8	Mar 24 (Mar 18 to Mar 30)	Apr 14 (Apr 2 to May 3)
S9	May 10 (Apr 21 to Jun 1)	Jun 19 (May 18 to Jul 19)

The estimates were displayed as mean date (95% CrI) based on 10,000 stochastic simulations with parameter values from MCMC sampling.
First day of no ascertained infections means the first day of $I = 0$.
Clearance of all infections means the first day of $E = A = I = 0$.

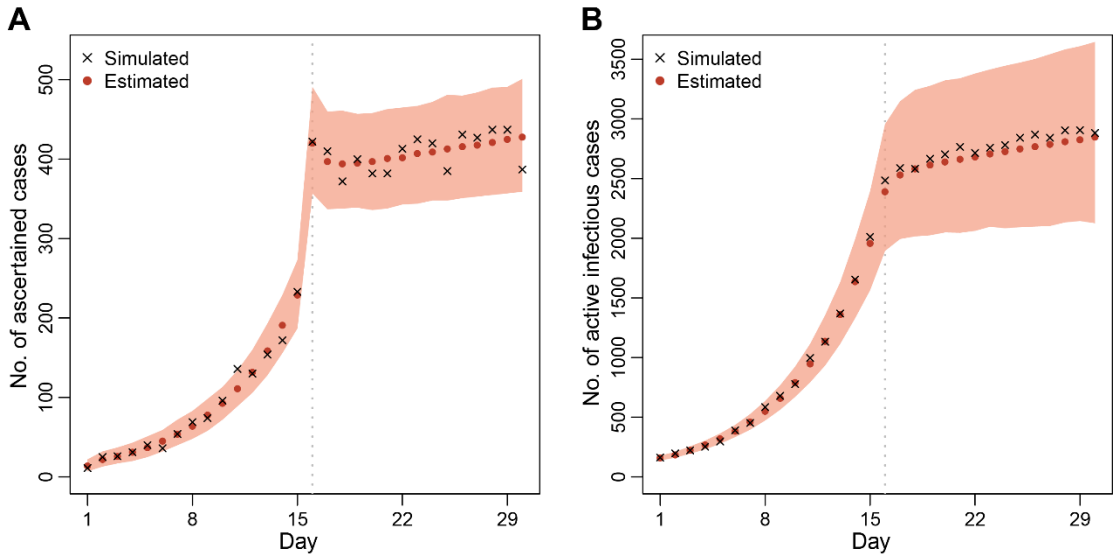

Fig. S1. Simulated data with two periods. We estimated b_1, b_2, r_1, and r_2 when the other parameters and initial states were specified to their true values. (A) Daily incidences. (B) Number of active infectious cases per day, including both ascertained and unascertained cases. The shaded areas indicate 95% CrIs of the estimated values.

