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

Sang Woo Park

Abstract

1 Introduction

Since the emergence of the novel coronavirus disease (COVID-19), a significant amount of research has focused on estimating its reproduction number $\mathcal{R}$. The reproduction number, which is defined as the average number of secondary cases caused by a primary case, allows us to predict the extent to which a disease will spread in the population and the amount of intervention to prevent an outbreak. Since the reproduction number cannot be measured directly, it is often estimated from the observed exponential growth rate using generation- and serial-interval distributions.

The generation interval is defined as the time between when an individual (infector) is infected and when an individual infects another person (infectee). On the other hand, the serial interval is defined as the time between when an infector and an infectee become symptomatic. Previous studies have often alluded to differences in their variances, despite having identical expected values, and noted that using serial-interval distribution can give different estimates of $\mathcal{R}$. Even though these distinctions were first made 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, and some even mislabel them.

The lack of clear understanding of their differences can be largely attributed 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 $\mathcal{R}$. Likewise, since infection typically leads to symptoms, 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. Current theory seems to contradict our biological intution.

Here, we provide an answer to the decade-old paradox. We provide a new framework for understanding serial intervals and show that both serial- and generation-interval distributions give identical estimates of $\mathcal{R}$ during the exponential growth phase of an epidemic.

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 *any* two epidemiological events; these events can be defined either within an infected individual (e.g., infection and symptom onset of an individual) or between infected individuals (e.g., symptom onsets of an infector and an infectee). Then, we can further divide these events into *primary* and *secondary* events. When we measure an epidemiological time delay within an infected individual (e.g., time between infection and symptom onset), the

primary event always occurs before the secondary event. When we measure an epidemiological time delay between infected individuals (e.g., 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 can be defined similarly based on the secondary events. For example, when we are measurign serial intervals, a primary cohort s consists of all infectors who became symptomatic at time s. Then, for each primary cohort s, 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 $f_s(\tau)$. The forward delay distributions can vary across primary cohorts.

Likewise, we can define a backward delay distribution $b_s(\tau)$ for a secondary cohort s: The backward delay distribution for secondary host s describes the time delays between a primary and secondary host given that the secondary event occurred at time s. It follows that:

$$b_s(\tau) \propto C(s-\tau) f_{s-\tau}(\tau) \tag{1}$$

where $C(s)$ is the size of the primary cohort s. Therefore, changes in the backward delay distribution depends on the changes in cohort size $C(s)$ (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. Here, we denote realized serial intervals τ as:

$$\tau = -x_0 + \sigma + x_1 \tag{2}$$

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 assumed that (i) x_0 and x_1 have the same means and (ii) therefore the serial and generation intervals have the same mean; however, these results implicitly assume that incidence stays constant.

Using the cohort-based framework provides a clear way of understanding the serial-interval distribution. Given that an infector became symptomatic at time ℓ, we have to first go backward in time by asking when the infector was infected and go forward in time by asking when the infectee became symptomatic. Then, it is clear that x_0 follows the backward incubation period distribution for secondary cohort ℓ; σ follows the forward generation-interval distribution for a primary cohort $\ell - x_0$ conditional on the incubation period x_0 of

the infector; and x_1 follows forward incubation period distribution for a primary cohort $\ell - x_0 + \sigma$. Assuming that the forward incubation distribution does not vary across cohorts, the forward serial interval distribution of a primary cohort ℓ can be written as follows:

$$f_\ell(\tau) \propto \int_0^\infty \int_0^\infty i(\ell - x_0) h_{\ell - x_0}(x_0, \sigma) k(\tau - \sigma + x_0) \mathrm{d}x_0 \, \mathrm{d}\sigma \tag{3}$$

where i is the incidence of infection, h is the joint probability distribution of the forward incubation period and forward generation interval of a primary cohort $\ell - x_0$, and k is a marginal probability distribution of h describing the forward incubation periods:

$$k(x_0) = \int_0^\infty h_\ell(x_0, \sigma) \mathrm{d}\sigma. \tag{4}$$

Here, we assume that the forward incubation period distribution does not vary across cohorts over the course of an epidemic as it represents an intrinsic property of the natrual history of a disease.

