On a Statistical Transmission Model in Analysis of the Early Phase of COVID-19 Outbreak

Yifan Zhu[1] · Ying Qing Chen[1]

Abstract

Since December 2019, a disease caused by a novel strain of coronavirus (COVID-19) had infected many people and the cumulative confirmed cases have reached almost 180,000 as of 17, March 2020. The COVID-19 outbreak was believed to have emerged from a seafood market in Wuhan, a metropolis city of more than 11 million population in Hubei province, China. We introduced a statistical disease transmission model using case symptom onset data to estimate the transmissibility of the early-phase outbreak in China, and provided sensitivity analyses with various assumptions of disease natural history of the COVID-19. We fitted the transmission model to several publicly available sources of the outbreak data until 11, February 2020, and estimated lock down intervention efficacy of Wuhan city. The estimated R_0 was between 2.7 and 4.2 from plausible distribution assumptions of the incubation period and relative infectivity over the infectious period. 95% confidence interval of R_0 were also reported. Potential issues such as data quality concerns and comparison of different modelling approaches were discussed.

✉ Yifan Zhu
yzhu2@fredhutch.org

Ying Qing Chen
yqchen@fredhutch.org

[1] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109, USA

🖄 Springer

1 Introduction

Growing number of cases from an outbreak caused by novel coronavirus (COVID-19) have been found since December 2019, and has reached almost 180,000 confirmed infections as of 17, March 2020 [16], including 7426 deaths globally. It has been declared to be pandemic by WHO on March 11. The outbreak size has surpassed infections caused by two other major coronavirus to date, severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV). It appears the transmission rate of the COVID-19 is the highest in all circulating human CoVs [17], and the disease is still spreading at fast pace albeit Chinese authorities taking very stringent measures including city lock down with suspending all travel by air, train and highways. The lock down of Wuhan city, the epicentre of this outbreak in Hubei province, took place at 10 am local time on 23, January, 2020. The same strict interventions were used in several other cities in Hubei province and cities in other provinces with high count of secondary input cases. The transmission rate was likely accelerated by the national migration event *ChunYun* before lunar new year, with enormous traffic volumes from and to Wuhan and other major cities in China.

As many aspects of COVID-19 is not yet fully understood, it is important to evaluate its transmissibility during the initial phase of outbreak. Epidemiological reviews, case reports and contact risk histories available provide crucial information for establishing dynamic transmission models to evaluate the basic reproduction number (R_0) of COVID-19. R_0 is defined as the average number of secondary cases in an uninfected population that an diseased person would infect, during the persons' whole infectious window. It is used as a measure of transmission rate for infectious diseases, and containment of an outbreak requires to reduce effective R_0 to below 1. In this paper, we fitted a Poisson transmission model with either symptomatic infection only or potential asymptomatic infection assumptions, together with current knowledge of the disease's natural history of COVID-19, to the case symptom onset time reports available from China CDC website [2] as well as report delay time augmented onset report from daily confirmed case ascertainment available at World Health Organization (http://who.int) [16] and National Health Commission of the People's Republic of China website (http://www.nhc.gov.cn/).

Researchers from several institutions had reported initial modelling results of the early-phase outbreak of COVID-19. Read et al. [13] and Wu et al. [17] relied on traffic data across Chinese cities and international flights incorporated in deterministic SEIR transmission model of infection. Disease natural history was based on early hypothesis of the COVID-19 as well as that of SARS and MERS. Li et al. [9] provided detailed disease onset record of the first 425 confirmed cases and estimates of disease incubation period and serial interval from the contact tracing, and applied these knowledge into renewal equations epidemic model to estimate R_0. Zhao et al. [18] applied an exponential growth model to fit report rate-adjusted lab-confirmed case report data from Chinese authorities.

The estimated R_0's were generally above 2 in all these modelling work, however, we have noticed estimates with value as high as 3.8 [13] and 3.58 [18]. The wide range of estimated R_0 is likely caused by insufficient high-quality data available and over-simplified model assumptions. We expect the estimates from various approaches to become relatively close with better understanding of the emerging disease and availability of incoming high-quality outbreak data.

The paper is organized as follows: Sect. 2 describes the data structure, assumption of COVID-19 natural disease history and likelihood-based Poisson model for evaluating transmissibility. Section 3 explains the data source for analyses, the obtained estimates of R_0 and impact of different model assumptions using sensitivity analyses. Section 4 applies our models to updated onset data following the first submission. Section 5 summarizes the findings from the transmission model, discussed advantages and drawbacks of the current approach, and provided directions for potential model expansion as the outbreak progresses.

2 The Statistical Transmission Model

Consider the outbreak observation was made on a daily basis with integer day t comprised of the continuous time $(t-1, t]$. Assume the available outbreak data were from day 1 and ended at day T. For each infected individual i, $i = 1, \dots, N$, let $\tilde{t}_i$ be the symptom onset time, and the corresponding infection time $\hat{t}_i$ that leads to the onset, where $\hat{t}_i < \tilde{t}_i$. The case confirmation time $\bar{t}_i$ is greater than $\tilde{t}_i$ due to several reasons, such as the lack of understanding during the emerging phase of novel disease outbreak and lab processing and reporting time delay.

