Cohort-based approach to understanding the roles of generation and serial intervals in shaping epidemiological dynamics

Sang Woo Park[1,*] Kaiyuan Sun[2] David Champredon[3] Michael Li[4] Benjamin M. Bolker[4,5,6] David J. D. Earn[5,6] Joshua S. Weitz[7, 8] Bryan T. Grenfell[1,2,9] Jonathan Dushoff[4,5,6]

1 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA
2 Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
3 Department of Pathology and Laboratory Medicine, University of Western Ontario, London, Ontario, Canada
4 Department of Biology, McMaster University, Hamilton, ON, Canada
5 Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada
6 M. G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON, Canada
7 School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
8 School of Physics, Georgia Institute of Technology, Atlanta, GA, USA
9 Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA

*Corresponding author: swp2@princeton.edu

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the U.S. National Institutes of Health or Department of Health and Human Services.

Abstract

Generation intervals and serial intervals are critical quantities for characterizing outbreak dynamics. Generation intervals characterize the time between infection and transmission, while serial intervals characterize the time between the onset of symptoms in a chain of transmission. They are often used interchangeably, leading to misunderstanding of how these intervals link the epidemic growth rate r and the reproduction number $\mathcal{R}$. Generation intervals provide a mechanistic link between r and $\mathcal{R}$ but are harder to measure via contact tracing. While serial intervals are easier to measure from contact tracing, recent studies suggest that the two intervals give different estimates of $\mathcal{R}$ from r. We present a general framework for characterizing epidemiological delays based on cohorts (i.e., a group of individuals that share the same event time, such as symptom onset) and show that *forward-looking* serial intervals, which correctly link $\mathcal{R}$ with r, are not the same as "intrinsic" serial intervals, but instead change with r. We provide a heuristic method for addressing potential biases that can arise from not accounting for changes in serial intervals across cohorts and apply the method to estimating $\mathcal{R}$ for the COVID-19 outbreak in China using serial-interval data — our analysis shows that using incorrectly defined serial intervals can severely bias estimates. This study demonstrates the importance of early epidemiological investigation through contact tracing and provides a rationale for reassessing generation intervals, serial intervals, and $\mathcal{R}$ estimates, for COVID-19.

Significance Statement

The generation- and serial-interval distributions are key, but different, quantities in outbreak analyses. Recent theoretical studies suggest that two distributions give different estimates of the reproduction number $\mathcal{R}$ from the exponential growth rate r; however, both intervals, by definition, describe disease transmission at the individual level. Here, we show that the serial-interval distribution, defined from the correct reference time and cohort, gives the same estimate of $\mathcal{R}$ as the generation-interval distribution. We then apply our framework to serial-interval data from the COVID-19 outbreak in China. While our study supports the use of serial-interval distributions in estimating $\mathcal{R}$, it also reveals necessary changes to the current understanding and applications of serial-interval distribution.

1 Introduction

The reproduction number $\mathcal{R}$ is one of the most important characteristics of an emerging epidemic, including the current pandemic of coronavirus disease 2019 (COVID-19) (Majumder and Mandl, 2020). The reproduction number is defined as the average number of secondary cases caused by a primary case; the value in a fully susceptible population — the "basic" reproduction number $\mathcal{R}_0$— allows us to predict the extent to which an infection will spread in the population, and the amount of intervention necessary to eliminate it (Anderson and May, 1991). Near the beginning of an outbreak, $\mathcal{R}$ is often estimated from the observed exponential growth rate using generation- and serial-interval distributions (e.g., Wallinga and Lipsitch (2007); Fraser et al. (2009); Hampson et al. (2009); Chunara et al. (2012); Chowell et al. (2014); Du et al. (2020); Jung et al. (2020)) but the differences among two distributions are often neglected.

The generation interval is defined as the time between when an individual (infector) is infected and when that individual infects another person (infectee); the generation-interval distribution determines the relationship between the exponential growth rate r and the reproduction number $\mathcal{R}$ (Anderson and May, 1991; Ferguson et al., 2005; 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).

Although these distributions were clearly and distinctly defined 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 make inferences about COVID-19, either without making a clear distinction (Abbott et al., 2020; Du et al., 2020; He et al., 2020; Wu et al., 2020; Zhao et al., 2020), or explicitly conflating the two intervals (Anderson et al., 2020; Hellewell et al., 2020).

One source of confusion arises from an apparent discrepancy between the serial-interval and generation-interval viewpoints. 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. It is well established that this renewal formulation allows us to link the exponential growth rate of an epidemic r with its reproduction number $\mathcal{R}$ using the generation-interval distribution (Wallinga and Lipsitch, 2007). By definition, the serial-interval distribution describes the renewal process between symptomatic cases based on their symptom onset dates; since both renewal processes, based on generation- and serial-interval distributions, describe the same underlying system of exponential growth, both should provide the same correct link between the reproduction number $\mathcal{R}$ and the epidemic growth rate r. In contexts where the distributions are expected to be different, current theory has no explanation for how these differing distributions could provide identical estimates of $\mathcal{R}$

from r. Some studies have further suggested that using serial intervals can underestimate $\mathcal{R}$ because the serial-interval distribution is wider than the generation-interval distribution (Britton and Scalia Tomba, 2019; Ganyani et al., 2020).

Here, we resolve this apparent discrepancy by showing that the relevant interval for the renewal framework is what is called the "forward" interval, and that the forward serial intervals are different from the "intrinsic" serial intervals that previous studies have relied on (Svensson, 2007; Klinkenberg and Nishiura, 2011; te Beest et al., 2013; Champredon et al., 2018; Britton and Scalia Tomba, 2019). We develop a new framework for characterizing and comparing serial intervals, as well as any other epidemiological delays, and show that the initial forward serial-interval distribution correctly estimates $\mathcal{R}$ from r. Conversely, using inaccurately defined serial intervals or failing to account for changes in the observed serial-interval distributions over the course of an epidemic can considerably bias estimates of $\mathcal{R}$. 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 to correctly infer the reproduction number during the early stages of an outbreak.

2 Methods

2.1 Intrinsic, forward, and backward delay distributions

The time delays between two epidemiological events can be defined either within an infected individual (e.g., incubation period: infection and symptom onset of an individual) or between infected individuals (e.g., serial interval: symptom onsets of an infector and an infectee). 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 usually occurs before the secondary event — most epidemiological events that can be observed within an individual have clear direction (e.g., infection to 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 in the infector. Again, some of these delays are always positive (the infector is always infected before the infectee) and some are not (it is possible for the infectee to develop symptoms before the infector (He et al., 2020)).

At the individual level, we can define the time distribution between a primary and a secondary event that we expect to observe (averaged across individual characteristics) for an infected individual — we refer to this distribution as the intrinsic distribution. For example, the intrinsic incubation period distribution describes the expected time distribution from infection to symptom onset of an infected individual. Likewise, the intrinsic generation-interval distribution describes the expected time distribution of infectious contacts made by an infected individual. However, the intrinsic time distributions are not always equivalent to the corresponding realized time distributions at the population level (i.e., the distribution of time between actual primary and secondary events that occur during an epidemic). For example, an infectious contact results in infection only if the contacted individual is susceptible

(and has not already been infected); this is one mechanism that causes realized generation intervals (time between actual infection events) to differ from from the intrinsic generation intervals (time between infection and infectious contacts).

At the population level, we model realized time delays between a primary and a secondary event from a cohort perspective. A cohort consists of *all* individuals whose (primary or secondary) event occurred at a given time. For example, when we are measuring incubation periods, a primary cohort consists of all individuals who became infected at time p, while a secondary cohort consists of all individuals whose symptom onset occurred at time s. Similarly, when we are measuring serial intervals, a primary cohort consists of all infectors who became symptomatic at time p. Then, for primary cohort at time p, 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 $f_p(\tau)$.

Likewise, we define the backward delay distribution $b_s(\tau)$ for a secondary cohort at time s: The backward delay distribution describes the time delays between a primary and secondary host given that the secondary event occurred at time s. For example, the backward incubation period distribution at time s describes incubation periods for a *cohort* of individuals who became symptomatic at time s. Likewise, the backward serial-interval distribution at time s describes serial intervals for a *cohort* of infectees who became symptomatic at time s.

Both forward and backward perspectives must yield identical *measurement* (e.g., the length of the incubation period of a given individual is the same whether measured forward from the time of infection or backward from the time of symptom onset). Consequently, no matter how delays are distributed, if $\mathcal{P}$ and $\mathcal{S}$ represent the sizes of primary and secondary cohorts then we can express the total density of intervals between time p and s as follows:

$$W(p)\mathcal{P}(p)f_p(\tau) = \mathcal{S}(s)b_s(\tau)\,, \tag{1}$$

where $W(p)$, the "weight" of the primary cohort, represents the average number of forward intervals that an individual in cohort $\mathcal{P}(p)$ produces over the course of their infection. When we measure within-individual delays, we expect $W(p) \leq 1$ because only a subset of individuals who experience the primary event (e.g., infection) will eventually experience the secondary event (e.g., symptom onset). For between-individual delays, we expect $W(p)$ to change throughout an epidemic, because individuals infected earlier in an epidemic will infect more individuals on average than those infected later.