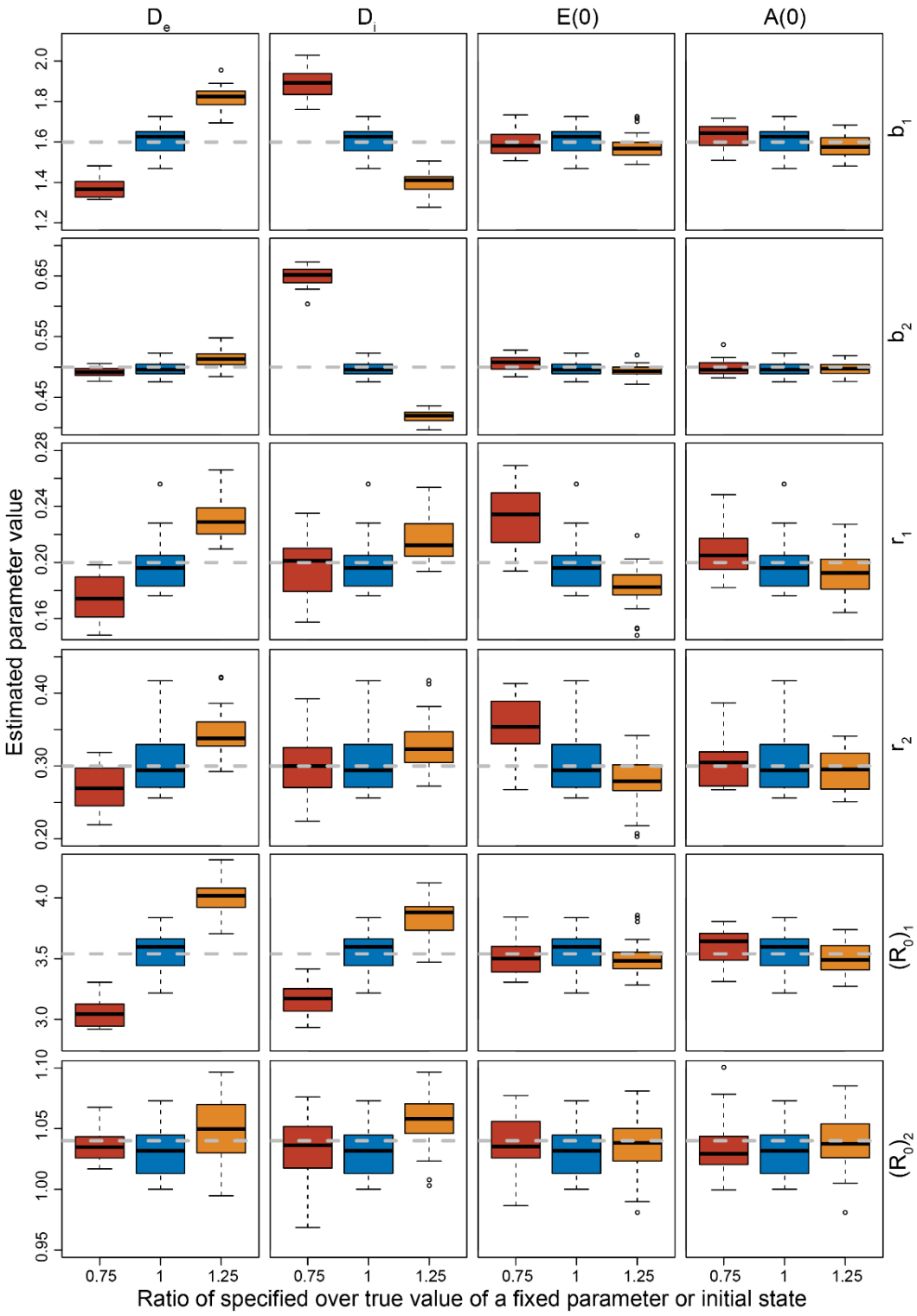

Fig. S2. Parameter estimation on simulated epidemic curves with two periods. Each row represents an estimated parameter as indicated on the right, including b_1, b_2, r_1, r_2, and R_0 for each period. The grey dashed line in each row represents the true value of the parameter to be estimated. Each column represents a specified parameter or initial state as indicated on the top, including D_e, D_i, $E(0)$, $A(0)$, which we specified by the true values or 25% lower or higher than the true values. Each box represents estimates from 25 replicates.

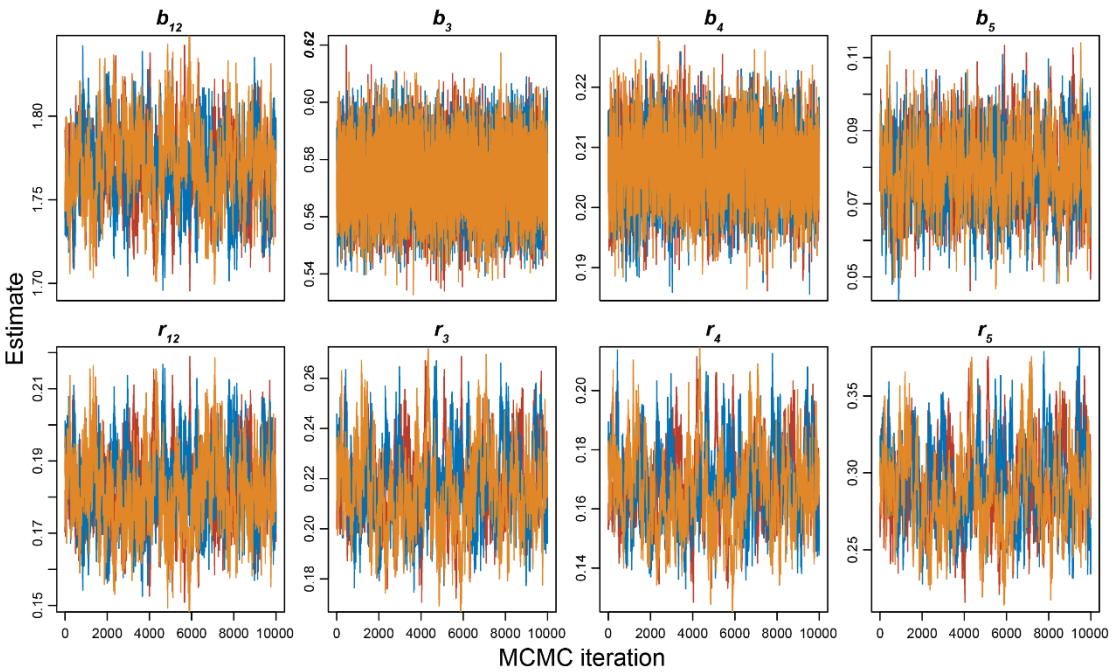

Fig. S3. Trace plots of MCMC for the main analysis of real data. Each panel represents the trajectory of 10,000 sampled values for a parameter indicated on the top of the panel. We generated three Markov chains with different initial values, which were colored by orange, red, and blue. The Gelman-Rubin diagnostic was 1.02, indicating convergence of MCMC.

