

# Exploring how generation intervals link strength and speed of epidemics

Sang Woo Park      David Champredon      Joshua Weitz  
Jonathan Dushoff

March 15, 2018: In preparation for *PRSB*

## Abstract

Infectious-disease outbreaks are often characterized by the reproductive number  $\mathcal{R}$  and exponential rate of growth  $r$ . While an outbreak is ongoing, the rate of growth  $r$  is often estimated directly from incidence data. The reproductive number  $\mathcal{R}$  is often of particular interest, because it provides information about how hard an outbreak will be to control, and about predicted final size. During an outbreak  $\mathcal{R}$  is typically estimated from  $r$  by using information about generation intervals – that is, the amount of time between when an individual is infected by an infector, and when that infector was infected. In practice, it is impossible to obtain the exact shape of a generation-interval distribution and it is not always qualitatively clear how changes in estimates of the distribution translate to changes in the estimate of  $\mathcal{R}$ . Here, we show that parameterizing a generation interval distribution using its mean and variance provides a clear biological intuition into how its shape affects the relationship between  $\mathcal{R}$  and  $r$ . We explore approximations based on estimates of mean and variance, and a gamma approximation, and find that use of these two moments is often sufficient to capture the  $r$ – $\mathcal{R}$  relationship and provide robust estimates of  $\mathcal{R}$ .

# 1 Introduction

Infectious disease research often focuses on estimating the reproductive number, i.e., the number of new infections caused on average by a single infection. This number is termed the reproductive number –  $\mathcal{R}$ . The reproductive number provides information about the disease’s potential for spread and the difficulty of control. It is described in terms of an average [2] or an appropriate sort of weighted average [6].

The reproductive number has remained a focal point for research because it provides information about how a disease spreads in a population, on the scale of disease generations. As it is a unitless quantity, it does not, however, contain information about *time*. Hence, another important quantity is the population-level *rate of spread*,  $r$ . The initial rate of spread can often be measured robustly early in an epidemic, since the number of incident cases at time  $t$  is expected to follow  $i(t) \approx i(0) \exp(rt)$ . The rate of growth can also be described using the “characteristic time” of exponential growth  $C = 1/r$ . This is closely related to the doubling time (given by  $T_2 = \ln(2)C \approx 0.69C$ ).

In disease outbreaks, the rate of spread,  $r$ , is often inferred from case-incidence reports, by fitting an exponential function to the incidence curve [14, 17, 13]. [JSW: *Talk about cumulative incidence?*] [JD: *None of these cites use cumulative incidence, and the point about how maybe you can use it seems like a detour.*] Estimates of the initial exponential rate of spread,  $r$ , can then be combined with a mechanistic model that includes unobserved features of the disease to estimate the initial reproductive number,  $\mathcal{R}$ . In particular,  $\mathcal{R}$  is often calculated from  $r$  and the generation-interval distribution using the generating function approach popularized by [23].

The *generation interval* is the amount of time between when an individual is infected by an infector, and the time that the infector was infected [19]. While  $r$  measures the speed of the disease at the population level, the generation interval measures speed at the individual level. Generation interval distributions are typically inferred from contact tracing, sometimes in combination with clinical data [4, 11, 8]. Generation interval distributions can be difficult to ascertain empirically, and the generation-function approach depends on entire distribution – which makes it difficult to ascertain which features of the distributions are essential to connect measurements of the rate of spread  $r$ , with the reproductive number,  $\mathcal{R}$ .

Here, we explore the qualitative relationship between generation time, initial rate of spread  $r$ , and initial reproductive number  $\mathcal{R}$  using means, vari-

ance measures and approximations. By doing so, we hope to shed light on the underpinnings of the relationship between  $r$  and  $\mathcal{R}$ , and on the factors underlying its robustness and its practical use when data on generation intervals is limited or hard to obtain.

## 2 Relating $\mathcal{R}$ and $r$

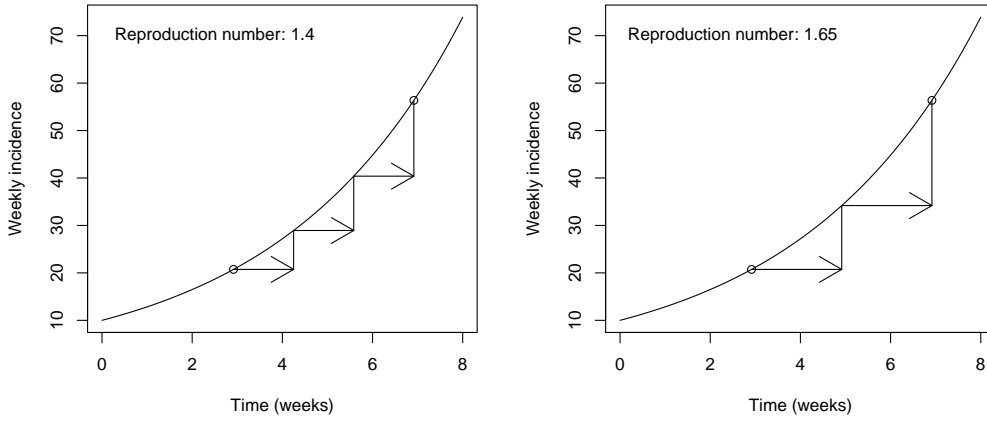


Figure 1: Two hypothetical epidemics with the same growth rate ( $r = 0.25/\text{week}$ ) and fixed generation intervals. Assuming a short generation interval (fast transmission at the individual level) implies a smaller reproductive number  $\mathcal{R}_0$  (panel A) when compared to a longer generation interval (slow transmission at the individual level, panel B). [Who did this first]

