



Estimating the generation interval and inferring the latent period of COVID-19 from the contact tracing data

Shi Zhao^{a,b,**}, Biao Tang^{c,d}, Salihu S Musa^{e,f}, Shujuan Ma^g, Jiayue Zhang^g, Minyan Zeng^h, Qingping Yunⁱ, Wei Guo^h, Yixiang Zheng^j, Zuyao Yang^a, Zhihang Peng^k, Marc KC Chong^{a,b}, Mohammad Javanbakht^l, Daihai He^{e,*}, Maggie H. Wang^{a,b}

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

^b CUHK Shenzhen Research Institute, Shenzhen, China

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

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

^e Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China

^f Department of Mathematics, Kano University of Science and Technology, Wudil, Nigeria

^g Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

^h Department of Neurology, Peking University Shenzhen Hospital, Shenzhen, China

ⁱ Department of Social Medicine and Health Education, School of Public Health, Peking University, Beijing, China

^j Department of Infectious Diseases, Key Laboratory of Viral Hepatitis of Hunan, Xiangya Hospital, Central South University, Changsha, China

^k Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China

^l Nephrology and Urology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Keywords:

COVID-19

Generation interval

Latent period

Serial interval

Incubation period

Contact tracing

Statistical inference

ABSTRACT

The coronavirus disease 2019 (COVID-19) emerged by end of 2019, and became a serious public health threat globally in less than half a year. The generation interval and latent period, though both are of importance in understanding the features of COVID-19 transmission, are difficult to observe, and thus they can rarely be learnt from surveillance data empirically. In this study, we develop a likelihood framework to estimate the generation interval and incubation period simultaneously by using the contact tracing data of COVID-19 cases, and infer the pre-symptomatic transmission proportion and latent period thereafter. We estimate the mean of incubation period at 6.8 days (95 %CI: 6.2, 7.5) and SD at 4.1 days (95 %CI: 3.7, 4.8), and the mean of generation interval at 6.7 days (95 %CI: 5.4, 7.6) and SD at 1.8 days (95 %CI: 0.3, 3.8). The basic reproduction number is estimated ranging from 1.9 to 3.6, and there are 49.8 % (95 %CI: 33.3, 71.5) of the secondary COVID-19 infections likely due to pre-symptomatic transmission. Using the best estimates of model parameters, we further infer the mean latent period at 3.3 days (95 %CI: 0.2, 7.9). Our findings highlight the importance of both isolation for symptomatic cases, and for the pre-symptomatic and asymptomatic cases.

1. Introduction

The dynamics of the transmission of an infectious disease is largely determined by the pathogen's infectiousness and the time interval between the transmission generations of cases (Tuite and Fisman, 2020; Zhao et al., 2020a; Riou and Althaus, 2020).

Serial interval (SI) is defined as the time interval between the onset of symptoms of an infector and that of the associated infectee (Fine, 2003; White et al., 2009; Milwid et al., 2016; Vink et al., 2014), which is commonly observable. By contrast, generation interval (GI) is defined as the time interval between the time of infection (or time of being exposed) in a primary case (i.e., infector) and that in the associated

* Corresponding author.

** Corresponding author at: JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China.

E-mail addresses: zhaoshi.cmsa@gmail.com (S. Zhao), btang66@yorku.ca (B. Tang), salihu-sabiu.musa@connect.polyu.hk (S.S. Musa), 790039940@qq.com (S. Ma), zhangjiayue@csu.edu.cn (J. Zhang), zmy6691@gmail.com (M. Zeng), yunqingping@bjmu.edu.cn (Q. Yun), 176912045@csu.edu.cn (W. Guo), yxzheng@csu.edu.cn (Y. Zheng), zyang@cuhk.edu.hk (Z. Yang), zhihangpeng@njmu.edu.cn (Z. Peng), marc@cuhk.edu.hk (M.K. Chong), mhmjvbt81@gmail.com (M. Javanbakht), daihai.he@polyu.edu.hk (D. He), maggiew@cuhk.edu.hk (M.H. Wang).

<https://doi.org/10.1016/j.epidem.2021.100482>

Received 29 August 2020; Received in revised form 13 June 2021; Accepted 16 June 2021

Available online 24 June 2021

1755-4365/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

secondary case (i.e., infectee) (Milwid et al., 2016; Wallinga and Lipsitch, 2007). Since it is generally difficult to determine the time points of infection in two consecutive cases, i.e., in a pairwise transmission chain, the GI is rarely observable. Fig. 1 illustrates the timeline in a pairwise transmission chain, and the difference between GI and SI in a transmission chain. It is worth noting that the SI could be negative when the incubation period has a wide range, which has been reported in the COVID-19 pandemic recently (Du et al., 2020; He et al., 2020; Zhao, 2020; Ali et al., 2020; Xu et al., 2020; You et al., 2020), but GI is strictly positive by its definition.

Incubation period is defined as the time interval between the infection and onset of symptoms of a typical case (Yan, 2008). Although the time of infection for an infector is difficult to observe, an infectee's time of infection can be determined by the contact tracing history of a 'infector-infectee' pair. Here, the 'infector-infectee' pair indicates there is one and only one infector with epidemiological link with the infectee. Hence, the incubation periods of infectees may be traceable. By contrast, the latent period is defined as the time interval between the infection and onset of infectiousness of one case (Milwid et al., 2016). Since the onset of infectiousness is unobservable, the latent period is untraceable. In most of the infectious diseases, the mean latent period is less than or equal to the mean incubation period, which is shown in Fig. 1. In infectious disease modelling, the latent period is of interest and used in the differential equations based compartmental modelling frameworks (Svensson, 2007).