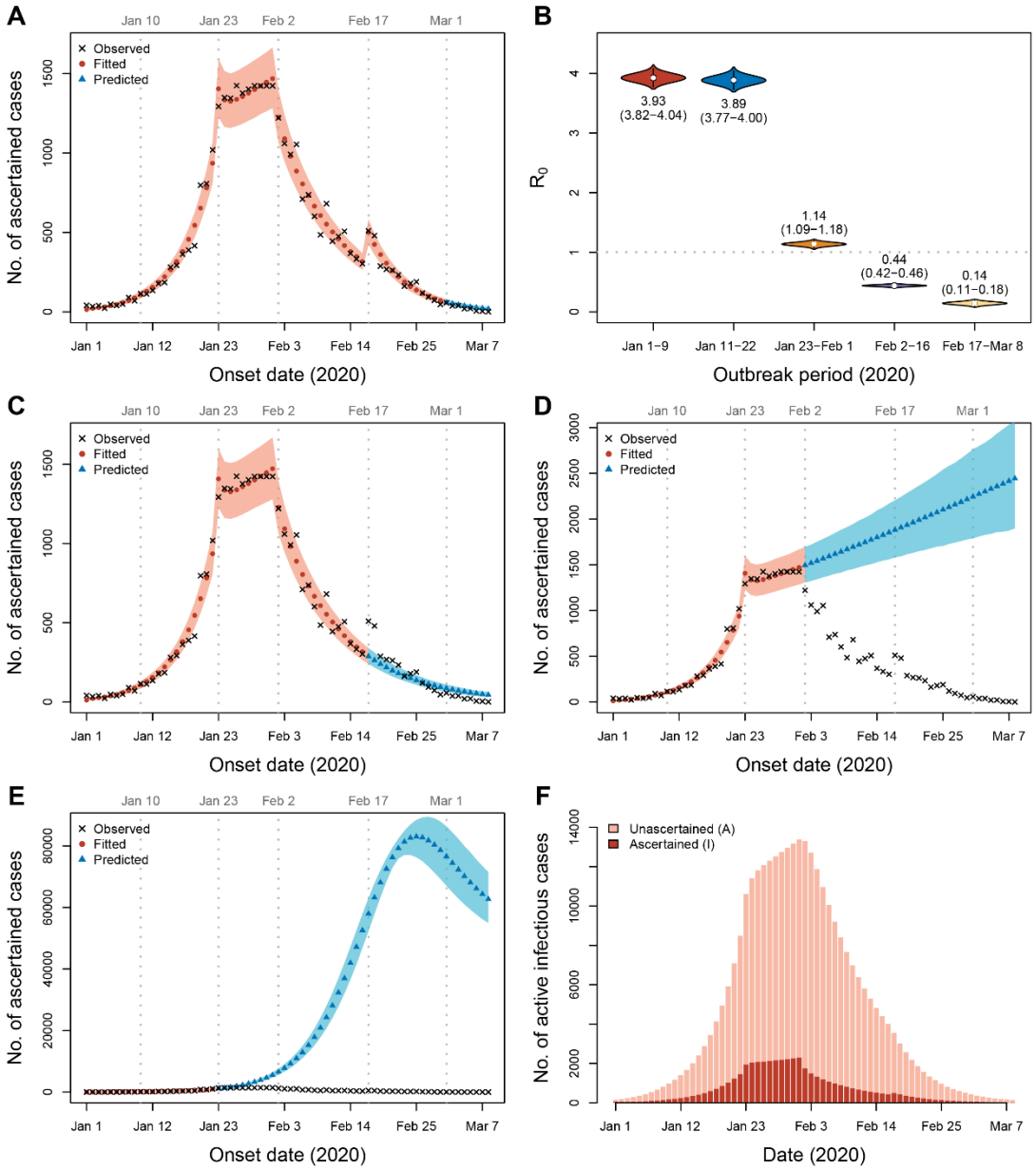

Fig. S4. Sensitivity analysis by adjusting the daily incidences from January 29 to February 1 to their average (sensitivity analysis S1). Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

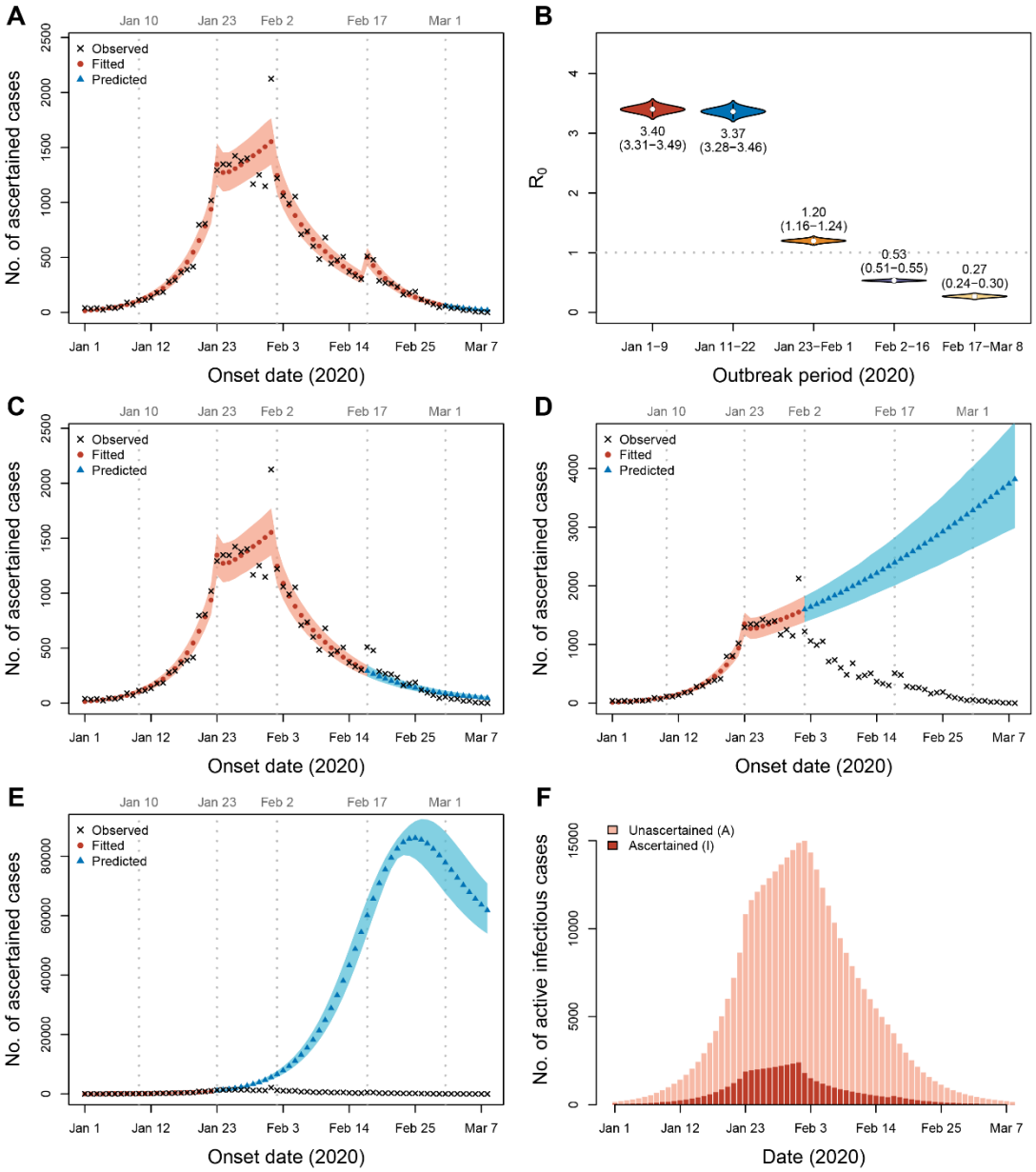

Fig. S5. Sensitivity analysis assuming a latent period of 4 days (sensitivity analysis S2). Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

