

# Exploring how generation intervals link strength and speed of epidemics

Sang Woo Park      David Champredon      Joshua Weitz  
Jonathan Dushoff

July 2017

## Abstract

## 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 basic reproductive number –  $\mathcal{R}_0$  – in the event that a single infection emerges in an otherwise susceptible population. The basic reproductive number provides information about the disease’s potential for spread and the difficulty of control. It is described in terms of an average [3] or an appropriate sort of weighted average [5]. 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 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,  $r_0$ , 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  $T_2 = \ln(2)C \approx 0.69C$ ).

[DC: *Champredon comment example*]

In disease outbreaks, the rate of spread,  $r$ , is often inferred from case-incidence reports, by fitting an exponential function to the incidence curve [10, 12, 9]. Estimate of the rate of spread,  $r$ , can then be combined with a mechanistic model that includes unobserved features of the disease to estimate the 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 [17].

The *generation interval* denotes the time that elapses between when an individual is infected by an infector, and the time that the infector was infected [13]. While the rate of spread 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 [GenerationMeasurement]. 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 [17] 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 [JSW: and its practical use even when data on generation intervals is limited or hard to obtain].

## 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  $\mathcal{R}$ , 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 this model:

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

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

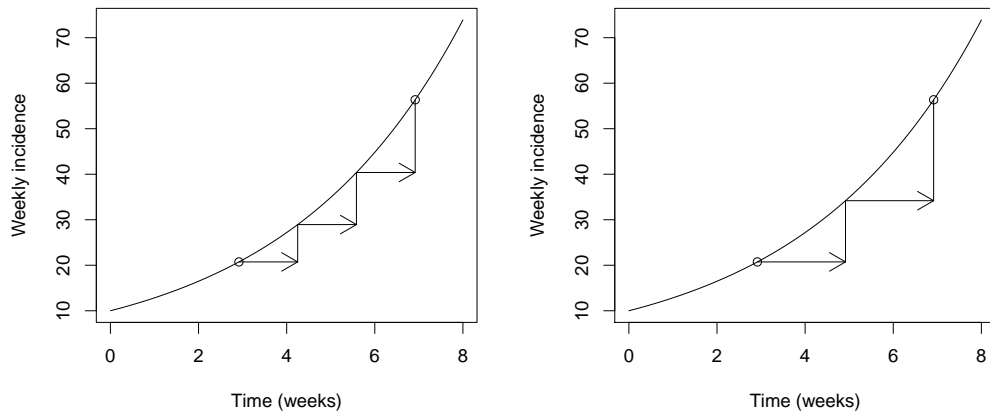


Figure 1: Two hypothetical epidemics with the same growth rate ( $r$ ). Assuming a short generation interval (fast transmission at the individual level) implies a smaller value for the basic reproduction number  $\mathcal{R}_0$  (panel A) when compared to a longer generation interval (slow transmission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

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 [4] (see also earlier work [13, 11]).

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 approximate  $S(t)$  as 1, and write:

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

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

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

The present formulation from equations (1)–(5) summarizes the work of [17] as a prelude to our approximation framework.

### 3 Approximation framework

Here, we propose an approximation approach to estimate  $\mathcal{R}$  from (5). The approximation approach was inspired by the work of [17] who used a normal approximation to construct such a moment approximation. Instead, we use the case study of H1N1 influenza in Japan [12] as an example and approximate the generation interval with a gamma distribution. 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 of the gamma distribution). Substituting the gamma distribution,  $g_\gamma(\tau)$ , into (5) then yields:

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

We try to interpret this equation further by writing:

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

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 define the generalized exponential function  $X$  above – it is equivalent to the Tsallis “q-exponential”, with  $q = 1 - \kappa$  [16] – its shape determines how the estimate of  $\mathcal{R}$  changes with the estimate of normalized generation length  $\rho$ .

