

# Exploring how generation intervals link strength and speed of epidemics

Sang Woo Park      David Champredon      Joshua Weitz  
Jonathan Dushoff

June 2017

## Abstract

## 1 Introduction

Much infectious disease modeling focuses on estimating the reproductive number – the number of new cases caused on average by each case. In the specific case where the case is introduced in a fully susceptible population, we talk about the basic reproductive number  $\mathcal{R}_0$ . The reproductive number provides information about the disease’s potential for spread and the difficulty of control. It is often thought of a single number: an average [1] or an appropriate sort of weighted average [3]. But the reproductive number can also be thought of as a distribution across the population of possible infectors: different hosts may have different tendencies to transmit disease.

The reproductive number provides information about how a disease spreads, 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,  $r_0$ , can often be measured robustly early in an epidemic, since the number of incident cases at time  $t$  is expected to follow  $i(t) \approx i(0) \exp(r_0 t)$ . The rate of growth can also be described using the “characteristic time” of exponential growth  $C = 1/r_0$ . This is closely related to the doubling time (given by  $T_2 = \ln(2)C \approx 0.69C$ ).

In a disease outbreaks, the rate of spread is often inferred from case-incidence reports and used to estimate the reproductive number, by means of the *generation-interval distribution*.

In particular,  $\mathcal{R}$  is often calculated from  $r$  and the generation-interval distribution using the generating function approach popularized by [12].

The reproductive number and the rate of spread are linked by the *generation interval* – the interval between the time that an individual is infected by an infector, and the time that the infector was infected [9].

Whereas the rate of spread measures the speed of the disease at the population level, the generation interval measures speed at the individual level. It is typically inferred from contact tracing, sometimes in combination with clinical data. Like the reproductive number, the generation interval can be thought of as a single number (typically its mean), or as a distribution.

Here, we re-interpret the work of [12] using means, variance measures and approximations. By doing so, we hope to shed light on the underpinnings of the relationship between  $r$  and  $\mathcal{R}$ , and to shed light on the factors underlying its robustness.

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

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 somewhat counter-intuitively, if  $r$  is known, then faster disease generations imply a *lower* value of  $\mathcal{R}$  (see Fig. 1).

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

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

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:

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

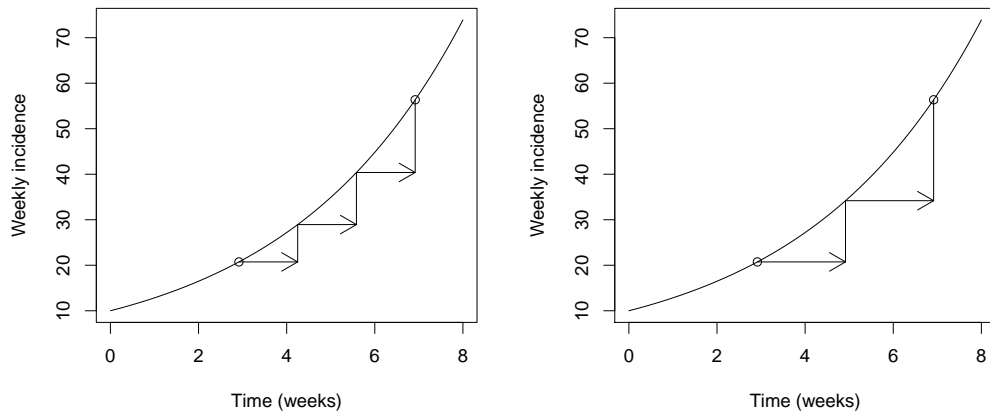


Figure 1: Two hypothetical epidemics with the same growth rate ( $r$ ). Assuming a short generation interval (fast transmission at the individual level) implies a smaller value for the basic reproduction number  $\mathcal{R}_0$  (panel A) when compared to a longer generation interval (slow transmission at the individual level, panel B).

and the *intrinsic* generation-interval distribution:

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

the “intrinsic” interval can be distinguished from “realized” intervals, which can look “forward” or “backward” in time [2] (see also earlier work [9, 7]).

