Report

Author for correspondence:
Sonja Lehtinen
e-mail: sonja.lehtinen@env.ethz.ch

THE ROYAL SOCIETY
PUBLISHING

On the relationship between serial interval, infectiousness profile and generation time

Sonja Lehtinen, Peter Ashcroft and Sebastian Bonhoeffer

Institute for Integrative Biology, Department of Environmental System Science, ETH Zürich, Zürich, Switzerland

SL, 0000-0002-4236-828X; PA, 0000-0003-4067-7692; SB, 0000-0001-8052-3925

The timing of transmission plays a key role in the dynamics and controllability of an epidemic. However, observing generation times—the time interval between the infection of an infector and an infectee in a transmission pair—requires data on infection times, which are generally unknown. The timing of symptom onset is more easily observed; generation times are therefore often estimated based on serial intervals—the time interval between symptom onset of an infector and an infectee. This estimation follows one of two approaches: (i) approximating the generation time distribution by the serial interval distribution or (ii) deriving the generation time distribution from the serial interval and incubation period—the time interval between infection and symptom onset in a single individual—distributions. These two approaches make different—and not always explicitly stated—assumptions about the relationship between infectiousness and symptoms, resulting in different generation time distributions with the same mean but unequal variances. Here, we clarify the assumptions that each approach makes and show that neither set of assumptions is plausible for most pathogens. However, the variances of the generation time distribution derived under each assumption can reasonably be considered as upper (approximation with serial interval) and lower (derivation from serial interval) bounds. Thus, we suggest a pragmatic solution is to use both approaches and treat these as edge cases in downstream analysis. We discuss the impact of the variance of the generation time distribution on the controllability of an epidemic through strategies based on contact tracing, and we show that underestimating this variance is likely to overestimate controllability.

