Reassessing global historical $\mathcal{R}_0$ estimates of canine rabies

Michael Li[1*], Katie Hampson[2], Benjamin M. Bolker[1,3,4], Jonathan Dushoff[1,3,4]

* Corresponding author: lim88@mcmaster.ca
1 Department of Biology, McMaster University, Hamilton, Ontario, Canada
2 Institute of BAH&CM, University of Glasgow, Glasgow, UK
3 Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
4 Institute for Infectious Diseases Research, McMaster University, Hamilton, Ontario, Canada

Keywords: incubation period, infectious period, generation interval

Abstract

Rabies spread by domestic dogs continues to cause tens of thousands of human deaths every year in low- and middle-income countries. Despite this heavy mortality burden, rabies is often neglected, perhaps because it has been eliminated from high-income countries through mass dog vaccination. Estimates of the intrinsic reproductive number ($\mathcal{R}_0$) of canine rabies from a wide range of times and locations are low (values <2), with narrow confidence intervals. The persistence of rabies in environments that vary enormously in ecological conditions is thus surprising. We combined incidence data from historical outbreaks of canine rabies from around the world (1917-2003) with high-quality contact-tracing data from Tanzania (2002-present) to investigate initial growth rates (r), generation intervals distributions (G) and reproductive numbers ($\mathcal{R}_0$). We used a novel combination of bootstrap and Bayesian methods to estimate uncertainties for $\mathcal{R}_0$, combining the uncertainties associated with r and G. Our $\mathcal{R}_0$ estimates are larger than previous estimates (on average closer to 2), with wider confidence intervals. Our novel hybrid approach for estimating $\mathcal{R}_0$ and its uncertainty

is applicable to other disease systems where researchers estimate $\mathcal{R}_0$ by combining estimates of r and G.

Introduction

Canine rabies, primarily spread by domestic dogs, is a vaccine-preventable disease that continues to cause tens of thousands of human deaths every year in low- and middle-income countries (LMICs) (**??**). Canine rabies has been effectively eliminated from high-income countries by mass dog vaccination (**?**). Despite the effectiveness of vaccinating dogs, rabies continues to cause many deaths and large economic losses in LMICs due to the limited implementation of rabies control strategies (**?**). The past two decades have seen an increase in rabies control efforts — including dog vaccination campaigns and improvements in surveillance (**?????**). More recently, the World Health Organization (WHO) and partners (OIE, FAO, GARC) joined forces to support LMICs in eliminating human deaths from dog-mediated rabies by 2030 (**??**). Mass dog vaccination campaigns have begun in some LMICs and are being scaled up (**??**). An understanding of rabies epidemiology — in particular, reliable estimates of the basic reproductive number ($\mathcal{R}_0$), a quantitative measure of disease spread that is often used to guide vaccination strategies — could inform rabies control efforts.

$\mathcal{R}_0$ is defined as the expected number of secondary cases generated from each primary case in a fully susceptible population (**?**). Estimates of $\mathcal{R}_0$ using various methods (i.e., direct estimates from infection histories, epidemic tree reconstruction, and epidemic curve methods) based on historical outbreaks of rabies have generally been surprisingly low, typically between 1 and 2 with narrow confidence intervals (**???**). It is surprising that a wide variety of regions and time periods with different densities and population structures of domestic dogs would have such similar values of $\mathcal{R}_0$. In contrast to diseases with large $\mathcal{R}_0$ (e.g. measles with $\mathcal{R}_0 >10$ (**?**)), $\mathcal{R}_0$ estimates

for rabies imply that control through vaccination should be relatively easy (compared to e.g., rinderpest with $\mathcal{R}_0 \approx 4$ (**?**)). With such a low $\mathcal{R}_0$ one might expect rabies to fade out from behavioural control measures combined with stochastic fluctuations, even in the absence of vaccination.

Here we revisit and explore why rabies, with apparently low $\mathcal{R}_0$, nonetheless persists in many countries around the world. Such persistence suggests that rabies's potential for spread, and therefore the difficulty of rabies control, may have been underestimated. In this paper, we will combine information derived from epidemic curves with a high-resolution contact tracing dataset that provides large number of observed generation intervals (which is rare for infectious disease studies) to estimate $\mathcal{R}_0$. This reassessment can improve the estimation of $\mathcal{R}_0$ and understanding of disease control more generally.