We provide an hypothetical example of case event time review and contact tracing process to help readers understand the natural history of an infectious disease. The incubation period $(\hat{t}, \tilde{t})$ is the time to develop symptoms after being infected. As
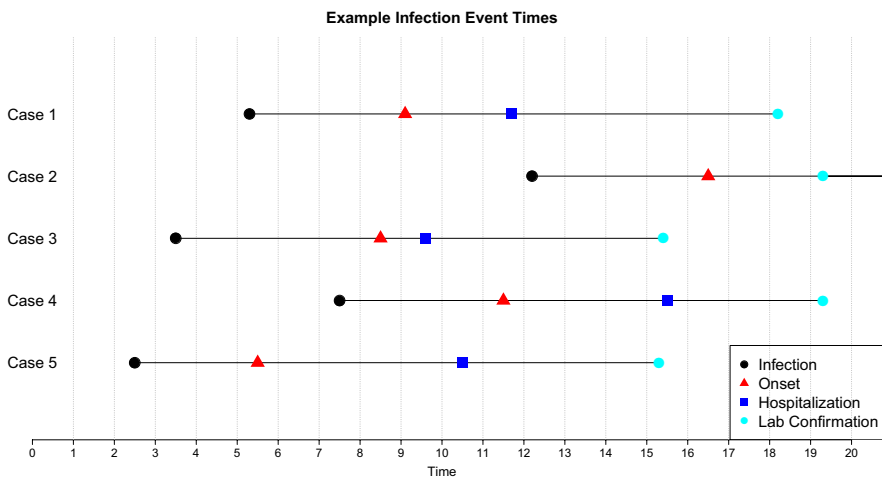

Fig. 1 Illustrative examples for infection, symptom onset, hospitalization and virus lab confirmation time

illustrated in Fig. 1, the lengths of incubation period are 4, 4, 5, 4 and 3 days for cases 1 to 5. The accurate estimate of these event times from contact history evaluation is important for understanding the natural history of disease. For instance, if we know that there was close contact between cases 1, 2 and cases 4, 5, but case 3 had no contact with all other cases, a patient would be unlikely to transmit the disease before symptom onset. Otherwise, if cases 3 and 4 were in close contact and after hospitalization case 3 was isolated from case 4, it would indicate that the infected case may be infectious before showing any symptom. Such conclusions are, however, very challenging during emerging outbreak investigations, as $\hat{t}_i$'s are often unavailable for most cases and the contact history between cases can be complex. On the other hand, there is usually delay between symptom onset time, hospitalization time and lab confirmation time. The incidence number of confirmed cases available as daily report usually consists of the lab confirmation time, thus is less accurate for recent infections.

2.1 Assumption of Disease Natural History

Exact $\hat{t}$ is usually difficult to determine for most infectious pathogens, necessitating the reasonable assumption of incubation period distribution. We assume the length incubation period of COVID-19 to be beta-distributed on $(\text{Inc}_{\min}, \text{Inc}_{\min} + \Delta_{\text{inc}})$, where $\text{Inc}_{\min}$ and $\text{Inc}_{\min} + \Delta_{\text{inc}}$ are the minimum and maximum lengths of the incubation period. The probability of observing symptom onset at t given infection at $\hat{t}$ is

$$p_{\text{inc}}(t \mid \hat{t}) = f_{\text{beta}}\left(\frac{t - \hat{t} - \text{Inc}_{\min}}{\Delta_{\text{inc}}} \mid \alpha_{\text{inc}}, \beta_{\text{inc}} \right),$$
$$t \in (\hat{t} + \text{Inc}_{\min}, \hat{t} + \text{Inc}_{\min} + \Delta_{\text{inc}}),$$

where $f_{\text{beta}}(\cdot \mid \alpha_{\text{inc}}, \beta_{\text{inc}})$ is the density function of beta-distribution. The mean incubation length is $\text{Inc}_{\min} + \frac{\Delta_{\text{inc}} \alpha_{\text{inc}}}{\alpha_{\text{inc}} + \beta_{\text{inc}}}$. Density functions from other distributions such as log-normal [9], Weibull [5] or gamma distributions may be used as well, as long as these fit the observed incubation periods of COVID-19. We used beta-distribution since it has finite support that reflects the shortest/longest incubation period without the need to truncate the right tail for distribution with an infinite support.

We also assume that latent period (time length between infection and start of infectiousness) is the same as the incubation period, thus the case is only able to infect other people after symptom onset. The relative infectivity during the infectious period $(\tilde{t}, \tilde{t} + \Delta_{\text{inf}})$ is assumed to be

$$p_{\text{inf}}(t \mid \tilde{t}) = f_{\text{beta}}\left(\frac{t - \tilde{t}}{\Delta_{\text{inf}}} \mid \alpha_{\text{inf}}, \beta_{\text{inf}} \right), \quad t \in (\tilde{t}, \tilde{t} + \Delta_{\text{inf}}).$$

2.2 Poisson Transmission Model for Daily Disease Onset Record

Let γ be the average hazard of secondary infection along the infectious period. The overall infection hazard from all infected cases in the community is

$$\hat{\lambda}(t) = \sum_{i\,:\,t\in[\tilde{t}_i,\tilde{t}_i+\Delta_{\mathrm{inf}}]} p_{\mathrm{inf}}(t|\tilde{t}_i)\Delta_{\mathrm{inf}}\gamma. \tag{1}$$

With incubation period distribution $p_{\mathrm{inc}}(t|\hat{t})$, the intensity of disease onset at time t is thus

$$\tilde{\lambda}(t) = \int_{t-\mathrm{Inc}_{\max}}^{t-\mathrm{Inc}_{\min}} \hat{\lambda}(\tau)p_{\mathrm{inc}}(t|\tau)\mathrm{d}\tau.$$

Let $\tilde{\Lambda}(t) = \int_{s-1}^{s} \tilde{\lambda}(s)\mathrm{d}s$ be the daily cumulative intensity rate of onset. Assume the number observed symptom onsets for day t is Poisson distributed with intensity $\tilde{\Lambda}(t)$, the likelihood of observing onset record $\tilde{\mathbf{N}} = (\tilde{N}_1,\dots,\tilde{N}_T)$ is

$$p(\tilde{\mathbf{N}}|\gamma,\theta_{\mathrm{inc}},\theta_{\mathrm{inf}}) = \prod_{t=1}^{T} \frac{e^{-\tilde{\Lambda}(t)}\tilde{\Lambda}(t)^{\tilde{N}_t}}{\tilde{N}_t!}. \tag{2}$$

R_0 by definition is the expected number of infections caused by one case, thus can be estimated as $\hat{R}_0 = \hat{\gamma}\Delta_{\mathrm{inf}}$ using the maximum likelihood estimate (MLE) for average hazard of secondary infection $\hat{\gamma}$. The likelihood approach is similar to [3], however, instead of making assumptions about the serial interval (time between onset of the index case and the secondary case), we differentiate the contribution of incubation period and relative infectivity in the transmission model.