Likewise, we can define backward serial interval distribution for a secondary cohort ℓ. Given that an infectee became symptomatic at time ℓ, 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 and σ follow the backward incubation period and generation-interval distributions, respectively, and x_0 follows the forward incubation period distribution. Therefore, the backward serial interval distribution of a secondary cohort ℓ can be written as follows

$$\begin{aligned} b_\ell(\tau) &\propto j(\ell) f_\ell(\tau) \\ &\propto \int_0^\infty i(\ell - x_0) k(x_0) f_\ell(\tau) \mathrm{d}x_0 \end{aligned} \tag{5}$$

where j represents the incidence of symptomatic cases.

2.3 Epidemic model

Here, use a renewal equation to model the spread of disease in a population:

$$i(t) = \mathcal{R}_0 S(t) \int_0^\infty i(t - \tau) g(\tau) \mathrm{d}\tau, \tag{6}$$

where $\mathcal{R}_0$ is the basic reproduction number (i.e., the average number of secondary cases caused by a primary case in a fully susceptible population), $g(\tau)$ is the intrinsic generation-interval distribution (i.e., the forward generation-interval distribution of a primary case in a fully susceptible population), $S(t)$ is the proportion of susceptible individuals. Then, the forward generation-interval for a primary cohort ℓ follows (Champredon and Dushoff, 2015):

$$g_\ell(\tau) \propto g(\tau) S(\ell + \tau), \tag{7}$$

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

$$h_\ell(x_0, \tau) \propto h(x_0, \tau) S(\ell + \tau), \tag{8}$$

which satisfies the following:

$$g(\tau) = \int_0^\infty h(x_0, \tau) \mathrm{d}x_0. \tag{9}$$

2.4 Linking r and $\mathcal{R}$

During the initial phase of an epidemic, the proprotion 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:

$$\frac{1}{\mathcal{R}} = \int_0^\infty \exp(-r\tau) g(\tau) \mathrm{d}\tau. \tag{10}$$

Like forward generation-interval distributions, forward serial-interval distributions describe the renewal process of symptomatic cases. Therefore, the forward serial-interval distribution $f_{\exp}(\tau)$ during the exponential growth phase provides the identical r–$\mathcal{R}$ link as the intrinsic generation-interval distribution:

$$\frac{1}{\mathcal{R}} = \int_{-\infty}^\infty \exp(-r\tau) f_{\exp}(\tau) \mathrm{d}\tau, \tag{11}$$

where the forward serial-interval distribution during the exponential growth phase is defined as:

$$f_{\exp}(\tau) \propto \int_0^\infty \int_0^\infty \exp(-r x_0) h(x_0, \sigma) k(\tau - \sigma + x_0) \mathrm{d}x_0 \, \mathrm{d}\sigma. \tag{12}$$

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

We note that the forward serial-interval distribution depends on the exponential growth rate r. When the epidemic grows fast (high r), we expect the backward incubation period to be short, and therefore, the forward serial-interval distribution will generally have a larger mean than the forward 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 generation-interval 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 have the same distributions during the exponential growth phase.