Materials and Methods

$\mathcal{R}_0$ is often estimated by combining two other epidemiological quantities: the initial growth rate of an epidemic (r) and the generation interval (G) distribution, where a G is defined as the time between successive infections along a transmission chain. The growth rate r is often estimated by fitting a growth rate to time series data from the early stages of epidemics. G is an individual level quantity that measures the time between an individual getting infected to infecting another individual. The generation interval distribution is the natural way to link r and $\mathcal{R}_0$ (**??**). During an outbreak in a fully susceptible population, $\mathcal{R}_0$ can be calculated from r and the G distribution by the Euler-Lotka equation (**?**)

$$\mathcal{R}_0 = \frac{1}{\sum_{t=1}^{\infty} G(t)e^{-rt}}, \tag{1}$$

where t is time, and $G(t)$ is the generation interval distribution. This formula is convenient to calculate point estimates of $\mathcal{R}_0$; however, propagating uncertainty from the estimates of r and the G distribution can be difficult.

Initial growth rate

Disease incidence typically increases approximately exponentially during the early stages of an epidemic. The initial growth rate r is often estimated by fitting exponential curves from near the beginning to near the peak of an epidemic. However, growth rates estimated from an exponential model can be biased downward, overconfident, and sensitive to the choice of fitting windows (**?**). Here we used the logistic curves provide more robust estimates of r (**??**). We select our fitting window consistently for each outbreak as follows: ML: Feed me with the new method of window selection.

Observed Generation intervals

Transmission events are generally hard to observe for most diseases. In an earlier, influential paper, estimated generation intervals were constructed by summing two quantities: a latent period (the time from infection to infectiousness), and a wait time (time from infectiousness to transmission) (**?**). Since clinical signs and infectiousness appear at nearly the same time in rabies, the incubation period (the time from infection to clinical signs) is routinely used as a proxy for the latent period. In their analysis, latent (really, incubation) periods and infectious periods were randomly and independently resampled from empirically observed distributions (**?**), and then waiting times sampled uniformly from the selected infection periods.

However, this approach for constructing G values (i.e., summing independently resampled values of incubation and infectious periods) does not account for the possibility of multiple transmissions from the same individual and the correlation between time distributions and biting behaviour. Figure 1 illustrates the generation intervals

of a single transmission event from a rabid animal (comprising a single incubation period plus a waiting time) and multiple transmission events from a rabid animal (comprising a single incubation period and three waiting times). For diseases like rabies, where transmissions links (and generation intervals) are observable, multiple transmissions and possible correlation structures are all accounted for within the observation processes [[BB: unclear]].

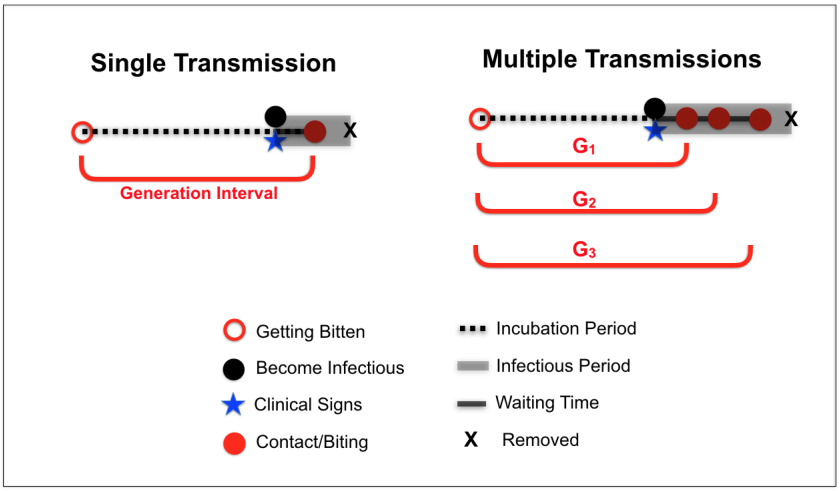

Figure 1: **Decomposing generation intervals.** Generation intervals start when a focal animal acquires infection (open red circle) and end after a period of viral replication (dashed line) when an animal shows clinical signs (blue star), becomes infectious (solid black circle) and infects another animal — in rabies the onset of clinical signs and of becoming infectious are closely synchronized. Once the infectious period (grey block) starts, there is a wait time (solid black line) until a susceptible host (solid red circles) is bitten. The infectious period ends with the death of the focal host (black X). The generation interval is the interval between getting infected and infecting a new case (red interval between open and solid circles). (right) If a single biter transmits multiple times, the wait times are generally different, but the incubation period is the same for each transmission event.

