Exploring how generation intervals link strength and speed of epidemics

Sang Woo Park David Champredon
Joshua S. Weitz; School of Biological Sciences, School of Physics; Georgia Institu
Jonathan Dushoff

April 3, 2018: In preparation for *PRSB*

Abstract

Infectious-disease outbreaks are often characterized by the reproductive number $\mathcal{R}$ and exponential rate of growth r. 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. However, directly estimating $\mathcal{R}$ is often difficult. In contrast, the rate of growth r can be estimated directly from incidence data while an outbreak is ongoing. $\mathcal{R}$ 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 the mean and variance of an underlying gamma distribution, and find that use of these two moments is often sufficient to capture the r–$\mathcal{R}$ relationship and provide robust estimates of $\mathcal{R}$while an outbreak is ongoing.

1 Introduction

[**SWP:** *Need to put author information (including JSW's)*] [**JD:** *Low priority*]

[**JSW:** *Don't be so mean to Wallinga.*] [**JD:** *We have revised in several places.*]

[**JSW:** *Obviously, I like the gamma over the normal. First real generation intervals can't be normal (because have negative values). Let's use Wallinga Lipsitch to our advantage, and point out that they already noted that normals can be okay, in certain limits where infectiousness is strongly delayed but not in others (e.g., when generation interval is skewed and there is 'early' infectiousness). The gamma does not have that limitation. The positive support alone is worth the switch. Then, our new 3.2. can have the point that in addition to good interpretability features (shown in 3.1) it has good computational features (shown in 3.2)... and then we move on to 3.3 explaining how we will put this to practice.*] [**JD:** *Seems like a mine field; a lot of obstacles to practice, and want to get this out.*]

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]. [**DC:** *is there a problem with the formatting of references?*] [**JD:** *I'm not sure. What makes you ask?*]

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 [**DC:** *do you really want to use "robustly"? sounds too strong because of observation error, etc.*] [**JD:** *I think "often" makes it OK .*] 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, e.g., by fitting an exponential function to the incidence curve [17, 20, 16]. 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 esimate 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 [27].

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 [23]. 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, 14, 10]. Generation interval distributions can be difficult to ascertain empirically, and the generation-function approach depends on an 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, variance 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

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 [7, 21] (see Fig. 2). [**JSW:** *See FIgure 1 of Weitz/Dushoff Sci REports 2015 for one example (I think it's a good practical example).*] [**JD:** *Cited sufficiently, IMO.*]

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

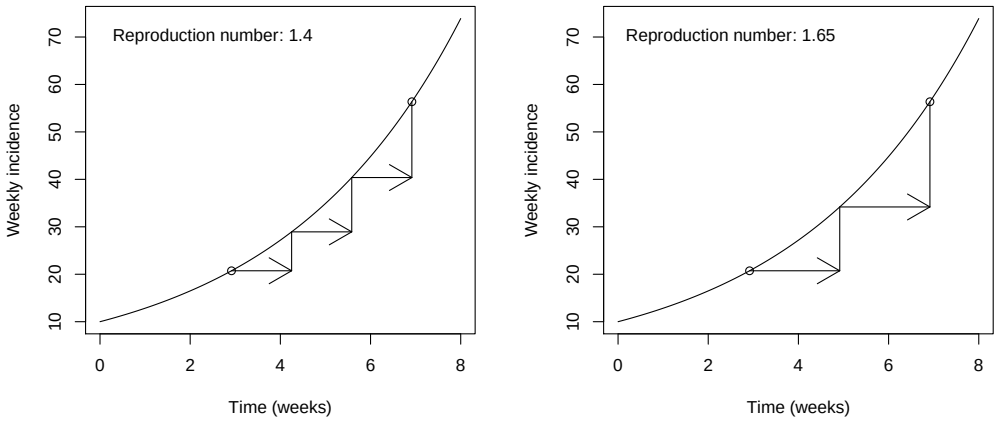

Figure 1: Two hypothetical epidemics with the same growth rate ($r = 0.25$/week) and fixed generation intervals. Assuming a short generation interval (fast transmission at the individual level) implies a smaller reproductive number $\mathcal{R}_0$ (panel A) when compared to a longer generation interval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

[**ChampInPrep**, 9, 27]:

$$i(t) = S(t) \int K(s) i(t-s)\, ds, \tag{1}$$

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.

