medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Improving estimates of waning immunity rates in stochastic SIRS models with a hierarchical framework

Punya Alahakoon*, James M. McCaw[1,2,3], Peter G. Taylor[1]

[1]School of Mathematics and Statistics , The University of Melbourne, Melbourne, Australia.
[2]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia.
[3]Peter Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne, Australia.

Abstract

Most disease pathogens require onward transmission for their continued persistence. It is necessary to have continuous replenishment of the population of susceptibles, either through births, immigration, or waning immunity in recovered individuals.

Consider the introduction of an unknown infectious disease into a fully susceptible population where it is not known how long immunity is conferred once an individual recovers. If the disease takes off, the number of infectives will typically decrease to a low level after the first major outbreak. During this period, the disease dynamics will be highly influenced by stochastic effects and there is a non-zero probability that the epidemic will die out. This is known as an epidemic fade-out. If the disease does not die out, the susceptible population may be replenished by the waning of immunity, and a second wave may start.

In this study, we describe an experiment where we generated synthetic outbreak data from independent stochastic $SIRS$ models in multiple communities. Some of the outbreaks faded-out and some did not. By conducting Bayesian parameter estimation independently on each outbreak, as well as under a hierarchical framework, we investigated if the waning immunity rate could be correctly identified. When the outbreaks were considered independently, the waning immunity rate was incorrectly estimated when an epidemic fade-out was observed. However, the hierarchical approach improved the parameter estimates. This was particularly the case for those communities where the epidemic faded out.

1 Introduction

Infectious diseases do not always provide life-long or long-term protective immunity after infection (Heffernan & Keeling, 2009; Mathews, McCaw, McVernon, McBryde, & McCaw, 2007). Some common examples are pertussis (Mooi, Van Der Maas, & De Melker, 2014), flu (Camacho & Cazelles, 2013), and the A/H3N2 epidemic that occurred on the remote island of Tristan da Cunha in 1971 (Camacho et al., 2011). Furthermore, for emerging infectious diseases, sufficient biological evidence to hypothesise that either re-infection or life-long immunity is possible is often limited. One example is the early stages of the recent COVID-19 pandemic (Lavine, Bjornstad, & Antia, 2021; Telenti et al., 2021).

Mathematical epidemic models rely on compartmentalisation of the population into different states that are related to the infectious diseases of interest (Camacho & Cazelles, 2013; Heffernan & Keeling, 2009; Kermack & McKendrick, 1927). Deterministic epidemic models that allow for replenishment of susceptibles via re-infection, births or immigration display damped oscillatory behaviour (Keeling & Rohani, 2011). The simplest model for such a situation is the one that allows for re-infection, the $SIRS$ model. In this model, recovered individuals have immunity that wanes resulting in them becoming susceptible again.

In contrast to a deterministic model, the number of infectives in a stochastic $SIRS$ model can drop to zero. This occurs due to random effects at low disease prevalence levels. Once an outbreak avoids the initial fade-out (disease extinction during the start of the epidemic), there exists a trough following the first major outbreak Lloyd (2004). There is a non-zero probability that disease extinction will occur during this trough (Lloyd-Smith et al., 2005). This is known as an epidemic fade-out (Alahakoon, McCaw, & Taylor, 2022;

*palahakoonmu@student.unimelb.edu.au

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Ballard, Bean, & Ross, 2016; Bartlett, 1960; Camacho et al., 2011; Camacho & Cazelles, 2013; Lloyd-Smith et al., 2005; Meerson & Sasorov, 2009; van Herwaarden, 1997). If epidemic fade-out does not take place, the susceptible fraction will increase due to the waning of immunity, and once the effective reproduction number is greater than one, a second wave may be started. The likelihood of the occurrence of epidemic fade-out or non-fade-out depends on the model parameters of which the outbreak is modeled Anderson and May (1992).

Here, we consider a hypothetical pathogen where there is insufficient evidence to suggest that recovered individuals remain immune from infection forever. We consider outbreaks of this disease observed in small closed communities (sub-populations) during short periods of time where demographic factors can be ignored. We assume outbreaks in these sub-populations take place independently. Due to stochastic effects and variability in the characteristics of the sub-populations, we might observe epidemic fade-out in some sub-populations and not in others. This setup is an extension of our previous work Alahakoon et al. (2022).

Conditional on an epidemic taking off, we attempt to recover parameters of waning immunity rates when epidemic fade-outs are observed. We conduct a simulation-based experiment where multiple outbreaks of synthetic data are generated from a stochastic $SIRS$ model. Some outbreaks display fade-outs while others display non-fade-outs. We estimate model parameters by considering each outbreak independently as well as under a Bayesian hierarchical framework. We further consider two assumptions for the prior distributions under the Bayesian framework: one where we allow the waning immunity rate of an $SIRS$ model to be zero and one where we constrain the waning immunity rate to be non-zero. We illustrate that when the estimation is conducted independently, the waning immunity rate is incorrectly estimated under both assumptions when an epidemic fade-out is observed. We further show that the estimates of the waning immunity rate are improved when the estimation is carried out under a hierarchical framework irrespective of the assumption used. Additionally, we show that the waning immunity rate is still identifiable under a hierarchical framework even under incomplete data conditions.

