https://doi.org/10.1016/j.jtbi.2021.110861

Title page

Full title: Shrinkage in serial intervals across transmission generations of COVID-19

Shi Zhao[1,2,+,*], Yu Zhao[3,+], Biao Tang[4,5], Daozhou Gao[6], Zihao Guo[1], Marc KC Chong[1,2], Salihu S Musa[7,8], Yongli Cai[9], Weiming Wang[9], Daihai He[7,*], and Maggie H Wang[1,2]

1 JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China

2 CUHK Shenzhen Research Institute, Shenzhen, China

3 School of Public Health and Management, Ningxia Medical University, Yinchuan, Ningxia, China

4 School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China

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

6 Department of Mathematics, Shanghai Normal University, Shanghai, China

7 Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China

8 Department of Mathematics, Kano University of Science and Technology, Wudil, Nigeria

9 School of Mathematics and Statistics, Huaiyin Normal University, Huaian, China

+ These authors contribute to this study equally, and thus they are joint first authors.

* Correspondence to: zhaoshi.cmsa@gmail.com (SZ), or daihai.he@polyu.edu.hk (DH).

Email addresses of all authors

SZ: zhaoshi.cmsa@gmail.com; YZ: zhaoyuzy123@163.com; BT: btang66@yorku.ca; DG: dzgao@shnu.edu.cn; ZG: guozihao9602@163.com; MKCC: marc@cuhk.edu.hk; SSM: salihu-sabiu.musa@connect.polyu.hk; YC: yonglicai@hytc.edu.cn; WW: weimingwang2003@163.com; DH: daihai.he@polyu.edu.hk; MHW: maggiew@cuhk.edu.hk.

Abstract

One of the key epidemiological characteristics that shape the transmission of coronavirus disease 2019 (COVID-19) is the serial interval (SI). Although SI is commonly considered following a probability distribution at a population scale, recent studies reported slight shrinkage (or contraction) of the mean of effective SI across transmission generations or over time. Here, we develop a likelihood-based statistical inference framework with truncation to explore the change in SI across transmission generations after adjusting the impacts of case isolation. The COVID-19 contact tracing surveillance data in Hong Kong are used for exemplification. We find that for COVID-19, the mean of individual SI is likely to shrink with a factor at 0.72 per generation (95%CI: 0.54, 0.96) as the transmission generation increases, where a threshold may exist as the lower boundary of this shrinking process. We speculate that one of the probable explanations for the shrinkage in SI might be an outcome due to the competition among multiple candidate infectors within a cluster of cases. Thus, the nonpharmaceutical interventive strategies are crucially important to block the transmission chains, and mitigate the COVID-19 epidemic.

Keywords: COVID-19; serial interval; transmission generation; contact tracing; statistical modelling.

1 Introduction

The transmission dynamics of an infectious disease are partially determined by the pathogen's infectiousness and the course of the transmission (He et al., 2020b; Kutter et al., 2018; Riou and Althaus, 2020; Tuite and Fisman, 2020; Wallinga and Lipsitch, 2007; Xu et al., 2020; Yan, 2008; Zhao, 2020a; Zhao et al., 2020e). The serial interval (SI), which is defined as the time interval between the symptoms onset dates of an infector and of the associated infectee (Fine, 2003; Milwid et al., 2016; Vink et al., 2014; White et al., 2009), is widely used to measure the duration of the transmission generation. As the most efficient proxy of the generation time (GT) (Wallinga and Teunis, 2004), SI is one of the crucial epidemiological parameters in describing the transmission process as well as the growth patterns of an outbreak (Champredon and Dushoff, 2015a; Kenah et al., 2008; Wallinga and Lipsitch, 2007; Yan, 2008).

As a contagious disease, the coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was firstly reported in 2019 (Huang et al., 2020; Leung et al., 2020; Li et al., 2020b; Parry, 2020; Zhao et al., 2020c), and rapidly spread to over 200 countries and territories, which poses a serious threat to global health. In response to the large-scale COVID-19 outbreaks, the World Health Organization (WHO) declared a public health emergency of international concern on January 30, 2020 (World Health Organization, 2020), which soon became a pandemic. As of February 14, 2021, there have been over 100 million confirmed COVID-19 cases worldwide with over 2 million associated deaths (2021).

To date, the transmission process of COVID-19 has been characterized and reconstructed both empirically and theoretically (Adam et al., 2020; He et al., 2020b; Kwok et al., 2020; Li et al., 2020b; Luo et al., 2020; Ren et al., 2021; Tindale et al., 2020; Wu et al., 2020a; Xu et al., 2020). In a number of existing literature, SI is commonly considered following a universal distribution at the population (or herd) scale for many well-known respiratory infectious diseases (Assiri et al., 2013; Cowling et al., 2009; Leung et al., 2004; Vink et al., 2014), which also occurs for COVID-19 (He et al., 2020b; Li et al., 2020b; Nishiura et al., 2020; Wang et al., 2020). In other words, SI was considered as a fixed distribution across transmission generations. However, two recent studies reported that SI appears with slight discrepancies across different transmission generations according to the summary statistics at populational scale (Li et al., 2020a; Ma et al., 2020). Inspiring by their findings, we suspect there may exist a solid difference in the mean SI in consecutive generations in a transmission chain.

In this study, we develop a statistical framework to explore the change in the SI across transmission generations after adjusting the impacts of case isolation. For exemplification, we quantify the change in SI by using the COVID-19 contact tracing surveillance data in Hong Kong. We explore the mechanism that drives the change in SI, and we also demonstrate its effects on shaping the transmission of COVID-19.

2 Methods

2.1 Conceptualization and statistical parameterization

We denote the SI of an infected individual, i.e., infector, by τ that follows a probability density function (PDF) $h(\tau)$ with mean μ and standard deviation (SD) σ. A transmission chain is composed by two consecutive transmission pairs, in which the infectee in the former transmission

pair acts as the infector in the latter transmission pair, see Fig 1. Regarding each case, we name the transmission pair between the infector of this case and himself by 'former transmission pair', and name the transmission pair between this case and his infectee by 'latter transmission pair'. As such, for convenience, we name the SI in the former transmission pair by 'former SI' and denoted by $\tau^{(F)}$, and the SI in the latter transmission pair by 'latter SI' and denoted by $\tau^{(L)}$. Here, we note that the superscript, i.e., '(F)' or '(L)', is used merely as a label instead of as a power.

We explore the changing patterns in SI across transmission generations. In the same transmission chain, an intuitive statistical relation between $\tau^{(F)}$ and $\tau^{(L)}$ in Eqn (1),

$$\mathbf{E}[\tau^{(L)}] = \lambda \cdot \mathbf{E}[\tau^{(F)}], \quad (1)$$

is considered, where $\mathbf{E}[\cdot]$ denotes the expectation function. The parameter λ is the change ratio between the means of two consecutive SIs, which is a positive constant to be determined. Straightforwardly, there exist iterative changes in mean SI across transmission generations, if $\lambda \neq 1$, while the mean SI may be a constant, if $\lambda = 1$. Hence, the relation in Eqn (1) can be examined by checking whether $\lambda = 1$ holds under the null hypothesis.

2.2 Likelihood-based inference framework

With the PDF $h(\tau)$ for the individual SI, the (baseline) likelihood framework, denoted by L_0, can be formulated in Eqn (2). That is

$$L_0(\lambda) = \prod_i \left[h^{(L)}\left(\tau_i^{(L)} \middle| \lambda, \tau_i^{(F)}\right) \cdot h^{(F)}\left(\tau_i^{(F)} \middle| \lambda, \tau_i^{(L)}\right) \right], \quad (2)$$

where the subscript i denotes the i-th transmission chain. For $h^{(L)}(\cdot|\lambda, \tau^{(F)})$, the mean of $h^{(L)}$ is given as $\mu^{(L)} = \tau^{(F)} \cdot \lambda$ according to the relation in Eqn (1). By contrast, for $h^{(F)}(\cdot|\lambda, \tau^{(L)})$, the mean of $h^{(F)}$ is given as $\mu^{(F)} = \tau^{(L)}/\lambda$. The SD of $h(\cdot)$, i.e., σ, is modelled as a function of μ. Due to the lack of information about the dispersion of the individual SI, as well as small sample size, we consider three scenarios of σ that cover a wide range of the possible situations. For a given individual infector, they include

- scenario (**I**), a large SD: $\sigma = |\mu|$, which refers to the scale of the coefficient of variation (CV) estimated in previous studies (Adam et al., 2020; Ali et al., 2020; Du et al., 2020; He et al., 2020b; Kwok et al., 2020; Nishiura et al., 2020; Tindale et al., 2020; Xu et al., 2020; You et al., 2020; Zhao et al., 2020f) and considered as an upper bound of SD;
- scenario (**II**), a moderate SD: $\sigma^2 = |\mu|$, which is assumed having a Poisson-like feature; and
- scenario (**III**), a small SD: $\sigma = 1$, which is assumed and considered as a lower bound of SD.

The script "$^{(L)}$" or "$^{(F)}$" is omitted here for simply convenience. We remark that, on one hand, for scenarios (**I**) and (**II**), the SD is depended on the mean SI of the generation, which indicates SD is not same between generations. Since the mean SI shrinks across transmission generations, the SD under these two scenarios will also change. On the other hand, the scenarios (**III**) reflected a condition that SD is fixed across transmission generation. The three scenarios here covered a wide range of SD of SI for COVID-19, which should include the most realistic situation. We acknowledge that the information about the SD of individual SI may improve the analysis. As our research target is focusing on the mean SI (μ), the settings in SD will not affect our conclusions.