Correcting for vaccination

In a population where some animals are not susceptible, calculations based on estimates of r and the G distribution (1) estimate the *realized* average number of cases

per case, also known as the effective reproductive number $\mathcal{R}_e$. In the case of rabies, vaccination is the only known cause of immunity (case fatality in dogs is believed to be 100%). For a given population with ν vaccination proportion, the estimated $\mathcal{R}_0$ with vaccination correction is the following:

$$\mathcal{R}_0 = \frac{\mathcal{R}_e}{(1-\nu)}. \tag{2}$$

Data and material

We used data from December 2002 – November 2022, from an ongoing contact tracing project in Tanzania (**??**). Since 2002, there are 8636 domestic dog recorded events (i.e., domestic dogs bitten by an animal), and 3552 suspected rabid dogs in the Serengeti, Tanzania. Transmission events were documented through retrospective interviews with witnesses, applying diagnostic epidemiological and clinical criteria from the six-step method (**?**). Each dog was given a unique identifier and date of the bite and clinical signs were recorded if applicable and available. 2132 of dog transmissions were from unidentified domestic animals or wildlife. We restricted our analysis in this paper to domestic dog transmissions (i.e., dog to dog), and obtained 293 directly observed generation intervals (i.e. both biter and secondary case have time bitten recorded). There were four observed dogs with multiple exposures (i.e., bitten by different identified biters), generating extra generation intervals, but it is unclear which transmission event transmitted rabies to these dogs. For simplicity, we omitted these four dogs and their generation intervals from our analysis.

Fitting and Propagating Parameter Uncertainties

To propagate uncertainties for both r and G, we used a hybrid approach. We first fitted logistic models, with negative binomial observation error, to incidence data to estimate r implemented in the R package "epigrowthfit". ML: cite We then compute a

sample of 300 $\hat{\mathcal{R}}_0$ values using equation (2); for each value of $\hat{\mathcal{R}}_0$, we first draw a value of $\hat{r}$ from a normal distribution from the estimates of the logistic fit and 100 G from the empirical contact tracing data. ML: fix me, use R to fill in the sample numbers. We then matched bootstrap samples of G to samples from the r samples to produce a range of estimates for $\mathcal{R}_0$. This hybrid bootstrap approach incorporates both sources of uncertainties from r and G (frequentist empirical bootstrap) when calculating $\mathcal{R}_0$ estimates. Finally, we take the 25, 50, 97.5% percentiles of the distribution of $\mathcal{R}_0$ estimates for each rabies outbreak.