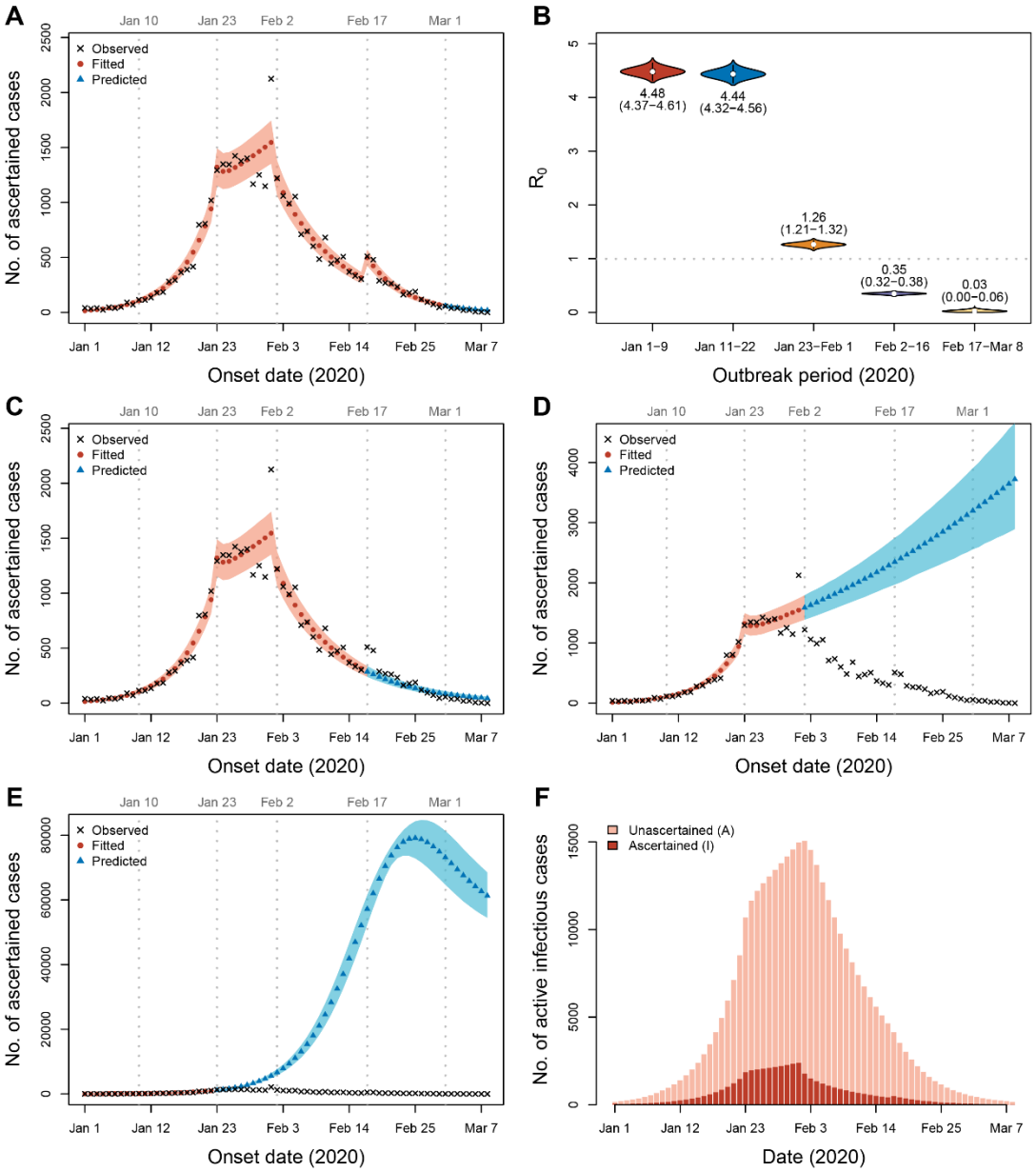

Fig. S6. Sensitivity analysis assuming a latent period of 6.4 days (sensitivity analysis S3). Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