Substituting $p = s - \tau$, it follows that

$$b_s(\tau) = \frac{W(s-\tau)\mathcal{P}(s-\tau)f_{s-\tau}(\tau)}{\mathcal{S}(s)}\,. \tag{2}$$

If we are considering incubation periods, the left hand side of this equation is the probability density that an individual who became symptomatic at time s had an incubation period of length τ. From the right hand side, we see that this probability density depends on the weight parameter $W(p-\tau)$ (in this case, the proportion of symptomatic infection), the time-varying primary cohort size at the earlier time $\mathcal{P}(s-\tau)$ (in this case, the number of

individuals infected at time $s - \tau$), and the forward delay distribution $f_{s-\tau}(\tau)$ (in this case, the probability density that an incubation period that starts at time $s - \tau$ ends at time s).

Several different mechanisms drive the changes in forward and backward delay distributions over time. Typically, within-individual forward delay distributions are not directly affected by epidemic dynamics. Some realized distributions, like incubation period distributions, are expected to be equivalent to their intrinsic distributions and remain invariant at the time scale of an outbreak. Other realized distributions, like the distribution of time from symptom onset to testing, may change over the course of an epidemic due to changes in public health policies. Between-individual forward delay distributions, such as generation- or serial-interval distributions, depend on epidemic dynamics; for example, forward generation intervals contract during the outbreak because infected individuals are less likely to infect others later in the epidemic due to factors that drive time-dependent decreases in transmission such as susceptible depletion, behavioral change, and interventions (Champredon and Dushoff, 2015).

Eq. (2) suggests that backward delay distributions change over time even if their corresponding forward delay distribution does not change. Backward delay distributions depend on changes in the primary cohort size over time due to conditionality of observations: Conditioning on individuals whose secondary events have occurred at the same time means that we tend to observe shorter (or longer) inter-event delays when cohort size has been increasing (decreasing) through time. When incidence is growing exponentially, we can calculate the amount of bias exactly. Assuming that the forward delay distribution ($f_p(\tau) \approx f_0(\tau)$) and the weight parameter ($W(p) \approx W(0)$) remain constant during the exponential growth phase, we can substitute $\mathcal{P}(t) = \mathcal{P}(0)\exp(rt)$ in Eq. (2) to obtain:

$$b_0(\tau) = [W(0)\mathcal{P}(0)/\mathcal{S}(0)]\exp(-r\tau)f_0(\tau) \tag{3}$$

where r is the exponential growth rate (and $[W(0)\mathcal{P}(0)/\mathcal{S}(0)]^{-1} = \int_{-\infty}^{\infty}\exp(-r\tau')f_0(\tau')\,\mathrm{d}\tau'$ since b_0 is a probability distribution). Thus, for faster epidemics (higher r), the backward delay distribution will have a shorter mean. In general, the backward delay distribution will differ from the forward delay distribution (unless the disease is at equilibrium), even if we are measuring time delays that are intrinsic to the life history of a disease (e.g., the incubation period). 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 realized generation intervals from the perspective of an infector and an infectee, respectively, as well as the work by Britton and Scalia Tomba (2019) who showed that Eq. (3) holds for the backward generation-interval distributions.

2.2 Realized serial-interval distributions

The serial interval is defined as the time between when an infector becomes symptomatic and when an infectee becomes symptomatic (Svensson, 2007). Previous studies have often expressed serial intervals τ_s in the form (Fig. 1A):

$$\tau_\mathrm{s} = (\tau_\mathrm{g} + \tau_\mathrm{i2}) - \tau_\mathrm{i1} \tag{4}$$

where $\tau_{\mathrm{i}1}$ and $\tau_{\mathrm{i}2}$ represent incubation periods of an infector and an infectee, respectively, and τ_{g} represents the generation interval between the infector and the infectee. These studies assume that $\tau_{\mathrm{i}1}$ and $\tau_{\mathrm{i}2}$ are drawn from 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); however, this conclusion is based on the implicit assumption that distributions of realized incubation periods, $\tau_{\mathrm{i}1}$ and $\tau_{\mathrm{i}2}$, as well as generation interval, τ_{g}, are intrinsic to individuals (and not dependent on epidemic dynamics at the population-level) — something that is generally true of forward but not backward incubation-period distributions. We refer to the definition Eq. (4) as the intrinsic serial interval (Fig. 1A).

To correctly link the realized serial-interval distribution to the renewal process between cases based on symptom onset dates, we must use the forward serial interval (i.e., use the perspective of a cohort of infectors that share the same symptom onset time). Given that an infector became symptomatic at time p, to calculate the forward serial interval we first go backward in time to when the infector was infected, and then forward in time to when the infectee was infected, and then forward to when the infectee became symptomatic. In Fig. 1B, we see that $\tau_{\mathrm{i}1}$ is drawn from the backward incubation period distribution of the cohort of infectors who became symptomatic at time p; τ_{g} is drawn from the forward generation-interval distribution of the cohort of infectors who became infected at time $p - \tau_{\mathrm{i}1}$; and $\tau_{\mathrm{i}2}$ is drawn from the forward incubation period distribution of the cohort of infectees who became infected at time $p - \tau_{\mathrm{i}1} + \tau_{\mathrm{g}}$. Likewise, we can define the backward serial-interval distribution for a cohort of infectees who became symptomatic at time s (Fig. 1C). This conceptual framework demonstrates that the distributions of $\tau_{\mathrm{i}1}$, τ_{g}, and $\tau_{\mathrm{i}2}$ (and therefore the distributions of realized serial intervals) depend on the reference cohort, which is defined by an event type (primary or secondary), temporal direction (forward or backward), and a particular reference time.

To calculate realized serial-interval distributions, we begin by modeling $\mathcal{T}(p, s)$: the total density of serial intervals that start (when infectors develop symptoms) at time p and end (when infectees develop symptoms) at time s. The density of serial intervals between time p and s, given that the infectors became infected at time $\alpha_1 \leq p$ and the infectees became infected at time $\alpha_2 \leq s$, depends on the amount of infection that occurs between time α_1 and α_2 as well as the density of forward incubation periods between α_1 and p (realized incubation periods of infectors) and between α_2 and s (realized incubation periods of infectees):

$$\underbrace{\mathcal{R}_{\mathrm{c}}(\alpha_1)}_{\substack{\text{case}\\ \text{reproduction}\\ \text{number}}} \times \underbrace{i(\alpha_1)}_{\text{incidence}} \times \underbrace{h_{\alpha_1}(p-\alpha_1, \alpha_2-\alpha_1)}_{\substack{\text{joint density of}\\ \text{the forward incubation}\\ \text{period } p-\alpha_1 \text{ and the forward}\\ \text{generation interval } \alpha_2-\alpha_1\\ \text{(of infectors)}}} \times \underbrace{\ell_{\alpha_2}(s-\alpha_2)}_{\substack{\text{marginal density of}\\ \text{the forward incubation}\\ \text{period } s-\alpha_2\\ \text{(of infectees)}}} \quad , \tag{5}$$

where the case reproduction number $\mathcal{R}_{\mathrm{c}}(\alpha_1)$ is defined as the average number of secondary cases caused by a primary case infected at time α_1 over the course of their infection (Fraser, 2007). We describe the forward incubation periods and the forward generation intervals using a joint probability distribution because onset of symptoms and transmission potential jointly depend on the life history of a disease; for example, if an infected individual can only

