



OPEN

Transmission dynamics and control measures of COVID-19 outbreak in China: a modelling study

Xu-Sheng Zhang^{1,2}✉, Emilia Vynnycky^{1,3}, Andre Charlett¹, Daniela De Angelis^{1,4}, Zhengji Chen⁵ & Wei Liu⁵✉

COVID-19 is reported to have been brought under control in China. To understand the COVID-19 outbreak in China and provide potential lessons for other parts of the world, in this study we apply a mathematical model with multiple datasets to estimate the transmissibility of the SARS-CoV-2 virus and the severity of the illness associated with the infection, and how both were affected by unprecedented control measures. Our analyses show that before 19th January 2020, 3.5% (95% CI 1.7–8.3%) of infected people were detected; this percentage increased to 36.6% (95% CI 26.1–55.4%) thereafter. The basic reproduction number (R_0) was 2.33 (95% CI 1.96–3.69) before 8th February 2020; then the effective reproduction number dropped to 0.04 (95% CI 0.01–0.10). This estimation also indicates that control measures taken since 23rd January 2020 affected the transmissibility about 2 weeks after they were introduced. The confirmed case fatality rate is estimated at 9.6% (95% CI 8.1–11.4%) before 15 February 2020, and then it reduced to 0.7% (95% CI 0.4–1.0%). This shows that SARS-CoV-2 virus is highly transmissible but may be less severe than SARS-CoV-1 and MERS-CoV. We found that at the early stage, the majority of R_0 comes from undetected infectious people. This implies that successful control in China was achieved through reducing the contact rates among people in the general population and increasing the rate of detection and quarantine of the infectious cases.

An outbreak of severe pneumonia, an infectious disease now named COVID-19, was reported in China in December 2019. The aetiological agent, SARS-CoV-2, a novel coronavirus, was isolated by Chinese authorities on 7th January 2020 and reported by WHO on 9th January 2020. The first case of COVID-19 was reported to have symptom onset on 1st December 2019 in Wuhan city of Hubei province, China¹, after which the virus quickly spread to other parts of China^{2,3} and the first case outside China was reported on 13th January 2020 (see Table 1). Because of the rapid spread of the disease, WHO announced the outbreak of COVID-19 as a “public health emergency of international concern” on 30th January 2020 and assessed COVID-19 as a pandemic on 11th March 2020⁴. By 25th March 2020, 405,742 cases and 18,791 fatalities associated with COVID-19 had been reported globally, of which, 81,218 cases and 3281 fatalities were reported in mainland China. By 10th April 2020 (25 days later), whilst the corresponding numbers in China had remained relatively unchanged (81,953 cases and 3339 deaths), the global numbers increased dramatically to 1,521,252 cases and 92,798 deaths. The comparatively small increase in the numbers of cases and deaths in mainland China can be attributed to draconian and rapid control measures implemented since late January 2020, starting from the lockdown of the epicentre Wuhan city on 23rd January 2020 and then extending to all of mainland China³. These included the complete shutdown and isolation of whole cities, cancellation of Chinese New Year celebrations, and prohibition of school and work attendance, frequent multi-media broadcasts of critical information (e.g., promoting hand washing, mask wearing, and care seeking), massive mobilization of health and public health personnel and military medical units, rapid construction of hospitals for patients with severe symptoms, and reconstruction of shelters for patients with no or mild symptoms. Control measures included travel bans and restrictions, contact reductions and social distancing, screening and contact tracing, early case identification and isolation^{2,5} to curb the epidemic.

¹Centre for Infectious Disease Surveillance and Control, National Infection Service, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK. ²Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College Faculty of Medicine, Norfolk Place, London W2 1PG, UK. ³TB Modelling Group, TB Centre, Centre for Mathematical Modelling of Infectious Diseases and Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. ⁴Medical Research Council Biostatistics Unit, University Forvie Site, Robinson Way, Cambridge CB2 0SR, UK. ⁵School of Public Health, Kunming Medical University, Kunming, Yunnan, People’s Republic of China. ✉email: xu-sheng.zhang@phe.gov.uk; liuweikm@qq.com

Date	Event
November 2019	Several pneumoniae of unknown aetiology were discovered in Wuhan city, Hubei province, China
01/12/2019	Symptom onset of the first case reported
31/12/2019	An alert was issued by the Wuhan Municipal Health Commission and a rapid response team was sent to Wuhan by Chinese Centre for Disease Control and Prevention (China CDC) The WHO China Country Office was informed of 27 cases of pneumonia of unknown aetiology detected in Wuhan city, Hubei Province of China
01/01/2020	Wuhan's Huanan Seafood Wholesale Market, the most probable index source of zoonotic COVID-19 infection, was shut down and disinfected China emergency response team was constructed
02/01/2020	The first 41 confirmed cases were identified as laboratory-confirmed COVID-19 in Wuhan city
03/01/2020	A total of 44 patients with pneumonia of unknown aetiology have been reported to World Health Organisation (WHO) by Chinese authorities Chinese National Health Commission (CNHC) issued the Diagnosis and Treatment plan for pneumonia of unknown aetiology
04/01/2020	One exported case from Wuhan city into Weng Zhou, Zhejiang Province, China
05/01/2020	A total of 59 patients with pneumonia of unknown aetiology have been reported to WHO by the Chinese authorities
07/01/2020	The causative pathogen was identified as a novel coronavirus and genomic characterisation and test method development ensued
12/01/2020	Notification of the novel coronavirus and its sequence was made to WHO
13/01/2020	First exported case was reported in Thailand
17/01/2020	A total of 62 patients with pneumonia of unknown aetiology have been reported
19/01/2020	A total of 198 patients with pneumonia of unknown aetiology have been reported
20/01/2020	Person-to-person transmission was confirmed and announced in China. China's "National Infectious Disease Law" was amended to make COVID-19 a class B notifiable disease and its "Frontier Health and Quarantine Law" was amended to support the COVID-19 outbreak response effort. By doing this, Chinese law required all cases to be immediately reported to China's Infectious Disease Information System CNHC started daily situation report of COVID-19 on their official website
23/01/2020	Lockdown of the epicentre Wuhan city, limiting mobility of people in and out of Wuhan, and then the control measures expanded quickly to the neighbouring areas
25/01/2020	The Chinese government raised the response level of the "Preparedness and Response Plan for Novel Infection Disease of Public Health Significance" to the Emergency level, based on the assessment that the risk of health impact caused by COVID-19 on the local population is high and imminent The Chinese government announced its highest-level commitment and mobilized all sectors to respond to the epidemic and prevent further spread of COVID-19
30/01/2020	WHO announced the outbreak of novel coronavirus as a "public health emergency of international concern"
12/02/2020	Definition of confirmed cases for Hubei province was changed to include clinically diagnosed cases or PRC tested positive. The extra high daily number of cases 15,152 was reported from Hubei province On 29/02/2020 this changed back to "clinically diagnosed plus PCR tested positive"
11/03/2020	WHO assessed COVID-19 as a pandemic
17/04/2020	CNHC raised the total death number in Wuhan city from 2579 to 3869

Table 1. Timeline of outbreak in mainland China.

As a novel coronavirus, it is important to understand its transmissibility and severity and to judge how effectively control measures in China have helped stop the spread of SARS-CoV-2 virus which may help inform control of COVID-19 outbreaks in the world and its future potential re-emergence. In previous analyses of transmission dynamics in China, Kucharski et al.⁶ fitted a stochastic transmission dynamic model to four datasets of cases from Wuhan and just considered the period up to 11th February 2020. They estimated that the median daily reproduction number (R_t) in Wuhan declined from 2.35 (95% CI 1.15–4.77) 1 week before travel restrictions were introduced on 23rd January 2020, to 1.05 (95% CI 0.41–2.39) 1 week after. Other modelling studies^{2,3,7,8} also used the confirmed case data from the early pandemic stage to model and estimate the spread, dynamics and control of the COVID-19 outbreak in China. As the number of new confirmed case in mainland China reached a peak only after 17th February 2020 and other relevant data such as those relating to fatality and hospital discharges are becoming available, a further study including more data streams is needed to show the overall effectiveness of the integrated control measures executed in the whole nation.