With the mean and SD, the function $h(\cdot)$ can be formulated by some widely adopted PDFs. We consider three different PDFs. They are

- Normal distribution as a representative of symmetric distributions defined on all real numbers (Ali et al., 2020; Du et al., 2020; Forsberg White and Pagano, 2008; Ma et al., 2020; Xu et al., 2020; Yang et al., 2020; You et al., 2020);
- Gumbel distribution as a representative of asymmetric distributions defined on all real numbers (Ali et al., 2020; Xu et al., 2020); and
- Gamma distribution as a representative of asymmetric distributions defined on positive numbers (Ali et al., 2020; Cowling et al., 2009; Du et al., 2020; Ferretti et al., 2020; Ganyani et al., 2020; He et al., 2020b; Li et al., 2020b; Ma et al., 2020; Nishiura et al., 2020; Ren et al., 2021; Tindale et al., 2020; Vink et al., 2014; Wang et al., 2020; Xu et al., 2020; Zhao, 2020b; Zhao et al., 2020f).

We select the scenario of SD and distribution of $h(\cdot)$ according to the fitting performance in terms of the Akaike information criterion with a correction for small sample sizes (AICc).

In addition, as pointed out in (Nishiura et al., 2020), the baseline likelihood in Eqn (2) might lead to an underestimation of SI due to the interval-censoring issue. Hence, according to the truncation scheme previously developed in (Zhao et al., 2020f), which accounts for the effects of each infector's isolation, we adjust for the truncation bias by an improved likelihood function, L, in Eqn (3). We have

$$L(\lambda) = \prod_i \left[\frac{h^{(\mathrm{L})}\left(\tau_i^{(\mathrm{L})} \middle| \lambda, \tau_i^{(\mathrm{F})}\right)}{H^{(\mathrm{L})}\left(d_i^{(\mathrm{L})} \middle| \lambda, \tau_i^{(\mathrm{F})}\right)} \cdot \frac{h^{(\mathrm{F})}\left(\tau_i^{(\mathrm{F})} \middle| \lambda, \tau_i^{(\mathrm{L})}\right)}{H^{(\mathrm{F})}\left(d_i^{(\mathrm{F})} \middle| \lambda, \tau_i^{(\mathrm{L})}\right)} \right], \tag{3}$$

where $H(\cdot)$ is the cumulative distribution function (CDF) of $h(\cdot)$, and the letter d denotes the duration from the onset date of an infector to the person's isolation date. All other notations are the same as those in Eqn (2).

The parameter λ is estimated under both truncated and non-truncated schemes by using the maximum likelihood estimation (MLE). The AICc is employed for model selection. The 95% confidence interval (95%CI) is calculated by using the profile likelihood estimation framework with the cutoff threshold of a Chi-square quantile (Cai et al., 2021; Fan and Huang, 2005; He et al., 2020a; Lin et al., 2018; Zhao et al., 2020a).

All analyses are conducted in the **R** statistical software (version 3.5.1), and no specific package is used.

2.3 COVID-19 surveillance data in Hong Kong

The COVID-19 surveillance data are originally released by the Centre for Health Protection (CHP) of Hong Kong (Centre for Health Protection, 2020), and used in (Adam et al., 2020) previously. According to the data description in (Adam et al., 2020), a total of 1038 laboratory-confirmed SARS-CoV-2 infections as of May 7, 2020, were initially screened. In Hong Kong, each contact of a confirmed COVID-19 case, defined as who has prolonged face-to-face interaction with a case, is traced and mandatorily quarantined for 14 days, regardless of symptom appearance. Then, each transmission pair, i.e., the 'infector-and-infectee' pair, can be reconstructed from the contact tracing records. A total of 169 transmission pairs including 27 asymptomatic transmission pairs for either infector or infectee, which are directly collected via https://github.com/dcadam/covid-19-sse/blob/master/data/transmission_pairs.csv, are identified for further screening.

In this study, we focus on the (169 – 27 =) 142 symptomatic transmission pairs in Hong Kong. We identify the infectee who acts as an infector in other transmission pairs, i.e., the 'secondary case' in Fig 1, by matching all combinations of the 142 transmission pairs. We reconstructed the transmission chain with 3 generations including primary case, secondary case, and tertiary case, which is illustrated as the 'secondary case' in Fig 1. A total of 21 transmission chains are extracted, and presented in Fig 2.

Since the isolation period of each infector is unavailable, we consider the case confirmation date as a proxy of the isolation starting time with the presumption that the isolation starts immediately after confirmation. Hence, the change ratio of SI, λ, can be estimated from these transmission chain data in Hong Kong by using the analytical framework in Section 2.2.

2.4 Sensitivity analysis

To evaluate the estimating sensitivity, an alternative formulation, similar to the relationship in Eqn (1), is adopted to repeat the estimation with the dataset from Hong Kong. The alternative relationship between the former and latter SIs is formulated in Eqn (4).

$$\mathbf{E}[\tau^{(\mathrm{L})}] = \lambda \cdot [\mathbf{E}[\tau^{(\mathrm{F})}] - T_{\mathrm{c}}] + T_{\mathrm{c}}, \quad (4)$$

where the term T_{c} (≥ 0) indicates the lower bound of the SI as generation increases. Other terms have the same meanings as those in Eqn (1). Straightforwardly, Eqn (1) and Eqn (4) will be equivalent, if $T_{\mathrm{c}} = 0$. Thus, the intuition of Eqn (4) is of the same fashion as that of the Eqn (1).

We estimate both T_{c} and λ simultaneously with the likelihood profiles and estimation procedures in Section 2.2. The model selection is conducted referring to the lowest AICc. We check the consistency of the λ estimates, and whether T_{c} is significantly larger than 0.

2.5 Exploratory explanation of the mechanisms behind the change in SI

In this section, we develop statistical models to explore two possible, but not verified, mechanisms behind the change in SI, their effects in shaping the transmission process, and their reasonability.

2.5.1 Exploration #1: changes in latent period and infectious period

In exploration #1, we consider a hypothetical scenario that the change in mean GT (= mean SI) is an intrinsic feature of the pathogen, which is due to change in latent period and infectious period across cluster generations. Then, according to the classic 'susceptible-exposed-infectious-removed' (SEIR) framework, where exponential distributions are assumed for most of the epidemiological parameters (Gatto et al., 2020; Lipsitch et al., 2003; Svensson, 2007; Wu et al., 2020b; Zhao et al., 2020b), we have

$$X^{(k)} + Y^{(k)} = \mathbf{E}[\tau^{(k)}], \text{ and } X^{(k+1)} + Y^{(k+1)} = \mathbf{E}[\tau^{(k+1)}], \quad (5)$$

where X (unit: day) denotes the mean latent period, and Y (unit: day) denotes the mean infectious period. Note that Eqn (5) originally holds for the relationship among latent period, incubation period and GT (Svensson, 2007). It can be extended to the situation of SI with the assumption that the infector and infectee have the same distribution for the incubation period, such that GT and SI have the same expectation. The superscript (k) is the label of transmission generation rather than a power. This relationship is derived in (Svensson, 2007) theoretically, and adopted in (Champredon and Dushoff, 2015a; Gatto et al., 2020; Lipsitch et al., 2003; Wu et al., 2020b; Zhao et al., 2020b). When

$\lambda < 1$, we assume $0 \le X^{(k+1)} \le X^{(k)}$, and $0 \le Y^{(k+1)} \le Y^{(k)}$ for Eqn (5). We define $\rho^{(k)} = \frac{Y^{(k)} - Y^{(k+1)}}{\mathbf{E}[\tau^{(k)}] - \mathbf{E}[\tau^{(k+1)}]} \times 100\%$ as the percentage of SI reduction due to the reduction in infectious period.

We explore the potential effects of the change in SI on the individual reproduction number, R, across transmission generations. Referring to the SEIR framework, the individual reproduction number can be modelled as the product of the mean effective contact rate and the mean infectious period, i.e., $R^{(k)} = \beta^{(k)} \cdot Y^{(k)}$ for the infector in the k-th generation in a transmission chain. Here, β (unit: per day) denotes the effective contact rate.

By fixing β as a constant, we explore the effects of the change in τ on R in the k-th transmission generation. To set up, we fix the mean SI of the infector, $\mathbf{E}[\tau^{(k=0)}]$, at 7.5 days referring to the estimates from the earliest COVID-19 data (Li et al., 2020b), and the mean latent period, $X^{(k=0)}$, at 3.3 days (Li et al., 2020c; Zhao, 2020b; Zhao et al., 2021b) for the initial, i.e., 0-th, generation. Thus, the mean infectious period, $Y^{(k=0)}$, is derived at (7.5 − 3.3=) 4.2 days by using Eqn (5), which is in line with the results in literatures (Kucharski et al., 2020; Li et al., 2020c; Wu et al., 2020b). We further fix the initial individual reproduction number, $R^{(k=0)}$, at 2.2, which is generally consistent with previous estimates (Ali et al., 2020; Chinazzi et al., 2020; Gatto et al., 2020; He et al., 2020b; Jung et al., 2020; Li et al., 2020b; Musa et al., 2020; Ran et al., 2020; Riou and Althaus, 2020; Wu et al., 2020b; Xu et al., 2020; Zhao et al., 2020c; Zhao et al., 2020f), and also for the situation in Hong Kong (Cowling et al., 2020). Then, we fix $\beta = 2.2 / 4.2 \approx 0.5$ individual per day. Thus, the relationship among k, ρ, and R can be solved numerically.

