Stochastic discrete epidemic modeling of COVID-19 transmission in the Province of Shaanxi incorporating public health intervention and case importation

Sanyi Tang¹, Biao Tang^{2,3}, Nicola Luigi Bragazzi³, Fan Xia^{2,4}, Tangjuan Li^{2,4}, Sha He¹, Pengyu Ren⁶, Xia Wang¹, Zhihang Peng⁷, and Yanni Xiao^{2,4,*} & Jianhong Wu^{2,3,5,*}

¹ School of Mathematics and Information Science, Shaanxi Normal University, Xi'an, 710119, China

² The Interdisciplinary Research Center for Mathematics and Life Sciences, Xi'an Jiaotong University, Xi'an 710049, China

³ Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, Ontario, M3J 1P3 Canada

⁴ School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China

⁵ Fields-CQAM Laboratory of Mathematics for Public Health, York University, Toronto, Ontario, M3J 1P3 Canada

⁶ The Second Affiliated Hospital of Xi'an Jiaotong University, West five Road 157, Xi'an, 710004, China.

⁷ Department of Epidemiology & Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, 211166, China

The first four authors made the same contributions.

* Correspondence: yxiao@mail.xjtu.edu.cn, wujh@yorku.ca

Received: date; Accepted: date; Published: date

Abstract:

Before the lock-down of Wuhan/Hubei/China, on January 23rd 2020, a large number of individuals infected by COVID-19 moved from the epicenter Wuhan and the Hubei province due to the Spring Festival, resulting in an epidemic in the other provinces including the Shaanxi province. The epidemic scale in Shaanxi was comparatively small and with half of cases being imported from the epicenter. Based on the complete epidemic data including the symptom onset time and transmission chains, we calculate the control reproduction number (1.48-1.69) in Xi'an. We could also compute the time transition, for each imported or local case, from the latent, to infected, to hospitalized compartment, as well as the effective reproduction number. This calculation enables us to revise our early deterministic transmission model to a stochastic discrete epidemic model with case importation and parameterize it. Our model-based analyses reveal that the newly generated infections decay to zero quickly; the cumulative number of case-driven quarantined individuals via contact tracing stabilize at a manageable level, indicating that the intervention strategies implemented in the Shaanxi province have been effective. Risk analyses, important for the consideration of "resumption of work", show that a large second outbreak is expected if the level of case importation remains at the same level as between January 10th and February 4th 2020. However, if the case importation decreases by 30%, 60% and 90%, the second outbreak if happening will be of small-scale assuming contact tracing and quarantine/isolation remain as effective as before. Finally, we consider the effects of intermittent inflow with a Poisson distribution on the likelihood of multiple outbreaks. We believe the developed methodology and stochastic model provide an important model framework for the evaluation of revising travel restriction rules in the consideration of resuming social-economic activities while managing the disease control with potential case importation.

Keywords: COVID-19; case importation, stochastic discrete model; the effective reproduction number; resumption of work

---

Introduction

Coronaviruses are single-stranded RNA viruses, with a spherical shape and crown-like spines, from which their name derives. They generally cause mild respiratory infections but can sometimes be fatal. Since their first discovery and molecular characterization, three major outbreaks have occurred, in mainland China in 2003, in the Kingdom of Saudi Arabia in 2012 and in South Korea in 2015. The fourth large-scale outbreak, caused by a novel coronavirus termed as COVID-19 and with the capital city Wuhan of the Hubei province in China as the epicenter, has spread out globally [2, 6, 8, 11], causing approximately 80,000 cases and 2,400 deaths as of February 22nd 2020 [13].

Since January 23rd 2020, the Chinese authorities have been implementing unprecedented and increasingly stringent public health interventions, including a complete lock-down of Wuhan and neighboring cities, strict contact tracing and travel restrictions in the epicenter and across the country [13]. Incorporating these interventions into the classical epidemic models has resulted in reliable estimates of epidemiological characteristics of COVID-19 transmission in the epicenter and accurate near-casting the epidemic trend and peak time [18-20], in comparison with other earlier modeling studies focusing on analyzing the transmission risk based on data at the early outbreak stage [1, 3, 4, 17]. One important lesson gained from our recent experience suggests the need of updating the model parameters and model setting to reflect the rapid improvement of screening, diagnosing and testing techniques/procedures, accompanied by intensive contact tracing, quarantine and isolation.

The feasibility of enhancing model capacity to assess transmission risk and evaluating intervention effectiveness has also been substantially increased due to the improvement of the quality of the data, specially from other Chinese cities and provinces where the complete lock-down of major cities such as Wuhan in the Hubei province has prevented a significant portion of infection importation, so the reduction of imported cases coupled with intensive contact tracing produced some high quality data with rather complete information on symptom onset time and transmission-chain, as will be discussed in the data section below. This enhanced model capacity for near-casting is particularly important, since the gradually relaxing of the travel restriction demands proactive- and real-time assessment of the risk of a potential second outbreak in other Chinese cities. An objective of our study is to provide this assessment, and to use this assessment along with the incorporation of feasible public health interventions into our stochastic disease transmission model to suggest the scale of interventions for these cities to prevent a second outbreak. As COVID-19 spread has occurred in many regions globally, we believe our study provides an important tool to inform public health decisions regarding travel restriction and social distancing in these regions according to their public health capacity to maintain an effective

contact tracing, quarantine and isolation procedure.