Incubation period is of clinical importance for individual medication and treatment, whereas latent period is important for disease transmission at population scale. Pre-symptomatic transmission may occur if latent period is sufficiently shorter than the incubation period, or in the other words, the onset of infectiousness is prior to the onset of symptoms (Tindale et al., 2020; Kong et al., 2020). When the onset of infectiousness is coincident with the onset of symptoms, i.e., latent and incubation periods match, pre-symptomatic transmission will not occur. In most situations, latent period may be shorter than incubation period, and thus, pre-symptomatic transmission may occur. Occasionally, when the onset of symptoms in the infectee occurs earlier than that in the infector, the observed SI will be negative. In the real-world situation, the data of SI are commonly used as a proxy to approximate the true patterns of GI (Cori et al., 2013). This simple adaptation fails if there are negative SI observations, and leads to biased estimates of reproduction number. This approach may be biased or inefficient if there are negative SI observations (Wallinga and Lipsitch, 2007; Zhao, 2020; Yan, 2008). In the following, we formulate an analytical approach to tackle this issue and implemented on the COVID-19 dataset.

COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was firstly reported in late 2019 (Huang et al., 2020; Li et al., 2020a; Zhao et al., 2020b), and spread to over 200 countries globally in a short period of time (Wu et al., 2020; Zhao et al., 2020c). In response to the COVID-19 pandemic, the World Health Organization (WHO) declared a public health emergency of international concern on January 30, 2020 (ZZZZZ, 2021a). As of January 2021, there are over 100 million COVID-19 confirmed cases globally including over 2 million associated deaths (ZZZZZ, 2021b). In previous studies (Du et al., 2020; He et al., 2020; Zhao, 2020; Xu et al., 2020; You et al., 2020), negative SIs are observed in the transmission events of COVID-19 (Ali et al., 2020; Adam et al., 2020), and considered as an evidence for the pre-symptomatic transmission as well as a short latent period thus difficult to control (Tindale et al., 2020; Kong et al., 2020; Liu et al., 2020a). Hence, estimating the GI and inferring the latent period is essential for understanding the features of the COVID-19 transmission.

In this study, we develop a likelihood-based framework to estimate the incubation period and GI simultaneously by using the pairwise contact tracing data of COVID-19 as an example. We infer the risk of pre-symptomatic transmission and the latent period based on the best estimates of incubation period and GI.

2. Data and methods

2.1. Data

We use the secondary COVID-19 contact tracing data retrieved from official news releases of government health authorities and public surveillance platforms through April 8, 2020. The dataset was previously collected in (Ma et al., 2020) with systematic and strict 'inclusion-and-exclusion' screening criteria based on plausible epidemiological evidences, and rigorous consistency checking by several researchers independently under the supervision of a senior author. Please see Ma et al. for the detailed introduction of this dataset (Ma et al., 2020).

From this dataset, we extract the information that contains the contact tracing history of 254 'infector-infectee' pairs recorded in exact date (instead of ranges of date), where there is one and only one infector with epidemiological link with each infectee. For each 'infector-infectee' pair, the detailed information includes the time of infection of infectee, and the onset of symptoms in both infector and infectee. Thus, we extract the incubation period of each infectee, and the SI of each transmission pair. Note that the time of infection is equivalent to the time of infection.

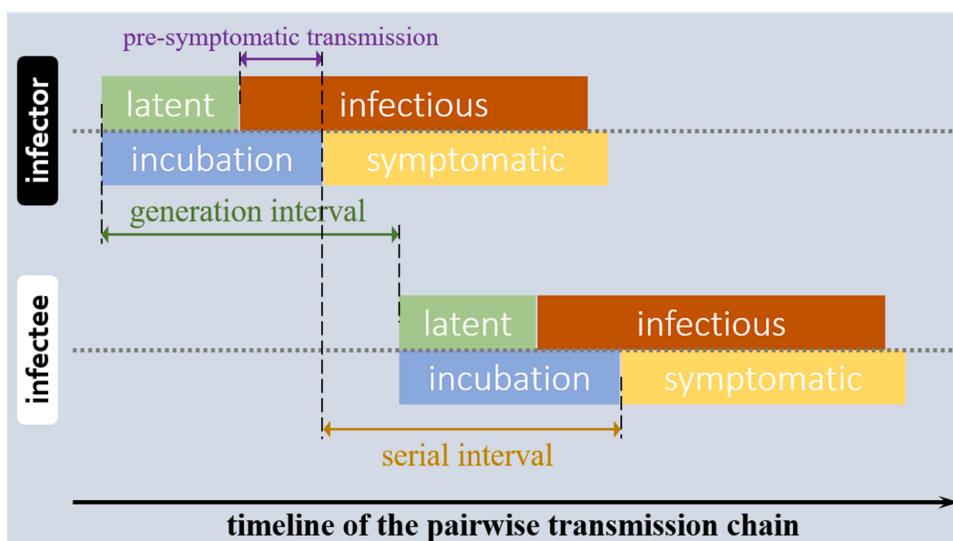


Fig. 1. The demonstrative timeline of COVID-19's transmission chain. The latent periods (ζ , in green), incubation periods (ξ , in blue), infectious periods (ν , in red), and symptomatic period (in orange) of both infector and infectee are illustrated in the timeline. The SI (s , in orange), GI (g , in green), and pre-symptomatic transmission (η) period (in purple) are also marked on the timeline of pairwise transmission chain. For COVID-19, we expect the latent period to be shorter than the infectious period, which is illustrated here. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Estimation of incubation period and generation interval

