Unraveling the paradox between generation and serial intervals: applications to the COVID-19 pandemic

Abstract

Generation and serial intervals are under-appreciated — often treated as auxilliary variables for estimating the reproduction number $\mathcal{R}$, which is defined as the average number of secondary cases caused by a primary case — yet important quantities in understanding the outset of an outbreak. The generation and serial intervals describe the time between infection and symptom onset, respectively, of an infector and an infectee. Due to their similar definitions, generation and serial intervals are often used interchangeably even though two distributions are different. We argue that a part of the confusion can be attributed to an apparent paradox. As the generation-interval distribution describes the renewal process of infection, it provides a link between the epidemic growth rate r and the reproduction number $\mathcal{R}$; likewise, the serial-interval distribution should describe the renewal process of symptomatic infection and therefore, provide the same link between the epidemic growth rate r and the reproduction number $\mathcal{R}$. However, the current theory does not explain how two different distributions can give the same link between r and $\mathcal{R}$ link. We provide an answer to the paradox. We show that serial intervals have to be defined with reference perspective as well as reference time; the forward serial interval distribution during the exponential provides the correct linke between r and $\mathcal{R}$.

1 Introduction

Since the emergence of the novel coronavirus disease (COVID-19), a significant amount of research has focused on estimating its reproduction number $\mathcal{R}$ (Majumder and Mandl, 2020). The reproduction number is defined as the average number of secondary cases caused by a primary case; the corresponding reproduction number in a fully susceptible population — referred to as the basic reproduction number $\mathcal{R}_0$ — allows us to predict the extent to which a disease will spread in the population and the amount of intervention to prevent an outbreak (Anderson and May, 1991). Since reproduction number cannot be measured directly, particularly during the outset of an outbreak, it is often estimated from the observed exponential growth rate using generation- and serial-interval distributions (e.g., Du et al. (2020); Jung et al. (2020); Li et al. (2020); Zhao et al. (2020)).

The generation interval is defined as the time between when an individual (infector) is infected and when an individual infects another person (infectee); the generation-interval distributions plays a key role in shaping the relationship between the exponential growth rate r and the reproduction number $\mathcal{R}$ (Wallinga and Lipsitch, 2007). Similarly, the serial interval is defined as the time between when an infector and an infectee become *symptomatic* (Svensson, 2007). While serial intervals are similar to generation intervals, previous studies have noted that, in many contexts, serial intervals are expected to have larger variances than generation intervals but have the same mean (Svensson, 2007; Klinkenberg and Nishiura, 2011; te Beest et al., 2013; Champredon et al., 2018); some studies further suggested that using serial intervals can give different estimates of $\mathcal{R}$ (Britton and Scalia Tomba, 2019). Even these distributions were clearly distinguished over a decade ago (Svensson, 2007), the need for a better conceptual and theoretical framework for understanding their differences is becoming clearer as the COVID-19 pandemic unfolds: ~~Researchers continue to rely on both generation and serial intervals to estimate~~ $\mathcal{R}$ ~~for COVID-19 without making a clear distinction.~~

One important source of confusion comes down to an apparent paradox. When the epidemic is growing exponentially, the spread of infection can be characterized as a "renewal process" based on previous incidence of infection, the associated generation-interval distribution, and the average infectiousness of an infected individual. This renewal formulation allows us to link the exponential growth rate of an epidemic r with its reproduction number $\mathcal{R}$ (Wallinga and Lipsitch, 2007). Likewise, we should be able to describe the renewal process of symptomatic cases using the serial-interval distribution. Therefore, both generation- and serial-interval distributions should give us identical estimates of $\mathcal{R}$ based the observed epidemic growth rate r. In contexts where the distributions are expected to be different, current theory has no explanation for how they might link r to $\mathcal{R}$ in the same way.

Here, we provide an answer to this paradox, by showing that the relevant interval for the renewal framework is what is called the "forward" interval, and that the forward serial (but not generation) interval is affected by the rate of growth r during the exponential-growth phase. We develop a new framework for understanding serial intervals and show that the initial forward serial-interval distribution ~~gives the~~ correct value of $\mathcal{R}$. However, using inaccurately defined serial intervals or failing to account for changes in the observed serial-

2

interval distributions over the course of an epidemic can bias the estimate of $\mathcal{R}_t$. We apply our framework to serial intervals of COVID-19 and lay out several principles to consider in using information about serial intervals and other epidemiological time delays in the analysis of the ongoing pandemic.

2 Methods

2.1 Backward and forward delay distributions

We first begin by describing a general framework for characterizing a distribution of time delays between two epidemiological events; these events can be defined either within an infected individual (e.g., infection and symptom onset of an individual, the incubation period) or between infected individuals (e.g., symptom onsets of an infector and an infectee, the serial interval). We can further divide these events into *primary* and *secondary* events. When we measure an epidemiological time delay within an infected individual (e.g., the incubation period), the primary event is the event that always or usually occurs before the secondary event — most epidemiological events that can be observed within an individual have clear direction (e.g., infection and onset of symptoms) but some may not (e.g., onset of infectiousness and onset of symptoms). When we measure an epidemiological time delay between infected individuals (e.g., the serial interval), the primary and secondary events are defined in terms of the direction of transmission: The primary event refers to the event that occurs within an infector and does not necessarily occur before the secondary event.

We model time delays between a primary and a secondary event from a cohort perspective. A primary cohort consists of *all* individuals whose primary event occurred at a given time; a secondary cohort is defined similarly based on the secondary events. For example, when we are measuring serial intervals, a primary cohort π consists of all infectors who became symptomatic at time π. Then, for each primary cohort π, we can define the expected time distribution between primary and secondary events. We refer to this distribution as the forward delay distribution and denote it as $f_\pi(\tau)$.

Likewise, we define the backward delay distribution $b_\delta(\tau)$ for a secondary cohort δ: The backward delay distribution describes the time delays between a primary and secondary host given that the secondary event occurred at time δ. Since both forward and backward perspectives provide valid ways of measuring time delays, we can express the total density of primary and secondary events occurring at time π and δ, respectively, using both forward and backward delay distributions:

$$P(\pi)f_\pi(\delta - \pi) = D(\delta)b_\delta(\delta - \pi), \tag{1}$$

where P and D represent the sizes of primary and secondary cohorts, respectively. Substituting $\tau = \delta - \pi$, it follows that:

$$b_\delta(\tau) = \frac{P(\delta - \tau)f_{\delta-\tau}(\tau)}{D(\delta)} \tag{2}$$

3

Therefore, the backward delay distribution depends on the changes in primary cohort size P (therefore incidence of infection) as well as changes in the forward delay distribution. These ideas apply to all epidemiological delay distributions and generalize the work by (Champredon and Dushoff, 2015) who compared forward and backward generation-interval distributions to describe the realized generation intervals from the perspective of an infector and an infectee, respectively.

2.2 Realized serial interval distributions

The serial interval is defined as the time between when an infector becomes symptomatic and when and infectee becomes symptomatic. Serial intervals τ have been typically written in the form:

$$\tau = -x_0 + \sigma + x_1 \qquad (3)$$

where x_0 and x_1 represent the realized time from infection to symptom onset of an infector and an infectee, respectively, and σ represents the realized generation interval. Previous studies have often assumed that x_0 and x_1 follow the same distributions and concluded that the serial and generation intervals have the same mean (Svensson, 2007; Klinkenberg and Nishiura, 2011; Champredon et al., 2018; Britton and Scalia Tomba, 2019).