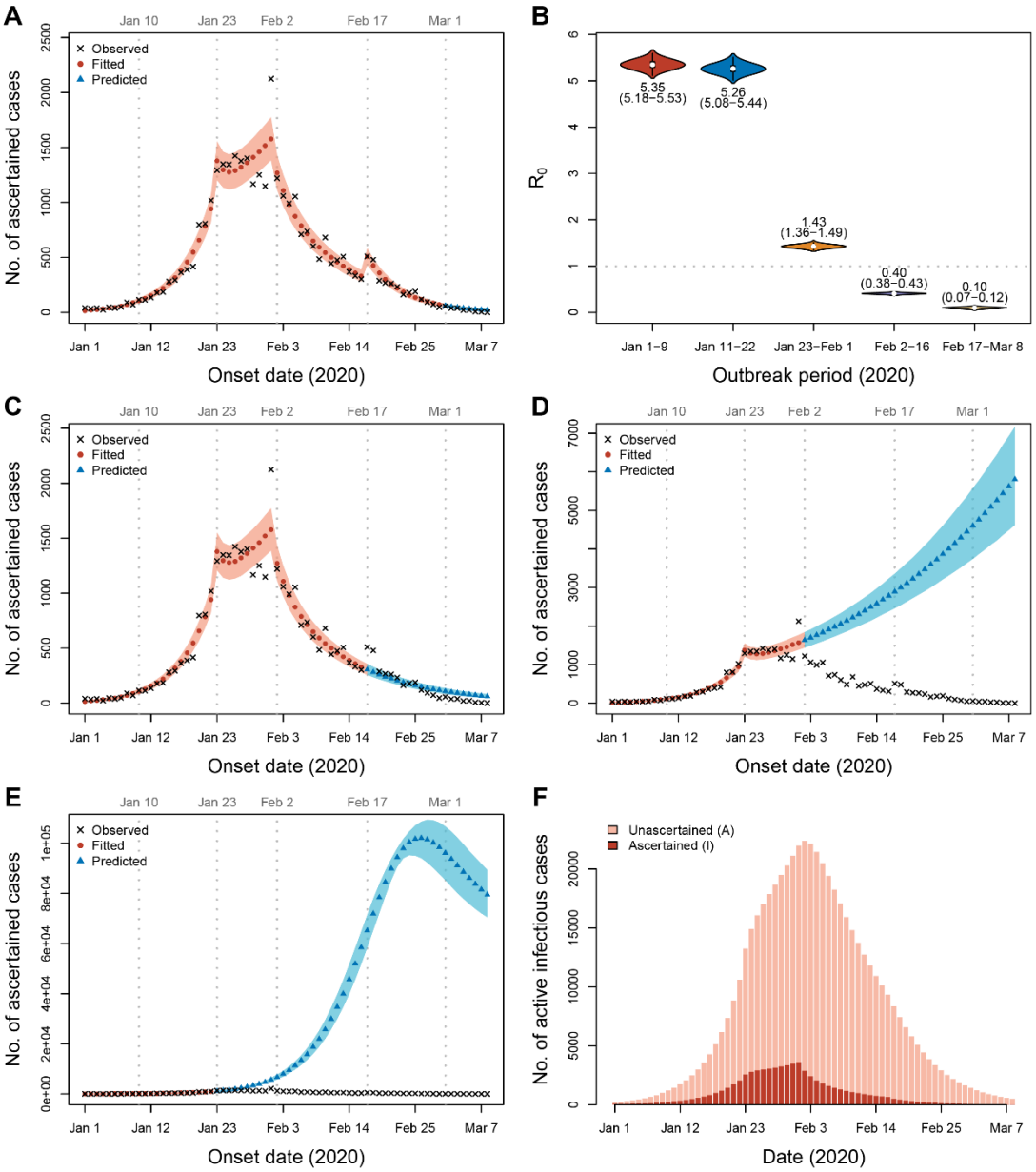

Fig. S7. Sensitivity analysis assuming an infectious period of 4.6 days (sensitivity analysis S4). Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

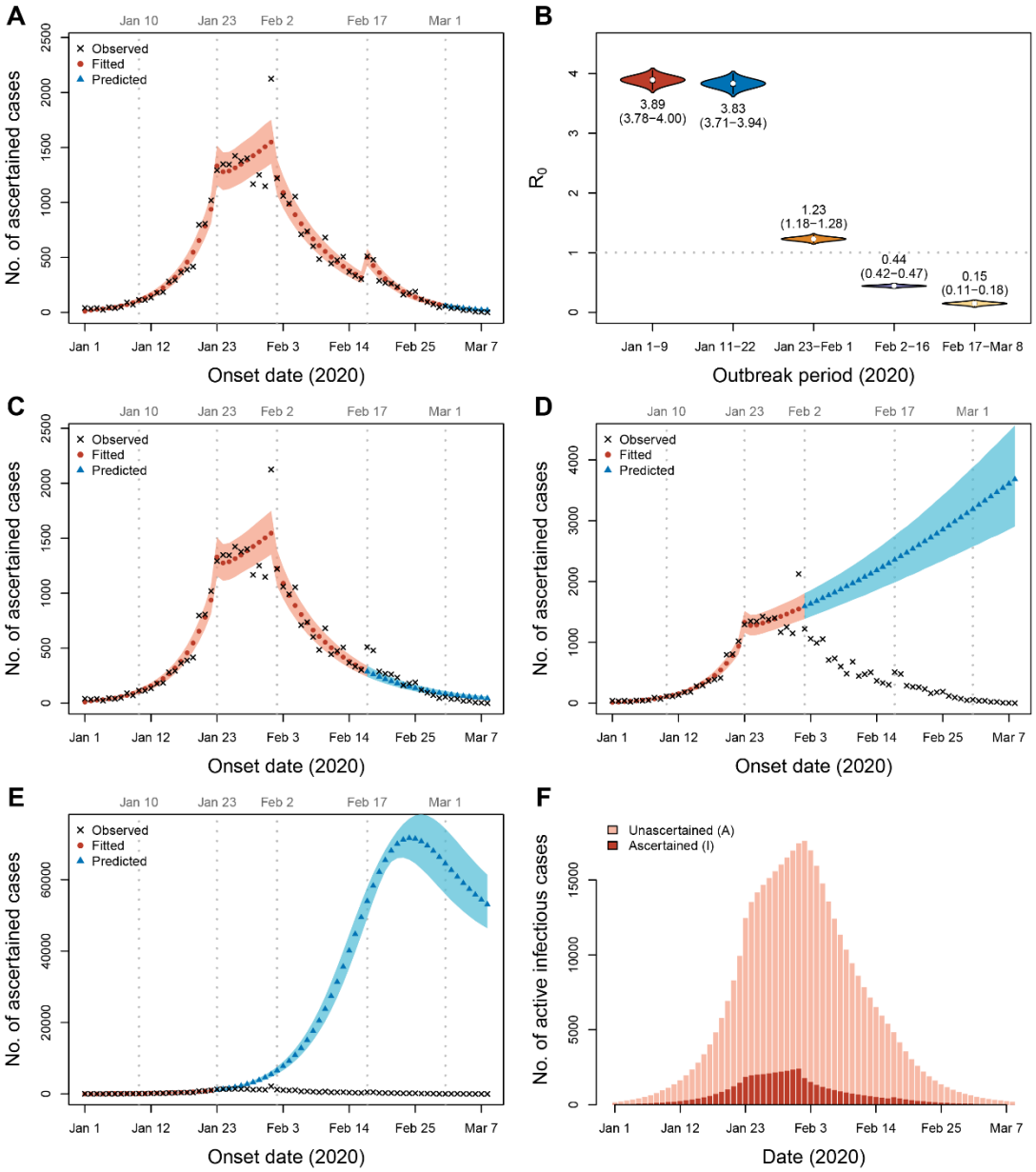

Fig. S8. Sensitivity analysis assuming the transmissibility of unascertained cases was half of that of ascertained cases ($\alpha = 0.5$, sensitivity analysis S5). Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