We denote the incubation period as ξ , and thus $\xi \geq 0$, see Fig. 1. Let ξ_1 and ξ_2 denote the incubation periods for a pair of infector and infectee in consecutive transmission generations, respectively. The ξ_1 and ξ_2 are considered as independent and identically distributed (IID) random variables determined by a probability density function (PDF) denoted by $h(\cdot)$. Following previous studies (Li et al., 2020a; Backer et al., 2020; Lauer et al., 2020), we consider $h(\cdot)$ as a Gamma distribution, which is a model assumption, with mean μ_ξ and standard deviation (SD) σ_ξ . Thus, the associated log likelihood profile to estimate the parameters in $h(\cdot)$, denoted by $l_\xi(\cdot)$, is defined as in Eq. (1).

$$l_\xi(\mu_\xi, \sigma_\xi | y_1, \dots, y_n) = \sum_i \log h(y_i | \mu_\xi, \sigma_\xi) \quad (1)$$

where y_i is the observed incubation period of the i -th infectee.

The GI, denoted by g , is determined by a PDF $\delta(\cdot)$ with mean μ_g and SD σ_g . Hence, the SI, denoted by s , is defined by $s = g + \xi_2 - \xi_1$, and straightforwardly, the expectations of g and s are equal, i.e., $E[g] = E[s]$. The distribution of s is determined by a PDF denoted as $f(\cdot)$, and by convolution, the $f(s)$ can be formulated in Eq. (2).

$$f(s) = \int \delta(s - y) \cdot \left[\int h(z) \cdot h(y + z) dz \right] dy \quad (2)$$

Following previous studies (Du et al., 2020; Cowling et al., 2009; Kwok et al., 2020; Nishiura et al., 2020a), we assume $\delta(\cdot)$ following a Gamma distribution. Thus, the distribution of $f(\cdot)$ can be obtained empirically by large-sample-size Monte Carlo. Then, the log likelihood profile of SI, $l_s(\cdot)$, can be defined as in Eq. (3).

$$l_s(\mu_\xi, \sigma_\xi, \mu_g, \sigma_g | x_1, \dots, x_n) = \sum_j \log f(x_j | \mu_\xi, \sigma_\xi, \mu_g, \sigma_g) \quad (3)$$

where the x_j s are the observations of SI.

With Eqs. (1) and (3), the overall log likelihood function of $\mu_\xi, \sigma_\xi, \mu_g$ and σ_g is summated as in Eq. (4).

$$l(\mu_\xi, \sigma_\xi, \mu_g, \sigma_g | y_1, \dots, y_n; x_1, \dots, x_n) = l_\xi(\mu_\xi, \sigma_\xi | y_1, \dots, y_n) + l_s(\mu_\xi, \sigma_\xi, \mu_g, \sigma_g | x_1, \dots, x_n) \quad (4)$$

As such, by fitting to the observations of incubation period and SI in Ma et al. (2020), the means and SDs of ξ and g , i.e., $\mu_\xi, \sigma_\xi, \mu_g$ and σ_g , can be estimated by using the maximum likelihood estimation approach. The 95 % confidence intervals (95 %CI) are calculated by using the profile likelihood estimation framework with a cutoff threshold determined using a Chi-square quantile (Fan and Huang, 2005; Bolker, 2008), which is also adopted in (Tariq et al., 2020; He et al., 2010; Zhao et al., 2018; Lin et al., 2018; Zhao et al., 2021; Ran et al., 2020).

2.3. Estimation of pre-symptomatic transmission

We are interested in the proportion of secondary infections due to pre-symptomatic transmission, denoted by η . By definition, η is the probability that the time of infection of an infectee is prior to the time when the associated infector symptoms onset occurs, which is given in Eq. (5).

$$\eta = \Pr(o^{(0)} > e^{(1)}) \approx \frac{\sum_{i=1}^N \mathbf{I}(o_i^{(0)} > e_i^{(1)})}{N} \times 100\% \quad (5)$$

Here, o and e denote the time (or date) of symptoms onset and infection, respectively. The superscript '(k)' indicates infector (for $k = 0$) or infectee (for $k = 1$) in one transmission pair. The subscript i denotes the i -th transmission pair in the total of $N = 254$ transmission pairs. The function $\mathbf{I}(\cdot)$ is the indicator function that equals 1 if the condition is true or 0 otherwise. Using Eq. (5), the η can be calculated empirically.

With simple transformations, the η can be estimated by the probability of $s < \xi$ as formulated in Eq. (6).

$$\eta = \Pr((o^{(1)} - o^{(0)}) < (e^{(1)} - e^{(0)})) = \Pr(s < \xi) \times 100\% \quad (6)$$

Eq. (6) or its other equivalent formulation is also adopted in previous works (Zhao, 2020; Tindale et al., 2020; Ganyani et al., 2020). We estimate η by Monte Carlo simulation of a bivariate distribution accounting for the dependency between the s and ξ . The bivariate distribution constructed by PDFs $f(s)$ and $h(\xi)$ with a covariance matrix of $\begin{bmatrix} \sigma_s^2 & \text{cov}(s, \xi) \\ \text{cov}(s, \xi) & \sigma_\xi^2 \end{bmatrix}$ between s and ξ . Here, σ_s^2 and σ_ξ^2 denote the variances of SI and incubation period, respectively, which are estimated from the likelihood framework in Section 2.2. The $\text{cov}(s, \xi)$ denotes the covariance between s and ξ , which is calculated at 20.2 day². The random variables are generated by using the inverse transform sampling approach.