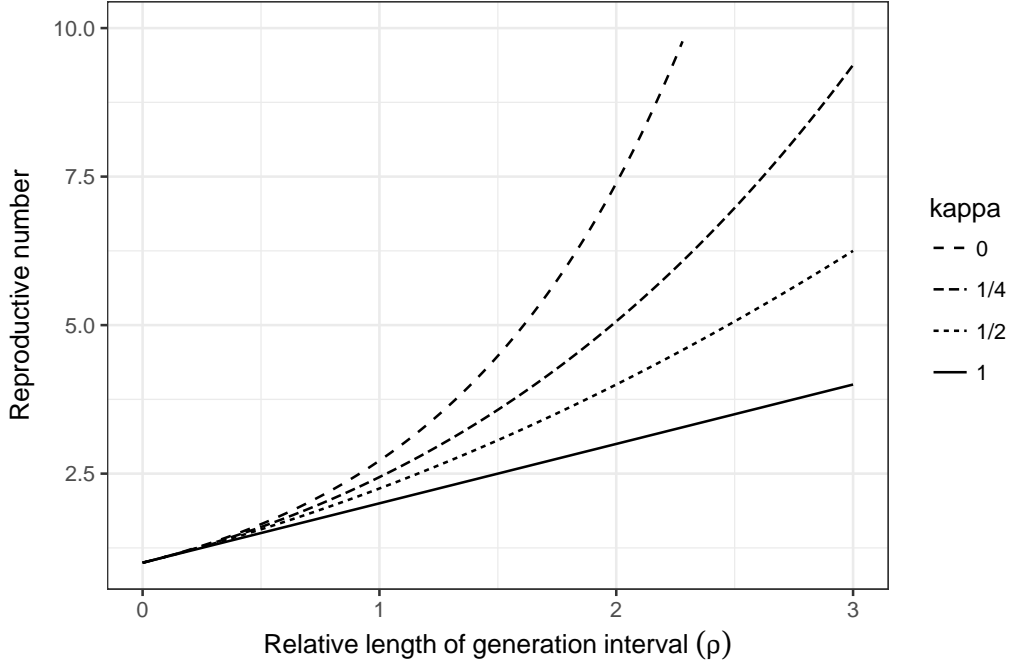


Figure 2: The approximate relationship between mean generation time (relative to the characteristic time of exponential growth) and reproductive number, for different amounts of variation in the **[SWP: gamma]** generation-interval distribution. **[SWP: As JSW pointed out, we need to make it clear that we are changing  $\rho$ , which is equivalent to scaling either  $r$  or  $\hat{g}$  in the gamma approximation.]**

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  $1/\kappa$ ,  $X$  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  $\mathcal{R} < \exp(\rho)$ .

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 is then given directly by the net per capita increase in infections –  $\mathcal{R} - 1$ , where one represents the individual’s recovery – divided by the generation time.

## 4 Examples

We investigate this approximation approach using three different examples. In doing so, we demonstrate the practicality of using our gamma approximation to estimate the basic reproductive number. 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 the West African Ebola Outbreak [18], 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).

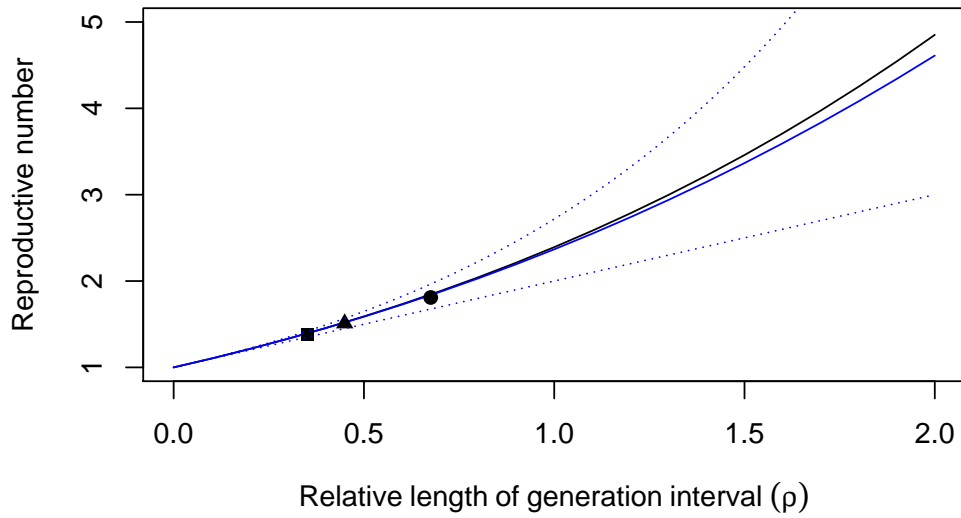


Figure 3: Estimating  $\mathcal{R}$  from Ebola infectious case data. (black curve) the relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution based on [WHO14]. (blue curve) the same relationship, approximated using the observed mean and CV. The blue dotted 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 [WHO14].