The cohort-based framework allows us to understand the distribution of serial intervals we expect to observe when incidence is changing. Given that an infector became symptomatic at time π, we first go backward in time by asking when the infector was infected, and then go forward in time by asking first when the infectee was infected and then when the infectee became symptomatic; this defines the forward serial interval. In Fig. 1A, we see that x_0 represents the backward incubation period of the infector who became symptomatic at time π; σ represents the forward generation interval of the infector who came infected at time $\pi - x_0$; and x_1 represents the forward incubation period for the infectees who became infected at time $\pi - x_0 + \sigma$.
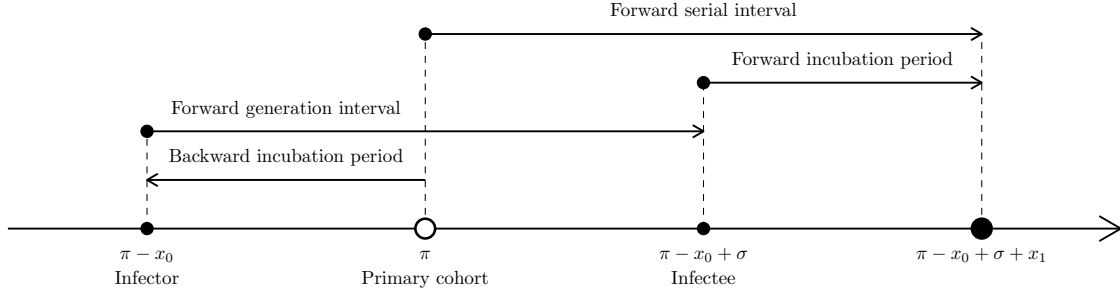
Likewise, we can define the backward serial interval distribution for a secondary cohort δ (Fig. 1B). Given that an infectee became symptomatic at time δ, we have to first go backward in time by asking when the infectee became infected and when the infector became infected; then, we have to go forward in time by asking when the infector became symptomatic. In this case, x_1 represents the backward incubation period of the infectee who became symptomatic at time δ; σ represents the backward generation interval of the infectee who became infected at time $\delta - x_1$; and x_0 represents the forward incubation period of the infector who became infected at time $\delta - x_1 - \sigma$. This conceptual framework clearly demonstrates that the distributions of x_0, δ, and x_1 (and therefore the distributions of serial intervals) depend on the perspective as well as time and cannot be treated statically.

In order to define forward and backward serial interval distributions, we begin by writing the total number of serial intervals between time π (when the infector developed symptoms) and δ (when the infectee developed symptoms), which can be expressed as an integral over all of the possibilities for the infection times of the infector α_1 and the infectee α_2:

$$T(\pi, \delta) = \int_{-\infty}^{\pi} \int_{\alpha_1}^{\delta} \mathcal{R}_c(\alpha_1) i(\alpha_1) h_{\alpha_1}(\pi - \alpha_1, \alpha_2 - \alpha_1) k_{\alpha_2}(\delta - \alpha_2) \mathrm{d}\alpha_2 \, \mathrm{d}\alpha_1, \qquad (4)$$
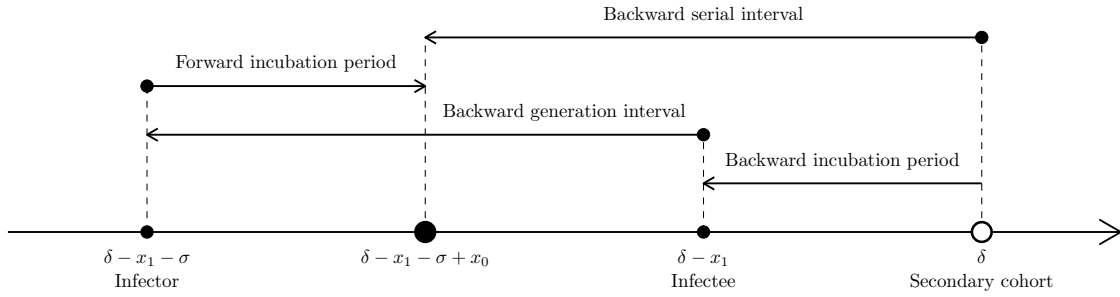
4

Figure 1: **Illustration of forward and backward serial intervals.** (A) The forward serial interval for a primary cohort π (i.e., an infector who became symptomatic at time π). In this case, x_0 represents the backward incubation period of the infector; σ represents the forward generation interval; and x_1 represents the forward incubation period of the infectee. (B) The backward serial interval for a secondary cohort δ (i.e., an infectee who became symptomatic at time δ). In this case, x_0 represents the forward incubation period of the infector; σ represents the backward generation interval; and x_1 represents the backward incubation period of the infectee.

where $\mathcal{R}_c(\alpha_1)$ is the case reproduction number (i.e., the average number of secondary cases caused by a primary case infected at time α_1, Fraser (2007)), $i(\alpha_1)$ is incidence, $h_{\alpha_1}(\pi - \alpha_1, \alpha_2 - \alpha_1)$ is the joint probability distribution describing the forward incubation period $\pi - \alpha_1$ and forward generation interval $\alpha_2 - \alpha_1$ of an individual infected at time α_1 (in this case, the infector), and $k_{\alpha_2}(\delta - \alpha_2)$ describes the marginal probability distribution of h_{α_2} describing the forward incubation period $\delta - \alpha_2$ of an individual infected at time α_2 (in this case, the infectee). Then, the forward serial-interval distribution is proportional to $T(\pi, \pi + \tau)$. Substituting $x = \pi - \alpha_1$ and $\sigma = \alpha_2 - \alpha_1$, we have:

$$f_\pi(\tau) \propto \int_0^\infty \int_0^{\max(0, x+\tau)} \mathcal{R}_c(\pi - x) i(\pi - x) h_{\pi-x}(x, \sigma) k_{\pi-x+\sigma}(x - \sigma + \tau) \mathrm{d}\sigma \, \mathrm{d}x \qquad (5)$$

Likewise, the backward serial-interval distribution is proportional to $T(\delta - \tau, \delta)$. This time,

we substitute $x = \delta - \alpha_2$ and $\sigma = \alpha_2 - \alpha_1$ to get:

$$b_\delta(\tau) \propto \int_0^\infty \int_{\max(0,\tau-x)}^\infty \mathcal{R}_c(\delta - x - \sigma)i(\delta - x - \sigma)h_{\delta-x-\sigma}(x + \sigma - \tau, \sigma)k_{\delta-x}(x)\mathrm{d}\sigma\,\mathrm{d}x. \quad (6)$$

2.3   Epidemic model

We validate the theory by applying it to a specific example of an epidemic model. We model disease spread with a renewal-equation model (Heesterbeek and Dietz, 1996; Diekmann and Heesterbeek, 2000; Roberts, 2004; Aldis and Roberts, 2005; Roberts and Heesterbeek, 2007; Champredon et al., 2018). Ignoring births and deaths, changes in the proportion of susceptible individuals $S(t)$ and incidence of infection $i(t)$ can be written as:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -i(t) \\
i(t) &= \mathcal{R}_0 S(t) \int_0^\infty i(t - \tau)g(\tau)\mathrm{d}\tau,
\end{aligned} \quad (7)$$

where $\mathcal{R}_0$ is the basic reproduction number, and $g(\tau)$ is the intrinsic generation-interval distribution (i.e., the forward generation-interval distribution of a primary case in a fully susceptible population; Champredon and Dushoff (2015)). Then, the forward generation-interval for a primary cohort π follows (Champredon and Dushoff, 2015):

$$g_\pi(\tau) \propto g(\tau)S(\pi + \tau), \quad (8)$$

which allows us to separate the joint probability distribution h_π of the forward incubation period and the forward generation-interval distribution as a product of the proportion of susceptible individuals S and the joint probability distribution h of the forward incubation period and the intrinsic generation intervals:

$$h_\pi(x, \tau) \propto h(x, \tau)S(\pi + \tau). \quad (9)$$

We further assume that the forward incubation period distributions does not vary across cohorts over the course of an epidemic, as they represent the natural history of a disease, and use k without subscript instead. Then, we have:

$$\begin{aligned}
k(x) &= \int_0^\infty h(x, \tau)\mathrm{d}\tau \\
g(\tau) &= \int_0^\infty h(x, \tau)\mathrm{d}x.
\end{aligned} \quad (10)$$