2.4. Inference of latent period

Base on the η estimate, we approximate the latent period, denoted by ζ . In this approximation, we construct a formula based on two relationships that include

- relationship (I): $\Pr[\nu \leq (E[\xi] - E[\zeta])] = \eta$, where ν denotes the infectious period; and
- relationship (II): $E[\zeta] + [(\alpha_\nu + 1) / (2\alpha_\nu)] \cdot E[\nu] = E[g]$, where α_ν denotes the shape factor of a Gamma-distributed PDF of ν .

The relationship (I) follows the definition of pre-symptomatic transmission straightforwardly, which refers to the time interval between incubation period (ξ) minus latent period (ζ) illustrated in Fig. 1. The ' $\Pr[\cdot]$ ' represents the Gamma cumulative distribution function with mean $E[\nu]$ and shape parameter α_ν for the infectious period. Namely, η is the probability that transmission occurs before symptom onset, which also follows definition in Eq. (5). In relationship (I), a constant risk of transmission is assumed across the infectious period. The relationship (II) is derived in section 5.1 of (Svensson, 2007) as well as adopted in (Krylova and Earn, 2013), which captures the relationship among latent period, infectious period and GI. To ensure both relationships being biologically reasonable, we strict $0 < \zeta < \min(\xi, g)$. We solve the latent period, ζ , based on the best estimates of $h(\xi)$ and $\delta(g)$ in Section 2.2 numerically.

3. Results and discussion

The summary statistics of incubation period and SI are calculated from the 254 selected transmission pairs empirically, see Table 1. Assuming a Gamma distribution, we estimate the mean of incubation period (μ_ξ) at 6.8 days (95 %CI: 6.2, 7.5) and SD (σ_ξ) at 4.1 days (95 %CI: 3.7, 4.8), see Fig. 2A. The μ_ξ estimate is generally consistent with the mean incubation period estimated in (Li et al., 2020a; Backer et al., 2020; Lauer et al., 2020; Lin et al., 2020). We estimate the mean of GI (μ_g) at 6.7 days (95 %CI: 5.4, 7.6) and SD (σ_g) at 1.8 days (95 %CI: 0.3, 3.8), see Fig. 2B. The μ_g estimate is largely consistent with the mean SI or GI of COVID-19 estimated in (He et al., 2020; Zhao, 2020; Ali et al., 2020; Li et al., 2020a; Nishiura et al., 2020a; Ferretti et al., 2020). The σ_g estimate is smaller than the empirical SD of SI estimated at 5.6 days in this study or 4.8 days (95 %CI: 4.5, 5.1) in (Du et al., 2020). The larger SD of SI implies that using SI as a proxy of GI directly is likely an inefficient estimation, which may be due to the large variation in the incubation period of COVID-19 in terms of coefficient of variable at ($CV = SD / \text{mean} = 4.4 / 7.1 = 0.62$). Assuming the intrinsic growth rate of COVID-19 epidemic curve from 0.1 to 0.2 per day (Li et al., 2020a; Zhao et al., 2020d; Musa et al., 2020; Zhuang et al., 2020), the basic reproduction number (R_0) is estimated from 1.9 to 3.6 by using the formula

Table 1

The summary statistics of the epidemiological parameters calculated from the 254 selected transmission pairs.

parameter	notation	mean	median	SD	1Q	3Q	95 %P	unit
incubation period	ξ	7.1	6	4.4	4	10	15	day
serial interval	s	6.9	6	5.6	3	10	18	day
pre-symptomatic transmission	η	0.53	not applicable					unit-free

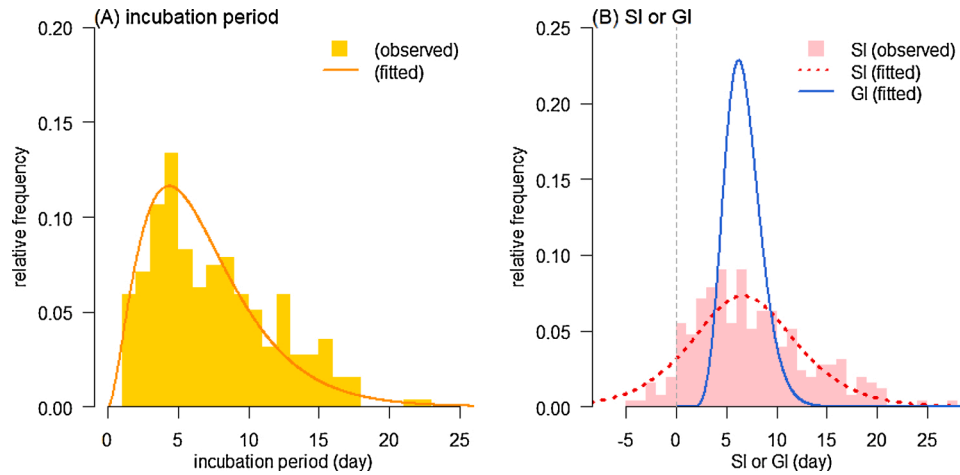


Fig. 2. The observed and fitted (or estimated) probability density distributions of incubation period (panel A), and serial interval (SI) and generation interval (GI, panel B) of COVID-19. In each panel, the histograms are the observed empirical distributions, and the dashed curves are the fitted distributions.

derived from the Euler-Lotka equation in (Wallinga and Lipsitch, 2007), which is consistent with previous R_0 estimates (Riou and Althaus, 2020; Du et al., 2020; Li et al., 2020a; Zhao et al., 2020b; Wu et al., 2020; Ran et al., 2020; Lin et al., 2020; Ferretti et al., 2020; Nishiura et al., 2020b; Wang et al., 2020; Zhao et al., 2020e).