## 4.1 Ebola

We first estimated a realistic generation-interval distribution for Ebola virus disease (EVD) using information from [WHO14] and a lognormal assumption for both the incubation and infectious periods. The lognormal was chosen as a simple distribution over positive numbers that is not too similar to the gamma, in order to probe the 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. We matched the serial interval distribution parameter as it is expected to be very similar to the generation interval distribution for EVD [WHO14]. We then compared the relationship between  $r$  and  $\mathcal{R}$  implied by our realistic distribution (5), and the approximate relationship (7) based 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

To test whether gamma moment matching works for high  $\mathcal{R}$  values, we applied the moment approximation to outbreaks of measles in country  $X$  during year  $Y$  [SWP: *assuming that they all share same generation interval distributions*]. The estimates values of  $\mathcal{R}$  for measles are repeatedly amongst the highest of studied infectious diseases [CITE]. Here, we found that the approximation matches the theoretical distribution almost perfectly across our range of interest ( $\mathcal{R}$  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.



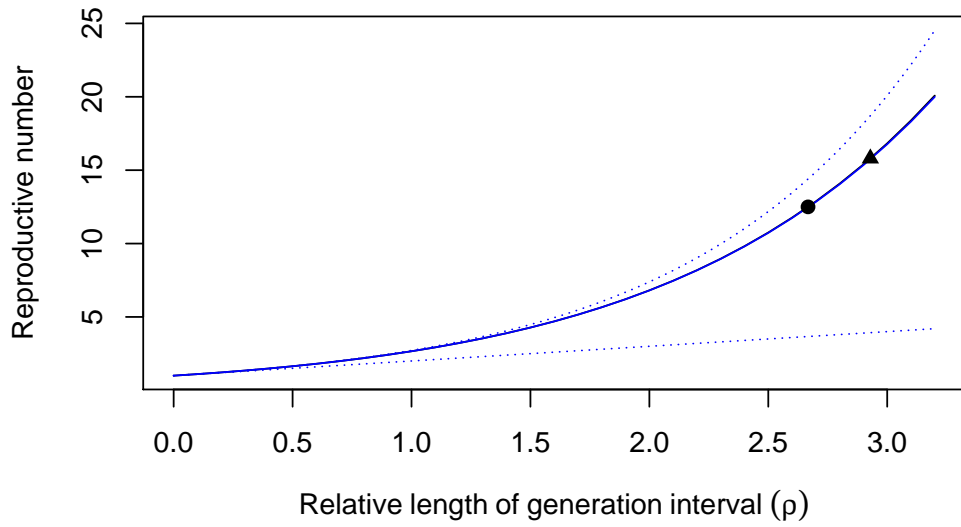


Figure 4: Estimating  $\mathcal{R}$  from measles data. (solid curve) the relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution. (dashed curve) the same relationship approximated using the observed mean and CV (this curve is almost invisible because it overlaps the solid curve) The blue dotted curves show the approximations based on exponential (lower) and fixed (upper) generation distributions. Points are estimates from [2].