Based on the published data associated with the confirmed cases, death and recovery, in this study we use a Synthesis model (Fig. 1) which covers both the transmission dynamics of the SARS-CoV-2 virus and the disease reporting process of COVID-19 to estimate the transmissibility of SARS-CoV-2 and the severity of COVID-19 in China and the impact on the spread of COVID-19 of control measures the Chinese government has taken from late January 2020.

Results

The estimates of the model parameters are shown in Table 2. The proportion of COVID-19 cases in mainland China that were detected and reported was about 3.5% (95% confidence interval (CI): 1.7–8.3%) before 19th January 2020, and then increased to 36.6% (95% CI 26.1–55.4%) afterwards. At the early stage before 8th February 2020, the transmissibility of COVID-19 was high with $R_0 = 2.33$ (95% CI 1.96–3.69); however, it reduced

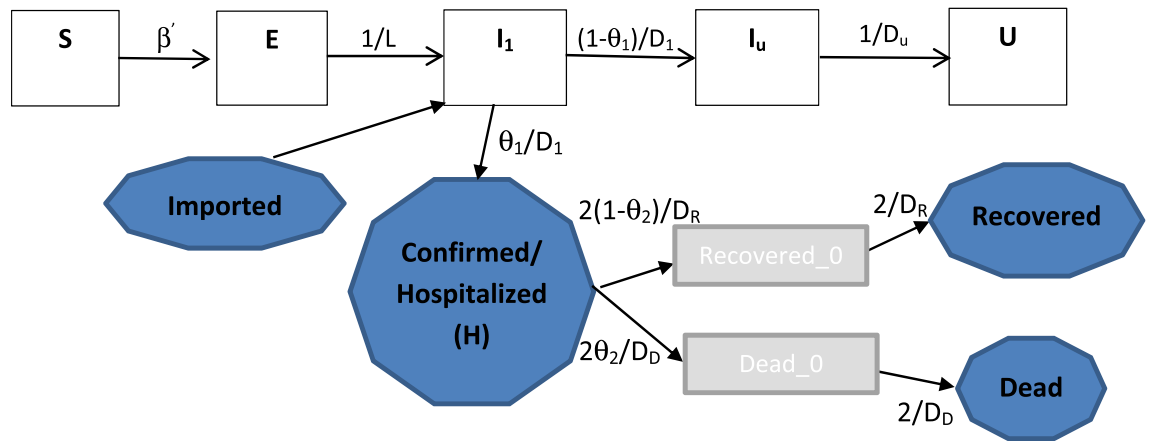


Figure 1. Flow chart of the synthesis model: transmission dynamics and disease reporting processes. The five rectangular boxes represent the hidden transmission dynamics process with U representing the number of people who have recovered from an undetected infection and the four blue shaded polygons represent the quantities upon which observations were made. The quantities above the arrows represents the rate at which people go from one compartment to the next and $\beta' = \beta(t)[I_1(t)(\theta_1 + \xi(1 - \theta_1)) + \xi I_u(t)]$.

Name	Definition	Prior	Posterior after distributing 1290 deaths to:	
			Before 20 February	Whole period
τ_β	Time point when the transmission rate is estimated to change because of control measures	U[60,143]*	70.6 (69.2,71.9)	71.4 (70.2,72.0)
β_a	Daily transmission rate before τ_β	U[0.020,1.00]	0.523 (0.311,0.691)	0.512 (0.292,0.674)
β_b	Daily transmission rate after τ_β	U[0.001,0.20]	0.009 (0.001,0.031)	0.007 (0.001,0.031)
$R_{0,1}$	Reproduction number before τ_β	-	2.33 (1.96,3.69)	2.20 (1.85,4.10)
$R_{0,con}$	R_0 due to confirmed cases before τ_β	-	0.042 (0.017,0.111)	0.037 (0.013,0.107)
$R_{0,und}$	R_0 due to undetected cases before τ_β	-	2.27 (1.92,3.65)	2.16 (182,4.07)
$R_{0,2}$	Reproduction number after τ_β	-	0.036 (0.006,0.100)	0.027 (0.006,0.094)
I_0	Initial number of infectious people on 1/12/2019	U[1,500]	8.7 (2.7,49.1)	14.3 (2.4,67.8)
	Initial total number of people infected with the virus on 1/12/2019	-	40.9 (12.8,207.2)	71.1 (12.4,304.4)
τ_θ	Time point when the case ascertainment rate is estimated to change	U[40,62]*	49.3 (47.1,51.6)	50.1 (48.0,52.8)
$\theta_{1,a}$	Case ascertainment rate before τ_θ	U[1%,25%]	3.48% (1.74%,8.33%)	3.47% (1.35%,8.96%)
$\theta_{1,b}$	Case ascertainment rate after τ_θ	U[25%,100%]	36.61% (26.07%,55.39%)	38.29% (30.44%,55.67%)
τ_F	Time point when the confirmed case-fatality rate (cCFR) is estimated to change	U[65,95]*	77.0 (75.3,78.0)	77.4 (76.0,78.9)
$\theta_{2,a}$	Mortality rate among confirmed cases (cCFR ₁) before τ_F	U[0.5%,50%]	9.61% (8.12%,11.36%)	8.56% (7.51%,10.01%)
$\theta_{2,b}$	Mortality rate among confirmed cases (cCFR ₂) after τ_F	U[0.1%,50%]	0.67% (0.45%,0.97%)	1.28% (0.96%,2.35%)
D_1	Average infectious period before hospitalization (days)	U[2.0,10]	2.28 (2.01,3.12)	2.12 (2.0,2.71)
D_1+D_u	Average infectious period of undetected infections (days)	U[3.0,25.0]	4.44 (3.19,11.71)	4.31 (3.11,13.43)
η^{HOS}	Dispersion parameter for reported cases	U[1.01,1500]	76.0 (50.2,122.8)	82.8 (53.9,134.1)
η^{Death}	Dispersion parameter for deaths	U[1.01,5000]	3.5 (2.4,5.2)	4.0 (2.8,5.9)
$\eta^{Recovery}$	Dispersion parameter for recoveries	U[1.01,1500]	144.8 (102.8,205.6)	150.3 (104.7,215.37)
IFR ₁	Infection fatality rate ($\theta_{1,a},\theta_{2,a}$) before τ_θ	-	0.33% (0.17%,0.85%)	0.30% (0.11%,0.81%)
IFR ₂	Infection fatality rate ($\theta_{1,b},\theta_{2,b}$) between τ_θ and τ_F	-	3.51% (2.60%,5.12%)	3.31% (2.47%,4.89%)
IFR ₃	Infection fatality rate (IRF ₂ = $\theta_{1,b},\theta_{2,b}$) after τ_F	-	0.24% (0.15%,0.41%)	0.50% (0.34%,0.86%)

Table 2. Estimates of model parameters for the COVID-19 outbreak within mainland China. The 1290 deaths within Hubei province added on 17th April 2020 were distributed over either the period before 17th April (i.e., the whole period) or before 20th February 2020 in proportional to the daily number of deaths reported before 17th April 2020. The relative infectiousness of undetected cases compared to confirmed cases $\xi = 1$ (i.e., both undetected and confirmed infections are of the same infectiousness). The estimated values of model parameters are nearly the same for both ways of distributing 1290 deaths added on 17th April 2020 by Chinese National Health Commission except for estimates of cCFR₂ and IFR₃. *The epidemic was assumed to start from 1st Dec 2019¹.

