

1
2 Forward-looking serial intervals correctly link epidemic growth to
3 reproduction numbers

4
5 Sang Woo Park^{1,*} Kaiyuan Sun² David Champredon³ Michael Li⁴ Benjamin M. Bolker^{4,5,6}
6 David J. D. Earn^{5,6} Joshua S. Weitz^{7, 8} Bryan T. Grenfell^{1,2,9} Jonathan Dushoff^{4,5,6}

7 **1** Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ,
8 USA

9 **2** Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

10 **3** Department of Pathology and Laboratory Medicine, University of Western Ontario,
11 London, Ontario, Canada

12 **4** Department of Biology, McMaster University, Hamilton, ON, Canada

13 **5** Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada

14 **6** M. G. DeGroot Institute for Infectious Disease Research, McMaster University,
15 Hamilton, ON, Canada

16 **7** School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

17 **8** School of Physics, Georgia Institute of Technology, Atlanta, GA, USA

18 **9** Woodrow Wilson School of Public and International Affairs, Princeton University,
19 Princeton, NJ, USA

20 *Corresponding author: swp2@princeton.edu

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

24 Abstract

25 The reproduction number \mathcal{R} and the growth rate r are critical epidemiological quantities.
26 They are linked by generation intervals, the time between infection and onward transmission.
27 Because generation intervals are difficult to observe, epidemiologists often substitute serial
28 intervals, the time between symptom onset in successive links in a transmission chain. Recent
29 studies suggest that such substitution biases estimates of \mathcal{R} based on r . Here we explore how
30 these intervals vary over the course of an epidemic, and the implications for \mathcal{R} estimation.
31 *Forward-looking* serial intervals, measuring time forward from symptom onset of an infector,
32 correctly describe the renewal process of symptomatic cases and therefore reliably link \mathcal{R}
33 with r . In contrast, *backward-looking* intervals, which measure time backward, and *intrinsic*
34 intervals, which neglect population-level dynamics, give incorrect \mathcal{R} estimates. Forward-
35 looking intervals are affected both by epidemic dynamics and by censoring, changing in
36 complex ways over the course of an epidemic. We present a heuristic method for addressing
37 biases that arise from neglecting changes in serial intervals. We apply the method to early (21
38 January – 8 February 2020) serial-interval-based estimates of \mathcal{R} for the COVID-19 outbreak
39 in China outside Hubei province; using improperly defined serial intervals in this context
40 biases estimates of initial \mathcal{R} by up to a factor of 2.6. This study demonstrates the importance
41 of early contact-tracing efforts and provides a framework for reassessing generation intervals,
42 serial intervals, and \mathcal{R} estimates for COVID-19.

43 Significance Statement

44 The generation- and serial-interval distributions are key, but different, quantities in outbreak
45 analyses. Recent theoretical studies suggest that the two distributions give different estimates
46 of the reproduction number \mathcal{R} as inferred from the observed exponential growth rate r .
47 Here, we show that estimating \mathcal{R} based on r and the serial-interval distribution, when
48 defined from the correct reference time and cohort, gives the same estimate as using r and
49 the generation-interval distribution. We apply our framework to serial-interval data from
50 the COVID-19 outbreak in China, outside Hubei province (January 21–February 8, 2020),
51 revealing systematic biases in prior inference methods. Our study provides the theoretical
52 basis for practical changes to the principled use of serial interval distributions in estimating
53 \mathcal{R} during epidemics.

54 1 Introduction

55 The reproduction number \mathcal{R} is one of the most important characteristics of an emerging
56 epidemic, such as the current pandemic of coronavirus disease 2019 (COVID-19) (Majumder
57 and Mandl, 2020). The reproduction number is defined as the average number of secondary
58 cases caused by a primary case. The value in a fully susceptible population — the “basic”
59 reproduction number \mathcal{R}_0 — allows us to predict the extent to which an infection will spread
60 in the population, and the amount of intervention necessary to eliminate it in simple cases
61 (Anderson and May, 1991). Since the reproduction number represents an average (Diekmann
62 et al., 1990; Anderson and May, 1991), it fails to capture heterogeneity among individuals
63 or across space. The reproduction number also fails to provide any information about the
64 time scale of disease transmission.

65 Estimating the reproduction number \mathcal{R} is often challenging. Direct estimates based on
66 observed infections will typically be biased down when some infections cannot be observed.
67 A common method of estimating \mathcal{R} near the beginning of an epidemic is based on the
68 population-level exponential growth rate r , which can often be estimated robustly from case
69 reports (Mills et al., 2004; Ma et al., 2014). The growth rate r and the reproduction number
70 \mathcal{R} are linked by the generation-interval distribution Wallinga and Lipsitch (2007), where the
71 generation interval is defined as the time between when an individual (infector) is infected
72 and when that individual infects another person (infectee) (Svensson, 2007).

73 Since generation intervals measure time between infection events, which can be difficult
74 to observe in practice, generation intervals are often replaced with serial intervals. The
75 serial interval is defined as the time between when an infector and an infectee develop
76 symptoms (Svensson, 2007). While generation and serial intervals both measure the time
77 scale of disease transmission, they measure fundamentally different quantities. In particular,
78 previous studies have noted that, in many contexts, serial intervals are expected to have larger
79 variances than generation intervals but have the same mean in many contexts (Svensson,
80 2007; Klinkenberg and Nishiura, 2011; te Beest et al., 2013; Champredon et al., 2018). Serial
81 intervals can in some cases even take negative values in the presence of presymptomatic
82 transmission (He et al., 2020), whereas generation intervals must be positive.

83 Although these distributions were clearly and distinctly defined over a decade ago (Svens-
84 son, 2007), the need for a better conceptual and theoretical framework for understanding
85 their differences is becoming clearer as the COVID-19 pandemic unfolds. Researchers con-
86 tinue to base inferences about COVID-19 on both generation and serial intervals without
87 clearly distinguishing between them (e.g., Abbott et al. (2020); Du et al. (2020); He et al.
88 (2020); Wu et al. (2020); Zhao et al. (2020)), and, in some cases, explicitly conflate the
89 definitions of the two intervals (e.g., Anderson et al. (2020); Hellewell et al. (2020)). This
90 confusion is apparent even in standard software for estimating \mathcal{R} , such as `EpiEstim`, in which
91 the serial-interval distribution is used to infer time-dependent \mathcal{R} (Thompson et al., 2019).
92 These studies are examples of many—indeed, it is a common practice to use the serial and
93 generation intervals interchangeably.

94 One source of confusion arises from an apparent discrepancy between the generation-
95 interval and serial-interval viewpoints. While the epidemic is growing exponentially, the

96 spread of infection can be characterized as a *renewal process* based on previous incidence
97 of infection, the associated generation-interval distribution, and the average infectiousness
98 of an infected individual. It is well established that this renewal formulation allows us to
99 link the exponential growth rate of an epidemic r with its reproduction number \mathcal{R} using the
100 generation-interval distribution (Wallinga and Lipsitch, 2007). However, the serial-interval
101 distribution also describes a renewal process — in this case, the creation of a new *symp-*
102 *tomatic* case based on a symptomatic case in the previous generation. Since both renewal
103 processes, based on either generation- or serial-interval distributions, describe the same un-
104 derlying exponentially growing system, both should provide the same correct link between
105 the reproduction number \mathcal{R} and the epidemic growth rate r .

106 In contexts where the serial- and generation-interval distributions differ, current theory
107 has no explanation for how two different distributions could provide identical estimates of
108 \mathcal{R} from r . In fact, recent theory suggest that using the serial-interval can underestimate
109 the reproduction number (Britton and Scalia Tomba, 2019; Ganyani et al., 2020). However,
110 these studies rely on *intrinsic* distributions of incubation periods and generation intervals
111 that neglect the population-level dynamics of disease spread.

112 Here we show that, by correctly defining and calculating the “forward” serial-interval
113 distribution (i.e., a distribution of serial intervals from a cohort of infectors that developed
114 symptoms at the same time) that connects symptom onset dates, we can resolve this discrep-
115 ancy. These forward intervals are different from the “intrinsic” serial intervals that previous
116 studies have relied on (Svensson, 2007; Klinkenberg and Nishiura, 2011; te Beest et al., 2013;
117 Champredon et al., 2018; Britton and Scalia Tomba, 2019). During an ongoing epidemic, all
118 observed epidemiological delays (e.g., incubation period) between primary (e.g., infection)
119 and secondary (e.g., symptom onset) events are subject to backward biases: when the inci-
120 dence of primary events is increasing (or decreasing), we are more likely to observe shorter
121 (respectively longer) intervals. In particular, when we consider forward serial-interval distri-
122 butions, the incubation periods of the infectors are subject to backward biases because we
123 have to look backward in time from their symptom onset to infection. Therefore, the realized
124 incubation period distributions of the infector and the infectee can differ dynamically, even
125 if the intrinsic analogues of the same distributions are expected to be equivalent.

126 We develop a cohort-based framework for characterizing and comparing realized serial
127 intervals, as well as any other epidemiological delays, and show that the initial forward serial-
128 interval distribution correctly estimates \mathcal{R} from r . Conversely, using inaccurately defined
129 serial intervals or failing to account for changes in the observed serial-interval distributions
130 over the course of an epidemic can considerably bias estimates of \mathcal{R} . For example, in our
131 analysis of the COVID-19 serial intervals from China, outside Hubei province, we find that
132 the original \mathcal{R}_0 estimates based on aggregated serial-interval data underestimated \mathcal{R}_0 by
133 a factor of 2.0–2.6. We further lay out several principles to consider in using information
134 about serial intervals and other epidemiological time delays to correctly infer the initial
135 reproduction number during the early stages of an outbreak.

2 Methods

2.1 Intrinsic, forward, and backward delay distributions

A time delay between two epidemiological events can involve either one infected individual (e.g., incubation period: infection and symptom onset of an individual) or two — an infector and an infectee (e.g., generation and serial intervals). We define the delay as the time difference between the *primary* event and the *secondary* event. In some cases, the primary event always occurs before the secondary event (e.g., the time from infection to onset of symptoms in a single individual, or the generation interval between two individuals). In other cases, the delay can sometimes be negative (e.g., the time from onset of infectiousness to onset of symptoms in a single individual, or the serial interval between two individuals).

At the individual level, we can define the time distribution between a primary and a secondary event that we expect to observe for a single infected individual by averaging across individual characteristics — 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; see Fig. 1). 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 the intrinsic generation intervals (time between infection and infectious contacts) (Park et al., 2020). In this example, the difference between intrinsic and realized time distributions can be attributed to the fact that the fraction of susceptible individuals is itself dynamic.

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 a primary cohort at time p , we can define the distribution of realized delays 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 events 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 .

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

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

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

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

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

191 If we are considering incubation periods, the left hand side of this equation is the probability
192 density that an individual who became symptomatic at time s had an incubation period
193 of length τ . From the right hand side, we see that this probability density depends on
194 the weight parameter $W(s - \tau)$ (in this case, the proportion of symptomatic infection), the
195 time-varying primary cohort size at the earlier time $\mathcal{P}(s - \tau)$ (in this case, the number of
196 individuals infected at time $s - \tau$), and the forward delay distribution $f_{s-\tau}(\tau)$ (in this case,
197 the probability density that an incubation period that starts at time $s - \tau$ ends at time s).

