Modelling and estimating the transmissibility of COVID-19: Transmission dynamics in Shaanxi Province of China as a case study

Xu-Sheng Zhang¹², Huan Xiong³, Zhengji Chen³, and Wei Liu³,*

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
³School of Public Health, Kunming Medical University, Kunming, Yunnan, P. R. China

*Corresponding author:
Email: liuweikm@qq.com
Abstract

To control and contain the outbreaks of emerging infectious diseases such as COVID-19, it is important to know how easy and fast they transmit among people. To explore the essential information of the novel infectious agents, people always confront an inverse problem: using (partially) observed numbers of infected people by time and region to dig up the underlying characteristics of unknown infectious agents. Epidemics armed with advanced statistical inference and mathematical theory has been developed to help reconstruct transmission dynamics processes and to estimate key features of infectious diseases. In this study we use COVID-19 outbreak in Shaanxi province as an example to illustrate how the infectious disease dynamics method can be used to help build the transmission process and to estimate the transmissibility of COVID-19. Three transmission dynamics models were proposed for this. By separating continuous importation from local transmission and treating imported cases as the source rather than results of local transmission, the basic reproduction number of COVID-19 in Shaanxi province was estimated in the range from 0.46 to 0.61, well below the critical value of 1.0. This indicates that COVID-19 cannot self-sustain in Shaanxi province and reflects the timely and strong control measures taken in Shaanxi province.

Keywords: Basic reproduction number, Bayesian inference, COVID-19, Epidemics, Mathematical modelling
Introduction

The emerging coronavirus infectious disease, COVID-19, has been circulated worldwide since January 2020. To control its spread, it is crucial to accurately estimate its important epidemiological characteristics such as transmissibility and to predict its further spread. In epidemiology, mathematical equations have been used to describe the transmission dynamic processes and to predict how infectious diseases spread within a population. It has a long history and goes back to Ronald Ross (1911). The problems people face in understanding infectious disease outbreaks are always inverse: we observe the numbers of infected people by time and region, but we don’t know the underlying characteristics of the infectious disease, especially for novel emerging infectious diseases such as COVID-19. With the progress of advanced statistical inference methodology and mathematical theory, epidemics has become a well-developed field which targets quantitatively describing infectious disease dynamics (Anderson and May, 1991; Heesterbeek and Diekmann, 2000; Keeling and Rohani, 2007). Based on the observed data and inference methods and modelling techniques, epidemics estimate the key parameters of epidemic processes and provide policy makers timely and accurate information of infectious disease and their spread within the population (e.g., Wu J et al 2020).

The transmissibility of an infectious agent describes how easy and fast the infectious disease can spread within a population. It is usually measured by the basic reproduction number (denoted it as $R_0$), which is defined as the average number of secondary infections generated by an infectious person introduced into a completely susceptible population (Anderson and May, 1991). Although many methods of estimating $R_0$ have been developed (Heffernan et al., 2005; Vynnycky and White, 2010), the difficulty in measuring $R_0$ of COVID-19 lies in the fact that it is a novel coronavirus. The knowledge of the well-known coronaviruses such as SARS and MERS have been borrowed for understanding the early transmission dynamics of COVID-19 (e.g., Wu J et al., 2020). Nevertheless, epidemiological characteristics of COVID-19 appear quite different from those of both SARS and MERS (Biggerstaff et al., 2020). The basic knowledge of COVID-19 epidemiological features should be obtained from the epidemic data during outbreaks. During the initial outbreaks of COVID-19 in China from January to March 2020, the national and provincial governments and public health authorities have collected lots of data about the outbreaks and individual cases’ information. These data undoubtedly provide a good chance for us to understand the transmissibility of COVID-19 and the impact of control measures taken on stopping the spread. The reliable and accurate estimates depend on our understanding of how SARS-CoV-2 is transmitted in the population and the appropriate inference methods to calibrate the transmission models. We noticed that although a few modelling studies have been published in China (Bai et al 2020; Wu W et al 2020), this has not been correctly addressed. In this study we propose three epidemic models for the transmission dynamics of SAR-CoV-2: renewal equation model, SEDAR and SEEDAR models (Figure 1), and use advanced inference methods to estimate $R_0$ by applying the three models to observed data in the outbreak in Shaanxi province of China. We found that having separated continuous importation from local transmission, $R_0$ of COVID-19 in the Shaanxi outbreak is well below the critical value 1.0; among the three models, the simplest model (renewal equation) gives the best model fitting and prediction.