Empirically, we calculated the percentage of the secondary infections due to pre-symptomatic transmission (η) at 53.4%. By using Eq. (6), we estimate η at 49.8% (95%CI: 33.3, 71.5). This η estimate appears largely in line with 32% (95%CI: 10, 74) in (Zhao, 2020), 37% (95%CI: 28, 45) estimated in (Ferretti et al., 2020), 40% (32 out of 80 transmission events) in (Ren et al., 2020), 44% (95%CI: 25, 69) in (He et al., 2020), and from 40% to 80% in (Tindale et al., 2020), respectively, with the 95%CIs largely aligned. In particular, the fraction of pre-symptomatic transmission was estimated around 50% after adjusting the impacts of case isolation (Sun et al., 2021). On one hand, η of 49.8% indicates that there will still be more than 49.8% of the secondary COVID-19 infections if isolation is merely implemented on

the symptomatic cases. On the other hand, if the isolation is implemented on all symptomatic COVID-19 cases immediately after the onset of symptoms, the controlled reproduction number less than $(1 / 49.8\% =) 2.0$ (95%CI: 1.4, 3.0) is thus required for controlling the COVID-19 outbreaks. Given the basic reproduction number of COVID-19 is likely close or larger than 2.0 (Riou and Althaus, 2020; Du et al., 2020; Li et al., 2020a; Zhao et al., 2020b; Wu et al., 2020; Ran et al., 2020; Lin et al., 2020; Ferretti et al., 2020; Zhuang et al., 2020; Nishiura et al., 2020b; Wang et al., 2020; Zhao et al., 2020e; Liu et al., 2020b; Li et al., 2020b), we note that even timely and effectively isolation of all symptomatic cases is necessary but may be insufficient to mitigate the COVID-19 outbreaks, and thus quarantine for the pre-symptomatic and asymptomatic cases, e.g., close contacts, is crucial (Fig. 3).

By using the best estimates of incubation period and GI of COVID-19, we infer the mean latent period, $E[\zeta]$, at 3.9 days (95%CI: 0.5, 8.1), 3.3 days (95%CI: 0.2, 7.9), and 3.4 days (95%CI: 0.2, 8.0) with the Gamma shape (α_ν) of infectious period (ν) at 0.1, 1, and 10, respectively, which is

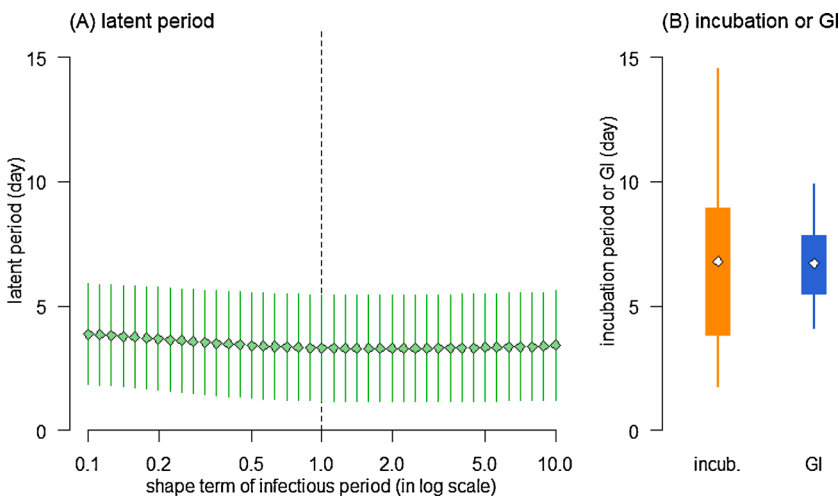


Fig. 3. The inferred latent period (panel A), and estimated incubation period and generation interval (GI, panel B). In panel A, the dots are the mean latent period estimates, and the bars represent the mean \pm SD. The vertical dashed line indicates the Gamma shape term (α_ν) of infectious period (ν) at 1, i.e., when a Gamma distribution is equivalent to an exponential distribution as assumed in the classic ‘SEIR’ compartmental modelling framework. In panel B, the summary statistics of incubation period and GI are shown according to the best estimates. The diamond dots are the estimated means, rectangles are the interquartile ranges, and the bars are the 90% centiles according to the Gamma distributions with best parameter estimates.

consistent with the previous estimate at 3.3 days in (Zhao, 2020) and from 3.4 to 3.7 in (Li et al., 2020b). We note that α_ν commonly ranges from 0.1–10 in the real-world situations. The relationship between α_ν and inferred ζ is shown in Fig. 2. We find the inferred mean latent period converges to 3.4 days as α_ν increasing, i.e., the variation in the infectious period decreases. The mean latent period is lower than those of incubation period and GI, which supports the pre-symptomatic transmission. Additionally, according to the relationship (II) in Section 2.4, the mean infectious period can also be inferred at 6.9 days when α_ν is assumed same as $\alpha_g = 13.8$, where α_g is the estimated shape of $\delta(\cdot)$ for GI. We note this inference is generally consistent with the findings in (Bi et al., 2020). If we set $\alpha_\nu = 1$ that is commonly adopted in the compartmental models of infectious diseases (Sahin, 2020), the mean infectious period is inferred at 3.7 days, which is in line with the previous calculations in (Wu et al., 2020; Li et al., 2020b; Kucharski et al., 2020).