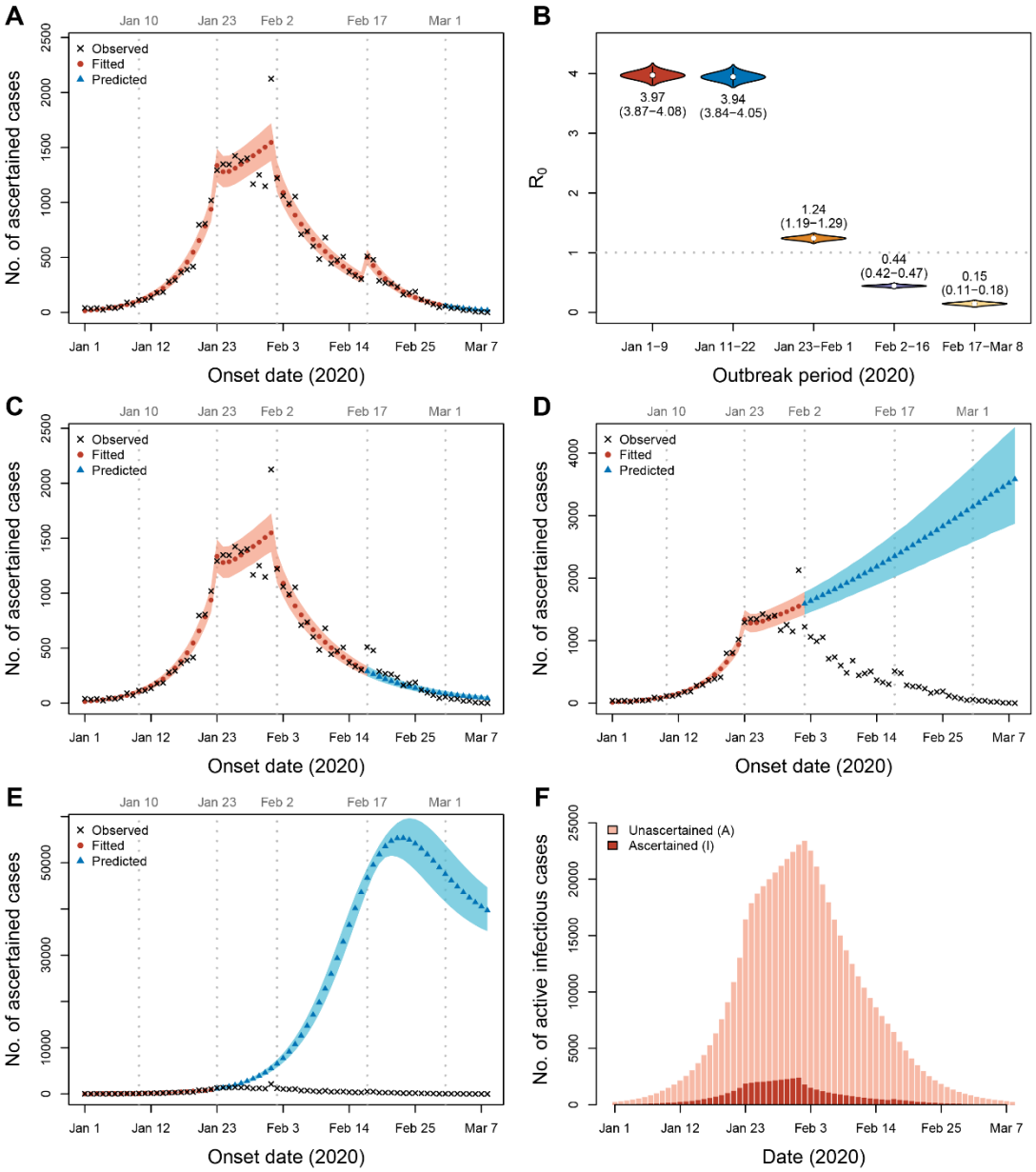

Fig. S9. Sensitivity analysis assuming the early ascertainment rate was 1/3 by setting $A(0) = 176$ and $E(0) = 555$ (sensitivity analysis S6). Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