1. Background

1.1. Motivation

Estimating the generation time (the timing between successive infections in a transmission chain) distribution in an emerging epidemic is both extremely important and extremely challenging. Generation time is key to assessing the controllability of the epidemic: it determines the relationship between the basic reproductive number R_0 and the epidemic's growth rate [1,2], as well as how much delays in the isolation of infected individuals impede epidemic control [3,4]. However, the timing of transmission events is often unknown. The distribution of generation times is therefore typically estimated based on the timing of symptom onset, which requires assumptions about the relationship between infectiousness and symptoms. These assumptions are not always explicitly stated and their plausibility is rarely discussed. Here, we illustrate how assumptions about infectiousness and symptom onset affect the relationship between the generation time and serial interval distributions, and the implications this has for assessing epidemic controllability.
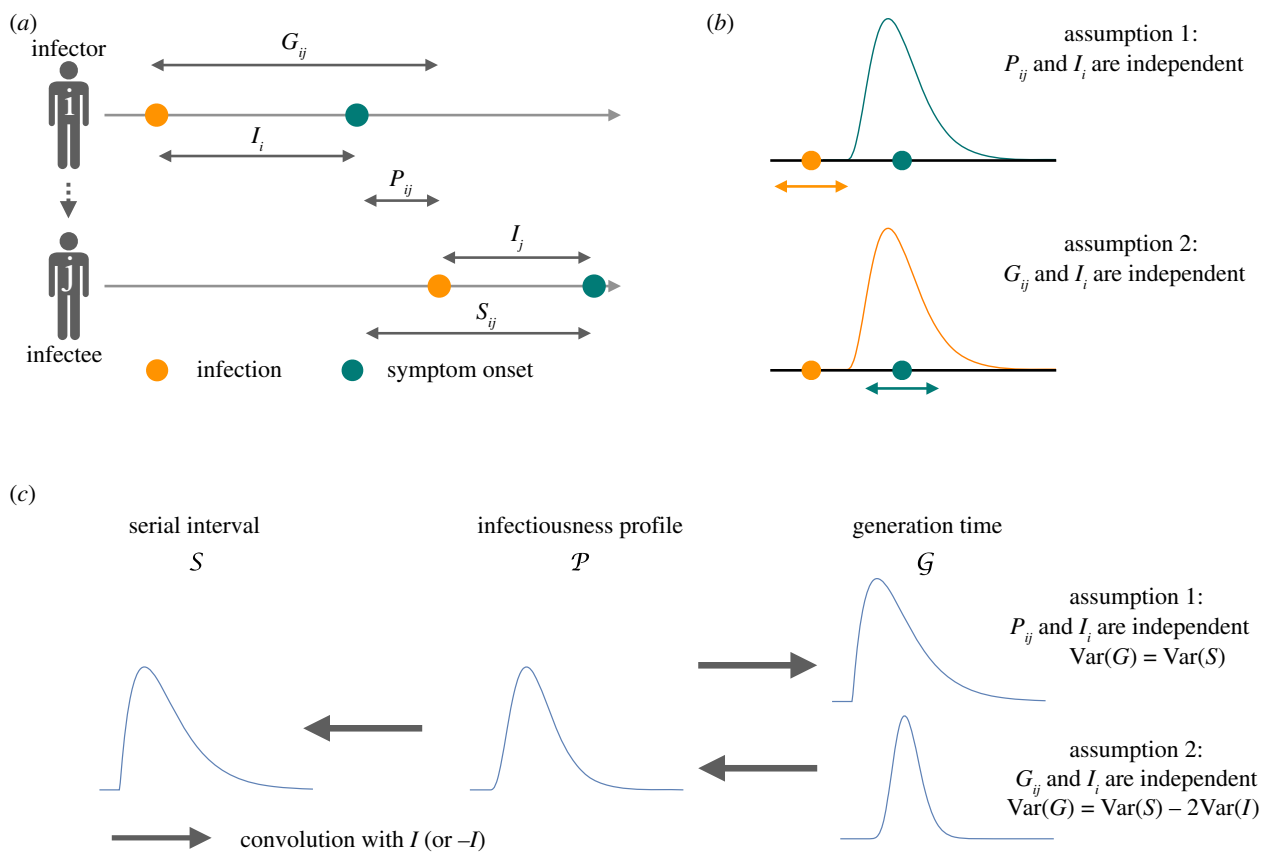
Figure 1. A schematic of how the assumptions about infectiousness and symptoms affect the relationship between the serial interval and generation time distributions. (a) Definitions of: serial interval S_{ij}, time from symptom onset of infector i to symptom onset of infectee j; generation time G_{ij}, time from infection of i to infection of j; incubation time I_i, time from infection of i to symptom onset of i; and P_{ij}, time from symptom onset of i to infection of j. (b) Illustration of how infectiousness relates to the point of infection and onset of symptoms under the two different assumptions. Under assumption 1 (P_{ij} and I_i independent), the infectiousness is fixed with reference to symptom onset. Under assumption 2 (G_{ij} and I_i independent), the infectiousness is fixed with reference to the point of infection. (c) The relationship between the generation time distribution, the infectiousness profile and the serial interval distribution under assumptions 1 and 2.

1.2. Definitions

We consider an infector i and infectee j (figure 1a) and define: S_{ij} as the serial interval (time interval between symptom onset of infector i and symptom onset of infectee j); G_{ij} as the generation time (time interval from infection of i to infection of j); P_{ij} as the time interval from symptom onset of i to infection of j; and I_i as the incubation period of i (and I_j is the incubation period of j). For clarity, we drop the indices when they are not necessary. We use calligraphic letters to denote the probability density functions—i.e. distributions—of these time variables (e.g. $\mathcal{S}$ is the serial interval distribution). The generation time distribution $\mathcal{G}$ describes infectiousness relative to the point of infection, while $\mathcal{P}$ describes infectiousness relative to symptom onset. We refer to $\mathcal{P}$ as the infectiousness profile [5,6].

1.3. Estimation based on transmission pairs

The distributions $\mathcal{I}$, $\mathcal{G}$, $\mathcal{P}$ and $\mathcal{S}$ are typically derived from contact tracing data during epidemic outbreaks. Such data consist of transmission pairs, usually with symptom onset times for infector and infectee, and an exposure window for the infection time of the infectee. These data allow $\mathcal{I}$ and $\mathcal{S}$ to be estimated without further assumptions, but not $\mathcal{P}$ and $\mathcal{G}$.