In order to achieve the above research objectives, we examine the transmission characteristics of the COVID-19 epidemic in other areas outside Hubei province and quantify the risk of case importation into the current outbreak which is expected to end soon, and into a potential secondary outbreak due to the “resumption of work” in these areas. We focus on the effectiveness and implementation timeliness of prevention and control strategies, as well as the choice of timing for the “resumption of work” and its impact on the risk of a secondary outbreak.

Our analysis is based on the complete “tracking and recording” information of imported cases and local cases detected in the province of Shaanxi province [7], as this allows us to generate a data set that includes the time series of daily latent, infected and hospitalized cases. Since the outbreak scale in the Province of Shanxi was relatively small and this is subject to stochasticity in both case importation and local cases generated by the importation and since the Provincial data has detailed information on symptom onset and serial intervals, we adopt a method developed by Nishiura and Chowell [15, 16] to estimate the effective reproduction number, and we revise some of our early deterministic models to a novel discrete stochastic model to address the key issues relevant to preventing a secondary outbreak of COVID-19 in other provinces except for Hubei, due to the anticipated gradually relaxing of travel restriction and returning to normal social and economic activities.

Methods

Data

We obtained the data of laboratory-confirmed 2019-nCov cases in Shaanxi province from the “National Health Commission” of the People’s Republic of China and the “Health Commission” of the Shaanxi Province [7, 13]. It is worth mentioning that very detailed data information for each confirmed case has been released. This strict contact tracing strategy gave detailed information of illness onset, the first medical visit, quarantined period and confirmation. Then the data do not only include numbers, such as the number of newly reported cases, the cumulative number of reported confirmed cases, the cumulative number of cured cases and the number of death cases, and the cumulative numbers of quarantined and suspected cases (shown in Figure 1(A)), but also include the information such as the dates of illness onset, first medical visit, isolation, laboratory confirmation and discharge, and/or date of coming into Shaanxi (for imported cases). The data were released and analyzed anonymously.

Based on the recorded information, we can get the daily number of latent, infected and hospitalized individuals, and we call these data sets as traced data, as shown in Figure 1(A-C).

Note that that daily number of hospitalized cases means the first medical visit and then being quarantined rather than daily number of confirmed cases. Further, we collect the data of population mobility from the website of Baidu Qianxi (<http://qianxi.baidu.com>). It contains the trend of inflow of population to Shaanxi before the 2020 Spring Festival and after 2019 Spring Festival, as shown in Figure 2(B).

Estimation of the effective reproduction number $R(t)$

Let M_t and C_t^M be the number of newly imported cases and the number of newly confirmed imported cases on day t in Shaanxi province. T_M represents the duration from importation to confirmation for imported cases. For k consecutive days $j = t_1, \dots, t_k$, M_j is assumed to follow the Poisson distribution with mean λ_j (that has to be estimated). Given the daily number of newly confirmed imported cases $C_{s_1}^M, \dots, C_{s_m}^M$ on m consecutive days $s_1, \dots, s_m$ and the probability $p_{ij} = P(i - j \leq T_M < i - j + 1)$ that an imported case entered Shaanxi province on day j and was confirmed on day i . We assume that for each j , $q_j = \sum_{i=\max\{j, s_1\}}^{s_m} p_{ij}$ is positive. Then parameters $\lambda_{t_1}, \dots, \lambda_{t_k}$ can be estimated by $C_{s_1}^M, \dots, C_{s_m}^M$ and p_{ij} through deconvolution method. We use the Richardson-Lucy iterative algorithm [5] to solve this problem. The procedure is iterative according to the following formulas:

$$C_i^{M(n)} = \sum_{j=t_1}^i p_{ij} \lambda_j^{(n)} \quad , \quad \lambda_j^{(n+1)} = \frac{\lambda_j^{(n)}}{q_j} \sum_{i=\max\{j, s_1\}}^{s_m} \frac{p_{ij} C_i^M}{C_i^{M(n)}}$$

where $C_i^{M(n)}$ and $\lambda_j^{(n)}$ are fitted values of C_i^M and λ_j in the n -th iteration respectively. We stop the iteration when the error of fitting

$$\chi^2 = \frac{1}{m} \sum_{i=s_1}^{s_m} \frac{(C_i^{M(n)} - C_i^M)^2}{C_i^{M(n)}}$$

becomes small and the value of $\lambda_{t_1}^{(n)}, \dots, \lambda_{t_k}^{(n)}$ are reasonable. Because the method above requires that $q_j = \sum_{i=\max\{j, s_1\}}^{s_m} p_{ij}$ is positive, which means that at least some fraction of imported cases who entered Shaanxi province on day j would be confirmed during the period $s_1, \dots, s_m$, we then can determine the proper t_1 and t_2 such that $q_j > 0$ for $t_1 \leq j \leq t_2$ in terms of the distribution of T_M . By using the same method, given the daily number of confirmed local cases and the probability distribution of the duration from infection to confirmation for local cases, we can also estimate the daily number of newly infected local cases L_t on day t during some time period.