198 Several different mechanisms drive the changes in forward and backward delay distri-
199 butions over time. Typically, within-individual forward delay distributions are not directly
200 affected by epidemic dynamics. Some realized forward distributions, like incubation pe-
201 riod distributions, are equivalent to their intrinsic distributions and remain invariant at the
202 time scale of an outbreak. Other realized distributions, like the distribution of time from
203 symptom onset to testing, may change over the course of an epidemic due to changes in
204 public-health policies or individual behavior. Between-individual forward delay distribu-
205 tions, such as generation- or serial-interval distributions, depend on epidemic dynamics. For
206 example, forward generation intervals often become shorter as an epidemic progresses due to
207 the dynamical process of susceptible depletion, as well as due to other factors like behavioral
208 change or interventions (Kenah et al., 2008; Nishiura, 2010; Champredon and Dushoff, 2015):
209 if it is harder to infect later in the course of infection, then proportionally more intervals will
210 be short.

Eq. (2) suggests that backward delay distributions change over time even if their corre-
sponding forward delay distribution does not change. Backward delay distributions depend
on changes in the primary cohort size over time due to conditionality of observations: Con-
ditioning 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:

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

211 where r is the exponential growth rate. Since b_0 is a probability distribution, $[W(0)\mathcal{P}(0)/\mathcal{S}(0)]^{-1} =$
212 $\int_{-\infty}^{\infty} \exp(-r\tau')f_0(\tau') d\tau'$ corresponds to the normalization constant. Therefore, the backward
213 delay distribution during the exponential growth phase depends only on the exponential
214 growth rate r and the initial forward delay distribution f_0 .

215 The mean backward interval will be always shorter than the mean forward interval as
216 long as $r > 0$. Even for different epidemics of the same disease, we expect to observe shorter
217 backward intervals within a fast-growing epidemic (high r), all else being equal. In general,
218 the backward delay distribution will differ from the forward delay distribution (unless the
219 disease is at equilibrium), even if we are measuring time delays that are intrinsic to the
220 life history of a disease (e.g., the incubation period). These ideas apply to all epidemiolog-
221 ical delay distributions and generalize the work by Champredon and Dushoff (2015) who
222 compared forward and backward generation-interval distributions to describe realized gener-
223 ation intervals from the perspective of an infector and an infectee, respectively, as well as the
224 work by Britton and Scalia Tomba (2019) who showed that Eq. (3) holds for the backward
225 generation-interval distribution.

226 2.2 Realized serial-interval distributions

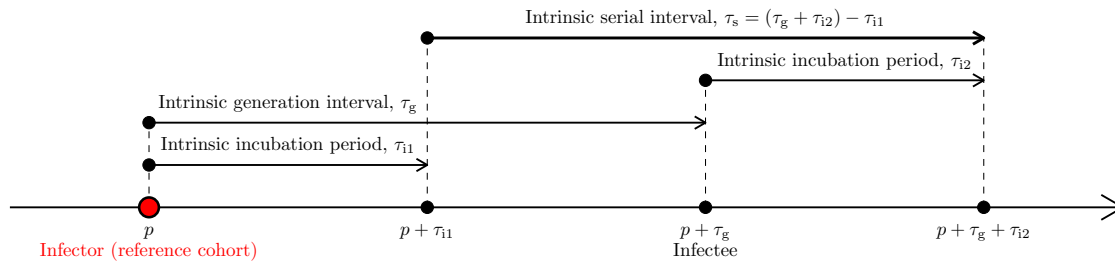
227 The serial interval is defined as the time between when an infector becomes symptomatic
228 and when their infectee becomes symptomatic (Svensson, 2007). Previous studies have often
229 expressed serial intervals τ_s in the form (Fig. 1A):

$$\tau_s = (\tau_g + \tau_{i2}) - \tau_{i1} \quad (4)$$

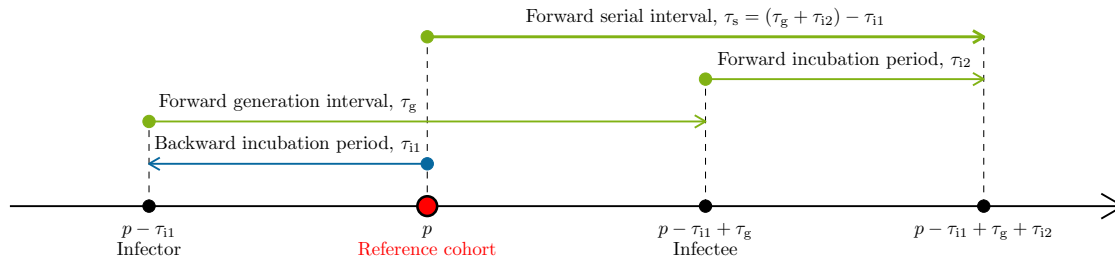
230 where τ_{i1} and τ_{i2} represent incubation periods of an infector and an infectee, respectively, and
231 τ_g represents the generation interval between the infector and the infectee. These studies
232 concluded that the serial and generation intervals have the same mean when τ_{i1} and τ_{i2}
233 are drawn from the same distributions (Svensson, 2007; Klinkenberg and Nishiura, 2011;
234 Champredon et al., 2018; Britton and Scalia Tomba, 2019). However, distributions of realized
235 incubation periods, τ_{i1} and τ_{i2} will be identical only if we assume that they are intrinsic to
236 individuals (and not dependent on epidemic dynamics at the population-level) — something
237 that is generally true of forward but not backward incubation-period distributions. We refer
238 to the definition Eq. (4) as the intrinsic serial interval (Fig. 1A).

239 To correctly link the realized serial-interval distribution to the renewal process between
240 cases based on symptom onset dates, we must use the forward serial interval (i.e., use the
241 perspective of a cohort of infectors that share the same symptom onset time). Given that

A. Intrinsic serial interval



B. Forward serial interval



C. Backward serial interval

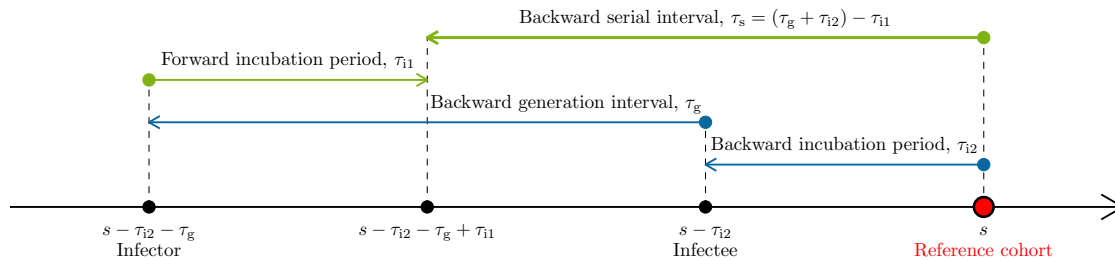


Figure 1: **Illustration of intrinsic, forward and backward serial intervals.** (A) The intrinsic serial interval for a cohort of individuals infected at time p . In this case, τ_{i1} is drawn from the intrinsic incubation period distribution; τ_g is drawn from the intrinsic generation-interval distribution; 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. Intrinsic intervals (black) reflect average of individual characteristics and are not dependent on population-level dynamics. Forward intervals (green) can change due to epidemiological dynamics (e.g., contraction of generation intervals through susceptible depletion). Backward intervals (blue) can change due to changes in cohort sizes even when forward intervals remain time-invariant.

242 an infector became symptomatic at time p , to calculate the forward serial interval we first
 243 go *backward* in time to when the infector was infected, and then forward in time to when
 244 the infectee was infected, and then forward again to when the infectee became symptomatic.
 245 In Fig. 1B, we see that τ_{i1} is drawn from the backward incubation period distribution of the
 246 cohort of infectors who became symptomatic at time p ; τ_g is drawn from the forward
 247 generation-interval distribution of the cohort of infectors who became infected at time $p - \tau_{i1}$;
 248 and τ_{i2} is drawn from the forward incubation period distribution of the cohort of infectees who
 249 became infected at time $p - \tau_{i1} + \tau_g$. Likewise, we can define the backward serial-interval
 250 distribution for a cohort of infectees who became symptomatic at time s (Fig. 1C). This
 251 conceptual framework demonstrates that the distributions of τ_{i1} , τ_g , and τ_{i2} (and therefore
 252 the distributions of realized serial intervals) depend on the reference cohort, which is defined
 253 by temporal direction (forward or backward) and a particular reference time.

254 To calculate realized serial-interval distributions, we begin by modeling $\mathcal{T}(p, s)$: the total
 255 density of serial intervals that start (when infectors develop symptoms) at time p and end
 256 (when infectees develop symptoms) at time s . For simplicity, we assume that all infected
 257 individuals eventually develop symptoms. Then, the density of serial intervals between time
 258 p and s , given that the infectors became infected at time $\alpha_1 \leq p$ and the infectees became
 259 infected at time $\alpha_2 \leq s$, depends on the amount of infection that occurs between time α_1 and
 260 α_2 as well as the density of forward incubation periods between α_1 and p (realized incubation
 261 periods of infectors) and between α_2 and s (realized incubation periods of infectees):

$$\mathcal{T}(p, s | \alpha_1, \alpha_2) = \underbrace{\mathcal{R}_c(\alpha_1)}_{\substack{\text{case} \\ \text{reproduction} \\ \text{number}}} \times \underbrace{i(\alpha_1)}_{\substack{\text{incidence} \\ \text{of} \\ \text{infection}}} \times \underbrace{h_{\alpha_1}(p - \alpha_1, \alpha_2 - \alpha_1)}_{\substack{\text{joint density of} \\ \text{forward incubation} \\ \text{periods } p - \alpha_1 \text{ and forward} \\ \text{generation intervals } \alpha_2 - \alpha_1 \\ \text{(of infectors)}}} \times \underbrace{\ell_{\alpha_2}(s - \alpha_2)}_{\substack{\text{marginal density of} \\ \text{forward incubation} \\ \text{periods } s - \alpha_2 \\ \text{(of infectees)}}, \quad (5)$$

262 where the case reproduction number $\mathcal{R}_c(\alpha_1)$ is defined as the average number of secondary
 263 cases that a primary case infected at time α_1 will generate over the course of their infection
 264 (Fraser, 2007). We describe the forward incubation periods and the forward generation
 265 intervals using a joint probability distribution because onset of symptoms and transmission
 266 potential jointly depend on the life history of a disease; for example, if an infected individual
 267 can only transmit the disease after symptom onset, the forward generation interval will
 268 necessarily be longer than the forward incubation period.

269 The total density of serial intervals between time p and s can now be obtained by inte-
 270 grating over all possible infection times for the infector and the infectee:

$$\mathcal{T}(p, s) = \int_{-\infty}^p \int_{\alpha_1}^s \mathcal{T}(p, s | \alpha_1, \alpha_2) d\alpha_2 d\alpha_1. \quad (6)$$

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

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

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

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

276 The denominator of the forward serial-interval distribution (Eq. (7)) then corresponds
 277 to the total number of infections generated by primary cases who themselves developed
 278 symptoms at time p . Dividing this quantity by the number of individuals who developed
 279 symptoms at time p , $j(p) = \int_{-\infty}^{\infty} \mathcal{T}(p - \tau', p) d\tau'$, we obtain the serial reproduction number:

$$\mathcal{R}_s(p) = \frac{\int_{-\infty}^{\infty} \mathcal{T}(p, p + \tau') d\tau'}{j(p)}, \quad (9)$$

280 which we define as the average number of infections generated by an individual who developed
 281 symptoms at time p . Combining the forward serial-interval distribution with the serial
 282 reproduction number completes the renewal process between symptomatic cases:

$$j(t) = \int_{-\infty}^{\infty} \mathcal{R}_s(t - \tau) j(t - \tau) f_{t-\tau}(\tau) d\tau. \quad (10)$$

283 This framework allows us to understand changes in the realized serial intervals for any epi-
 284 demic model and properly link serial-interval distributions with the renewal process. In
 285 addition, assuming that the reproduction number as well as the forward serial-interval dis-
 286 tribution remain constant during the exponential growth phase, we can substitute $j(t) \approx$
 287 $j(0) \exp(rt)$, $\mathcal{R}_s(t) \approx \mathcal{R}_s(0)$, and $f_{t-\tau}(\tau) \approx f_0(\tau)$ to obtain:

$$\frac{1}{\mathcal{R}_s(0)} = \int_{-\infty}^{\infty} \exp(-r\tau) f_0(\tau) d\tau. \quad (11)$$

288 Therefore, the initial forward serial-interval distribution, $f_0(\tau)$, provides the correct link
 289 between the exponential growth rate r and the initial serial reproduction number $\mathcal{R}_s(0)$.
 290 We re-visit this idea later in Section 2.4 and show that the initial forward serial-interval
 291 distribution provides the same r - \mathcal{R} link as the intrinsic generation-interval distribution.