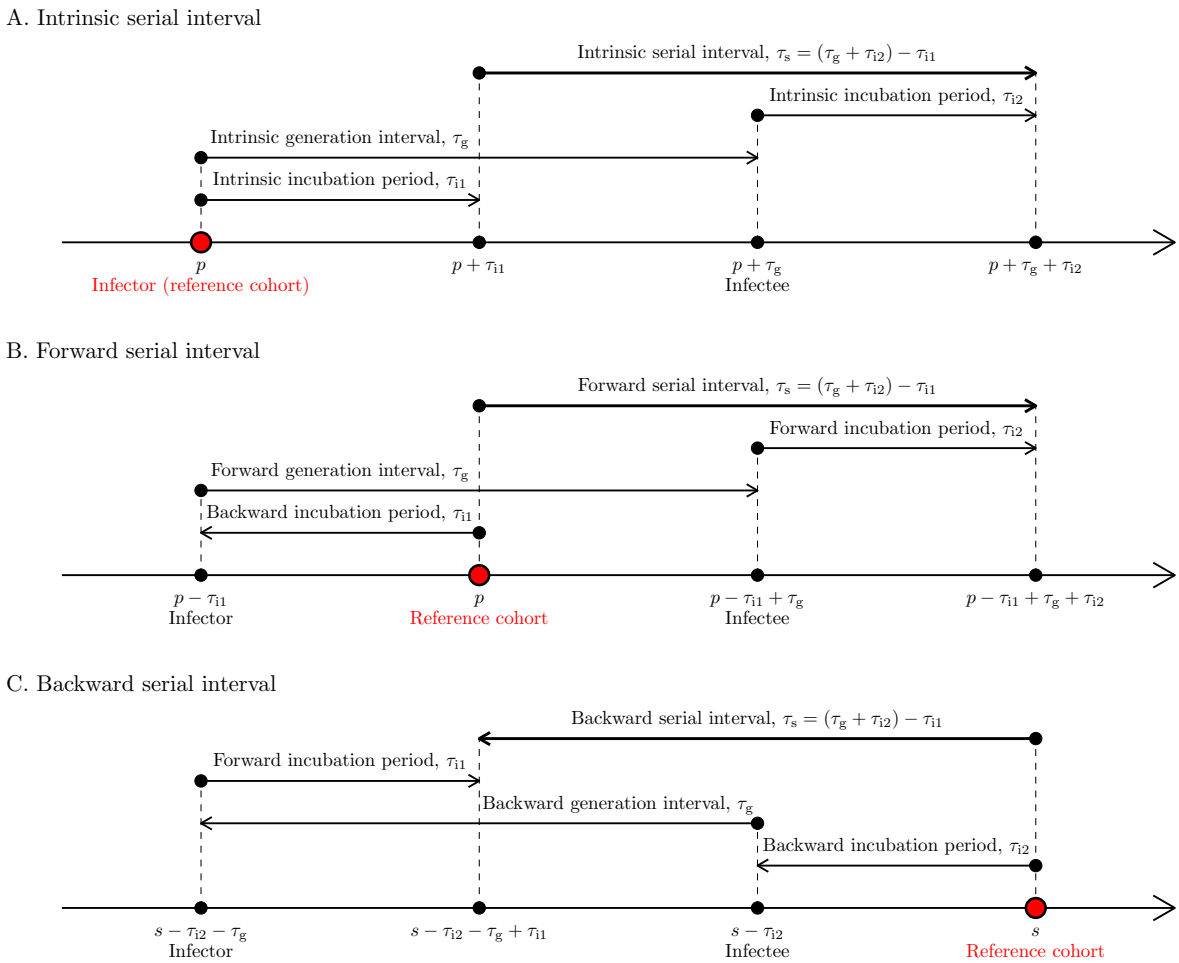


Figure 1: **Illustration of intrinsic, forward and backward serial intervals.** (A) The intrinsic serial interval for a cohort of individuals infected at time p. In this case, τ_{i1} is drawn from the intrinsic incubation period distribution; τ_g is drawn from the intrinsic generation-interval distribution; and τ_{i2} is drawn from the intrinsic incubation period distribution. (B) The forward serial interval for a cohort of infectors who became symptomatic at time p. In this case, τ_{i1} is drawn from the backward incubation period distribution; τ_g is drawn from the forward generation-interval distribution; and τ_{i2} is drawn from the forward incubation period distribution. (C) The backward serial interval for a cohort of infectees who became symptomatic at time s. In this case, τ_{i1} is drawn from the forward incubation period distribution; τ_g is drawn from the backward generation-interval distribution; and τ_{i2} is drawn from the backward incubation period distribution.

transmit the disease after symptom onset, the forward generation interval will necessarily be longer than the forward incubation period.

The total amount of serial intervals can now be obtained by integrating over all possible

infection times for the infector and the infectee:

$$\mathcal{T}(p,s) = \int_{-\infty}^{p}\int_{\alpha_1}^{s} \mathcal{R}_c(\alpha_1) i(\alpha_1) h_{\alpha_1}(p-\alpha_1, \alpha_2-\alpha_1)\ell_{\alpha_2}(s-\alpha_2)\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_1. \tag{6}$$

Then, the forward serial-interval distribution $f_p(\tau)$ is given by the density of intervals of length τ starting at p, relative to the total number of serial intervals starting at p:

$$f_p(\tau) = \frac{\mathcal{T}(p, p+\tau)}{\int_{-\infty}^{\infty}\mathcal{T}(p,p+\tau')\,\mathrm{d}\tau'}. \tag{7}$$

Likewise, the backward serial-interval distribution $b_s(\tau)$ is given by the density of intervals of length τ ending at s, relative to the total number of serial intervals ending at s:

$$b_s(\tau) = \frac{\mathcal{T}(s-\tau, s)}{\int_{-\infty}^{\infty}\mathcal{T}(s-\tau',s)\,\mathrm{d}\tau'}. \tag{8}$$

This framework allows us to understand changes in the realized serial intervals for any deterministic epidemic model.

2.3 Epidemic model

We illustrate changes in the forward and backward serial intervals over the course of an epidemic 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:

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= -i(t) \\ i(t) &= \mathcal{R}_0 S(t)\int_0^{\infty} i(t-\tau)g(\tau)\,\mathrm{d}\tau, \end{aligned} \tag{9}$$

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 population where the number of susceptibles remains constant). Then, the forward generation-interval for a cohort of individuals that were infected at time p follows (Champredon and Dushoff, 2015):

$$g_p(\tau) = \frac{g(\tau)S(p+\tau)}{\int_0^{\infty} g(\tau')S(p+\tau')\,\mathrm{d}\tau'}, \tag{10}$$

which allows us to separate the joint probability distribution h_p 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:

$$h_p(x,\tau) = \frac{h(x,\tau)S(p+\tau)}{\int_0^{\infty}\int_0^{\infty} h(x',\tau')S(p+\tau')\,\mathrm{d}\tau'\,\mathrm{d}x'}. \tag{11}$$

We further assume that the forward incubation period distribution does not vary across cohorts over the course of an epidemic, as it represents the life history of a disease; we denote it as ℓ. Then, we have:

$$\begin{aligned} \ell(x) &= \int_0^\infty h(x,\tau)\,\mathrm{d}\tau \\ g(\tau) &= \int_0^\infty h(x,\tau)\,\mathrm{d}x \end{aligned} \tag{12}$$

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

$$\mathcal{R}_\mathrm{c}(t) = \mathcal{R}_0 \int_0^\infty g(\tau) S(t+\tau)\,\mathrm{d}\tau. \tag{13}$$

The forward and backward serial-interval distributions are then calculated by substituting these quantities into Eq. (7) and Eq. (8). We use this framework to illustrate how the realized epidemiological time distributions vary over the course of an epidemic and depend on the perspective (i.e., forward vs. backward).

2.4 Linking r and $\mathcal{R}$

During the initial phase of an epidemic, the proportion susceptible remains approximately constant ($S(t) \approx S(0)$) and incidence of infection grows exponentially: $i(t) \approx i_0 \exp(rt)$. During this period, the intrinsic generation-interval distribution provides the correct link between the exponential growth rate r and the initial reproduction number $\mathcal{R} = \mathcal{R}_0 S(0)$ based on the Euler-Lotka equation (Wallinga and Lipsitch, 2007). Here, we focus on the estimates of the basic reproduction number $\mathcal{R}_0$ (the value of $\mathcal{R}$ in a fully susceptible population, $S(t) \approx 1$):

$$\frac{1}{\mathcal{R}_0} = \int_0^\infty \exp(-r\tau) g(\tau)\,\mathrm{d}\tau. \tag{14}$$

Analogous to the intrinsic generation-interval distribution, forward serial-interval distributions describe the renewal process of infection based on symptom onset dates. 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 estimate the same value of $\mathcal{R}_0$ for a given r as the intrinsic generation-interval distribution (note, however, that the forward serial interval is not necessarily positive):

$$\frac{1}{\mathcal{R}_0} = \int_{-\infty}^\infty \exp(-r\tau) f_0(\tau)\mathrm{d}\tau, \tag{15}$$

where the initial forward serial-interval distribution is given by:

$$f_0(\tau) = \frac{1}{\phi} \int_{-\infty}^0 \int_{\alpha_1}^\tau \exp(r\alpha_1) h(-\alpha_1, \alpha_2 - \alpha_1) \ell(\tau - \alpha_2)\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_1, \tag{16}$$

where the normalization constant ϕ is determined by the requirement that $\int_{-\infty}^{\infty} f_0(\tau)\ \mathrm{d}\tau = 1$. In Supplementary Materials, we provide a mathematical proof of this relationship. We further compare this with the estimate of $\mathcal{R}_0$ based on the intrinsic serial-interval distribution $q(\tau)$:

$$\frac{1}{\mathcal{R}_{\text{intrinsic}}} = \int_{-\infty}^{\infty} \exp(-r\tau) q(\tau) \mathrm{d}\tau. \tag{17}$$

Here, the intrinsic serial-interval distribution $q(\tau)$ does not depend on epidemic dynamics:

$$q(\tau) = \frac{1}{\phi_q} \int_{-\infty}^{0} \int_{\alpha_1}^{\tau} h(-\alpha_1, \alpha_2 - \alpha_1) \ell(\tau - \alpha_2)\, \mathrm{d}\alpha_2\, \mathrm{d}\alpha_1, \tag{18}$$

where the normalization constant ϕ_q is determined by the requirement that $\int_{-\infty}^{\infty} q(\tau)\ \mathrm{d}\tau = 1$. Rather than numerically integrating over closed forms of g, f_0, and q to estimate $\mathcal{R}_0$, we use simulation-based approaches for simplicity (Supplementary Materials).