To estimate the effective reproduction number $R(t)$, we make an adaptation of the method in [15,16] to allow for imported cases. With the same notations and assumptions above, for $R(t)$ we have

$$E[L_t] = R(t) \sum_{\tau=1}^{t-1} (g_{\tau} L_{t-\tau} + h_{\tau} M_{t-\tau})$$

where $g_{\tau} = G(\tau) - G(\tau - 1)$ is the discretized distribution of the serial interval with $G(\tau)$ being the cumulative distribution function. Let T_h be the duration from importation of a primary imported case to infection of the secondary local case, $h_{\tau} = H(\tau) - H(\tau - 1)$ is the discretized distribution of T_h with $H(\tau)$ being the cumulative distribution function.

The model

On the basis of the clinical progression of the disease, epidemiological status of the individuals, and intervention measures we stratify the populations as susceptible (S), exposed (E), infected (I), hospitalized (H), recovered (R) compartments, the quarantined susceptible (S_q) and quarantined suspected individuals (B). The flow diagram is shown in Figure 3. Note that here the model framework extends the model structure in our previous study by including the quarantined suspected compartment [18-20], which consists of exposed infectious individuals resulting from contact tracing and individuals with influenza like illness and common fever needing clinical medication. The deterministic model is replaced by a discrete system with stochastic importation since the case numbers are relatively small with most of the early cases being imported in Shaanxi province [10, 12]. This modelling approach can well describe demographic stochasticity. The model equations are

$$\begin{cases} S_{t+1} = S_t - D_{11}(t) - D_{12}(t) - D_{13}(t) + D_{51}(t) + (1-f)D_{41}(t), \\ E_{t+1} = E_t + (1-q(t))D_{11}(t) - D_{21}(t) + P_E(\lambda_E), \\ I_{t+1} = I_t + D_{21}(t) - D_{31}(t) - D_{32}(t) - D_{33}(t) + P_I(\lambda_I), \\ B_{t+1} = B_t + q(t)D_{11}(t) + D_{13}(t) - D_{41}(t) + P_B(\lambda_B), \\ S_{qt+1} = c + D_{12}(t) - D_{51}(t), \\ H_{t+1} = H_t + D_{31}(t) + fD_{41}(t) - D_{61}(t) - D_{62}(t) + P_H(\lambda_H), \\ R_{t+1} = R_t + D_{33}(t) + D_{61}(t), \end{cases} \quad (1)$$

where the random variables in (1) can be defined by binomial $Bin(n, p)$ distributions as follows [10,12]:

$$\begin{aligned} D_{11}(t) &\sim Bin(S_t, P_{11}(t)), & D_{12}(t) &\sim Bin(S_t, P_{12}(t)), & D_{13}(t) &\sim Bin(S_t, P_{13}), \\ D_{21}(t) &\sim Bin(E_t, P_{21}), & D_{31}(t) &\sim Bin(I_t, P_{31}(t)), & D_{32}(t) &\sim Bin(I_t, P_{32}), \\ D_{33}(t) &\sim Bin(I_t, P_{33}), & D_{41}(t) &\sim Bin(B_t, P_{41}), & D_{51}(t) &\sim Bin(H_t, P_{51}), \\ D_{61}(t) &\sim Bin(H_t, P_{61}), & D_{62}(t) &\sim Bin(H_t, P_{62}) \end{aligned}$$

with probabilities:

$$\begin{aligned} P_{11}(t) &= 1 - \exp[-\beta c(t) \frac{I}{N} h], & P_{12}(t) &= 1 - \exp\left[-(1-\beta)c(t)q(t) \frac{I}{N} h\right], \\ P_{13} &= 1 - \exp(-mh), & P_{21} &= 1 - \exp(-\sigma h), \\ P_{31}(t) &= 1 - \exp[-\delta_I(t)h], & P_{32} &= 1 - \exp(-\alpha h), \\ P_{33} &= 1 - \exp(-\gamma_I h), & P_{41} &= 1 - \exp(-bh), \\ P_{51} &= 1 - \exp(-\lambda h), & P_{61} &= 1 - \exp(-\gamma_H h). \end{aligned}$$

Let constant m be the transition rate from susceptible class to the suspected compartment via general clinical medication due to fever or illness-like symptoms. The suspected individuals leave this compartment at a rate of b , with a proportion, f , if has been confirmed to be infected by the COVID-19, going to the hospitalized compartment, whilst the other proportion, $1 - f$, has been proven to be not infected by the COVID-19 and goes back to the susceptible class once recovery. Let the transmission probability be β and the contact rate be c . With contact tracing, we assume a proportion, q , of individuals exposed to the virus is quarantined, and can either move to the compartment B or S_q at a rate of $\beta c q$ or $(1 - \beta) c q$, depending on whether they are infected or not, while the other proportion, $1 - q$, consists of individuals exposed to the virus who are missed from contact tracing and move to the exposed compartment E at a rate of $\beta c (1 - q)$ once infected or stay in compartment S otherwise. All definitions of variables and parameters have been listed in Table 1.