We are interested in the relationship between  $r$ ,  $\mathcal{R}$  and the generation-interval distribution, which describes the interval between the time an individual becomes *infected* and the time that they *infect* another individual. This distribution links  $r$  and  $\mathcal{R}$ . In particular, if  $\mathcal{R}$  is known, a shorter generation interval means a faster epidemic (larger  $r$ ). Conversely (and perhaps counter-intuitively), if  $r$  is known, then faster disease generations imply a *lower* value of  $\mathcal{R}$ , because more *individual* generations are required to realize the same *population* spread of disease (see Fig. 1).

We define the generation-interval distribution using a renewal-equation approach. A wide range of disease models can be described using the model [Hees, ChampInPrep]:

$$i(t) = S(t) \int K(s) i(t-s) ds, \quad (1)$$

where  $t$  is time,  $i(t)$  is the incidence of new infections,  $S(t)$  is the *proportion* of the population susceptible, and  $K(s)$  is the intrinsic infectiousness of individuals who have been infected for a length of time  $s$ .

We then have the basic reproductive number:

$$\mathcal{R}_0 = \int K(s) ds, \quad (2)$$

and the *intrinsic* generation-interval distribution:

$$g(s) = \frac{K(s)}{\mathcal{R}_0}. \quad (3)$$

The “intrinsic” interval can be distinguished from “realized” intervals, which can look “forward” or “backward” in time [5] (see also earlier work [19, 15]).

Disease growth is predicted to be approximately exponential in the early phase, because the depletion in the effective number of susceptibles is relatively small. Thus, for the exponential phase, we write:

$$i(t) = \mathcal{R} \int g(\tau) i(t-\tau) d\tau, \quad (4)$$

where  $\mathcal{R} = \mathcal{R}_0 S$ .

We then solve for the characteristic time  $C$  by assuming that the population is growing exponentially: i.e., substitute  $i(t) = i(0) \exp(t/C)$  to obtain

$$1/\mathcal{R} = \int g(\tau) \exp(-\tau/C) d\tau. \quad (5)$$

This fundamental relationship dates back to the work of Euler [Euler]. We will explore the shape of this relationship using parameters based on some human infectious diseases, and investigate approximations based on gamma-distributed generation intervals.

## 3 Approximation framework

### 3.1 Approximation method, theory

We do not expect to know the full distribution  $g(\tau)$ , so we are interested in approximations to  $\mathcal{R}$  based on limited information. We follow the approach of [17] and approximate the generation interval with a gamma distribution. This is a better-grounded starting point than the earlier normal-distribution approach of [23], since the gamma distribution is confined to non-negative values (see Appendix for comparisons).

For biological interpretability, we describe the distribution using the mean  $\bar{G}$  and the squared coefficient of variation  $\kappa$  (thus  $\kappa = 1/a$ , and  $\bar{G} = a\theta$ , where  $a$  and  $\theta$  are the shape and scale parameters under the standard parameterization of the gamma distribution). Substituting the gamma distribution into (5) then yields:

$$\mathcal{R} \approx (1 + \kappa r \bar{G})^{1/\kappa}. \quad (6)$$

We write:

$$\mathcal{R} \approx (1 + \kappa \rho)^{1/\kappa} \equiv X(\rho; 1/\kappa), \quad (7)$$

where  $\rho = \bar{G}/C = r\bar{G}$  measures how fast the epidemic is growing (on the time scale of the mean generation interval) – or equivalently, the length of the mean generation interval (in units of the characteristic time of exponential growth). The longer the generation interval is compared to  $T_c$ , the higher the estimate of  $\mathcal{R}$  (see Fig. 1). We then explore the behaviour of the generalized exponential function  $X$  defined above (equivalent to the Tsallis “q-exponential”, with  $q = 1 - \kappa$  [22]): its shape determines how the estimate of  $\mathcal{R}$  changes with the estimate of normalized generation length  $\rho$ .