Results
Renewal equation model: Bayesian inference suggests the basic reproduction number ($R_0$) during the COVID-19 outbreak in Shaanxi province has a median 0.61 and 95% confidence interval (CI) from 0.54 to 0.68 (Table 1). The serial interval (SI), the lag in onset dates of symptoms between an infector and its infectee, is estimated to have a mean 4.6 days and standard deviation 11.7 days, which is comparable with the direct observation of SI from the outbreak (Figure S3). Model projection into next month (right panel of Figure 2) indicates that the outbreak will die out within one week (i.e., the end of February 2020) and is unlikely to generate any further local cases under the current restriction measures. This prediction is consistent with what was happened in Shaanxi province.

The time-varying reproduction number $R_t$ shown in Figure 3 demonstrates how the transmissibility changed along the course of the outbreak. Assuming $SI_{mean}=3.9$ days and $SI_{sd}=5.1$ days which are directly estimated from the observed SIs, $R_t$ increased to about 3.0 within the first week and then reduced to low values but occasionally exceeded the critical value 1.0. Its overall average is 0.60 which is nearly identical to the median of posterior of $R_0$ in the above model fitting. If assuming $SI_{mean}=4.6$ and $SI_{sd}=11.7$ (i.e., the estimate of SI from the above model fitting), we found the similar changing patterns of $R_t$ and nearly the same overall mean $R_t$. The change of $R_t$ reflects the enhancement of intervention measures with the outbreak in Shaanxi province.

SEDAR model: The estimates of model parameters are shown in Table 1. The calibration of the SEDAR model under the situation with equal incubation and infectious periods for both symptomatic and asymptomatic infection shows that: $R_0$ is estimated at 0.59 with 95%CI from 0.51 to 0.71. The incubation period is 1.8 days (95%CI: 1.6, 2.9), and the infectious period is 3.8 days (95%CI: 3.5, 5.3). The SEDAR model fitting to the observed data of local cases and its prediction over one month ahead is shown in the middle panel of Figure 2. Sensitivity analyses show that nearly the same estimates of model parameters are obtained when considering different values of $L_2$ and $D_2$ and $\xi$. This reflects the fact of the outbreak in Shaanxi province that the asymptomatic infections occupied a very small proportion of all infections (1.1%) (Bai et al., 2020) and therefore had small effects on model performance.

SEEDAR model: The estimates of parameters of SEEDAR model are also shown in Table 1. The MCMC samples of model parameters under the situation (i.e., $L_2=L_1$, $D_2=D_1$ and $\xi=0.5$) are shown in Figure S5. The $R_0$ is estimated at 0.46 with 95%CI from 0.30 to 0.79. The incubation period of symptomatic infection is 5.0 days (95%CI: 1.3, 9.7) which is consistent with the observed values (mean= 4.4 days and sd=5.5 days), and the infectious period of symptomatic infections is 4.8 days (95%CI: 1.6, 14.1), which is longer than the delay from onset date of symptoms to hospitalization: mean= 2.6 days and sd=3.7 days (Figure S4). The duration of pre-symptomatic transmission ($L_3$) is estimated at 1.5 day (95%CI: 1.0 to 4.4 days), which suggests the fraction of transmission from strictly pre-symptomatic infections was 1.5/(1.5+4.8) =24% (c.f. Ferretti et al., 2020). The SEEDAR model well fits to the observed data and predicts that the outbreak will die out within about half month (the right panel of Figure 2).

Discussion
In this study, three dynamics models were proposed to model the transmission dynamics of COVID-19 epidemic within Shaanxi province from the early January to late February 2020. By distinguishing imported and local cases in their contribution to local transmission dynamics, we show the basic reproduction number $R_0$ of COVID-19 in the Shaanxi outbreak was well below the critical value 1.0. This indicates that SAR-CoV-2 cannot self-sustain under the current control measures within Shaanxi and will stop once the importation of COVID-19 cases is halted. Our model predictions reflect the real situation in Shaanxi province since the late of February 2020.