[**JSW:** *No idea why this reference as 'in prep' article, when it's a standard approach. Shouldn't this cite Wallinga/Lipsitch as well?*] [**JD:** *We should let DC cite himself here, it's an interesting extension. Not feeling the need to cite Wallinga, but don't object.*]

We then have the basic reproductive number:

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

and the *intrinsic* generation-interval distribution:

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

The "intrinsic" interval can be distinguished from "realized" intervals, which can look "forward" or "backward" in time [5] (see also earlier work [23, 18]). In particular, it is important to correct for biases that shorten the intrinsic interval when generation intervals are observed through contact tracing during an outbreak.

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:

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

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 the exact speed-strength relationship:

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

[**JSW:** *Let's call this something which we can invoke later. "The exact speed-strength relationship" or "Euler's exact speed-strength relstionship"*] [**SWP:** *Yes.*]

This fundamental relationship dates back to the work of Euler and Lotka [**Euler**, **Lotka**]. 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 $g(\tau)$ – particularly while an epidemic is ongoing – so we are interested in approximations to $\mathcal{R}$ based on limited information. We follow the approach of [20] and approximate the generation interval with a gamma distribution. This is a biologically more realistic starting point than the standard normal approximation used in many applications, since the gamma distribution is confined to non-negative values.

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

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

[**SWP:** *Calling it the "Gamma-approximated" speed-strength relationship as suggested by JSW.*]

We write:

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

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). [**DC:** *could we make a table of observed values for ρ from real epidemics? we should take the ones used in Examples later on, i.e. Ebola, Measles and Rabies, but also other epidemics. I would say people (including me) are not too familiar thinking in terms of ρ, and that would be helpful to have some numerical values from the real world.*] [**JD:** *I like the idea of adding this to the table for now.*] The longer the generation interval is compared to T_c, the higher the estimate of $\mathcal{R}$ (see Fig. 2). We then explore the behaviour of the

generalized exponential function X defined above (equivalent to the Tsallis "q-exponential", with $q = 1 - \kappa$ [26]): its shape determines how the estimate of $\mathcal{R}$ changes with the estimate of normalized generation length ρ.