As pointed out in (Champredon and Dushoff, 2015; Park et al., 2021; Britton and Scalia Tomba, 2019), several definitions of SI and GI, e.g., forward or backward delay distributions, were studied based on a cohort perspective at different stages during epidemic, as well as their impacts on shaping the linkage between growth rate and reproduction number. With respect to the infectees, the observations of SI or incubation period are considered as random samples drawn from the backward delay distributions as conceptualized in (Park et al., 2021). For the COVID-19 epidemic in China outside Hubei, the epidemic curve (by case onset date) peaks by the end January 2020, which roughly occurred on January 29 (Leung et al., 2020). We calculate the median of the onset dates of all selected infectees on January 31, which matches the peak of the epidemic curve. As such, the samples of SI and incubation period extracted from the 254 selected samples (i.e., infectees) are considered as good representatives of the infectees cohort during the complete course of the early COVID-19 outbreak (as of April 8, 2020) in China outside Hubei. As our dataset covered the case cohort during the entire outbreak, the potential sampling bias occurred during periods of exponential growth, or exponential decay, is less likely.

The study has limitations. First, the SI observations collected from the public domains might be biased toward cases with ascertainable symptoms, which was also pointed out in (Du et al., 2020). This limitation might lead to an underestimation of the mean GI. In data collection, the online source screening was conducted till April 8, 2020 when either mild or severe COVID-19 cases would be publicly reported. There were 14062 out of 15358 (91.6 %) COVID-19 cases included in the initial screening, and thus the initial samples are considered as good representatives of the ascertained (or known) cases cohort during the complete course of COVID-19 outbreak (as of April 8, 2020) in China outside Hubei. As such, we consider this limitation is minor. Second, recalling biasness might also be another factor causing an underestimation of the mean GI because people are more likely to have precise memory about recent events, e.g., contact history. Third, although there existed unascertained COVID-19 infections, the difference between observed and unobserved cases cannot be assessed in this study due to lack of information. Similarly, the cases with identifiable contacts might appear different from those without traceable contact in terms of epidemiological characteristics. In this study, more than 90 % of the individual cases were excluded due to lack of the information of contact-tracing history or timeline of illness. Hence, the interpretation of our estimates is limited to the COVID-19 cases that are ascertained and with identifiable contacts. Fourth, we presumed the distributions of the incubation periods of infector and infectee are IIDs, and thus $E[\xi_1] = E[\xi_2]$. This is a simplification of the real-world situation because the infector's ξ_1 is difficult to observe. We admit that difference in $E[\xi_1]$ and $E[\xi_2]$ may affect the GI estimates, but we presume this difference is unlikely large. Fifth, the estimation of η and inference of latent period (ζ) are conducted under the assumption that individuals are equally infectious before and after the onset of symptoms. Last but not the least, the effects of nonpharmaceutical interventions (NPI), including the case isolation, may shorten the mean of the observed SI samples (Ali et al.,

2020; Sun et al., 2021; Zhao et al., 2020f), which is considered as the effective SI rather than the intrinsic SI as distinguished in (Park et al., 2020). If the isolation date of each infector were available, the likelihood function of SI, $l_s(\cdot)$, in Eq. (3) could be extended to a right-truncated version (Zhao et al., 2020g), and thus the intrinsic SI and GI can be inferred. Similarly, the likelihood function of incubation period, $l_\zeta(\cdot)$, in Eq. (1) can be revised to an interval-censoring version when the infectee's time of infection is uncertain.

4. Conclusions

For COVID-19, we estimate mean of incubation period at 6.8 days (95 %CI: 6.2, 7.5), and the mean of generation interval at 6.7 days (95 %CI: 5.4, 7.6). We estimate the basic reproduction number ranging from 1.9 to 3.6, and there are 49.8 % (95 %CI: 33.3, 71.5) of the secondary COVID-19 infections are linked to the pre-symptomatic transmission. The mean latent period is inferred at 3.3 days (95 %CI: 0.2, 7.9). Moreover, we highlight the importance of both isolation for symptomatic cases, and quarantine for the pre-symptomatic and asymptomatic cases.

Ethics approval and consent to participate

The COVID-19 contact tracing surveillance data were previously used in Ma et al. (2020), which were all collected originally via the public domains, and thus neither ethical approval nor individual consent was not applicable.

Availability of materials

All data used in this work were publicly available via the data sources in Ma et al. (2020). The extracted dataset and R code are availability as supplementary materials of this study.

Consent for publication

Not applicable.

Funding

DH was supported by General Research Fund (Grant Number 15205119) of the Research Grants Council (RGC) of Hong Kong, China, and an Alibaba (China) Co. Ltd. Collaborative Research grant.

Author's contributions

SZ conceived the study, and carried out the analysis. All authors discussed the results, drafted, read and revised the manuscript, and approved for publishing.

Declaration of Competing Interest

DH received support from an Alibaba (China) Co. Ltd. Collaborative Research grant. MHW is a shareholder of Beth Bioinformatics Co., Ltd. Other authors have no conflict of interest. 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.

Acknowledgements

We thank the reviewers for their helpful comments, which improve this manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.epidem.2021.100482>.