The estimate of $R_0$ from renewal equation and SEDAR models are close to each other, and its estimate from SEEDAR model is lower; nevertheless, their 95% CIs are closely overlapped. Overall, the estimate of $R_0$ is in the range from 0.46 to 0.61. The model fittings to the local cases shown in Figure 2 indicates that the renewal equation model provides the best fit to the observations and SEDAR model is the worst. This is further confirmed by the values of DIC in Table 1 (Burnham and Anderson 2002): 126.9, 175.1 and 160.5 for Renewal equation, SEDAR and SEEDAR models, respectively. Furthermore, the better performance of the SEEDAR model than the SEDAR model confirms the existence of pre-symptomatic transmission (Ferretti et al., 2020).

It is worth emphasizing that although the renewal equation model is the simplest in model structure, it gives the best model fit. Given distribution of SI of COVID-19 which is specified by its mean and standard deviation, it is straightforward to obtain the estimate of the $R_0$ (Chen et al., 2020). In this study we perform a joint estimation of $R_0$ and SI, and the results agree well with two compartmental models in estimate of $R_0$ and the empirical knowledge of SI (Biggerstaff et al., 2020). However, it should bear in mind that the successful performance of the joint estimation of $R_0$ and SI in this study may be conditional on the very low proportion (i.e., 1.1%) of asymptomatic infections (c.f. Griffin et al., 2011). Another thing worth pointing out is the continuous importation along the course of an outbreak within one region except the epicentre Wuhan city, China. This spectacular feature, unlike infectious disease pandemics before (such as 2003 SARS pandemic and 2009 influenza pandemic), may reflect the rapid and huge movements of modern human beings. If assuming the earliest importation as the only index case(s), the transmission dynamics cannot be appropriately investigated and might be misled (Bai et al., 2020; Wu W et al., 2020).

In modelling the transmission dynamics of COVID-19 over the whole mainland China, we found that $R_0$ was estimated at 2.23 before 8th February 2020 and then it dropped to 0.04 (Zhang et al., 2020). To check whether there was any potential breaking point in transmissibility of COVID-19 within Shaanxi, we calculate the instantaneous reproduction number $R_t$ (Cori et al., 2013). The result shown in Figure 3 indicates that no clear pattern emerged that supported a potential breaking point in transmissibility although $R_t$ exceeded 1.0 on five days (8, 20, 31 January and 2, 10 February 2020). This may reflect the quick and strong control measures taken within Shaanxi province and the COVID-19 was under effective control soon since it was announced in public on 20 January 2020 that COVID-19 can be transmitted among people.
In conclusion, it is important to separate continuous importation from the local transmission when modelling the local transmission dynamics of COVID-19. Modern inference methodology and mathematical theory can help reveal the unobserved transmission dynamic process and hence provide valuable information for us to understand and control the spread of COVID-19. The renewal equation model, albeit being simple in model structure and of only three parameters, provides the good model fitting and therefore is a practical candidate for analysing transmission dynamics and estimating the transmissibility.

**Materials and methods**

**Data**

The outbreak data of COVID-19 in Shaanxi province, China were collected from Shaanxi provincial government website from 23 January to 20 February 2020. There are 245 cases reported during the period, among which 113 are imported from outside of Shaanxi and 132 are locally transmitted within Shaanxi province (left panel of Figure S1). The dates of symptom onset were recorded for only 210 cases. The delay from symptom onset to reporting from these 210 cases is estimated to have a mean of 7.54 days and a standard deviation of 4.12 days. The other 35 cases whose dates of symptom onset were missed are imputed from their reporting dates and the distribution of delay from symptom onset to reporting. The timeline of dates of symptom onset of 245 cases is shown in the right panel of Figure S1.

To provide the direct dates for modelling local transmission dynamics, we construct modified dates of symptom onset for imported cases. If the date of symptom onset of one imported case was earlier than the date of entry into Shaanxi province, then its modified date of symptom onset is its entry date; otherwise, the modified date of symptom onset is its date of symptom onset. There are 19 cases whose dates of symptom onset were earlier than their entry dates. With these arrangements, the timeline of modified dates of symptom onset is shown in Figure S2.

From 101 pairs of infector-infectee observed during the outbreak, the SI is estimated to have a mean of 3.88 days and standard deviation of 5.08 days (Figure S3). From 144 cases that had dates of exposure and symptom onset, the incubation period is estimated to have a mean of 4.51 days and standard deviation of 5.40 days and the delay from onset of symptoms to hospitalization has a mean of 2.63 days and a standard deviation of 3.56 days (Figure S4).