Here, we note some general caveats relating to the use of transmission pairs to estimate these distributions. These are not relevant to the relationship between $\mathcal{G}$ and $\mathcal{S}$ discussed here, but should nevertheless be considered when working with these data. Firstly, in a growing epidemic, contact tracing data will underestimate generation times and serial intervals: when prevalence is increasing, sampled cases will be biased towards recent infections. This bias can be corrected by explicitly accounting for the growth when deriving the distributions [2], as done, for example, in Ferretti et al. [4]. Secondly, as prevalence increases and the number of susceptible individuals becomes limiting, generation times and serial intervals will contract: each susceptible can only be infected once, resulting in fewer longer intervals [7]. Thirdly, sampled transmission pairs may not be representative of the overall population—for example, asymptomatic cases will be under-represented. Furthermore, contacts who are exposed to infection but not infected contribute information about infectiousness that is not captured in these analyses.

2. Relationship between $\mathcal{G}$, $\mathcal{P}$ and $\mathcal{S}$

2.1. Deriving $\mathcal{G}$ and $\mathcal{P}$ from $\mathcal{S}$

The relationships between the time intervals G_{ij}, P_{ij} and S_{ij} are illustrated in figure 1a and are captured by the following equations:

$$S_{ij} = P_{ij} + I_j, \tag{2.1a}$$
$$G_{ij} = P_{ij} + I_i \tag{2.1b}$$

and

$$G_{ij} = S_{ij} + I_i - I_j. \tag{2.1c}$$

Deriving $\mathcal{P}$ from $\mathcal{S}$ does not require strong assumptions; P_{ij} and I_j are plausibly independent: it is reasonable to assume

that the interval between the infector's symptom onset and onward transmission does not affect the incubation period of the infectee. From equation (2.1a), we can then write S as the convolution between $\mathcal{P}$ and $\mathcal{I}$, i.e.

$$S = \mathcal{P} * \mathcal{I}. \qquad (2.2)$$

The infectiousness profile $\mathcal{P}$ can therefore be derived by deconvolution of the serial interval and incubation period distributions [5,6].

Deriving $\mathcal{G}$ is not as straightforward: as the intervals P_{ij} and I_i relate to the same individual, independence of the two is a more debatable assumption than for P_{ij} and I_j. Progress can be made by assuming that I_i and I_j are independent and identically distributed (i.i.d): under this assumption, the intervals S and G have the same mean and their variances are related by the covariance of P_{ij} and I_i (from equations (2.1a,b)) [8]

$$\mathrm{Var}(G) = \mathrm{Var}(S) + 2\mathrm{Cov}(P_{ij}, I_i). \qquad (2.3)$$

Deriving the generation time distribution $\mathcal{G}$ requires further assumptions: typically, either the independence of P_{ij} and I_i or the independence of G_{ij} and I_i.

2.2. Assumption 1: independence of incubation period of the infector (I_i) and time from symptoms of infector to infection of infectee (P_{ij})

Under this assumption, infectiousness is fixed with reference to symptom onset (figure 1b): there is no correlation between how long it takes an individual to develop symptoms and the interval between symptom onset and onward transmission. Such a situation would arise, for example, if individuals have a variable period between infection and the onset of infectiousness, the duration of which does not affect subsequent infectiousness or onset of symptoms (see also [2]).

Using equation (2.1b), the independence of P_{ij} and I_i means that $\mathcal{G}$ can be derived as the convolution of $\mathcal{P}$ and $\mathcal{I}$ (i.e. $\mathcal{G} = \mathcal{P} * \mathcal{I}$) and is thus identical to S (equation (2.2)). Thus, the often-used approach of approximating the generation time distribution by the serial interval implicitly makes this assumption (figure 1c).

In line with the above, under this assumption, the variance of G is equal to the variance of S,

$$\begin{aligned}\mathrm{Var}(G) &= \mathrm{Var}(S) + 2\mathrm{Cov}(P_{ij}, I_i)\\ &= \mathrm{Var}(S).\end{aligned} \qquad (2.4)$$

This assumption is biologically implausible: it requires the incubation period to be independent of processes affecting infectiousness. Yet infectiousness and symptom onset are both likely to depend on pathogen load; it is therefore unlikely that assumption 1 holds for most pathogens. Furthermore, unlike serial intervals, generation times cannot be negative. When observed, negative serial intervals are empirical evidence against assumption 1.