References

- 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 (11), 1714–1719.
- Ali, S.T., Wang, L., Lau, E.H.Y., Xu, X.K., Du ZW, Wu Y., Leung, G.M., Cowling, B.J., 2020. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* 369 (6507), 1106–1109.
- Backer, J.A., Klinkenberg, D., Wallinga, J., 2020. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill.* 25 (5), 2000062.
- Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S.A., Zhang, T., 2020. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* 20 (8), 911–919.
- Bolker, B.M., 2008. *Ecological Models and Data* in R. Princeton University Press.
- Britton, T., Scalia Tomba, G., 2019. Estimation in emerging epidemics: biases and remedies. *J. R. Soc. Interface* 16 (150), 20180670.
- Champreton, D., Dushoff, J., 2015. Intrinsic and realized generation intervals in infectious-disease transmission. *Proc. Biol. Sci.* 282 (1821), 20152026.
- Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S., 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* 178 (9), 1505–1512.
- Cowling, B.J., Fang, V.J., Riley, S., Malik Peiris, J.S., Leung, G.M., 2009. Estimation of the serial interval of influenza. *Epidemiology (Cambridge, Mass)* 20 (3), 344–347.
- 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 (6), 1341–1343.
- Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* 11 (6), 1031–1057.
- 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 (6491), eabb6936.
- Fine, P.E., 2003. The interval between successive cases of an infectious disease. *Am. J. Epidemiol.* 158 (11), 1039–1047.
- 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 (17), 2000257.
- He, X., Lau, E.H.Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., X, Tan, 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* 1–4.
- He, D., Ionides, E.L., King, A.A., 2010. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *J. R. Soc. Interface* 7 (43), 271–283.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)* 395 (10223), 497–506.
- Kong, D., Zheng, Y., Wu, H., Pan, H., Wagner, A.L., Zheng, Y., Gong, X., Zhu, Y., Jin, B., Xiao, W., et al., 2020. Pre-symptomatic transmission of novel coronavirus in community settings. *Influenza Other Respir. Viruses* 14 (6), 610–614.
- Krylova, O., Earn, D.J.D., 2013. Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *J. R. Soc. Interface* 10 (84), 20130098.
- Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R.M., 2020. Centre for Mathematical Modelling of Infectious Diseases C-wg: Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* 20 (5), 553–558.
- Kwok, K.O., Wong, V.W.Y., Wei, W.L., 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 (16), 2000155.
- Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler, J., 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* 172 (9), 577–582.
- 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 (London, England)* 395 (10233), 1382–1393.
- 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., et al., 2020a. Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* 382 (13), 1199–1207.
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020b. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 368 (6490), 489–493.
- Lin, Q., Chiu, A.P., Zhao, S., He, D., 2018. Modelling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia. *Stat. Methods Med. Res.* 27 (7), 1968–1978.
- Lin, Y.F., Duan, Q., Zhou, Y., Yuan, T., Li, P., Fitzpatrick, T., Fu, L., Feng, A., Luo, G., Zhan, Y., et al., 2020. Spread and impact of COVID-19 in China: a systematic review and synthesis of predictions from transmission-dynamic models. *Front. Med. (Lausanne)* 7, 321.
- Liu, Y., Centre for Mathematical Modelling of Infectious Diseases nCo VWG, Funk, S., Flasche, S., 2020a. The contribution of pre-symptomatic infection to the transmission dynamics of COVID-2019. *Wellcome Open Res.* 5 (58), 58.
- Liu, Y., Gayle, A.A., Wilder-Smith, A., Rocklöv, J., 2020b. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* 27 (2).
- Ma, S., Zhang, J., Zeng, M., Yun, Q., Guo, W., Zheng, Y., Zhao, S., Wang, M.H., Yang, Z., 2020. Epidemiological parameters of COVID-19: case series study. *J. Med. Internet Res.* 22 (10), e19994.
- Milwid, R., Steriù, A., Arino, J., Heffernan, J., Hyder, A., Schanzer, D., Gardner, E., Haworth-Brockman, M., Isfeld-Kiely, H., Langley, J.M., et al., 2016. Toward standardizing a lexicon of infectious disease modeling terms. *Front. Public Health* 4 (213), 213.
- Musa, S.S., Zhao, S., Wang, M.H., Habib, A.G., Mustapha, U.T., He, D., 2020. Estimation of exponential growth rate and basic reproduction number of the coronavirus disease 2019 (COVID-19) in Africa. *Infect. Dis. Poverty* 9 (1), 96.
- Nishiura, H., Linton, N.M., Akhmetzhanov, A.R., 2020a. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* 93, 284–286.
- Nishiura, H., Kobayashi, T., Yang, Y., Hayashi, K., Miyama, T., Kinoshita, R., Linton, N.M., Jung, S.M., Yuan, B., Suzuki, A., et al., 2020b. The rate of underascertainment of novel coronavirus (2019-nCoV) infection: estimation using Japanese passengers data on evacuation flights. *J. Clin. Med.* 9 (2).
- Park, S.W., Sun, K., Champredon, D., Li, M., Bolker, B.M., Earn, D.J.D., Weitz, J.S., Grenfell, B.T., Dushoff, J., 2021. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. *Proc. Natl. Acad. Sci.* 118 (2), e2011548118.
- Park, S.W., Champredon, D., Dushoff, J., 2020. Inferring generation-interval distributions from contact-tracing data. *J. R. Soc. Interface* 17 (167), 20190719.
- Ran, J., Zhao, S., Han, L., Liao, G., Wang, K., Wang, M.H., He, D., 2020. A re-analysis in exploring the association between temperature and COVID-19 transmissibility: an ecological study with 154 Chinese cities. *Eur. Respir. J.* 56 (2), 2001253.
- Ren, X., Li, Y., Yang, X., Li, Z., Cui, J., Zhu, A., Zhao, H., Yu, J., Nie, T., Ren, M., et al., 2020. Evidence for pre-symptomatic transmission of coronavirus disease 2019 (COVID-19) in China. *Influenza Other Respir. Viruses* 15 (1), 19–26.
- Riou, J., Althaus, C.L., 2020. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill.* 25 (4), 2000058.
- Sahin, M., 2020. Impact of weather on COVID-19 pandemic in Turkey. *Sci. Total Environ.* 728, 138810.
- Sun, K., Wang, W., Gao, L., Wang, Y., Luo, K., Ren, L., Zhan, Z., Chen, X., Zhao, S., Huang, Y., 2021. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* 371 (6526).
- Svensson, A., 2007. A note on generation times in epidemic models. *Math. Biosci.* 208 (1), 300–311.
- Tariq, A., Lee, Y., Roosa, K., Blumberg, S., Yan, P., Ma, S., Chowell, G., 2020. Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020. *BMC Med.* 18 (1), 1–14.
- Tindale, L.C., Stockdale, J.E., Coombe, M., Garlock, E.S., Lau, W.Y.V., Saraswat, M., Zhang, L., Chen, D., Wallinga, J., Colijn, C., 2020. Evidence for transmission of COVID-19 prior to symptom onset. *Elife* 9, e57149.
- Tuite, A.R., Fisman, D.N., 2020. Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. *Ann. Intern. Med.* 172 (8), 567–568.
- Vink, M.A., Bootsma, M.C., Wallinga, J., 2014. Serial intervals of respiratory infectious diseases: a systematic review and analysis. *Am. J. Epidemiol.* 180 (9), 865–875.
- Wallinga, J., Lipsitch, M., 2007. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B: Biol. Sci.* 274 (1609), 599–604.
- Wang, K., Zhao, S., Liao, Y., Zhao, T., Wang, X., Zhang, X., Jiao, H., Li, H., Yin, Y., Wang, M.H., et al., 2020. Estimating the serial interval of the novel coronavirus disease (COVID-19) based on the public surveillance data in Shenzhen, China, from 19 January to 22 February 2020. *Transbound. Emerg. Dis.* 67 (6), 2818–2822.
- White, L.F., Wallinga, J., Finelli, L., Reed, C., Riley, S., Lipsitch, M., Pagano, M., 2009. Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA. *Influenza Other Respir. Viruses* 3 (6), 267–276.
- Wu, J.T., Leung, K., Leung, G.M., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet (London, England)* 395 (10225), 689–697.
- 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. Reconstruction of transmission pairs for novel coronavirus disease 2019 (COVID-19) in mainland China: estimation of super-spreading events, serial interval, and hazard of infection. *Clin. Infect. Dis.* 71 (12), 3163–3167.
- Yan, P., 2008. Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. *J. Theor. Biol.* 251 (2), 238–252.
- You, C., Deng, Y., Hu, W., Sun, J., Lin, Q., Zhou, F., Pang, C.H., Zhang, Y., Chen, Z., Zhou, X.H., 2020. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int. J. Hyg. Environ. Health* 228, 113555.
- Zhao, S., 2020. Estimating the time interval between transmission generations when negative values occur in the serial interval data: using COVID-19 as an example. *Math. Biosci. Eng.* 17 (4), 3512–3519.
- Zhao, S., Lou, Y., Chiu, A.P.Y., He, D., 2018. Modelling the skip-and-resurgence of Japanese encephalitis epidemics in Hong Kong. *J. Theor. Biol.* 454, 1–10.
- Zhao, S., Cao, P., Gao, D., Zhuang, Z., Cai, Y., Ran, J., Chong, M.K.C., Wang, K., Lou, Y., Wang, W., et al., 2020a. Serial interval in determining the estimation of reproduction