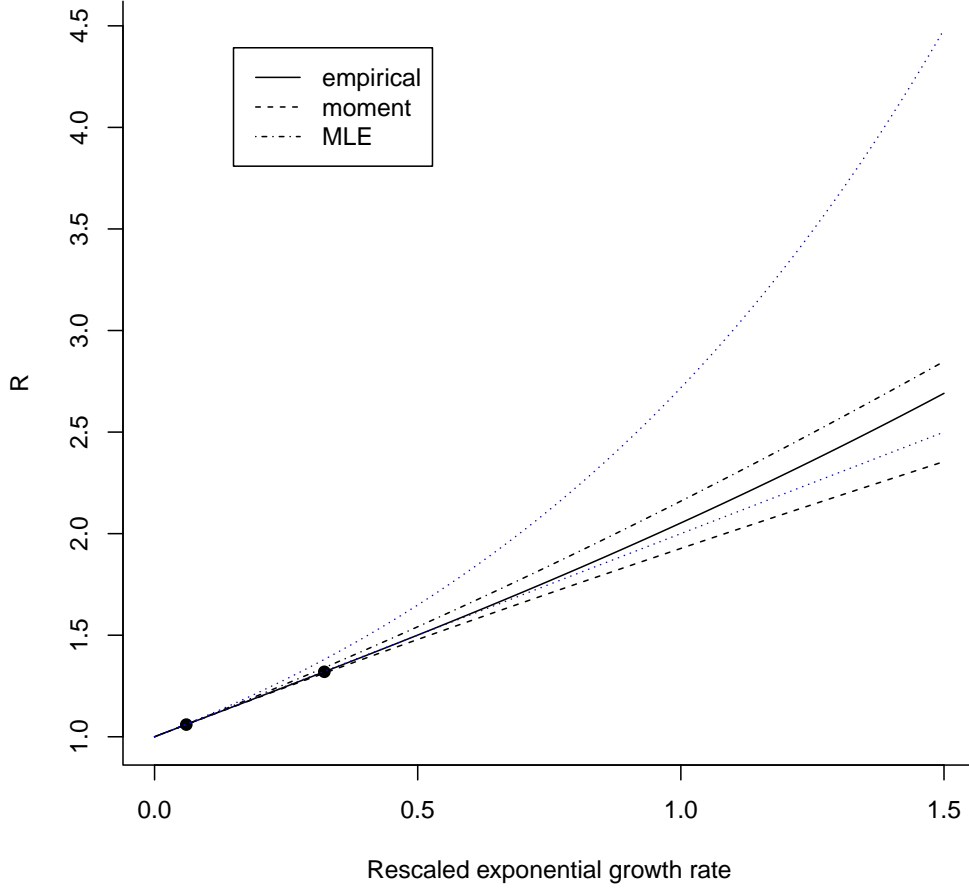


Figure 5: Estimating  $\mathcal{R}$  from rabies infectious case data. (solid curve) the relationship between growth rate and  $\mathcal{R}$  using a realistic generation-interval distribution. (dashed curve) the same relationship approximated using the observed mean and CV. (dash-dotted curve) the same relationship approximated using the mean and CV calculated from a maximum-likelihood fit. The blue dotted curves show the approximations based on exponential (lower) and fixed (upper) generation distributions. Points are estimates from [6].

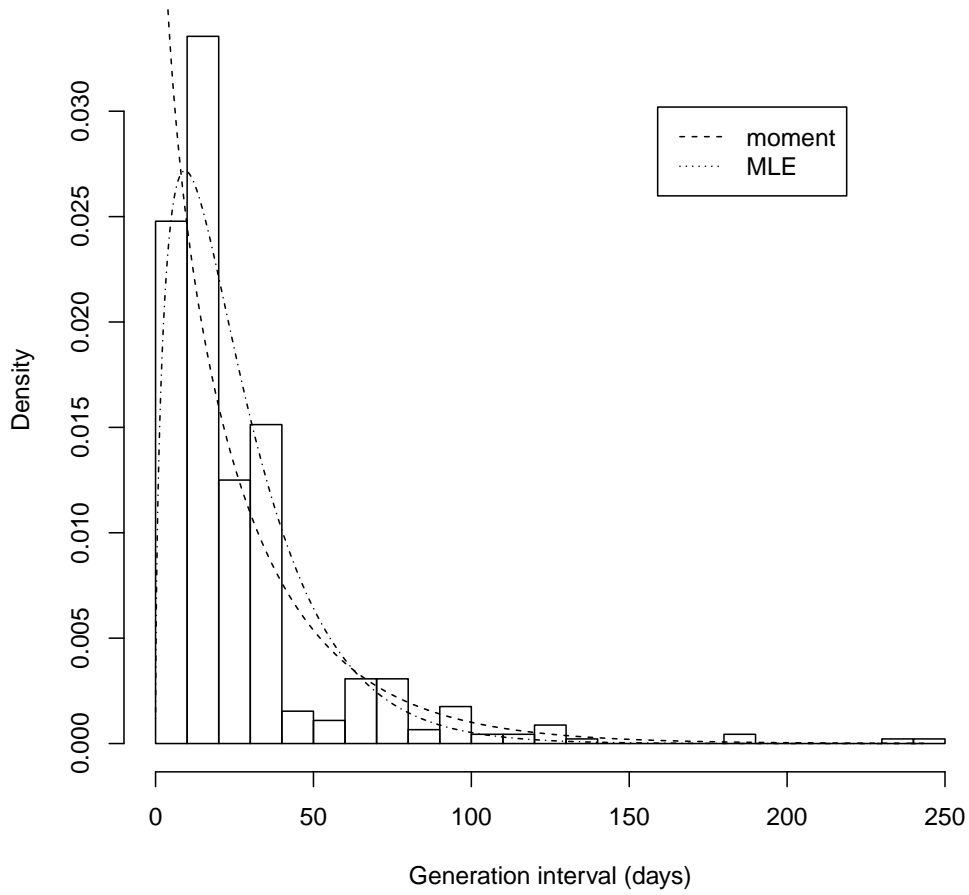


Figure 6: Fitting gamma distributions to generation intervals. Rabies generation distributions estimated from incubation and infectious periods observed by [6], with gamma approximations based on moments (dashed curve) and a maximum-likelihood fit (dotted curve).