Disease growth is approximately exponential in the early phase because the depletion in the effective number of susceptibles is relatively small. Thus, for the exponential phase, we approximate  $S(t)$  as 1, and write:

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

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

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

### 3 Approximation framework

One way to explore solutions to this equation is by using approximations based on the mean and variance of the generation interval.

[12] used a normal approximation to construct such a moment approximation; here we follow [8] and approximate the generation interval with a gamma distribution, which gives a simpler and more robust answer. For biological interpretability, we describe the distribution using the mean  $\bar{G}$  and the squared coefficient of variation  $\kappa$  (thus  $\kappa = 1/a$ , and  $\bar{G} = \theta/\kappa$ , where  $a$  and  $\theta$  are the shape and scale parameters under the standard parameterization). Substituting into (1) then yields:

$$\mathcal{R} \approx (1 + \kappa\rho)^{1/\kappa} \equiv X(\rho; 1/\kappa), \quad (2)$$

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).

**[SW: This is how Daniel should make comments.]** We define the generalized exponential function  $X$  in an attempt to convey an intuitive understanding. **[JD: Why is Tsallis suppressed?]** For small  $\rho$  it always looks like

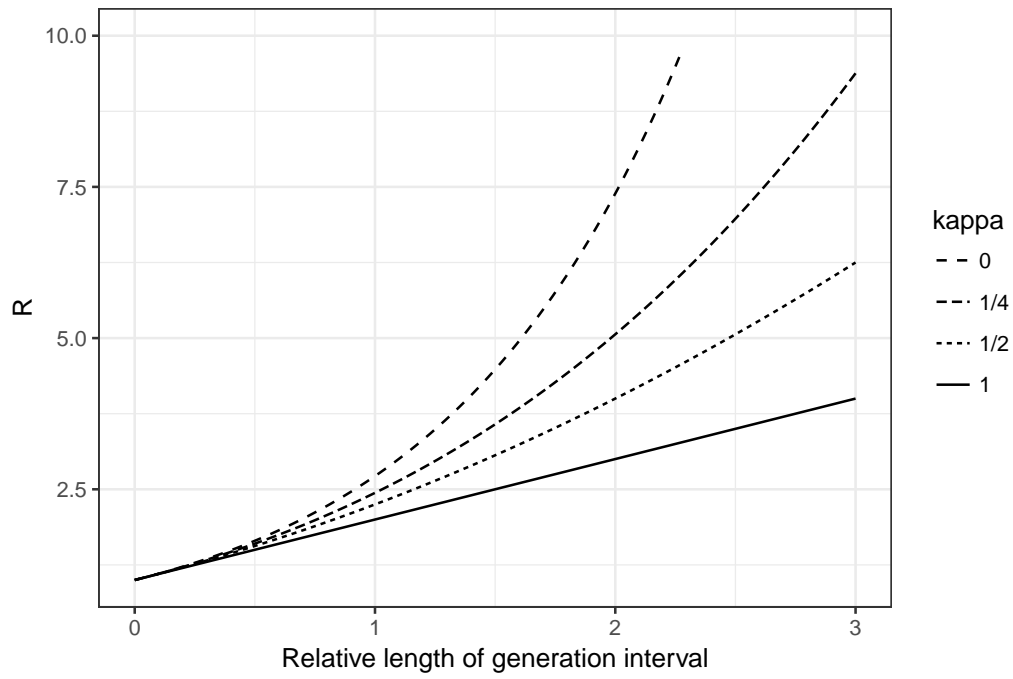


Figure 2: The approximate relationship between mean generation time (relative to the characteristic time of exponential growth) and reproductive number, for different amounts of variation in the generation-interval distribution

$1 + \rho$ , 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  $1/\kappa$ ,  $X$  looks like the “compound-interest approximation” to the exponential; and when  $\kappa \rightarrow 0$ ,  $X(\rho)$  converges to  $\exp(\rho)$ .

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 the generation interval is fixed, then this 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  $\mathcal{R} < \exp(\rho)$ .

**[JD: *Is there a nice interpretation for the  $\kappa = 1$  case?*]**

## 4 Examples

We investigate this approximation approach using three different examples. Our initial investigation of this question was motivated by the West African Ebola Outbreak [Weitz], 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).