2.5.2 Exploration #2: competition among multiple candidate infectors

In exploration #2, we consider a statistical mechanism that the shrinkage in SI may be an outcome of a competition among multiple candidate infectors, which was previously pointed out in (Kenah et al., 2008). The SI is recorded pairwisely as the duration between onset dates of an infectee and the infector who triggers the infection. In a cluster of cases, contacts are likely to occur in most pairs of infected and susceptible individuals simultaneously. Here, a cluster is defined as a group of cases who are seeded to the same (traceable) source of infection. The size of the cluster is determined by the number of cases within the same cluster. For example, 1 seed case without causing further infection would be a cluster of size 1, and 1 seed case transmits to 3 cases in the first generation, who further transmit to 5 cases in the second generation, would be a cluster of size (1 + 3 + 5 =) 9. The candidate infector is defined as those cases who contribute to the exposure of an infectee but may or may not trigger the infection eventually. We speculate the competitions among multiple candidate infectors may shorten the SI.

For the competition among a total of J candidate infectors for one infectee, the onset time, t, of the infectee who is triggered by the j-th candidate infector follows a PDF denoted by $g(t = t_j + \tau_j)$. Here, t_j denotes onset date of the j-th candidate infector, and τ_j denotes the candidate SI, if occurs, between the j-th candidate infector and the infectee. The parameter t_j is observable from the surveillance data, and thus is considered as a constant. The parameter τ_j is modelled as independent and identically distributed (IID) random variable following the PDF $h(\tau)$ as defined in Section 2.1. Hence, the $g(t)$ appears a shifted version of $h(\tau)$ with a shift term of $-t_j$. The candidate infector who triggers the infectee is recognized as the infector. Thus, the observed SI of the infectee is τ_j that associates with the smallest $(t_j + \tau_j)$ for all indexes j.

We simulate this candidate infector competition framework stochastically. To set up, we consider a cluster starting with one seed case whose onset date is day (or time) 0. The PDF $h(\tau)$ is modelled as a Gamma distribution with mean 5.5 and standard deviation (SD) 3.3 days, which is in line with many existing estimates (Ferretti et al., 2020; Ganyani et al., 2020; Tindale et al., 2020; Zhao, 2020b). With h, the PDF g can also be determined by shifting. For the reproducibility, we restrict the number of offsprings generated by each infector following a Poisson distribution with rate parameter fixed at 2.2, which is consistent to the predefined value of reproduction number (R) in Section 2.5.1. Alternatively, the Poisson distribution adopted here can be extended to a Negative Binomial distribution to further account for the overdispersion feature in the individual reproduction number, i.e., superspreading potential (Adam et al., 2020; Lloyd-Smith et al., 2005; Zhao et al., 2021a). The number of offsprings from the initial seed case, namely number of primary offsprings, is a criterion to identify the superspreading events, and thus is importance to explore its effect in shaping SI across transmission generation. For simplicity, we neglect the isolation in the simulation framework, such that the Poisson rate is fixed at 2.2. Namely, we investigate the theoretical outcomes under an intervention-free scenario. More realistic scenarios can be explored with time-varying values of reproduction number, which can be calculated by using the approach in previous studies empirically (Ma et al., 2014; Park et al., 2019; Wallinga and Lipsitch, 2007; Zhao et al., 2019).

In the model simulation, we record the cluster size in terms of the cumulative number of cases, onset dates of each case, infector of each infectee (except the initial seed case), SI, generation of cases, and number of offsprings for each infector. The generation of cases is traced by the transmission chain linked to the initial seed case, and we defined the generation of initial seed case as generation 0. For convenience, the transmission generation between a case in generation 0 and another case in generation 1 as the first transmission generation, and thus the index of transmission generation can be ranked subsequently.

For each simulation, we extract the SIs from first and second transmission generations, and treat these consecutive SIs as pairs of former and latter SIs that is illustrated in Fig 1. We generate 30 pairs of former and latter SIs, and conduct the estimation of λ using the framework in Eqn (2). We explore the effects of cluster size, number of primary offsprings and generation numbers in changing the scale of SI.

3 Results and discussion

For the 21 identified COVID-19 transmission chains in Hong Kong, the pairs of former and latter SIs are presented in Fig 2. We report the descriptive statistics as follows. For the former SI, we report a mean of 5.4 days, median of 6.0 days, interquartile range (IQR) between 3.0 and 7.0 days, 95% centile from 1.5 to 10.5 days, 95% percentile of 8.0 days, and a range from 1.0 to 13.0 days. For the latter SI, we report a mean of 4.8 days, median of 4.0 days, IQR between 3.0 and 7.0 days, 95% centile from 1.0 to 9.5 days, 95% percentile of 9.0 days, and a range from 1.0 to 10.0 days. We observe that the mean (and median) SI decreases when generation increases, and this finding was reported previously in (Li et al., 2020a; Ma et al., 2020). With the sample means, we calculate the ratio of latter SI over former SI at (4.8 / 5.4 =) 0.89, which is roughly the same scale as 0.73 in (Ma et al., 2020) and 0.94 or 0.75 in (Li et al., 2020a). Empirically, the pairwise difference of latter SI minus former SI has a mean of −0.7 days, median of 0.0 day, and IQR between −3.0 and 2.0 days.

The pairwise ratio of latter SI over former SI has a mean of 1.1, median of 1.0, and IQR between 0.5 and 1.5. By using the nonparametric bootstrapping approach, the crude change ratio of SI across generations is calculated at 1.00 with 95%CI: (0.57, 1.43).

Considering the theorical probability profile of individual SI in Eqn (2), we estimate the λ at 0.77 and 95%CI: (0.51, 1.16) selected with the lowest AICc among all non-truncated scenarios, see Table 1. For all scenarios in Table 1, we find that the Gamma distribution with $\sigma^2 = |\mu|$ outperformed against other scenarios in terms of the lowest AICc. As such, we estimate the λ at 0.72 and 95%CI: (0.54, 0.96), which are considered as the main results. Besides the fitting performance, we also consider the biological feasibility of probability profile in governing the real-world observations of the SI of COVID-19. Referring to the previous literatures (Adam et al., 2020; Ali et al., 2020; Du et al., 2020; Ganyani et al., 2020; Tindale et al., 2020; Xu et al., 2020; You et al., 2020; Zhao, 2020b), the SI of COVID-19 might be negative, i.e., $\tau < 0$. Although the Gamma distribution outperforms, the negative SI observations cannot be governed by a Gamma-distributed $h(\cdot)$. In this case, the scenario with the second lowest AICc is considered as another main results. As such, we estimate the λ at 0.74 and 95%CI: (0.61, 0.91) with a Gumbel distribution, which is also highlighted in Table 1. The best fitting performance from Gamma distribution is probably because all our SI observations appear positive, see Fig 2. We remark that with negative SI observations, Gumbel distribution is likely to yield a better fitting performance than Gamma distribution.

Consistently, the estimates of λ using Gamma and Gumbel distributions are almost the same, and significantly less than 1. Thus, the individual intrinsic SI is likely to shrink when the transmission generation increases with rate at 0.72 per generation. Remarkably, we distinguish effective SI and intrinsic SI. The intrinsic SI measure the SI when the effect of control measure is not in place, while the effective SI emphases the SI under the control measure. It is important to reveal the fundamental change of the parameter and associated external factors. According to our truncated likelihood framework in Eqn (3), the estimated shrinkage can be understood in regarding the intrinsic SI (Champredon and Dushoff, 2015b; Nishiura et al., 2020; Zhao et al., 2020f), of which the "*distribution depends only on the average infectiousness of an individual*" and the incubation period of secondary case as defined in (Champredon and Dushoff, 2015a). We have adjusted the impacts of case isolation in SI estimating, and thus the shrinkage in intrinsic SI is interpreted as an across-generation feature. The shrinkage in SI is also unlikely due to the effects of other types of nonpharmaceutical interventions (e.g., social distancing, facemask, sterilization, suspension of gathering, and city lockdown), which may shorten the realized SI as pointed out in (Ali et al., 2020; Nishiura, 2010; Park et al., 2020; Zhao et al., 2020d), because the decrease in effective transmission rate of each source case is unlikely to affect the mean intrinsic SI estimate. The effective SI is inferred in (Ali et al., 2020; Ma et al., 2020). Furthermore, the issue that right censoring due to sampling bias may lead to estimation bias in SI, which is pointed out in recent study (Park et al., 2021), may not affect our conclusion because our dataset covers a complete epidemic wave before May 2020 in Hong Kong. We also ignore the possible difference in the incubation periods of infector and infectee, which are considered following the same distribution in (Ganyani et al., 2020; Tindale et al., 2020). However, slight changes could occur due to several factors including case definitions, cohort assumptions, and changes in contact tracing strategies, which might impact the SI estimates and need further investigations.

Regardless of the number of direct offspring in each generation, the shrinkage in SI implies that the transmission is likely to occur more rapidly since the exposure of each infector. Then, the infectee is more likely exposed before the symptom onset of the infector in the late generations, comparing to the situation in the early generations. In other words, pre-symptomatic transmission may occur more frequently in the late generations. In addition, we find both main estimates of λ are under the scenario (**II**) of individual SI's SD (σ). Since the SIs from a population may have an SD as in scenario (**I**), this finding indicates that the SI of a population is more dispersive than the individual SI.