Finally, the case reproduction for this model is defined as follows:

$$\mathcal{R}_c(t) = \mathcal{R}_0 \int_0^\infty g(\tau)S(t + \tau)\mathrm{d}\tau. \quad (11)$$

6

2.4 Linking r and $\mathcal{R}$

During the initial phase of an epidemic, the proportion susceptible remains constant ($S(t) = S(0)$) and incidence of infection grows exponentially: $i(t) = i_0 \exp(rt)$. Then, we can estimate the reproduction number from the exponential growth rate r via the Euler-Lotka equation:

$$\frac{1}{\mathcal{R}} = \int_0^\infty \exp(-r\tau)g(\tau)\mathrm{d}\tau. \tag{12}$$

Like forward generation-interval distributions, forward serial-interval distributions describe the renewal process of symptomatic cases. Therefore, we expect the forward serial-interval distribution during the exponential growth phase — which we refer to as the *initial* forward serial-interval distribution f_0 — to provide the identical r–$\mathcal{R}$ link as the intrinsic generation-interval distribution:

$$\frac{1}{\mathcal{R}} = \int_{-\infty}^\infty \exp(-r\tau)f_0(\tau)\mathrm{d}\tau, \tag{13}$$

where the initial forward serial-interval distribution is defined as:

$$f_0(\tau) \propto \int_0^\infty \int_0^{\max(0,x+\tau)} \exp(-rx)h(x,\sigma)k(x-\sigma+\tau)\mathrm{d}\sigma\,\mathrm{d}x \tag{14}$$

In the Appendix, we provide a mathematical proof that this relationship holds.

The initial forward serial-interval distribution depends on the exponential growth rate r. For a fast-growing epidemic (high r), we expect the backward incubation periods to be short, and therefore, the forward serial-interval distribution will generally have a larger mean than the intrinsic generation-interval distribution. The Susceptible-Exposed-Infected-Recovered model, which assumes that incubation and exposed periods are equivalent, is a special case where the conditional forward generation-interval distribution cancels out with the backward incubation period distribution exactly because (i) infected individuals can only transmit after symptom onset and (ii) the time between symptom onset to infection is independent of the incubation period of an infector; in this case, the forward serial- and generation-intervals follow the same distributions during the exponential growth phase.

We use a simulation-based approach to compare the estimates of $\mathcal{R}$ based on the serial- and generation-interval distributions. To do so, we model the intrinsic generation-interval distribution and the incubation period using a multivariate log-normal distribution with log means μ_G, μ_I, log standard variances σ_G^2, σ_I^2, and correlation ρ; the multivariate log-normal distribution is parameterized based on parameter estimates for COVID-19 (Table 1). We construct forward serial intervals during the exponential growth period as follows:

$$S_i = -B_i + (G_i|B_i) + I_i, \tag{15}$$

where the backward incubation period B_i of an infector is simulated by drawing random log-normal samples A_i with log mean μ_I and log variance σ_I^2 and resampling A_i, each weighted by the inverse of the exponential growth function $\exp(-rA_i)$; the intrinsic generation interval conditional on the incubation period of the infector $(G_i|B_i)$ is drawn from a log-normal

Parameter	Values	Source
Mean forward incubation period	5.5 days	Lauer et al. (2020)
SD forward incubation period	2.4	Lauer et al. (2020)
Mean intrinsic generation interval	5 days	Ferretti et al. (2020)
SD intrinsic generation interval	2	Ferretti et al. (2020)

Table 1: **Parameter values used for simulations.** The intrinsic generation-interval distribution is parameterized using a log-normal distribution with log mean $\mu_G = 1.54$ and log standard deviation $\sigma_G = 0.37$. The forward incubation period distribution is parameterized using a log-normal distribution with log mean $\mu_I = 1.62$ and log standard deviation $\sigma_I = 0.42$. The joint probability distribution is modeled using a multivariate log-normal distribution with correlations $\rho = -0.5, 0, 0.5$.

distribution with log mean $\mu_G + \sigma_G \rho (\log(B_i) - \mu_I)/\sigma_I$ and log variance $\sigma_G^2 (1 - \rho^2)$; the forward incubation period I_i of an infectee is drawn from a log-normal distribution with log mean μ_I and log variance σ_I^2. We then calculate the reproduction number $\mathcal{R}$ using the empirical estimator:

$$\mathcal{R} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \exp(-rS_i)}. \tag{16}$$

We compare this with an estimate of $\mathcal{R}$ based on naive serial-interval distribution that assumes that the backward and the forward incubation periods are identically distributed (Svensson, 2007; Klinkenberg and Nishiura, 2011; Champredon et al., 2018; Britton and Scalia Tomba, 2019):

$$\mathcal{R}_{\text{naive}} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \exp(-rQ_i)}, \tag{17}$$

where

$$Q_i = -A_i + (G_i|A_i) + I_i. \tag{18}$$

3 Results

3.1 Realized serial interval distributions

The initial forward serial-interval distributions $f_0(\tau)$ and the intrinsic generation-interval distribution $g(\tau)$ give identical estimates of $\mathcal{R}$ regardless of the correlation ρ between the incubation period distribution and the intrinsic generation-interval distribution (Fig. 2A). However, using naive serial-interval distributions that do not account for disease dynamics (i.e., assuming that the backward and the forward incubation period distributions are identical) underestimates $\mathcal{R}$; as r increases, $\mathcal{R}_{\text{naive}}$ saturates and eventually decreases due to negative serial intervals (Fig. 2B). While the forward serial intervals during the exponential growth phase can also be negative, the proportion of negative intervals are appropriately