### 4.1 Ebola

We first estimated a realistic generation-interval distribution for Ebola virus disease (EVD) using information from [11] and a lognormal assumption for both the incubation and infectious periods. We used the reported standard deviation for the infectious period, and used chose the standard deviation for the incubation period to match the reported standard deviation for the serial interval distribution. We matched the serial interval distribution parameter as it is expected to be very similar to the generation interval distribution for EVD [11]. We then compared the relationship between  $r$  and  $\mathcal{R}$  implied by our realistic distribution (1), and the approximate relationship (2) based

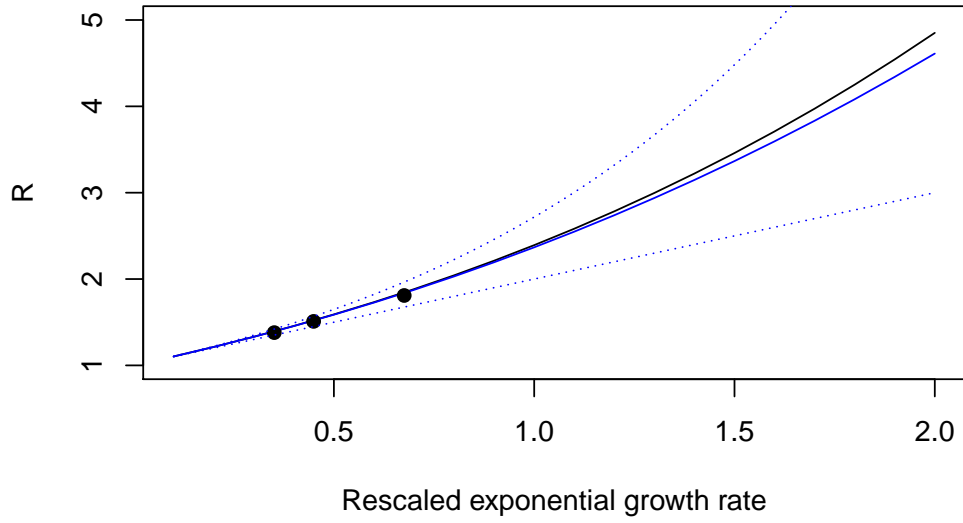


Figure 3: Estimating  $\mathcal{R}$  from Ebola data. (black curve) the relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution based on [11]. (blue curve) the same relationship, approximated using the observed mean and CV. The blue dotted curves show the 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 [11].

only on the mean and CV (see Fig. 3). The approximation works well over relevant parameter ranges, implying that it may be sufficient to understand the mean and CV of the generation-interval distribution when investigating this relationship.

## 4.2 Measles

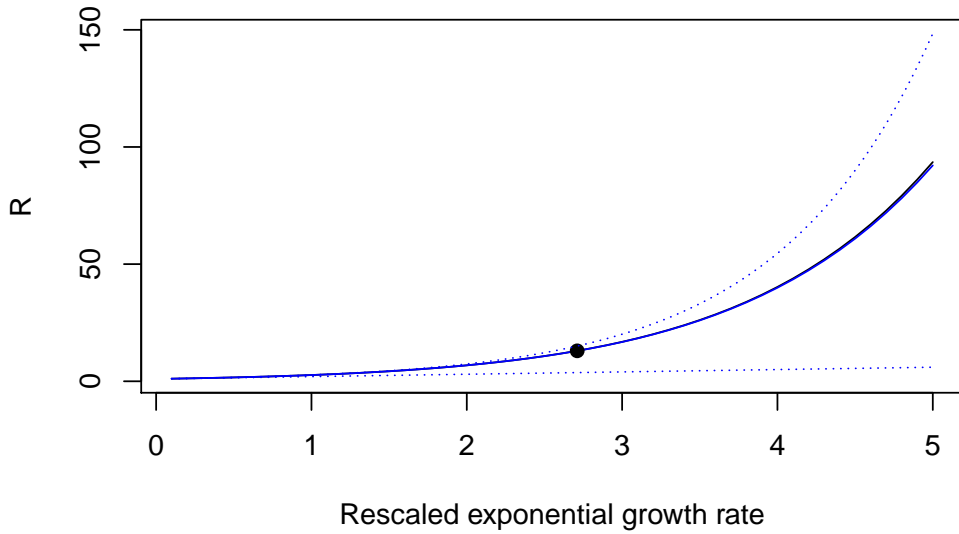


Figure 4: Estimating  $\mathcal{R}$  from measles data. (solid curve) the relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution. (dashed curve) the same relationship approximated using the observed mean and CV (this curve is almost invisible because it overlaps the solid curve). The blue dotted curves show the approximations based on exponential (lower) and fixed (upper) generation distributions. Points are estimates from [Katie]. [JD: *Not Katie.*]