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).

#### 4.4 Robust estimation

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 in statistical estimation to that of the more sophisticated maximum likelihood method.

## 5 Methods

Disease	Ebola	Measles	Rabies
Parameter	Values		
Reproduction number	1.38, 1.51, 1.81 [WHO14]	12.5, 15.8 [2]	1.06, 1.32 [6]
Mean incubation period (days)	11.4 [WHO14]	12.77 [LessReic09inc]	24.24 [6]
SD incubation period (days)	7.5 [(see 4.1)]	2.67 [LessReic09inc]	29.49 [6]
Mean infectious period (days)	5 [WHO14]	3.65 [8]	3.57 [6]
SD infectious period (days)	4.7 [WHO14]	1.63 [8]	2.26 [6]

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.

In order to obtain realistic example generation-interval distributions, we first gathered empirical information about each disease. For simplicity, we assumed that latent and infectious periods are independently distributed and (for rabies and Ebola) that incubation period is equivalent to latent period.

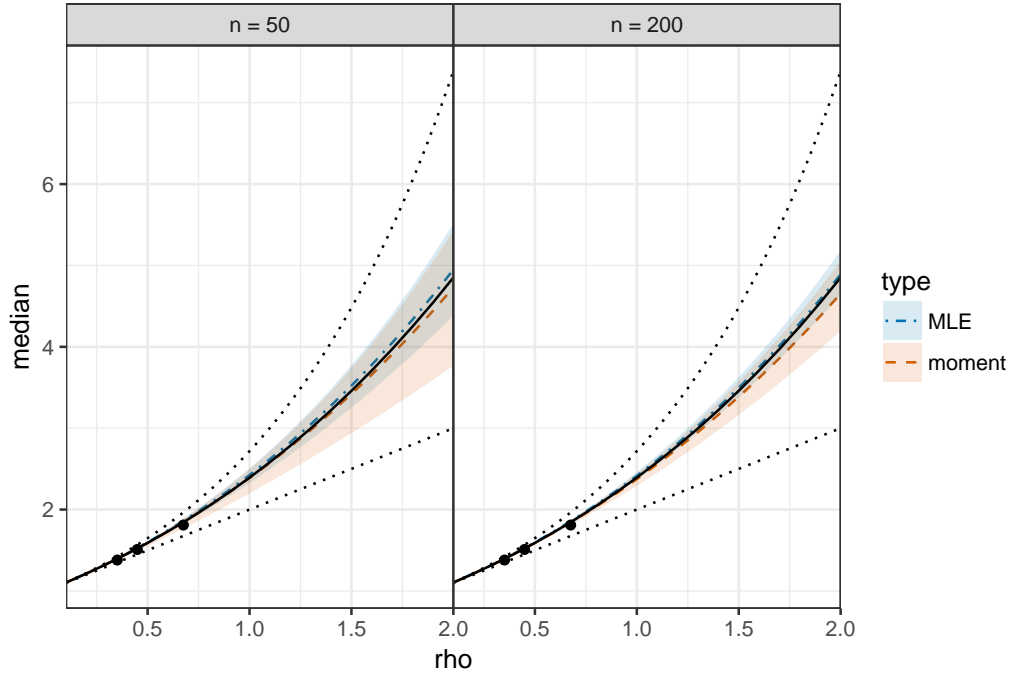


Figure 7: The effect of sample size on estimates of  $\mathcal{R}$ . (black) the relationship between growth rate and  $\mathcal{R}$  using a known generation-interval distribution (see Fig. 3). (colors) estimates based on finite samples from this distribution: curves show the median of 1000 sampling 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.

We also assumed that transmission rate remains constant over the infectious period.