Results

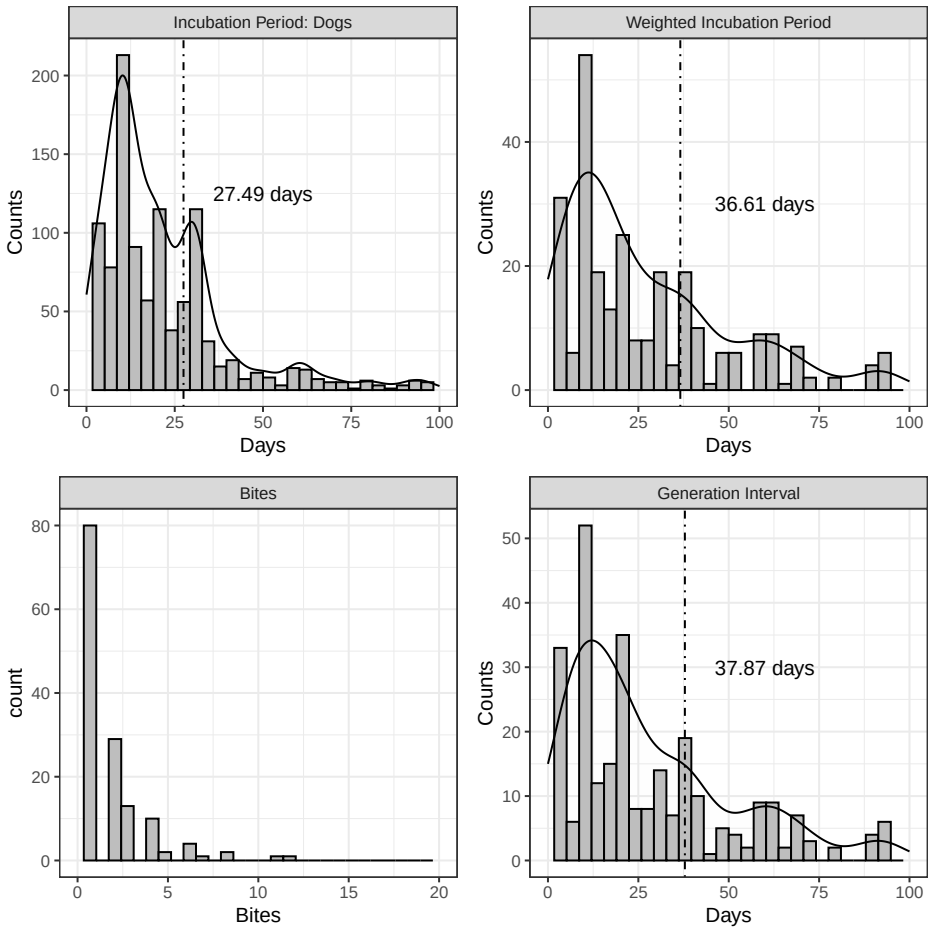

Figure 2: **Empirical distributions from contact tracing data.** Top left is the distribution of observed incubation periods. Top right is the distribution of infectious periods weighted by each dog's biting frequency (biting frequency shown bottom left). The weighted distribution corresponds to the contribution of incubation periods to generation intervals, which are shown bottom right. Dash-dotted lines show the means of each time-interval distribution. [[BB: one less sig digit in mean values? Lower left ("bites" panel) seems like maybe it should be of the same style as type = h for base plots (i.e. geom_segment from 0 to number of counts, with a point at the top?) Why is the top right facet label "incubation period: dogs" ? Can we have letter labels on the subplots so we can refer to a, b, c, d, rather than top left/top right/etc.?]]

Figure 2 shows the empirical distributions of the observed incubation periods, rabid dog biting frequency, and generation intervals from contact tracing data. The mean observed incubation period is 27.5 days (n = 1109 dogs), the mean biting frequency is 1.65 bites per rabid dog, and the weighted mean incubation period is 36.6 days (n = 143 biting dogs). The mean observed generation interval is 37.9 days

(n = 143 primary infections resulting in 293 secondary cases), which is significantly larger than the mean generation interval constructed from independently summing incubation periods and wait times (24.9 days (**?**)). The weighted incubation period distribution resembles much closer to the generation interval distribution than the incubation period of all dogs.