To test whether gamma moment matching works for high  $\mathcal{R}$  values, we applied the moment approximation to measles generation distribution [JD: *Confirm: really moments, right?*]. We found that the approximation matches



the theoretical distribution almost perfectly even when  $\mathcal{R}$  is very high ( $\approx 13$ ).  
[JD: Let's change the graph so that  $R_0$  goes up to 20.]

### 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 Fig. 6. 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. Short intervals will be much more important in driving the speed of the epidemic, and therefore in determining the relationship between  $r$  and  $\mathcal{R}$ . The maximum-likelihood fit to the same data does a better job of matching the observed pattern of short generation intervals (Fig. 6) and of predicting the simulated relationship between  $r$  and  $\mathcal{R}$ , particularly for larger values of  $r$ , where estimates diverge more (Fig. 5).

## 5 Methods

Disease	Ebola	Measles	Rabies
Parameter	Values		
Reproduction number	1.38, 1.51, 1.81 [11]		1.06, 1.32 [4]
Mean incubation period (days)	11.4 [11]	12.77 [5]	24.24 [4]
SD incubation period (days)	7.5 [assumption]	2.67 [5]	29.49 [4]
Mean infectious period (days)	5 [11]	3.65 [6]	3.57 [4]
SD infectious period (days)	4.7 [11]	1.63 [6]	2.26 [4]

Table 1: This is not a caption.

[JD: Can you make the table into a real table (csv or tsv), and then use make or something to make the latex version? Also, let's discuss the table.]