2.3. Assumption 2: independence of incubation period of the infector (I_i) and time from infection of infector to infection of infectee (G_{ij})

Under this assumption, infectiousness is fixed with reference to the point of infection (figure 1b): the timing of transmission is uncorrelated with the timing of symptom onset. As $P_{ij} =$

$G_{ij} - I_i$ (equation (2.1b)), $\mathcal{P}$ would then be the convolution of $\mathcal{G}$ and $-\mathcal{I}$, $\mathcal{P} = \mathcal{G} * (-\mathcal{I})$. The generation time distribution G could therefore be derived from S by deconvolving first with $\mathcal{I}$ and then with $-\mathcal{I}$ (figure 1c), i.e. solving $S = \mathcal{G} * (-\mathcal{I}) * \mathcal{I}$ for $\mathcal{G}$. The functional form of $\mathcal{G}$ would therefore depend on both empirical distributions S and $\mathcal{I}$. This is the approach adopted for deriving the generation interval of severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) in Ferretti et al. [4] and Ganyani et al. [9].

In line with the above, under this assumption, the variance of G is smaller than the variance of S,

$$\begin{aligned}\mathrm{Var}(G) &= \mathrm{Var}(S) + 2\mathrm{Cov}(P_{ij}, I_i)\\ &= \mathrm{Var}(S) + 2\mathrm{Cov}(G_{ij} - I_i, I_i)\\ &= \mathrm{Var}(S) + 2[\mathrm{Cov}(G_{ij}, I_i) - \mathrm{Cov}(I_i, I_i)]\\ &= \mathrm{Var}(S) - 2\mathrm{Var}(I)\\ &\leq \mathrm{Var}(S).\end{aligned} \qquad (2.5)$$

This assumption is also biologically implausible. If infectiousness and symptom onset both depend on pathogen load, individuals with a rapid increase in pathogen load will develop symptoms early (short I_i) and transmit sooner after infection (small G_{ij}), leading to $\mathrm{Cov}(G_{ij}, I_i) > 0$. Furthermore, symptom onset itself is likely to affect infectiousness. Depending on the pathogen, the effect could be in either direction (symptomatic individuals transmitting more because symptoms contribute to transmission, or symptomatic individuals transmitting less because they self-isolate). However, either scenario would lead to a positive correlation between the timing of symptom onset and transmission (electronic supplementary material, figure S1).

2.4. Assumptions 1 and 2 bound the variance of G

Although neither assumption 1 nor assumption 2 is plausible, they are still informative: the variances of the generation time distribution derived under these assumptions can reasonably be considered as upper and lower bounds for $\mathrm{Var}(G)$,

$$\mathrm{Var}(S) - 2\mathrm{Var}(I) \leq \mathrm{Var}(G) \leq \mathrm{Var}(S). \qquad (2.6)$$

Assumption 1 leads to the upper bound $\mathrm{Var}(G) = \mathrm{Var}(S)$. A greater variance would require $\mathrm{Cov}(P_{ij}, I_i) > 0$ (see equation (2.3)), i.e. transmission occurring late with reference to symptoms for individuals with a longer incubation period—for example, a greater proportion of transmission being post-symptomatic when symptoms appear late. The notion that $\mathrm{Cov}(P_{ij}, I_i) > 0$ is unlikely has also been previously suggested in the literature [2]. Furthermore, if negative serial intervals are observed, this suggests $\mathrm{Var}(G) < \mathrm{Var}(S)$ (assuming that the serial interval distribution and generation time distribution have a similar shape), since the distributions have the same mean and negative generation times are not possible.

Assumption 2 leads to the lower bound $\mathrm{Var}(G) = \mathrm{Var}(S) - 2\mathrm{Var}(I)$. A lower variance would require $\mathrm{Cov}(G_{ij}, I_i) < 0$ (see equation (2.5)), i.e. transmission occurring soon after infection for individuals with a longer incubation period. However, as discussed above, individuals with a faster increase in pathogen load are likely to start transmitting earlier and also have a shorter incubation period, leading to $\mathrm{Cov}(G_{ij}, I_i) > 0$. Furthermore, if the appearance of symptoms leads to a change in infectiousness (in either direction), earlier symptoms will correlate with earlier transmission, again leading to $\mathrm{Cov}(G_{ij}, I_i) > 0$.