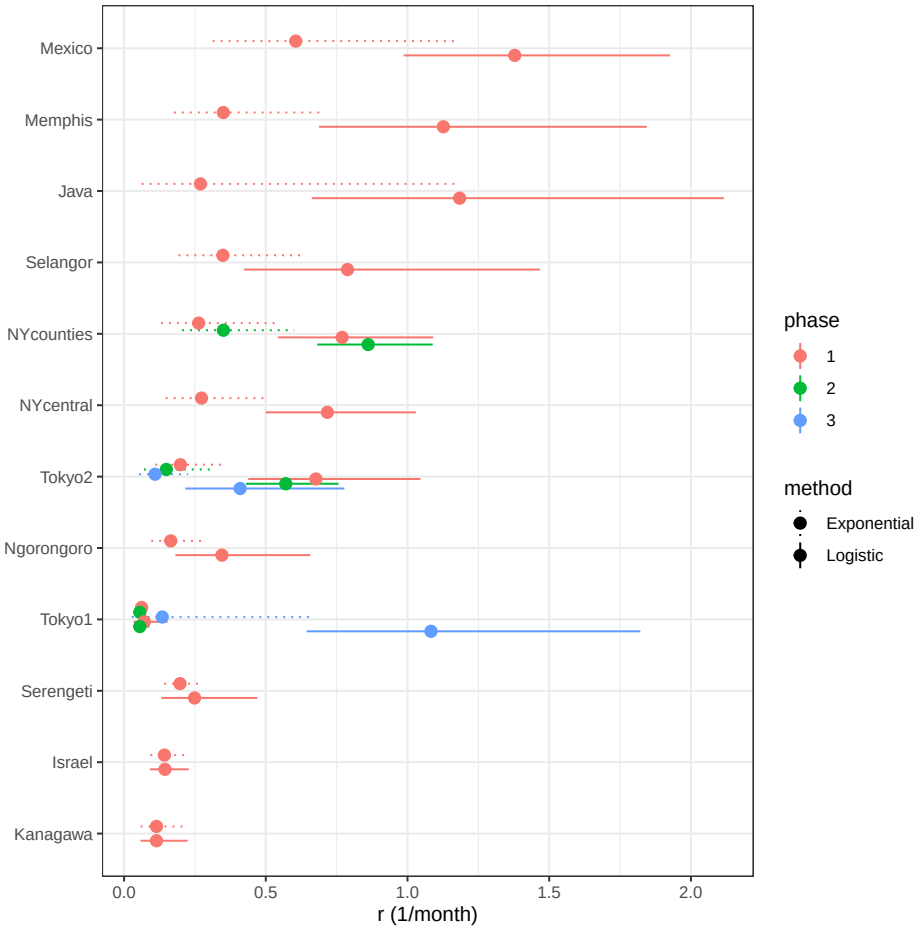

Figure 3: **Growth rate estimates for global historical outbreaks of rabies.** Estimates and 95% confidence intervals of r in global historical outbreaks estimated from exponential (dotted) and logistic (solid) model fits.
Different colors represents different phases from the times series data.

We estimated r from historical outbreak data (Figure 3). For a direct comparison, we also estimated r from an exponential model. Overall, r estimates from the logistical

model are larger with wider confidence intervals compared to r estimates from the exponential model.