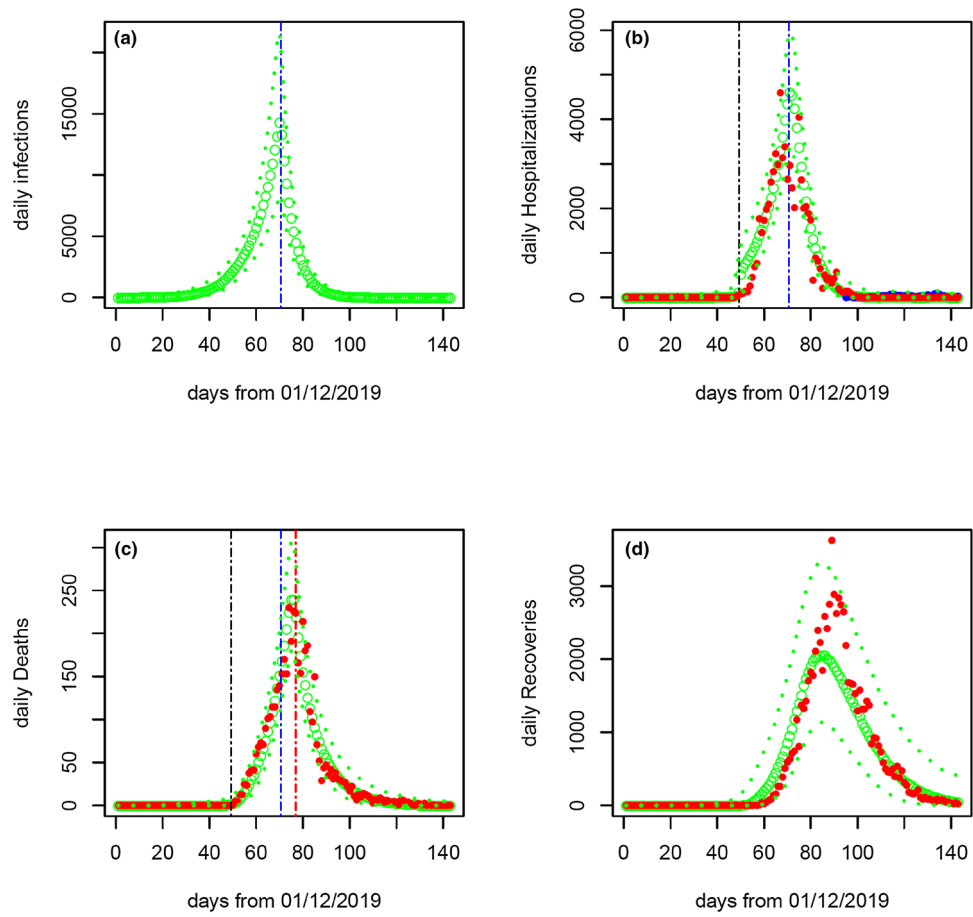


Figure 2. Epidemic curves from 1st December 2019 to 21st April 2020. Model predictions of infections (panel a) and model fitting to the daily numbers of confirmed cases (panel b), deaths (panel c) and recovered cases (panel d) are shown. The model predictions are obtained after redistributing the 1290 deaths added to the official data on 17th April 2020 to the period before 20th February 2020 in accordance with the daily number of deaths reported before 17th April 2020. The green dotted lines represent the model predictions (large green circles for median and thin green dotted lines for lower and upper levels of 95% confidence interval). The red points are the observed data. The black, blue and red vertical lines denote the estimates of times when the case ascertainment rate, transmission rate and confirmed case fatality rate respectively are estimated to have changed. In panel (b), the blue points represent daily number of imported cases which are nearly invisible. Note that the daily number of confirmed cases (15,152) on 12th February 2020 (day 74) is beyond the range shown in panel (b).

dramatically to 0.04 (95% CI 0.01–0.10) from 8th February 2020 (95% CI 7th–9th February 2020). By fixing the duration from onset of symptoms to death and recovery at 17.8 days and 22.6 days respectively⁹, the average infectious period for those that are symptomatic and confirmed and then quarantined is 2.28 days (95% CI 2.01–3.12 days) and the average infectious period for undetected infections is 4.44 days (95% CI 3.91–11.71 days).

In our model, the basic reproduction number is contributed by two parts: the confirmed cases and undetected infections. Confirmed cases and undetected infections are assumed to be equally infectious in this study; combining with the estimate that only about 3.5% of infections were confirmed and reported at the early stage, this suggests that at the early stage, about 98% of R_0 (Table 2) is due to the undetected infections. If the infectiousness of undetected infections is only half or one third of that of confirmed cases⁷, R_0 in the absence of interventions remains nearly the same, and about 96% of R_0 is due to the undetected infections (Supplementary Table S2.1). This indicates that only isolating confirmed cases and their contacts is not enough to stop the spread and that the main factor that stopped the COVID-19 outbreak in mainland China was the dramatic drop in contact rates among the general population (c.f.⁷).

The model fitting to the daily number of hospitalizations (confirmed cases), the daily number of deaths, and daily number of recovered people are shown in Fig. 2. Our model analysis suggests that the death rate could have changed around 15 February 2020 (95% CI 14th–16th February 2020). Before this date the confirmed case fatality rate (cCFR) is 9.6% (95% CI 8.1–11.4%), after this date it reduces to 0.7% (95% CI 0.4–1.0%). Based on the estimates of the case ascertainment rate and case fatality rate, the infection fatality rate (IFR) is 0.33% (95% CI 0.17–0.85%) before 19th January 2020 and increases to 3.51% (2.60–5.12%) during the period from 20th January to 15th February 2020; after this, it decreases to 0.24% (95% CI 0.15–0.41%). Our model inference suggests

Start date	Number of infections	Number of confirmed cases	Number of deaths
2 nd Jan 2020	17,536 (4501, 47,147)	5642 (936, 20,617)	301 (33, 1198)
9th Jan 2020	41,856 (15,845, 87,826)	14,156 (4,696, 35,681)	769 (205, 2072)
16th Jan 2020	97,184 (47,351, 171,881)	33,547 (16,736, 64,441)	1,828 (828, 3705)
23 rd Jan 2020	223,111 (125,333, 361,357)	78,247 (49,893, 122,761)	4280 (2600, 6964)
30th Jan 2020*	509,036 (301,752, 810,468)	181,200 (126,400, 253,818)	9,936 (6,806, 14,269)
6th Feb 2020*	1,164,089 (668,901, 1,898,473)	420,146 (277,738, 583,748)	23,082 (15,068, 32,260)
13th Feb 2020*	2,671,685 (1,394,762, 4,577,990)	971,916 (564,900, 1,449,202)	53,490 (30,710, 78,960)

Table 3. Impact of changing the start time of control measures during the outbreak in mainland China on the number of infections, confirmed cases and deaths from 1st December 2019 to 21st April 2020. *China started its Chinese New Year holidays from 25th January 2020, when the number of contacts made between people became much higher than during the other periods of the year. Hence the estimated numbers of cases are likely to be conservative.

that on 1st December 2019, there were 41 (95% CI 13–207) people infected with SARS-CoV-2 virus. These estimates are based on the assumption that the 1290 deaths added on 17th April 2020 were distributed before 20th February 2020. If we assume that the 1290 deaths were distributed before 17th April 2020, the estimates of most model parameters remain nearly the same except for the death-related quantities (Table 2). For example, after 15 February 2020 cCFR changes from 0.67 to 1.28%, and the IFR from 0.24 to 0.50%.