For the sensitivity analysis, the relationship in Eqn (4) is examined. We find that, consistent with the main results, the Gamma-distributed $h(\cdot)$ with scenario (**II**) of σ and likelihood truncation outperforms among other scenarios, see Fig 3. The SI lower bound, T_c, is estimated at 0.0 exactly, which means Eqn (4) becomes equivalent to Eqn (1), and implies relationship in Eqn (1) holds consistently.

We explore the impacts of the shrinkage in SI in shaping the individual reproduction number modelled as exploration #1 in Section 2.5.1. We find that the R decreases when the transmission generations increase, see Fig 4. With a higher percentage of the reduction in SI due to the reduction in infectious period (ρ), the R may decrease more rapidly. As a geometric sequence, if the absolute value of the common ratio, i.e., λ, is less than one, the sequence defined in Eqn (1) will converge. Thus, the mean GT will decrease and approach 0 theoretically, when the number of transmission generations becomes sufficiently large. As such, under exploration #1, a discrepancy between the theoretical outcome and the real-world fact occurs as follows.

> When the GT decreases, the individual R of each infector will also decrease, which leads to an outcome that the transmission of COVID-19 may vanish after several generations. In contradiction, as a matter of fact, the pandemic of COVID-19 continuous in many places (2021).

We note that this discrepancy may imply a restriction of exploration #1 in explaining the real-world observation. Thus, exploration #1 is less in favored comparing to exploration #2, which will be discussed next.

For the suspected candidate infector competition mechanism proposed as exploration #2 in Section 2.5.2, we find the shrinkage in SI is likely occur when the cluster size increases, see Fig 5, and when the number of offsprings increases, see Fig 6. Under the mechanism in exploration #2, the mean individual reproduction number holds as a constant. In other words, the outbreak maintains with substantial offspring cases in each transmission generation, and thus the discrepancy under exploration #1 vanishes. Therefore, we consider exploration #2 as the main discussion, which may be more reasonable than exploration #1. The mechanism of competition among multiple potential infectors is supported by previous study, and the findings of the shrinkage in the individual SI in this study provide a real-world evidence, which validates the theoretical framework in (Kenah et al., 2008). Furthermore, the scenario under exploration #2 would become evident only when there is sufficient number of seed cases serving as sources of infections, e.g., an outbreak is near to or has passed its peak, and within a cluster of cases and their close contacts. By contrast, under intensive nonpharmaceutical interventions, the infectors are typically isolated and close contact are quarantined timely, the competition among candidate infectors would be difficult to happen.

We observe the SI shrinks as generation increases, and approaches a boundary level in the late generations, see Fig 7. As such, we argue that the alternative relation in Eqn (4) may be more biologically reasonable, even though the simpler formulation in Eqn (1) slightly outperforms. We speculate the outperformance of Eqn (1) is possibly because most transmission chains (16 out of 21) are the chains of cases from 'zero-first-second' generations in each cluster of COVID-19 cases. This character of our COVID-19 dataset makes the simple geometric relation in Eqn (1) an optimal fit to the observations from early generations. In other words, if more SI observations from late generations would be included, Eqn (4) may replace Eqn (1) as the optimal relationship. To verify, we repeat the estimation in Section 2.4 by merely using the (21 − 16 =) 5 transmission chains that are from late generations, i.e., secondary, tertiary, or quaternary. In this case, we estimate the T_c at 1.4 days (data not shown), which indicates Eqn (4) appears more feasible than Eqn (1). Hence, The estimate can be benefit from a larger sample size. We remark that the data with more generations observed from each transmission chain will probably improve the estimation of the change in SI across generations. In addition, the simulation results in Fig 7 are considerably limited to several simplified (but unrealistic) modelling assumptions, and they include a 'perfectly mixed' system, i.e., everyone contacts all others constantly, the absence of nonpharmaceutical interventions, and SARS-CoV-2 transmits for infinitely many generations. Under these restrictions, the results in Fig 7 can merely be considered as an ideal scenario, which seldom occurs in the real-world setting.

We note the following limitation of our analysis. Although the effects of isolation time were adjusted in the inference framework in Eqn (3), the shrinkage of SI might also arise from other artificial factors, e.g., self-isolation, change in social activities or behaviors due to the awareness of virus circulation, and recalling bias such that short-term events are more precisely memorized, which occurred in the backward contact tracing exercise (Du et al., 2020). These potential confounders cannot be fully ruled out in the current analytical framework mainly due to lack of data. Therefore, although exploration #2 outperforms exploration #1, and is supported theoretically (Kenah et al., 2008), it cannot guarantee either true causality or a full causal effect even the causality is verified, i.e., a partial contribution.

The nonpharmaceutical interventive strategies (Fraser et al., 2004), which can cut off the transmission chain, e.g., case isolation, quarantine, social distancing, and personal protective equipment (PPE), are thus crucially important to mitigate the cluster size and flattening the epidemic curve. The statistical mechanism in exploration #2 may be applicable to study the transmission dynamics of other infectious diseases. Future studies on verifying the exploration #2, or on exploring other clinical or biological mechanisms that affects the individual SI across transmission generations are desired.

4 Conclusions

The mean of individual SI of COVID-19 is likely to shrink as the transmission generation increases with a threshold as the lower boundary. We speculate that the shrinkage in SI is an outcome of the competition among multiple candidate infectors within the same cluster of cases. The shrinkage in SI may speed up the transmission process, and thus the nonpharmaceutical interventive strategies are crucially important to mitigate the epidemic.

Declarations

Ethics approval, consent to participate, and consent for publication

All data used in this work are publicly available, and thus neither ethical approval nor consent is applicable.

Availability of materials

The COVID-19 surveillance data are collected via https://github.com/dcadam/covid-19-sse/blob/master/data/transmission_pairs.csv, which are originally released by the Centre for Health Protection (CHP) of Hong Kong (Centre for Health Protection, 2020), and previously used in (Adam et al., 2020).

Funding

DH was supported by General Research Fund (grant number: 15205119) of the Research Grants Council (RGC) of Hong Kong, China, and Alibaba (China) Co. Ltd. Collaborative Research project. YC and WW were supported by the National Natural Science Foundation of China (grant numbers 61672013, and 12071173), and the Huaian Key Laboratory for Infectious Diseases Control and Prevention (HAP201704).

Acknowledgements

We thank the insightful comments from three reviewers that help improve the manuscript.

Conflict of interests

DH received funding from Alibaba (China) Co. Ltd. Collaborative Research project. MHW is a shareholder of Beth Bioinformatics Co., Ltd. Other authors declared no competing interests. The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Authors' contributions

SZ conceived the study, carried out the analysis, and drafted the first manuscript. SZ and DH discussed the results. All authors critically read and revised the manuscript, and gave final approval for publication.

References

2021. World Health Organization, Coronavirus disease 2019 (COVID-19) situation reports. . Vol. 2021.

Adam, D. C., Wu, P., Wong, J. Y., Lau, E. H. Y., Tsang, T. K., Cauchemez, S., Leung, G. M., Cowling, B. J., 2020. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat Med 26, 1714-1719, doi:10.1038/s41591-020-1092-0.

Ali, S. T., Wang, L., Lau, E. H. Y., Xu, X. K., Du, Z., Wu, Y., Leung, G. M., Cowling, B. J., 2020. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science 369, 1106-1109, doi:10.1126/science.abc9004.

Assiri, A., McGeer, A., Perl, T. M., Price, C. S., Al Rabeeah, A. A., Cummings, D. A., Alabdullatif, Z. N., Assad, M., Almulhim, A., Makhdoom, H., Madani, H., Alhakeem, R., Al-Tawfiq, J. A., Cotten, M., Watson, S. J., Kellam, P., Zumla, A. I., Memish, Z. A., Team, K. M.-C. I., 2013. Hospital outbreak of Middle East respiratory syndrome coronavirus. N Engl J Med 369, 407-16, doi:10.1056/NEJMoa1306742.

Cai, Y., Zhao, S., Niu, Y., Peng, Z., Wang, K., He, D., Wang, W., 2021. Modelling the effects of the contaminated environments on tuberculosis in Jiangsu, China. J Theor Biol 508, 110453, doi:10.1016/j.jtbi.2020.110453.

Centre for Health Protection, 2020. Summary of data and outbreak situation of the Severe Respiratory Disease associated with a Novel Infectious Agent, Centre for Health Protection, the government of Hong Kong. . Vol. 2020.

Champredon, D., Dushoff, J., 2015a. Intrinsic and realized generation intervals in infectious-disease transmission. Proc Biol Sci 282, 20152026, doi:10.1098/rspb.2015.2026.

Champredon, D., Dushoff, J., 2015b. Intrinsic and realized generation intervals in infectious-disease transmission. Proceedings of the Royal Society B: Biological Sciences 282, 20152026.

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore, Y. P. A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M. E., Longini, I. M., Jr., Vespignani, A., 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 368, 395-400, doi:10.1126/science.aba9757.

Cowling, B. J., Fang, V. J., Riley, S., Malik Peiris, J. S., Leung, G. M., 2009. Estimation of the serial interval of influenza. Epidemiology 20, 344-7, doi:10.1097/EDE.0b013e31819d1092.

Cowling, B. J., Ali, S. T., Ng, T. W. Y., Tsang, T. K., Li, J. C. M., Fong, M. W., Liao, Q., Kwan, M. Y., Lee, S. L., Chiu, S. S., Wu, J. T., Wu, P., Leung, G. M., 2020. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 5, e279-e288, doi:10.1016/S2468-2667(20)30090-6.

Du, Z., Xu, X., Wu, Y., Wang, L., Cowling, B. J., Meyers, L. A., 2020. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. Emerg Infect Dis 26, 1341-1343, doi:10.3201/eid2606.200357.