Table 1. Shape, mean and variance of incubation period and serial interval distributions of SARS-CoV-2 from a range of studies. N indicates the sample size.

study	distribution	shape	mean (days)	standard deviation (days)	N
Zhang *et al.* [11]	incubation	lognormal	5.2	2.6	49
Li *et al.* [12]	incubation	lognormal	5.2	3.9	10
Lauer *et al.* [13]	incubation	lognormal	5.5	2.4	181
Backer *et al.* [14]	incubation	Weibull	6.4	2.3	88
Linton *et al.* [15]	incubation	lognormal	5.6	2.8	158
Ganyani *et al.* [9]	serial interval (Singapore)	gamma	5.2	4.3	54
Ganyani *et al.* [9]	serial interval (Tianjin)	gamma	4.0	4.2	114
Zhang *et al.* [11]	serial interval	gamma	5.1	2.7	34
Li *et al.* [12]	serial interval	gamma	7.5	3.4	6
He *et al.* [5]	serial interval	gamma (shifted)	5.8	4.5	77
Nishiura *et al.* [16]	serial interval	lognormal	4.7	2.9	28
Ali *et al.* [17]	serial interval (all)	normal	5.1	5.3	677
Ali *et al.* [17]	serial interval (pre-peak)	normal	7.8	5.2	162

3. Possible solutions

3.1. Empirical testing of assumptions

A priori, there is no reason to consider either assumption 1 or assumption 2 as more plausible than the other. With appropriate data, the assumptions can be tested empirically. For example, such analysis for SARS-CoV-2 suggests a strong positive correlation between G_{ij} and I_i, and a weak negative correlation between P_{ij} and I_i [10]. In other words, for SARS-CoV-2, neither assumption holds, but assumption 1 (independence of P_{ij} and I_i) is a better approximation.

The empirical testing of these assumptions requires transmission pairs for which I_i and G_{ij} (or, equivalently, P_{ij}) can be estimated. This can be done with either: (i) data on the infection time for both infector i and infectee j and the symptom onset time for i or (ii) data on the symptom onset time for both i and j and the infection time for i, as the assumption that P_{ij} and I_j are independent allows the infection time for j to be estimated. Therefore, an interesting corollary here is that, for transmission pairs with a known serial interval, data on the infection time of the infector is more informative than the infection time of the infectee.

In practice, when such data are available, the generation interval distribution can simply be directly estimated from the data [10]. The reason for deriving $\mathcal{G}$ from $\mathcal{S}$ is precisely the lack of such data; an alternative approach for assessing the plausibility of the assumptions underlying this derivation is therefore necessary.

3.2. Assumptions 1 and 2 as edge cases

As assumptions 1 and 2 bound the variance of G, a solution when data are lacking is to derive $\mathcal{G}$ under both assumptions, and treat these as boundary cases in downstream analysis (e.g. best and worst case scenarios). This approach may not always be entirely straightforward. If $\text{Var}(S) < 2\text{Var}(I)$, assumption 2 would lead to negative variance of G. In these cases, the lower bound for $\text{Var}(G)$ is zero. If the serial interval distribution includes negative values, deriving $\mathcal{G}$ under assumption 1 is problematic. A pragmatic approach in these cases would be to use $\text{Var}(G) = \text{Var}(S)$ and to assume a non-negative functional form for $\mathcal{G}$ (e.g. lognormal, gamma or Weibull), although the resulting distribution will not be the correct distribution under assumption 1. The key point is that evidence against assumption 1, such as negative serial intervals, is not, in itself, evidence in favour of assumption 2.

4. Implications for the modelling of contact tracing

Finally, we explore the impact of the variance and functional form of the generation time distribution on the modelling of contact tracing, using the example of SARS-CoV-2. Table 1 shows empirical estimates for the mean and standard deviation (s.d.) of the serial interval and incubation period. Both have a mean of around 5 days. The s.d. of the incubation period is generally estimated to be in the range of 2.3–2.8 days, although some studies have also reported considerably higher values (table 1). With the exception of some smaller studies, the s.d. of the serial interval is generally estimated to be of the order of 4.2–5.5 days. Assuming the s.d. of S to be 5 days [$\text{Var}(S) = 25$] and the s.d. of I to be 2.8 days [$\text{Var}(I) = 8$], a plausible range for the s.d. of G would thus be 5 to 3 days [$\text{Var}(G) = 25$ and $\text{Var}(G) = 9$] under assumptions 1 and 2 respectively—though lower values cannot be excluded if the lower estimates of the s.d. of S or higher estimates of the s.d. of I hold.