2.3 Probable Infectiousness Before Symptom Onset

We previously assumed that the latent period is the same as the incubation period, however, with COVID-19 there had been reports of confirmed cases showing mild or even no symptoms who infected their family members. Thus, if we instead assume the infectiousness already develops since time of infection, the modified relative infectivity is $p_{\mathrm{inf}}^*(t|\hat{t}) = f_{\mathrm{beta}}(\frac{t-\hat{t}}{\Delta_{\mathrm{inf}}}|\alpha_{\mathrm{inf}}^*,\beta_{\mathrm{inf}}^*)$, $t \in (\hat{t},\hat{t}+\Delta_{\mathrm{inf}}^*)$. Consequently Eq. (1) is modified as follows:

$$\hat{\lambda}(t) = \sum_{i\,:\,t\in[\hat{t}_i,\hat{t}_i+\Delta_{\mathrm{inf}}^*]} p_{\mathrm{inf}}^*(t|\hat{t}_i)\Delta_{\mathrm{inf}}^*\gamma. \tag{3}$$

Thus the likelihood (2) must be calculated with both the onset record $\tilde{\mathbf{N}}$ and the infection record $\hat{\mathbf{N}} = (\hat{N}_1,\dots,\hat{N}_T)$. We may treat $\hat{\mathbf{N}}$ as missing data that are linked to $\tilde{\mathbf{N}}$ through $p_{\mathrm{inc}}(t|\hat{t})$. Given known $p_{\mathrm{inc}}(t|\hat{t})$ multiple imputation approach can produce MLE estimates for γ with repeated augmented $\hat{\mathbf{N}}^{(k)}$, and summarizing over the $k = 1,\dots,K$ estimates from complete data $(\tilde{\mathbf{N}}, \hat{\mathbf{N}}^{(k)})$. The model with the asymptomatic infection assumption is referred as modified Poisson transmission model.

2.4 City Lock Down Intervention

We assume the lock down intervention had an immediate effect to reduce the disease transmissibility, plus another lasting effect that strengthens following the lock down. The intervention effect can be modelled as

$$\theta(t) = \left(\theta_1 e^{-\theta_2(t-t_0)}\right)^{I(t>t_0)}, \tag{4}$$

where t_0 is the start time of intervention, $\theta_1 \in [0, 1]$ is the immediate infectivity reduction and $\theta_2 > 0$ is the lasting effect rate. To incorporate the intervention effect in the previous transmission model Eq. (1) will be modified as

$$\hat{\lambda}(t) = \sum_{i : t \in [\tilde{t}_i, \tilde{t}_i + \Delta_{\text{inf}}]} p_{\text{inf}}(t|\tilde{t}_i) \Delta_{\text{inf}} \gamma \theta(t). \tag{5}$$

The intervention parameters θ_1, θ_2 can be estimated using MLE and their Wald confidence intervals can be obtained via inverting the observed Fisher information from the likelihood.

3 Analysis of the Outbreak's Early Phase

We fitted the proposed transmission models to infer the basic reproduction number of COVID-19 in China (including Hong Kong, Macao and Taiwan) using the disease onset sizes recorded daily fro1, December 2019 to 23, January 2020. Sources of COVID-19 onset data were (1) confirmed and suspected daily onset record from China CDC Epidemic update and risk assessment of 2019 Novel Coronavirus [2]; (2) augmented daily onset record using confirmed cases reports from National Health Commission of the People's Republic of China using confirmation delay distribution [5]. We restricted case onset data until 23 January 2020 to be fitted, due to the following considerations: (a) the onset record after the cutoff is likely inaccurate with the delayed confirmation process at the manuscript preparation time; (b) insufficient evidence for transmission reduction effect with the strict lock down quarantine interventions in Greater Wuhan Region implemented since 23–24, January 2020.