If the public had been made aware of COVID-19 earlier and the control measures started earlier, the size of outbreak could have been much smaller. This can be quantitatively analysed by assuming the same epidemiological characteristics but moving the start times at which the reporting, transmission and death rates changed. The results are listed in Table 3. It shows that if the control measures started 1 week, 2 weeks, or 3 weeks earlier, about 57% (57%), 81% (82%) and 93% (93%) of the confirmed cases (and deaths) would have been averted. However, if the control measures had started 1 week, 2 weeks, or 3 weeks later, the numbers of the confirmed cases (and deaths) would have increased 2.3-fold (2.3), 5.3-fold (5.4), and 12.4-fold (12.5) across mainland China, respectively.

To investigate how the epidemic within the epicentre differs from that over the whole nation, we also obtain estimates of model parameters for Hubei province and Wuhan city (Supplementary Table S3.1). The results show that within the epicentre, the transmissibility and case ascertainment rate differ only slightly. However, there was a large difference in the estimated confirmed case fatality rate. Before the time point τ_F when the confirmed case-fatality was estimated to have changed (15 February 2020), the estimated cCFR is 9.6% (95% CI 8.1–11.4%) across the whole country, 11.5% (95% CI 10.4–13.5%) over Hubei province, and 15.1% (95% CI 13.3–17.6%) within Wuhan city. However, the estimate of cCFR appears to be nearly the same in all three settings after τ_F . These estimates imply that the fatality rate before 15 February 2020 is significantly higher at the epicentre than the national average, and many deaths of COVID-19 patients were due to insufficient treatments at the epicentre.

Discussion

Understanding the transmissibility and severity of the novel coronavirus (SARS-CoV-2) is paramount; understanding how its rapid spread was brought under control in mainland China is of practical importance for other countries now facing ongoing outbreaks of COVID-19. In this study we have studied the transmissibility and control of SARS-CoV-2 in China by using a mathematical model to reconstruct the COVID-19 outbreak in mainland China from 1st December 2019 to 21st April 2020. Our analyses indicate that the SARS-CoV-2 had a basic reproduction number of 2.33 (95% CI 1.96–3.69) in the absence of intervention measures and therefore it is highly transmissible. The fatality rate among those that are symptomatic and confirmed was about 9.6% before 15 February 2020, and then reduced to 0.7%. The draconian control measures taken by the Chinese government from 23rd January 2020 brought the spread of SARS-CoV-2 under control in mainland China. However, it took more than 2 weeks for the effect of control measures to emerge.

Our study shows that the early reporting rate (3.5%) was very low, suggesting that about 96.5% of all infections were undetected prior to 19th January 2020 (c.f.^{7,10,11}). This might be mainly due to the limited knowledge and unclear definition of the novel disease¹². The reporting rate of 36.6% at the late stage reflects increased awareness of the virus and consequently escalating rate of medical help seeking behaviour for respiratory symptoms; however, this rate appears relatively low and suggests that many infections were associated with mild symptoms or were asymptomatic. Analyses by China CDC¹³ using data up to 11th February 2020 estimated that COVID-19 has been mild for 81% of infected people. The high proportion of mild symptomatic or asymptomatic infections was further confirmed by a well-investigated outbreak on the Diamond Princess cruise ship during February 2020: among 696 confirmed cases, 410 (58.9%; c.f.¹⁴) were asymptomatic¹⁵.

The actions taken by the Chinese government appear to have stopped the spread of COVID-19 in mainland China. This was achieved under very strict control measures, which dropped the transmissibility, as reflected by the effective reproduction number to 0.04, which is 1.8% of its initial value of $R_0 = 2.33$ (c.f.³). This estimate of R_0 at the early stage is consistent with most previous estimates^{3,6–8,10,11,16–22}. We found that the infectious period before isolation for confirmed cases is about 2.3 days, which is comparable with the estimates^{23–26}; though it appears to be shorter than that obtained by Tian et al.³: 5.19 days (95% CI 4.51–5.86 days). Our estimate of the infectious

period for infections that were not detected and not quarantined (4.4 days ranging from 3.2 to 11.7 days) agrees with the range identified in previous studies^{27,28} of 1–14 days.

Although increasing the rate of detection and quarantine of symptomatic cases can help reduce the number of sources of infection (c.f.²), the main force of transmission is from the undetected cases which contributed to most of the transmissibility of SARS-CoV-2 during the early stage of the outbreak in mainland China (c.f.⁷). Hence it is likely that limiting the mobility of the general population made the biggest contribution to stopping the spread of SARS-CoV-2 virus in mainland China. This may further explain why the draconian control measures implemented in China since the 23rd January 2020 took about 16 days (i.e., delay from 23rd January to 8th February 2020) to influence transmission within community.

The impact of actions taken by the Chinese government is also reflected on the change of mortality associated with COVID-19. Before 15th February 2020, cCFR is estimated at 9.6% and then it reduced to about its 7% (i.e., 0.67%). This huge reduction in cCFR may be due to the significant improvement of medical treatment and availability of medical resources in China, especially in the epicentre Hubei province.

It is interesting to compare our estimate of the mean case fatality rate against that from other studies, which, after averaging the values estimated before and after the 15th February, is 5.14%. This is slightly less than that of Deng et al.²⁹: 5.65% (95% CI 5.50–5.81%). Including death data up to 17th April 2020, Deng et al.²⁹ used individual-level data for cases to obtain their estimate which is close to both the corresponding crude or naïve confirmed case fatality risk: $4632/82,758 = 5.60\%$ and the approximator of deaths/(deaths + recoveries) = $4632/(4632 + 78,112) = 5.60\%$ as of 21st April 2020. Based on the data up to 11th February 2020, Verity et al.⁹ suggested that the overall CFR for the outbreak in China was 1.38% (95% CI 1.23–1.53%). Using data up to 29th February 2020, Wu et al.³⁰ found the overall symptomatic case fatality rate (the probability of dying after developing symptoms) to be 1.4% (95% CI 0.9–2.1%). Russell et al.³¹ used several simplifying assumptions to obtain an estimate of the CFR in China to be 1.2% (95% CI 0.3–2.7). Our estimate of CFR is higher than these, which largely results from the fact that these three studies did not include the 1290 deaths added to the official data on 17th April 2020 by the Chinese National Health Commission³².

Through model fitting to observed data in mainland China, we found that the model predicted that there had been twenty eight times more cases than were reported at the early stage of the outbreak; even at the late stage since 19th January 2020, the estimated number of infected people were more than twice those reported. Based on the epidemiological characteristics obtained from COVID-19 outbreak data, our analyses suggest that even if the control measures had started 1 week earlier, they would have averted 57% of confirmed cases and 57% of deaths. If they had started 3 weeks earlier, then 93% confirmed cases and 93% of deaths would have been averted. This estimate of the effect of start time of control measures is smaller than but comparable with what Lai et al.² and Yang et al.⁵ found, although both studies estimated only the number of infections. These findings highlight the importance of an early response in controlling transmission in the population.

As in Kucharski et al.⁶, we assumed the latent period is equal to the incubation period. In view of evidence that there is pre-symptomatic transmission^{28,33,34}, it is interesting to know whether this will alter the results of our model. For this, we modified the model system (Fig. 1) by dividing exposure stage *E* into two equal sub-classes *E*₁ and *E*₂ and assumed that people in *E*₂ can transmit the virus with the same infectiousness as ill cases (SI Sect. 4). We obtain similar results (Supplementary Table S4.1): For example, the basic reproduction number before the control measures is 2.22 (95% CI 1.95–2.92) of which 98% was due to undetected cases, and the cCFR is 9.5% (95% CI 8.0–11.2%) before 15 February 2020 and then reduces to 0.7% (95% CI 0.4%, 1.1%). This indicates our conclusions are relatively unaffected by assumptions relating to pre-symptomatic transmission.