balanced because faster epidemic growth will lead to shorter backward incubation periods
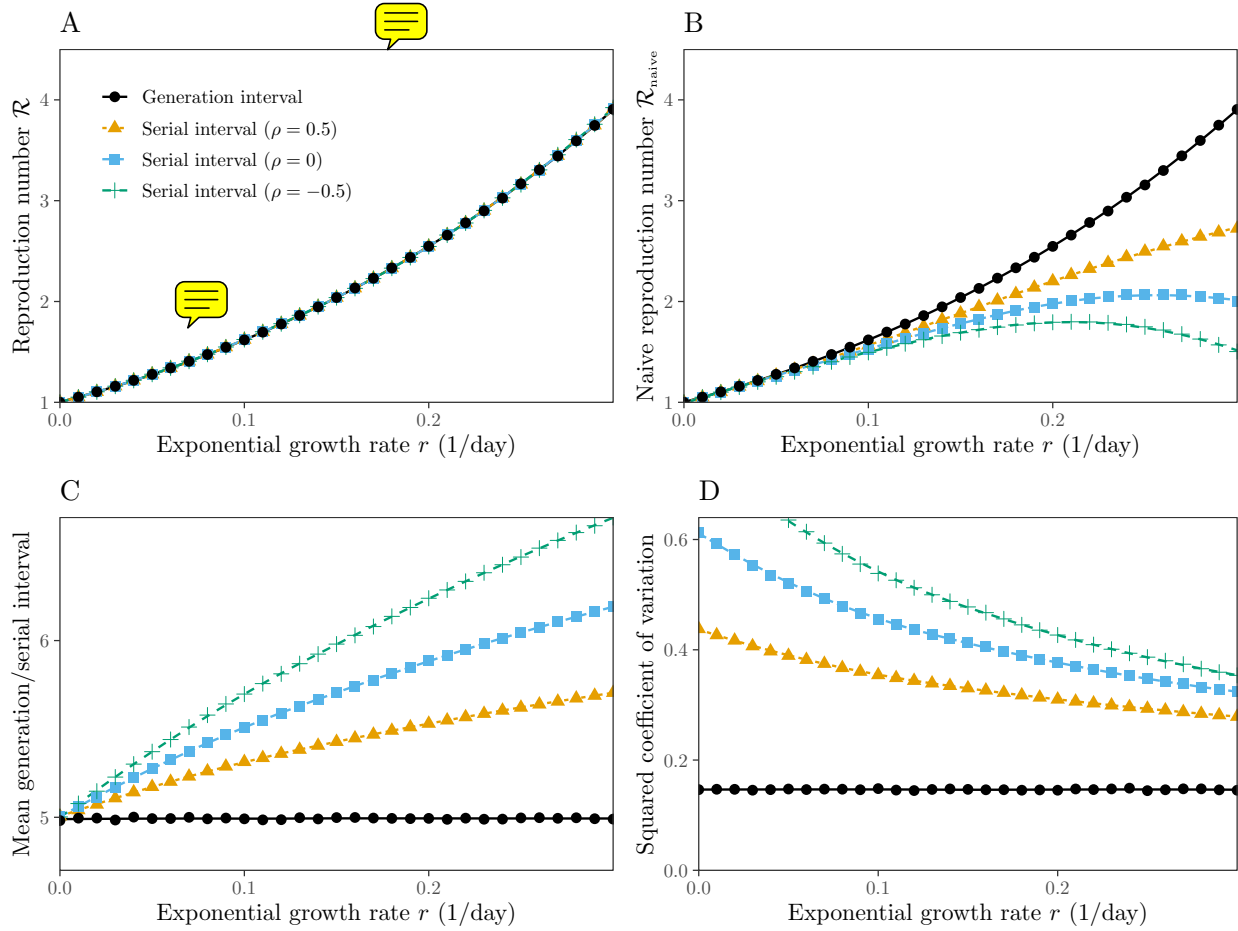(and therefore a lower proportion of negative serial intervals).

Figure 2: **Estimates of the reproduction number from the exponential growth rate based on serial- and generation-interval distributions.** (A). The forward serial-interval distribution during the exponential growth phase give a correct link between the exponential growth rate r and the reproduction number $\mathcal{R}$. (B) The serial-interval distributions that assume that the backward incubation period of an infector and the forward incubation period of an infectee are identically distributed give an incorrect link between r and $\mathcal{R}$. (C) The mean forward serial interval during the exponential growth phase depends on r. (D) The squared coefficient of variation of forward serial intervals during the exponential growth phase depends on r.

Comparing the shapes of initial forward serial-interval distributions and the intrinsic generation-interval distribution allows us to better understand how different distributions are able to give identical estimates of $\mathcal{R}$. In general, generation-interval distributions with higher means and less variability are expected to give higher $\mathcal{R}$ for a given r (Wallinga and Lipsitch, 2007; Park et al., 2019). In this case, the initial forward serial intervals during the

exponential growth phase have higher means (Fig. 2C) and squared coefficients of variation (Fig. 2D) than the intrinsic generation-interval distribution. The effects of higher means (which increases $\mathcal{R}$) and higher variability (which decreases $\mathcal{R}$) cancel out exactly; therefore, we can estimate the same $\mathcal{R}$ using both serial and generation intervals.

The initial forward serial interval distribution only applies to the exponential growth phase of an epidemic. Now, we characterize how forward and backward serial intervals vary over the course of an epidemic when incubation periods and generation intervals are independent. Fig. 3.1 compares the epidemiological dynamics (A) with the mean forward (B–D) and the mean backward (E–F) delay distributions of a deterministic model based on the renewal equation (Eq. 7) and of the corresponding stochastic realizations based on the Gillespie algorithm. The mean forward incubation period remains constant throughout an epidemic as assumed (Fig. 3.1B). The mean forward generation interval decreases slightly as the epidemic progresses because an infected individual is less likely to infect another person as the proportion of susceptible individuals decreases (Fig. 3.1C; Kenah et al. (2008); Champredon and Dushoff (2015)). In contrast, the mean forward serial interval decreases over time (Fig. 3.1D).

The forward serial interval distributions depend on three intervals (Fig. 1): (i) the backward incubation periods of infectors, (ii) the forward generation intervals, and (iii) the forward incubation periods of infectees. Since both forward incubation period (Fig. 3.1B) and generation-interval (Fig. 3.1C) distributions remain roughly constant, changes in the forward serial-interval distributions are predominantly driven by changes in the backward incubation period distributions, whose mean increases over time (Fig. 3.1D). In this case, the increase in the mean backward incubation period directly translates to the decrease in the mean forward serial interval. We note that the relative contributions of the three distributions depend on their shapes and correlations between incubation periods and generation intervals — as discussed earlier, the backward incubation period has *no* effect on the forward serial intervals in the SEIR model, because transmission always occurs after symptom onset.

Qualitative patterns in the mean backward delays is robust across all delay distributions because they are predominantly driven by the changes in incidence (Fig. 3.1D–F). When incidence increases exponentially, individuals are more likely to have been infected more recently and therefore, we are more likely to observe shorter intervals. When incidence decreases, we are more likely to observe longer intervals for similar but opposite reasons. Since changes in incidence can occur even in the absence of significant susceptible depletion (e.g., due to intervention), we expect qualitative patterns in the changes of backward delay distributions to be robust across many outbreaks.