**[JSW: Okay! But, for readability, why do you need to define this as  $X$ . By defining in terms of  $X$  you have sentences like ...]** **[JD: I see the point, but I don’t think that’s much abstraction.  $R$  is approximated by the function  $X$ , and we’re investigating how  $X$  behaves. I have changed the text a bit to try to get this idea across.]** For small  $\rho$ ,  $X$  always looks like  $1 + \rho$ , regardless of the shape parameter  $1/\kappa$ , which determines the curvature: if  $1/\kappa = 1$ , we get a straight line, for  $1/\kappa = 2$  the curve is quadratic, and so on (see Fig. 2). For large values of  $1/\kappa$ ,  $X$  converges toward  $\exp(\rho)$ . **[JSW: specify this using a formula, don’t just allude to the compound interest approximation to the exponential. Just go ahead and specify it.]** **[JD: I thought about this, but it would look just like the formula for  $X$  ... I just simplified]**

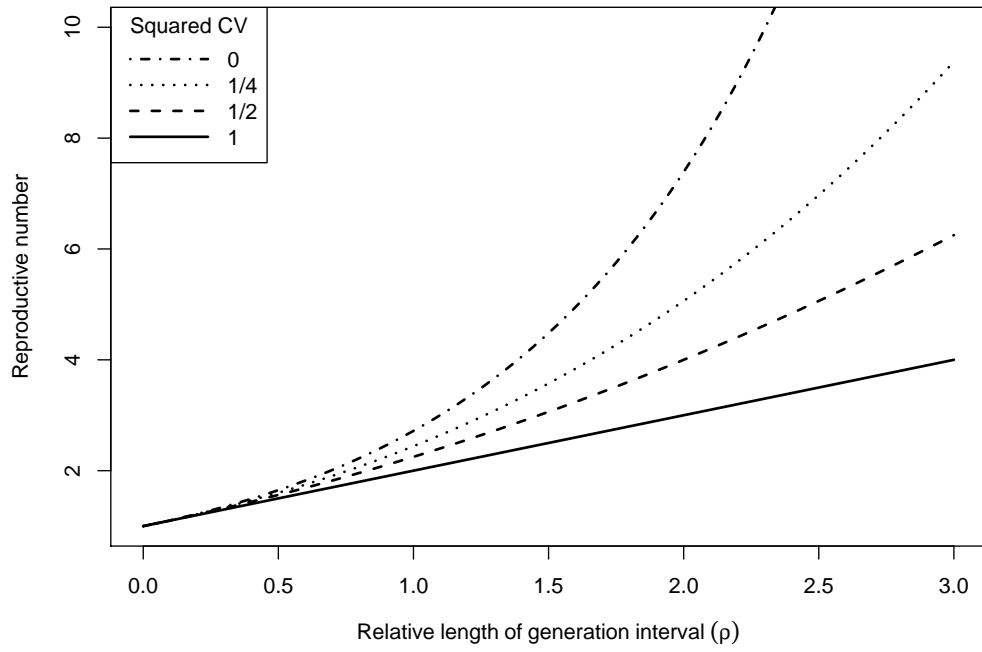


Figure 2: The approximate relationship (6) between mean generation time (relative to the characteristic time of exponential growth,  $\rho = r\bar{G} = \bar{G}/C$ ) and reproductive number. The curves correspond to different amounts of variation in the generation-interval distribution.

The limit as  $\kappa \rightarrow 0$  is reasonably easy to interpret. The incidence is increasing by a factor of  $\exp(\rho)$  in the time it takes for an average disease generation. If  $\kappa = 0$ , the generation interval is fixed, so the means the average case must cause  $\mathcal{R} = \exp(\rho)$  new cases. If variation in the generation time (i.e.,  $\kappa$ ) increases, then some new cases will be produced before, and some after, the mean generation time. Since we assume the disease is increasing exponentially, infections that occur early on represent a larger proportion of the population, and thus will have a disproportionate effect: individuals don't have to produce as many lifetime infections to sustain the growth rate, and thus we expect  $\mathcal{R} < \exp(\rho)$ .

The straight-line relationship for  $\kappa = 1$  also has a biological interpretation. In our approximation, this corresponds to a generation distribution that is approximated by an exponential distribution. In this case, recovery rate and infection rate are constant for each individual. The rate of exponential growth per generation is then given directly by the net per capita increase in infections:  $\mathcal{R} - 1$ , where one represents the recovery of an infectious individual.

## 3.2 Approximation method, in practice

We test our approximation method by generating pseudo-realistic generation-interval distributions, using observations (or estimated parameters) of generation intervals (or else latent and infectious intervals) from specific diseases. For each pseudo-realistic distribution, we calculate the “true” relationship between  $r$  and  $\mathcal{R}$  based on our distribution assumptions, and compare it with one or more distributions based on gamma approximations. These approximations are first done with large amounts of data, allowing us to evaluate how well the approximations describe the  $r$ – $\mathcal{R}$  relationship under ideal conditions, and then tested with smaller amounts of data.

Estimating generation intervals is complex; our goal with pseudo-realistic distributions is not to precisely match real diseases, but to generate distributions that are likely to be roughly as challenging for our approximation methods as real distributions would be. To generate generation intervals we sample from observations (for empirical distributions) or generate from uniform quantiles (for parametric distributions). We sample latent periods at random. We then choose infectious periods by length-weighted resampling (since longer infectious period implies more opportunities to infect), then choose infection delays uniformly within resampled infectious periods. Gen-

eration interval is the sum of latent period and infection delay [7]. For our examples, we drew 10000 values from each distribution.