292 2.3 Epidemic model

We illustrate changes in forward and backward serial intervals over the course of an epi-
 demic by applying our framework 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 suscep-
 tible individuals $S(t)$ and incidence of infection $i(t)$ can be described as:

$$\begin{aligned} \frac{dS}{dt} &= -i(t) \\ i(t) &= \mathcal{R}(t) \int_0^{\infty} i(t - \tau) g(\tau) d\tau, \end{aligned} \quad (12)$$

293 where $\mathcal{R}(t)$ is the instantaneous reproduction number (i.e., the average number of secondary
 294 cases that a primary case infected at time t will generate if conditions at time t remain
 295 unchanged (Fraser, 2007)), and $g(\tau)$ is the intrinsic generation-interval distribution (i.e., the
 296 forward generation-interval distribution of a primary case in a population where changes in
 297 $\mathcal{R}(t)$ is negligible (Champredon and Dushoff, 2015)). This model assumes that $g(\tau)$ remains
 298 constant through time – in other words, that epidemic dynamics are driven by changes in
 299 transmission rate. This assumption may not be well suited to individual-based intervention
 300 such as case isolation (Fraser, 2007); nonetheless, this form has been widely used in the
 301 literature and has been successfully applied in modeling the current COVID-19 pandemic
 302 (Gostic et al., 2020).

303 Here, changes in reproduction number can be modeled as a product of the basic repro-
 304 duction number \mathcal{R}_0 , proportion susceptible $S(t)$, and a time-dependent factor $M(t)$ (for
 305 example, accounting for nonpharmaceutical interventions and behavioral changes): $\mathcal{R}(t) =$
 306 $\mathcal{R}_0 S(t) M(t)$; Flaxman et al. (2020) used a similar framework to evaluate the impact of
 307 nonpharmaceutical interventions on the spread of COVID-19 in 11 countries. Then, the
 308 forward generation-interval for a cohort of individuals that were infected at time p follows
 309 (Champredon and Dushoff, 2015):

$$g_p(\tau) = \frac{g(\tau)S(p + \tau)M(p + \tau)}{\int_0^\infty g(\tau')S(p + \tau')M(p + \tau') d\tau'}, \quad (13)$$

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

$$h_p(x, \tau) = \frac{h(x, \tau)S(p + \tau)M(p + \tau)}{\int_0^\infty \int_0^\infty h(x', \tau')S(p + \tau')M(p + \tau') d\tau' dx'}. \quad (14)$$

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:

$$\begin{aligned} \ell(x) &= \int_0^\infty h(x, \tau) d\tau, \\ g(\tau) &= \int_0^\infty h(x, \tau) dx. \end{aligned} \quad (15)$$

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

$$\mathcal{R}_c(t) = \mathcal{R}_0 \int_0^\infty g(\tau)S(t + \tau)M(p + \tau) d\tau. \quad (16)$$

315 The forward and backward serial-interval distributions are then calculated by substituting
 316 these quantities into Eq. (7) and Eq. (8). We use this framework to illustrate how the realized

317 epidemiological time distributions vary over the course of an epidemic and depend on the
318 perspective (i.e., forward vs. backward).

319 For simplicity, we let $M = 1$ and assume that epidemic dynamics depend only on sus-
320 ceptible depletion in our simulations. Since we are interested in the initial epidemic growth
321 phase (i.e., linking r to \mathcal{R}), we expect $\mathcal{R}(t)$ to remain roughly constant during this period. In
322 addition, qualitative effects of M that reduces $\mathcal{R}(t)$ monotonically over time will be similar
323 to the impact of susceptible depletion under this modeling framework. Therefore, general
324 conclusions we draw from our analysis is expected to be robust—however, detailed shape of
325 the epidemic curve and changes in generation- and serial-intervals can still depend on the
326 shape of M .

327 2.4 Linking r and \mathcal{R}

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

$$\frac{1}{\mathcal{R}_0} = \int_0^\infty \exp(-r\tau)g(\tau) d\tau. \quad (17)$$

335 Analogous to the intrinsic generation-interval distribution, forward serial-interval distribu-
336 tions describe the renewal process between symptomatic cases. Therefore, we expect the
337 forward serial-interval distribution during the exponential growth phase — which we refer to
338 as the *initial* forward serial-interval distribution f_0 — to estimate the same value of \mathcal{R}_0 for
339 a given r as the intrinsic generation-interval distribution (note, however, that the forward
340 serial interval is not necessarily positive):

$$\frac{1}{\mathcal{R}_0} = \int_{-\infty}^\infty \exp(-r\tau)f_0(\tau)d\tau. \quad (18)$$

341 Here, the initial forward serial-interval distribution is given by:

$$f_0(\tau) = \frac{1}{\phi} \int_{-\infty}^0 \int_{\alpha_1}^\tau \exp(r\alpha_1)h(-\alpha_1, \alpha_2 - \alpha_1)\ell(\tau - \alpha_2) d\alpha_2 d\alpha_1, \quad (19)$$

342 where the normalization constant ϕ is determined by the requirement that $\int_{-\infty}^\infty f_0(\tau) d\tau = 1$.
343 We provide a mathematical proof of this relationship in Supplementary Materials. Since we
344 do not make any assumptions about the shape of the joint distribution h between incubation
345 periods and the generation intervals, Eq. (18) holds in general whether or not there is a
346 presymptomatic transmission period.

347 We further compare this with the estimate of \mathcal{R}_0 based on the intrinsic serial-interval
348 distribution $q(\tau)$:

$$\frac{1}{\mathcal{R}_{\text{intrinsic}}} = \int_{-\infty}^\infty \exp(-r\tau)q(\tau)d\tau. \quad (20)$$

349 The intrinsic serial-interval distribution $q(\tau)$ does not depend on epidemic dynamics, and is
350 given by:

$$q(\tau) = \frac{1}{\phi_q} \int_{-\infty}^0 \int_{\alpha_1}^{\tau} h(-\alpha_1, \alpha_2 - \alpha_1) \ell(\tau - \alpha_2) d\alpha_2 d\alpha_1, \quad (21)$$

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

354 The initial forward serial-interval distribution depends on the exponential growth rate
355 r . For a fast-growing epidemic (high r), we expect the backward incubation periods to be
356 short (Eq. (3)), meaning that presymptomatic transmission is less likely to occur. Therefore,
357 the initial forward serial-interval distribution will generally have a larger mean than the in-
358 trinsic generation- and serial-interval distributions. However, the exact shape of the initial
359 forward serial-interval distribution depends on the shape of the joint distribution. For ex-
360 ample, the Susceptible-Exposed-Infected-Recovered model, under the additional assumption
361 that the incubation and exposed periods are equivalent (i.e. that onset of symptoms and
362 infectiousness occur simultaneously), provides a special case. In this case, the forward serial-
363 and generation-intervals follow the same distributions during the exponential growth phase
364 because (i) infected individuals can only transmit after symptom onset and (ii) the time be-
365 tween symptom onset and infection is independent of the incubation period of an infector (see
366 Supplementary Materials). Everywhere else in this paper, however, we do not assume that
367 the incubation and exposed periods are equivalent. Instead, we allow for presymptomatic
368 transmission in the model in order to reflect the transmission dynamics of COVID-19.

369 2.5 Model parameterization

370 We have shown that the dynamics of the serial-interval distribution depend on the joint
371 distribution between incubation periods and generation intervals. Here, we use a bivariate
372 lognormal distribution to model the joint probability distribution h of intrinsic incubation
373 periods and intrinsic generation intervals (in the renewal equation model, Eq. (12)) while
374 allowing for the possibility that they might be correlated. Given that the viral load of SARS-
375 CoV-2 peaks around the time of symptom onset (He et al., 2020), we generally expect the
376 generation intervals to be positively correlated with the incubation period: that is, individu-
377 als who develop symptoms later are more likely to transmit later. Marginal distributions of
378 incubation periods and generation intervals are parameterized based on parameter estimates
379 for COVID-19 (Table 1). For simplicity, we consider four values for the correlation coeffi-
380 cients (on the log scale) of the bivariate lognormal distribution: $\rho = 0, 0.25, 0.5, 0.75$. This
381 parameterization allows for generation intervals to be shorter than the incubation period,
382 allowing for presymptomatic transmission.

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	5.0 days	Ferretti et al. (2020)
SD intrinsic generation interval	1.9 days	Ferretti et al. (2020)

Table 1: **Parameter values used for simulations.** 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 intrinsic generation-interval distribution is parameterized using a log-normal distribution with log mean $\mu_G = 1.54$ and log standard deviation $\sigma_G = 0.37$. Log mean and log standard deviations represent the mean and standard deviations of the underlying normal distributions, which are later exponentiated. The joint probability distribution is modeled using a bivariate log-normal distribution with correlations (on the log scale) $\rho = \{0, 0.25, 0.5, 0.75\}$. The intrinsic incubation period and generation-interval distributions are chosen to match characteristic of COVID-19 to illustrate realistic magnitudes of time-varying/perspective effects in the current pandemic.

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 an epidemic and to 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 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 serial-interval data from the COVID-19 epidemic in China, outside Hubei province, based on 468 transmission events reported between January 21–February 8, 2020, under our framework.

3.1 Realized serial-interval distributions during the exponential growth phase

Fig. 2 shows Euler-Lotka estimates of \mathcal{R}_0 based on different definitions of the serial interval. When the initial forward serial-interval distribution $f_0(\tau)$ is used, estimates (from Eq. (18)) exactly match the (correct) generation-interval-based estimates (Eq. (17)) 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. (20)) 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).