Fan, J. Q., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11, 1031-1057, doi:DOI 10.3150/bj/1137421639.

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dorner, L., Parker, M., Bonsall, D., Fraser, C., 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 368, eabb6936, doi:10.1126/science.abb6936.

Fine, P. E., 2003. The interval between successive cases of an infectious disease. Am J Epidemiol 158, 1039-47, doi:10.1093/aje/kwg251.

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, 2999-3016.

Fraser, C., Riley, S., Anderson, R. M., Ferguson, N. M., 2004. Factors that make an infectious disease outbreak controllable. Proc Natl Acad Sci U S A 101, 6146-51, doi:10.1073/pnas.0307506101.

Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., Hens, N., 2020. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Euro Surveill 25, 2000257, doi:10.2807/1560-7917.ES.2020.25.17.2000257.

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., Rinaldo, A., 2020. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc Natl Acad Sci U S A 117, 10484-10491, doi:10.1073/pnas.2004978117.

He, D., Zhao, S., Lin, Q., Musa, S. S., Stone, L., 2020a. New estimates of the Zika virus epidemic attack rate in Northeastern Brazil from 2015 to 2016: A modelling analysis based on Guillain-Barre Syndrome (GBS) surveillance data. PLoS Negl Trop Dis 14, e0007502, doi:10.1371/journal.pntd.0007502.

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., 2020b. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med, 1-4.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497-506, doi:10.1016/S0140-6736(20)30183-5.

Jung, S. M., Akhmetzhanov, A. R., Hayashi, K., Linton, N. M., Yang, Y., Yuan, B., Kobayashi, T., Kinoshita, R., Nishiura, H., 2020. Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. J Clin Med 9, doi:10.3390/jcm9020523.

Kenah, E., Lipsitch, M., Robins, J. M., 2008. Generation interval contraction and epidemic data analysis. Math Biosci 213, 71-9, doi:10.1016/j.mbs.2008.02.007.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Centre for Mathematical Modelling of Infectious Diseases, C.-w. g., 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 20, 553-558, doi:10.1016/S1473-3099(20)30144-4.

Kutter, J. S., Spronken, M. I., Fraaij, P. L., Fouchier, R. A., Herfst, S., 2018. Transmission routes of respiratory viruses among humans. Curr Opin Virol 28, 142-151, doi:10.1016/j.coviro.2018.01.001.

Kwok, K. O., Wong, V. W. Y., Wei, W. I., Wong, S. Y. S., Tang, J. W., 2020. Epidemiological characteristics of the first 53 laboratory-confirmed cases of COVID-19 epidemic in Hong Kong, 13 February 2020. Euro Surveill 25, 2000155, doi:10.2807/1560-7917.ES.2020.25.16.2000155.

Leung, G. M., Hedley, A. J., Ho, L. M., Chau, P., Wong, I. O., Thach, T. Q., Ghani, A. C., Donnelly, C. A., Fraser, C., Riley, S., Ferguson, N. M., Anderson, R. M., Tsang, T., Leung, P. Y., Wong, V., Chan, J. C., Tsui, E., Lo, S. V., Lam, T. H., 2004. The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients. Ann Intern Med 141, 662-73, doi:10.7326/0003-4819-141-9-200411020-00006.

Leung, K., Wu, J. T., Liu, D., Leung, G. M., 2020. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. Lancet 395, 1382-1393, doi:10.1016/S0140-6736(20)30746-7.

Li, M., Liu, K., Song, Y., Wang, M., Wu, J., 2020a. Serial Interval and Generation Interval for Imported and Local Infectors, Respectively, Estimated Using Reported Contact-Tracing Data of COVID-19 in China. Front Public Health 8, 577431, doi:10.3389/fpubh.2020.577431.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., 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, T. T. Y., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., Feng, Z., 2020b. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 382, 1199-1207, doi:10.1056/NEJMoa2001316.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020c. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 368, 489-493, doi:10.1126/science.abb3221.

Lin, Q., Chiu, A. P., Zhao, S., He, D., 2018. Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia. Stat Methods Med Res 27, 1968-1978, doi:10.1177/0962280217746442.

Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., Fisman, D., Murray, M., 2003. Transmission dynamics and control of severe acute respiratory syndrome. Science 300, 1966-70, doi:10.1126/science.1086616.

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., Getz, W. M., 2005. Superspreading and the effect of individual variation on disease emergence. Nature 438, 355-9, doi:10.1038/nature04153.

538 Luo, L., Liu, D., Liao, X., Wu, X., Jing, Q., Zheng, J., Liu, F., Yang, S., Bi, H., Li, Z., Liu, J., Song, W., Zhu, W., Wang,
539 Z., Zhang, X., Huang, Q., Chen, P., Liu, H., Cheng, X., Cai, M., Yang, P., Yang, X., Han, Z., Tang, J., Ma, Y.,
540 Mao, C., 2020. Contact Settings and Risk for Transmission in 3410 Close Contacts of Patients With
541 COVID-19 in Guangzhou, China : A Prospective Cohort Study. Ann Intern Med 173, 879-887,
542 doi:10.7326/M20-2671.
543 Ma, J., Dushoff, J., Bolker, B. M., Earn, D. J. D., 2014. Estimating initial epidemic growth rates. Bull Math Biol
544 76, 245-260.
545 Ma, S., Zhang, J., Zeng, M., Yun, Q., Guo, W., Zheng, Y., Zhao, S., Wang, M. H., Yang, Z., 2020. Epidemiological
546 Parameters of COVID-19: Case Series Study. J Med Internet Res 22, e19994, doi:10.2196/19994.
547 Milwid, R., Steriu, A., Arino, J., Heffernan, J., Hyder, A., Schanzer, D., Gardner, E., Haworth-Brockman, M.,
548 Isfeld-Kiely, H., Langley, J. M., Moghadas, S. M., 2016. Toward Standardizing a Lexicon of Infectious
549 Disease Modeling Terms. Front Public Health 4, 213, doi:10.3389/fpubh.2016.00213.
550 Musa, S. S., Zhao, S., Wang, M. H., Habib, A. G., Mustapha, U. T., He, D., 2020. Estimation of exponential
551 growth rate and basic reproduction number of the coronavirus disease 2019 (COVID-19) in Africa.
552 Infectious Diseases of Poverty 9, 96, doi:10.1186/s40249-020-00718-y.
553 Nishiura, H., 2010. Time variations in the generation time of an infectious disease: implications for sampling
554 to appropriately quantify transmission potential. Math Biosci Eng 7, 851-69,
555 doi:10.3934/mbe.2010.7.851.
556 Nishiura, H., Linton, N. M., Akhmetzhanov, A. R., 2020. Serial interval of novel coronavirus (COVID-19)
557 infections. Int J Infect Dis 93, 284-286, doi:10.1016/j.ijid.2020.02.060.
558 Park, S. W., Champredon, D., Dushoff, J., 2020. Inferring generation-interval distributions from contact-
559 tracing data. J R Soc Interface 17, 20190719, doi:10.1098/rsif.2019.0719.
560 Park, S. W., Champredon, D., Weitz, J. S., Dushoff, J., 2019. A practical generation-interval-based approach to
561 inferring the strength of epidemics from their speed. Epidemics 27, 12-18,
562 doi:https://doi.org/10.1016/j.epidem.2018.12.002.
563 Park, S. W., Sun, K., Champredon, D., Li, M., Bolker, B. M., Earn, D. J. D., Weitz, J. S., Grenfell, B. T., Dushoff, J.,
564 2021. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. Proc
565 Natl Acad Sci U S A 118, e2011548118, doi:10.1073/pnas.2011548118.
566 Parry, J., 2020. China coronavirus: cases surge as official admits human to human transmission. BMJ 368,
567 m236, doi:10.1136/bmj.m236.
568 Ran, J., Zhao, S., Han, L., Liao, G., Wang, K., Wang, M. H., He, D., 2020. A re-analysis in exploring the
569 association between temperature and COVID-19 transmissibility: an ecological study with 154
570 Chinese cities. Eur Respir J 56, 2001253, doi:10.1183/13993003.01253-2020.
571 Ren, X., Li, Y., Yang, X., Li, Z., Cui, J., Zhu, A., Zhao, H., Yu, J., Nie, T., Ren, M., Dong, S., Cheng, Y., Chen, Q.,
572 Chang, Z., Sun, J., Wang, L., Feng, L., Gao, G. F., Feng, Z., Li, Z., 2021. Evidence for pre-symptomatic
573 transmission of coronavirus disease 2019 (COVID-19) in China. Influenza Other Respir Viruses 15, 19-
574 26, doi:10.1111/irv.12787.
575 Riou, J., Althaus, C. L., 2020. Pattern of early human-to-human transmission of Wuhan 2019 novel
576 coronavirus (2019-nCoV), December 2019 to January 2020. Euro Surveill 25, 2000058,
577 doi:10.2807/1560-7917.ES.2020.25.4.2000058.
578 Svensson, A., 2007. A note on generation times in epidemic models. Math Biosci 208, 300-11,
579 doi:10.1016/j.mbs.2006.10.010.
580 Tindale, L. C., Stockdale, J. E., Coombe, M., Garlock, E. S., Lau, W. Y. V., Saraswat, M., Zhang, L., Chen, D.,
581 Wallinga, J., Colijn, C., 2020. Evidence for transmission of COVID-19 prior to symptom onset. Elife 9,
582 e57149, doi:10.7554/eLife.57149.
583 Tuite, A. R., Fisman, D. N., 2020. Reporting, Epidemic Growth, and Reproduction Numbers for the 2019 Novel
584 Coronavirus (2019-nCoV) Epidemic. Annals of Internal Medicine 172, 567-568, doi:10.7326/M20-
585 0358.
586 Vink, M. A., Bootsma, M. C., Wallinga, J., 2014. Serial intervals of respiratory infectious diseases: a systematic
587 review and analysis. Am J Epidemiol 180, 865-75, doi:10.1093/aje/kwu209.