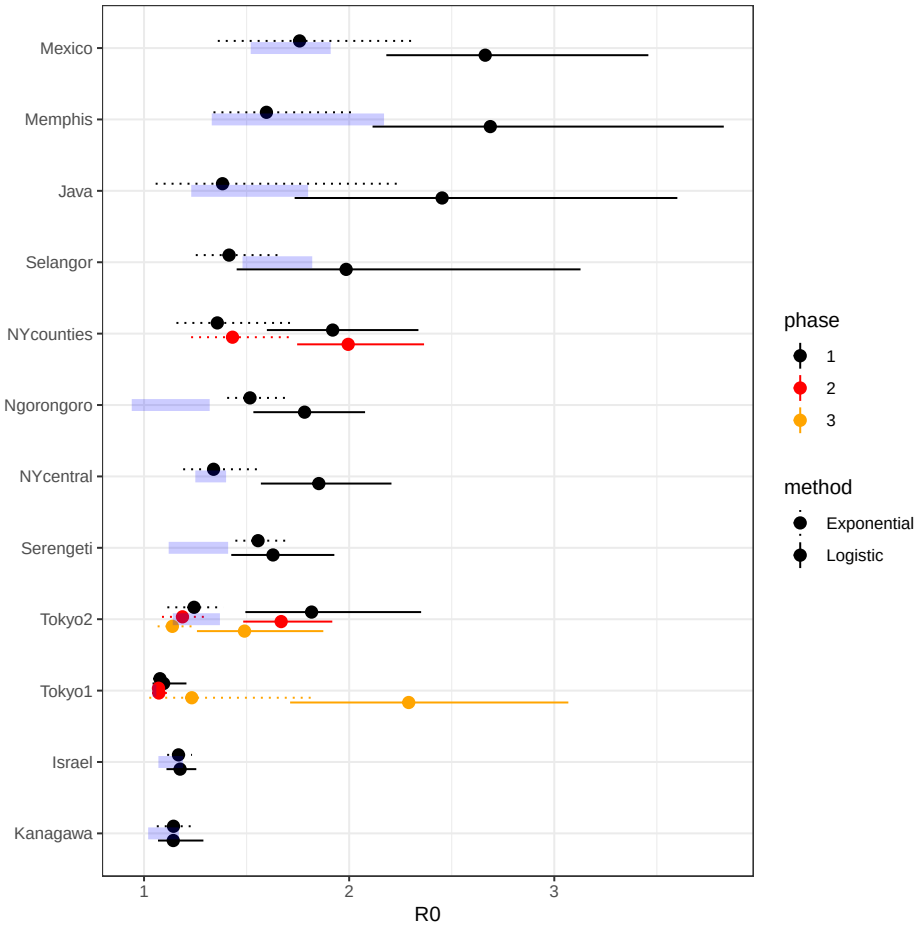

Figure 4: **Reproductive number estimates for global historical outbreaks of rabies**
Previous estimates of $\mathcal{R}_0$ are shown in blue highlights; $\mathcal{R}_0$ estimates and confidence interval (95% quantiles from the estimated $\mathcal{R}_0$ sample) from our hybrid approach using exponential (dotted lines) and logistic (solid lines). $\mathcal{R}_0$ are corrected for vaccination coverage.ML: todo: add vaccination coverage like KH's paper in the location name.

We combined our estimates of r from the logistic model with the empirical G from our detailed Tanzanian data to produce $\mathcal{R}_0$ estimates. Of the listed historical outbreaks, four occurred in locations with prior rabies vaccination coverage: Mem-

phis, US (1947, 10% vaccine coverage); Serengeti, Tanzania (2003, 20% coverage); Ngorongoro, Tanzania (2003, 20% coverage); and Sultan Hamad, Kenya (1992, 24% coverage). Our estimates of $\mathcal{R}_0$ ranged from 1.08 to 2.66 ML: check and use program, with upper confidence intervals greater than 2 for most locations. The hybrid approach provides larger values of $\mathcal{R}_0$ and wider confidence intervals than previous $\mathcal{R}_0$ estimates after propagating uncertainty from both r and generation interval distributions (see Figure **??**).