Figure 2 illustrates how the variance and functional form of the generation time distribution impact how quickly infected individuals need to be isolated to prevent a significant portion of onward transmission, that is, how quickly contact tracing needs to operate for the epidemic to be controllable. For example, assuming that the generation time is gamma distributed with a mean of 5 days, preventing 80% of onward transmission requires isolation of an infected individual within 1.1 days if the s.d. of G is 5 days, and 2.5 days if the s.d. of G is 3 days. On the other hand, if the variance is large, isolating individuals even with considerable delay will still have an impact on onward transmission. For example, isolating an infected individual 10 days after infection will
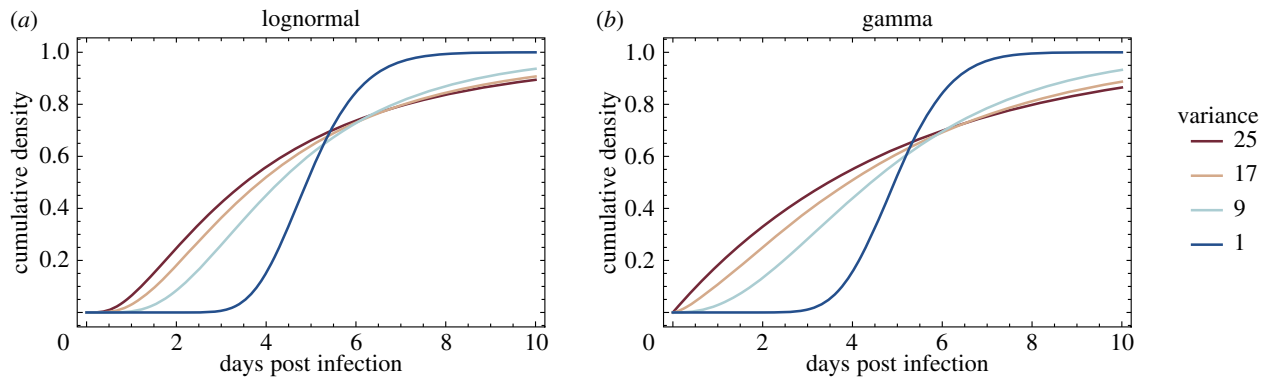
Figure 2. A schematic showing the impact of functional form and variance on the timing of onward transmission. The plots show cumulative generation time distributions, i.e. the proportion of transmission occurring within x days of infection. All distributions have a mean of 5 days. The illustrated variances correspond to standard deviations of 5.0, 4.1, 3 and 1 days. Note that lognormal (a) and gamma (b) distributions have support on (0, ∞), implying an infinite infectious period, which is not correct. However, in practice, this is an acceptable approximation when the probability density in the tail of the distribution is very low.

prevent 14% of onward transmission if the s.d. of G is 5 days, but only 7% if the s.d. of G is 3 days. In practice, if the goal of contact tracing is to control the epidemic, the former scenario is more relevant [4]. Thus underestimating the variance of the generation time distribution (assumption 2) risks overestimating the effectiveness of contact tracing.

5. Conclusion

Neither of these two commonly used approaches for estimating the generation time distribution from the serial interval distribution is based on plausible assumptions for most pathogens. The two approaches yield generation time distributions with the same mean, but different variances. This difference in variance can have a considerable impact on estimating the controllability of an epidemic through contact tracing. The two variances are plausible upper and lower bounds for the variance of the generation time distribution. We therefore suggest that a pragmatic solution is to treat the distributions derived through the two approaches as edge cases in downstream analysis. When implementing this solution, it remains important to correct for the bias towards short intervals arising in a growing epidemic and to consider the limitations of analyses based on contact tracing data.