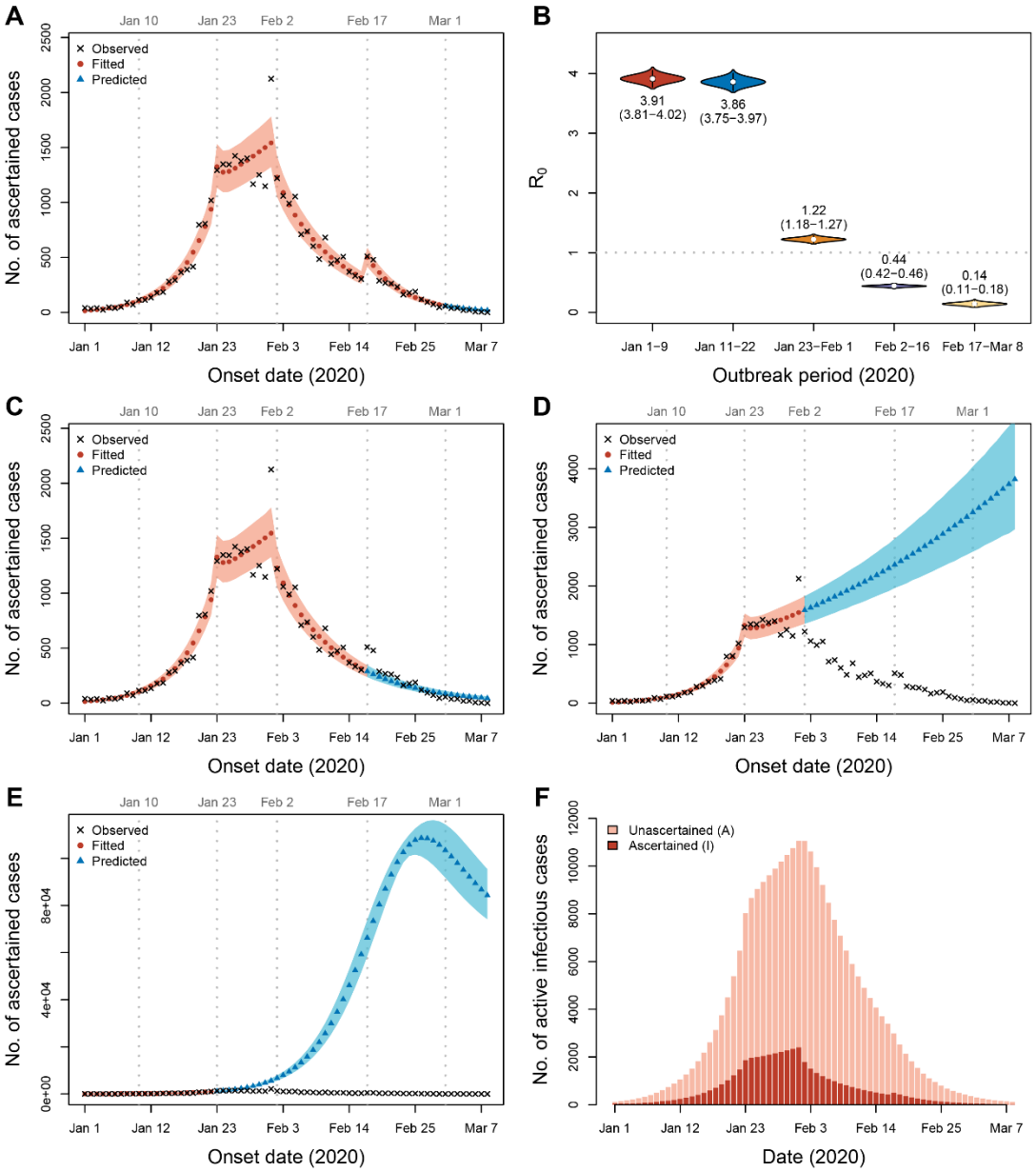

Fig. S10. Sensitivity analysis assuming the early ascertainment rate was 2/3 by setting $A(0) = 44$ and $E(0) = 278$ (sensitivity analysis S7). Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

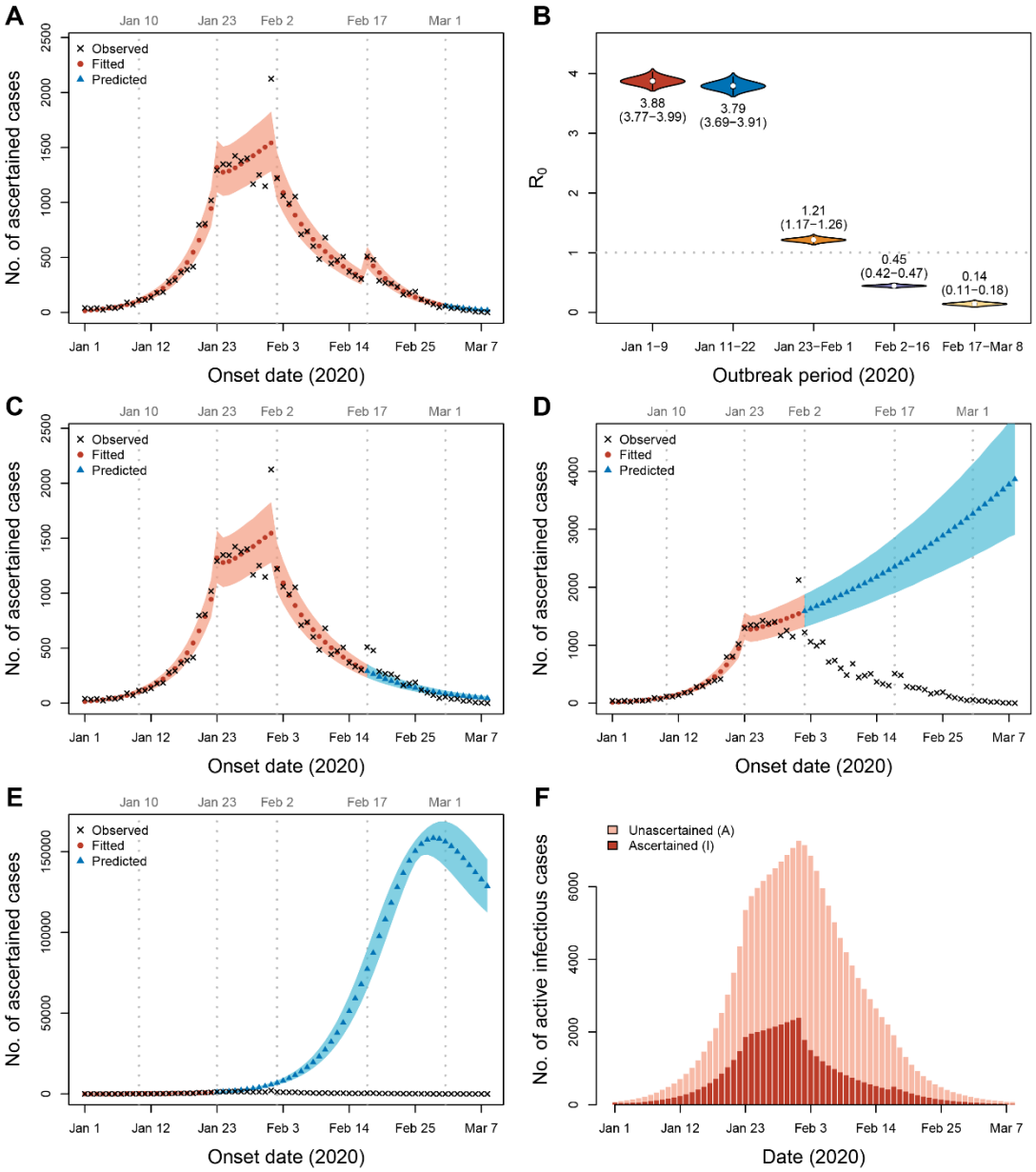

Fig. S11. Sensitivity analysis assuming the early ascertainment rate was 1 by setting $A(0) = 0$ and $E(0) = 185$ (sensitivity analysis S8). Parameters were estimated by fitting data from January 1 to February 29. (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