We then calculate “exact” relationships (for our pseudo-realistic distributions) by substituting sampled generation intervals into (5). This relationship is then compared to the approximate relationship based on (6)

## 4 Examples

We investigate this approximation approach using three different examples. These examples also serve to demonstrate that robust estimates could be made with less data and potentially earlier in an outbreak – a point we revisit in the Discussion. Our initial investigation of this question was motivated by our work on the West African Ebola Outbreak [24], so we start with that example. To probe the approximation more thoroughly, we also chose one disease with high variation in generation interval (canine rabies), and one with a high reproductive number (measles). For simplicity, we assumed that latent and infectious periods are equivalent to incubation and symptomatic periods for Ebola virus disease (EVD) and canine rabies.

### 4.1 Ebola

We generated a pseudo-realistic generation-interval distribution for Ebola virus disease (EVD) using information from [4] and a lognormal assumption for both the incubation and infectious periods. We used a lognormal assumption for our components because it is straightforward and should provide a challenging test of our gamma approximation. We used the reported standard deviation for the infectious period, and chose the standard deviation for the incubation period to match the reported standard deviation for the serial interval distribution, since this value is available and is expected to be similar to the generation interval distribution for EVD [4].

We then calculated the exact relationship between  $r$  and  $\mathcal{R}$  implied by our pseudo-realistic distribution (5), and the approximate relationship (6) based only on the mean and CV (see Fig. 3). The approximation is within 1% of the pseudo-realistic distribution it is approximating across the range of country estimates, and within 5% across the range shown. It is also within 2% of the WHO estimates, which were done using a different approach.

**[JSW:** *What do you mean “simulated”... a distribution. The reader*



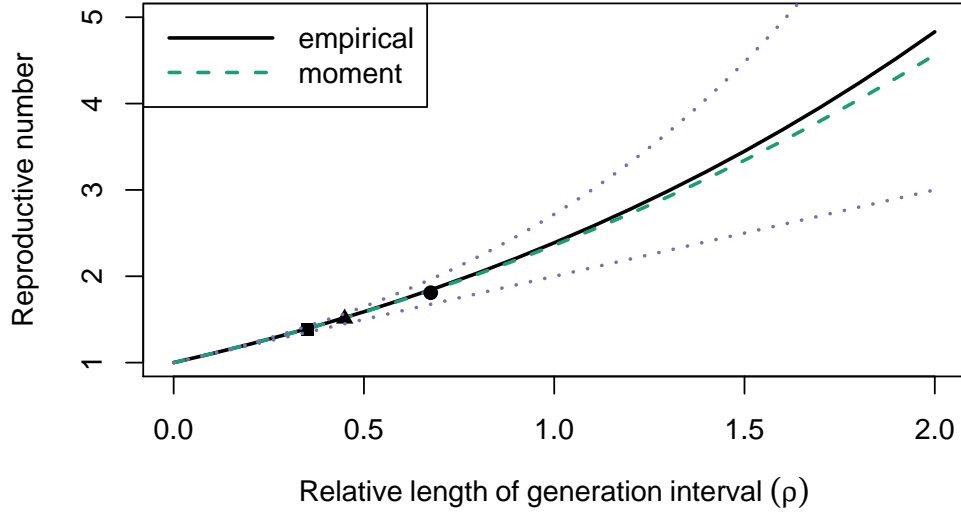


Figure 3: Estimating  $\mathcal{R}$  for the West African Ebola Outbreak. (solid curve) The relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution based on [4]. (dashed curve) The same relationship, approximated using the observed mean and CV. (dotted curves) Naive approximations based on exponential (lower) and fixed (upper) generation distributions. Points indicate estimates for the three focal countries of the West African Ebola Outbreak calculated by [4]: Sierra Leone (square,  $\mathcal{R} = 1.38$ ), Liberia (triangle,  $\mathcal{R} = 1.51$ ), and Guinea (circle,  $\mathcal{R} = 1.81$ ). Initial growth rate for each outbreak was inferred from doubling periods reported by [4] ( $r = \ln(2)/T_2$ ).

*expects you are using real data... and will be concerned that you are doing gamma in gamma out.] [JD: I have been working hard on the language about simulated distributions, let me know what you think.] [JD: We explain that we picked a lognormal to make sure it was different. I'm pretty sure that last time you suggested that we not mess with the lognormal (that is, you suggested that we should do gamma in/gamma out, which is not actually stupid, but would probably be confusing.) [JD: Using an estimate from observed intervals might in fact be possible, but raises issues that I might rather deal with in another paper.]*

## 4.2 Measles

We also applied the moment approximation to a pseudo-realistic generation-interval distribution based on information about measles from [10], [12], and [3]. Here, we found surprisingly close agreement between the exact and approximate relationships between  $r$  and  $\mathcal{R}$  across a much wider range of interest (a difference of  $< 1\%$  for  $\mathcal{R}$  up to  $> 20$ ). On examination, this closer agreement is due to the smaller overall variation in generation times in measles: when overall variation is small, differences between distributions have less effect.