The non-pharmaceutical interventions introduced in mainland China appeared to have stopped the spread of the virus, but the risk of another potential outbreak remains. As estimated in our analyses, at most 223,111 individuals were infected up to 21st April 2020 during the outbreak (Table 3), which is less than 0.02% of the Chinese population. Even if all those recovered from infection have developed complete immunity³⁵, this level of immunity in the Chinese population is considerably lower than the herd immunity threshold of 55% which is required for control of transmission^{36,37}. This simply implies that once the strict quarantine measures currently in place in mainland China are relaxed, transmission of SARS-CoV-2 in China is very likely to rebound, especially given the extent of ongoing transmission across the rest of the world (c.f.³⁸). Nevertheless, the non-pharmaceutical interventions appear to have halted the spread of SARS-CoV-2 virus in mainland China and bought time for vaccines and drugs to be developed and used later on.

It is of interest to compare the novel coronavirus SARS-CoV-2 with other two coronaviruses: SARS-CoV-1 and MERS-CoV which caused large outbreaks in human populations. SARS-CoV-1 has a *R*₀ from 2 to 5³⁹ and case-fatality rate of 9.6% among probable cases in mainland China⁴⁰. MERS-CoV in the 2015 outbreak in South Korea has been estimated to have a *R*₀ from 2 to 7⁴¹ and case fatality rate of 34.5% among laboratory-confirmed cases⁴². This suggests that that SARS-CoV-2 is nearly as transmissible as SARS-CoV-1 and MERS-CoV, but it is potentially less severe particularly when sufficient treatments are available. Furthermore, there was no evidence of a super-spreader event occurring in any of the Chinese health facilities serving COVID-19 patients, which is distinct from the 2003 outbreak of SARS-CoV and 2015 MERS-CoV outbreak in South Korea. Due to advanced modern transportation, people can move easily and quickly across the world, which allows SARS-CoV-2 to quickly spread to other countries. This is in a sharp contrast with both SARS-CoV-1 and MERS-CoV which have been controlled and confined within relatively limited areas of the world. This sharp difference may be attributed to another aspect of the coronaviruses: a large proportion of SARS-CoV-2 infections are associated with mild symptoms or are asymptomatic^{13–15}, while both SARS-CoV-1 and MERS-CoV are highly symptomatic^{21,33}. In the absence of high coverage with highly effective vaccines, this characteristic of the SARS-CoV-2 virus will make it challenging for humans to control and manage it.

Compared with previous studies^{2,3,5,6,21,22}, our analyses have two key strengths: our investigations are based on three datasets (confirmed cases, deaths and recovered cases) and we model the outbreak over a long period

(143 days) which should avoid any bias and confounding arising due to observations over a short period. This study also has several limitations. To model the transmission dynamics and disease reporting, our synthesis model has included several simplifications. To reflect the temporal change in the ascertainment rate, transmissibility and case fatality of COVID-19, two different values are assumed for each. These quantities are likely to change gradually during the outbreak, as is the case for public awareness and interventions³. For example, Tsang et al.¹² found the ascertainment rate changed as the case definition for COVID-19 changed from being initially narrow to becoming gradually wider during the period from 15th January to 3rd March. The time-to-event intervals such as the delay from symptom onset to death may also change as epidemic grows^{9,24,26} and the time from symptom onset to hospitalization may reduce as more reliable medical resources become available^{22,24}. Further, in this study we ignore the heterogeneity in both geography and age^{3,7,13}. To provide more specific and practically useful information for control measures, it needs to look at variations between regions^{3,7} and in different age groups¹³. A further limitation is that we model the overall effectiveness of integrated intervention measures rather than the different types of control measures and therefore cannot provide insight into their relative impacts in stopping the spread of infection (c.f.^{2,3,16}).

In conclusion, our finding that the main driver of transmission of SARS-CoV-2 at the early stage of the outbreak in mainland China came from the undetected infections provides helpful information for policy makers when designing optimal intervention strategies. In the absence of vaccination and effective drugs, early detection and isolation are essential for containing and controlling the spread of SARS-CoV-2 but reducing contact rates among people in the general population is likely to have the biggest impact.

Methods

Data. We extracted the following data relating to COVID-19 for mainland China from 3 datasets on the website of Chinese National Health Commission for the period from 1st December 2019 to 21st April 2020: the daily number of confirmed cases who were confirmed/admitted to hospital, the daily number of deaths and the number of patients who had recovered each day. Here the data are given by the symptom onset date during the period 1st December 2019–1st January 2020 from¹ and by reporting date thereafter from the website of Chinese National Health Commission, due to difficulties in collecting the onset dates of those cases through the website. The reporting date is assumed to be the same as the date that cases were diagnosed. Before 12th February 2020, confirmed cases were defined as those who were positive for SARS-CoV-2 on PCR; from 12th February 2020 confirmed cases were defined (for the epicentre Hubei province) as those who were either clinically diagnosed or positive for SARS-CoV-2 on PCR¹², and as those who were PCR positive for SARS-CoV-2 for the rest of China. This case definition resulted in the number of confirmed cases increasing from 2015 to 15,151 between the 11th and 12 February 2020.

The daily numbers of reported deaths were adjusted to account for changes to the numbers of deaths that were made by the Chinese government, as follows. Specifically, on 17th April 2020, Chinese National Health Commission revised its estimates of the total numbers of COVID-19-related deaths in Wuhan city that had occurred since the start of the outbreak from 2579 to 3869³². It stated that these additional 1290 deaths probably occurred before 20th February 2020, given that, after this date, the number of hospitals that could treat COVID-19 patients increased from 2 to 48, and the HOUSHEN and LEISHEN hospitals, and FANGCANG shelters that have been constructed since then also provided an increased number of beds meeting the needs of COVID-19 patients with different symptoms. Furthermore, the system of data collection improved rapidly, and the number of missed cases/deaths decreased greatly during the course of the outbreak. As detailed information for these 1290 deaths is not available, for the sake of model fitting, these 1290 deaths were distributed in proportion to the number of deaths reported each day from the date of the first death (10th January 2020) to 20th February 2020. For example, the number of deaths reported on 18th January 2020 was 26, the number is now corrected to $26 + 26 \times 1290/2236 = 26 + 15 = 41$. Here 2236 is the cumulative number of deaths reported up to 20th February 2020 before the government had revised its estimates in April. In sensitivity analysis, we also explored the effect of assuming that these 1290 deaths added on 17th April 2020 were distributed in proportion to the daily number of deaths reported over the entire period from 10th January to 17th April 2020. Except for estimates of mortality rate, the results shown in Table 2 (and described in the “Results” section) are very similar for both ways of distributing the 1290 deaths added to the official statistics on 17th April 2020.