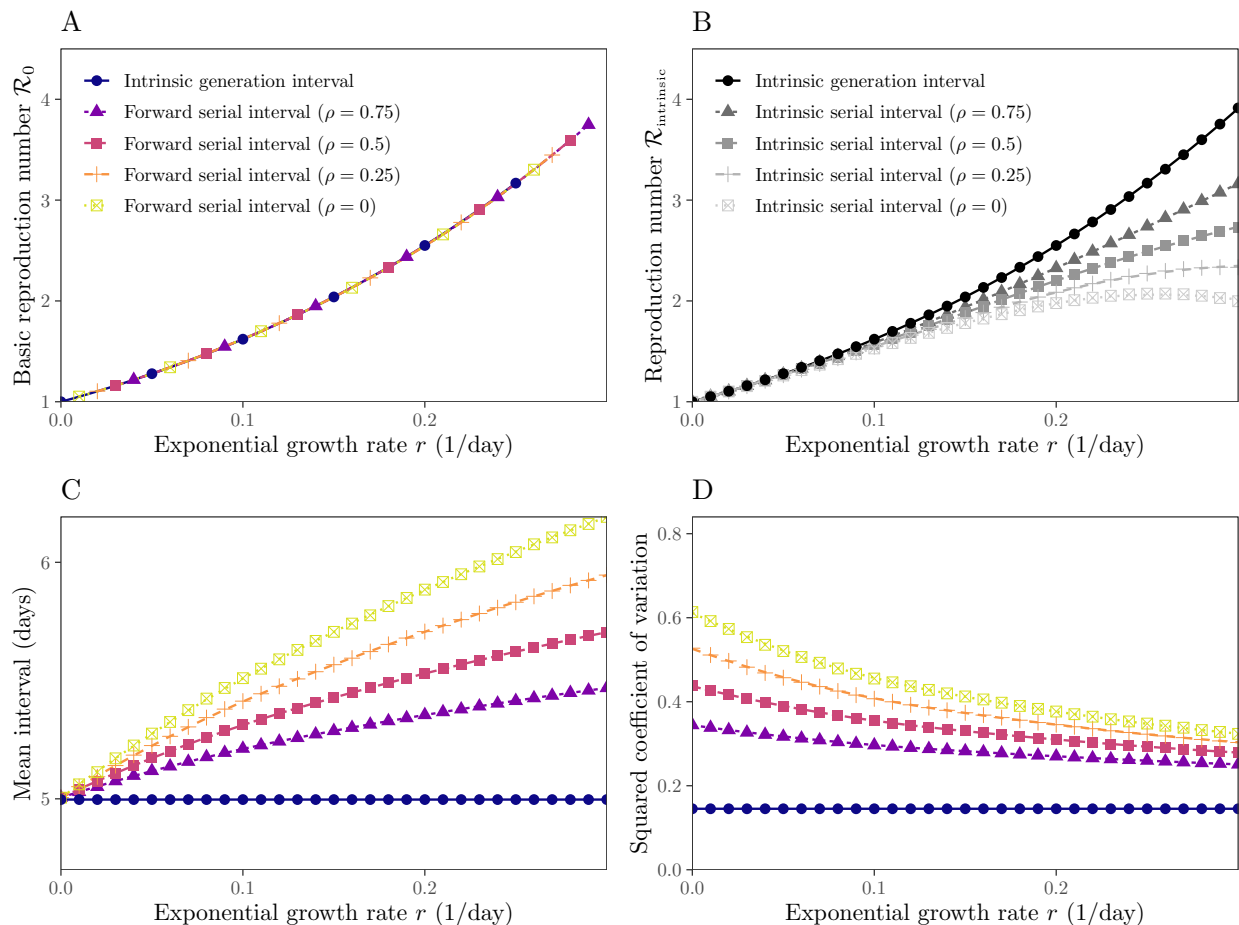


Figure 2: **Estimates of the reproduction number from the exponential growth rate based on serial- and generation-interval distributions.** (A). The initial forward serial-interval distributions give the correct link between the exponential growth rate r and the reproduction number \mathcal{R}_0 , for 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 .

405 Comparing the shapes of the initial forward serial-interval distribution (Eq. (19)) and the
 406 intrinsic generation-interval distribution allows us to better understand how different forward
 407 distributions lead to identical estimates of \mathcal{R}_0 . In general, distributions with higher means
 408 and less variability lead to higher \mathcal{R}_0 for a given r (Wallinga and Lipsitch, 2007; Weitz and
 409 Dushoff, 2015; Park et al., 2019). When incidence is growing exponentially, forward serial
 410 intervals have higher means (Fig. 2C) and squared coefficients of variation (Fig. 2D) than the
 411 intrinsic generation-interval distribution. The effects of higher means (which increase \mathcal{R}_0)

412 exactly cancel those of higher variability (which decrease \mathcal{R}_0). On the other hand, *intrinsic*
413 serial intervals (Eq. (21)) have the same mean (equal to the mean initial forward serial at
414 $r = 0$ in Fig. 2C) as the intrinsic generation intervals but are more variable (also see squared
415 coefficient of variation of the initial forward serial-interval distribution at $r = 0$ in Fig. 2D);
416 therefore, we underestimate \mathcal{R}_0 when we use the intrinsic serial-interval distribution.

417 **3.2 Realized serial-interval distributions during an ongoing epi-** 418 **demic**

419 The initial forward serial-interval distribution captures the exponential growth phase of an
420 epidemic. We now explore how forward and backward serial intervals can vary over the course
421 of an epidemic using deterministic and stochastic simulations based on the renewal equations
422 (see Supplementary Materials) using parameters in Table 1; we further assume $\mathcal{R}_0 = 2.5$
423 to reflect the transmission dynamics of COVID-19 in China (Park et al., 2020). While
424 the forward serial-interval distribution is our primary focus, understanding the differences
425 between the forward and the backward distributions is important because the observed
426 intervals during an ongoing epidemic are often the backward ones: we typically identify
427 infected individuals and ask when and by whom they were infected. Similarly, when we are
428 estimating the incubation period of an individual, we typically observe their symptom onset
429 date and try to estimate when they were infected (e.g., Backer et al. (2020)).

430 Fig. 3 shows the epidemiological dynamics (A) together with the mean forward (B–D) and
431 the mean backward (E–G) delay distributions of a deterministic model based on the renewal
432 equation (Eq. (12)) and of the corresponding stochastic realizations based on individual-
433 based simulations. The mean forward incubation period remains constant throughout an
434 epidemic by assumption (Fig. 3B). The mean forward generation interval decreases slightly
435 when incidence is high, which is when the susceptible population declines rapidly (Fig. 3C;
436 Kenah et al. (2008); Champredon and Dushoff (2015)). In contrast, the mean forward serial
437 interval decreases over time (Fig. 3D).

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

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

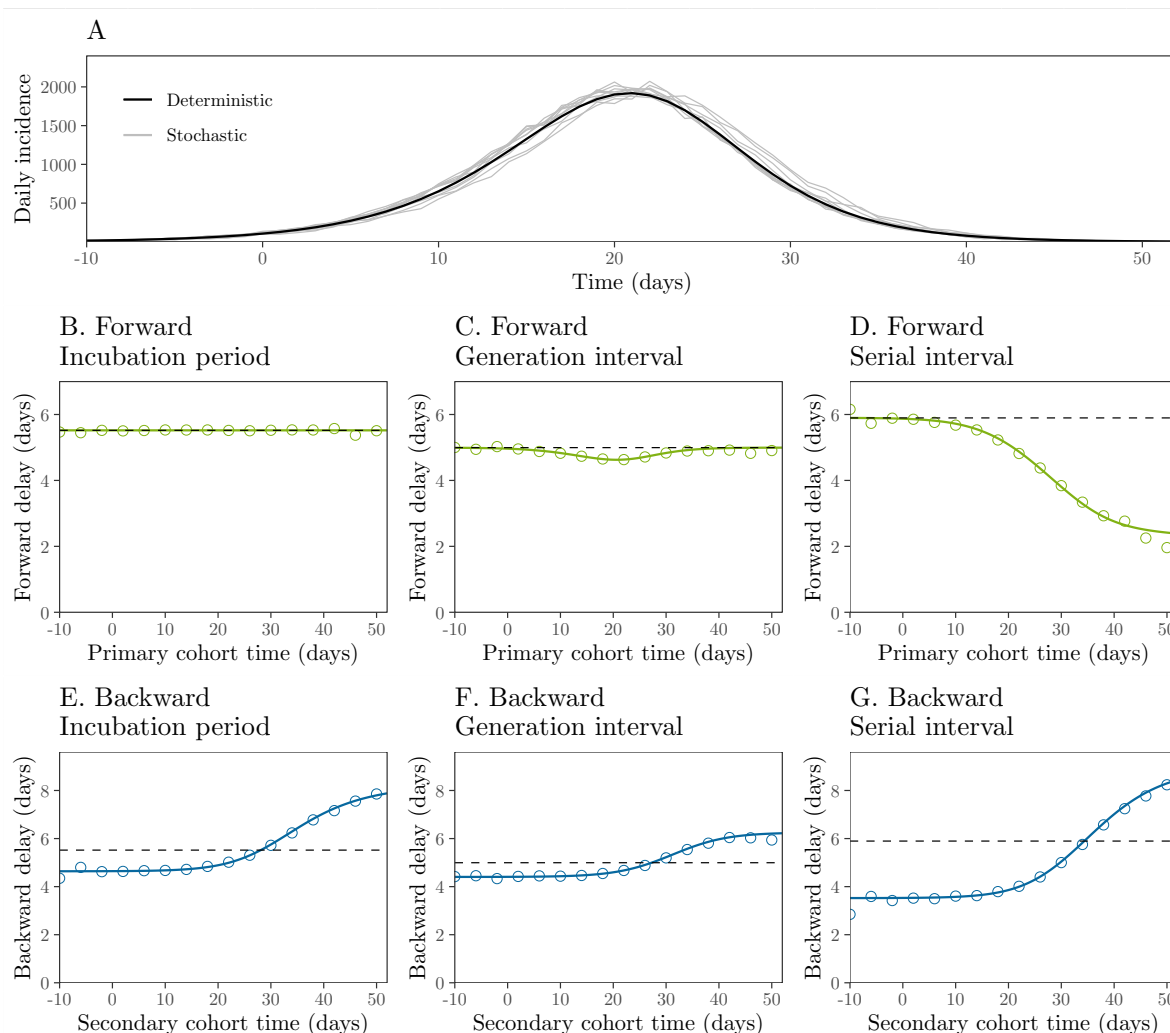


Figure 3: Epidemiological dynamics and changes in mean forward and backward delay distributions. (A) Daily incidence over time. (B–D) Changes in the mean forward incubation period, generation interval, and serial interval. (E–G) Changes in the mean backward incubation period, generation interval, and serial interval. Black (A) and colored (B–G) lines represent the results of a deterministic simulation. Gray lines (A) represent the results of 10 stochastic simulations. Colored points (B–G) represent the average of 10 stochastic simulations. Dashed lines represent the mean initial forward delay. Forward and backward delays are colored according to Fig. 1. In order to remove possible transient dynamics (e.g., left-censoring of time delays and initial stochasticity due to low number of infections), we set $t = 0$ to the first time point when daily incidence is greater than 100. Intrinsic incubation periods and intrinsic generation intervals are assumed to be independent of each other for simplicity. See Supplementary Materials for simulations with correlated incubation periods and generation intervals. See Table 1 for parameter values.