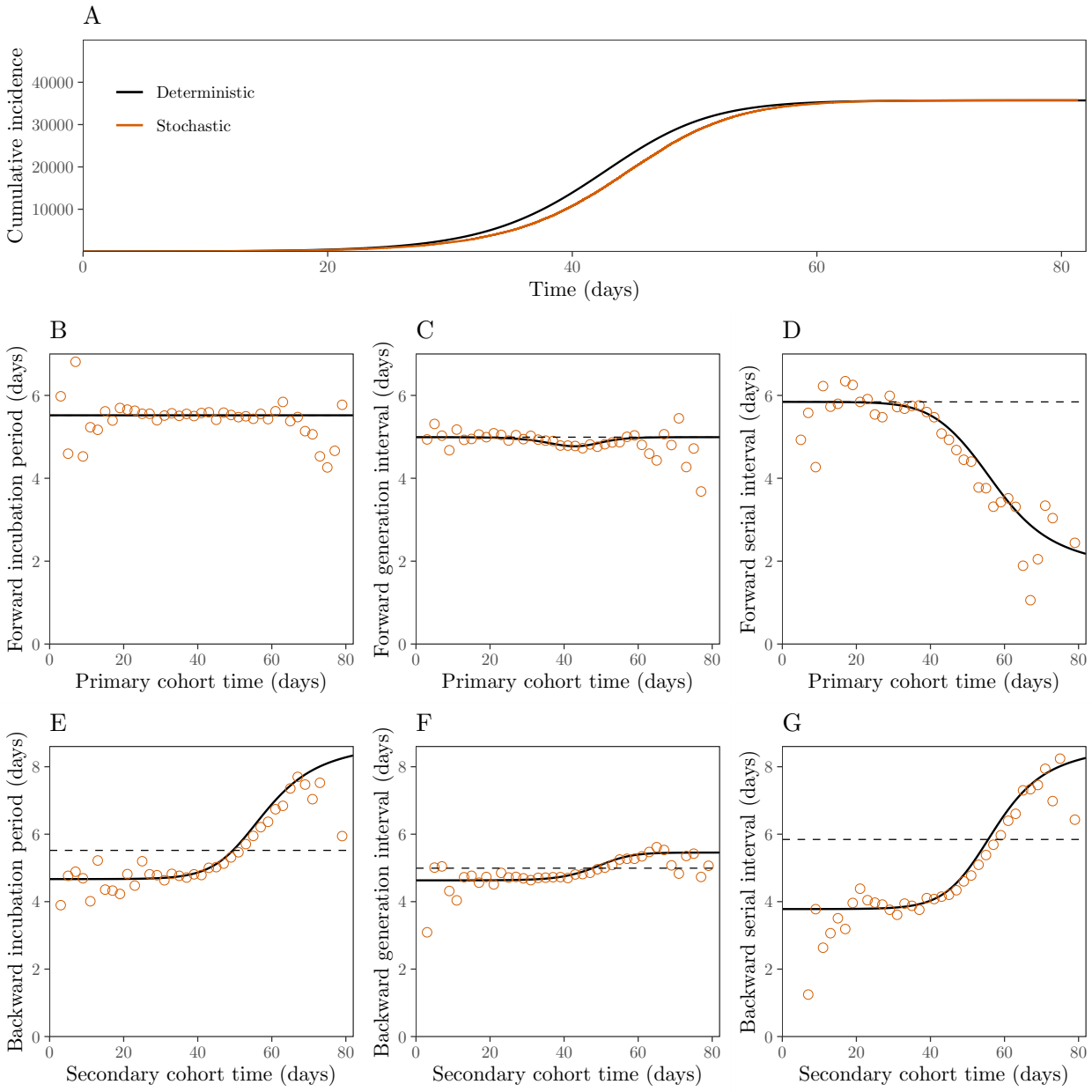

Figure 1: **Epidemiological dynamics and changes in mean forward and backward delay distributions.**

3 Results

Fig. 1 compares the epidemiological dynamics (A) with the mean forward (B–D) and the mean backward (E–F) delay distributions. The mean forward incubation period remains constant throughout an epidemic as expected (Fig. 1B). The mean forward generation interval contracts as epidemic progresses because an infected individual is less likely to infect

another person as the proportion of susceptible individuals decreases (Fig. 1C; Champredon and Dushoff (2015)). In contrast to mean forward incubation period or generation interval, the mean serial interval depends on previous incidence and therefore decreases over time (Fig. 1D): When incidence is increasing, a symptomatic individual is more likely to have been infected more recently (shorter backward incubation period and therefore longer forward serial interval) whereas when incidence is decreasing, a symptomatic individual is more likely to have been infected later (longer backward incubation period and therefore shorter forward serial interval). Qualitative patterns in the changes in the mean backward delays is robust across all delay distributions because they are largely driven by changes in incidence (Fig. 1D).