**Models**

**Renewal equation model**

It is assumed that, once infected, individuals have an infectivity profile given by a probability distribution $w_s$, dependent on time since infection of the case, $s$, but independent of calendar time, $t$. The distribution $w_s$ typically depends on individual biological factors such as pathogen shedding or symptom severity. For simplicity, the distribution $w_s$ is approximated by the distribution of serial interval (SI). In the original renewal equation model, Fraser (2007) considers a situation where the only importation is index case(s) at the beginning of
the outbreak and other cases are generated by local transmission. During the spread of COVID-19 in 2020, the outbreak within a region (except Wuhan the epicentre) took place with continuous importation. Let \( c_t \) be the number of cases whose symptoms onset at day \( t \), its expected value is approximated by

\[
E(c_t) = R_0 \sum_{j=1}^{\min(t-1,SI_{\text{max}})} w_s(c_{t-s} + I_{t-s})
\]  
(1)

Here \( I_{t-s} \) is the number of imported cases that have the onset date of symptoms on day \( t-s \) and \( w_s \) represents the probability mass function of the SI of length \( s \) days, which can be obtained by \( w_s = G(s) - G(s - 1) \), with \( G(.) \) representing the cumulative distribution function of the gamma distribution. The gamma distribution is characterised by its mean \( SI_{\text{mean}} \) and standard deviation \( SI_{sd} \), both of which are to be estimated jointly with \( R_0 \) from the collected data during an outbreak (Griffin et al., 2011). Because only 19 cases among 113 imported cases had symptom onset before entering Shaanxi province, the assumption that all cases started their infectivity duration within Shaanxi province, which is implicitly required in equation (1), should be approximately satisfied.

Equation (1) can be rearranged to give an expression for instantaneous reproduction number \( R_t \):

\[
R_t = c_t / \sum_{j=1}^{\min(t-1,SI_{\text{max}})} w_s(c_{t-s} + I_{t-s})
\]  
(1a)

That is, \( R_t \) can be estimated by the ratio of the number of new infections produced at time step \( t \), \( c_t \), to the total infectiousness of infected individuals at time \( t \), given by \( \sum_{j=1}^{\min(t-1,SI_{\text{max}})} w_s(c_{t-s} + I_{t-s}) \), the sum of infection incidence including both imported and locally generated up to time step \( t-1 \) or the maximum of SI (whichever is the smallest), weighted by the infectivity function \( w_s \). \( R_t \) is the average number of secondary cases that each infected individual would infect if the conditions remained as they were at time \( t \) (Cori et al., 2013). \( R_t \) can be used to monitor the change of transmissibility along the course of outbreak.

**SEDAR transmission model**

We assume the transmission dynamics of SARS-CoV-2 virus is described by an SEDAR compartmental model as shown in Figure 1A. That is, a susceptible person (\( S \)) contracts SARS-CoV-2 virus from infectious people and then enters the latent class (\( E \)); a fraction (\( \theta \)) of these exposed after an average latent period (\( L_1 \)) progress to become diseased (\( I \)) and the other fraction (1-\( \theta \)) remain asymptomatic but become infectious after an average infective period (\( L_2 \)). The diseased infections will be detected and admitted to hospital and isolated from the community after an average period \( D_1 \) and the asymptomatic cases recover after an average infectious period \( D_2 \). The model can be described by the following set of differential equations:

\[
\frac{d}{dt} S(t) = - \beta S(t)(I(t) + \xi A(t))/N
\]
\[
\frac{d}{dt} E(t) = \beta S(t)(I(t) + \xi A(t))/N - \theta E(t)/L_1 - (1 - \theta)E(t)/L_2
\]

\[
\frac{d}{dt} I(t) = \theta E(t)/L_1 - I(t)/D_1 + \text{Imported}(t)
\]

\[
\frac{d}{dt} A(t) = (1 - \theta)E(t)/L_2 - A(t)/D_2
\]

\[
\frac{d}{dt} R(t) = I(t)/D_1 + A(t)/D_2
\]

Here \( N \) is the total size of the population under investigation and is assumed to be constant during the outbreak. The definitions of model parameters are given in Table 1. Importantly, the model includes imported cases (i.e., \( \text{Imported}(t) \) in equation for \( I \)) from outside the population as reported by public health authorities (c.f., Bai et al., 2020). This is to treat imported cases as the source rather than the results of local transmission from the region under investigation.