Model. In this study we use a synthesis model⁴³ (Fig. 1), which combines the hidden transmission dynamics of SARS-CoV-2 virus and the reporting system of COVID-19, to investigate the transmissibility and severity of COVID-19 and the efforts to contain and control the spread of SARS-CoV-2. We assume the transmission dynamics of SARS-CoV-2 virus are described by an SEIR compartmental model, with a few modifications, as shown in Fig. 1. That is, a susceptible person (S) can contract SARS-CoV-2 virus from infectious people and then enter the latent class (E); after an average latent period (L), the exposed person progresses to become infectious (I_1). A fraction (θ_1) of these people (with severe symptoms) will be detected and admitted to hospital (H) and will then be treated and isolated from the community, with a fraction ($1 - \theta_2$) recovering after an average period D_R and the other fraction dying after an average period D_D . The other fraction ($1 - \theta_1$) of infected people, typically with mild or asymptomatic infections (I_u) will not be detected and hence remain in the community as sources of infection for a further average period D_i before recovering. For simplicity, we assume that the average latent period equals the incubation period, which is fixed at $L = 5.2$ days as estimated by^{18,26,44}. SARS-CoV-2 virus can be transmitted by three possible modes: respiratory transmission (through respiratory droplets when symptomatic people sneeze or cough), aerosol transmission (through fine virus particles that were aerosolized) and contact transmission (through contacting the contaminated surface). For simplicity and consistency with these easy and quick modes of transmission, we assume that people mix randomly. Although China is a large

Name	Definition
Transmission dynamics	
$S(t)$	Number of susceptible people at time t
$E(t)$	Number of exposed (infected but not yet infectious) people at time t
$I_1(t)$	Number of all infectious people at time t who have not yet been detected
$I_u(t)$	Number of infectious people at time t who have not been detected and will remain undetected
Disease reporting	
$H(t)$	Number of people who were hospitalized/reported due to COVID-19 at time t
$Dead_0(t)$	Number of people in the “Dead_0” compartment. This is defined as an intermediate compartment between those hospitalised (H) and $Dead$ at time t . This compartment does not have any epidemiological interpretation, but including it in the model results in the distribution of the time interval from hospitalisation to death following the Gamma distribution
$Dead(t)$	Number of people who have died from COVID-19 at time t
$Recovered_0(t)$	Number of people in the “Recovered_0” compartment. This is defined as an intermediate compartment between those hospitalised (H) and $Recovered$ at time t . This compartment does not have any epidemiological interpretation, but including it in the model results in the distribution of the time interval from hospitalisation to recovery following the Gamma distribution
$Recovered(t)$	Number of recovered people at time t

Table 4. Definitions of model compartments.

country with a population of 1,400,050,000 people residing on a huge area of 9,596,960 km², its recent urbanization and development of rapid transport systems make it easy and quick for people to move around the country. This makes it reasonable to model the transmission of SARS-CoV-2 within the whole country as a well mixing population. For comparison, we also model the spread of COVID-19 within Hubei province and Wuhan city where the outbreak started (SI Sect. 3).

The synthesis model is described by Eq. (1). The 9 compartments are defined in Table 4 and the definitions of model parameters are given in Table 2.

$$\begin{aligned}
 \frac{d}{dt}S(t) &= -\beta(t)S(t)(I_1(t)(\theta_1 + \xi(1 - \theta_1)) + \xi I_u(t))/N \\
 \frac{d}{dt}E(t) &= \beta(t)S(t)(I_1(t)(\theta_1 + \xi(1 - \theta_1)) + \xi I_u(t))/N - E(t)/L \\
 \frac{d}{dt}I_1(t) &= E(t)/L - I_1(t)/D_1 + Imported(t) \\
 \frac{d}{dt}I_u(t) &= (1 - \theta_1)I_1(t)/D_1 - I_u(t)/D_u \\
 \frac{d}{dt}H(t) &= \theta_1 I_1(t)/D_1 - 2\theta_2 H(t)/D_D - 2(1 - \theta_2)H(t)/D_R \\
 \frac{d}{dt}Recovered_0(t) &= 2(1 - \theta_2)H(t)/D_R - 2Recovered_0(t)/D_R \\
 \frac{d}{dt}Recovered(t) &= 2Recovered_0(t)/D_R \\
 \frac{d}{dt}Dead_0(t) &= 2\theta_2 H(t)/D_D - 2Dead_0(t)/D_D \\
 \frac{d}{dt}Dead(t) &= 2Dead_0(t)/D_D
 \end{aligned} \tag{1}$$

Here $N = 1,400,050,000$ is the total population size in mainland China and is assumed to be constant during the outbreak. The model also includes imported cases (i.e., $Imported(t)$ in equation for I_1) from outside China as reported by Chinese National Health commission since 24th March 2020, which is likely to help with improving the reliability of our estimates of the transmissibility at the late stage of the epidemic and the effectiveness of intervention measures. Including the numbers of imported cases in the model means that we implicitly include people entering China at the latent stage. For example, due to the stronger and improved testing and tracing services at the late stage of the epidemic, they would have been put into isolation for 2 weeks after entering China and would have been detected as cases during that time.

Consistent with our assumption that people mix randomly, the force of infection (the rate at which susceptibles are infected per unit time) is assumed to be proportional to the number of infectious people at time t , which is given by the following expression in the equation for the rate of change in the number of susceptible individuals (i.e. $dS(t)/dt$):

$$(I_1(t)(\theta_1 + \xi(1 - \theta_1)) + \xi I_u(t))$$

This expression comprises three terms: $I_1(t)\theta_1$ —the number of infectious people at time t who have severe symptoms and will be detected and hospitalised, $I_1(t)(1 - \theta_1)$ —the number of infectious people who have mild or no symptoms and will not be detected and therefore will transfer to the compartment I_u , and $I_u(t)$ —the number of infectious people who have not been detected and will remain so. The parameter ξ is the relative infectiousness of people with undetected infections compared to confirmed cases.

To model the potential changes in the transmission rate (the rate at which two specific people come into effective contact per unit time) due to the combined and dramatic control measures taken by Chinese government, we simply assume that the transmission coefficient, β , differed before and after they are introduced as shown in the Eq. (2)

$$\beta(t) = \begin{cases} \beta_a & t \leq \tau_\beta \\ \beta_b & t > \tau_\beta \end{cases} \quad (2)$$

(c.f.^{3,41}). Here τ_β is the time point when the transmission rate changed as a result of the interventions and will be estimated in our analyses.

In addition, the limited knowledge and unclear definition of COVID-19 in the early stages of the outbreak means that the proportion of cases that were detected and reported was probably low and increased as knowledge about COVID-19 improved and the availability of advanced techniques to test SARS-CoV-2 increased¹². In particular, case-finding, diagnosis and reporting have sped up since 20th January 2020. Local governments across China encouraged and supported routine screening and quarantine of travellers from Hubei Province in order to identify COVID-19 infections as early as possible⁶. To reflect these changes, we assume that the proportion of people that were detected and reported (θ_1) (or rate of case ascertainment) varied with time as follows

$$\theta_1 = \begin{cases} \theta_{1,a} & t \leq \tau_\theta \\ \theta_{1,b} & t > \tau_\theta \end{cases} \quad (3)$$

Here τ_θ is the time point when the rate of case ascertainment changed.

It is also likely that the case fatality rate changed over time, as treatment improved with increasing provision of medical resources. For simplicity, we introduce a time point τ_F so that before that time, the confirmed case fatality rate is $\theta_{2,a}$ and after this time point it becomes $\theta_{2,b}$, as follows:

$$\theta_2 = \begin{cases} \theta_{2,a} & t \leq \tau_F \\ \theta_{2,b} & t > \tau_F \end{cases} \quad (4)$$

$\theta_{2,a}$ and $\theta_{2,b}$ and τ_F will be estimated in our analyses.

Although there may have been some sporadic cases of COVID-19 that might not have had the chance to be hospitalized and so would die at home, especially at the early stage of the outbreak, our assumed flow of confirmed COVID-19 patients should approximate actual procedures during the outbreak in mainland China reasonably well⁴⁵. In this study we fix the average durations from onset of symptoms to death and to hospital discharge at $(D_1 + D_D =)$ 17.8 days and $(D_1 + D_R =)$ 22.6 days, respectively, consistent with the estimates⁹ from the China outbreak data. To allow the durations from hospitalization to death and from hospitalization to recovery to follow the Gamma distribution rather than the usual exponential distribution⁴⁶, we introduce the intermediate compartments *Dead_0*, and *Recovered_0*. As we use (mostly) data of reported dates, the time-event-length (duration) should be thought to include such reporting delay, which is not explicitly treated in this study⁶.