## 4.3 Rabies

We did a similar analysis for rabies, and found that approximation is generally harder for this high-variance case (see Fig. 5). Since rabies estimates point to a value of  $\mathcal{R}$  near 1, results are not very sensitive to any tested assumption about the relationship. But, looking at the relationship more broadly, we see that the moment-based approximation would do a poor job of predicting the relationship for intermediate or large values of  $\mathcal{R}$  – in fact, a poorer job than if we use the approximation based on exponentially distributed generation times.

The reason for this can be seen in the histogram shown in Fig. 5. The moment approximation is strongly influenced by rare, very long generation intervals, and does a poor job of matching the observed pattern of short generation intervals (in particular, the moment approximation misses the fact that the distribution has a density peak at a finite value). We expect short intervals to be particularly important in driving the speed of the epidemic, and therefore in determining the relationship between  $r$  and  $\mathcal{R}$ . We

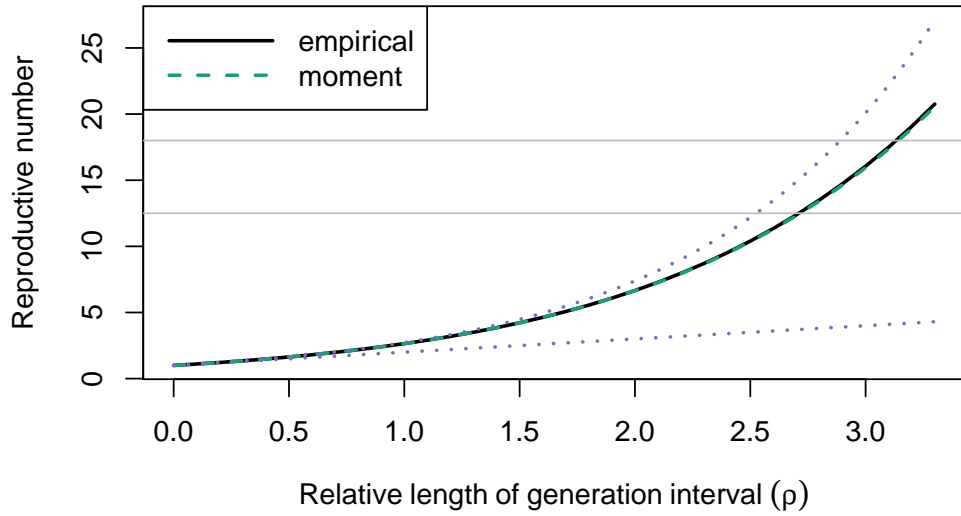


Figure 4: Estimating  $\mathcal{R}$  for measles. (solid curve) the relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution. (dashed curve) the same relationship approximated using the estimated mean and CV (this curve is almost invisible because it overlaps the solid curve) The dotted curves show the approximations based on exponential (lower) and fixed (upper) generation distributions. Gray horizontal lines represent  $\mathcal{R}$  ranges: 12.5-18 [3].