References

1. Wallinga J, Lipsitch M. 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604. (doi:10.1098/rspb.2006.3754)

2. Britton T, Scalia Tomba G. 2019 Estimation in emerging epidemics: biases and remedies. *J. R. Soc. Interface* **16**, 20180670. (doi:10.1098/rsif.2018.0670)

3. Fraser C, Riley S, Anderson RM, Ferguson NM. 2004 Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151. (doi:10.1073/pnas.0307506101)

4. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Parker M, Bonsall D, Fraser C. 2020 Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936. (doi:10.1126/science.abb6936)

5. He X et al. 2020 Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675. (doi:10.1038/s41591-020-0869-5)

6. Ashcroft P, Huisman JS, Lehtinen S, Bouman JA, Althaus CL, Regoes RR, Bonhoeffer S. 2020 COVID-19 infectivity profile correction. *Swiss. Med. Wkly.* **150**, w20336. (doi:10.4414/smw.2020.20336)

7. Kenah E, Lipsitch M, Robins JM. 2008 Generation interval contraction and epidemic data analysis. *Math. Biosci.* **213**, 71–79. (doi:10.1016/j.mbs.2008.02.007)

8. Svensson Å. 2007 A note on generation times in epidemic models. *Math. Biosci.* **208**, 300–311. (doi:10.1016/j.mbs.2006.10.010)

9. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, Hens N. 2020 Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance* **25**, 2000257. (doi:10.2807/1560-7917.ES.2020.25.17.2000257)

10. Ferretti L et al. 2020 The timing of COVID-19 transmission. (https://www.medrxiv.org/content/10.1101/2020.09.04.20188516v2).

11. Zhang J et al. 2020 Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.* **20**, P793–P802.

12. Li Q et al. 2020 Early transmission dynamics in Wuhan, China, of novel coronavirus—infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207.

13. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. 2020 The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* **172**, 577–582. (doi:10.7326/M20-0504)

14. Backer JA, Klinkenberg D, Wallinga J. 2020 Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance* **25**, 2000062. (doi:10.2807/1560-7917.ES.2020.25.5.2000062)

15. Linton NM et al. 2020 Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J. Clin. Med.* **9**, 538. (doi:10.3390/jcm9020538)

16. Nishiura H, Linton NM, Akhmetzhanov AR. 2020 Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, P284–P286.

17. Ali ST, Wang L, Lau EH, Xu XK, Du Z, Wu Y, Leung GM, Cowling BJ. 2020 Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* **369**, 1106–1109. (doi:10.1126/science.abc9004)

Supporting Figure

Sonja Lehtinen, Peter Ashcroft & Sebastian Bonhoeffer
0n the relationship between serial interval, infectiousness profile and generation time
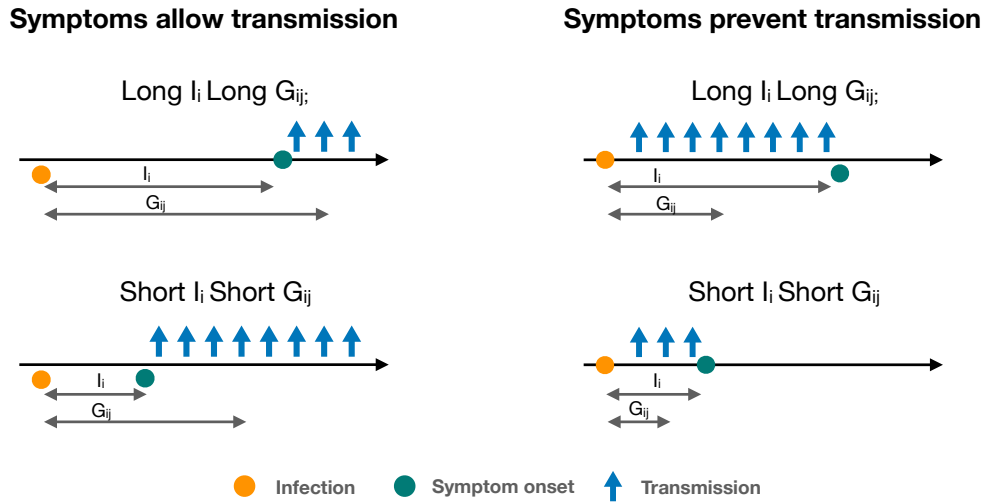*Journal of the Royal Society Interface*

Figure 1: Schematic illustrating how symptom onset affecting transmission leads to a positive correlation between incubation period and generation time, whether symptoms increase or decrease transmission. In the left-hand panels, transmission only occurs after symptom onset. In the right-hand panels, transmission only occurs prior to symptom onset. We make the additional assumption that the length of the incubation period does not affect when infectiousness ends (for the left-hand panels) or starts (for the right-hand panel).