The steady-state solution of the equation system (2) can be easily obtained. The expression for \( S^* \) (the size of the population susceptible to infection at equilibrium) is

\[
S^* = N \frac{\theta/L_1 + (1-\theta)/L_2}{\beta[\theta D_1/L_1 + \xi(1-\theta)D_2/L_2]}.
\]

From this, we can obtain the expression of basic reproduction number

\[
R_0 = N/S^* = \beta[\theta D_1/L_1 + \xi(1-\theta)D_2/L_2]/[\theta/L_1 + (1 - \theta)/L_2]
\]

(Vynnycky and White 2010). In the special situation where \( L_1=L_2 \) and \( D_1=D_2 \), expression (3) reduces to

\[
R_0 = \beta[\theta D_1 + \xi(1-\theta)D_2]
\]

(SEEDAR transmission model)

Ferretti et al. (2020) shows that 30% to 50% of all transmission are pre-symptomatic transmission. To take the pre-symptomatic transmission into account, we modify the above SEDAR model by including a secondary exposure compartment (see Figure 1B). For simplicity, this new compartment is assumed to be asymptomatic but of the same infectivity as the symptomatic infections. The corresponding equations are modified as

\[
\frac{d}{dt} S(t) = - \beta S(t)(E_2(t) + I(t) + \xi A(t))/N
\]

\[
\frac{d}{dt} E_1(t) = \beta S(t)(E_2(t) + I(t) + \xi A(t))/N - \theta E_1(t)/L_1 - (1 - \theta)E_1(t)/L_2
\]

\[
\frac{d}{dt} E_2(t) = \theta E_1(t)/L_1 - 2E_2(t)/L_3
\]
\[
\frac{d}{dt} I(t) = \frac{E_2(t)}{L_3} - I(t) / D_1 + \text{Imported(t)} \quad (4)
\]

\[
\frac{d}{dt} A(t) = (1 - \theta)E_1(t)/L_2 - A(t)/D_2
\]

\[
\frac{d}{dt} R(t) = I(t)/D_1 + A(t)/D_2
\]

Compared with the SEDAR model, a new parameter \( L_3 \), the duration of the late incubation period on which the infected person can pass the virus on, is introduced and is to be estimated (See Table 1).

Similarly, the basic reproduction number \( R_0 \) for SEEDAR model can be obtained by deriving the expression of equilibrium number of susceptible people, and it is given by

\[
R_0 = \beta [\theta(L_3 + D_1)/L_1 + \xi(1 - \theta)D_2/L_2]/[\theta/L_1 + (1 - \theta)/L_2] \quad (5)
\]

**Inference method**

Inference is carried out within the Bayesian framework (Bettencourt and Ribeiro 2008; Cauchemez et al., 2004), based upon the posterior distributions of parameters and derived quantities of interest, obtained through the combination of the prior distributions and the likelihood function. We denote the set of model parameters to be inferred as \( \Theta = \{ R_0, \text{SI}_{\text{mean}}, \text{SI}_{\text{ad}} \} \) for renewal equation model, \( \Theta = \{ \beta, L_1, D_1 \} \) for SEDAR model and \( \Theta = \{ \beta, L_1, L_3, D_1 \} \) for SEEDAR model under the special situation where both asymptomatic and symptomatic infections are of the same latent period and infectious period (i.e., \( L_2 = L_1 \) and \( D_2 = D_1 \)). For simplicity, the proportion of symptomatic infections (\( \theta \)) is set at 98.9% as reported (Bai et al., 2020). Given values of parameters \( \Theta \) for renewal equation model (1), simulating the time series of infections, denoted as \( \mu(t), t = t_{\text{start}},..., t_{\text{end}} \), is straightforward. Here \( t_{\text{start}} \) and \( t_{\text{end}} \) represent the start day and end day of the outbreak data collected, respectively. For each set of parameter values of SEDAR and SEEDAR models, the Runge-Kutta fourth order method is used to solve the model equations and to obtain model predicted time series of infections. In the inference of model parameters, directly observed dataset of confirmed/hospitalized/reported cases are used as illustrated in the following. The likelihood function for the observed times series of local cases \( x(t) \), \( t = t_{\text{start}},..., t_{\text{end}} \), is given as

\[
L(\Theta|\text{Data}) = \prod_{t=t_{\text{start}}}^{t_{\text{end}}} \frac{\Gamma(x(t)+r(t))}{\Gamma(r(t))\Gamma(x(t)+1)} \frac{1}{\eta} r(t) (1 - \frac{1}{\eta}) x(t) \quad (6)
\]

Here \( r(t) = \frac{\mu(t)}{\eta-1} \) with \( \eta \) being the dispersion parameter of the negative binomial distribution.