453 3.3 Observed serial-interval distributions

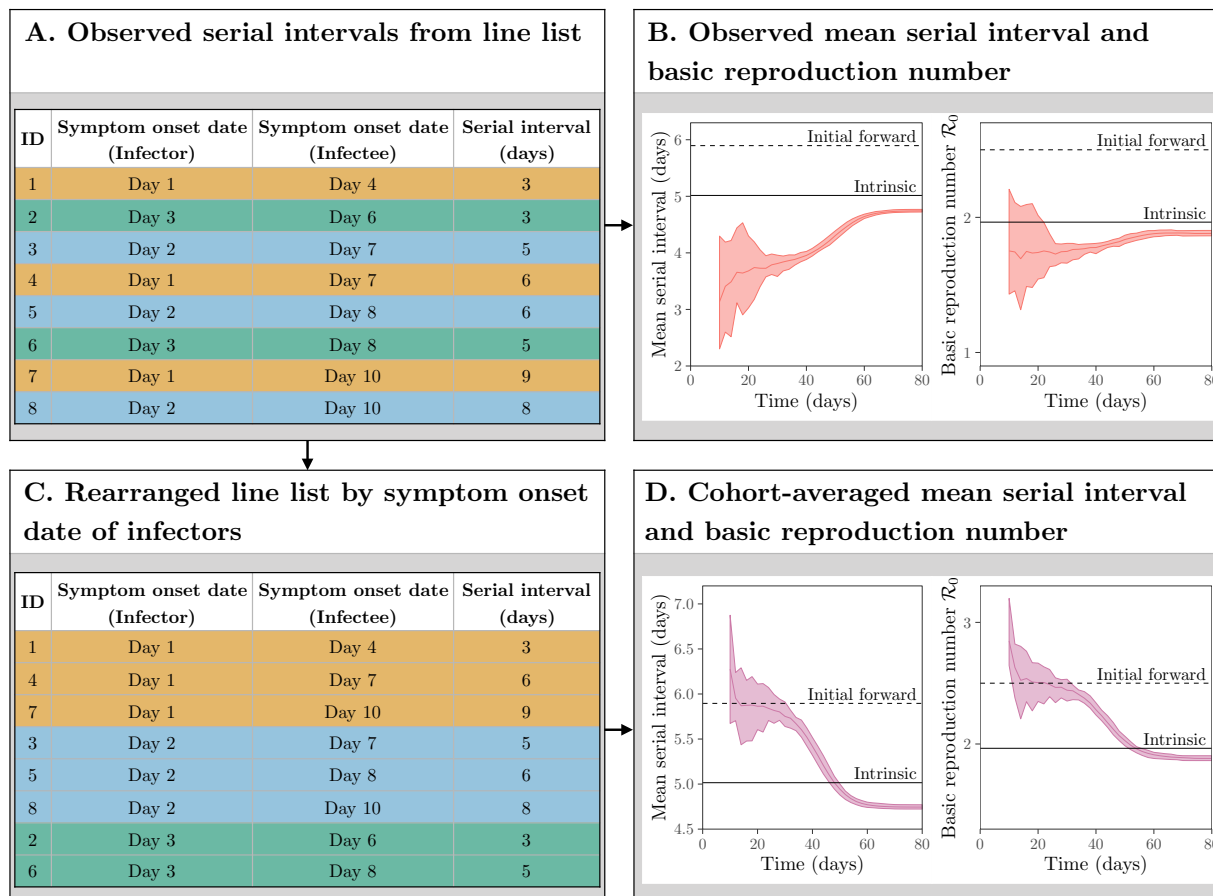


Figure 4: **Estimating the reproduction number from the observed serial intervals.** (A) Schematic representation of line list data collected during an epidemic. (B) Estimates of \mathcal{R}_0 based on all observed serial intervals completed by a given time. (C) Schematic representation of line list 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.

454 Now, we turn to practical issues of estimating the reproduction number from the observed
 455 serial-interval data during an ongoing epidemic. In order to have an unbiased estimate of the
 456 basic reproduction number, we need to estimate the initial forward serial-interval distribution
 457 — i.e., serial intervals based on cohorts of infectors who share the same symptom onset time,
 458 at the early stage of the epidemic. However, researchers typically use all available information
 459 to estimate epidemiological parameters (e.g., aggregating all serial intervals observed until

460 certain time of an epidemic). For example, Thompson et al. (2019) recently suggested that
461 up-to-date serial-interval data are necessary to accurately estimate the reproduction number.
462 We explore the consequences of neglecting changes in the realized serial-interval distribution
463 on estimates of the basic reproduction number.

464 When an epidemic is ongoing, the observed serial intervals are subject to right-censoring
465 because we cannot observe a serial interval if either an infector or an infectee has not yet
466 developed symptoms. For example, if we were to measure serial intervals on Day 8 as in
467 Fig. 4A, we will only be able to observe the first 6 events (ID 1–6). Fig. 4B demonstrates how
468 the effect of right-censoring in the observed serial intervals translates to the underestimation
469 of the basic reproduction number \mathcal{R}_0 in our stochastic simulations (assuming $\mathcal{R}_0 = 2.5$
470 as in Fig. 3). Notably, even if we could observe and aggregate *all* serial intervals across all
471 transmission pairs after the epidemic has ended, we would still underestimate the initial mean
472 forward serial interval (and therefore \mathcal{R}_0), likely by a large amount. The observed serial-
473 interval distribution converges to the intrinsic serial-interval distribution as the incubation
474 periods and generation intervals will no longer be subject to backward biases. In fact,
475 we would even underestimate the intrinsic value slightly due to contraction of the forward
476 generation-interval distribution during the susceptible depletion phase (Fig. 3C). Therefore,
477 aggregated distributions of serial intervals that have been collected throughout different
478 periods of an epidemic must be interpreted with care.

479 Here, we provide a heuristic way of assessing potential biases in the estimate of the
480 mean initial forward serial interval and therefore \mathcal{R}_0 retrospectively. We can rearrange the
481 line list and group observed serial intervals based on the symptom onset date of infectors
482 (Fig. 4C)—as we showed earlier, serial intervals that share the same symptom onset date of
483 a primary case give us the forward serial-interval distribution. Then, we can compare how
484 the shape of the serial-interval distribution (particularly its mean) as well as the estimate
485 of \mathcal{R}_0 change as we incorporate more recent cohorts into the analysis: that is, we analyze
486 observed serial intervals from infectors who became symptomatic before time t and evaluate
487 how the estimates change as we increase t . This approach is analogous to averaging over a
488 set of forward intervals, just as using all information up to a certain time is analogous to
489 averaging over a set of backward intervals (Fig. 4D); the major difference is that we focus
490 on serial intervals that begin in a certain period, rather than those that end in a certain
491 period. During the exponential growth phase, the estimates of the mean serial interval and
492 \mathcal{R}_0 are consistent with the true value (see ‘initial forward’ in Fig. 4B,D); adding more data
493 allows us to make more precise inference during this period. However, the cohort-averaged
494 estimates decrease rapidly soon after the exponential growth period, reflecting changes in the
495 forward serial-interval distributions. This approach allows us to detect dynamical changes
496 in the forward serial-interval distributions and their effect on the estimates of \mathcal{R}_0 .

497 **3.4 Applications to the COVID-19 pandemic**

498 Finally, we re-analyze serial intervals of COVID-19 collected by Du et al. (2020) from main-
499 land China, outside Hubei province, based on 468 transmission events reported between
500 January 21–February 8, 2020. Du et al. (2020) estimated the mean serial interval of 3.96

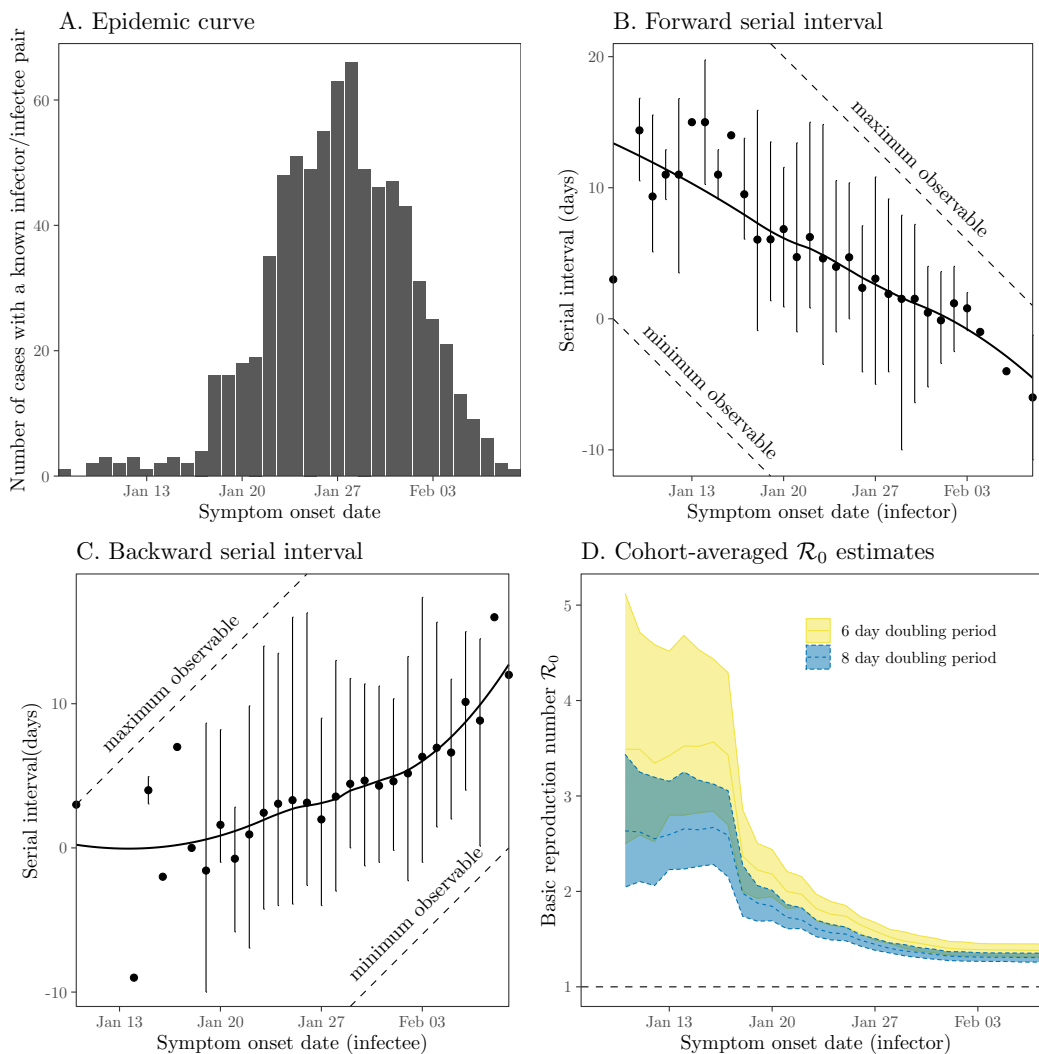


Figure 5: **Observed serial intervals of COVID-19 and cohort-averaged estimates of \mathcal{R} .** (A) Symptom onset dates of all individuals within 468 transmission pairs included in the contact tracing data. (B–C) forward and backward serial intervals over time. Serial interval data have been grouped based on the symptom onset dates of primary (B) and secondary (C) cases. 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. (D) 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).

501 days (95% CI 3.53–4.39 days) and \mathcal{R}_0 of 1.32 (95% CI 1.16–1.48). Fig. 5A shows the distri-

502 bution of symptom onset dates of all individuals within 468 transmission pairs (consisting a
503 total of 752 unique individuals), resembling a COVID-19 epidemic curve in China (cf. Fig. 1
504 in Pan et al. (2020)). In order to quantify changes in serial intervals, we group them by the
505 symptom onset dates of the primary (Fig. 5B) and secondary (Fig. 5C) cases—corresponding
506 to forward and backward serial-interval distributions, respectively—and compute their mean
507 and 95% quantiles. Fig. 5B shows that the mean forward serial interval decreases over time.
508 While the decrease is likely to be affected by the right-censoring (indicated by the close-
509 ness between the quantiles of the observed serial intervals and maximum observable serial
510 intervals), the increase in the proportion of negative serial intervals indicates changes in the
511 forward serial-interval distribution; this proportion is unlikely to be affected by left-censoring
512 (based on the gap between the quantiles of the observed serial intervals and minimum ob-
513 servable serial intervals). The decrease in the mean forward serial interval was probably
514 driven by interventions against spread. Interventions during this time period both decreased
515 (and then reversed) the growth rate of COVID-19 cases — thus increasing the backward
516 incubation period — and also reduced generation intervals, by preventing infections once
517 cases were identified. Both of these would have acted to reduce the forward serial interval.
518 Fig. 5C shows that the mean backward serial interval increased over time, also likely driven
519 directly by the decrease in COVID-19 infections.