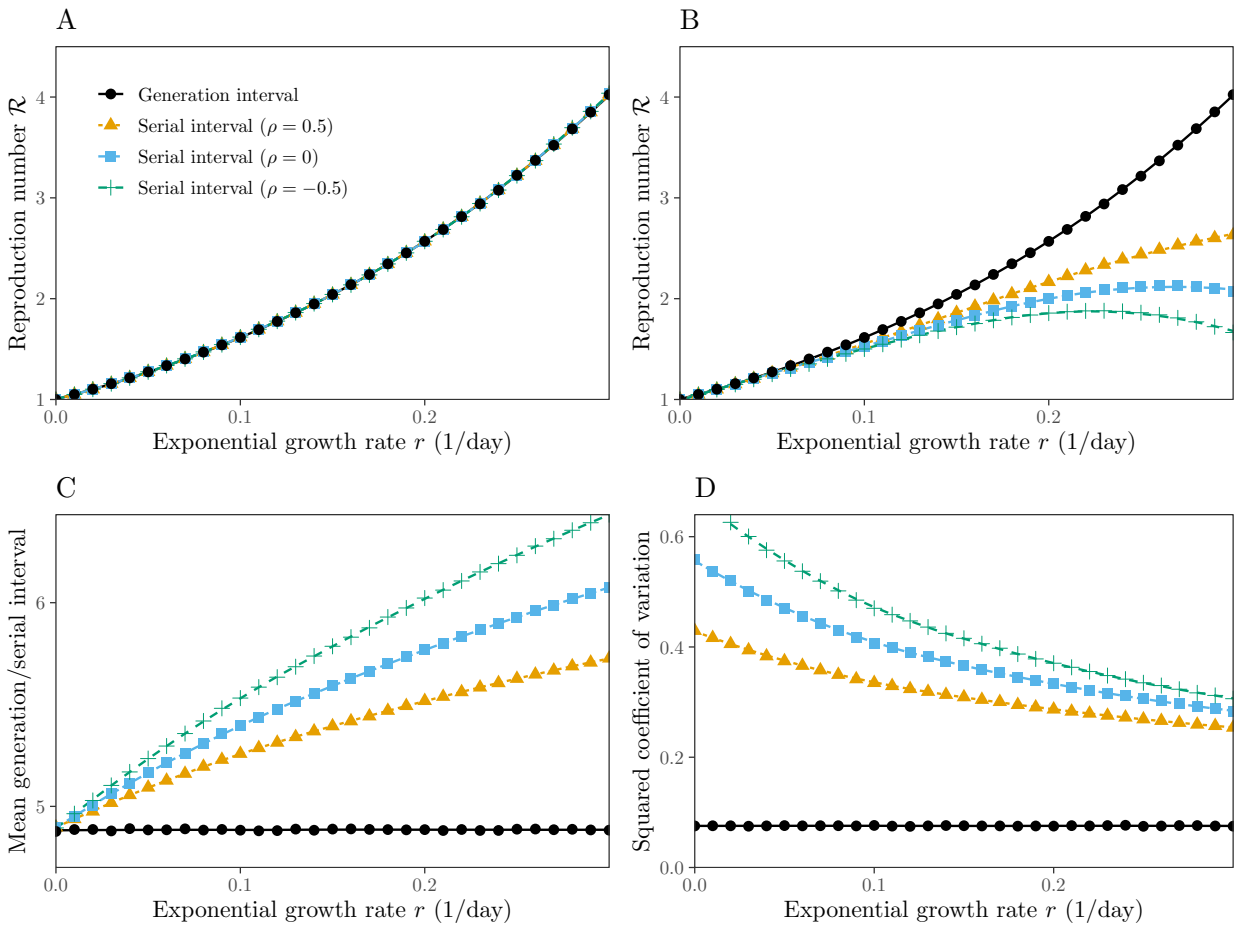

Figure 2: **Estimates of the reproduction number from the exponential growth rate.**

Fig. 2 compares the estimates of the reproduction number $\mathcal{R}$ from the exponential growth rate r using forward serial-interval distributions and naive serial-interval distributions that do not account for the differences between backward and forward incubation periods. When we use the forward serial-interval distributions to estimate $\mathcal{R}$, we obtain identical values

regardless of the correlation between incubation periods and generation intervals; as expected, these estimates match $\mathcal{R}$ estimes using the intrinsic generation-interval distributions (Fig. 2A). On the other hand, when naive serial-interval distributions are used, we underestimate $\mathcal{R}$; $\mathcal{R}$ even decreases for high r because negative serial intervals are given too much weight (Fig. 2B). While the forward serial intervals can be also negative, the proportion of negative intervals are appropriately balances because faster epidemic growth will lead to shorter backward incubation period (and therefore less negative serial intervals).