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 (Eq. (3)), and therefore, the forward serial-interval distribution will generally have a larger mean than the intrinsic generation- and serial-interval distributions. The Susceptible-Exposed-Infected-Recovered model, under the additional assumption that the incubation and exposed periods are equivalent (i.e. that onset of symptoms and infectiousness occur simultaneously), provides a special case where the forward serial- and generation-intervals follow the same distributions during the exponential growth phase because (i) infected individuals can only transmit after symptom onset and (ii) the time between symptom onset and infection is independent of the incubation period of an infector (see Supplementary Materials).

2.5 Model parameterization

We parameterize our model based on parameter estimates for COVID-19 (Table 1). We use a bivariate lognormal distribution to model the joint probability distribution of the intrinsic incubation periods and the intrinsic generation intervals while allowing for the possibility that they might be correlated. For simplicity, we consider three values for the correlation coefficients (on the log scale) of the bivariate lognormal distribution: $\rho = -0.5, 0, 0.5$.

3 Results

We use parameter estimates for COVID-19 to characterize the degree to which the realized serial-interval distribution can change over the course of the epidemic and evaluate how different definitions of the serial-interval distribution can affect the Euler-Lotka estimates of $\mathcal{R}_0$. We further address how the observed serial intervals, measured through contact tracing, are affected by the right censoring during an ongoing epidemic and provide a heuristic method for addressing biases that can arise from using serial interval data to estimate $\mathcal{R}_0$. Finally, we analyze the serial interval data from the COVID-19 epidemic in China under our framework.

Parameter	Values	Source
Mean intrinsic incubation period	5.5 days	Lauer et al. (2020)
SD intrinsic incubation period	2.4 days	Lauer et al. (2020)
Mean intrinsic generation interval	5 days	Ferretti et al. (2020)
SD intrinsic generation interval	2 days	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 intrinsic 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 bivariate log-normal distribution with correlations (on the log scale) $\rho = -0.5, 0, 0.5$. The intrinsic generation-interval and incubation period distributions are chosen to match characteristic of COVID-19 to illustrate realistic magnitudes of time-varying/perspective effects in the current pandemic.

3.1 Realized serial-interval distributions

Fig. 2 shows Euler-Lotka estimates of $\mathcal{R}_0$ on different definitions of the serial interval. When the initial forward serial-interval distributions $f_0(\tau)$ is used, estimates (from Eq. (15)) exactly match the (correct) generation-interval-based estimates (Eq. (14)) for all values of the correlation ρ between the intrinsic incubation period and the intrinsic generation interval (Fig. 2A). When the intrinsic distributions are used, however, estimates based on the serial interval (Eq. (17)) underestimate $\mathcal{R}_0$: as r increases, $\mathcal{R}_{\text{intrinsic}}$ saturates and eventually *decreases* due to the increasing inferred importance of negative serial intervals (Fig. 2B). While the initial forward serial intervals during the exponential growth phase can also be negative, their effects are appropriately balanced because faster epidemic growth leads to longer serial intervals (and a corresponding lower proportion of negative intervals).

Comparing the shapes of the initial forward serial-interval distribution (Eq. (16)) and the intrinsic generation-interval distribution allows us to better understand how different forward distributions lead to identical estimates of $\mathcal{R}_0$. In general, distributions with higher means and less variability lead to higher $\mathcal{R}_0$ for a given r (Wallinga and Lipsitch, 2007; Weitz and Dushoff, 2015; Park et al., 2019). When incidence is growing exponentially, 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 increase $\mathcal{R}_0$) exactly cancel those of higher variability (which decrease $\mathcal{R}_0$). On the other hand, *intrinsic* serial intervals (Eq. (18)) have the same mean (equal to the mean initial forward serial at $r = 0$ in Fig. 2C) as the intrinsic generation intervals but are more variable (also see squared coefficient of variation of the initial forward serials at $r = 0$ in Fig. 2D); therefore, we underestimate $\mathcal{R}_0$ when we use the intrinsic serial-interval distribution.

The initial forward serial-interval distribution captures the exponential growth phase of an epidemic. We explore how forward and backward serial intervals can vary over the course of an epidemic using deterministic and stochastic simulations based on the renewal equations (see Supplementary Materials) using parameters in Table 1; we further assume

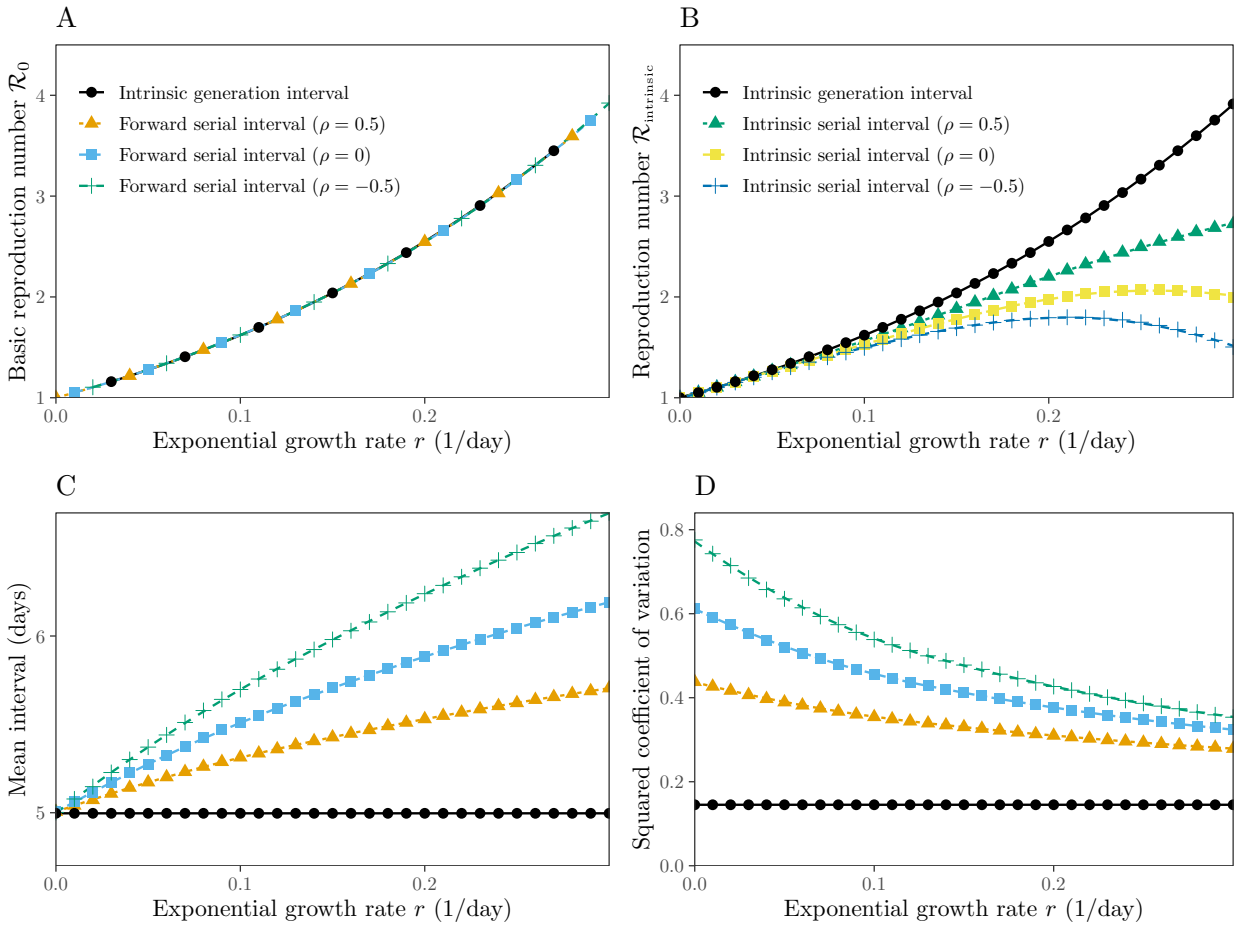


Figure 2: **Estimates of the reproduction number from the exponential growth rate based on serial- and generation-interval distributions.** (A). The initial forward serial-interval distributions give a correct link between the exponential growth rate r and the reproduction number $\mathcal{R}_0$, for any correlation ρ between intrinsic incubation period and intrinsic generation interval of the underlying bivariate log-normal distribution. (B) The intrinsic serial-interval distributions give an incorrect link between r and $\mathcal{R}_0$. (C) The mean initial forward serial interval during the exponential growth phase increases with r. (D) The squared coefficient of variation of the initial forward serial intervals during the exponential growth phase decreases with r.

$\mathcal{R}_0 = 2.5$. While the forward serial-interval distribution is our primary focus, understanding the differences between the forward and the backward distributions is important because the observed intervals during an ongoing epidemic are often the backward ones: we typically identify infected individuals and ask when and by whom they were infected. Similarly, when we are estimating the incubation period of an individual, we typically observe their symptom onset date and try to estimate when they were infected (e.g., Backer et al. (2020)).