Several preliminary estimates of disease natural history and R_0 were available for the COVID-19 outbreak [9, 10, 13, 17, 18]. We assumed $\alpha_{\text{inc}} = 6$ and $\beta_{\text{inc}} = 12$ for the beta-distributed incubation period, with a range of 1 to 14 days. The assumed mean incubation is 5.67 days which is similar to the previous literature on SARS and MERS incubation period [7, 8, 15]. The relative infectivity is assumed to be varying over the infectious period (assumed to be 0–10 days after onset) since the patients may reduce their mobility and contact with others with alarming symptoms such as fever, fatigue and coughing. We performed sensitivity analyses by varying the distribution of incubation period and relative infectivity over time. These assumptions are summarized in Fig. 2, including shorter or longer mean incubation
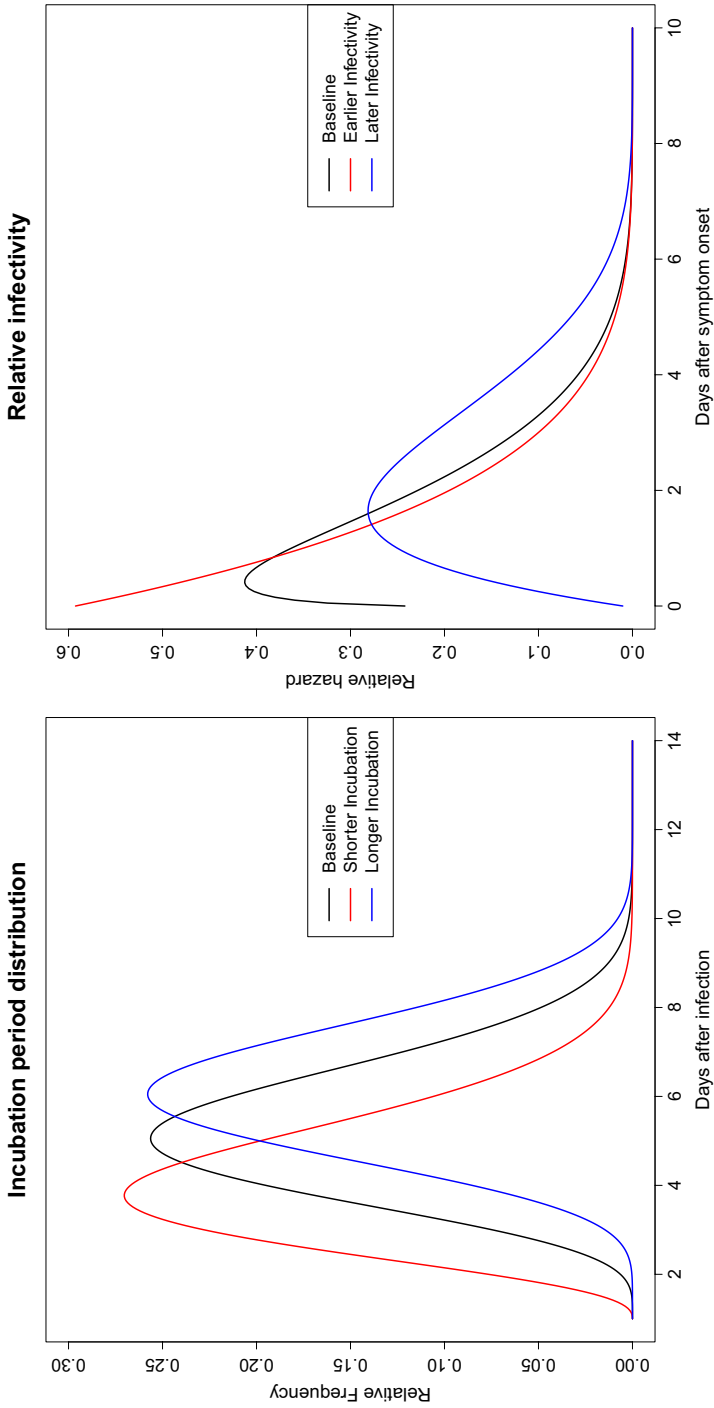
Fig. 2 Assumptions used to model the incubation period length distribution and relative infectivity following patient symptom onset for COVID-19 outbreak in China from 1, December 2019 to 23 January 2020. No asymptomatic infectiousness was possible in this model as cases may infect others only after symptom onset

period as well as different time after symptom where the relative infectivity is the highest. Evaluation of model fit was performed by comparing observed and fitted values of onset records.

The different parameter assumptions for the modified transmission model where cases may show infectiousness before symptom onset include $\Delta_{\text{inf}}^* = 24$ days after $\hat{\iota}$, and the relative infectivity was assumed to be the highest at or after symptom onset. The incubation period assumptions were the same as the original model. Figure 3 summarizes the assumed incubation and relative infectivity distribution for the modified model.

3.1 Data Processing

The daily onset record from "China CDC Epidemic update and risk assessment of 2019 Novel Coronavirus" is publicly available in graphical form [2]. The daily reports of both confirmed and suspected cases were plotted. We used Engauge Digitizer [11] to export the numeric onset record for MLE estimation. To evaluate the impact of potential error during the data exporting process, sensitivity analyses were performed by adding error counts with random generated integers between -3 and 3 with equal probability to the daily number of cases after 1, January 2020. The onset counts in 2019 were available in [9] and should be accurate at the time of our modelling effort.

The second source of daily onset record is to augment onset time $\tilde{\iota}$ from case report confirmation time $\bar{\iota}$ using the confirmation delay distribution. We used a publicly available toolkit in GitHub [5] to generate bootstrap samples of augmented onset record. The delay distributions between case onset times and case ascertainment times were estimated using geometric distribution fitted to "Kudos line list data" [14]. Sensitivity analyses were performed to account for the variability of augmented onset records.