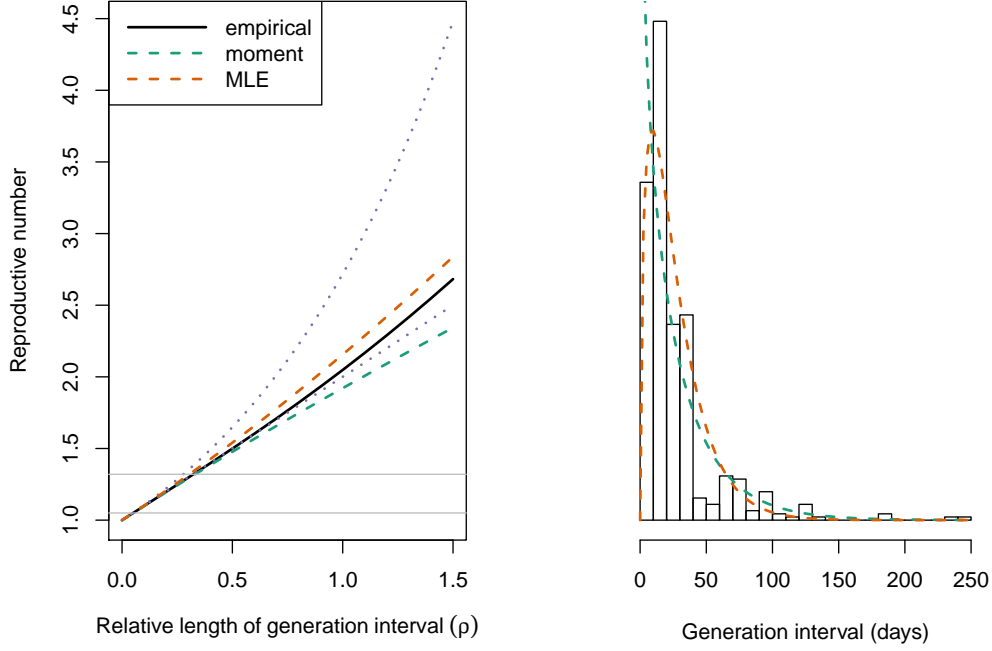


Figure 5: (Left) estimating  $\mathcal{R}$  from rabies infectious case data. (solid curve) the relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution. (dashed curve) the same relationship approximated using the observed mean and CV. (dash-dotted curve) the same relationship approximated using the mean and CV calculated from a maximum-likelihood fit. The dotted curves show the approximations based on exponential (lower) and fixed (upper) generation distributions. Gray horizontal lines represent  $\mathcal{R}$  ranges estimated by [7]: 1.05 - 1.32. (Right) histogram represents rabies generation interval distributions simulated from incubation and infectious periods observed by [7]. Curves represent estimated distribution of generation intervals using method of moments and MLE (corresponding to approximate  $r - \mathcal{R}$  relationships of the left figure).

can address this problem by estimating gamma parameters formally using a maximum-likelihood fit to the pseudo-realistic generation intervals. This fit does a better job of matching the observed pattern of short generation intervals and of predicting the simulated relationship between  $r$  and  $\mathcal{R}$  across a broad range (Fig. 5).

Disease	Ebola	Measles	Rabies
Parameter	Values		
Reproduction number	1.38, 1.51, 1.81 [4]	12.5-18 [3]	1.05-1.32 [7]
Mean incubation period (days)	11.4 [4]	12.77 [10]	24.24 [7]
SD incubation period (days)	7.5 (see Sec. 4.1)	2.67 [10]	29.49 [7]
Mean infectious period (days)	5 [4]	6.5 [3]	3.57 [7]
SD infectious period (days)	4.7 [4]	2.9 [12]	2.26 [7]

Table 1: Parameters that were used to obtain theoretical generation distributions for each disease. Reproduction numbers are represented as points in figure Fig. 3–5.

## 5 Discussion

Estimating the reproductive number  $\mathcal{R}$  is a key part of characterizing and controlling infectious disease spread. The initial value of  $\mathcal{R}$  for an outbreak is often estimated by estimating the initial exponential rate of growth, and then using a generation-interval distribution to relate the two quantities [23, 19, 15, 20]. However, detailed estimates of the full generation interval are difficult to obtain, and the link between uncertainty in the generation interval and uncertainty in estimates of  $\mathcal{R}$  are often unclear. Here we introduced and analyzed simple framework for *estimating* the relationship between  $\mathcal{R}$  and  $r$ , using only the estimated mean and CV of the generation interval. The framework is based on the gamma distribution. We use examples to test the robustness of the framework. We also compare estimates based directly on estimated mean and variance of the generation interval to estimates based on maximum-likelihood fits.

The gamma approximation for calculating  $\mathcal{R}$  from  $r$  was introduced by [17], and provides estimates that are simpler, more robust and more realistic than those from the normal approximations developed by [23] (see Appendix). Here, we attempted to present the gamma approximation in a form

conducive to intuitive understanding of the relationship between speed,  $r$ , and strength,  $\mathcal{R}$  (See Fig. 2). We discuss the general result that estimates of  $\mathcal{R}$  increase with mean generation, but decrease with *variation* in generation times [23]. We also provide mechanistic interpretations: when generation intervals are longer, more infection is needed per generation (larger  $\mathcal{R}$ ) in order to produce a given rate of increase  $r$ . Similarly, when variance in generation time is low, there is less early infection, and thus slower exponential growth, also meaning that a larger  $\mathcal{R}$  is needed.

We tested the gamma approximation framework by applying it to parameter regimes based on three diseases: Ebola, measles, and rabies. We found that approximations based on observed moments closely match true answers (based on known, pseudo-realistic distributions, see Sec. 4 for details) when the generation-interval distribution is not too broad (as is the case for Ebola and measles, but not for rabies), but that using maximum likelihood to estimate the moments provides better estimates for a broader range of parameters Sec. 4.3, and also when data are limited (see Appendix).

Our key finding is that summarizing an entire generation interval distribution using two moments can give sensible and robust estimates of the relationship between  $r$  and  $\mathcal{R}$ . This framework has potential advantages for understanding the likely effects of parameter changes, and also for parameter estimation with uncertainty: since  $\mathcal{R}$  can be estimated from three simple quantities ( $\bar{G}$ ,  $\kappa$  and  $r$ ), it should be straightforward to propagate uncertainty from estimates of these quantities to estimates of  $\mathcal{R}$ .

For example, during the Ebola outbreak in West Africa, many researchers tried to estimate  $\mathcal{R}$  from  $r$  [1, 4, 16, 18, 9] but uncertainty in the generation-interval distribution was often neglected (but see [21]). During the outbreak, [24] used a generation-interval argument to show that neglecting the effects of post-burial transmission would be expected to lead to underestimates of  $\mathcal{R}$ . Our generation interval framework provides a clear interpretation of this result: as long as post-burial transmission tends to increase generation intervals, it should result in higher estimates of  $\mathcal{R}$  for a given estimate of  $r$ . Knowing the exact shape of the generation interval distribution is difficult, but quantifying how various transmission routes and epidemic parameters affect the moments of the generation interval distribution will help researchers better understand and predict the scope of future outbreaks.