The imported latent, infected and suspected and hospitalized cases for each day can be described by Poisson distributions $P_E(\lambda_E)$, $P_I(\lambda_I)$, $P_B(\lambda_B)$ and $P_H(\lambda_H)$ with parameters $\lambda_E, \lambda_I, \lambda_B$ and λ_H , respectively. Therefore, we have

$$P_E(E_t = m) = \frac{\lambda_E^m e^{-\lambda_E}}{m!}, \quad P_I(I_t = m) = \frac{\lambda_I^m e^{-\lambda_I}}{m!},$$

$$P_B(B_t = m) = \frac{\lambda_B^m e^{-\lambda_B}}{m!}, \quad P_H(H_t = m) = \frac{\lambda_H^m e^{-\lambda_H}}{m!}.$$

Fortunately, the number of imported cases to the above four compartments in Shaanxi province can be fully determined according to the traced data, as shown in Figures 2(A).

Since the implementation of the lock-down strategy in Wuhan on January 23rd 2020, the prevention strategies in other parts of mainland China have been continuously strengthened. In order to truly describe the constantly strengthening of control strategies, we should consider the contact rate, quarantine rate, diagnosis rate and cure rate as the functions of time t . In this model, we assume that the contact rate $c(t)$ is a decreasing function with respect to time t , given by

$$c(t) = \begin{cases} c_0, & \text{before Jan 23nd, 2020} \\ (c_0 - c_b)e^{-r_1 t} + c_b, & \text{else} \end{cases},$$

where c_0 denotes the baseline contact rate at the initial time (i.e. before January 23rd, 2020) with $c(0) = c_0$, c_b denotes the minimum contact rate under the current control strategies with $\lim_{t \rightarrow \infty} c(t) = c_b$, where $c_b < c_0$, and r_1 denotes the exponential decreasing rate in the contact rate.

Similarly, to characterize the enhanced contact tracing we define $q(t)$ as an increasing function with respect to time t , written as

$$q(t) = \begin{cases} q_0, & \text{before Jan 23nd, 2020} \\ (q_0 - q_m)e^{-r_2 t} + q_m, & \text{else} \end{cases}$$

where q_0 is the initial quarantined rate of exposed individuals with $q(0) = q_0 = 0$ in Shaanxi

province, q_m is the maximum quarantined rate under the current control strategies with $\lim_{t \rightarrow \infty} q(t) = q_m$ and $q_m > q_0$, and r_2 represents the exponential increasing rate in quarantined rate. We also set the transition rate $\delta_I(t)$ as an increasing function with respect to time t , thus the detection period $1/\delta_I(t)$ is a decreasing function of t with the following form:

$$\frac{1}{\delta_I(t)} = \begin{cases} \frac{1}{\delta_{I0}}, & \text{before Jan 23rd, 2020} \\ \left(\frac{1}{\delta_{I0}} - \frac{1}{\delta_{If}} \right) e^{-r_3 t} + \frac{1}{\delta_{If}}, & \text{else} \end{cases},$$

where δ_{I0} is the initial rate of confirmation, δ_{If} is the fastest confirmation rate, and r_3 is the exponential decreasing rate of the detection period. Definitely, $\delta_I(0) = \delta_{I0}$ and $\lim_{t \rightarrow \infty} \delta_I(t) = \delta_{If}$ with $\delta_{If} > \delta_{I0}$. Let T_c be the time point at which stringent control measures are introduced on Jan 23rd, 2020. Until February 4th 2020, there was no cured case in Shaanxi province, so the cure rate γ_H should be zero before February 4th 2020. The number of discharged patients has increased rapidly since February 7th 2020. Therefore, the cure rate is defined as a constant piecewise function based on the real situation.

Simulation of the resumption of work

When simulating the resumption of work, we assume that the movement trend of Shaanxi population after resumption of work is the same as that in same period of 2019. Let the average proportions of E, I, E_q, H that flow in the Shaanxi during January 17th to 22nd, 2020 be p_E , p_I , p_B , p_H , which can be calculated by the data of imported cases and population mobility. Let the proportions of E, I, E_q, H that flow in Shaanxi province after resumption of work be $\hat{p}_E$, $\hat{p}_I$, $\hat{p}_B$, $\hat{p}_H$. We initially assume that $\hat{p}_E$, $\hat{p}_I$, $\hat{p}_B$, $\hat{p}_H$ are the same as those of p_E , p_I , p_B , p_H , when the resumption of work begins on February 24th, then we consider the proportions $\hat{p}_E$, $\hat{p}_I$, $\hat{p}_B$, $\hat{p}_H$ may reduce 30%, 60% and even 90% corresponding to weaker and weaker epidemic.

Main results

Data implications

From the detailed description about traced data, we can know when and what is the state of each infected person and obtain the number of who are in latent (or quarantined latent), infected and hospitalized classes, as shown in Figure 1, which shows the evolution process of the epidemic in Shaanxi province. Figure 1(A) gives both newly reported cases and new cases from traced data, and comparison of two time series indicates that the new cases from traced data is about 7 days ahead of the reported cases data. In particular, Figure 1(B) and (C) give the epidemic status of cases confirmed before February 7th and 15th, respectively. Note that although Shaanxi province began to release the case numbers on January 23rd, 2020, the first imported latent case came to

Shaanxi as early as January 10th, 2020, and the majority of imported cases were latent and came into Shaanxi before the end of January, 2020, as shown in Figure 2(A). Figure 2(B) shows the correlation between the total imported cases and the number of population flow into the Shaanxi province, from which we calculate the proportion of those inflows that are infected.