Fig. 3 shows the epidemiological dynamics (A) together with the mean forward (B–D) and

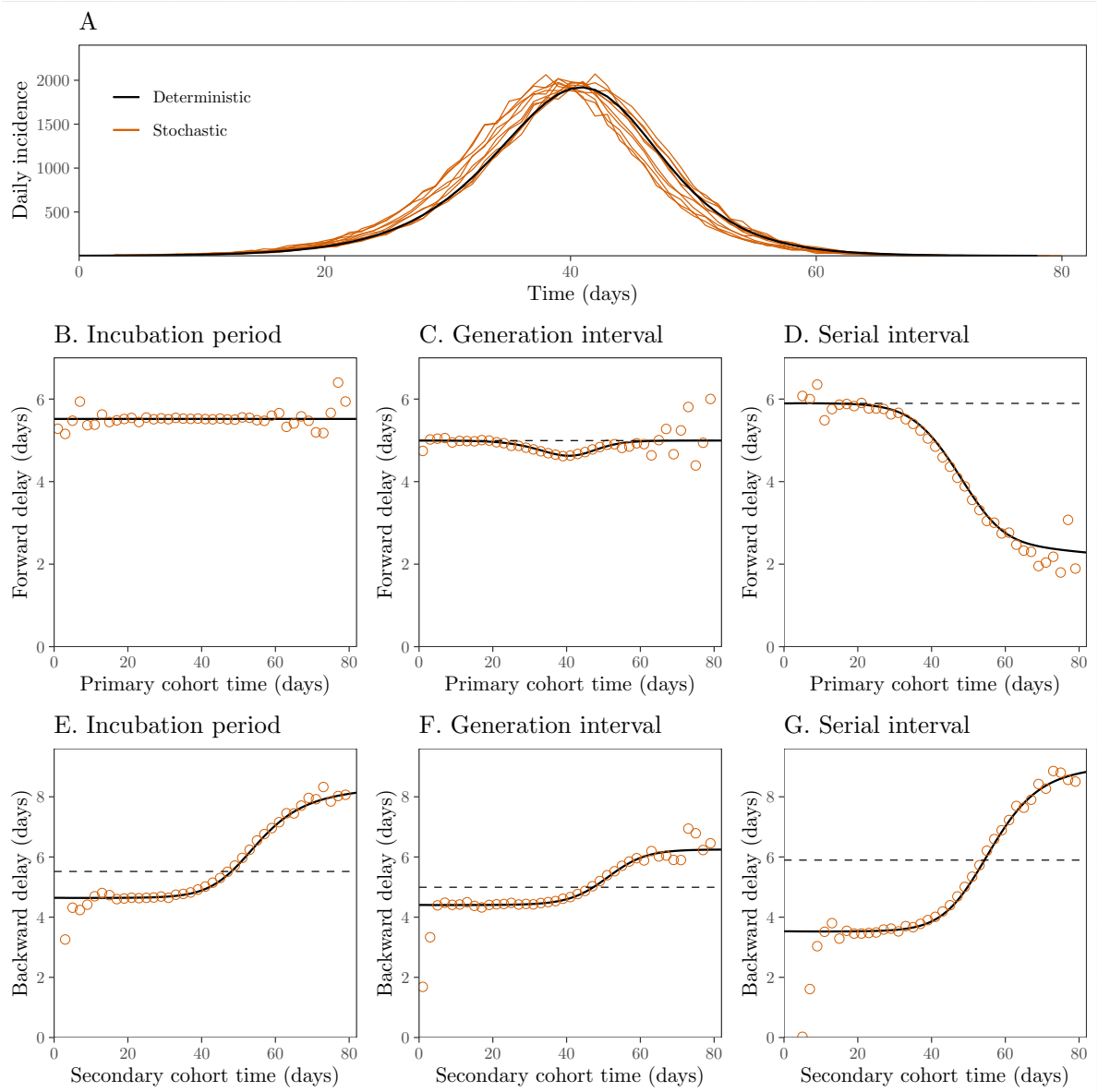


Figure 3: **Epidemiological dynamics and changes in mean forward and backward delay distributions.** (A) Daily incidence over time. (B–D) Changes in the mean forward incubation period, generation interval, and serial interval. (E–G) Changes in the mean backward incubation period, generation interval, and serial interval. Black lines represent the results of a deterministic simulation. Orange lines (A) represent the results of 10 stochastic simulations. Orange points (B-G) represent the average of 10 stochastic simulations. Dashed lines represent the mean initial forward delay. Intrinsic incubation periods and intrinsic generation-intervals are assumed to be independent of each other. See Table 1 for parameter values.

the mean backward (E–G) delay distributions of a deterministic model based on the renewal equation (Eq. (9)) and of the corresponding stochastic realizations based on individual-based simulations. The mean forward incubation period remains constant throughout an epidemic by assumption (Fig. 3B). The mean forward generation interval decreases slightly when incidence is high, which is when the susceptible population declines rapidly (Fig. 3C; Kenah et al. (2008); Champredon and Dushoff (2015)). In contrast, the mean forward serial interval decreases over time (Fig. 3D).

The forward serial-interval distributions depend on distributions of three intervals (Fig. 1): (i) the backward incubation period, (ii) the forward generation interval, and (iii) the forward incubation period. In these simulations, both forward incubation period (Fig. 3B) and generation-interval (Fig. 3C) distributions remain roughly constant; therefore, changes in the forward serial-interval distributions (Fig. 3D) are predominantly driven by changes in the backward incubation period distribution, whose mean increases over time as the growth rate of disease incidence slows, then becomes negative. In general, relative contributions of the three distributions depend on their shapes, correlations between intrinsic incubation periods and generation intervals, and overall epidemiological dynamics.

We see similar qualitative patterns in all three backward delays (Fig. 3E–G; Eq. (2)), because they are predominantly driven by the rate of change in incidence, which in turn affects relative cohort sizes. When incidence is increasing, individuals are more likely to have been infected recently and therefore, we are more likely to observe shorter intervals (Eq. (3)). Similarly, when incidence decreases, we are more likely to observe longer intervals. Neglecting these changes in the backward distributions will necessarily bias the inference of the intrinsic distributions from the observed distributions.

3.2 Observed serial-interval distributions

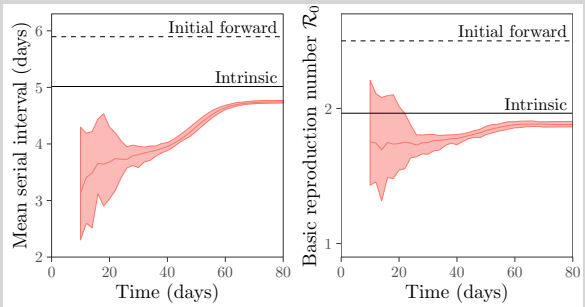
Now, we turn to issues of estimating the reproduction number from the observed serial-interval data during on ongoing epidemic. In order to have an unbiased estimate of the basic reproduction number, we need to estimate the initial forward serial-interval distribution — i.e., estimate based on cohorts of infectors who share the same symptom onset time, at the early stage of the epidemic. However, researchers typically use all available information to estimate epidemiological parameters; in particular, Thompson et al. (2019) recently suggested that up-to-date serial-interval data are necessary to accurately estimate the reproduction number. We explore the consequences of neglecting changes in the realized serial-interval distribution on estimates of the basic reproduction number.

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; for example, if we were to measure serial intervals on Day 8 as in Fig. 4A, we will only be able to observe the first 6 events (ID 1–6). Fig. 4B demonstrates how the effect of right-censoring in the observed serial intervals translates to the underestimation of the basic reproduction number $\mathcal{R}_0$ in our stochastic simulations (assuming $\mathcal{R}_0 = 2.5$ as in Fig. 3). Notably, even if we could observe, and average, *all* serial intervals across all transmission pairs after the epidemic has ended, we would still underestimate the

A. Observed serial intervals from line list

ID	Symptom onset date (Infector)	Symptom onset date (Infectee)	Serial interval (days)
1	Day 1	Day 4	3
2	Day 3	Day 6	3
3	Day 2	Day 7	5
4	Day 1	Day 7	6
5	Day 2	Day 8	6
6	Day 3	Day 8	5
7	Day 1	Day 10	9
8	Day 2	Day 10	8

B. Observed mean serial interval and basic reproduction number



C. Rearranged line list by symptom onset date of infectors

ID	Symptom onset date (Infector)	Symptom onset date (Infectee)	Serial interval (days)
1	Day 1	Day 4	3
4	Day 1	Day 7	6
7	Day 1	Day 10	9
3	Day 2	Day 7	5
5	Day 2	Day 8	6
8	Day 2	Day 10	8
2	Day 3	Day 6	3
6	Day 3	Day 8	5

D. Cohort-averaged mean serial interval and basic reproduction number

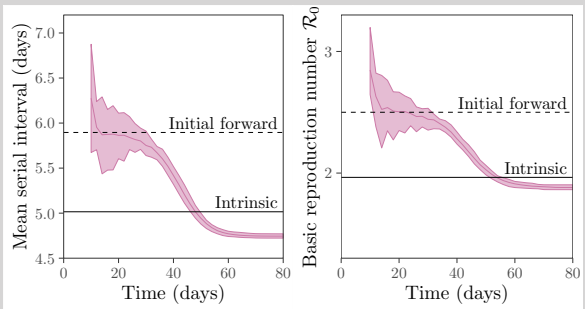


Figure 4: **Estimating the reproduction number from the observed serial intervals.** (A) Schematic representation of line list data collected during an epidemic. (B) Estimates of $\mathcal{R}_0$ based on all observed serial intervals completed by a given time. (C) Schematic representation of line list data rearranged by symptom onset date of infectors. (D) Estimates of $\mathcal{R}_0$ based on all observed serial intervals started by a given time. Black dashed lines represent the mean initial forward serial interval and $\mathcal{R}_0$. Black solid lines represent the mean intrinsic serial interval and $\mathcal{R}_{\text{intrinsic}}$. Colored solid lines represent the mean estimates of $\mathcal{R}_0$ across 10 stochastic simulations. Colored ribbons represent the range of estimates of $\mathcal{R}_0$ across 10 stochastic simulations.

initial mean forward serial interval (and therefore $\mathcal{R}_0$), likely by a large amount, because the observed serial-interval distribution converges to the intrinsic serial-interval distribution; in fact, we would even underestimate the intrinsic value slightly due to contraction of the forward generation-interval distribution during the susceptible depletion phase (Fig. 3C).