## Acknowledgments

We thank every god that anyone has ever thought of. Just in case.

## References

- [1] C. L. Althaus. “Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa.” In: *PLoS Curr* 6 (2014 Sep 02), 2014 Sep 02.
- [2] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press, 1991.
- [3] Roy M Anderson and Robert M May. “Directly transmitted infections diseases: control by vaccination”. In: *Science* 215.4536 (1982), pp. 1053–1060.
- [4] B. Aylward et al. “Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections.” In: *N Engl J Med* 371 (2014 Oct 16), pp. 1481–95.
- [5] D. Champredon and J. Dushoff. “Intrinsic and realized generation intervals in infectious-disease transmission.” In: *Proc Biol Sci* 282 (2015 Dec 22), p. 20152026.
- [6] O. Diekmann, J. A. Heesterbeek, and J. A. Metz. “On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations.” In: *J Math Biol* 28 (1990), pp. 365–82.
- [7] K. Hampson et al. “Transmission dynamics and prospects for the elimination of canine rabies.” In: *PLoS Biol* 7 (2009 Mar 10), e53.
- [8] J. H. Huber et al. “Quantitative, model-based estimates of variability in the generation and serial intervals of *Plasmodium falciparum* malaria.” In: *Malar J* 15 (2016 Sep 22), p. 490.
- [9] A. A. King et al. “Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola”. In: *Proceedings of the Royal Society B: Biological Sciences* 282 (2015), p. 2015.
- [10] J. Lessler et al. “Incubation periods of acute respiratory viral infections: a systematic review.” In: *Lancet Infect Dis* 9 (2009 May), pp. 291–300.

- [11] J. Lessler et al. “Times to key events in Zika virus infection and implications for blood donation: a systematic review.” In: *Bull World Health Organ* 94 (2016 Nov 01), pp. 841–849.
- [12] A. L. Lloyd. “Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics.” In: *Theor Popul Biol* 60 (2001 Aug), pp. 59–71.
- [13] J. Ma et al. “Estimating initial epidemic growth rates.” In: *Bull Math Biol* 76 (2014 Jan), pp. 245–60.
- [14] C. E. Mills, J. M. Robins, and M. Lipsitch. “Transmissibility of 1918 pandemic influenza.” In: *Nature* 432 (2004 Dec 16), pp. 904–6.
- [15] H. Nishiura. “Time variations in the generation time of an infectious disease: Implications for sampling to appropriately quantify transmission potential”. In: *Mathematical Biosciences and Engineering* 7 (2010), p. 2010.
- [16] H. Nishiura and G. Chowell. “Theoretical perspectives on the infectiousness of Ebola virus disease.” In: *Theor Biol Med Model* 12 (2015 Jan 06), p. 1.
- [17] H. Nishiura et al. “Transmission potential of the new influenza A(H1N1) virus and its age-specificity in Japan.” In: *Euro Surveill* 14 (2009 Jun 04), 2009 Jun 04.
- [18] C. M. Rivers et al. “Modeling the impact of interventions on an epidemic of ebola in sierra leone and liberia.” In: *PLoS Curr* 6 (2014 Oct 16), 2014 Oct 16.
- [19] A. Svensson. “A note on generation times in epidemic models.” In: *Math Biosci* 208 (2007 Jul), pp. 300–11.
- [20] A. Svensson. “The influence of assumptions on generation time distributions in epidemic models.” In: *Math Biosci* 270 (2015 Dec), pp. 81–9.
- [21] B. P. Taylor, J. Dushoff, and J. S. Weitz. “Stochasticity and the limits to confidence when estimating  $R_0$  of Ebola and other emerging infectious diseases.” In: *J Theor Biol* 408 (2016 Nov 07), pp. 145–54.
- [22] Constantino Tsallis. “What are the numbers that experiments provide”. In: *Quimica Nova* 17.6 (1994), pp. 468–471.



- [23] J. Wallinga and M. Lipsitch. “How generation intervals shape the relationship between growth rates and reproductive numbers.” In: *Proc Biol Sci* 274 (2007 Feb 22), pp. 599–604.
- [24] J. S. Weitz and J. Dushoff. “Modeling post-death transmission of Ebola: challenges for inference and opportunities for control.” In: *Sci Rep* 5 (2015 Mar 04), p. 8751.