3.2 Observed serial interval distributions

Now, we turn to practical issues. When an epidemic is ongoing, the observed serial intervals are subject to right-censoring because we cannot observe a serial interval if either an infector or an infectee has not yet developed symptoms. Fig. 4 demonstrates how the effect of right-censoring in the observed serial intervals translates to the underestimation of $\mathcal{R}$. Notably, even if we can observe *all* serial intervals across all transmission pairs after the epidemic has
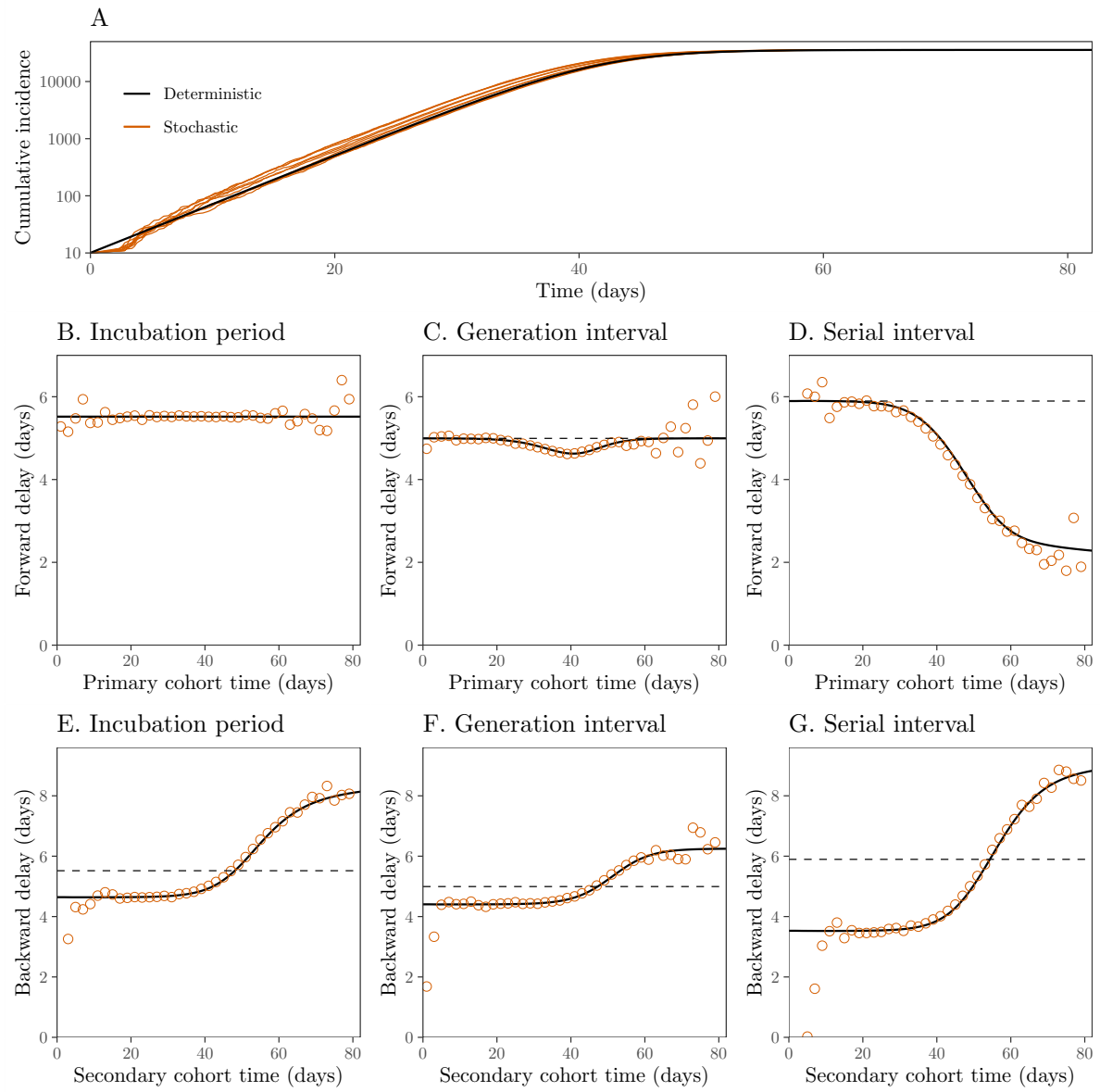
10

Figure 3: **Epidemiological dynamics and changes in mean forward and backward delay distributions.** (A) Cumulative incidence over time. (B–D) Changes in the mean forward incubation period, generation interval, and serial interval. (E–G) Changes in the mean backward incubation period, generation interval, and serial interval. Black lines represent the results of a deterministic simulation. Orange lines and points represent the average of 10 stochastic simulations; cumulative incidence from the stochastic simulations in panel A are shown without averaging. Forward incubation periods and intrinsic generation-intervals are assumed to be independent of each other: $h(x_0, \sigma) = k(x_0)g(\sigma)$. See Table 1 for parameter values.
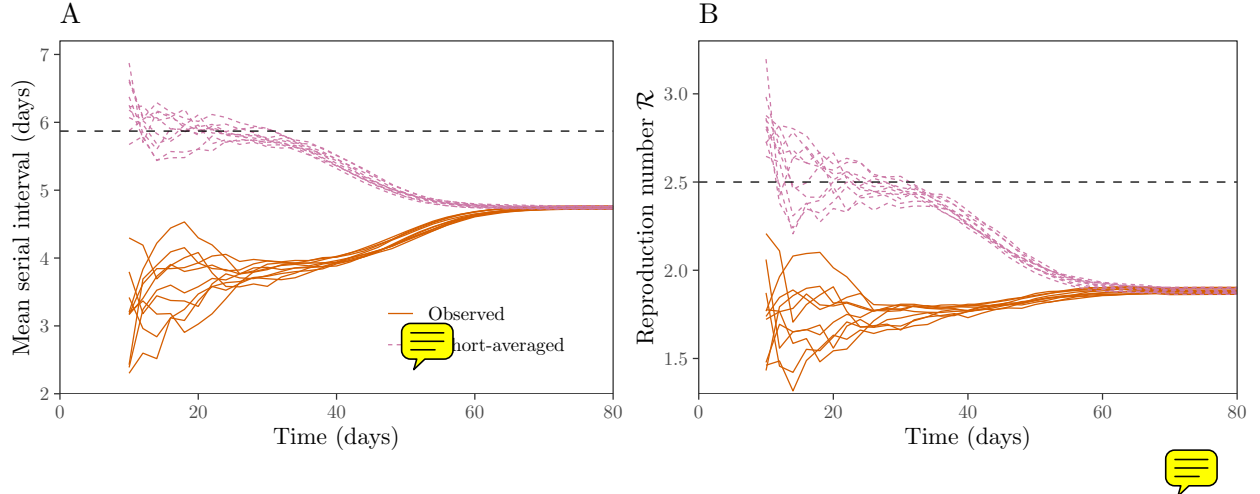
Figure 4: **Estimates of the reproduction number from the observed serial intervals.** (A) Changes in the observed and cohort-averaged mean serial interval over time. (B) Changes in the estimate of $\mathcal{R}$ based on the observed and cohort-averaged mean serial interval over time. Each line represents an independent stochastic realization.

ended, we still underestimate the initial mean forward serial interval (Fig. 4A) and therefore $\mathcal{R}_0$ (Fig. 4) by a large amount because the observed serial-interval distribution does not account for changes in the forward serial-interval distribution. As the mean forward serial-interval distribution decreases over time, taking the average of all observed serial intervals throughout an epidemic will underestimate the mean of the ~~ial forward serial-interval distribution, which provides the correct link between r and $\mathcal{R}$.~~

We provide a simple, heuristic way of assessing potential biases in the estimate of the mean initial forward serial interval and therefore $\mathcal{R}$ retrospectively. Once serial intervals have been observed after the epidemic has been sufficiently progressed, we can group observed serial intervals by their primary cohort times (i.e., symptom onset times of infectors). Then, we can compare how estimates of the mean serial interval as well as $\mathcal{R}$ change as we include more recent cohorts into the analysis (see 'cohort-averaged' in Fig. 4). During the exponential growth phase, the estimates of the mean serial interval and $\mathcal{R}$ are consistent with the target value; adding more data allows us to make more precise inference during this period. However, the cohort-averaged estimates decreases rapidly soon after the exponential growth period. This approach allows us to characterize potential biases in the estimates of $\mathcal{R}$ cuased by changes in the forward serial-interval distributions.