The basic reproduction number R_0 , which is defined as the average number of secondary infectious people generated by an infectious person introduced into a completely susceptible population, is an important quantity which characterises the transmissibility of infectious agents³⁶. We can obtain the expression for R_0 by considering the situation without imported cases and obtain the steady-state solution of the equation system (1), with its S^* (the size of the population susceptible to infection at equilibrium) being given by

$$S^* = N \frac{1/D_1}{\beta_a[\theta_{1,a} + \xi(1 - \theta_{1,a})] + \xi(1 - \theta_{1,a})D_u/D_1}$$

At the steady state, $R_0 \times (S^*/N) = 1$ ^{36,37}, the above expression gives the basic reproduction number at the early stage as follows

$$R_{0,1} = (\theta_{1,a}D_1 + \xi(1 - \theta_{1,a})(D_1 + D_u))\beta_a \quad (5a)$$

Similarly, at the late stage after the time point defined as $\max(\tau_\beta, \tau_\theta)$, the effective reproduction number under interventions is

$$R_{0,2} = (\theta_{1,b}D_1 + \xi(1 - \theta_{1,b})(D_1 + D_u))\beta_b \quad (5b)$$

From Eq. (5a), the contribution from undetected infections to the transmissibility at the early stage is

$$R_{0,1}(\text{undetected cases}) = \xi(1 - \theta_{1,a})(D_1 + D_u)\beta_a \quad (5c)$$

Here $D_1 + D_u$ is the average infectious period of the infected people who were not detected. The parameter ξ is introduced to measure the relative infectiousness of undetected infections to confirmed cases, and it will take the value of 1.0 (i.e., both undetected and confirmed infection are of the same infectiousness). The above equations for the basic reproduction number can also be obtained using the Next Generation Matrix approach described in Diekmann et al.⁴⁷ (see SI Sect. 5).

To see how the relative infectiousness affects our findings, estimates of R_0 and other model parameters for $\xi = 1/2, 1/3$ are given in Supplementary Table S2.1.

Initial seeding. In this study we model the transmission process from 1st December 2019 with initial infections of $I_1(0) = I_0$ which is to be estimated from model fitting to data (see below). The initial values of the other variables were obtained from I_0 and the total population size by assuming that the epidemic around 1st December 2019 was at its early exponential growth stage, as described in SI Sect. 1. In contrast, Wu et al.²¹ assumed that the epidemic during 1st–31st December 2019 was seeded by a constant zoonotic force of infection that caused 86 cases (twice the number of 43 confirmed cases with zoonotic exposure) per day before market closure on 1st January 2020. Kucharski et al.⁶ assumed the outbreak started with a single infectious case or 10 cases on 22nd November 2019.

Inference method. We denote the set of model parameters to be inferred as $\Theta = \{I_0, \beta_a, \beta_b, \tau_\beta, \theta_{1,a}, \theta_{1,b}, \tau_\theta, \theta_{2,a}, \theta_{2,b}, \tau_\beta, D_1, D_u\}$ which is listed in Table 2. For each set of parameter values, the Runge–Kutta fourth order method was used to solve the model equations and to obtain model predicted time series of infections, confirmed cases, deaths, and recovered cases. In the inference of model parameters, directly observed dataset of confirmed/hospitalized/reported cases (denoting it as HOS for short in the following), Death and Recovery are used as illustrated in the following. To capture the large dispersion in the daily numbers of these observations, the negative binomial likelihood function was assumed. The likelihood for number $x^C(t)$ of observations on day t is given as

$$l(x^C(t)|\Theta, \eta^C) = \frac{\Gamma(x^C(t) + r^C(t))}{\Gamma(r^C(t))\Gamma(x^C(t) + 1)} \left(\frac{1}{\eta^C}\right)^{r^C(t)} \left(1 - \frac{1}{\eta^C}\right)^{x^C(t)} \quad (6a)$$

where

$$r^C(t) = \frac{\mu^C(t)}{\eta^C - 1} \quad (6a)$$

Here η^C is the dispersion parameter, which is estimated and $\mu^C(t)$ are the predictions of the number of hospitalisations, deaths or newly recovered people, as appropriate on day t from synthesis model (1). Superscript C represents three different datasets: HOS, Death and Recovery.

Special attention was paid to the extra high daily number of cases (15,152) on 12th February 2020 (day 74 from 1st December 2019) due to the change in the case definition in Hubei province^{12,48}. In principle, most of these cases might have been accumulated over the days before 12th February 2020. To deal with this complexity, only the cumulative numbers of cases on the 12th February and daily numbers of cases after that will be used in the model inference. Let the reported daily number of HOS be represented by $x(1), x(2), \dots, x(T)$, with T being the number of days from 1st December 2019 to 21st April 2020. The likelihood of the cumulative number of cases on 12th February $X = \sum_{t=1}^{74} x(t)$ is assumed to be

$$L_{74} = \frac{\exp(-M)}{X!} M^X \quad (7)$$

Here $M = \sum_{t=1}^{74} \mu^{\text{HOS}}(t)$ represents the cumulative number of confirmed cases predicted by model. Assuming that the observed daily number of deaths: $y(1), y(2), \dots, y(T)$ and daily number of recovered people: $z(1), z(2), \dots, z(T)$ are conditionally independent, the total likelihood given model parameters Θ is

$$L(\Theta, \eta^{\text{HOS}}, \eta^{\text{Death}}, \eta^{\text{Recovery}}) = L_{74} \times \prod_{t=75}^T l(x(t)|\Theta, \eta^{\text{HOS}}) \times \prod_{t=1}^T l(y(t)|\Theta, \eta^{\text{Death}}) l(z(t)|\Theta, \eta^{\text{Recovery}}) \quad (8)$$

Parameter inference: We assume the uninformative prior distributions $f(\Theta)$ which are uniform for parameters (Table 2). Employing a Bayesian framework through the combination of the prior distribution $f(\Theta)$ and the likelihood $L(\Theta, \eta^{\text{HOS}}, \eta^{\text{Death}}, \eta^{\text{Recovery}}; x, y, z)$, the posterior distribution can be obtained by Markov Chain Monte Carlo simulations (MCMC)⁴⁹. From these samples, we obtain the medians and the 95% confidence intervals for the model parameters. The posteriors of the model parameters will provide the estimates of the transmissibility and the severity of SARS-CoV-2 in mainland China and the effects on transmissibility of control measures implemented by the Chinese government. The same inference method will also be used when modelling the outbreak in epicentre Hubei province and Wuhan city.

To explore the effect of the timing of control measures, we calculate the numbers of infections, cases and deaths that would have occurred if the control measures had been implemented 1, 2, or 3 weeks earlier than they had been, or if they had been implemented 1, 2, or 3 weeks later than they had been. To do this, we simply change the start times at which the reporting, transmission and death rates changed while the transmission parameters, ascertainment rate and fatality rate remain unchanged.