Here, we provide a heuristic way of assessing potential biases in the estimate of the mean initial forward serial interval and therefore $\mathcal{R}_0$ retrospectively. We can rearrange the line list and group observed serial intervals based on the symptom onset date of infectors

(Fig. 4C). Then, we can compare how estimates of the mean serial interval as well as $\mathcal{R}_0$ change as we incorporate more recent cohorts into the analysis; that is, we analyze observed serial intervals from infectors who became symptomatic before time t and evaluate how the estimates change as we increase t. This approach is analogous to averaging over a set of forward intervals, just as using all information up to a certain time is analogous to averaging over a set of backward intervals (Fig. 4D). During the exponential growth phase, the estimates of the mean serial interval and $\mathcal{R}_0$ are consistent with the true value (see 'initial forward' in Fig. 4B,D); adding more data allows us to make more precise inference during this period. However, the cohort-averaged estimates decrease rapidly soon after the exponential growth period, reflecting changes in the forward serial-interval distributions. This approach allows us to detect dynamical changes in the forward serial-interval distributions and their effect on the estimates of $\mathcal{R}_0$.

3.3 Applications to the COVID-19 pandemic

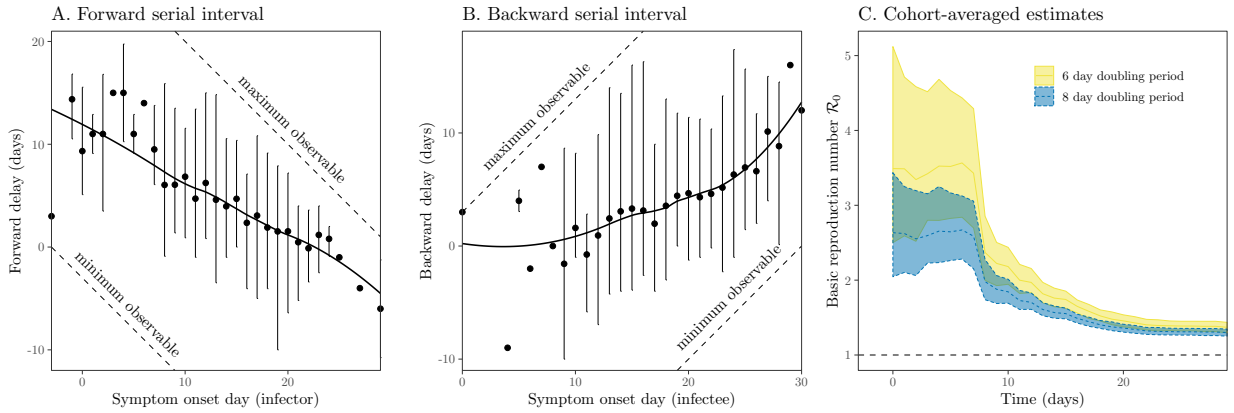


Figure 5: **Observed serial intervals of COVID-19 and cohort-averaged estimates of $\mathcal{R}$.** (A–B) forward and backward serial intervals over time. Points represent the means. Vertical error bars represent the 95% equi-tailed quantiles. Solid lines represent the estimated locally estimated scatterplot smoothing (LOESS) fits. The dashed line represents the maximum and minimum observable delays across the range of reported symptom onset dates. (C) Cohort-averaged estimates of $\mathcal{R}_0$ assuming doubling period of 6 and 8 days (Li et al., 2020; Wu et al., 2020). Ribbons represent the associated 95% bootstrap confidence intervals. The data were taken from Supplementary Materials of Du et al. (2020).

Finally, we re-analyze 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 interval 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 (indicated by the closeness between the quantiles of the observed serial intervals and maximum observable serial intervals), the increase in the proportion of negative

serial intervals indicates changes in the forward serial-interval distribution; this proportion is unlikely to be affected by left-censoring (based on the gap between the quantiles of the observed serial intervals and minimum observable serial intervals). The decrease in the mean forward serial interval was probably driven by interventions against spread. Interventions during this time period both decreased (and then reversed) the growth rate of COVID-19 cases — thus increasing the backward incubation period — and also reduced generation intervals, by preventing infections once cases were identified. Both of these would have acted to reduce the forward serial interval. Fig. 5B shows that the mean backward serial interval increased over time, likely driven directly by the decrease in COVID-19 cases.

While the qualitative changes in the mean forward and backward serial interval are consistent with our earlier simulations (Fig. 3), the initial mean forward serial interval (Fig. 5A) appears to be larger than what we calculated based on previously estimated incubation period and generation-interval distributions (Fig. 2C). These may imply that the incubation period and generation interval (Table 1) were underestimated, as neither study explicitly accounted for the fact that the observed intervals were drawn from the backward distributions and were likely to have been censored.

Fig. 5C shows the cohort-averaged estimates of $\mathcal{R}_0$, which remain roughly constant until day 7 and suddenly decreases; this sudden decrease is due changes in the forward serial intervals consistent with the dynamics seen in our simulations (Fig. 4). The cohort-averaged estimates of $\mathcal{R}_0$ based on the early forward serial intervals are also consistent with previous estimates of $\mathcal{R}_0$ of the COVID-19 epidemic in China (Majumder and Mandl, 2020; Park et al., 2020): $\mathcal{R}_0 = 2.6$ (95% CI: 2.2–3.1) and $\mathcal{R}_0 = 3.4$ (95% CI: $2.7 - 4.3$) based on 8 and 6 doubling periods, respectively, using serial-interval data from infectors who developed symptoms by day 7. These early cohort-averaged estimates of $\mathcal{R}_0$ are unlikely to be affected by the right-censoring as we expect the degree of right-censoring to be low (Fig. 5A). Therefore, the $\mathcal{R}_0$ estimate of 1.32 (95% CI 1.16–1.48), which neglects the changes in the forward serial-interval distribution, is likely to be an underestimate. This example demonstrates the danger of using the observed serial intervals to calculate the reproduction number without organizing serial intervals into cohorts.

4 Discussion

Generation and serial intervals determine the time scale of disease transmission, and are therefore critical to dynamical modeling of infectious outbreaks. Here, we show that the initial *forward* serial-interval distribution — measured during the exponential growth phase of an epidemic — provides the correct link between the exponential growth rate r and the reproduction number $\mathcal{R}$. In general, the forward serial-interval distributions will not match the intrinsic or the backward serial-interval distributions. In particular, the mean forward serial interval can decrease over time due to epidemic dynamics. Failing to account for these effects can result in underestimation of $\mathcal{R}$.

Our study underlines the importance of carefully defining measured epidemiological time distributions. Previous studies have shown that it matters whether generation intervals are

measured forward or backward (Nishiura, 2010; Champredon and Dushoff, 2015; Britton and Scalia Tomba, 2019); we generalize these ideas and show that they apply to other epidemiological distributions. Changes in the backward delay distributions due to changing cohort sizes are expected to be a pervasive feature of outbreak dynamics. Recent analyses of COVID-19 epidemics have attempted to reconstruct epidemic time series from by using time-independent delay distributions from infection or symptom onset to reporting (e.g., Abbott et al. (2020); Park et al. (2020); Sciré et al. (2020); Shim et al. (2020)); these studies effectively assume constant backward delay distributions. Although such approaches may be able to roughly match the time scale of an epidemic, we show here that they are subject to bias. Dynamical approaches that model explicitly model incidence as well as reporting delays (e.g., Flaxman et al. (2020)) can avoid this problem, but at the cost of more complexity in estimation. There is a need for simple methods that can complement more explicit approaches of addressing this problem.

While our results support the use of serial-interval distributions for calculating $\mathcal{R}$, they also reveal gaps in current practices in incorporating serial-interval distributions into outbreak analyses. Thompson et al. (2019) recently emphasized the importance of using up-to-date serial-interval data for accurate estimation of time-varying reproduction numbers. We show here, however, that if changes the forward serial interval through time are not accounted for, using up-to-date serial-interval data can, in fact, exacerbate the underestimation of initial $\mathcal{R}$. Future studies should explore how neglecting changes in the forward serial-interval distribution can affect the estimates of $\mathcal{R}$ beyond the exponential growth phase and potentially re-assess existing estimates of $\mathcal{R}$.