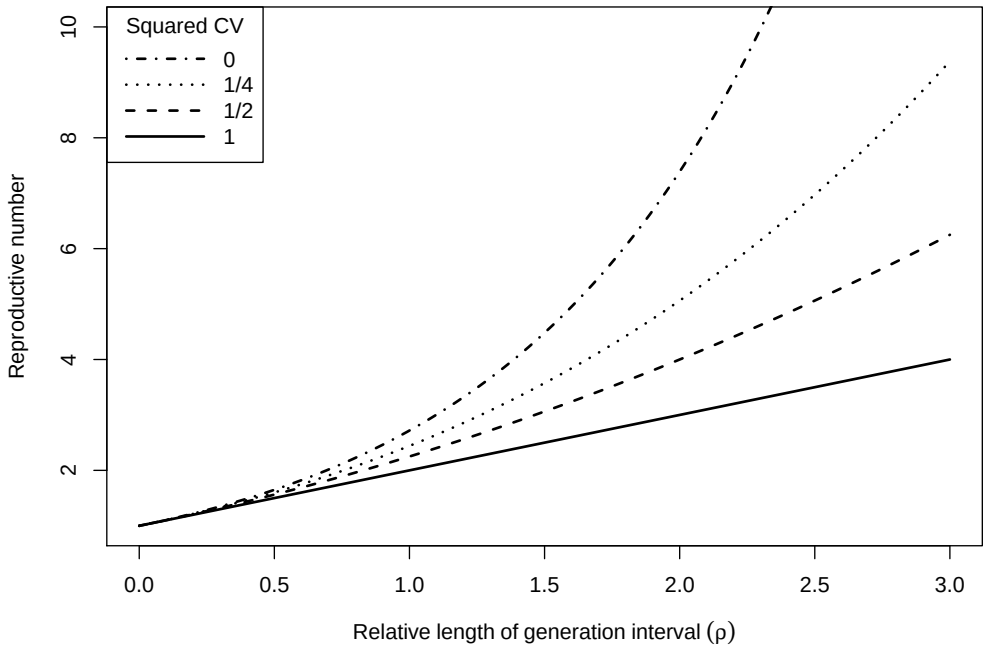

Figure 2: The approximate relationship (6) between mean generation time (relative to the characteristic time of exponential growth, $\rho = r\bar{G} = \bar{G}/C$) and reproductive number. The curves correspond to different amounts of variation in the generation-interval distribution.

For small ρ, 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 $\exp(\rho)$.

The limit as $\kappa \to 0$ 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 average case must cause exactly $\mathcal{R} = \exp(\rho)$ new cases. If variation in the generation time

(i.e., κ) 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

[**JSW:** *Okay I think you need a new 3.2 - in which you evaluate the Gamma compared to other distributions. 3.1 Approximation method, in theory 3.2 Approximation method, in simulation 3.3 Approximation method, in practice (the point of 3.2 is to show how robust the approximation is given totally synthetic data, i.e., on top of being appealing it is robust, this is your current part 7) then 3.3. tells us what you are about to do in the case studies.*]
[**SWP:** *I don't think robust can come before examples. We need to show simple things first and then introduce robustness. Also, all examples are somewhat synthetic because we don't have observed GI here. GI distributions have been synthesized from estimated/observed infectious and latent periods. If we were to add robustness, it should be 3.3.*]

We test our approximation method by generating a pseudo-realistic generation-interval distributions using previously estimated/observed latent and infectious period distributions for different diseases. For each pseudo-realistic distribution, we calculate the "true" relationship between r and $\mathcal{R}$ and compare it with a relationship inferred based on gamma distribution 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 distri-

butions that are likely to be roughly as challenging for our approximation methods as real distributions would be. We construct pseudo-realistic intervals from sampled latent and infectious periods by adding the sampled latent period to a an infection delay chosen uniformly from the sampled infectious period:

$$G_i = E_i + \mathrm{U}(0, I_i), \tag{8}$$

where G_i, E_i and I_i are the sampled intrinsic generation interval, latent period, and infectious period, respectively, and U represents a uniform random deviate. This implicitly assumes that infectiousness is constant across the infectious period [8]. We sample from latent and infectious periods obtained from observations (for empirical distributions), or by using a uniform set of quantiles (for parametric distributions). For the purpose of constructing pseudo-realistic distributions, we do *not* attempt to correct for the fact that observed intervals may be sampled in a context more relevant to backward than to intrinsic generation intervals (see [5]). [**JD:** *DC: I hope you're happy that I had to write this terrible sentence. Feel free to improve it.*] We sample latent periods at random, and infectious periods by length-weighted resampling (since longer infectious period implies more opportunities to infect). For our examples, we used 10000 quantiles for each parametric distribution and 10000 sampled generation intervals for each disease.

We then calculate "exact" relationships (for our pseudo-realistic distributions) by substituting sampled generation intervals into the exact speed-strength relationship (5). This relationship is then compared to the corresponding Gamma-approximated relationship (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 work on the West African Ebola Outbreak [28], 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), measles, and canine rabies.