The parameters are estimated using Markov Chain Monte Carlo methods with Gibbs sampling and non-informative flat prior. The boundaries of uniformly distributed priors are set upon from the literature (Biggerstaff et al., 2020) and the data collected from Shaanxi province (Figures S3 and S4). We give the details of the MCM sampling method below.
To propose new values for parameters, we use normal random walk. Suppose the current value of the \(j\)th parameter of \(\Theta\) is \(\Theta_j^{(t-1)}\), the new proposal is
\[
\Theta_j^* = \Theta_j^{(t-1)} + \sigma_j z
\]
Here \(z\) is a standard normal variable and \(\sigma_j\) is the step size of the \(j\)th parameter. The normal proposal density is given by
\[
q(\Theta^* | \Theta^{(t-1)}) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{(\Theta^* - \Theta^{(t-1)})^2}{2\sigma^2}\right]
\]
That is, \(\Theta_j^*\) follows \(N(\Theta_j^{(t-1)}, \sigma_j^2)\) (normal distribution with mean= \(\Theta_j^{(t-1)}\), and standard deviation =\(\sigma_j\)). The proposal is accepted as the next step of the Markov chain with probability \(\alpha = \min(A,1)\), where
\[
A = \frac{\pi(\Theta^*) L(\Theta^* | y) q(\Theta^{(t-1)} | \Theta^*)}{\pi(\Theta^{(t-1)}) L(\Theta^{(t-1)} | y) q(\Theta^* | \Theta^{(t-1)})}
\]
Here \(\pi(.)\) denote the prior density, \(L(\Theta_j^* | y)\) the likelihood of parameter \(\Theta_j^*\) given data \(y\). For a truncated normal walk on the range \((a,b)\), the proposal density is given by
\[
q(\Theta^* | \Theta^{(t-1)})_{(a,b)} = \frac{q(\Theta^* | \Theta^{(t-1)})}{\Phi\left(\frac{b - \Theta^{(t-1)}}{\sigma}\right) - \Phi\left(\frac{a - \Theta^{(t-1)}}{\sigma}\right)}
\]
Where \(\Phi(.)\) is the cumulative distribution function of standard normal. The expression for \(A\) is consequently modified as
\[
A = \frac{\pi(\Theta^*) L(\Theta^* | y) \Phi\left(\frac{b - \Theta^{(t-1)}}{\sigma}\right) - \Phi\left(\frac{a - \Theta^{(t-1)}}{\sigma}\right)}{\pi(\Theta^{(t-1)}) L(\Theta^{(t-1)} | y) \Phi\left(\frac{b - \Theta^*}{\sigma}\right) - \Phi\left(\frac{a - \Theta^*}{\sigma}\right)}
\]
Sample a uniformly distributed random number \((r)\) between 0 and 1,
\[
\Theta_j^{(0)} = \Theta_j^* \quad \text{if} \quad r < \alpha; \quad \text{(accepted)}
\]
\[
\Theta_j^{(0)} = \Theta_j^{(t-1)} \quad \text{otherwise. (rejected)}
\]
To generate nearly independent samples of model parameters, the Markov chain samples are to be thinned every 400th observations. To respond to the acceptance rate, the following adaptive procedure is applied: if the acceptance ratio over 400*200 iterations is less than 12\%, then decrease the jump step to 80\% of its current size (i.e., \(\sigma_j = 0.8\sigma_j\)); if it exceeds 40\%, then \(\sigma_j = 1.2\sigma_j\). Otherwise the jump step \(\sigma_j\) remains unchanged. To allow the MCMC process to fully converge, a burn-in period of 400,000 iterations is chosen and the estimates of model parameters are obtained from the further 400,000 iterations.
To compare the performance of the above three models (Burnham and Anderson, 2002), the deviance information criterion (DIC), which combines goodness of fit and model complexity (Spiegelhalter et al., 2002), is used. It measures fit via the deviance \( \text{Dev}(\Theta) = -2\log L(\Theta|\text{Data}) \) and complexity by estimate of the ‘effective number of parameters’ \( p_D = \text{mean}(\text{Dev}(\Theta)) - \text{Dev}(\text{mean}(\Theta)) \) (i.e., posterior mean deviance minus deviance evaluated at the posterior mean of the parameters). The DIC is calculated as

\[
\text{DIC} = \text{Dev}(\text{mean}(\Theta)) + 2p_D = \text{mean}(\text{Dev}(\Theta)) + p_D.
\]

The model that has the smallest DIC is the best.