3.3   Applications to the COVID-19 pandemic

Finally, we reanalyze serial intervals of COVID-19 collected by Du et al. (2020) from mainland China, outside Hubei province, based on transmission events reported between January 21–February 8, 2020; Du et al. (2020) estimated the mean serial of 3.96 days (95% CI 3.53–4.39 days) and $\mathcal{R}_0$ of 1.32 (95% CI 1.16–1.48). Fig. 5A shows that the mean forward serial interval decreases over time. While the decrease is likely to be affected by the right-censoring
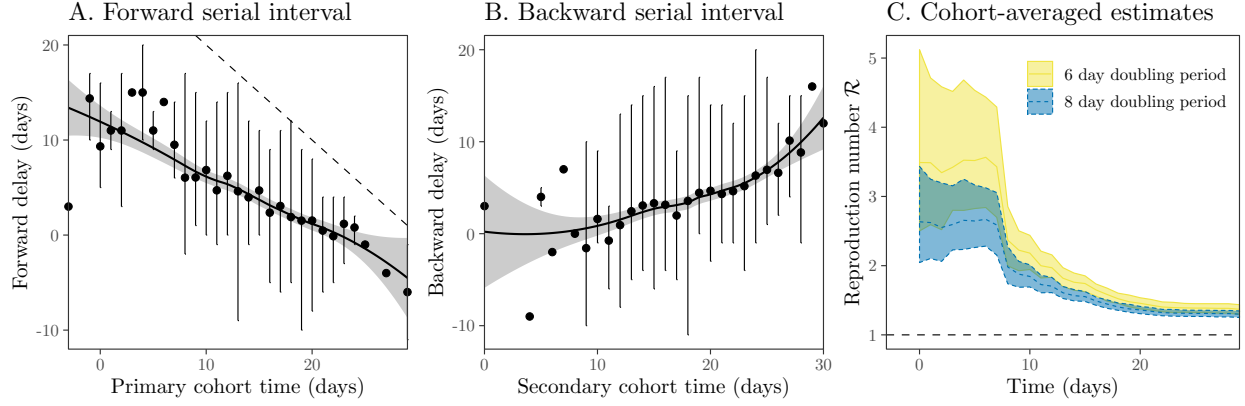
Figure 5: **Observed serial intervals of COVID-19 and cohort-averaged estimates of $\mathcal{R}$.** (A–B) forward and backward serial intervals over time. Points represent the means. Vertical error bars represent the observed maximum and minimum values. Solid lines and ribbons represent the estimated locally estimated scatterplot smoothing (LOESS) fits and associated 95% confidence intervals. The dashed line represents the maximum observable delay for each primary cohort calculated from the most recent observed symptom onset date. (C) Cohort-averaged estimates of $\mathcal{R}$ assuming doubling period of 6 and 8 days (Li et al., 2020; Wu et al., 2020). Ribbons represent the associated 95% bootstrap confidence intervals.

²⁸⁴ (indicated by the closeness between the maximum observed serial intervals and maximum
²⁸⁵ *observable* serial intervals), increase in the proportion of negative serial intervals are clearly
²⁸⁶ indicative of the changes in the forward serial-interval distribution. Fig. 5B shows that
²⁸⁷ the mean backward serial interval increases over time. While the qualitative changes in
²⁸⁸ the mean forward and backward serial interval are consistent with our earlier simulations
²⁸⁹ (Fig. 3.1), the initial mean forward serial interval (Fig. 5A) appears to be larger than what
²⁹⁰ we calculate earlier based on previously estimated incubation period and generation-interval
²⁹¹ distributions (Fig. 2C). These results indicate that previous estimates of incubation period
²⁹² and generation-intervals may have been underestimated as neither study explicitly accounts
²⁹³ for right-censoring (Table 1).

²⁹⁴ Fig. 5C shows the cohort-averaged estimates of $\mathcal{R}_0$, which remain roughly constant until
²⁹⁵ day 7 and suddenly decreases, as expected from our simulation study (Fig. 4). The estimates
²⁹⁶ of $\mathcal{R}_0$ based on the early forward serial intervals are also consistent with previous estimates
²⁹⁷ of $\mathcal{R}_0$ of the COVID-19 epidemic in China (Majumder and Mandl, 2020; Park et al., 2020).
²⁹⁸ We note that early cohort-averaged estimates of $\mathcal{R}_0$ are unlikely to be affected by the right-
²⁹⁹ censoring as we expect the degree of right-censoring to be low (Fig. 5A). This example clearly
³⁰⁰ demonstrates the danger of ~~simply~~ averaging the observed serial intervals and calculating
³⁰¹ the reproduction number using them.

4   Discussion

Characterizing generation- and serial-interval distributions is critical to understanding the outset of an outbreak as they determine the time scale of disease transmission. Generation intervals measure the time difference between infection of a transmission pair, whereas serial intervals measure the time difference between symptom onset of a transmission pair (Svensson, 2007). Due to their similar definitions, we show that their differences have been inaccurately captured previously.

Our study demonstrate the importance of defining a cohort for studying any epidemiological time delay distributions. Previous studies have shown that generation interval can be either measured forward or backward (Champredon and Dushoff, 2015); we generalize their ideas and show that these ideas can be applied to all distributions. Changes in the backward delay distributions is particularly robust across all distributions because they depend on previous changes in cohort sizes, which, in turn, depend on incidence of infection. Previous analyses of COVID-19 epidemics have attempted to reconstruct the time series of symptomatic cases or infected cases from the reports of confirmed cases by assuming a constant backward delay distribution (e.g., Abbott et al. (2020); Park et al. (2020); Shim et al. (2020)). Although such approaches may be able to roughly match the time scale of an epidemic, we recommend using deconvolution approaches instead for proper reconstruction of epidemic time series (Goldstein et al., 2009).

Using cohort-based approaches, we define forward and backward serial intervals. The forward serial-interval distribution describes the renewal process of the symptomatic individuals, and therefore, its initial distribution provides the correct link between r and $\mathcal{R}$. While our results support the use of serial interval distributions for calculating $\mathcal{R}$, they also reveal gaps in current approaches to characterizing serial intervals. Previous studies have (i) recommended using up-to-date serial intervals (Thompson et al., 2019), (ii) assumed that the serial interval distribution remains constant throughout an epidemic (Wallinga and Teunis, 2004; Cori et al., 2013), and (iii) applied serial interval distributions estimated from one location to analyzing epidemics in other locations (Abbott et al., 2020); however, all of these approaches implicitly ignore spatiotemporal variation in incidence, which affects the serial-interval distributions, and may therefore give systematically biased estimates of $\mathcal{R}$. For example, given large geographical differences in the observed exponential growth rates of the COVID-19 epidemic (Abbott et al., 2020), the initial forward serial interval distributions is likely to vary across different countries. Future studies should aim to characterize spatiotemporal variation in serial intervals and understand the drivers of their differences.

We also show that there are multiple serial-interval distributions that correspond to the same set of generation-interval and incubation period distributions depending on their correlations. Conversely, this means that not explicitly accounting for their correlations can bias the estimate of the intrinsic generation-interval distribution from the serial-interval distribution. Furthermore, previous studies that tried to estimate the generation-interval distributions from the observed serial intervals often ignored the differences between the backward incubation period of an infector and the forward incubation period of an infectee (e.g., Klinkenberg and Nishiura (2011); Ganyani et al. (2020)). There is currently a need for

14

³⁴⁴ better statistical tools for teasing apart the intrinsic generation-interval distributions from
³⁴⁵ the observed serial intervals.

³⁴⁶ Our study is not without limitations. Here, we assumed that all individuals develop symp-
³⁴⁷ toms and that the the entire transmission process (e.g., direction of transmission, symptom
³⁴⁸ onset dates, infection dates, etc.) are observable. In practice, asymptomatic and presymp-
³⁴⁹ tomatic transmission of COVID-19 makes measuring serial intervals difficult (Bai et al.,
³⁵⁰ 2020; He et al., 2020; Wei, 2020). For example, symptom onset dates cannot be used as a
³⁵¹ reliable proxy for the direction of transmission as infectees may develop symptoms before
³⁵² their infectors. Based on the COVID-19 parameters (Table 1), we expect roughly 3% of
³⁵³ infectees to develop symptoms before their infectors during the exponential growth period;
³⁵⁴ this probability increases as epidemic progresses because backward incubation period of an
³⁵⁵ infector increases. Biases in the observed serial intervals will necessarily bias the estimates
³⁵⁶ of $\mathcal{R}$. Furthermore, serial intervals do not take into account asymptomatic transmission; not
³⁵⁷ explicitly accounting for asymptomatic transmission may also bias the estimates of $\mathcal{R}$ (Park
³⁵⁸ et al., 2020).