Received: 27 July 2020; Accepted: 6 January 2021

Published online: 29 January 2021

References

- Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) (2020).
- Lai, S. *et al.* Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. *Nature* **585**, 410–413. <https://doi.org/10.1038/s41586-020-2293-x> (2020).
- Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**(638–642), 2020. <https://doi.org/10.1126/science.abb6105> (2020).
- World Health Organization. Archived: WHO Timeline—COVI-19. 2020. <https://www.who.int/news-room/detail/27-04-2020-who-timeline-covid-19>.
- Yang, Z. *et al.* Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* **12**(3), 165–174. <https://doi.org/10.21037/jtd.2020.02.64> (2020).
- Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **20**(5), 553–558. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4) (2020).
- Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**, 489–493. <https://doi.org/10.1126/science.abb3221> (2020).
- Yang, Y. *et al.* Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *medRxiv* <https://doi.org/10.1101/2020.02.10.20021675> (2020).
- Verity, R. *et al.* Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7) (2020).
- Read, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A. & Jewell, C. P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* <https://doi.org/10.1101/2020.01.23.20018549.thisversionpostedJanuary28,2020> (2020).
- Zhao, S. *et al.* Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *Int. J. Infect. Dis.* **92**, 214–217. <https://doi.org/10.1016/j.ijid.2020.01.050> (2020).
- Tsang, T. K. *et al.* Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *Lancet Public Health* **5**(5), e289–e296. [https://doi.org/10.1016/S2468-2667\(20\)30089-X](https://doi.org/10.1016/S2468-2667(20)30089-X) (2020).
- The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristic of an outbreak of 2019 novel coronavirus disease (COVID-19)—China, 2020. *China CDC Wkly.* <https://doi.org/10.46234/ccdcw2020.032> (2020).
- 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**, 10. <https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180> (2020).
- https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_on_cruise_ships. Accessed 01 Apr 2020.
- Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (2019-nCoV) outbreak. *Science* **368**, 395–400. <https://doi.org/10.1126/science.aba9757> (2020).
- Imai, N. *et al.* Report 3(imperial): Transmissibility of 2019-nCoV. <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-transmissibility-25-01-2020.pdf> (2020).
- Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207. <https://doi.org/10.1056/NEJMoa2001316> (2020).
- Liu, T. *et al.* Transmission dynamics of 2019 novel coronavirus (2019-nCoV). *bioRxiv*. (published online Jan 26.) (preprint). doi: <https://doi.org/10.2139/ssrn.3526307> (2020)
- Riou, J. & Althaus, C. L. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill.* **25**, 4. <https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058> (2020).
- 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* **395**, 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9) (2020).
- Sun, G.-Q. *et al.* Transmission dynamics of COVID-19 in Wuhan, China: effects of lockdown and medical resources. *Nonlinear Dyn.* <https://doi.org/10.1007/s11071-020-05770-9> (2020).
- Linton, N. M. *et al.* Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *J. Clin. Med.* **17**(9), 2. <https://doi.org/10.3390/jcm9020538> (2020).
- Sanche, S. *et al.* High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**, 1470–1477. <https://doi.org/10.3201/eid2607.200282> (2020).
- Xu, X. W. *et al.* Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ* **368**, m606. <https://doi.org/10.1136/bmj.m606> (2020).
- Zhang, J. *et al.* Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.* **20**(7), 793–802. [https://doi.org/10.1016/S1473-3099\(20\)30230-9](https://doi.org/10.1016/S1473-3099(20)30230-9) (2020).
- Jing, Q.-L. *et al.* Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 1141–1150. [https://doi.org/10.1016/S1473-3099\(20\)30471-0](https://doi.org/10.1016/S1473-3099(20)30471-0) (2020).
- Siordia, J. A. Jr. Epidemiology and clinical features of COVID-19: A review of current literature. *J. Clin. Virol.* **127**(2020), 104357. <https://doi.org/10.1016/j.jcv.2020.104357> (2020).
- Deng, X. *et al.* Case fatality risk of the first pandemic wave of novel coronavirus disease 2019 (COVID-19) in China. *Clin Infect. Dis.* <https://doi.org/10.1093/cid/ciaa578> (2020).
- Wu, J. T. *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**(4), 506–510. <https://doi.org/10.1038/s41591-020-0822-7> (2020).
- Russell, T. W. *et al.* Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Euro Surveill.* **25**(12), 2000256. <https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256> (2020).
- National Health Commission of the People's Republic of China. Notice of the COVID-19 on the revision of the number of confirmed cases of pneumonia and the number of deaths in COVID-19, Wuhan City. <http://www.nhc.gov.cn/xcs/yqtb/202004/6f8eb064959f4ab7b56fe03236920be1.shtml> (2020). Accessed 17 March 2020 (in Chinese).
- He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**(5), 672–675. <https://doi.org/10.1038/s41591-020-0869-5> (2020).
- Savvides, C. & Siegel, R. Asymptomatic and presymptomatic transmission of SARS-CoV-2: A systematic review. *medRxiv* <https://doi.org/10.1101/2020.06.11.20129072> (2020).
- Xu, X. *et al.* Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. *Nat. Med.* **26**(8), 1193–1195. <https://doi.org/10.1038/s41591-020-0949-6> (2020).
- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991).
- Vynnycky, E. & White, R. G. *An Introduction to Infectious Disease Modelling* (Oxford University Press, Oxford, 2010).
- Ferguson, N. M. *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 16 March 2020 Imperial College modelling report. <https://doi.org/10.25561/77482> (2020)

39. Wallinga, J. & Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **160**, 509–516. <https://doi.org/10.1093/aje/kwh255> (2004).
40. Jia, N. *et al.* Case fatality of SARS in mainland China and associated risk factors. *Trop. Med. Int. Health* **14**(suppl 1), 21–27. <https://doi.org/10.1111/j.1365-3156.2008.02147.x> (2009).
41. Zhang, X.-S. *et al.* Estimating and modelling the transmissibility of Middle East Respiratory Syndrome CoronaVirus during the 2015 outbreak in the Republic of Korea. *Influenza Other Respir. Viruses* **11**(5), 434–444. <https://doi.org/10.1111/irv.12467> (2017).
42. Oh, M. D. *et al.* Middle East respiratory syndrome: what we learned from the 2015 outbreak in the Republic of Korea. *Korean J. Int. Med.* **232**(2), 233–246. <https://doi.org/10.3904/kjim.2018.031> (2018).
43. Birrell, P. *et al.* Forecasting the 2017/2018 seasonal influenza epidemic in England using multiple dynamic transmission models: a Case Study. *BMC Public Health* **20**, 486. <https://doi.org/10.1186/s12889-020-8455-9> (2020).
44. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* **172**, 577–582. <https://doi.org/10.7326/M20-0504> (2020).
45. Yan, L. & Li, Y. Identification and treatment strategies for patients with COVID-19. *New Med.* **51**(3), 161–167. <https://doi.org/10.3969/j.issn.0253-9802.2020.03.001> (2020).
46. Wearing, H. J., Rohani, P. & Keeling, M. J. Appropriate models for the management of infectious diseases. *PLoS Med.* **2**(7), e174. <https://doi.org/10.1371/journal.pmed.0020174> (2005).
47. Chinese National Health Commission release of 7th edition of case definitions. <http://www.nhc.gov.cn/zycj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml>. Accessed 12 March 2020 (in Chinese).
48. Diekmann, O., Heesterbeek, J. A. P. & Roberts, M. G. The construction of next generation matrices for compartmental epidemic models. *J. R Interface* **7**, 873–885 (2010).
49. Birrell, P. *et al.* Bayesian modelling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proc. Natl. Acad. Sci. USA* **108**, 18238–18243. <https://doi.org/10.1073/pnas.1103002108> (2011).

Acknowledgements

This work is jointly supported by Public Health England, UK and Innovative Research Team of Yunnan Province (2019(6)), China. The code of the synthesis model and data used in this study are available on request from the corresponding authors. X.-S.Z. would like to dedicate the paper to the momery of Professor Zi-Tian Yang of Yunnan University who passed away on 02/09/2019 for his teachings and encouragement.

Author contributions

Conceived and designed the mathematical model and analysis: X.-S.Z., E.V., A.C., D.D.A. and W. L.; collated data: Z.C.; performed the study: X.-S.Z. and W.L.; wrote the paper: X.-S.Z. and E.V.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-81985-z>.

Correspondence and requests for materials should be addressed to X.-S.Z. or W.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021