588 Wallinga, J., Teunis, P., 2004. Different epidemic curves for severe acute respiratory syndrome reveal similar
589 impacts of control measures. Am J Epidemiol 160, 509-16, doi:10.1093/aje/kwh255.
590 Wallinga, J., Lipsitch, M., 2007. How generation intervals shape the relationship between growth rates and
591 reproductive numbers. Proc Biol Sci 274, 599-604, doi:10.1098/rspb.2006.3754.
592 Wang, K., Zhao, S., Liao, Y., Zhao, T., Wang, X., Zhang, X., Jiao, H., Li, H., Yin, Y., Wang, M. H., Xiao, L., Wang, L.,
593 He, D., 2020. Estimating the serial interval of the novel coronavirus disease (COVID-19) based on the
594 public surveillance data in Shenzhen, China, from 19 January to 22 February 2020. Transbound
595 Emerg Dis 67, 2818-2822, doi:10.1111/tbed.13647.
596 White, L. F., Wallinga, J., Finelli, L., Reed, C., Riley, S., Lipsitch, M., Pagano, M., 2009. Estimation of the
597 reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic
598 in the USA. Influenza Other Respir Viruses 3, 267-76, doi:10.1111/j.1750-2659.2009.00106.x.
599 World Health Organization, 2020. Statement on the second meeting of the International Health Regulations
600 Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV), World Health
601 Organization (WHO). Vol. 2020.
602 Wu, J., Huang, Y., Tu, C., Bi, C., Chen, Z., Luo, L., Huang, M., Chen, M., Tan, C., Wang, Z., Wang, K., Liang, Y.,
603 Huang, J., Zheng, X., Liu, J., 2020a. Household Transmission of SARS-CoV-2, Zhuhai, China, 2020. Clin
604 Infect Dis 71, 2099-2108, doi:10.1093/cid/ciaa557.
605 Wu, J. T., Leung, K., Leung, G. M., 2020b. Nowcasting and forecasting the potential domestic and
606 international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study.
607 Lancet 395, 689-697, doi:10.1016/S0140-6736(20)30260-9.
608 Xu, X. K., Liu, X. F., Wu, Y., Ali, S. T., Du, Z., Bosetti, P., Lau, E. H. Y., Cowling, B. J., Wang, L., 2020.
609 Reconstruction of Transmission Pairs for Novel Coronavirus Disease 2019 (COVID-19) in Mainland
610 China: Estimation of Superspreading Events, Serial Interval, and Hazard of Infection. Clin Infect Dis 71,
611 3163-3167, doi:10.1093/cid/ciaa790.
612 Yan, P., 2008. Separate roles of the latent and infectious periods in shaping the relation between the basic
613 reproduction number and the intrinsic growth rate of infectious disease outbreaks. J Theor Biol 251,
614 238-52, doi:10.1016/j.jtbi.2007.11.027.
615 Yang, L., Dai, J., Zhao, J., Wang, Y., Deng, P., Wang, J., 2020. Estimation of incubation period and serial
616 interval of COVID-19: analysis of 178 cases and 131 transmission chains in Hubei province, China.
617 Epidemiol Infect 148, e117, doi:10.1017/S0950268820001338.
618 You, C., Deng, Y., Hu, W., Sun, J., Lin, Q., Zhou, F., Pang, C. H., Zhang, Y., Chen, Z., Zhou, X. H., 2020.
619 Estimation of the time-varying reproduction number of COVID-19 outbreak in China. Int J Hyg
620 Environ Health 228, 113555, doi:10.1016/j.ijheh.2020.113555.
621 Zhao, S., 2020a. To avoid the noncausal association between environmental factor and COVID-19 when using
622 aggregated data: Simulation-based counterexamples for demonstration. Sci Total Environ 748,
623 141590, doi:10.1016/j.scitotenv.2020.141590.
624 Zhao, S., 2020b. Estimating the time interval between transmission generations when negative values occur
625 in the serial interval data: using COVID-19 as an example. Math Biosci Eng 17, 3512-3519,
626 doi:10.3934/mbe.2020198.
627 Zhao, S., Musa, S. S., Fu, H., He, D., Qin, J., 2019. Simple framework for real-time forecast in a data-limited
628 situation: the Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example. Parasit Vectors
629 12, 344, doi:10.1186/s13071-019-3602-9.
630 Zhao, S., Musa, S. S., Meng, J., Qin, J., He, D., 2020a. The long-term changing dynamics of dengue infectivity
631 in Guangdong, China, from 2008–2018: a modelling analysis. Trans R Soc Trop Med Hyg 114, 62-71.
632 Zhao, S., Stone, L., Gao, D., Musa, S. S., Chong, M. K. C., He, D., Wang, M. H., 2020b. Imitation dynamics in
633 the mitigation of the novel coronavirus disease (COVID-19) outbreak in Wuhan, China from 2019 to
634 2020. Annals of Translational Medicine 8, 448, doi:10.21037/atm.2020.03.168.
635 Zhao, S., Shen, M., Musa, S. S., Guo, Z., Ran, J., Peng, Z., Zhao, Y., Chong, M. K. C., He, D., Wang, M. H., 2021a.
636 Inferencing superspreading potential using zero-truncated negative binomial model: exemplification
637 with COVID-19. BMC Med Res Methodol 21, 1-8.

638 Zhao, S., Musa, S. S., Lin, Q., Ran, J., Yang, G., Wang, W., Lou, Y., Yang, L., Gao, D., He, D., Wang, M. H., 2020c.
639 Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half
640 of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak. Journal of Clinical Medicine
641 9, 388, doi:10.3390/jcm9020388.
642 Zhao, S., Cao, P., Chong, M. K. C., Gao, D., Lou, Y., Ran, J., Wang, K., Wang, W., Yang, L., He, D., Wang, M. H.,
643 2020d. COVID-19 and gender-specific difference: Analysis of public surveillance data in Hong Kong
644 and Shenzhen, China, from January 10 to February 15, 2020. Infect Control Hosp Epidemiol 41, 750-
645 751, doi:10.1017/ice.2020.64.
646 Zhao, S., Cao, P., Gao, D., Zhuang, Z., Cai, Y., Ran, J., Chong, M. K. C., Wang, K., Lou, Y., Wang, W., Yang, L., He,
647 D., Wang, M. H., 2020e. Serial interval in determining the estimation of reproduction number of the
648 novel coronavirus disease (COVID-19) during the early outbreak. J Travel Med 27, taaa033,
649 doi:10.1093/jtm/taaa033.
650 Zhao, S., Gao, D. Z., Zhuang, Z., Chong, M. K. C., Cai, Y. L., Ran, J. J., Cao, P. H., Wang, K., Lou, Y. J., Wang, W.
651 M., Yang, L., He, D. H., Wang, M. H., 2020f. Estimating the Serial Interval of the Novel Coronavirus
652 Disease (COVID-19): A Statistical Analysis Using the Public Data in Hong Kong From January 16 to
653 February 15, 2020. Frontiers in Physics 8, 347, doi:ARTN 347

654 10.3389/fphy.2020.00347.
655 Zhao, S., Tang, B., Musa, S. S., Ma, S., Zhang, J., Zeng, M., Yun, Q., Guo, W., Zheng, Y., Yang, Z., Peng, Z.,
656 Chong, M. K. C., Javanbakht, M., He, D., Wang, M. H., 2021b. Estimating the generation interval and
657 inferring the latent period of COVID-19 from the contact tracing data. Epidemics 36, 100482,
658 doi:https://doi.org/10.1016/j.epidem.2021.100482.

659

660

Figures

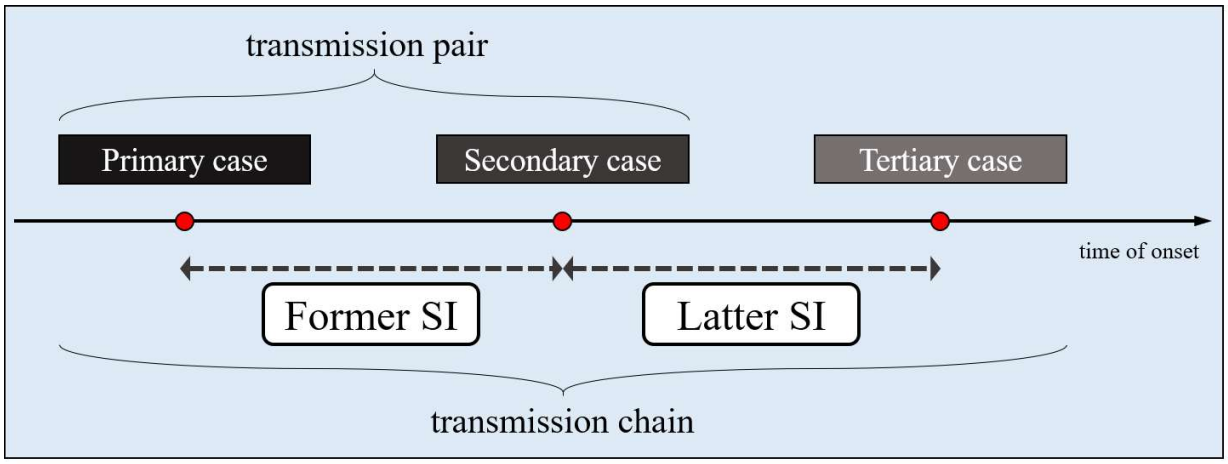

Figure 1.