[**DC:** *for all these examples, do you pay attention to the fact that the observed GIs are* backward *and not intrinsic? Is any adjustment made? If so, we should at least have a short paragraph about it.*] [**JD:** *No adjustment, now explained.*]

4.1 Ebola

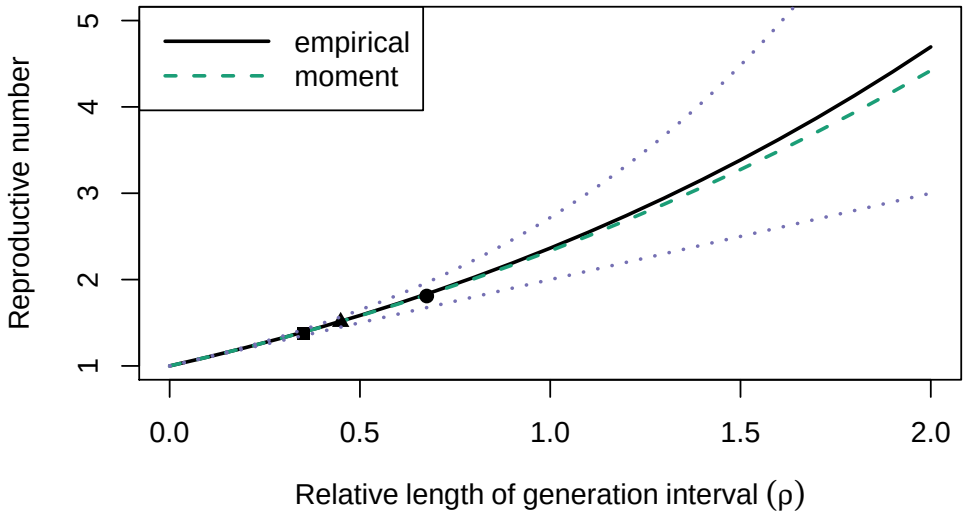

Figure 3: Estimating $\mathcal{R}$ for the West African Ebola Outbreak. (solid curve) The relationship between growth rate and $\mathcal{R}$ using a realistic generation-interval distribution based on [4]. (dashed curve) The same relationship, approximated using the observed mean and CV. (dotted curves) 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] ($r = \ln(2)/T_2$).

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. In contrast to gamma distributed incubation and infectious periods assumed by [4], we used a lognormal assumption for our components because it is straightforward and should provide a challenging test of our gamma approximation (see Appendix for results using gamma components). [**DC:** *I understand why you chose a lognormal. But wouldn't we also want to show that it works (even better?) in the simpler case where we have a gamma distribution for the "true" distribution? At least in appendix if not in the main text.*] [**JD:** *We've been back and forth about this. I kind of think it's as confusing (since readers may think falsely that isn't it trivial. Daniel kind of thinks we might as well. I wonder whether JSW has an opinion.*] [**SWP:** *I have tried adding this figure in appendix.*] We used the reported standard deviation for the infectious period, and chose the standard deviation for the incubation period to match the reported coefficient of variation 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 used our pseudo-realistic distribution to calculate both the exact (5) and Gamma-approximated (6) speed-strength relationships (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.

4.2 Measles

We also applied the moment approximation to a pseudo-realistic generation-interval distribution based on information about measles from [13], [15], and [3]. [**DC:** *What is the distribution you chose?? Please insert the values of ρ and CV for g in Table 1 and not let the reader figure it out.*] [**JD:** *Table updated.*] Incubation periods were assumed to follow a lognormal distribution [13]. Infectious periods were assumed to follow a gamma distribution with coefficient of variation of 0.2 [**Hope-Simpson**, 15, 11]. Since variation in infectious period is relatively low [**Hope-Simpson**, 11], and infectious period is short compared to incubation period, this choice is reasonable (and our results are not sensitive to the details).

Here, we found surprisingly close agreement between the exact and approximate relationships between r and $\mathcal{R}$ across a much wider range of inter-