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

527 Fig. 5D shows the cohort-averaged estimates of \mathcal{R}_0 , which remain roughly constant until
528 day January 17th and suddenly decreases; this sudden decrease is due to changes in the
529 forward serial intervals consistent with the dynamics seen in our simulations (Fig. 4). The
530 cohort-averaged estimates of \mathcal{R}_0 based on the early forward serial intervals are also consistent
531 with previous estimates of \mathcal{R}_0 of the COVID-19 epidemic in China (Majumder and Mandl,
532 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
533 on a doubling period of 8 or 6 days, respectively, using serial-interval data from infectors
534 who developed symptoms by January 17th. These early cohort-averaged estimates of \mathcal{R}_0
535 are unlikely to be affected by the right-censoring as we expect the degree of right-censoring
536 to be low (Fig. 5A). Therefore, the original \mathcal{R}_0 estimate of 1.32 (95% CI 1.16–1.48), which
537 neglects the changes in the forward serial-interval distribution, underestimates \mathcal{R}_0 by a factor
538 of 2.0–2.6. This example demonstrates the danger of using the observed serial intervals to
539 calculate the reproduction number without organizing serial intervals into cohorts.

540 4 Discussion

541 Generation and serial intervals determine the time scale of disease transmission, and are
542 therefore critical to dynamical modeling of infectious outbreaks. We have shown that the
543 initial *forward* serial-interval distribution — measured from the cohort of infectors who
544 developed symptoms during the exponential growth phase of an epidemic — provides the
545 correct link between the exponential growth rate r and the initial reproduction number \mathcal{R} .
546 In general, the forward serial-interval distributions will not match the intrinsic serial-interval
547 distribution (which has the same mean as the intrinsic generation-interval distribution) be-
548 cause the incubation period of the infectors (conditional on their symptom onset date of
549 the infector) will be subject to backward biases. In particular, the mean forward serial in-
550 terval can decrease over time for COVID-19 as individuals who develop symptoms later in
551 an epidemic are more likely to have longer incubation periods, and therefore have greater
552 opportunity to transmit presymptomatically. Failing to account for these effects can result
553 in underestimation of initial \mathcal{R} .

554 Recently, Ali et al. (2020) also showed that forward serial intervals of COVID-19 decreased
555 through time in China. They grouped serial intervals by the symptom onset date of infectors
556 across 14-day periods and found that the mean forward serial interval decreased from 7.8
557 days to 2.6 days. While they attributed the decrease in serial intervals to reduction of the
558 isolation delay, their regression analysis showed that isolation delays explain only 51.5% of
559 the variation in serial intervals (they could explain up to 72% of the variance by including
560 other intervention measures). Our framework provides an explanation for the remaining
561 variation: changes in the backward incubation period during the decreasing phase of an
562 epidemic act to further shorten serial intervals due to increased amount of presymptomatic
563 transmission (even in the absence of nonpharmaceutical interventions). Isolation delays and
564 other intervention measures affect the amount of onward transmission, and therefore the
565 distribution of realized (forward) generation intervals. They therefore are not expected to
566 explain all the variation in forward serial intervals, since these additionally depend on both
567 the backward incubation period of the infector and the forward incubation period of the
568 infectee (Fig. 1B).

569 Our results support the use of serial-interval distributions for calculating the \mathcal{R} during
570 the exponential growth phase, but they also reveal gaps in current practices in incorpo-
571 rating serial-interval distributions into outbreak analyses. For example, Thompson et al.
572 (2019) recently emphasized the importance of using up-to-date serial-interval data for ac-
573 curate estimation of time-varying reproduction numbers. However, our results show that if
574 observational biases in the forward serial interval through time are not accounted for, using
575 up-to-date serial-interval data can actually exacerbate the underestimation of \mathcal{R} in the ini-
576 tial growth phase of an outbreak. Future studies should explore how neglecting changes in
577 the forward serial-interval distribution can affect the estimates of \mathcal{R} beyond the exponen-
578 tial growth phase, and potentially re-assess existing estimates of \mathcal{R} . We also suggest that
579 modelers should aim to characterize spatiotemporal variation in forward serial-interval dis-
580 tributions. These modeling approaches should be coupled with epidemiological investigation
581 through contact tracing. Going forward, an additional advantage of early, intensive contact

582 tracing of emerging diseases is that it provides the best information to characterize the initial
583 forward serial-interval distribution.

584 Our study underlines the fact that the serial-interval distribution depends not only on
585 the generation-interval and incubation-period distributions, but also on the correlation be-
586 tween their duration in a given individual. Here, we use a bivariate lognormal distribution
587 to capture these correlations phenomenologically and to show that realized serial intervals
588 can decrease over time in the context of COVID-19. Although their true correlation will
589 depend on viral load dynamics, we expect our conclusions about decreasing serial intervals
590 of COVID-19 to be robust, as individuals with longer incubation periods will generally have
591 a longer time window to transmit before symptom onset. In general, the impact of increasing
592 backward incubation periods on the forward serial intervals are likely to be disease-specific—
593 for example, we show in Supplementary Materials that the initial forward serial-interval dis-
594 tribution can be equivalent to the intrinsic generation-interval distribution, regardless of the
595 growth rate r , due to independence between the incubation period and time from symptom
596 onset to transmission and the lack of presymptomatic transmission. Future studies trying
597 to interpret realized serial intervals should consider carefully the joint distribution between
598 the generation intervals and incubation periods.

599 In closing, we lay out a few practical principles for analyzing and interpreting serial-
600 interval data. First, serial intervals should be cohorted based on the symptom onset date of
601 the infector (and not of the infectee) whenever possible. Previous studies have often regarded
602 serial intervals as an intrinsic quantity, having the same mean as the intrinsic generation
603 interval (Svensson, 2007; Klinkenberg and Nishiura, 2011; Champredon et al., 2018; Britton
604 and Scalia Tomba, 2019), but the distribution (and the mean) of observed serial intervals
605 differs from this expectation, and changes through time due to epidemic dynamics. Second,
606 aggregating serial intervals across different cohorts and epidemic periods should be avoided
607 because the realized serial-interval distribution can be subject to different censoring and
608 epidemiological biases: Even when *all* realized serial intervals can be observed throughout
609 an unmitigated epidemic, we do not obtain the intrinsic serial interval distribution due to
610 susceptible depletion (Fig. 4). Third, applying serial-interval information across epidemics of
611 a given disease should be done with care, because serial intervals are epidemic-specific, rather
612 than disease-specific. Finally, serial-interval data should be accompanied by a trajectory of
613 the epidemic curve, whenever possible, to provide epidemiological context. In practice,
614 these recommendations will sometimes be hard to follow, due to limited data about serial
615 intervals, but these issues should be kept in mind when interpreting serial-interval data to
616 inform transmission dynamics.

617 More broadly, our study underlines the importance of carefully defining measured epi-
618 demiological time distributions. Previous studies have shown the importance of forward vs.
619 backward measurement of generation intervals (Nishiura, 2010; Champredon and Dushoff,
620 2015; Britton and Scalia Tomba, 2019); we generalize these ideas and show that they apply to
621 other epidemiological distributions. Some studies during the early phases of the COVID-19
622 epidemics have tried to correct for the backward biases (Verity et al., 2020), but changes in
623 the backward delay distributions due to changing cohort sizes are expected to be a pervasive
624 feature of outbreak dynamics. Cohorting epidemiological delays by the primary event time

625 can help avoid backward biases (although censoring biases can still exist) as well as detect
626 potential changes in the distribution.

627 Here, we assume that all individuals develop symptoms and that the entire transmission
628 process, including all relevant epidemiological delays, is known exactly. In practice, iden-
629 tifying who infected whom is difficult in general, and asymptomatic and presymptomatic
630 transmission of COVID-19 exacerbates this difficulty (Bai et al., 2020; He et al., 2020; Wei,
631 2020). Biases in the observed serial intervals will necessarily bias the estimates of \mathcal{R} . Fur-
632 thermore, when one of the individuals in a transmission pair is asymptomatic, there is no
633 symptom-based serial interval. Neglecting the time scale of asymptomatic transmission may
634 also bias the estimates of \mathcal{R} (Park et al., 2020).

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

643 **Data availability**

644 All data and code are stored in a publicly available GitHub repository ([https://github.](https://github.com/parksw3/serial)
645 [com/parksw3/serial](https://github.com/parksw3/serial)).

646 **Competing interests**

647 We declare no competing interests.

648 5 Supplementary Materials

649 5.1 Deterministic simulation

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

$$\begin{aligned} i(t) &= \mathcal{R}_0 S(t - \Delta t) \sum_{m=1}^{m_{\max}} i(t - m\Delta t) \hat{g}(m\Delta t) \\ S(t) &= S(t - \Delta t) - i(t) \end{aligned} \quad (22)$$

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

$$\hat{g}(m\Delta t) = \frac{g(m\Delta t)}{\sum_{i=1}^{\ell} g(m\Delta t)}, \quad m = 1, \dots, m_{\max}. \quad (23)$$

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

$$\hat{\ell}(m\Delta t) = \frac{\ell(m\Delta t)}{\sum_{i=1}^{\ell} \ell(m\Delta t)}, \quad m = 1, \dots, m_{\max}, \quad (24)$$

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

$$\hat{h}(m\Delta t, n\Delta t) = \hat{\ell}(m\Delta t) \hat{g}(n\Delta t), \quad m, n = 1, \dots, m_{\max}. \quad (25)$$

658 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:

$$\begin{aligned} i(m\Delta t) &= C \exp(rm\Delta t), \quad m = 1, \dots, m_{\max} \\ S(m\Delta t) &= N - \sum_{n=1}^m i(n\Delta t), \quad m = 1, \dots, m_{\max} \end{aligned} \quad (26)$$

659 where C is chosen such that $\sum_{n=1}^{m_{\max}} i(n\Delta t) = 10$. These initial conditions allow the model
660 to follow exponential growth from time $\Delta t(m_{\max} + 1)$ without any transient behaviors.

661 5.2 Stochastic simulation

662 We run stochastic simulations of the renewal equation model using an individual-based
663 model on a fully connected network (i.e., homogeneous population) based on the Gillespie
664 algorithm that we developed earlier (Park et al., 2020). First, we initialize an epidemic with
665 $I(0)$ infected individuals (nodes) in a fully connected network of size N . For each initially
666 infected individual, we draw number of infectious contacts from a Poisson distribution with
667 the mean of \mathcal{R}_0 and the corresponding generation intervals for each contact from a log-
668 normal distribution (Table 1). Contactees are uniformly sampled from the total population.

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

676 5.3 Linking r and \mathcal{R}_0 using serial-interval distributions

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

$$\frac{1}{\mathcal{R}_0} = \int_0^{\infty} \exp(-r\tau)g(\tau) d\tau. \quad (27)$$

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

$$\frac{1}{\mathcal{R}_0} = \int_{-\infty}^{\infty} \exp(-r\tau)f_0(\tau) d\tau. \quad (28)$$

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

$$f_0(\tau) = \frac{1}{\phi} \int_{-\infty}^0 \int_{\alpha_1}^{\tau} \exp(r\alpha_1)h(-\alpha_1, \alpha_2 - \alpha_1)\ell(\tau - \alpha_2) d\alpha_2 d\alpha_1, \quad (29)$$

683 where h is the joint probability distribution describing the intrinsic generation-interval dis-
 684 tribution g and the intrinsic incubation period distribution ℓ (see Eq. (15) in the main text),
 685 and the normalization constant ϕ is determined by the requirement that $\int_{-\infty}^{\infty} f_0(\tau) d\tau = 1$.