The illustration diagram of the timeline of a typical transmission chain. The former SI is denoted by $\tau^{(F)}$, and the latter SI and denoted by $\tau^{(L)}$.

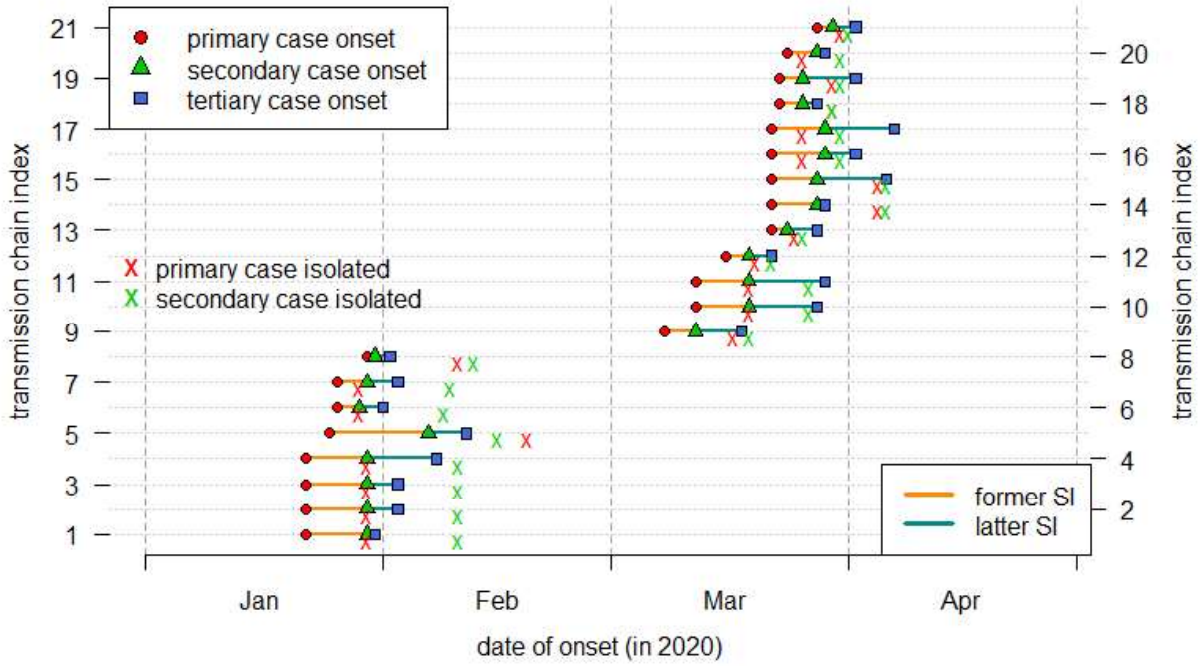

Figure 2.

The timeline of the transmission chains included in this study. The dot indicates the symptoms onset date of each case. The horizontal solid line represents the duration of each serial interval (SI). The transmission chains are indexed in the sequence of the onset dates of primary, secondary, and tertiary cases, which is merely for visualization purposed and will not affect the analytical procedures.

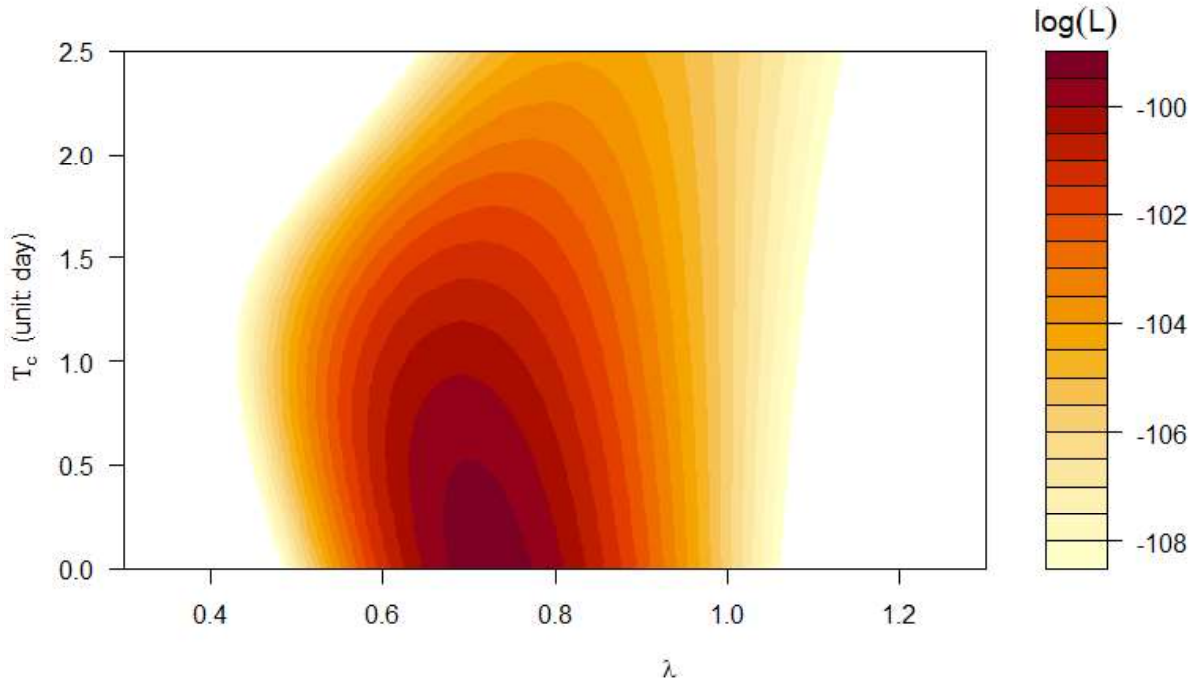

Figure 3.

The Gamma-distributed log-likelihood profile of the lower abound of the serial interval T_c (unit: day) and the scale of change in serial interval across generations λ, in Eqn (4), under scenario (**II**), which has the best fitting performance in terms of the (corrected Akaike information criterion) AICc = 202.8. The color scheme of the log-likelihood values is shown in the right column.

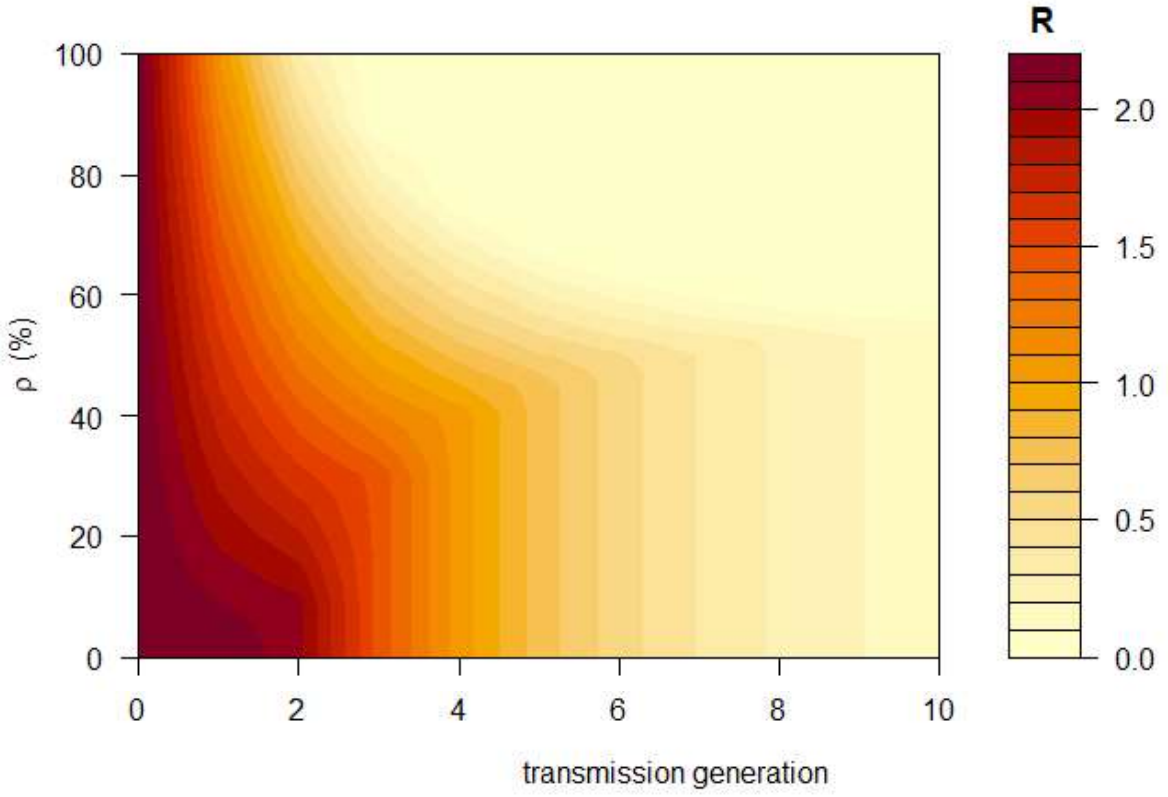

Figure 4.