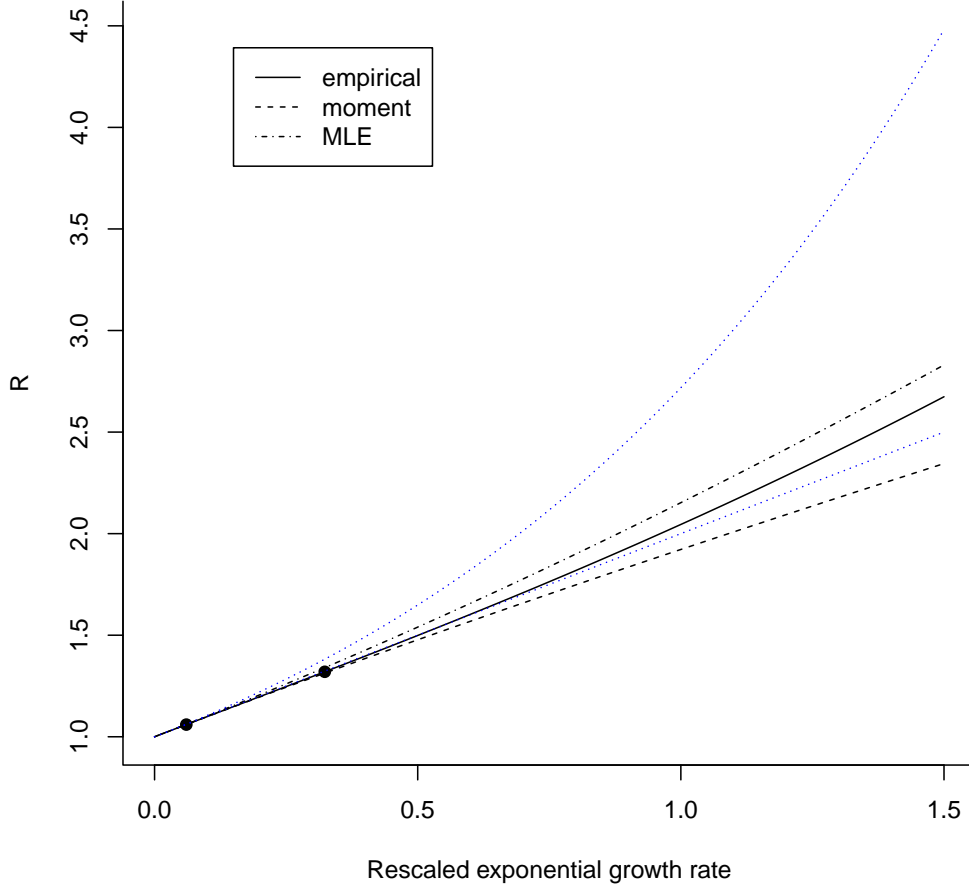


Figure 5: Estimating  $\mathcal{R}$  from rabies data. (solid curve) the relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution. (dashed curve) the same relationship approximated using the observed mean and CV. (dash-dotted curve) the same relationship approximated using the mean and CV calculated from a maximum-likelihood fit. The blue dotted curves show the approximations based on exponential (lower) and fixed (upper) generation distributions. Points are estimates from [4].

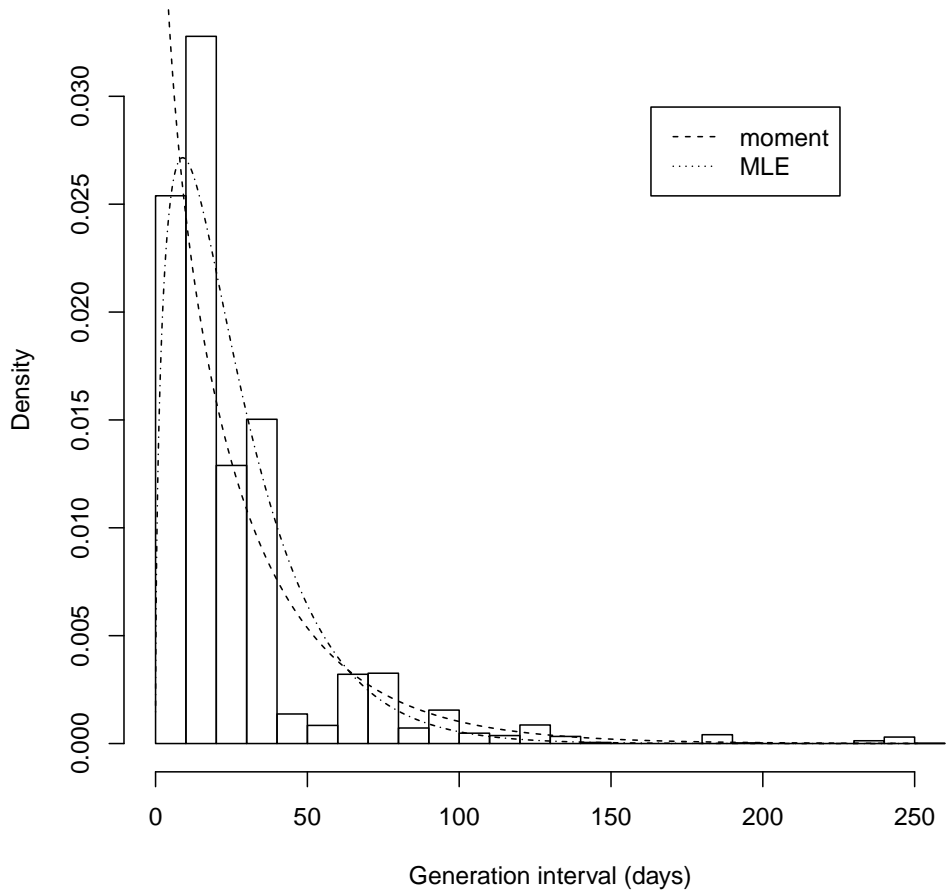


Figure 6: Fitting gamma distributions to generation intervals. Rabies generation distributions estimated from incubation and infectious periods observed by [4], with gamma approximations based on moments (dashed curve) and a maximum-likelihood fit (dotted curve).

In order to obtain a theoretical distribution of generation intervals of each disease, we first gathered information on empirical and estimated distributions of latent and infectious periods of the disease from various sources (see text above). For simplicity, we assumed that latent and infectious periods are independently distributed and that incubation period is equivalent to latent period. We also assumed that transmission rate remains constant over infectious period.