## 6 Appendix

[JD: SWP: Restart figure numbering]

### 6.1 Normal approximation

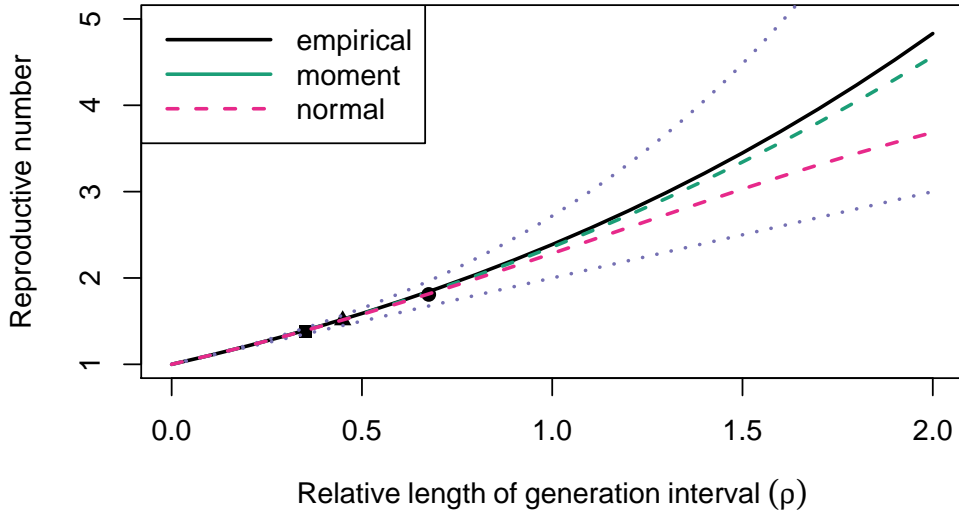


Figure 6: Approximating a generation intervals distribution with a normal distribution has two problems. First, the distribution extends to negative values, which are biologically impossible. Second, as a consequence, the normal approximation predicts a saturating and eventually a decreasing  $r - \mathcal{R}$  relationship for large  $r$ . Parameters and points as in Fig. 3.

### 6.2 Robustness of the gamma approximation

The moment-matching method (approximating  $\mathcal{R}$  based on estimated mean and variance of the generation interval) has an appealing simplicity, and works well for all of the actual disease parameters we tested (the breakdown

for rabies distributions occurs for values of  $\mathcal{R}$  well above observed values). We therefore wanted to compare its robustness given small sample size ( $n = 100$ ) along with that of the more sophisticated maximum likelihood method. Fig. 7 shows results of this experiment. When sample size is limited, estimates using MLE tend to be substantially close to the known true values in these experiments. However, it is worth noting that using the observed moment gives narrower estimates than the two naive estimates even when the sample size is very small.

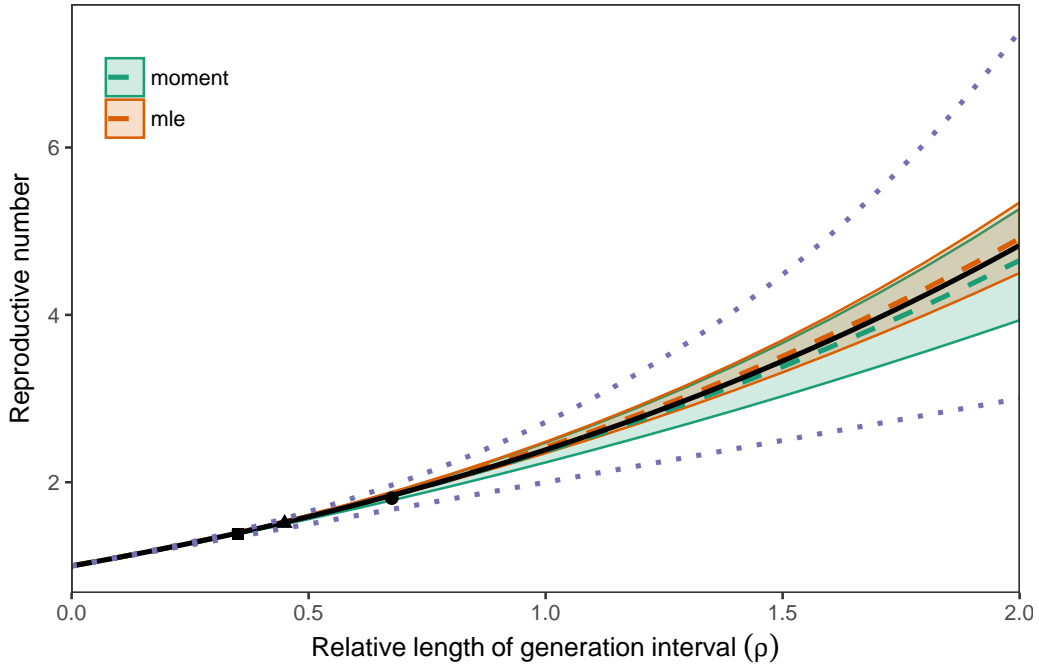


Figure 7: The effect of small sample size on approximated relationship between  $r$  and  $\mathcal{R}$ . (black solid curve) The relationship between growth rate and  $\mathcal{R}$  using a known generation-interval distribution (see Fig. 3). (colored curves) Estimates based on finite samples from this distribution: dashed curves show the median and solid curves show 95% quantiles of 1000 sampling experiments. (dotted curves) Naive approximations based on exponential (lower) and fixed (upper) generation distributions.

[JD: SWP: Add a rabies figure]