3.2 Results

With the baseline assumptions shown in black curves in Fig. 2, the estimated R_0 and its 95% confidence interval was 2.47(2.39, 2.55) from confirmed cases only, or 2.54(2.49, 2.60) from both confirmed and suspected cases onset records in [2]. We found little impact of the error counts on the estimated model parameters, thus we illustrate our results assuming the error counts were all 0. Assuming no error the reported number of confirmed onset cases was 3442 by 23, January 2020, and number of suspected plus confirmed cases were 8348. Estimated R_0 was 2.31(2.25, 2.38) using confirmation report augmented onset records, from an average of 4940 augmented confirmed cases. Using the repeatedly augmented onset records yielded similar estimates of R_0 and the variance estimate for the 95% CI was obtained by combining the estimates from the repeatedly augmented data, accounting for multiple augmentation variability. Figure 4 shows the time trend of number in observed onset cases, fitted onset cases as well as fitted infected cases each day for all three data sources, respectively. The fitted onset trend matches the observed data well,
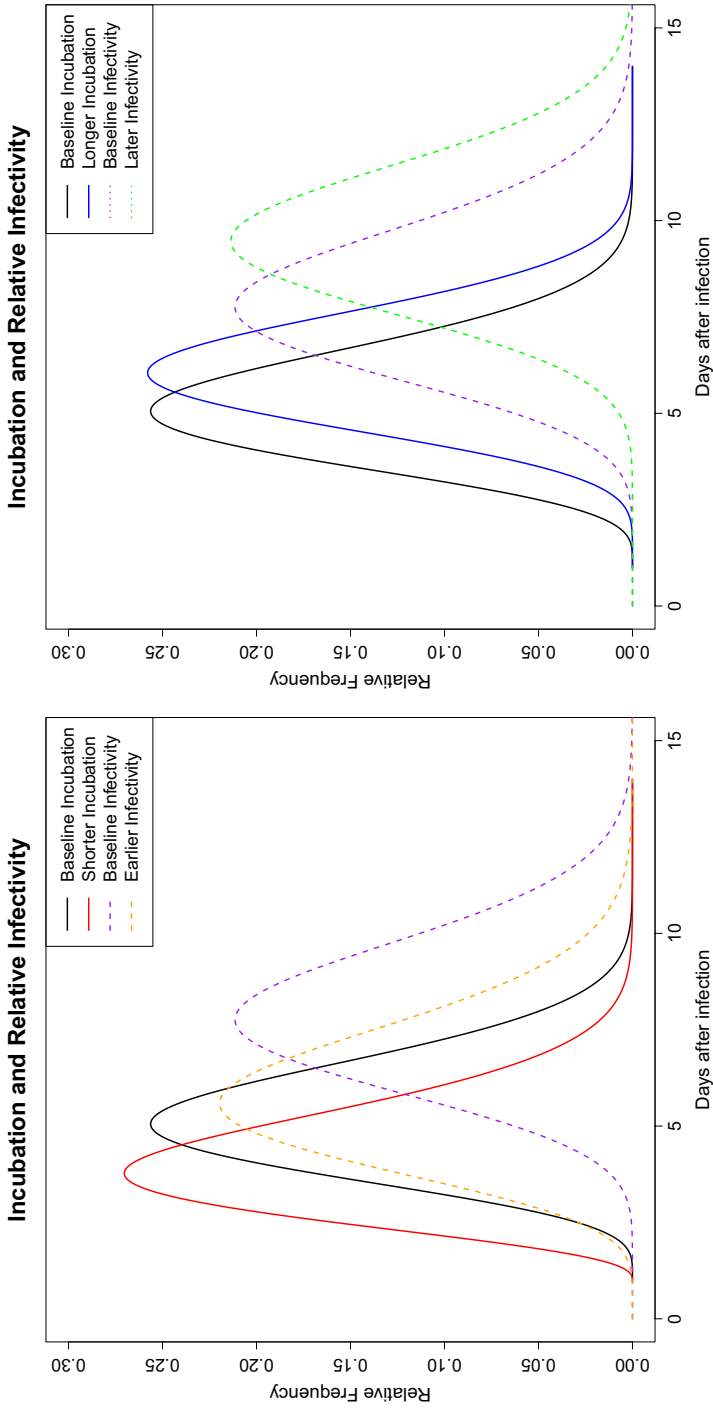
Fig. 3 Modified assumptions used to model the incubation period length distribution and relative infectivity following infection time for COVID-19 outbreak in China from 1, December 2019 to 23, January 2020. Asymptomatic infectiousness exist as the infectious period starts at time of infection

indicating the assumed parameters for incubation and relative infectivity were reasonable.

We summarize the R_0 estimates from sensitivity analyses in Table 1. Under these scenarios, the estimated values of R_0 fall between 2 and 3, which were in general agreement with modelling literature [9, 10, 17, 18] and China CDC report [2], despite various data sources and modelling approaches. The estimated R_0 tended to be higher with longer incubation period or with the later development of infectivity peak compared to the baseline assumptions. Shorter incubation period or earlier development of infectivity peak led to lower estimated R_0, while distribution assumptions with the opposite direction resulted in higher estimated R_0. The fitted versus observed number of daily onset cases were in good agreement in all scenarios for sensitivity analyses, as the assumptions for either incubation period or relative infectivity in these scenarios were generally close to these in the baseline scenario. The transmission model would be expected to have poor fit to the data if the disease natural history assumptions were very different from the truth.

The R_0 estimates from the modified Poisson transmission model were 2.88(2.78, 2.95) from the confirmed cases only, and 2.97(2.90, 3.04) from both confirmed and suspected cases [2]. With augmented onset records, R_0 was estimated to be 2.69(2.61, 2.77). The fitted versus observed symptom onset records were also plotted in Fig. 4. Interestingly, we noticed similarly goodness-of-fit performance from the modified model and that from the original model, but the estimates for R_0 is about 17% higher in the modified transmission model. The difference mainly consisted of the additional infectiousness in the incubation period compared to the original model.

The sensitivity analyses for the modified transmission model are shown in Table 1. The varying incubation period distribution had very little impact on R_0 estimates this time. On the other hand, different positions of relative infectivity peak after infection affected R_0 estimates a lot. Under the scenario where the relative infectivity reached its peak about 9-10 days after infection, estimated R_0 can be very high at 3.5-4.0.