We can also illustrate the spatial distribution of imported cases (circles) and local cases (squares) and disease evolution based on the traced data. It follows from Figure 4 that most cases are in Xi'an city, few cases are in the northern part of Shaanxi while relatively many cases are in the southern part of Shaanxi, which is adjacent to Hubei province. Therefore, we can estimate the impact of population migration on the epidemic situation in the neighboring areas of Hubei Province.

The detailed records reveal how long the patients stay in latent, infected and hospitalized compartments. Based on all cases reported in Shaanxi province we use the Kaplan-Meier method to estimate the median duration from illness onset to first medical visit (quarantined, denoted by T_2^W) is around 0.60 days, median duration from first medical visit (quarantined) to confirmation (denoted by T_3^W) is around 3.43 days, shown in Figure 5. In particular, for imported cases, the median duration from importation to illness onset (denoted by T_1^I) is around 2.38 days with longest incubation duration of 19 days, median duration from illness onset to first medical visit (quarantined, denoted by T_2^I) is around 1 days with the longest (or shortest) duration of 15 (or 0) days; median duration from the first medical visit to confirmation (denoted by T_3^I) is around 3.05 days with the longest (or shortest) duration of 13 (or 1 days). For local cases median duration from illness onset to first medical visit (quarantined, denoted by T_2^L) is around 0.29 days with the longest (or shortest) infectious duration of 10 (or 0) days; median duration from the first medical visit to confirmation (denoted by T_3^L) is around 3.9 days with the longest (or shortest) duration of 16 (or 1 days). This analysis indicates that local people went to medical visit when having illness onset much quicker than the imported individuals. Further, these durations are relatively smaller than those for mainland China or Hubei province [?], which implies that the control measures implemented in Shaanxi are timely and medical resources are relatively sufficient.

Control/ effective Reproduction numbers

From the detailed record we can find 21 transmission chains (or transmission trees) which involve 60 cases with 13 imported cases. For each transmission chain we obtained when and who infected whom, and calculated the mean control reproduction number of 1.48 with standard deviation (std) 0.98. In particular, it follows from 21 transmission chains that the number of secondary infections for each seeded case yields 2, 2, 1, 5, 1, 0, 2, 2, 2, 1, 1, 1, 2, 1, 2, 1, 1, 1, 1,

1. As mentioned before, the early reported cases in Shaanxi are mainly imported cases before February 3rd, 2020. Then, given seed infected individuals who may get infection from January 23rd to February 3rd, 2020, we had the transmission 13 chains (No. 1-13 as shown in Figure 6), which gives mean control reproduction number of 1.69 with std 1.18.

To investigate the new infections we need to know the duration from importation to confirmation (denoted by T_M for imported cases) and the duration from infection to confirmation for local cases (denoted by T_L). By fitting a Weibull distribution to the traced data imported cases confirmed before February 16th, 2020 we estimate the estimated mean and standard deviation of T_M as 10.68 days and 4.95 days. Let T_L be assumed to follow a Weibull distribution, which is the sum of the incubation period T_{L1} and the duration from illness onset to confirmation for local cases T_{L2} . It follows from [3, 4, 6, 8, 11, 14, 17] that the mean and standard deviation of T_{L1} were 5.2 days and 3.91 days. The mean and standard deviation of T_{L2} were estimated to be 7.28 days and 3.89 days on the data of detailed information of local cases confirmed before February 16th, 2020. Then the mean and standard deviation of T_L was calculated to be 12.48 days and 5.52 days. Then we can estimate the daily new infections at the same period on the basis of the daily number of reported confirmed cases and the distribution of T_M and T_L by deconvolution method Figure 7(A). It is interesting to note that we can get the part of new infections before February 12th, 2020 from the recoded traced data in Shaanxi province, shown in Figure 7(B). It shows that the estimated new infections agree well with the traced data in terms of trend. We note that for some cases, their information is not complete, only the dates of illness onset or first medical visit are recorded, and hence they cannot be considered as new infections at certain dates. That is why the real new infections from traced data are less than the estimated ones, shown in Figure 7(B).

To estimate the effective reproduction number we let the distribution of serial interval g_τ be chosen as gamma distribution with mean of 7.5 days and standard deviation of 3.4 days [4, 11]. Because the distribution of the duration from importation of a primary imported case to infection of the secondary local case h_τ is unknown, we initially set h_τ be equal to g_τ and then estimate the effective reproduction number $R(t)$, as shown in Figure 7(C). It shows the effective reproduction number started to decline on January 23rd, 2020 when the prediction and control strategies was strengthened, was less than 1 since around January 27th, 2020, and stabilized at almost 0 recently, which means few new infections occur. In order to investigate the variation of the effective reproduction number with T_h varies, sensitivity analysis was carried out by changing the mean of T_h from 1 to 7 days, as shown in Figure 7(D). It shows that the more the mean of T_h , the greater duration of the $R(t)$ being greater than 1 and the earlier the $R(t)$ exceeds the unity.