**Compliance and ethics**

The authors declare no Conflict of Interest.

**Acknowledgements**

This work is jointly supported by Public Health England, UK, National Nature Science Foundation of China (81860607) and Innovative Research Team of Yunnan Province (2019(6)), China.

**References**

Anderson, R.M., and May, R.M. (1991). Infectious Diseases of Humans: Dynamics and Control. (Oxford University Press, Oxford).

Bai, Y., Liu, K., Chen, Z., Chen, B., Shao, Z. (2020) Early transmission dynamics of novel coronavirus pneumonia Epidemic in Shaanxi Province. Chin J Nosocomiol 30, 834- 838.

Bettencourt, L.M.A., Ribeiro, R.M. (2008) Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS One 3:e2185.

Biggerstaff, M., Cowling, B.J., Cucunubá, Z.M., Dinh, L., Ferguson, N.M., Gao, H., Hill, V., Imai, N., Johansson, M.A., Kada, S., Morgan, O., Pastore y Piontti, A., Polonsky, J.A., Prasad, P.V., Quandelacy, T.M., Rambaut, A., Tapper, J.W., Vandemaele, K.A., Vespignani, A., Warmbord, K.L., Wong, J.Y., and for the WHO COVID-19 Modelling Parameters Group (2020) Early Insights from Statistical and Mathematical Modeling of Key Epidemiologic Parameters of COVID-19 Emerg Infect Dis 26(11) Nov 2020

Burnham, K.P., and Anderson, D.R. (2002). Model Selection and Multimodel Inference: a Practical Information—Theoretic Approach. 2nd edition. (New York: Springer).

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A.J., Boëlle, P.Y. (2004). A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. Stat Med 23, 3469–3487.

Chen, Z., Li, H., Feng, L., Gu, R., Zhang, L., Liu, H., Li, N., Liu, W., Wei, T., Zhang, X.-S. (2020). Dynamics of early transmission of covid-19 in Yunnan province. Submitted.
Cori, A., Ferguson, N.M., Fraser, C., and Cauchemez, S. (2013). A new Framework and software to estimate time varying reproduction numbers during epidemics. Am J Epidemiol 178, 1505-1512.

Ferretti, L., Ledda, A., Wymant, C., Zhao, L., Ledda, V., Abeler-Dörner, L., Kendall, M., Nurtay, A., Cheng, H.Y., Ng, T.-C., Lin, H.H., Hinch, R., Masel, J., Kilpatrick, A.M., Fraser, C. (2020). The timing of COVID-19 transmission. medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188516

Fraser, C. (2007). Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. PLoS ONE 2(8): e758.

Griffin, J.T., Garske, T., Ghani, A.C. and Clarke, P.C. (2011). Joint estimation of the basic reproduction number and generation time parameters for infectious disease outbreaks. Biostatistics. 12, 303-12.

Heesterbeek, J.A.P., and Diekmann, O. (2000). Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. (Chichester: John Wiley & Sons)

Heffernan, J.M., Smith, R.J., and Wahl, L.M. (2005). Perspectives on the Basic Reproductive Ratio. J Roy Soc Interface 2, 281–93.

Keeling, M.J., and Rohani, P. (2007). Modelling infectious diseases in human and animals. (Princeton, NJ: Princeton University Press)

Ross, R. (1911) The Prevention of Malaria. (London: John Murray).

Spiegelhalter, D.J., Best, N., Carlin, B.P., van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). J Roy Stat Soc B64, 1-34.

Vynnycky, E. and White, R.G. (2010). An Introduction to Infectious Disease Modelling. (Oxford: Oxford University Press)

Wu, J.T., Leung, K. and Leung, G.M. (2020). 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.

Wu, W., Bai, R., Li D., Feng, A., Xu A., Lü, J. (2020). Preliminary prediction of the epidemic trend of 2019 novel coronavirus (2019-nCoV) pneumonia in Guangdong province. J Jinan Univ (Natural Science & Medicine Edition) 41,1–6.