4 Results with Updated Case Onsets in China

Updated data from China CDC became available [12] during the review time of our first submission. The updated data include case symptom onset times until 11, February 2020. Compared to previous data, the updated data include more than 2 week's onset records following lock down of the greater Wuhan Region implemented since January 23–24. We could then estimate the intervention effect parameters with the updated transmission model. Recently published median incubation period in China [4] was estimated to be 4 days with interquartile range of 2–7 days. The shorter incubation period distribution is most close to these findings. Published viral load discoveries in [19] indicated that the levels were highest at or a few days following symptom onset. Observing that the relative infectivity is highly associated to viral load, we felt baseline or earlier relative infectivity assumptions might be more appropriate reflecting recent findings. Thus, we used shorter incubation assumptions
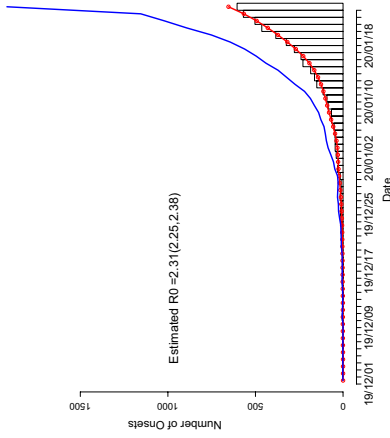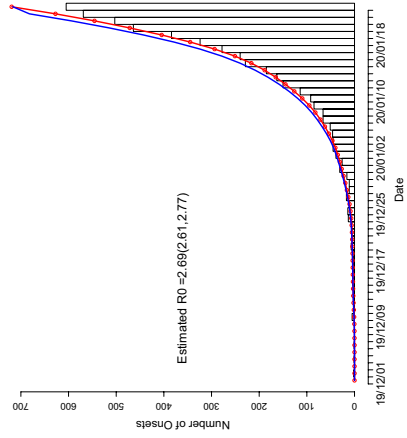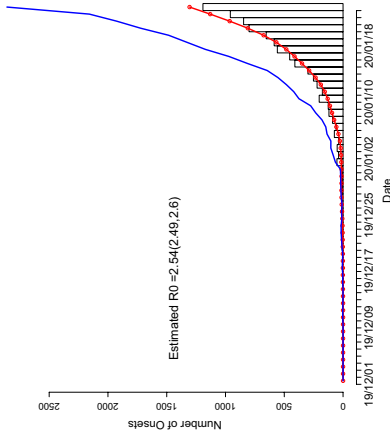
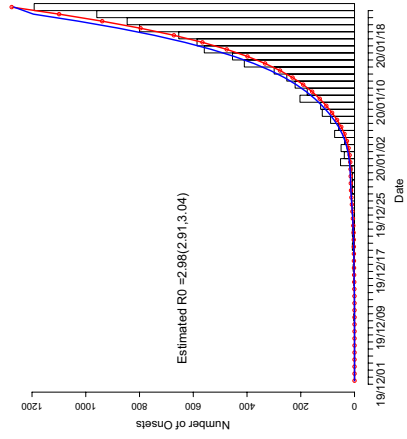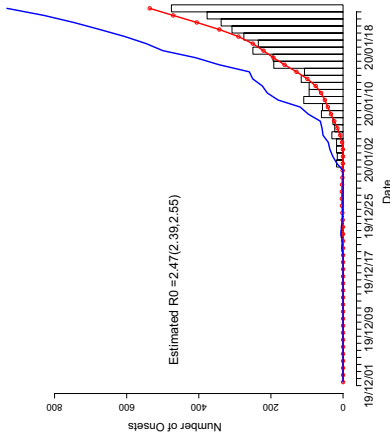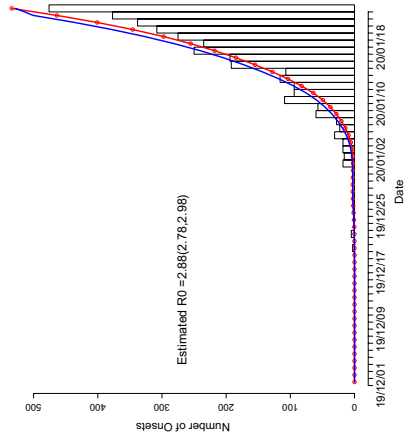◀ **Fig. 4** Observed (shown as bars) and fitted (red curve) number of cases with symptom onset over time, and fitted number of infected cases (blue curve) over time. The upper panels show the results from the model that assumed the incubation period and latent period are the same. The lower panels show the results from the modified model where infectiousness can develop before symptom onset

and either baseline or earlier infectivity assumptions in the transmission models to fit the updated onset data.

One reviewer suggested separating cases inside or outside Wuhan as regional differences may exist in the transmission. This is possible with the updated data thus we fitted models to two data sets separately reflecting the place of onset. However, we felt that the current model assumptions were not well suited to describe the transmission outside Wuhan. About 68.6% cases outside Wuhan had Wuhan related exposure history and bring the disease to other regions in China [12]. These cases should be treated as input cases in the model and more detailed data are required for further analysis. Thus, we only included results with cases within Wuhan in the main text. The results with cases outside Wuhan region can be found in the supplementary material.

Figure 5 shows the model fit to the observed onset curves from 8, December 2019 to 11, February 2020 inside Wuhan Region. Similar to previous results, fitted onset curves matched observed onset curves relatively well. The intervention effect was clearly shown as the daily onset cases started to decrease after 4 to 5 days following the lock down intervention shown as vertical purple dotted line at 23, January 2020.

Fitted model parameters are summarized in Table 2. Interestingly, the estimated R_0's with updated data tend to be higher than those from previous results, in the range of 2.7 to 4.2. As the new results were from more recent data and more solid disease natural history assumptions, we believe these estimates should be more convincing. The intervention effect of lock down took place gradually and we estimated after an average of 3.7-5.3 days, the effective reproduction number $R_{eff} = R_0\theta(t)$ fall below 1, when the outbreak became under control.

5 Discussion

Obtaining timely understanding of the transmission of emerging infectious disease is crucial for evaluating the damage and size from the outbreak, and to provide important information for informed intervention strategies to contain the outbreak. We have fitted a relatively simple transmission model requiring basic knowledge of disease natural history and daily records of symptom onset to estimate R_0. Our inferred R_0 fall in the same range as results from other researchers [9, 10, 17, 18], who performed modelling work using different sources of outbreak data. Our model made the assumption that the outbreak took place in a homogeneous population with random mixing of individuals, which reflects the fact that the majority of the early phase of COVID-19 outbreak occurred in the greater Wuhan area, and the input cases in other cities in China and internationally account for a small proportion of all infected by 23, January 2020. We did not rely on the traffic volume data in the current approach, however, these could be

Table 1 Estimated R_0 and its 95% CI for early outbreak phase of 2019 nCOV outbreak in China