Many studies have already estimated the (time-varying) reproduction numbers for COVID-19 epidemics using aggregated serial-interval data (e.g., Abbott et al. (2020); Du et al. (2020); Pan et al. (2020); Zhang et al. (2020)); these studies should also re-assess whether they appropriately considered the *forward* serial interval. Forward serial-interval distributions are also likely to vary across countries. We suggest that modelers should aim to characterize spatiotemporal variation in forward serial-interval distributions and understand their changes over time. These modeling approaches should be coupled with epidemiological investigation through contact tracing. Going forward, an additional advantage of early, intensive contact tracing of emerging diseases is that this is the best way to characterize the initial forward serial-interval distribution.

Our study also underlines the fact that the serial-interval distribution depends not only on the generation-interval and incubation-period distributions, but also on their intrinsic correlations. This implies that not considering this correlation can bias the estimates of the intrinsic generation-interval distribution from the serial-interval distribution. Previous studies that tried to estimate the generation-interval distributions from the observed serial intervals often ignored the dynamical differences between the realized incubation periods of infectors (backward-looking) and those of infectees (forward looking) (e.g., Klinkenberg and Nishiura (2011); Ganyani et al. (2020)). Better statistical tools for teasing apart the intrinsic generation-interval distributions from the observed serial intervals are needed.

Here, we assume that all individuals develop symptoms and that the entire transmission process, including all relevant epidemiological delays, is known exactly. In practice, iden-

tifying who infected whom is difficult in general, and asymptomatic and presymptomatic transmission of COVID-19 exacerbates this difficulty (Bai et al., 2020; He et al., 2020; Wei, 2020). Biases in the observed serial intervals will necessarily bias the estimates of $\mathcal{R}$. Furthermore, when one of the individuals is asymptomatic, there is no symptom-based serial interval; neglecting the time scale asymptomatic transmission may also bias the estimates of $\mathcal{R}$ (Park et al., 2020).

While we focus here on the effect of epidemiological dynamics on serial-interval distributions, other factors, such as intervention strategies, can also affect the generation- and serial-interval distributions. Individual-level interventions, such as case isolation, and behavioral responses directly affect individuals' ability to transmit and will shorten the forward generation- and serial-intervals. Population-level interventions, such as social distancing, can squeeze contacts into household and induce clustering, which will in turn shorten the realized generation interval due to local depletion of the susceptible pool within households (Park et al., 2019).

Despite these limitations, our analysis of serial intervals of COVID-19 from China provides further support for our theoretical framework, demonstrating temporal variation in serial intervals and its effect on the estimates of $\mathcal{R}$. Most existing estimates of the serial-intervals of COVID-19 implicitly or explicitly assume that the serial-interval distributions remain constant throughout the course of an epidemic (Du et al., 2020; He et al., 2020; Nishiura et al., 2020; Tindale et al., 2020; Zhao et al., 2020; Zhang et al., 2020). Our study provides a rationale for reassessing estimates of serial-interval distributions—and their use in estimating $\mathcal{R}$—during the COVID-19 pandemic.

5 Supplementary Materials

5.1 Deterministic simulation

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

$$
\begin{aligned}
i(t) &= \mathcal{R}_0 S(t-\Delta t) \sum_{m=1}^{m_{\max}} i(t-m\Delta t)\hat{g}(m\Delta t) \\
S(t) &= S(t-\Delta t) - i(t)
\end{aligned} \tag{19}
$$

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

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

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

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

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

$$
\hat{h}(m\Delta t, n\Delta t) = \hat{\ell}(m\Delta t)\hat{g}(n\Delta t), \quad m, n = 1, \ldots, m_{\max}. \tag{22}
$$

We use $\Delta t = 0.025$ days and $m_{\max} = 2001$ for discretization steps.

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

$$
\begin{aligned}
i(m\Delta t) &= C \exp(rm\Delta t), \quad m = 1, \ldots, m_{\max} \\
S(m\Delta t) &= N - \sum_{n=1}^{m} i(n\Delta t), \quad m = 1, \ldots, m_{\max}
\end{aligned} \tag{23}
$$

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

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 developed 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.

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 added 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 $\mathcal{R}_0$ using serial-interval distributions

The intrinsic generation-interval distribution $g(\tau)$ provides a link between r and $\mathcal{R}_0$ via the Euler-Lotka equation (Wallinga and Lipsitch, 2007):

$$\frac{1}{\mathcal{R}_0} = \int_0^\infty \exp(-r\tau) g(\tau)\, \mathrm{d}\tau\,. \tag{24}$$

In this section, we prove that the initial forward serial-interval distribution $f_0(\tau)$ also estimates the same $\mathcal{R}_0$ from r, except that integral extends to $\tau = -\infty$ rather than beginning at $\tau = 0$, because serial intervals can be negative:

$$\frac{1}{\mathcal{R}_0} = \int_{-\infty}^\infty \exp(-r\tau) f_0(\tau)\, \mathrm{d}\tau\,. \tag{25}$$

Here, the initial forward serial-interval distribution $f_0(\tau)$ is defined as:

$$f_0(\tau) = \frac{1}{\phi} \int_{-\infty}^0 \int_{\alpha_1}^\tau \exp(r\alpha_1) h(-\alpha_1, \alpha_2 - \alpha_1) \ell(\tau - \alpha_2)\, \mathrm{d}\alpha_2\, \mathrm{d}\alpha_1\,, \tag{26}$$

where h is the joint probability distribution describing the intrinsic generation-interval distribution g and the intrinsic incubation period distribution ℓ (see Eq. (12) in the main text), and the normalization constant ϕ is determined by the requirement that $\int_{-\infty}^\infty f_0(\tau)\ \mathrm{d}\tau = 1$.

In order to verify Eq. (25), we first rewrite the integral in Eq. (26) by substituting $-\alpha_1$ for α_1, and then changing the order of integration:

$$\begin{aligned} f_0(\tau) &= \frac{1}{\phi} \int_0^\infty \int_{-\alpha_1}^\tau \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) \ell(\tau - \alpha_2)\, \mathrm{d}\alpha_2\, \mathrm{d}\alpha_1\,, \\ &= \frac{1}{\phi} \int_{-\infty}^\tau \int_{\max(0, -\alpha_2)}^\infty \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) \ell(\tau - \alpha_2)\, \mathrm{d}\alpha_1\, \mathrm{d}\alpha_2\,. \end{aligned} \tag{27}$$

To further simplify the expression, we define $z(\alpha_2)$ as follows:

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

Substituting $z(\alpha_2)$ into Eq. (27) we obtain:

$$f_0(\tau) = \frac{1}{\phi} \int_{-\infty}^\tau z(\alpha_2) \ell(\tau - \alpha_2)\, \mathrm{d}\alpha_2\,, . \tag{29}$$

Writing $\hat{z}$ for a normalized version of z,

$$\hat{z}(\alpha_2) = \frac{z(\alpha_2)}{\int_{-\infty}^{\infty} z(x)\,\mathrm{d}x}\,, \tag{30}$$

we can now express the initial forward serial-interval distribution f_0 as a convolution of $\hat{z}$ and ℓ:

$$f_0(\tau) = \frac{1}{\hat{\phi}} \int_{-\infty}^{\tau} \hat{z}(\alpha_2)\ell(\tau - \alpha_2)\,\mathrm{d}\alpha_2\,, \tag{31}$$

where $\hat{\phi} = \phi / \int_{-\infty}^{\infty} z(x)\,\mathrm{d}x$.

Since the right hand side of Eq. (25) is also a Laplace transform of $f_0 = \hat{z} * \ell$, we can express it as the product of Laplace transforms of $\hat{z}$ and ℓ:

$$\int_{-\infty}^{\infty} \exp(-r\tau) f_0(\tau)\mathrm{d}\tau = \int_{-\infty}^{\infty} \exp(-r\tau)\hat{z}(\tau)\,\mathrm{d}\tau \int_{0}^{\infty} \exp(-r\tau)\ell(\tau)\,\mathrm{d}\tau\,. \tag{32}$$

In order to derive an expression for a Laplace transform of $\hat{z}$, we have to first derive an analytical expression for $\int_{-\infty}^{\infty} z(x)\,\mathrm{d}x$. By changing the order of integration, we have:

$$\begin{aligned}\int_{-\infty}^{\infty} z(\alpha_2)\mathrm{d}\alpha_2 &= \int_{-\infty}^{\infty} \int_{\max(0,-\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1)\,\mathrm{d}\alpha_1\,\mathrm{d}\alpha_2\,, \\ &= \int_{0}^{\infty} \int_{-\alpha_1}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1)\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_1\,. \end{aligned} \tag{33}$$

Since ℓ is a marginal probability distribution of h, it follows that:

$$\int_{-\infty}^{\infty} z(\alpha_2)\mathrm{d}\alpha_2 = \int_{0}^{\infty} \exp(-r\alpha_1)\ell(\alpha_1)\,\mathrm{d}\alpha_1\,. \tag{34}$$

Then, we have:

$$\hat{z}(\alpha_2) = \frac{\int_{\max(0,-\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1)\,\mathrm{d}\alpha_1}{\int_{0}^{\infty} \exp(-r\alpha_1)\ell(\alpha_1)\,\mathrm{d}\alpha_1}\,. \tag{35}$$