Comapring the shapes of forward serial-interval distributions and the intrinsic generation-interval distribution allow us to better understand how they are able to give identical estimates of $\mathcal{R}$. Generation-interval distributions with higher means and less variability are generally expected to give higher $\mathcal{R}$ for a given r. Here, we find that the forward serial intervals 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, the estimates of $\mathcal{R}$ remain unchanged.

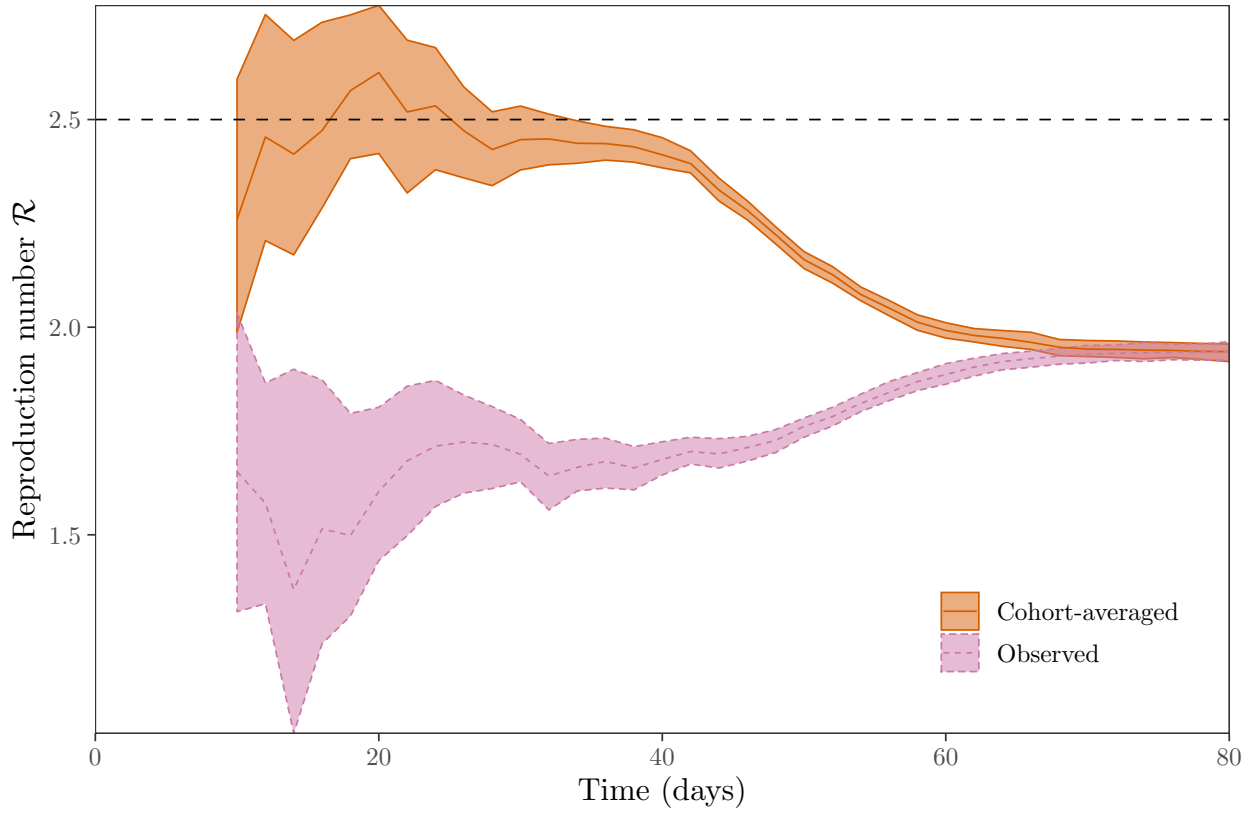

Figure 3: **Estimates of the reproduction number from the observed serial intervals.**

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. 3 demonstrates how the effect of right-censoring in the observed serial intervals translates to the underestimation of $\mathcal{R}$. However, even if we can observe *all* serial intervals after the epidemic has ended, we still underestimate $\mathcal{R}$ by a large amount because it 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 forward serial-interval distribution during the exponential growth phase, which provides the correct linke between r and $\mathcal{R}$.