Model with infectiousness after symptom onset

P_{inc}	Baseline	Shorter	Longer	Baseline	Baseline
P_{inf}	Baseline	Baseline	Baseline	Earlier	Later
Data source					
Confirmed [2]	2.47 (2.39, 2.55)	2.05 (1.98, 2.12)	2.89 (2.80, 2.99)	2.40 (2.32, 2.48)	2.82 (2.73, 2.92)
Confirmed/suspected [2]	2.54 (2.49, 2.60)	2.10 (2.06, 2.15)	2.99 (2.92, 3.05)	2.47 (2.42, 2.52)	2.92 (2.85, 2.98)
Augmented confirmed [5]	2.31 (2.25, 2.38)	1.95 (1.90, 2.01)	2.65 (2.58, 2.73)	2.25 (2.19, 2.32)	2.60 (2.52, 2.67)

Modified model with asymptomatic infection

P_{inc}	Baseline	Shorter	Longer	Baseline	Baseline
P_{inf}	Baseline	Baseline	Baseline	Earlier	Later
Data source					
Confirmed [2]	2.88 (2.78, 2.95)	2.88 (2.78, 2.99)	2.87 (2.77, 2.98)	2.09 (2.02, 2.17)	3.79 (3.66, 3.92)
Confirmed/suspected [2]	2.97 (2.90, 3.04)	2.98 (2.92, 3.05)	2.97 (2.90, 3.04)	2.15 (2.10, 2.19)	3.93 (3.84, 4.03)
Augmented confirmed [5]	2.69 (2.61, 2.77)	2.69 (2.62, 2.77)	2.69 (2.61, 2.78)	2.01 (1.95, 2.07)	3.41 (3.30, 3.53)

Results from the first model assumed same latent period and incubation period. Results from the modified model assumed infectiousness can develop before symptom onset. Estimates of R_0 from sensitivity analyses were also shown, where the assumed distribution of incubation period and relative infectivity was changed shown as in Figs. 2 and 3
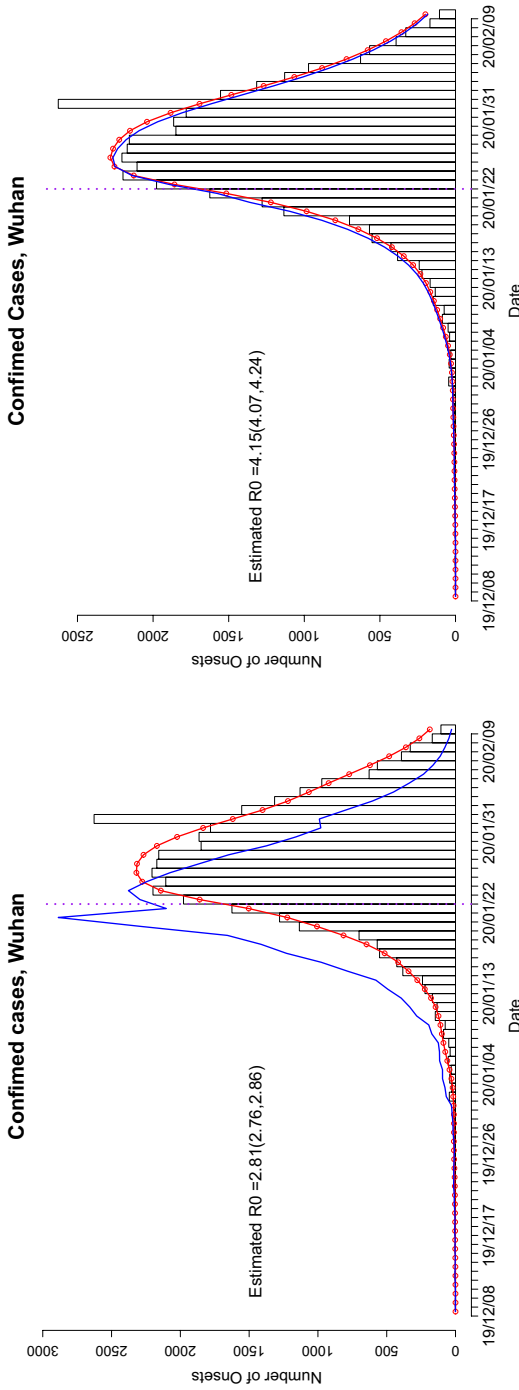
(a) Cases in Wuhan, without asymptomatic infection.

(b) Cases in Wuhan, with asymptomatic infection.

Fig. 5 Observed (shown as bars) and fitted (red curve) number of cases with symptom onset over time, and fitted number of infected cases (blue curve) over time from updated data [12]. Results were with shorter incubation period and baseline infectivity assumptions. The purple vertical dotted line indicates the start of lock down intervention

modelled as additional factors that affect the hazard contribution from geographical covariates when available. The inferred transmissibility were similar from epidemiological report when using China CDC and the augmented onset record from case confirmation time with potential report delay. Ignoring the report delay may lead to substantial bias for estimating R_0, as the exponential growth rate of cumulative cases could be a lot higher than reality. We considered potential asymptomatic infectiousness in the modified model. As more recent evidences indicating potential infection without symptom exists [1, 6], we want to emphasis that these modified assumptions may be crucial to understand not only the disease transmissibility, but to a greater extent in future model to evaluate intervention strategies with sensitive timing components.