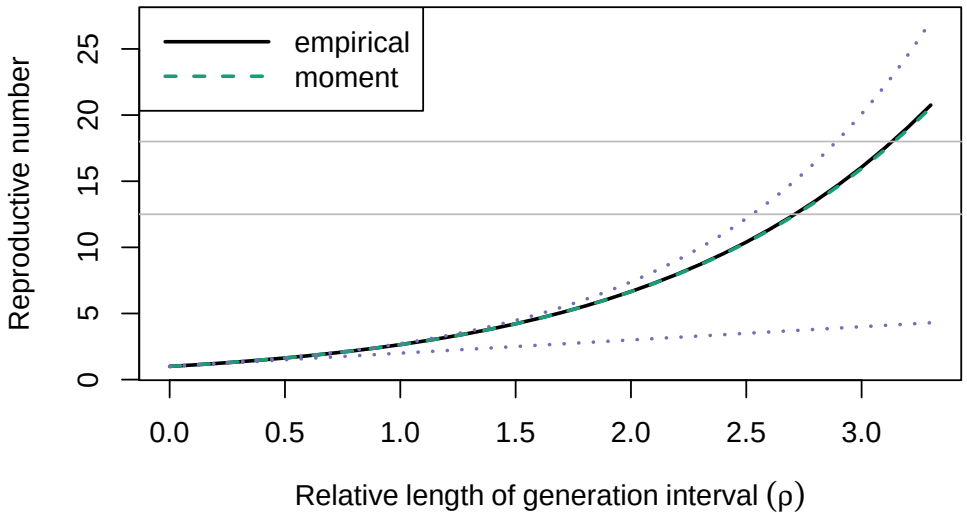

Figure 4: Estimating $\mathcal{R}$ for measles. (solid curve) the relationship between growth rate and $\mathcal{R}$ using a realistic generation-interval distribution. (dashed curve) the same relationship approximated using the estimated mean and CV (this curve is almost invisible because it overlaps 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 estimated by [3]: 12.5-18.

est (a difference of $< 1\%$ for $\mathcal{R}$ up to > 20) (see Fig. 4). 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. [**JD:** *Get Hope-Simpson (or else the summary from Bailey, in which case cite also Bailey, and fix citations above, then I think we can ship.)*]

4.3 Rabies

We did a similar analysis for rabies by constructing a pseudo-realistic generation-interval distribution from observed incubation and infectious period distributions (see Fig. **??**). Since estimates of $\mathcal{R}$ for rabies are near 1, there is small difference between the naive estimates and the gamma approximated speed-strength 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 poor predictions of the moment approximation for higher $\mathcal{R}$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 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).

[**JSW:** *Losing me... so it is not good in the specific case, but it would actually be worse in the general case? So we should be satisfied or unsatisfied? Or is the point that w/ rabies any kind of inference is hard because inferring R when near 1 is generally a hard problem?*] [**JD:** *Better now?*]