For rabies, we used the reported values from [6] as our sample. For other diseases, we use 10000 quantiles from idealized distributions of the latent and infectious periods corresponding to the estimated parameters. We then re-sampled the infectious periods from our samples by weighting each sample by the length of the infectious period to account for higher chance of transmission of infected individuals with longer infectious periods. Finally, a generation interval was drawn from a uniform distribution whose minimum was the duration of the latent period and whose maximum was 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 5.

$$1/\mathcal{R} = \frac{1}{N} \sum_{i=1}^N \exp(-x_i/C), \quad (8)$$

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

Then, we investigated how using estimated distribution affects the relationship by fitting gamma distributions to samples from this distribution, using either maximum likelihood or moment matching. We used our full samples for the curves in figures Fig. 3–5, and subsampled from these for the sampling example Fig. 7. We used (8) to draw curves related  $r$  to  $\mathcal{R}$ .

## 6 Discussion

Estimating the reproductive number  $\mathcal{R}$  is a key part of characterizing and controlling infectious disease spread.  $\mathcal{R}$  is often estimated by estimating the exponential rate of growth, and then using a generation-interval distribution to relate the two quantities. Such estimates require detailed information on the times between when individuals are first infected and when they infect other susceptible individuals. In this paper, we propose a **[JSW: novel]** approach to estimate  $\mathcal{R}$ , in practice, given minimal information about the full (unknown) generation interval distribution.

Here, we used a gamma approximation [12] to estimate how  $\mathcal{R}$  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 ( $r$ ) (7). We attempted to present these approximations in a form conducive to intuitive understanding of the relationship between speed,  $r$ , and strength,  $\mathcal{R}$  (See Fig. 2). The framework has another advantage when 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 [17]: the key difference between the two distributions is that the gamma distribution does not extend to negative values.

*[SWP: I think we can get rid of the sampling thing and just create gamma confidence interval. I think that will match our paper better.]*

*[SWP: We should mention that WL also derived this result.]*

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. Although simpler moment-based approximations give good estimates when the generation-interval distribution is not too broad (as is the case for Ebola and measles, but not for rabies Table 1), we showed that they are also more sensitive to small sample size.

Generation interval plays a crucial role in linking the exponential growth rate to the basic reproduction number of an outbreak. Much work has been done in exploring such links [17, 13, 14, 11]; here we attempt to broaden these ideas, and present them in an accessible fashion to help scientists build intuition on how the generation-interval distribution affects mediates the relationship between  $\mathcal{R}_0$ .

During the Ebola outbreak in West Africa, many researchers tried to estimate  $\mathcal{R}_0$  from  $r$  [WHO14, Others’p, 1, 7] but uncertainty in the generation-interval distribution was often neglected (but see [15]). During the outbreak, [18] used a generation-interval argument to show that neglecting the effects of post-burial transmission would be expected to lead to underestimates of

$\mathcal{R}_0$ . 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  $\mathcal{R}_0$  for a given estimate of  $r_0$ . 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 better understand future outbreaks.

## Acknowledgments

## 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. “Directly transmitted infections diseases: control by vaccination.” In: *Science* 215 (1982 Feb 26), pp. 1053–60.
- [3] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press, 1991.
- [4] D. Champredon and J. Dushoff. “Intrinsic and realized generation intervals in infectious-disease transmission.” In: *Proc Biol Sci* 282 (2015 Dec 22), p. 20152026.
- [5] 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.
- [6] K. Hampson et al. “Transmission dynamics and prospects for the elimination of canine rabies.” In: *PLoS Biol* 7 (2009 Mar 10), e53.
- [7] 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.
- [8] 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.



- [9] J. Ma et al. “Estimating initial epidemic growth rates.” In: *Bull Math Biol* 76 (2014 Jan), pp. 245–60.
- [10] C. E. Mills, J. M. Robins, and M. Lipsitch. “Transmissibility of 1918 pandemic influenza.” In: *Nature* 432 (2004 Dec 16), pp. 904–6.
- [11] 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.
- [12] 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.
- [13] A. Svensson. “A note on generation times in epidemic models.” In: *Math Biosci* 208 (2007 Jul), pp. 300–11.
- [14] A. Svensson. “The influence of assumptions on generation time distributions in epidemic models.” In: *Math Biosci* 270 (2015 Dec), pp. 81–9.
- [15] 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.
- [16] Constantino Tsallis. “What are the numbers that experiments provide”. In: *Quimica Nova* 17.6 (1994), pp. 468–471.
- [17] 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.
- [18] 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.