Substituting the expression into Eq. (32), we have:

$$\int_{-\infty}^{\infty} \exp(-r\tau) f_0(\tau)\,\mathrm{d}\tau = \int_{-\infty}^{\infty} \exp(-r\alpha_2) \int_{\max(0,-\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1)\,\mathrm{d}\alpha_1\,\mathrm{d}\alpha_2\,. \tag{36}$$

Recall that g is also a marginal probability distribution of h:

$$g(\tau) = \int_{0}^{\infty} h(x, \tau)\,\mathrm{d}x\,. \tag{37}$$

We can then substitute $\tau = \alpha_1 + \alpha_2$ into Eq. (36) and apply change of variables to obtain:

$$\int_{-\infty}^{\infty} \exp(-r\tau) f_0(\tau)\,\mathrm{d}\tau \tag{38}$$

$$= \int_{-\infty}^{\infty} \exp(-r\alpha_2) \int_{\max(0,-\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1)\,\mathrm{d}\alpha_1\,\mathrm{d}\alpha_2 \tag{39}$$

$$= \int_{0}^{\infty} \int_{0}^{\infty} \exp(-r\tau) h(\alpha_1, \tau)\,\mathrm{d}\alpha_1\,\mathrm{d}\tau \tag{40}$$

$$= \int_{0}^{\infty} \exp(-r\tau) g(\tau)\,\mathrm{d}\tau = \frac{1}{\mathcal{R}_0} \tag{41}$$

Therefore, the initial forward serial-interval distribution and the intrinsic generation-interval distribution give the same estimates of $\mathcal{R}_0$ from r. □

5.4 Comparing the estimates of $\mathcal{R}_0$ using the initial forward and the intrinsic serial-interval distributions

We use a simulation-based approach to compare the estimates of $\mathcal{R}_0$ 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 log-scale 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:

$$F_i = -X_{1,i} + (G_i | X_{1,i}) + X_{2,i}, \tag{42}$$

where the backward incubation period $X_{1,i}$ of an infector is simulated by drawing random log-normal samples Y_i with log mean μ_I and log variance σ_I^2 and resampling Y_i, each weighted by the inverse of the exponential growth function $\exp(-rY_i)$; the intrinsic generation interval conditional on the incubation period of the infector $(G_i|X_{1,i})$ is drawn from a log-normal distribution with log mean $\mu_G + \sigma_G\rho(\log(X_{1,i}) - \mu_I)/\sigma_I$ and log variance $\sigma_G^2(1-\rho^2)$; the forward incubation period $X_{2,i}$ 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 basic reproduction number $\mathcal{R}_0$ using the empirical estimator:

$$\mathcal{R}_0 = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \exp(-rF_i)}. \tag{43}$$

We compare this with an estimate of $\mathcal{R}_0$ based on the intrinsic serial-interval distribution which has the same mean as the intrinsic generation-interval distribution (Svensson, 2007; Klinkenberg and Nishiura, 2011; Champredon et al., 2018; Britton and Scalia Tomba, 2019):

$$\mathcal{R}_{\text{intrinsic}} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \exp(-rQ_i)}, \tag{44}$$

where

$$Q_i = -Y_i + (G_i|Y_i) + X_{2,i}. \tag{45}$$

5.5 Applications: SEIR model

Consider a Susceptible-Exposed-Infectious-Recovered model:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\beta SI \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \beta SI - \gamma_E E \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \gamma_E E - \gamma_I I \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma_I I
\end{aligned} \tag{46}$$

where β is the transmission rate, $1/\gamma_E$ is the mean latent period, and $1/\gamma_I$ is the mean infectious period. We further assume that the latent period is equivalent to incubation period; in other words, infected individuals can only transmit after symptom onset. Then, the generation interval will be always longer than the incubation period.

The joint probability distribution of the intrinsic incubation periods and intrinsic generation intervals for this model can be written as:

$$h(x,\tau) = \begin{cases} 0 & x > \tau \\ \gamma_I \gamma_E \exp(-\gamma_I(\tau - x) - \gamma_E x) & x \leq \tau \end{cases} \tag{47}$$

Then, the intrinsic generation-interval distribution is given by:

$$\begin{aligned}
g(\tau) &= \int_0^\tau h(x,\tau)\,\mathrm{d}x \\
&= \frac{\gamma_I \gamma_E}{\gamma_E - \gamma_I}(\exp(-\gamma_I \tau) - \exp(-\gamma_E \tau))
\end{aligned} \tag{48}$$

On the other hand, the initial forward serial-interval distribution is given by:

$$\begin{aligned}
f_0(\tau) &\propto \int_{-\infty}^0 \int_0^\tau \exp(r\alpha_1) h(-\alpha_1, \alpha_2 - \alpha_1) \ell(\tau - \alpha_2)\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_1 \\
&\propto \int_{-\infty}^0 \int_0^\tau \exp(r\alpha_1) \exp(-\gamma_I \alpha_2 + \gamma_E \alpha_1) \exp(-\gamma_E(\tau - \alpha_2))\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_1 \\
&\propto \exp(-\gamma_E \tau) \int_{-\infty}^0 \int_0^\tau \exp((\gamma_E - \gamma_I)\alpha_2) \exp((r + \gamma_E)\alpha_1)\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_1 \\
&\propto (\exp(-\gamma_I \tau) - \exp(-\gamma_E \tau)) \int_{-\infty}^0 \exp((r + \gamma_E)\alpha_1)\,\mathrm{d}\alpha_1 \\
&\propto \exp(-\gamma_I \tau) - \exp(-\gamma_E \tau)
\end{aligned} \tag{49}$$

Therefore, both the intrinsic generation intervals and the initial forward serial intervals are identically distributed and have the same mean.

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., H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet 395*(10228), 931–934.

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

Backer, J. A., D. Klinkenberg, and J. Wallinga (2020). Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance 25*(5).

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 323*(14), 1406–1407.

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

Chowell, G., L. Simonsen, C. Viboud, and Y. Kuang (2014). Is West Africa approaching a catastrophic phase or is the 2014 Ebola epidemic slowing down? Different models yield different answers for Liberia. *PLoS currents 6*.

Chunara, R., J. R. Andrews, and J. S. Brownstein (2012). Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American journal of tropical medicine and hygiene 86*(1), 39–45.

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).

Ferguson, N. M., D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature 437*(7056), 209–214.

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 368*(6491).

Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, H. Coupland, T. A. Mellan, H. Zhu, T. Berah, J. W. Eaton, P. N. Guzman, et al. (2020). Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. `https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-30-COVID19-Report-13.pdf`. Accessed April 30, 2020.

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

Fraser, C., C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons, et al. (2009). Pandemic potential of a strain of influenza A (H1N1): early findings. *science 324*(5934), 1557–1561.

Ganyani, T., C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens (2020). Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance 25*(17), 2000257.

Hampson, K., J. Dushoff, S. Cleaveland, D. T. Haydon, M. Kaare, C. Packer, and A. Dobson (2009). Transmission dynamics and prospects for the elimination of canine rabies. *PLoS biology 7*(3).

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.

Hellewell, J., S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, S. Flasche, B. J. Quilty, N. Davies, Y. Liu, S. Clifford, P. Klepac, M. Jit, C. Diamond, H. Gibbs, K. [van Zandvoort], S. Funk, and R. M. Eggo (2020). Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health 8*(4), e488 – e496.

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 172*(9), 577–582.

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 8*(5), e627–e630.

Nishiura, H. (2010). Time variations in the generation time of an infectious disease: implications for sampling to appropriately quantify transmission potential. *Mathematical Biosciences & Engineering 7*(4), 851–869.

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

Pan, A., L. Liu, C. Wang, H. Guo, X. Hao, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, S. Wei, and T. Wu (2020, 04). Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA*.

Park, S. W., B. M. Bolker, D. Champredon, D. J. D. 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*. https://doi.org/10.1101/2020.01.30.20019877.

Park, S. W., D. Champredon, and J. Dushoff (2019). Inferring generation-interval distributions from contact-tracing data. *bioRxiv*, 683326. https://doi.org/10.1101/683326.

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. *Epidemics*, 100392.

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*. https://doi.org/10.1101/2020.03.27.20045815.

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.

Sciré, J., S. A. Nadeau, T. G. Vaughan, B. Gavin, S. Fuchs, J. Sommer, K. N. Koch, R. Misteli, L. Mundorff, T. Götz, et al. (2020). Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft. *Swiss Medical Weekly 150*(19-20), w20271.

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*. https://doi.org/10.1101/2020.03.03.20029983.

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.

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

Weitz, J. S. and J. Dushoff (2015). Modeling post-death transmission of Ebola: challenges for inference and opportunities for control. *Scientific reports 5*, 8751.

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., P. Cao, D. Gao, Z. Zhuang, Y. Cai, J. Ran, M. K. Chong, K. Wang, Y. Lou, W. Wang, et al. (2020). Serial interval in determining the estimation of reproduction number of the novel coronavirus disease (COVID-19) during the early outbreak. *Journal of Travel Medicine 27*(3), taaa033.

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*. https://doi.org/10.1101/2020.02.21.20026559.