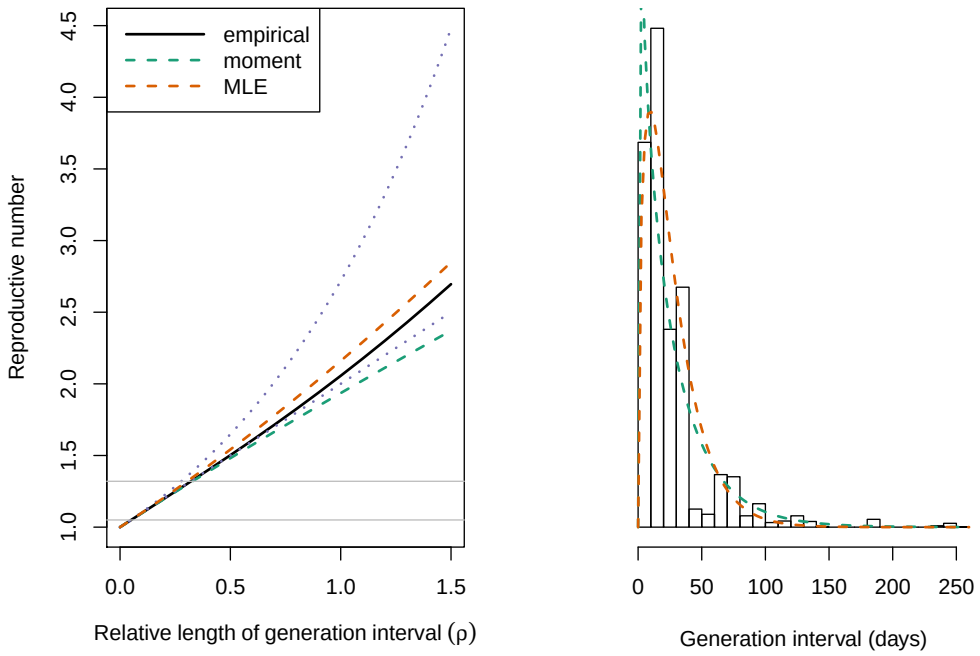

Figure 5: (Left) estimating $\mathcal{R}$ from rabies infectious case data. (solid curve) the relationship between growth rate and $\mathcal{R}$ using a realistic generation-interval distribution. (dashed curve) the same relationship approximated using the observed mean and CV. (dash-dotted curve) the 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 [8]: 1.05 - 1.32. (Right) histogram represents rabies generation interval distributions simulated from incubation and infectious periods observed by [8]. Curves represent estimated distribution of generation intervals using method of moments and MLE (corresponding to approximate $r - \mathcal{R}$ relationships of the left figure).

Disease	Ebola	Measles	Rabies
Parameter		Values	
Reproduction number	1.38, 1.51, 1.81 [4]	12.5-18 [3]	1.05-1.32 [8]
Mean incubation period (days)	11.4 [4]	12.77 [13]	24.24 [8]
SD incubation period (days)	8.1 (see Sec. 4.1)	2.67 [13]	29.49 [8]
Mean infectious period (days)	5 [4]	6.5 [3]	3.57 [8]
SD infectious period (days)	4.7 [4]	2.9 (see Sec. 4.2)	2.26 [8]
Relative length of generation interval (ρ)	0.35, 0.45, 0.68 [4]	NA	NA
Mean generation interval (days)	16.2	15.0	26.6
CV generation interval	0.58	0.21	1.09

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. Ebola parameters in triples represent Sierra Leone, Liberia, Guinea.

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 [27, 23, 18, 24]. 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 a 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 used examples to test the robustness of the framework. We also compared estimates based directly on estimated mean and variance of of the generation interval to estimates based on maximum-likelihood fits.

The gamma approximation for calculating $\mathcal{R}$ from r was introduced by [20], and provides estimates that are simpler, more robust and more realistic than those from normal approximations (see Appendix). [**JD:** *Good enough? It's a tiny bit of a straw person.*] 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 [27]. 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). [**DC:** *This is also something I said in earlier comments: the small sample size stuff should not be in Appendix, but deserves to be in the main text, as I think it's the main selling point of this method. Moreover, I suggest making a more thorough analysis on the sample size. In appendix, you only use* $n = 100$*, but what's really interesting is how well the gamma approx does as* n *increases (say from 10 to 100 by steps of 10), under various assumption for the "true" GI distribution. Once* n *is large, we can use the empirical GI distribution (again it's backward GI so we must be careful) and we do not really care anymore about the gamma approx.*] [**JD:** *We've discussed this, but can experiment with one or two alternatives. Daniel is not psyched about a detailed exploration, but maybe you can convince him.*]

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 ($\bar{G}$, κ 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, 19, 22, 12] but uncertainty in the generation-interval distribution was often neglected (but see [25]). During the outbreak, [28] 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 R0 in models for infectious diseases in heterogeneous populations." In: *J Math Biol* 28 (1990), pp. 365–82.

[7] J. W. Eaton and T. B. Hallett. "Why the proportion of transmission during early-stage HIV infection does not predict the long-term impact of treatment on HIV incidence." In: *Proc Natl Acad Sci U S A* 111 (2014 Nov 11), pp. 16202–7.

[8] K. Hampson et al. "Transmission dynamics and prospects for the elimination of canine rabies." In: *PLoS Biol* 7 (2009 Mar 10), e53.

[9] JAP Heesterbeek and K Dietz. "The concept of $\mathcal{R}_0$ in epidemic theory". In: *Statistica Neerlandica* 50.1 (1996), pp. 89–110.