In order to verify Eq. (28), we first rewrite the integral in Eq. (29) by substituting $-\alpha_1$
 for α_1 , and then changing the order of integration:

$$\begin{aligned} f_0(\tau) &= \frac{1}{\phi} \int_0^{\infty} \int_{-\alpha_1}^{\tau} \exp(-r\alpha_1)h(\alpha_1, \alpha_2 + \alpha_1)\ell(\tau - \alpha_2) d\alpha_2 d\alpha_1, \\ &= \frac{1}{\phi} \int_{-\infty}^{\tau} \int_{\max(0, -\alpha_2)}^{\infty} \exp(-r\alpha_1)h(\alpha_1, \alpha_2 + \alpha_1)\ell(\tau - \alpha_2) d\alpha_1 d\alpha_2. \end{aligned} \quad (30)$$

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

$$z(\alpha_2) = \int_{\max(0, -\alpha_2)}^{\infty} \exp(-r\alpha_1)h(\alpha_1, \alpha_2 + \alpha_1) d\alpha_1. \quad (31)$$

687 Substituting $z(\alpha_2)$ into Eq. (30) we obtain:

$$f_0(\tau) = \frac{1}{\phi} \int_{-\infty}^{\tau} z(\alpha_2)\ell(\tau - \alpha_2) d\alpha_2, \quad (32)$$

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

$$\hat{z}(\alpha_2) = \frac{z(\alpha_2)}{\int_{-\infty}^{\infty} z(x) dx}, \quad (33)$$

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

$$f_0(\tau) = \frac{1}{\hat{\phi}} \int_{-\infty}^{\tau} \hat{z}(\alpha_2) \ell(\tau - \alpha_2) d\alpha_2, \quad (34)$$

691 where $\hat{\phi} = \phi / \int_{-\infty}^{\infty} z(x) dx$.

692 Since the right hand side of Eq. (28) is also a Laplace transform of $f_0 = \hat{z} * \ell$, we can
693 express it as the product of Laplace transforms of \hat{z} and ℓ :

$$\int_{-\infty}^{\infty} \exp(-r\tau) f_0(\tau) d\tau = \int_{-\infty}^{\infty} \exp(-r\tau) \hat{z}(\tau) d\tau \int_0^{\infty} \exp(-r\tau) \ell(\tau) d\tau. \quad (35)$$

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

$$\begin{aligned} \int_{-\infty}^{\infty} z(\alpha_2) d\alpha_2 &= \int_{-\infty}^{\infty} \int_{\max(0, -\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) d\alpha_1 d\alpha_2, \\ &= \int_0^{\infty} \int_{-\alpha_1}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) d\alpha_2 d\alpha_1. \end{aligned} \quad (36)$$

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

$$\int_{-\infty}^{\infty} z(\alpha_2) d\alpha_2 = \int_0^{\infty} \exp(-r\alpha_1) \ell(\alpha_1) d\alpha_1. \quad (37)$$

695 Then, we have:

$$\hat{z}(\alpha_2) = \frac{\int_{\max(0, -\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) d\alpha_1}{\int_0^{\infty} \exp(-r\alpha_1) \ell(\alpha_1) d\alpha_1}. \quad (38)$$

696 Substituting the expression into Eq. (35), we have:

$$\int_{-\infty}^{\infty} \exp(-r\tau) f_0(\tau) d\tau = \int_{-\infty}^{\infty} \exp(-r\alpha_2) \int_{\max(0, -\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) d\alpha_1 d\alpha_2. \quad (39)$$

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

$$g(\tau) = \int_0^{\infty} h(x, \tau) dx. \quad (40)$$

We can then substitute $\tau = \alpha_1 + \alpha_2$ into Eq. (39) and apply change of variables to obtain:

$$\int_{-\infty}^{\infty} \exp(-r\tau) f_0(\tau) d\tau \quad (41)$$

$$= \int_{-\infty}^{\infty} \exp(-r\alpha_2) \int_{\max(0, -\alpha_2)}^{\infty} \exp(-r\alpha_1) h(\alpha_1, \alpha_2 + \alpha_1) d\alpha_1 d\alpha_2 \quad (42)$$

$$= \int_0^{\infty} \int_0^{\infty} \exp(-r\tau) h(\alpha_1, \tau) d\alpha_1 d\tau \quad (43)$$

$$= \int_0^{\infty} \exp(-r\tau) g(\tau) d\tau = \frac{1}{\mathcal{R}_0} \quad (44)$$

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

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:

$$F_i = -X_{1,i} + (G_i|X_{1,i}) + X_{2,i}, \quad (45)$$

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:

$$\mathcal{R}_0 = \frac{1}{\frac{1}{N} \sum_{i=1}^N \exp(-rF_i)}. \quad (46)$$

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):

$$\mathcal{R}_{\text{intrinsic}} = \frac{1}{\frac{1}{N} \sum_{i=1}^N \exp(-rQ_i)}, \quad (47)$$

where

$$Q_i = -Y_i + (G_i|Y_i) + X_{2,i}. \quad (48)$$

721 5.5 Applications: SEIR model

Consider a Susceptible-Exposed-Infectious-Recovered model:

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dE}{dt} &= \beta SI - \gamma_E E \\ \frac{dI}{dt} &= \gamma_E E - \gamma_I I \\ \frac{dR}{dt} &= \gamma_I I\end{aligned}\tag{49}$$

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

726 The joint probability distribution of the intrinsic incubation periods and intrinsic gener-
727 ation intervals for this model can be written as:

$$h(x, \tau) = \begin{cases} 0 & x > \tau \\ \gamma_I \gamma_E \exp(-\gamma_I(\tau - x) - \gamma_E x) & x \leq \tau \end{cases}\tag{50}$$

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

$$\begin{aligned}g(\tau) &= \int_0^\tau h(x, \tau) dx \\ &= \frac{\gamma_I \gamma_E}{\gamma_E - \gamma_I} (\exp(-\gamma_I \tau) - \exp(-\gamma_E \tau))\end{aligned}\tag{51}$$

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

$$\begin{aligned}f_0(\tau) &\propto \int_{-\infty}^0 \int_0^\tau \exp(r\alpha_1) h(-\alpha_1, \alpha_2 - \alpha_1) \ell(\tau - \alpha_2) d\alpha_2 d\alpha_1 \\ &\propto \int_{-\infty}^0 \int_0^\tau \exp(r\alpha_1) \exp(-\gamma_I \alpha_2 + \gamma_E \alpha_1) \exp(-\gamma_E(\tau - \alpha_2)) d\alpha_2 d\alpha_1 \\ &\propto \exp(-\gamma_E \tau) \int_{-\infty}^0 \int_0^\tau \exp((\gamma_E - \gamma_I)\alpha_2) \exp((r + \gamma_E)\alpha_1) d\alpha_2 d\alpha_1 \\ &\propto (\exp(-\gamma_I \tau) - \exp(-\gamma_E \tau)) \int_{-\infty}^0 \exp((r + \gamma_E)\alpha_1) d\alpha_1 \\ &\propto \exp(-\gamma_I \tau) - \exp(-\gamma_E \tau)\end{aligned}\tag{52}$$

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

730 5.6 Simulations with correlated intrinsic incubation periods and 731 intrinsic generation intervals.

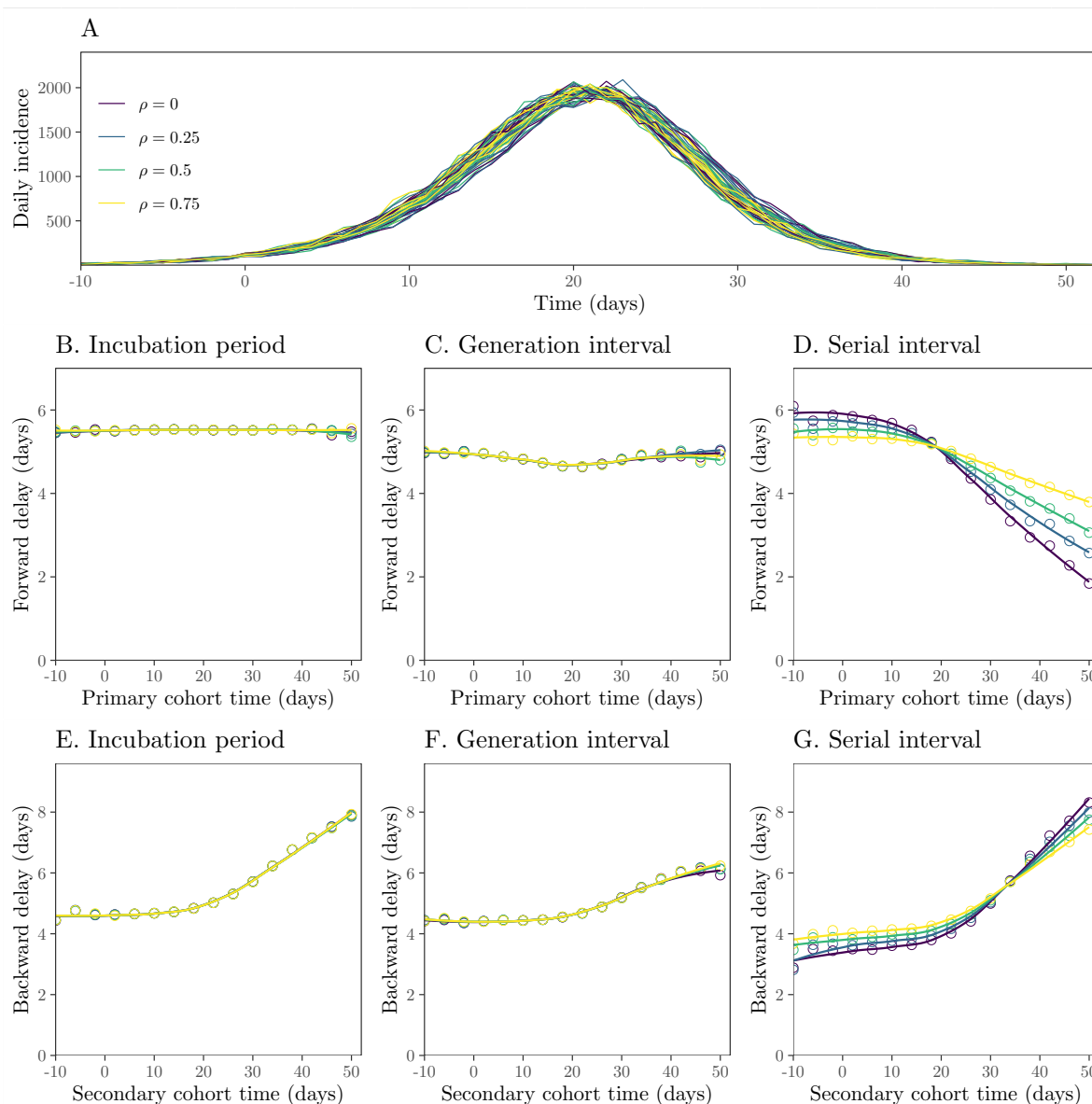


Figure S1: Epidemiological dynamics and changes in mean forward and backward delay distributions. (A) Daily incidence over time. (B–D) Changes in the mean forward incubation period, generation interval, and serial interval. (E–G) Changes in the mean backward incubation period, generation interval, and serial interval. Intrinsic incubation periods and intrinsic generation intervals are modeled using a correlated bivariate lognormal distribution; therefore, generation intervals are drawn from the corresponding conditional distributions (given a incubation period), instead of the marginal distribution. Higher correlation reduces the amount of changes in the mean forward serial interval because shorter (longer) backward incubation periods of infectors during the increasing (decreasing) phase of an epidemic are associated with shorter (longer) forward generation intervals. See Figure 3 in the main text for a detailed description.