³⁵⁹ Desipte these limitations, our analysis of serial intervals of COVID-19 provide further
³⁶⁰ support for our theoretical framework, demonstrating temporal variation in serial intervals
³⁶¹ and their effects on the estimates of $\mathcal{R}$. To our knowledge, most, if not all, existing estimates
³⁶² of the serial-intervals of COVID-19 implicitly or explicitly assume that the serial-interval
³⁶³ distributions remain constant throughout the course of an epidemic (Du et al., 2020; He
³⁶⁴ et al., 2020; Nishiura et al., 2020; Tindale et al., 2020; Zhao et al., 2020; Zhang et al., 2020).
³⁶⁵ Our study provides a rationale for reassessing estimates of serial-interval distributions of the
³⁶⁶ COVID-19 pandemic.

5 Appendix

5.1 Deterministic simulation

We simulate the renewal equation model using a discrete-time approximation:

$$i(t) = \mathcal{R}_0 S(t - \Delta t) \sum_{m=1}^{\ell} i(t - \Delta tm) \hat{g}(\Delta tm)$$
$$S(t) = S(t - \Delta t) - i(t) \tag{19}$$

where $\hat{g}$ is a discrete-time intrinsic generation-interval distribution that satisfies the following:

$$\hat{g}(\Delta tm) = \frac{g(\Delta tm)}{\sum_{i=1}^{\ell} g(\Delta tm)}, \quad m = 1, \ldots, \ell \tag{20}$$

The continuous-time intrinsic generation-interval distribution is parameterized using a log-normal distribution (Table 1). We define the forward incubation period distribution in a similar manner:

$$\hat{k}(\Delta tm) = \frac{k(\Delta tm)}{\sum_{i=1}^{\ell} k(\Delta tm)}, \quad m = 1, \ldots, \ell, \tag{21}$$

where its continuous-time analog is also based on a log-normal distribution. For brevity, we assume that the forward incubation periods and intrisic generation intervals are independent:

$$\hat{h}(\Delta tm, \Delta tn) = \hat{k}(\Delta tm) \hat{g}(\Delta tn), \quad m, n = 1, \ldots, \ell. \tag{22}$$

We use $\Delta t = 0.025$ days and $\ell = 2001$ for discretization steps.

We initialize the simulation with population size N=40,000 as follows:

$$i(\Delta tm) = C \exp(r\Delta tm), \quad m = 1, \ldots, \ell$$
$$S(\Delta tm) = N - \sum_{n=1}^{m} i(\Delta tm), \quad m = 1, \ldots, \ell \tag{23}$$

where C is chosen such that $\sum_{n=1}^{\ell} i(\Delta tm) = 10$. These initial conditions allow the model to follow exponentially growth from time $\Delta t(\ell + 1)$ without any transient behaviors. Results presented in the main text show simulations beginning from time $\Delta t(\ell + 1)$.

5.2 Stochastic simulation

We run stochastic simulations of the renewal equation model using an individual-based model on a fully connected network (i.e., homogeneous population) based on the Gillespie algorithm that we devloped earlier (Park et al., 2019). First, we initialize an epidemic with $I(0)$ infected individuals (nodes) in a fully connected network of size N. For each initially infected individual, we draw number of infectious contacts from a Poisson distribution with the mean of $\mathcal{R}_0$ and the corresponding generation intervals for each contact from a log-normal distribution (Table 1). Contactees are uniformly sampled from the total population.

16

All contactees are sorted into event queues based on their infection time. We update the current time to the infection time of the first person in the queue. Then, the first person in the queue makes contacts based on the Poisson offspring distribution described earlier and their contactees are added to the sorted queue. Whenever contactees are aded to the sorted queue, we remove all duplicated contacts (but keep the first one) as well as contacts made to individuals that have already been infected. Simulations continue until there are no more individuals in the queue. We simulate 10 epidemics with $I(0) = 10$ and $N=40,000$.

5.3   Linking r and $\mathcal{R}$ using serial-interval distributions

The intrinsic generation-interval distribution $g(\tau)$ provides a link between r and $\mathcal{R}$ via the Euler-Lotka equation:

$$\frac{1}{\mathcal{R}} = \int_0^\infty \exp(-r\tau)g(\tau)\mathrm{d}\tau. \tag{24}$$

Here, we prove that the initial forward serial-interval distribution f_0 also provides the same link:

$$\frac{1}{\mathcal{R}} = \int_{-\infty}^\infty \exp(-r\tau)f_0(\tau)\mathrm{d}\tau, \tag{25}$$

where the initial forward serial-interval distribution is defined as:

$$f_0(\tau) \propto \int_0^\infty \int_0^{\max(0,x+\tau)} \exp(-rx)h(x,\sigma)k(x-\sigma+\tau)\mathrm{d}\sigma\,\mathrm{d}x. \tag{26}$$

First, we rewrite the initial forward serial-interval distribution in the following form:

$$\begin{aligned}
f_0(\tau) &\propto \int_0^\infty \int_{-\alpha_1}^\tau \exp(-r\alpha_1)h(\alpha_1,\alpha_2+\alpha_1)k(\tau-\alpha_2)\mathrm{d}\alpha_2\,\mathrm{d}\alpha_1 \\
&= \int_{-\infty}^\tau \int_{\max(0,-\alpha_2)}^\infty \exp(-r\alpha_1)h(\alpha_1,\alpha_2+\alpha_1)k(\tau-\alpha_2)\mathrm{d}\alpha_1\,\mathrm{d}\alpha_2
\end{aligned} \tag{27}$$

It follows that $z(\alpha_2)$ describes the time between symptom onset of infector and infection of infectee:

$$z(\alpha_2) \propto \int_{\max(0,-\alpha_2)}^\infty \exp(-r\alpha_1)h(\alpha_1,\alpha_2+\alpha_1)\mathrm{d}\alpha_1. \tag{28}$$

The initial forward serial-interval distribution f_0 can be expressed as a convolution of z and k:

$$f_0(\tau) \propto \int_{-\infty}^\tau z(\alpha_2)k(\tau-\alpha_2)\mathrm{d}\alpha_2. \tag{29}$$

Then, we have

$$\int_{-\infty}^\infty \exp(-r\tau)f_0(\tau)\mathrm{d}\tau = \int_{-\infty}^\infty \exp(-r\tau)z(\tau)\mathrm{d}\tau \int_0^\infty \exp(-r\tau)k(\tau)\mathrm{d}\tau \tag{30}$$

17

Note that

$$
\int_{-\infty}^{\infty} \int_{\max(0,-\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) \mathrm{d}\alpha_1 \mathrm{d}\alpha_2
$$
$$
= \int_{0}^{\infty} \int_{-\alpha_1}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) \mathrm{d}\alpha_2 \, \mathrm{d}\alpha_1 \tag{31}
$$
$$
= \int_{0}^{\infty} \exp(-r\alpha_1) k(\alpha_1) d\alpha_1
$$

because k is a marginal probability distribution of h. Since $\int_{0}^{\infty} \exp(-r\alpha_1) k(\alpha_1) d\alpha_1$ is a normalization factor for probability distribution z, we have:

$$
\int_{-\infty}^{\infty} \exp(-r\tau) z(\tau) \mathrm{d}\tau = \frac{\int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0,-\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) \mathrm{d}\alpha_1 \mathrm{d}\tau}{\int_{0}^{\infty} \exp(-r\alpha_1) k(\alpha_1) d\alpha_1}. \tag{32}
$$

Then, we have:

$$
\int_{-\infty}^{\infty} \exp(-r\tau) f_0(\tau) \mathrm{d}\tau = \int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0,-\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) \mathrm{d}\alpha_1 \mathrm{d}\tau. \tag{33}
$$

We are left to show that

$$
\int_{0}^{\infty} \exp(-r\tau) g(\tau) \mathrm{d}\tau = \int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0,-\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) \mathrm{d}\alpha_1 \mathrm{d}\tau, \tag{34}
$$

where the intrisic generation-interval distribution g is also a marginal probability distribution of f:

$$
g(\tau) = \int_{0}^{\infty} h(\alpha_1, \tau) \mathrm{d}\alpha_1. \tag{35}
$$

Let $\sigma = \alpha_1 + \tau$. Then, by change of variables, it immediately follows that

$$
\int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0,-\tau)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \tau + \alpha_1) \mathrm{d}\alpha_1 \, \mathrm{d}\tau
$$
$$
= \int_{0}^{\infty} \int_{0}^{\infty} \exp(-r\sigma) h(\alpha_1, \sigma) \mathrm{d}\alpha_1 \, \mathrm{d}\sigma \tag{36}
$$
$$
= \int_{0}^{\infty} \exp(-r\tau) g(\tau) \mathrm{d}\tau
$$

Therefore, the initial forward serial–interval distribution and the intrinsic generation-interval distribution give the same link between r and $\mathcal{R}$.