Does resumption of work induce the second outbreak

Since the Spring Festival holiday most people have been at home, when do they return back to work/study is a challenging problem. Consequently, we investigate the effect of population movement on infections, especially on the possibility of inducing the second outbreak. To this end, we initially estimate the unknown parameters by least square methods, listed in Table 1, by fitting the discrete stochastic model to the traced data between January 10th to February 7th (purple curve with dots), and by using the mean parameter values of 1000 times of least square estimations (listed in Table 1) we get 500 stochastic trajectories of model (1), as shown in Figure 8(A-C). Figure 8(D) shows the estimated susceptible individuals vary significantly. Mean estimated values (green curves) and 95% confidence interval (CI) of E_t , I_t , H_t compartments are shown in Figure 8(E-G). Note that on February 4th, 2020 the first cured individual was covered, and the estimated recovery rate γ_H was so small that mean value of H_t persistently and slowly increases, shown in Figure 8(C). However, tripling the recovery rate can induce the value of H_t decline. By using the data between January 10th to February 16th, 2020 (purple curve with dots) we verify the model and find the traced data are within the 95% CI of estimated E_t , I_t and H_t . It is worth mentioning that the number of suspected cases B_t (upper green curve in Figure 8(D)) quickly increases since people, once having illness-like symptom or fever, were likely access to medical visit and were considered to be suspected individuals and isolated. While the case-driven cumulative quarantined individuals via contact tracing (lower green curve in Figure 8(D)) becomes to stabilize, and is less than the number of suspected individuals due to relatively small number of confirmed cases. The total number of these two classes indicates the prevention and control strategies implemented in Shaanxi province are strong.

To investigate the effect of resumption of work on the epidemic, especially on possible second outbreak, by which we may suggest the public health sectors when workers can return back to Shaanxi, we predicate the disease infection by using the data on imported cases and moving from other cities to Shaanxi province (shown in Figure 2). Suppose population to start to return back to Shaanxi province from February 24th to March 9th, 2020, we can see from Figure 9 (A-B) that there is a large second outbreak. If the proportions of imported cases are decreased by 30%, 60% and 90%, the second outbreak becomes smaller and smaller, shown in Figure 9(C-D), (E-F), and (G-H), respectively. Further, if the inflow mode is changed into the intermittent inflow, that is, population are allowed to move into Shaanxi every other week, we can from Figure 10 that there are one more outbreaks with low peak values. Decreasing the imported proportions greatly reduces the peak values of the following outbreaks. Comparing Figure 9 and Figure 10 indicates that intermittent inflow mode may induce small outbreaks and a relatively good choice for resumption of work.

Discussion

Shaanxi province is located northwest of China, adjacent to Hubei in the south. Therefore, except for Xi'an city, the capital of Shaanxi, many confirmed cases are in the southern part of Shaanxi province. It is interesting to note that Health Commission of Shaanxi Province released the case number every day with very detailed information, by which we can know when and who infected whom. Then except for the new reported (or cumulative) number of confirmed cases, we can get the daily number of individuals who are in latent (or quarantined latent), infected and hospitalized compartments, which gives the evolution process of the epidemic in Shaanxi province. We can get the durations from illness onset to the first medical visit, to the hospitalization, and to confirmation of the case. These durations are relatively smaller than those for mainland China or Hubei province [3, 4, 6, 8, 11, 14, 17], which implies that the control measures implemented in Shaanxi are timely and medical resources are relatively sufficient.

From the detailed data description, we can obtain the transmission chains over the time, which enable to calculate the control reproduction number under the implementation of stringent control strategies. The mean control reproduction number ranges from 1.48 to 1.69, and the distribution of these chains agrees well with the effective reproduction number. Note that the effective number has been declining, and is less than unity since January 27th 2020. Recently it has stabilized at almost zero, meaning that no new infections occur, which seem to be in line with real data, as shown in Figures 1 and 7 [7].

In this study we extend our deterministic model framework [9, 18-20] to a discrete system incorporating stochastic modeling of imported cases to represent the relatively small number of cases observed in the Shaanxi province, with about half of them being imported cases. It is worth noting that traced data is around 7 days ahead of the reported ones, as illustrated in Figure 1(A). Moreover, detailed data records revealed that the time interval between the first visit to a doctor and confirmation was longer, with a median of around 4 days. Then, the number of reported cases is not suitable to identify the model or estimate the parameters due to delays such as report delay, confirmation delay, etc. Therefore, it is not accurate to associate the number of confirmed cases with the number of cases in hospital, which may seriously overestimate the scale of the epidemic. Therefore, the multiple traced data sets have been employed to identify the proposed discrete stochastic model and estimate the unknown parameters in this work. Consequently, we use the number of individuals in E, I, and H classes from traced data rather than the newly (or cumulative) reported number to parameterize the model.

We investigate the impact of population movement (resumption of work) on the potential second outbreak by using movement data from other cities to Shaanxi province. Our main results indicate that the proportion of imported cases is a key variable in order to determine

whether the second outbreak is likely or not to occur. The greater the proportion of imported cases the larger the second outbreak has. Furthermore, intermittent importation may induce one further outbreak episode but with small peak values.

From our study, we can conclude that it is very important to study the epidemic situation of COVID-19 by using data from contact tracing rather than simply using the number of confirmed cases, especially in the areas where the imported cases are the main source of infection. This can not only provide important data information for the accurate evaluation of the epidemic situation in other areas outside Hubei Province, but also provide important and accurate data supporting the evaluation of the effectiveness and timeliness of public health intervention strategies in different areas.