For a disease whose estimates of the distribution parameters were given, we generated the theoretical distributions of the latent and infectious periods by taking 10000 samples from their quantile distributions corresponding to the estimated parameters. We then re-sampled the infectious periods from the 10000 samples by weighing each sample by the length of the period to account for higher chance of transmission of infected individuals with longer infectious periods. For a disease whose empirical distributions of the latent and infectious periods were available, we drew 10000 samples directly from the empirical distributions and re-sampled infectious period samples using the method described above. Finally, a generation interval of each sample was drawn from a uniform distribution whose minimum given by the duration of the latent period and maximum is given by the sum of the durations of the two periods.

We estimated the theoretical relationship between exponential growth rate and reproductive number by discretized version of the equation 1.

$$1/\mathcal{R} = \frac{1}{N} \sum_{i=1}^N \exp(-x_i/C), \quad (3)$$

where  $N$  is the number of samples,  $x_i$  is each sample, and  $C$  is the characteristic time. By varying  $C$ , we were able to obtain a smooth curve describing the relationship between  $r$  and  $\mathcal{R}$ .

Then, we investigated how using estimated distribution affects the relationship by fitting gamma distribution to the 10000 samples using maximum likelihood and method of moments. Using the estimated parameters, 10000 samples were drawn from the corresponding gamma quantile distribution. We applied equation 3 to the gamma samples again to draw out the curve.

To test for robustness of gamma fitting, we took small number of samples from estimated generation interval distribution and fitted gamma. See ?? for results.

## 6 Discussion

Estimating the reproductive number  $\mathcal{R}$  is a key part of calibrating models of infectious-disease spread and of applying them to evaluate control interventions.  $\mathcal{R}$  is often estimated by estimating the exponential rate of growth, and then using a generation-interval distribution to relate the two quantities.

Here, we used a gamma approximation [8] to develop a framework for how estimates of  $\mathcal{R}$  depend on estimates of the mean ( $\bar{G}$ ) and squared CV ( $\kappa$ ) of the generation-interval distribution, as well as on the exponential growth rate ( $r$ ) [eqref]. We attempted to present these approximations in a form conducive to both intuitive understanding [fref], and to propagating 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  $\mathcal{R}$ . The gamma approximation provides estimates that are simpler, more robust and more realistic than those from normal approximations [12]: the key difference between the two distributions is that the gamma distribution does not extend to negative values.

In particular, we show that estimates of  $\mathcal{R}$  increase when the generation time gets longer, and decrease when variation in generation times increase. We also provide mechanistic interpretations. If generation intervals are slower, 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 confirmed the effectiveness of the gamma approximation framework by applying it to three diseases: Ebola, measles, and rabies. We found that MLE-based approximations are generally better than approximations based on moment matching. Nonetheless, the simpler moment-based approximations give good estimates when the generation-interval distribution is not too broad (this is the case for Ebola and measles, but not for rabies Table ??).

Generation interval plays a crucial role in linking the exponential growth rate to the basic reproduction number of an outbreak. Despite much effort put into understanding how generation interval characterizes an epidemic, most of the work is too technical for non-mathematicians and lacks an intuitive interpretation [12, 9, 10, 7]. In this study, we intended to re-inform other researchers by providing a clearer intuition on how shape of the generation interval distribution affects  $\mathcal{R}_0$ .

During the Ebola outbreak in West Africa, many researchers tried to

estimate  $\mathcal{R}_0$  but generation interval distribution was often neglected [CITE Althaus and some other people]. Model based approach may be able to give accurate estimation of  $\mathcal{R}_0$  but are often complicated, and relying on early incidence data can be problematic [CITE Weitz and Dushoff].

On the other hand, estimation based on generation interval can give similar answers even with a much simpler framework. Looking back at the Ebola outbreak, many studies did not consider post burial transmission in their estimation of  $\mathcal{R}_0$  and [CITE Weitz and Dushoff] showed that not including the post burial transmission underestimates  $\mathcal{R}_0$ . Our generation interval framework provides a clear interpretation of the result as post burial transmission increases mean generation interval, resulting in higher  $\mathcal{R}_0$ . Knowing the exact shape of the generation interval distribution is difficult, but thinking about how various transmission routes and epidemic parameters affect the distribution will help researchers understand future outbreaks.