Nevertheless, there were challenges to the current approach, mainly due to the limited understanding of the emerging COVID-19 outbreak and the lack of full, accurate data. First, the R_0 estimates were sensitive to different assumptions of disease natural history. We rely on accurate quantification from contact history of the current outbreak and knowledge of previous coronavirus outbreaks such as SARS and MERS to formulate reasonable assumptions. Also, whether asymptomatic infection exists for COVID-19 is still unclear yet. From the comparison between model that assumed no asymptomatic infection and the modified one assuming possible infectiousness before symptom onset, the R_0 estimates were about 0.5 higher in the latter results. Updated case reports that indicate potential asymptomatic infection of COVID-19 would be important in making the sensible model assumption. If there is asymptomatic transmission for COVID-19 outbreak, much more stringent intervention efforts such as quarantining all individuals with potential exposure contacts are required, as compared to quarantining only these with exact contact history with confirmed cases. On the other hand, currently reported cases from China CDC may merely be a reflection of the limited resources of available medical services including hospital beds, health care work forces, medical protection equipment and virus RNA testing capability. There could be more cases of infection and death than

Table 2 Estimated model parameters and their 95% CIs for COVID-19 outbreak in Wuhan with updated data [12]

p_{inf}	Without asymptomatic infection		With asymptomatic infection	
	Baseline	Earlier	Baseline	Earlier
Confirmed cases, Wuhan[a]				
θ_1	0.579 (0.556, 0.599)	0.582 (0.563, 0.602)	0.635 (0.612, 0.659)	0.664 (0.640, 0.690)
θ_2	0.126 (0.122, 0.130)	0.121 (0.117, 0.125)	0.183 (0.179, 0.189)	0.137 (0.133, 0.142)
R_0	2.81 (2.76, 2.86)	2.70 (2.65, 2.75)	4.15 (4.07, 4.24)	2.73 (2.68, 2.79)
Days after intervention to $R_{eff} < 1$	3.85	3.74	5.29	4.35

Shorter incubation period assumption used

[a]See the results from cases outside Wuhan in the Supplementary Material

reported, acting as one of the biggest challenges for transmission modelling. The limited resources not only caused potential underestimated number of cases, but also could lead to additional distortion in the distribution of report delay and relative infectivity.

As the outbreak is continually expanding to more regions within China and internationally, the homogeneous spreading assumption must be adapted to account for the heterogeneity of transmission conditions in different time and space. Individual level covariates should be incorporated into transmission dynamics, and time-varying effect as well as spatial parameters to modify the local intensity of infection can be estimated when richer data from the outbreak become available. It is especially critical to evaluate the efficacy of traffic ban of the greater Wuhan region and future interventions including quarantine sites or promising medication used to reduce the effective reproductive number of COVID-19. We have included lock down effects in our model with updated data, and it was shown to effectively control the outbreak within 4 to 5 days following the intervention, albeit at massive economical and social costs. It is very unlikely other countries could copy the exact draconian interventions happened in China. It remains utterly important for careful evaluation and implementation of various strategies to control the outbreak.

References

1. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, Wang M (2020) Presumed asymptomatic carrier transmission of covid-19. JAMA. https://doi.org/10.1001/jama.2020.2565
2. ChinaCDC (2020) Epidemic update and risk assessment of 2019 novel coronavirus. http://www.chinacdc.cn/yyrdgz/202001/P020200128523354919292.pdf
3. Forsberg White L, Pagano M (2008) A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. Stat Med 27(16):2999–3016
4. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, Liu L, Shan H, Lei C, Hui D et al (2020) Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. https://doi.org/10.1056/NEJMoa2002032
5. Hay J (2020) Turning ncov case reports into infection incidence . https://github.com/jameshay218/case_to_infection
6. Hu Z, Song C, Xu C, Jin G, Chen Y, Xu X, Ma H, Chen W, Lin Y, Zheng Y et al (2020) Clinical characteristics of 24 asymptomatic infections with covid-19 screened among close contacts in Nanjing, China. Sci China Life Sci. https://doi.org/10.1007/s11427-020-1661-4
7. Lau EH, Hsiung CA, Cowling BJ, Chen CH, Ho LM, Tsang T, Chang CW, Donnelly CA, Leung GM (2010) A comparative epidemiologic analysis of sars in Hong Kong, Beijing and Taiwan. BMC Infect Dis 10(1):50
8. Leung GM, Hedley AJ, Ho LM, Chau P, Wong IO, Thach TQ, Ghani AC, Donnelly CA, Fraser C, Riley S et al (2004) The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients. Ann Intern Med 141(9):662–673
9. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TT, Wu JT, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus—infected pneumonia. N Engl J Med. https://doi.org/10.1056/NEJMoa2001316
10. Liu T, Hu J, Kang M, Lin L, Zhong H, Xiao J, He G, Song T, Huang Q, Rong Z et al (2020) Transmission dynamics of 2019 novel coronavirus (2019-ncov). bioRxiv
11. Mitchell M (2002) Software enguage digitizer (version 4.1)

12. Novel CPERE et al (2020) The epidemiological characteristics of an outbreak of 2019 novel corona-virus diseases (covid-19) in china. Zhonghua Liu Xing Bing Xue Za Zhi 41(2):145

13. Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP (2020) Novel coronavirus 2019-ncov: early estimation of epidemiological parameters and epidemic predictions. medRxiv

14. Sun K (2020) Kudos line list data. https://docs.google.com/spreadsheets/d/1jS24DjSPVWa4iu xuD4OAXrE3QeI8c9BC1hSlqr-NMiU

15. Virlogeux V, Fang VJ, Park M, Wu JT, Cowling BJ (2016) Comparison of incubation period distri-bution of human infections with mers-cov in South Korea and Saudi Arabia. Sci Rep 6:35839

16. WHO (2020) Novel coronavirus(2019-ncov) situation report-57. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200317-sitrep-57-ncov.pdf

17. Wu JT, Leung K, Leung GM (2020) Nowcasting and forecasting the potential domestic and inter-national spread of the 2019-ncov outbreak originating in Wuhan, China: a modelling study. Lancet 395:689–697

18. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Lou Y, Gao D, Yang L, He D et al (2020) Prelim-inary 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

19. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, Yu J, Kang M, Song Y, Xia J et al (2020) Sars-cov-2 viral load in upper respiratory specimens of infected patients. N Engl J Med 382(12):1177–1179