Zhang, X.-S., Vynnycky, E., Charlett, C., DeAngelis, D., Chen, Z., and Liu, W. (2020) Transmission dynamics and control measures of COVID-19 outbreak in China: a modelling study. https://doi.org/10.1101/2020.07.09.20150086
# Tables

Table 1 Parameter estimates of three transmission dynamics models

| parameter | Renewal equation model | SEDAR model* | SEEDAR model* |
|-----------|------------------------|--------------|--------------|
|           | prior                  | Posterior    | prior        | Posterior    | Prior        | Posterior    |
| $S_1$mean | $U[3.5,10.0]$          | 4.62[3.55,7.46] | –           | –           | –           | –           |
| $S_1$sd   | $U[3.0,15.0]$          | 11.7[6.04,14.88] | –           | –           | –           | –           |
| beta      | –                      | –            | $U[0.001,0.5]$ | 0.155[0.117,0.186] | $U[0.001,0.5]$ | 0.066[0.029,0.154] |
| $L_1$     | –                      | –            | $U[1.6,14.0]$ | 1.81[1.61,2.82] | $U[1.0,10.]$ | 5.04[1.25,9.65] |
| $L_3$     | –                      | –            | –            | –            | $U[1.0,10.]$ | 1.45[1.04,4.43] |
| $D_1$     | –                      | –            | $U[3.5,25.0]$ | 3.75[3.51,5.16] | $U[1.5,15.]$ | 4.78[1.61,14.06] |
| $\eta$    | $U[1.01,50.]$          | 1.57[1.05,2.99] | $U[1.01,50]$ | 2.47[1.56,4.431] | $U[1.01,50]$ | 1.73[1.08,3.26] |
| $R_0$     | $U[0.05,3.0]$          | 0.61[0.54,0.68] | –           | 0.589[0.500,0.702] | –           | 0.455[0.297,0.756] |
| DIC       | –                      | 126.86       | –            | 175.10       | –            | 160.49      |

*: For both SEDAR and SEEDAR models, the relative infectivity ($\xi$) of asymptomatic infections to symptomatic infection is set at 0.5, and the incubation and infectious period for asymptomatic infections are set to be equal to the counterparts of symptomatic infections (i.e. $L_2=L_1$ and $D_2=D_1$). As the proportion of asymptomatic infections is very small (i.e. $1-\theta=1.1\%$), the other choices of these three parameters (say $\xi=1$, $L_2=2L_1$ and $D_2=2D_1$) do not noticeably change the estimates of model parameters listed here.
Figure 1 Flow chart of A) SEDAR transmission model and B) SEEDAR transmission model.
Figure 2 Three models’ fitting to symptom onset dates: Renewal equation (left panel), SEDAR (middle panel) and SEEDAR (right panel). Red dots represent the imported cases, the blue triangles the cases locally transmitted in Shaanxi province. The thick green line represents the median of MCMC samples and thin lines their upper and lower levels of 95% confidence intervals. The vertical black line shows the deadline up to which the observed data were used to calibrate the renewal equation model. The top right inlet shows the one week ahead predictions.
Figure 3 The effective reproductive number along the course of outbreak in Shaanxi province, China under two SI distributions: SI_mean=3.88 days, SI_sd=5.08 days estimated from the observed SIs (upper panel) and SI_mean = 4.6 days, SI_sd = 11.7 days (the MLE of SI from model calibration of renewal equation) (lower panel).
Supporting Information

Data depository: line list data of 245 cases in Shaanxi province up to 20 February 2020

Figure S1 Timeline of COVID-19 outbreak in Shaanxi province, China based on reported dates (left panel) and dates of symptom onset (right panel).

Figure S2 Timeline of COVID-19 outbreak in Shaanxi province, China based on the modified dates of symptom onset.
Figure S3. Distribution of serial interval (SI) from 101 pairs of infector-infectee during the outbreak.

Figure S4. Distribution of incubation from 114 cases whose had both dates of exposures and dates of symptom onset during the outbreak (left panel) and distribution of delay from date of symptom onset to hospital visit (right panel).
Figure S5 MCMC samples of model parameters and its convergence under SEEDAR model under the condition that $L_2=L_1$, $D_2=D_1$ and $\xi=0.5$. 