## 7 Sampling

What do we want to do about this?

## Acknowledgments

## References

- [1] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press, 1991.
- [2] D. Champredon and J. Dushoff. “Intrinsic and realized generation intervals in infectious-disease transmission.” In: *Proc Biol Sci* 282 (2015 Dec 22), 2015 Dec 22.
- [3] 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.
- [4] K. Hampson et al. “Transmission dynamics and prospects for the elimination of canine rabies.” In: *PLoS Biol* 7 (2009 Mar 10), e53.
- [5] J. Lessler et al. “Incubation periods of acute respiratory viral infections: a systematic review.” In: *Lancet Infect Dis* 9 (2009 May), pp. 291–300.

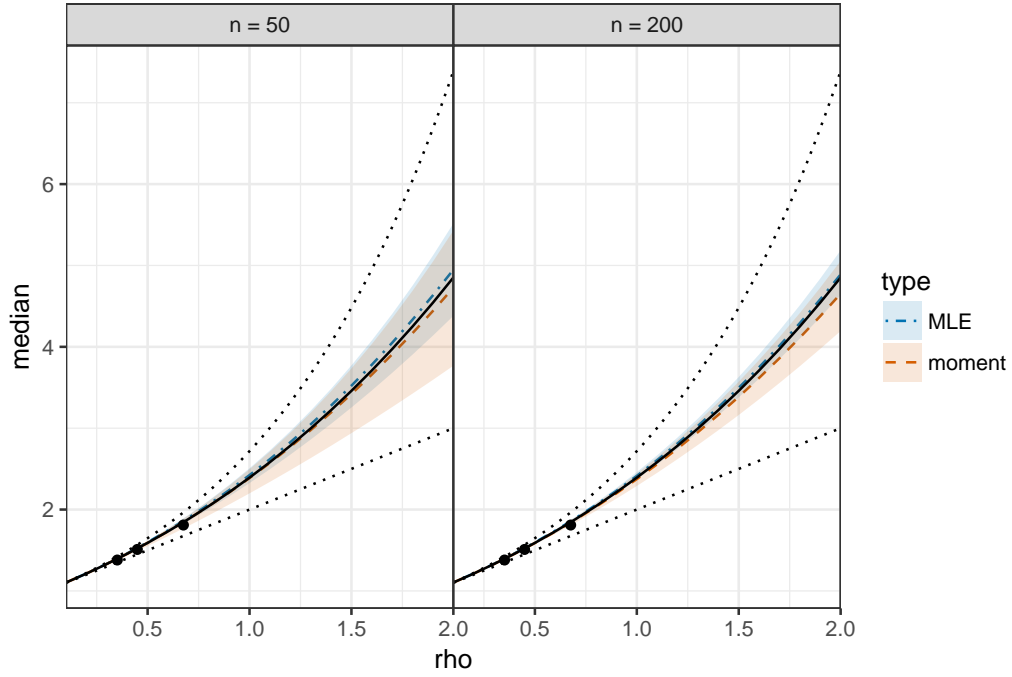


Figure 7: The effect of sample size on estimates of  $\mathcal{R}$ . (black) the relationship between growth rate and  $\mathcal{R}$  using a known generation-interval distribution (see Fig. 3). (colors) estimates based on finite samples from this distribution: curves show the median of 1000 sampling experiments, and shading shows the range where 95% of the results fall Blue shows estimates based on estimated mean and CV. Red shows estimates based on maximum likelihood fits.

- [6] 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.
- [7] H. Nishiura. “Time variations in the generation time of an infectious disease: Implications for sampling to appropriately quantify transmission potential”. In: *MBE* 7 (2010), p. 2010.
- [8] 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.
- [9] A. Svensson. “A note on generation times in epidemic models.” In: *Math Biosci* 208 (2007 Jul), pp. 300–11.
- [10] A. Svensson. “The influence of assumptions on generation time distributions in epidemic models.” In: *Math Biosci* 270 (2015 Dec), pp. 81–9.
- [11] WHO Ebola Response Team. “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.
- [12] 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.