- number of the novel coronavirus disease (COVID-19) during the early outbreak. *J. Travel Med.* 27 (3), taaa033.
- Zhao, S., Musa, S.S., Lin, Q., Ran, J., Yang, G., Wang, W., Lou, Y., Yang, L., Gao, D., He, D., et al., 2020b. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak. *J. Clin. Med.* 9 (2), 388.
- Zhao, S., Zhuang, Z., Cao, P., Ran, J., Gao, D., Lou, Y., Yang, L., Cai, Y., Wang, W., He, D., et al., 2020c. Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-nCoV) cases from Wuhan, China in 2020: a correlational analysis. *J. Travel Med.* 27 (2), taaa022.
- Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., et al., 2020d. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int. J. Infect. Dis.* (92), 214–217.
- Zhao, S., Stone, L., Gao, D., Musa, S.S., Chong, M.K.C., He, D., Wang, M.H., 2020e. Imitation dynamics in the mitigation of the novel coronavirus disease (COVID-19) outbreak in Wuhan, China from 2019 to 2020. *Ann. Transl. Med.* 8 (7), 448.
- Zhao, S., Cao, P., Chong, M.K.C., Gao, D., Lou, Y., Ran, J., Wang, K., Wang, W., Yang, L., He, D., et al., 2020f. COVID-19 and gender-specific difference: analysis of public surveillance data in Hong Kong and Shenzhen, China, from January 10 to February 15, 2020. *Infect. Control Hosp. Epidemiol.* 41 (6), 750–751.
- Zhao, S., Gao, D., Zhuang, Z., Chong, M.K.C., Cai, Y., Ran, J., Cao, P., Wang, K., Lou, Y., Wang, W., et al., 2020g. Estimating the serial interval of the novel coronavirus disease (COVID-19): a statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020. *Front. Phys.* 8, 347.
- Zhao, S., Lou, J., Cao, L., Zheng, H., Chong, M.K.C., Chen, Z., Chan, R.W.Y., Zee, B.C.Y., Chan, P.K.S., Wang, M.H., 2021. Quantifying the transmission advantage associated with N501Y substitution of SARS-CoV-2 in the United Kingdom: an early data-driven analysis. *J. Travel Med.* taab011.
- Zhuang, Z., Zhao, S., Lin, Q., Cao, P., Lou, Y., Yang, L., He, D., 2020. Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: a modelling analysis based on overseas cases and air travel data. *Int. J. Infect. Dis.* 94, 29–31.
- Statement on the Second Meeting of the International Health Regulations Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV)**, World Health Organization (WHO), 2021. [https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)).
- Novel Coronavirus (2019-nCoV) Situation Reports**, Released by the World Health Organization (WHO), 2021. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.