[10] 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.

[11] M. J. Keeling and B. T. Grenfell. “Disease extinction and community size: modeling the persistence of measles.” In: *Science* 275 (1997 Jan 3), pp. 65–7.

[12] 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.

[13] J. Lessler et al. “Incubation periods of acute respiratory viral infections: a systematic review.” In: *Lancet Infect Dis* 9 (2009 May), pp. 291–300.

[14] 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.

[15] 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.

[16] J. Ma et al. “Estimating initial epidemic growth rates.” In: *Bull Math Biol* 76 (2014 Jan), pp. 245–60.

[17] C. E. Mills, J. M. Robins, and M. Lipsitch. “Transmissibility of 1918 pandemic influenza.” In: *Nature* 432 (2004 Dec 16), pp. 904–6.

[18] 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.

[19] 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.

[20] 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.

[21] K. A. Powers et al. “Impact of early-stage HIV transmission on treatment as prevention.” In: *Proc Natl Acad Sci U S A* 111 (2014 Nov 11), pp. 15867–8.

[22] 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.

[23] A. Svensson. “A note on generation times in epidemic models.” In: *Math Biosci* 208 (2007 Jul), pp. 300–11.

[24] A. Svensson. “The influence of assumptions on generation time distributions in epidemic models.” In: *Math Biosci* 270 (2015 Dec), pp. 81–9.

[25] B. P. Taylor, J. Dushoff, and J. S. Weitz. “Stochasticity and the limits to confidence when estimating R0 of Ebola and other emerging infectious diseases.” In: *J Theor Biol* 408 (2016 Nov 07), pp. 145–54.

[26] Constantino Tsallis. “What are the numbers that experiments provide”. In: *Quimica Nova* 17.6 (1994), pp. 468–471.

[27] 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.

[28] 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

6.1 Ebola example

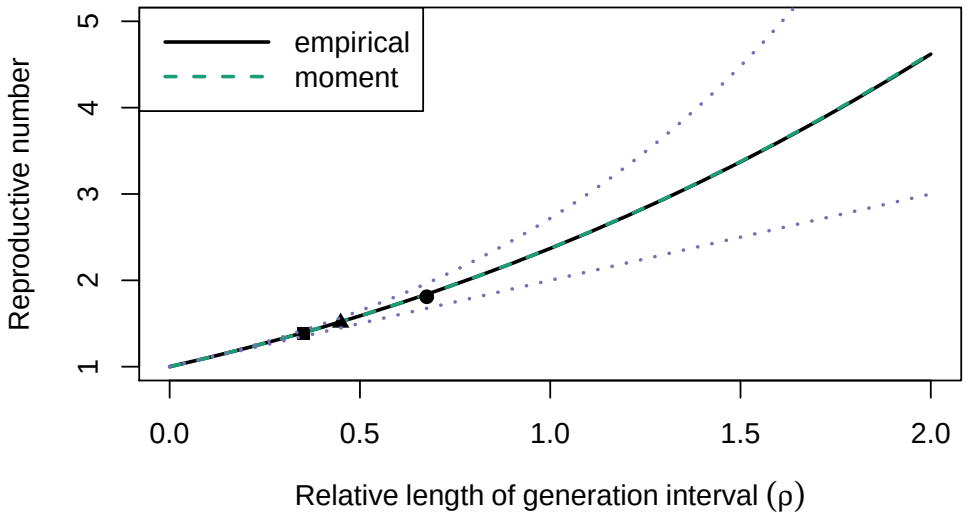

Figure S1: We perform the same analysis as we did in Sec. 4.1 assuming gamma distributed incubation and infectious periods. We find that the gamma approximated speed-strength relationship matches the true relationship almost perfectly in this case. Once again, we adjust the standard deviation of the incubation period to match the reported coefficient of variation in serial interval distributions. Rest of the parameters and points as in Fig. 3