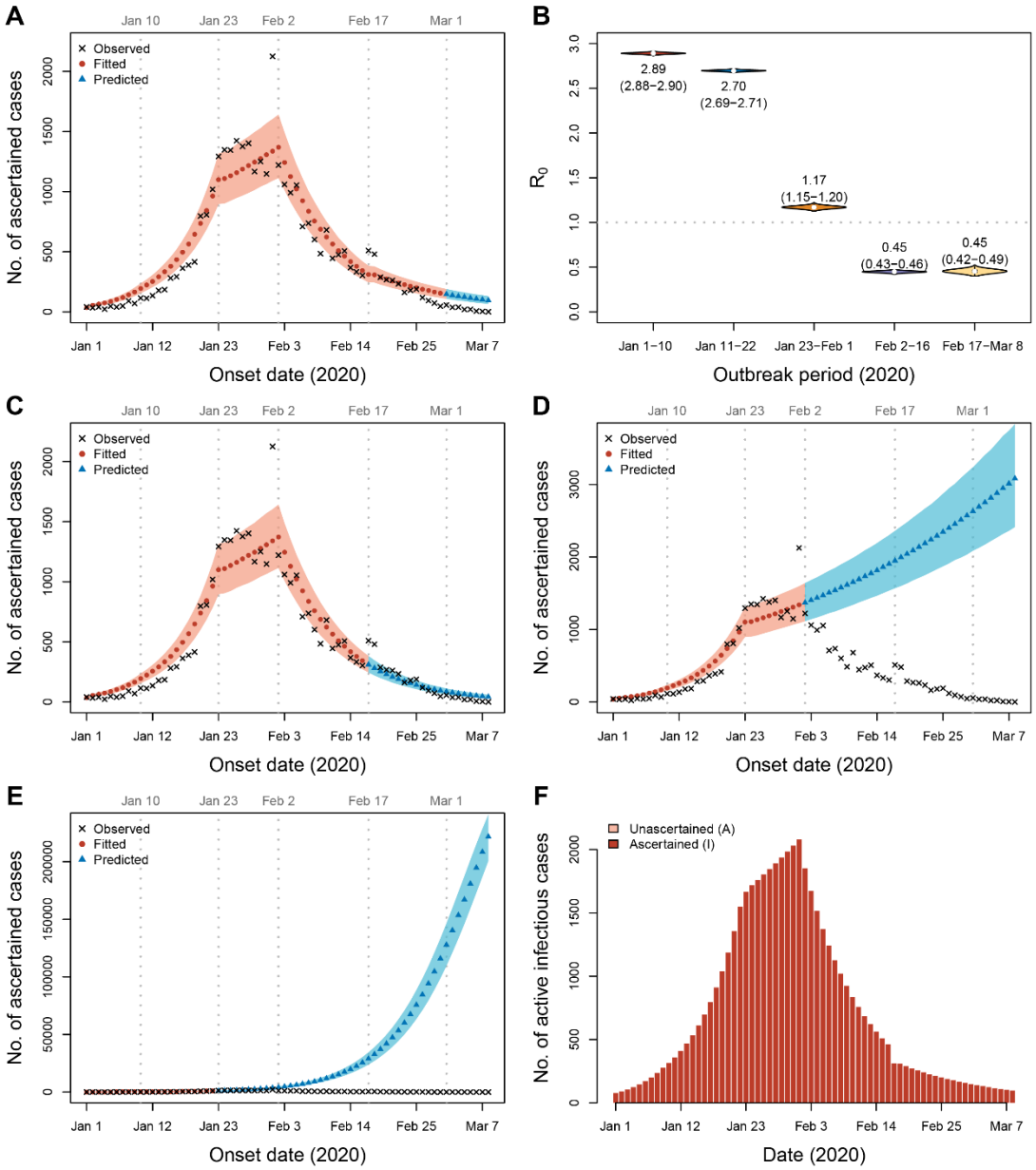

Fig. S12. Sensitivity analysis assuming complete ascertainment at any time (sensitivity analysis S10). Parameters were estimated by fitting data from January 1 to February 29. Compared to the full model, this simplified model fit the data significantly worse (likelihood ratio test, $\chi_4^2 = 19235$, $p = 0$). (A) Prediction using parameters from period 5 (February 17-29). (B) Estimated R_0 for each period. The mean and 95% CrI (in parentheses) are labeled below or above the violin plots. (C) Prediction using parameters from period 4 (February 2-16). (D) Prediction using parameters from period 3 (January 23-February 1). (E) Prediction using parameters from period 2 (January 10-22). The shaded areas in (A, C, D and E) are 95% CrI. (E) Estimated number of active infectious cases in Wuhan from January 1 to March 8.

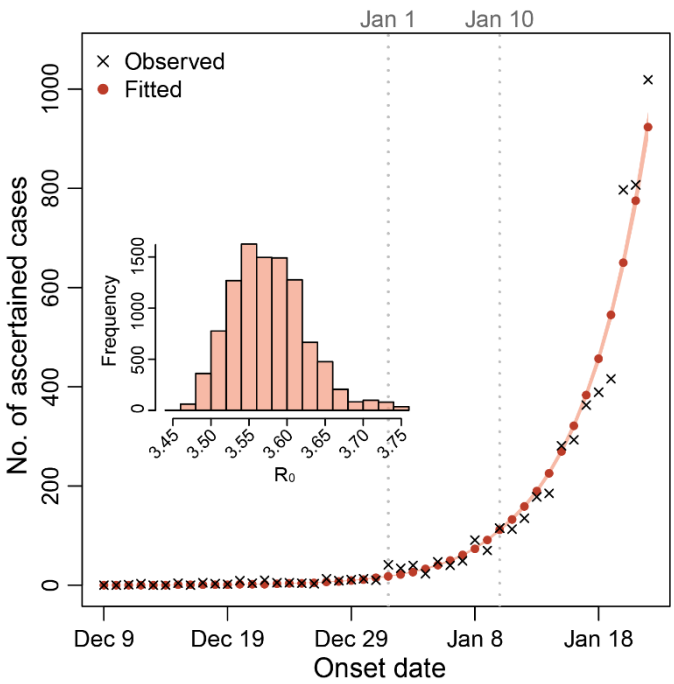

Fig. S13. Estimation of R_0 using daily incidence data starting from December 9. We set $I(0) = 1$ and $E(0) = A(0) = H(0) = R(0) = 0$ because the first case was reported to have illness onset on December 8, 2019. We assumed transmission rate b, ascertainment rate r, and duration from illness onset to hospitalization D_q (set to 10.5 days) were the same until January 22, 2020. All the other settings were the same as in the main analysis. The shaded area in the plot indicates 95% CrIs estimated by the deterministic SEIR model with 10,000 sets of parameter values sampled from MCMC. Unlike other analyses, we did not construct 95% CrIs by stochastic simulations, because stochastic fluctuations at the early days would have extremely large impacts due to low counts, leading to unreasonable CrIs. The inserted histogram shows the distribution of the estimated R_0 from December 9, 2019 to January 22, 2020, for which the mean estimate was 3.58 (95% CrI: 3.49-3.69).