Here, we provide a simple, heuristic way of assessing potential biases in the estimate of $\mathcal{R}$ retrospctively. Once a set of serial intervals has been observed from the line list, we can group observed serial intervals by their primary cohort times. Then, we can compare how estimates of $\mathcal{R}$ change as we include more cohorts into the analysis (see 'cohort-averaged' in Fig. 3). During the exponential growth phase, the estimates of $\mathcal{R}$ remains consistent; adding more data allows us to make more precise inference about $\mathcal{R}$ during this period. However, the cohort-averaged estimates of $\mathcal{R}$ decreases rapidly soon after the exponential growth period. Confidence intervals associated with the cohort-averaged estimates of $\mathcal{R}$ becomes narrower and do not overlap with earlier estimates. This approach allows us to visualize how changes in the forward serial intervals translate to biases in estimates of $\mathcal{R}$.

4 Discussion

Appendix

Recall that the forward serial interval can be written as:

$$-x_0 + \sigma + x_1 \tag{13}$$

Note that x_1 is independent of x_0 and σ. Then, we get:

$$M_{-x_0+\sigma+x_1}(-r) = M_{-x_0+\sigma}(-r)M_{x_1}(-r). \tag{14}$$

We want to show that

$$M_{-x_0+\sigma}(-r) = \frac{M_\sigma(-r)}{M_{x_1}(-r)} \tag{15}$$

for $r \geq 0$. Note that the time between symptom onset and infection of an infectee follows the following distribution:

$$z(\tau) \propto \int_{\max(0,-\tau)}^{\infty} i(\ell - x_0)h_{\ell-x_0}(x_0, \tau + x_0)\mathrm{d}x_0. \tag{16}$$

During the exponential growth phase, we have

$$z_{\exp}(\tau) = \frac{1}{N}\int_{\max(0,-\tau)}^{\infty} \exp(-rx_0)h(x_0, \tau + x_0)\mathrm{d}x_0 \tag{17}$$

where N is the normalization factor:

$$\begin{aligned} N &= \int_{-\infty}^{\infty} \int_{\max(0,-\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) \mathrm{d}x_0 \, \mathrm{d}\tau \\ &= \int_{0}^{\infty} \int_{-x_0}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) \mathrm{d}\tau \, \mathrm{d}x_0 \\ &= \int_{0}^{\infty} \exp(-rx_0) k(x_0) dx_0 \\ &= M_{x_1}(-r) \end{aligned} \tag{18}$$

Therefore,

$$\begin{aligned} &\int_{-\infty}^{\infty} \exp(-r\tau) z_{\exp}(\tau) \mathrm{d}\tau \\ &= \frac{1}{M_{-x_1}(-r)} \int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0,-\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) \mathrm{d}x_0 \, \mathrm{d}\tau \end{aligned} \tag{19}$$

Finally, we are left to prove that

$$\int_{0}^{\infty} \exp(-r\tau) g(\tau) \mathrm{d}\tau = \int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0,-\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) \mathrm{d}x_0 \, \mathrm{d}\tau \tag{20}$$

where g is a marginal probability distribution of h describing the forward generation intervals:

$$g(\tau) = \int_{0}^{\infty} h(x_0, \tau) \mathrm{d}x_0. \tag{21}$$

Let $a = x_0 + \tau$. Then, by change of variables, it immediately follows that

$$\begin{aligned} &\int_{-\infty}^{\infty} \exp(-r\tau) \int_{\max(0,-\tau)}^{\infty} \exp(-rx_0) h(x_0, \tau + x_0) \mathrm{d}x_0 \, \mathrm{d}\tau \\ &= \int_{0}^{\infty} \int_{0}^{\infty} \exp(-ra) h(x_0, a) \mathrm{d}x_0 \, \mathrm{d}a \\ &= \int_{0}^{\infty} \exp(-r\tau) g(\tau) \mathrm{d}\tau \end{aligned} \tag{22}$$

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

References

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.

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