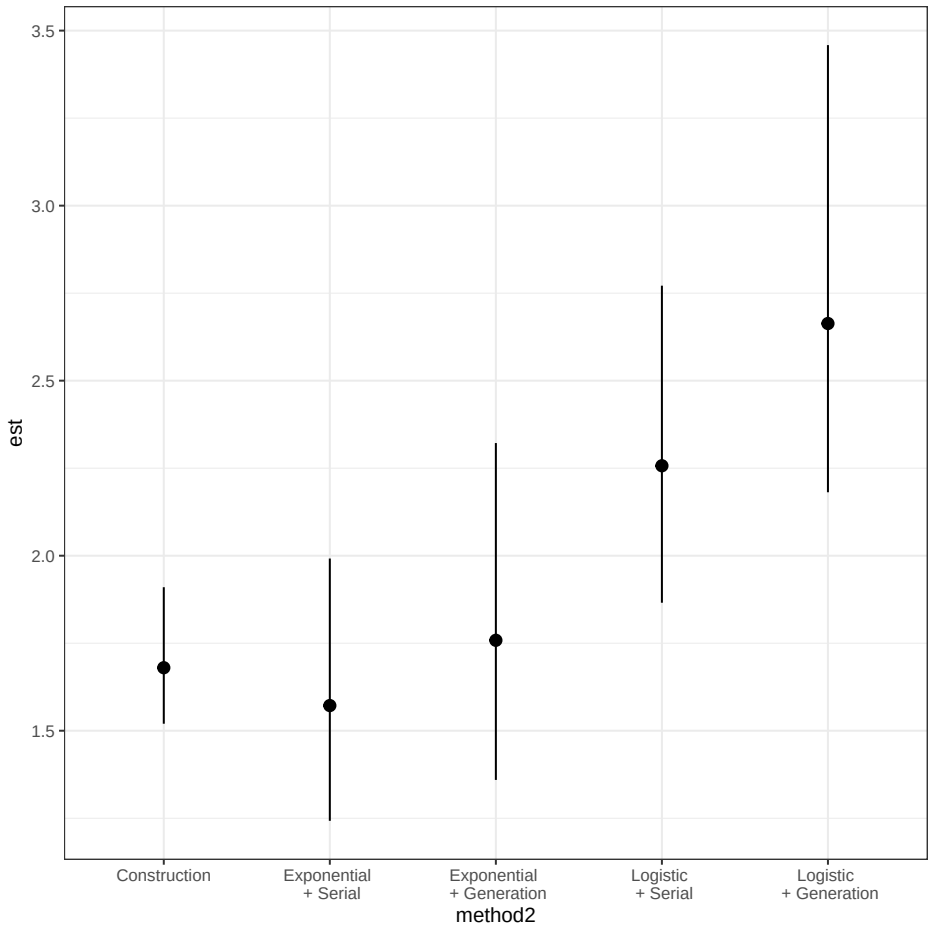

Figure 5: **Effects of r, corrected G on the estimates of $\mathcal{R}_0$ in Mexico outbreak.** [[BB: more complete caption? Make this horizontal to match the other figs? Extend R0-axis to have a lower limit at $\mathcal{R}_0$ =1?]]

Lastly, we compare the effects of different estimation techniques of r and G on estimates of $\mathcal{R}_0$ (Figure **??**). For illustrative purposes, we used the 1987 outbreak in Mexico where there was no vaccination; the supplementary material shows similar plots for the other historic outbreaks. Propagating uncertainty from both r and G generally leads to wider confidence intervals compared to previous $\mathcal{R}_0$ estimates in **?**. $\mathcal{R}_0$ estimate increases when we estimate r via the logistic model, or G instead of the constructed approach in **?** . The overall effect of $\mathcal{R}_0$ increases even more when included both corrections (i.e., in r and G) with even larger uncertainty.

Discussion

The basic reproductive number $\mathcal{R}_0$ is commonly used to summarize the risk of infectious disease and to inform control measures. Here, we used a relatively simple approach to estimate $\mathcal{R}_0$ by combining initial growth rate (r) estimates from incidence data and generation intervals from contact tracing data. We improved on earlier work by correcting for slowdown in growth in estimating r and by developing a hybrid approach to propagate uncertainty from both r and G, resulting in higher $\mathcal{R}_0$ estimates with wider confidence intervals.

Re-analysis of these data also allowed us to identify an overlooked fact about rabies generation intervals: observed generation intervals are longer on average than constructed ones [[BB: than intervals constructed by naively adding ...?]], because of within-individual correlations in time distributions and biting behaviour. The unexpected importance of these correlations could have implications for G-based studies of other infectious diseases. Further investigation of how these correlations affect the overall dynamics of rabies is warranted.

Estimates of $\mathcal{R}_0$ are strongly affected by estimates of the growth rate during the initial phase of the epidemic. The logistic model gives a better approximation of

the initial phase of the epidemic resulting in a larger estimate of r compared to the exponential model (**?**). Our estimates of r account for observation error (measurements may not perfectly match reality), but not for process error (the fundamental stochasticity of the system itself). Thus, there may be more uncertainty in r than we estimate (**?**), but this is not always true in practice (**?**). [[BB: don't really understand the second half of this sentence?]]

Nevertheless, our estimates suggest that rabies $\mathcal{R}_0$ may be larger, and more uncertain, than previously thought. This finding may explain some of the formerly unexplained variations in the success of rabies-control programs (e.g., low levels of coverage (30–50%) have been successful in some settings while high coverage 75% was not enough to control rabies in others (**?**)). While our primary goal was to understand why estimates of rabies $\mathcal{R}_0$ were small with narrow confidence intervals, our analysis also revealed an interesting biological process in rabies generation interval. [[BB: which is ... ?]] $\mathcal{R}_0$ is typically used as a first approximation for interventions such as vaccination to determine herd immunity thresholds. However, the main heterogeneity in contacts and correlations with incubation periods observed through the generation interval may suggest that simple $\mathcal{R}_0$ estimation methods may be inadequate and should be used with caution. [[BB: refine this sentence?]] A mechanistic fitting framework will be useful to study these patterns in more detail.

[[BB: really, really unimportant, but: check for appropriate capitalization of PLOS? (PLoS?)]]