In the present investigation, we devised a stochastic model that can be used as an important model framework and methodology for the evaluation of imported cases on local epidemic in the Shaanxi province or other similar provinces. From the data analysis, it seems that the stringent public health interventions adopted by the Chinese authorities are effective in controlling the infection. According to data from contact tracing, the first imported cases in the Shaanxi Province occurred earlier than January 10th 2020, and most of the imported cases were imported to the Shaanxi Province before the implementation of the national travel restrictions on January 23rd 2020. Therefore, the earlier the implementation of the travel restrictions, the greater the reduction of infected cases by COVID-19 in the Shaanxi province has. An intermittent population inflow with low proportion of imported cases may significantly mitigate the transmission risk. Thus, if the public health authorities can strictly monitor and reduce the number of imported cases, it would be possible for the Shaanxi province to return back and resume working continuously.

Conclusions

It is very important to study the epidemic situation of COVID-19 by using traced data rather than simply using the number of confirmed cases, especially in the areas where the imported cases are the main and the epidemic situation is not serious. This can not only provide important data information for the accurate evaluation of the epidemic situation in other areas outside Hubei Province, but also provide important and accurate data support for the evaluation of the effectiveness and timeliness of public health intervention strategies in different areas.

We proposed a stochastic model which can provide an important model framework and methodology for the evaluation of imported cases on local epidemic in the Shaanxi province (and can eventually be adapted for other similar provinces of China). Data analysis reveals that strengthening the public health interventions, tracing imported cases and improving the

confirmation rate are effective and timely after January 23rd 2020. According to the traced data shown in Figure 1, the first imported cases in Shaanxi Province were earlier than January 10, and most of the imported cases were imported to Shaanxi Province before the national travel restrictions were implemented on January 23rd, as shown in Figure 2. Therefore, the earlier implementation of travel restriction will greatly reduce the epidemic situation of COVID-19 in Shaanxi province, which was not serious.

The findings suggest that keeping an intermittent population inflow with a low proportion of imported cases may induce a small second outbreak and hence decrease the transmission risk. Thus, if we can strictly monitor and reduce the number of imported cases, it is possible for Shaanxi to resume work continuously, especially intermittently.

Acknowledgement: This research was funded by the National Natural Science Foundation of China (grant numbers: 11631012 (YX, ST), 61772017 (ST)), and by the Canada Research Chair Program (grant number: 230720 (JW) and the Natural Sciences and Engineering Research Council of Canada (Grant number:105588-2011 (JW)).

References

1. Boldog P, Tekeli T, Vizi Z, Dénes A, Bartha F A, Röst G. Risk assessment of novel coronavirus COVID-19 outbreaks outside China. *J Clin Med* 2020, 9:571.
2. Cohen J, Normile D. New SARS-like virus in China triggers alarm. *Science* 2020,367: 234-235.
3. Du Z, Wang L, Chauchemez S, Xu X, Wang X, Cowling BJ, et al. Risk for transportation of 2019 novel coronavirus disease from Wuhan to other cities in China. *Emerg Infect Dis* 2020, DOI: 10.3201/eid2605.200146.
4. Gilbert M, Pullano G, Pinotti F, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet* 2020, DOI: [https://doi.org/10.1016/S0140-6736\(20\)30411-6](https://doi.org/10.1016/S0140-6736(20)30411-6).
5. Goldstein E, Dushoff J, Ma J, Plotkin JB, Earn DJ, Lipsitch M. (2009). Reconstructing influenza incidence by deconvolution of daily mortality time series. *PNAS* 2009, 106 (51): 21825-21829.
6. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of 2019 novel coronavirus infection in China. *medRxiv* 2020 doi:10.1101/2020.02.06.20020974.
7. Health Commission of Shaanxi Province. Available at <http://sxwjw.shaanxi.gov.cn/col/col863/index.html> [Accessed 20 Feb 2020]

8. Heymann D L, Shindo N. COVID-19: what is next for public health? *The Lancet*, 2020, DOI:[https://doi.org/10.1016/S0140-6736\(20\)30374-3](https://doi.org/10.1016/S0140-6736(20)30374-3).
9. Keeling M J, Rohani P. *Modeling infectious diseases in humans and animals*. Princeton University Press. 2008..
10. Lekone P E, Finkenstädt B F. Statistical Inference in a stochastic epidemic SEIR model with control intervention: Ebola as a Case Study. *Biometrics*, 2006, 62: 1170 – 1177. DOI: 10.1111/j.1541-0420.2006.00609.x
11. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020, doi: 10.1056/NEJMoa2001316.
12. Mode C J, Sleeman C K. *Stochastic Processes in Epidemiology*. Singapore: World Scientific, 2000.
13. National Health Commission of the People's Republic of China. <http://www.nhc.gov.cn>.
14. Parry J. China coronavirus: cases surge as official admits human to human transmission. *BMJ* 2020; 368: m236.
15. Nishiura H. Correcting the actual reproduction number: a simple method to estimate R_0 from early epidemic growth data. *Int J Env Res Pub Health* 2010, 7 (1): 291-302.
16. Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In *Mathematical and statistical estimation approaches in epidemiology* (pp. 103-121), Springer, Dordrecht 2009.
17. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digital Health* 2020, DOI: [https://doi.org/10.1016/S2589-7500\(20\)30026-1](https://doi.org/10.1016/S2589-7500(20)30026-1).
18. Tang B, Bragazzi NL, Li Q, Tang S, Xiao Y, Wu J. An updated estimation of the risk of transmission of the novel coronavirus (COVID-19). *Infect Dis Model* 2020,5:248-255.
19. Tang B, Xia F, Tang S, Bragazzi N L, Li Q, Sun X, Liang J, Xiao Y, Wu J. The evolution of quarantined and suspected cases determines the final trend of the 2019-nCoV epidemics based on multi-source data analyses. Available at SSRN: <https://ssrn.com/abstract=3537099>.
20. Tang B, Wang X, Li Q, et al. Estimation of the Transmission Risk of the COVID-19 and Its Implication for Public Health Interventions. *J Clin Med* 2020, 9: 462.