The changing patterns of individual reproduction number (R) across the increasing transmission generations and the percentage of the reduction in serial interval (SI) due to the reduction in infectious period (ρ), see Section 2.5.1. For the initial (i.e., 0-th) transmission generation, the SI for the initial generation is fixed at 7.5 days, the latent period is fixed at 3.3 days, and the individual basic reproduction number (R_0) is fixed at 2.2.

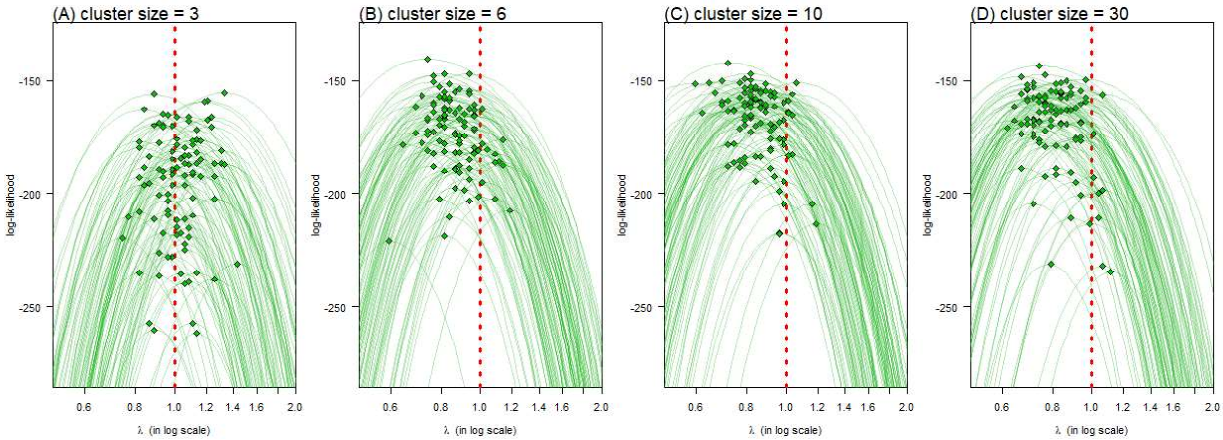

Figure 5.

The likelihood profiles of the scale of change in serial interval across generations (λ) when the cluster size is 3 (A), 6 (B), 10 (C), or 30 (D). In each panel, the green curves are the likelihood profiles of 100 set of samples (sample size of 30 for each set), and the green dots are the maximum likelihood estimates of λ. This simulation results are under the setting of exploration #2, see Section 2.5.2.

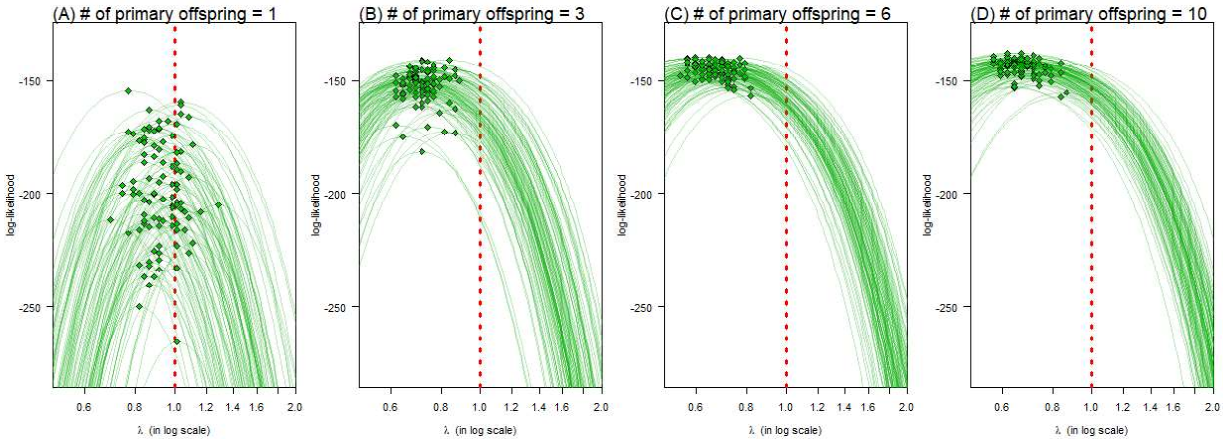

Figure 6.

The likelihood profiles of the scale of change in serial interval across generations (λ) when the number of primary offsprings is 1 (A), 3 (B), 6 (C), or 10 (D). In each panel, the green curves are the likelihood profiles of 100 set of samples (sample size of 30 for each set), and the green dots are the maximum likelihood estimates of λ. This simulation results are under the setting of exploration #2, see Section 2.5.2.

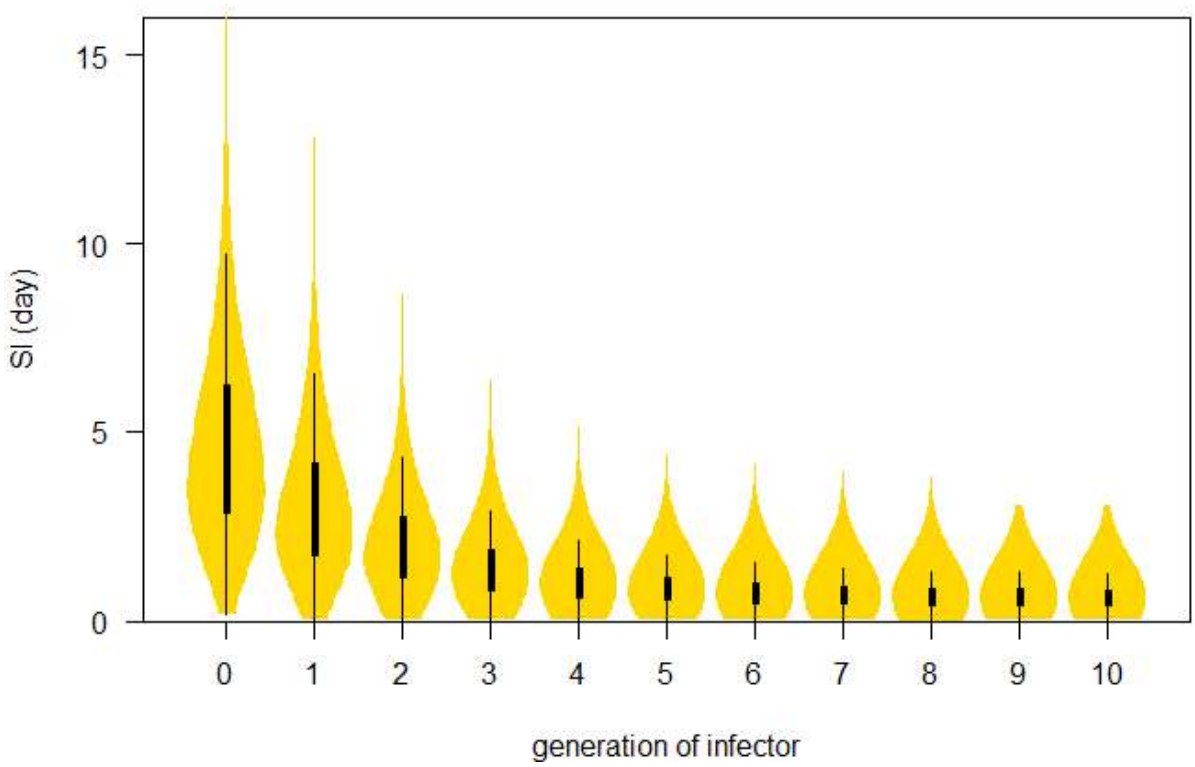

Figure 7.

The distribution of serial interval (SI) of infector in each (cluster) generations. The gold area indicates the distribution, the bold bars are the interquartile ranges (IQR), and the thin bars are the 95% centiles.

718 **Table**

719 Table 1.

720 Summary of the scale of change in serial interval across generations (λ) estimates (unit: per
721 transmission generation). The shaded estimates are considered as the main results.

SD of SI (σ)	Truncation	Distribution	scale of change in SI (λ)	AICc
scenario (**I**): large, i.e., SD = mean	No	Normal	0.66 (0.53, 0.82)	259.8
		Gumbel	0.78 (0.55, 1.11)	242.3
		Gamma	0.77 (0.51, 1.16)	237.7
	Yes	Normal	0.65 (0.53, 0.82)	224.7
		Gumbel	0.76 (0.52, 1.11)	212.5
		Gamma	0.72 (0.45, 1.16)	212.1
scenario (**II**): moderate, i.e., SD^2 = mean	No	Normal	0.79 (0.66, 0.95)	275.2
		Gumbel	0.86 (0.74, 0.99)	252.2
		Gamma	0.87 (0.72, 1.05)	242.5
	Yes	Normal	0.69 (0.55, 0.87)	228.3
		Gumbel	0.74 (0.61, 0.91)	206.9
		Gamma	0.72 (0.54, 0.96)	200.6
scenario (**III**): small, i.e., SD = 1	No	Normal	0.92 (0.86, 1.00)	552.6
		Gumbel	0.82 (0.81, 0.83)	6825.6
		Gamma	0.88 (0.83, 0.92)	634.0
	Yes	Normal	0.82 (0.73, 0.92)	452.2
		Gumbel	0.81 (0.80, 0.82)	5357.1
		Gamma	0.78 (0.73, 0.85)	494.8
crude estimate (bootstrapping without truncation)			1.00 (0.57, 1.43)	none

722

723

724