6.2 Normal approximation

6.3 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

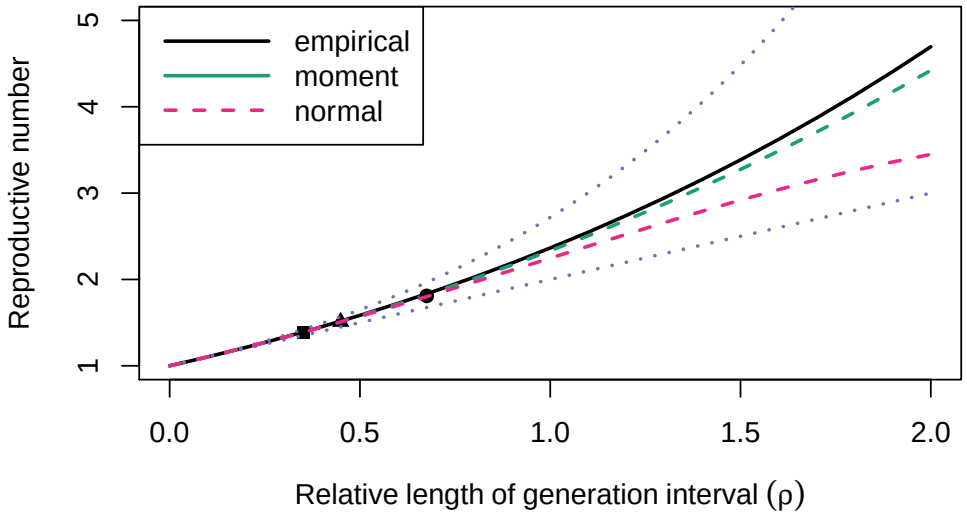

Figure S2: Approximating generation-interval distributions with a normal distribution has two problems. First, the distribution extends to negative values, which are biologically impossible. Second, as a consequence, the normal approximation predicts a saturating and eventually a decreasing $r-\mathcal{R}$ relationship for large r. Parameters and points as in Fig. 3.

for rabies distributions occurs for values of $\mathcal{R}$ well above observed values). We therefore wanted to compare its robustness given small sample sizes along with that of the more sophisticated maximum likelihood method. Fig. S3 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. As we increase sample size, our estimates become narrower. We also find that using the gamma approximated speed-strength relationship gives narrower estimates than the two naive estimates even when the sample size is extremely small ($n = 10$). It is important to note that Fig. S3 only conveys uncertainty in the estimate of coefficient of variation of generation interval distributions. Estimate of mean generation interval varies as well and relative length of generation interval depends on the estimate of the mean. [**SWP:** *Need to say something about why we shouldn't look at the plot vertically here*]

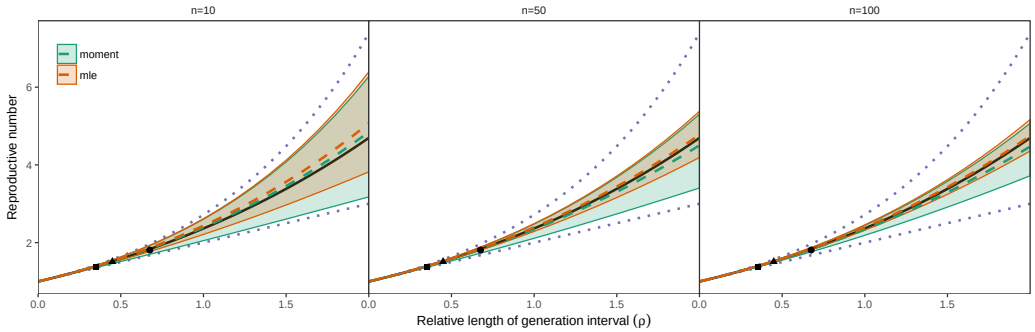

Figure S3: The effect of small sample size on approximated relationship between r and $\mathcal{R}$. (black solid curve) The relationship between growth rate and $\mathcal{R}$ using a known generation-interval distribution (see Fig. 3). (colored curves) Estimates based on finite samples from this distribution: 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.