Table 1: Parameter estimates for the 2019-nCov epidemics in Shaanxi Province, China.

Parameter	Definitions	Shaanxi	Source
c_0	Contact rate at the initial time	15	9
c_b	Minimum contact rate under the current control strategies	6.0371	Estimated
r_1	Exponential decreasing rate of contact rate	0.051	Estimated
β	Probability of transmission per contact	0.21	Estimated
q_0	Quarantined rate of exposed individuals at the initial time	0	Data
q_m	Maximum quarantined rate of exposed individuals under the current control strategies	0.5228	Estimated
r_2	Exponential increasing rate of quarantined rate of exposed individuals	0.0799	Estimated
m	Transition rate of susceptible individuals to the suspected class	0.007	Estimated
b	Detection rate of the suspected class	0.0397	Estimated
f	Confirmation ratio: Transition rate of exposed individuals in the suspected class to the quarantined infected class	0.01	Estimated
σ	Transition rate of exposed individuals to the infected class	1/2.38	Data
λ	Rate at which the quarantined uninfected contacts were released into the wider community	1/12	9
δ_{I0}	Initial transition rate of symptomatic infected individuals to the quarantined infected class	1/4.01	Data
δ_{If}	Fastest diagnose rate	0.4076	Estimated
r_3	Exponential decreasing rate of diagnose rate	0.0832	Estimated
γ_I	Recovery rate of infected individuals	0.0497	9
γ_H	Recovery rate of quarantined infected individuals	0.0199	Estimated
α	Disease-induced death rate	0	Data
Initial values	Definitions	Shaanxi	Source
$S(0)$	Initial susceptible population	9.9895×10^4	Estimated
$E(0)$	Initial exposed population	0	Data
$I(0)$	Initial infected population	0	Data
$B(0)$	Initial suspected population	0	Data
$S_q(0)$	Initial quarantined susceptible population	0	Data
$H(0)$	Initial quarantined infected population	0	Data
$R(0)$	Initial recovered population	0	Data

Figure 1. (A) Newly reported cases and daily new cases from traced data. The daily number of exposed, infected and hospitalized individuals in Shaanxi province based on the traced data by February 9th (B), and by February 16th 2020 (C)

Figure 2. (A) The imported cases for the four compartments during Jan 10th to Feb 7th, 2020. (B) The total imported cases and number of flow into the Shaanxi province from other regions from January 10th to February 3rd.

Figure 3. Diagram of the model adopted in the study for simulating the COVID-19 infection. Interventions including intensive contact tracing followed by quarantine and isolation are indicated. The gray compartment means suspected case compartment consisting of contact tracing Eq and fever clinics.

Figure 4 Spatial distribution of imported cases (circles) and local cases (squares) based on the traced data. The imported cases include exposed (yellow), infected (magenta), quarantined infected (red) and hospitalized (black) individuals, the local cases include infected (green), quarantined infected (red) and hospitalized (black) individuals.

Figure 5. Survival analysis by using Kaplan-Meier method to estimate the duration from importation to illness onset (denoted by T_1), the duration from illness onset to first medical visit (quarantined, denoted by T_2) and the duration from first medical visit (quarantined) to confirmation (denoted by T_3) for imported, local and whole cases in Shaanxi province.

Figure 6 Transmission chains along the time. Blue M_i represent the imported cases and greens L_j are local cases, and i, j denote the identifier numbers in the system of Health Commission of Shaanxi province.

Figure7 The back-calculated time series on number of new infections for imported/local individuals based on the reported confirmed cases (A), comparison of number of new infections from back calculation and traced data (B), the estimated reproduction number $R(t)$ based on the back-calculated number of illness onset (C).

Figure 8. Data fittings with 500 simulations (A-C). The variations of the number of susceptible population are significantly (D). The mean values (green curve) and 95% CI (grey curve) for $E(t)$, $I(t)$ and $H(t)$ compartments (E-G). D: the number of suspected individuals $B(t)$ (upper green curve) and the cumulative number of the quarantined individuals (lower green curve).

Figure 9: Effect of continuous inflow on the number of exposed individuals (A) and infected individuals (B) with baseline parameter values, with 30% decreased inflow (C-D), with 60% decreased inflow (E-F), with 90% decreased inflow (G-H). Here the inflow has been lasting for 24 days (from February 24th to March 18th).

Figure 10: Effect of intermittent inflow on the number of exposed individuals (A) and infected individuals (B) with baseline parameter values, with 30% decreased inflow (C-D), with 60% decreased inflow (E-F), with 90% decreased inflow (G-H). Here the inflow is on for the first (from February 24th to March 1st) and third weeks (from March 9th to 15th) and off for the second and fourth week and the following days.