732 References

- 733 Abbott, S., J. Hellewell, J. D. Munday, J. Y. Chun, R. N. Thompson, N. I. Bosse,
734 Y.-W. D. Chan, T. W. Russell, C. I. Jarvis, CMMID nCov working group, S. Flasche,
735 A. J. Kucharski, R. Eggo, and S. Funk (2020). Temporal variation in transmis-
736 sion during the COVID-19 outbreak. [https://cmmid.github.io/topics/covid19/
737 current-patterns-transmission/global-time-varying-transmission.html](https://cmmid.github.io/topics/covid19/current-patterns-transmission/global-time-varying-transmission.html). Ac-
738 cessed April 20, 2020.
- 739 Aldis, G. and M. Roberts (2005). An integral equation model for the control of a smallpox
740 outbreak. *Mathematical biosciences* 195(1), 1–22.
- 741 Ali, S. T., L. Wang, E. H. Lau, X.-K. Xu, Z. Du, Y. Wu, G. M. Leung, and B. J. Cowling
742 (2020). Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical
743 interventions. *Science* 369(6507), 1106–1109.
- 744 Anderson, R. M., H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth (2020). How
745 will country-based mitigation measures influence the course of the COVID-19 epidemic?
746 *The Lancet* 395(10228), 931–934.
- 747 Anderson, R. M. and R. M. May (1991). *Infectious diseases of humans: dynamics and
748 control*. Oxford university press.
- 749 Backer, J. A., D. Klinkenberg, and J. Wallinga (2020). Incubation period of 2019 novel
750 coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January
751 2020. *Eurosurveillance* 25(5).
- 752 Bai, Y., L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, and M. Wang (2020). Presumed
753 asymptomatic carrier transmission of COVID-19. *Jama* 323(14), 1406–1407.
- 754 Britton, T. and G. Scalia Tomba (2019). Estimation in emerging epidemics: Biases and
755 remedies. *Journal of the Royal Society Interface* 16(150), 20180670.
- 756 Champredon, D. and J. Dushoff (2015). Intrinsic and realized generation intervals in
757 infectious-disease transmission. *Proceedings of the Royal Society B: Biological Sci-
758 ences* 282(1821), 20152026.
- 759 Champredon, D., J. Dushoff, and D. J. D. Earn (2018). Equivalence of the Erlang-distributed
760 SEIR epidemic model and the renewal equation. *SIAM Journal on Applied Mathemat-
761 ics* 78(6), 3258–3278.
- 762 Diekmann, O. and J. A. P. Heesterbeek (2000). *Mathematical epidemiology of infectious
763 diseases: model building, analysis and interpretation*, Volume 5. John Wiley & Sons.
- 764 Diekmann, O., J. A. P. Heesterbeek, and J. A. Metz (1990). On the definition and the compu-
765 tation of the basic reproduction ratio \mathcal{R}_0 in models for infectious diseases in heterogeneous
766 populations. *Journal of mathematical biology* 28(4), 365–382.

- 767 Du, Z., X. Xu, Y. Wu, L. Wang, B. J. Cowling, and L. A. Meyers (2020). Serial Inter-
768 val of COVID-19 among Publicly Reported Confirmed Cases. *Emerging Infectious Dis-*
769 *eases* 26(6).
- 770 Ferretti, L., C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker,
771 D. Bonsall, and C. Fraser (2020). Quantifying SARS-CoV-2 transmission suggests epidemic
772 control with digital contact tracing. *Science* 368(6491).
- 773 Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker,
774 H. Zhu, T. Berah, J. W. Eaton, et al. (2020). Estimating the effects of non-pharmaceutical
775 interventions on covid-19 in europe. *Nature* 584(7820), 257–261.
- 776 Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging
777 epidemic. *PloS one* 2(8).
- 778 Ganyani, T., C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens (2020).
779 Estimating the generation interval for coronavirus disease (COVID-19) based on symptom
780 onset data, March 2020. *Eurosurveillance* 25(17), 2000257.
- 781 Gostic, K. M., L. McGough, E. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, R. Kahn,
782 R. Niehus, J. A. Hay, P. M. De Salazar, et al. (2020). Practical considerations for measuring
783 the effective reproductive number, \mathcal{R}_t . *medRxiv*.
- 784 He, X., E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan,
785 X. Tan, et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-
786 19. *Nature Medicine*, 1–4.
- 787 Heesterbeek, J. and K. Dietz (1996). The concept of \mathcal{R}_0 in epidemic theory. *Statistica*
788 *Neerlandica* 50(1), 89–110.
- 789 Hellewell, J., S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday,
790 A. J. Kucharski, W. J. Edmunds, F. Sun, S. Flasche, B. J. Quilty, N. Davies, Y. Liu,
791 S. Clifford, P. Klepac, M. Jit, C. Diamond, H. Gibbs, K. [van Zandvoort], S. Funk, and
792 R. M. Eggo (2020). Feasibility of controlling COVID-19 outbreaks by isolation of cases
793 and contacts. *The Lancet Global Health* 8(4), e488 – e496.
- 794 Kenah, E., M. Lipsitch, and J. M. Robins (2008). Generation interval contraction and
795 epidemic data analysis. *Mathematical biosciences* 213(1), 71–79.
- 796 Klinkenberg, D. and H. Nishiura (2011). The correlation between infectivity and incuba-
797 tion period of measles, estimated from households with two cases. *Journal of theoretical*
798 *biology* 284(1), 52–60.
- 799 Lauer, S. A., K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman,
800 N. G. Reich, and J. Lessler (2020). The incubation period of coronavirus disease 2019
801 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals*
802 *of internal medicine* 172(9), 577–582.

- 803 Li, Q., X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y.
804 Wong, et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-
805 infected pneumonia. *New England Journal of Medicine*.
- 806 Ma, J., J. Dushoff, B. M. Bolker, and D. J. D. Earn (2014). Estimating initial epidemic
807 growth rates. *Bulletin of mathematical biology* 76(1), 245–260.
- 808 Majumder, M. S. and K. D. Mandl (2020). Early in the epidemic: impact of preprints on
809 global discourse about covid-19 transmissibility. *The Lancet Global Health* 8(5), e627–
810 e630.
- 811 Mills, C. E., J. M. Robins, and M. Lipsitch (2004). Transmissibility of 1918 pandemic
812 influenza. *Nature* 432(7019), 904–906.
- 813 Nishiura, H. (2010). Time variations in the generation time of an infectious disease: im-
814 plications for sampling to appropriately quantify transmission potential. *Mathematical*
815 *Biosciences & Engineering* 7(4), 851–869.
- 816 Nishiura, H., N. M. Linton, and A. R. Akhmetzhanov (2020). Serial interval of novel coron-
817 avirus (COVID-19) infections. *International Journal of Infectious Diseases*.
- 818 Pan, A., L. Liu, C. Wang, H. Guo, X. Hao, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, S. Wei,
819 and T. Wu (2020, 04). Association of Public Health Interventions With the Epidemiology
820 of the COVID-19 Outbreak in Wuhan, China. *JAMA*.
- 821 Park, S. W., B. M. Bolker, D. Champredon, D. J. D. Earn, M. Li, J. S. Weitz, B. T. Grenfell,
822 and J. Dushoff (2020). Reconciling early-outbreak estimates of the basic reproductive
823 number and its uncertainty: framework and applications to the novel coronavirus (SARS-
824 CoV-2) outbreak. *Journal of The Royal Society Interface* 17(168), 20200144.
- 825 Park, S. W., D. Champredon, and J. Dushoff (2020). Inferring generation-interval distribu-
826 tions from contact-tracing data. *Journal of the Royal Society Interface* 17(167), 20190719.
- 827 Park, S. W., D. Champredon, J. S. Weitz, and J. Dushoff (2019). A practical generation-
828 interval-based approach to inferring the strength of epidemics from their speed. *Epi-*
829 *demics* 27, 12–18.
- 830 Park, S. W., D. M. Cornforth, J. Dushoff, and J. S. Weitz (2020). The time scale of asymp-
831 tomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak.
832 *Epidemics*, 100392.
- 833 Roberts, M. (2004). Modelling strategies for minimizing the impact of an imported ex-
834 otic infection. *Proceedings of the Royal Society of London. Series B: Biological Sci-*
835 *ences* 271(1555), 2411–2415.
- 836 Roberts, M. and J. Heesterbeek (2007). Model-consistent estimation of the basic repro-
837 duction number from the incidence of an emerging infection. *Journal of mathematical*
838 *biology* 55(5-6), 803.

- 839 Svensson, Å. (2007). A note on generation times in epidemic models. *Mathematical bio-*
840 *sciences* 208(1), 300–311.
- 841 te Beest, D. E., J. Wallinga, T. Donker, and M. van Boven (2013). Estimating the generation
842 interval of influenza A (H1N1) in a range of social settings. *Epidemiology*, 244–250.
- 843 Thompson, R., J. Stockwin, R. van Gaalen, J. Polonsky, Z. Kamvar, P. Demarsh,
844 E. Dahlqwist, S. Li, E. Miguel, T. Jombart, et al. (2019). Improved inference of time-
845 varying reproduction numbers during infectious disease outbreaks. *Epidemics* 29, 100356.
- 846 Tindale, L., M. Coombe, J. E. Stockdale, E. Garlock, W. Y. V. Lau, M. Saraswat,
847 Y.-H. B. Lee, L. Zhang, D. Chen, J. Wallinga, et al. (2020). Transmis-
848 sion interval estimates suggest pre-symptomatic spread of COVID-19. *medRxiv*.
849 <https://doi.org/10.1101/2020.03.03.20029983>.
- 850 Verity, R., L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-
851 Dannenburg, H. Thompson, P. G. Walker, H. Fu, et al. (2020). Estimates of the severity
852 of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases*.
- 853 Wallinga, J. and M. Lipsitch (2007). How generation intervals shape the relationship between
854 growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological*
855 *Sciences* 274(1609), 599–604.
- 856 Wei, W. E. (2020). Presymptomatic Transmission of SARS-CoV-2—Singapore, January
857 23–March 16, 2020. *MMWR. Morbidity and Mortality Weekly Report* 69.
- 858 Weitz, J. S. and J. Dushoff (2015). Modeling post-death transmission of Ebola: challenges
859 for inference and opportunities for control. *Scientific reports* 5, 8751.
- 860 Wu, J. T., K. Leung, and G. M. Leung (2020). Nowcasting and forecasting the potential do-
861 mestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China:
862 a modelling study. *The Lancet* 395(10225), 689–697.
- 863 Zhang, J., M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi,
864 X. Chen, et al. (2020). Evolving epidemiology and transmission dynamics of coronavirus
865 disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet*
866 *Infectious Diseases*.
- 867 Zhao, S., P. Cao, D. Gao, Z. Zhuang, Y. Cai, J. Ran, M. K. Chong, K. Wang, Y. Lou,
868 W. Wang, et al. (2020). Serial interval in determining the estimation of reproduction
869 number of the novel coronavirus disease (COVID-19) during the early outbreak. *Journal*
870 *of Travel Medicine* 27(3), taaa033.
- 871 Zhao, S., D. Gao, Z. Zhuang, M. Chong, Y. Cai, J. Ran, P. Cao, K. Wang, Y. Lou, W. Wang,
872 et al. (2020). Estimating the serial interval of the novel coronavirus disease (COVID-19):
873 A statistical analysis using the public data in Hong Kong from January 16 to February
874 15, 2020. *medRxiv*. <https://doi.org/10.1101/2020.02.21.20026559>.