References

Abbott, S., J. Hellewell, J. D. Munday, J. Y. Chun, R. N. Thompson, N. I. Bosse, Y.-W. D. Chan, T. W. Russell, C. I. Jarvis, CMMID nCov working group, S. Flasche,

A. J. Kucharski, R. Eggo, and S. Funk (2020). Temporal variation in transmission during the COVID-19 outbreak. `https://cmmid.github.io/topics/covid19/current-patterns-transmission/global-time-varying-transmission.html`. Accessed April 20, 2020.

Aldis, G. and M. Roberts (2005). An integral equation model for the control of a smallpox outbreak. *Mathematical biosciences 195*(1), 1–22.

Anderson, R. M. and R. M. May (1991). *Infectious diseases of humans: dynamics and control.* Oxford university press.

Bai, Y., L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, and M. Wang (2020). Presumed asymptomatic carrier transmission of COVID-19. *Jama*.

Britton, T. and G. Scalia Tomba (2019). Estimation in emerging epidemics: Biases and remedies. *Journal of the Royal Society Interface 16*(150), 20180670.

Champredon, D. and J. Dushoff (2015). Intrinsic and realized generation intervals in infectious-disease transmission. *Proceedings of the Royal Society B: Biological Sciences 282*(1821), 20152026.

Champredon, D., J. Dushoff, and D. J. Earn (2018). Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation. *SIAM Journal on Applied Mathematics 78*(6), 3258–3278.

Cori, A., N. M. Ferguson, C. Fraser, and S. Cauchemez (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology 178*(9), 1505–1512.

Diekmann, O. and J. A. P. Heesterbeek (2000). *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, Volume 5. John Wiley & Sons.

Du, Z., X. Xu, Y. Wu, L. Wang, B. J. Cowling, and L. A. Meyers (2020). Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerging Infectious Diseases 26*(6).

Ferretti, L., C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*.

Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PloS one 2*(8).

Ganyani, T., C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens (2020). Estimating the generation interval for COVID-19 based on symptom onset data. *medRxiv*.

Goldstein, E., J. Dushoff, J. Ma, J. B. Plotkin, D. J. Earn, and M. Lipsitch (2009). Reconstructing influenza incidence by deconvolution of daily mortality time series. *Proceedings of the National Academy of Sciences 106*(51), 21825–21829.

He, X., E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 1–4.

Heesterbeek, J. and K. Dietz (1996). The concept of $\mathcal{R}_0$ in epidemic theory. *Statistica Neerlandica 50*(1), 89–110.

Jung, S.-m., A. R. Akhmetzhanov, K. Hayashi, N. M. Linton, Y. Yang, B. Yuan, T. Kobayashi, R. Kinoshita, and H. Nishiura (2020). Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *Journal of clinical medicine 9*(2), 523.

Kenah, E., M. Lipsitch, and J. M. Robins (2008). Generation interval contraction and epidemic data analysis. *Mathematical biosciences 213*(1), 71–79.

Klinkenberg, D. and H. Nishiura (2011). The correlation between infectivity and incubation period of measles, estimated from households with two cases. *Journal of theoretical biology 284*(1), 52–60.

Lauer, S. A., K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*.

Li, Q., X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*.

Majumder, M. S. and K. D. Mandl (2020). Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility. *The Lancet Global Health*.

Nishiura, H., N. M. Linton, and A. R. Akhmetzhanov (2020). Serial interval of novel coronavirus (COVID-19) infections. *International Journal of Infectious Diseases*.

Park, S. W., B. M. Bolker, D. Champredon, D. J. Earn, M. Li, J. S. Weitz, B. T. Grenfell, and J. Dushoff (2020). Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. *medRxiv*.

Park, S. W., D. Champredon, and J. Dushoff (2019). Inferring generation-interval distributions from contact-tracing data. *bioRxiv*, 683326.

Park, S. W., D. Champredon, J. S. Weitz, and J. Dushoff (2019). A practical generation-interval-based approach to inferring the strength of epidemics from their speed. *Epidemics 27*, 12–18.

Park, S. W., D. M. Cornforth, J. Dushoff, and J. S. Weitz (2020). The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak. *medRxiv*.

Park, S. W., K. Sun, C. Viboud, B. T. Grenfell, and J. Dushoff (2020). Potential roles of social distancing in mitigating the spread of coronavirus disease 2019 (COVID-19) in South Korea. *medRxiv*.

Roberts, M. (2004). Modelling strategies for minimizing the impact of an imported exotic infection. *Proceedings of the Royal Society of London. Series B: Biological Sciences 271*(1555), 2411–2415.

Roberts, M. and J. Heesterbeek (2007). Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *Journal of mathematical biology 55*(5-6), 803.

Shim, E., A. Tariq, W. Choi, Y. Lee, and G. Chowell (2020). Transmission potential and severity of COVID-19 in South Korea. *International Journal of Infectious Diseases*.

Svensson, Å. (2007). A note on generation times in epidemic models. *Mathematical biosciences 208*(1), 300–311.

te Beest, D. E., J. Wallinga, T. Donker, and M. van Boven (2013). Estimating the generation interval of influenza A (H1N1) in a range of social settings. *Epidemiology*, 244–250.

Thompson, R., J. Stockwin, R. van Gaalen, J. Polonsky, Z. Kamvar, P. Demarsh, E. Dahlqwist, S. Li, E. Miguel, T. Jombart, et al. (2019). Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics 29*, 100356.

Tindale, L., M. Coombe, J. E. Stockdale, E. Garlock, W. Y. V. Lau, M. Saraswat, Y.-H. B. Lee, L. Zhang, D. Chen, J. Wallinga, et al. (2020). Transmission interval estimates suggest pre-symptomatic spread of COVID-19. *medRxiv*.

Wallinga, J. and M. Lipsitch (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences 274*(1609), 599–604.

Wallinga, J. and P. Teunis (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology 160*(6), 509–516.

Wei, W. E. (2020). Presymptomatic Transmission of SARS-CoV-2—Singapore, January 23–March 16, 2020. *MMWR. Morbidity and Mortality Weekly Report 69*.

Wu, J. T., K. Leung, and G. M. Leung (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet 395*(10225), 689–697.

Zhang, J., M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi, X. Chen, et al. (2020). Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases*.

Zhao, S., D. Gao, Z. Zhuang, M. Chong, Y. Cai, J. Ran, P. Cao, K. Wang, Y. Lou, W. Wang, et al. (2020). 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. *medRxiv*.

Zhao, S., Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He, et al. (2020). 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. *International journal of infectious diseases 92*, 214–217.