2 Background

2.1 The Markovian $SIRS$ model in a closed sub-population

In a well-mixed sub-population of size N, we will denote the number of susceptibles, infectious individuals and recovered individuals by $S(t)$, $I(t)$, and $R(t)$ respectively at time t. An $SIRS$ model is parameterised by β, the transmission rate, γ, the rate of recovery, and μ, the waning immunity rate. The stochastic $SIRS$ system can be formulated as a continuous-time Markov chain with bi-variate states $(S(t), I(t))$. The model structure is illustrated in Figure 1 and the transition rates are presented in Table 1. When $\mu = 0$, the dynamics of this model are identical to those of the SIR model.

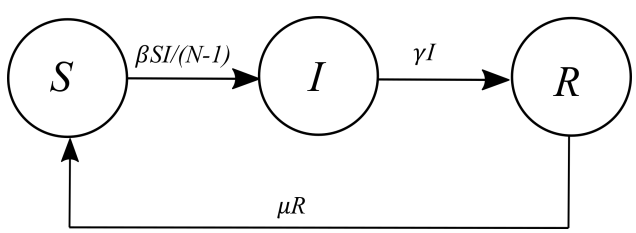

Figure 1: $SIRS$ model structure.

Table 1: Transition rates of an $SIRS$ model

Event	Transition	Rates
Infection	$(s, i) \rightarrow (s-1, i+1)$	$\beta si/(N-1)$
Recovery	$(s, i) \rightarrow (s, i-1)$	γi
Loss of immunity	$(s, i) \rightarrow (s+1, i)$	μr, ($r = N - s - i$)

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

2.2 A Bayesian hierarchical modelling framework

In this study, we consider the estimation of parameters under a hierarchical modeling approach. Let us assume that we are studying outbreaks corresponding to K sub-populations and each outbreak is modeled using a continuous-time Markov chain with transition rates defined in Table 1. Let the model parameter set corresponding each sub-population be denoted as a vector $\boldsymbol{\theta}_k$ $(k = 1, 2, \ldots, K)$. Under a hierarchical framework, we assume that model parameters corresponding to each outbreak are drawn from a common distribution (Gelman et al., 2013).

We construct our hierarchical framework with three levels similar to that of Alahakoon et al. (2022). Level I represents the observed prevalence data $\boldsymbol{y_k} = (I_k(1), I_k(2) \ldots, I_k(T_k))$ at T_k discrete time points for sub-populations $k = 1, 2, \ldots, K$. Level II represents the structural relationship between the sub-population specific parameters $\boldsymbol{\theta}_k$ $(k = 1, 2, \ldots, K)$ and the hyper-parameters, Ψ. The θ_k are independent random variables with common distribution $p(\theta|\Psi)$. A common distribution for the conditional prior distribution is is that the parameters are independently drawn from normal distributions with a given means and standard deviations. Finally, Level III represents the prior distributions for the hyper-parameters, which are generally known as hyper-prior distributions, $p(\boldsymbol{\Psi})$ (Gelman et al., 2013).

The joint posterior distribution for a population consisting of K sub-populations is,

$$p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K, \boldsymbol{\Psi}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_K, \boldsymbol{\Psi})p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_K, \boldsymbol{\Psi})}{p(\boldsymbol{y})} = \frac{\left[\prod_{k=1}^{K} p(\boldsymbol{y}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\boldsymbol{\Psi})\right] p(\boldsymbol{\Psi})}{p(\boldsymbol{y})}$$
$$\propto \left[\prod_{k=1}^{K} p(\boldsymbol{y}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\boldsymbol{\Psi})\right] p(\boldsymbol{\Psi}). \quad (1)$$

Within our $SIRS$ modelling framework for sub-population k, $\boldsymbol{\theta}_k = (\beta_k, \gamma_k, \mu_k)$. We define the conditional prior distribution for the model parameters as a multivariate normal distribution with means Ψ_β, Ψ_γ, and Ψ_μ, and standard deviations $\sigma_\beta, \sigma_\gamma$, and σ_μ, and correlations set to zero.

3 Materials and Methods

3.1 Synthetic data generation

We constructed synthetic data for 15 sub-populations each consisting of 1000 individuals. We fixed the initial conditions of each outbreak to include one infectious person in each sub-population at the start of the outbreak.

We generated synthetic data for all the sub-populations with the stochastic $SIRS$ model structure that was introduced in Section 2. We independently generated transmission, recovery, and waning immunity rates from three truncated normal distributions. We randomly generated the transmission rates, β_k, for the sub-populations from a normal distribution with a mean of 2.5 and standard deviation of 0.25, truncated on the interval $(1, 10)$. We generated the recovery rates, γ_k, from a normal distribution with a mean of 1 and a standard deviation of 0.05, truncated on the interval $(0, 4)$. We generated the waning immunity rates, μ_k, from a normal distribution with a mean of 0.06 and standard deviation of 0.01, truncated on the interval $(0.01, 1)$. See Table 2 for summary statistics of the actual values of the parameters that were generated for each of the sub-populations. Using these parameters for the $SIRS$ model, we generated sample paths from the Doob-Gillespie (Doob, 1945; Gillespie, 1977) algorithm for 35 days and retained the prevalence of infections each day. When generating sample paths, if the sample path produced an initial fade-out, we discarded that sample-path and repeatedly generated sample paths until an initial outbreak was observed. The criteria we used to identify an outbreak were similar to Alahakoon et al. (2022). Figure 2 shows the time-series data of the 15 sub-populations. Sub-populations 4, 8, 10, 12, 14, and 15 displayed an epidemic fade-out and other sub-populations displayed multiple waves.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2: Summary statistics of the parameters of the 15 outbreaks

Parameter	Mean	Standard deviation
β	2.5235	0.3476
γ	1.0198	0.0498
μ	0.0606	0.0073

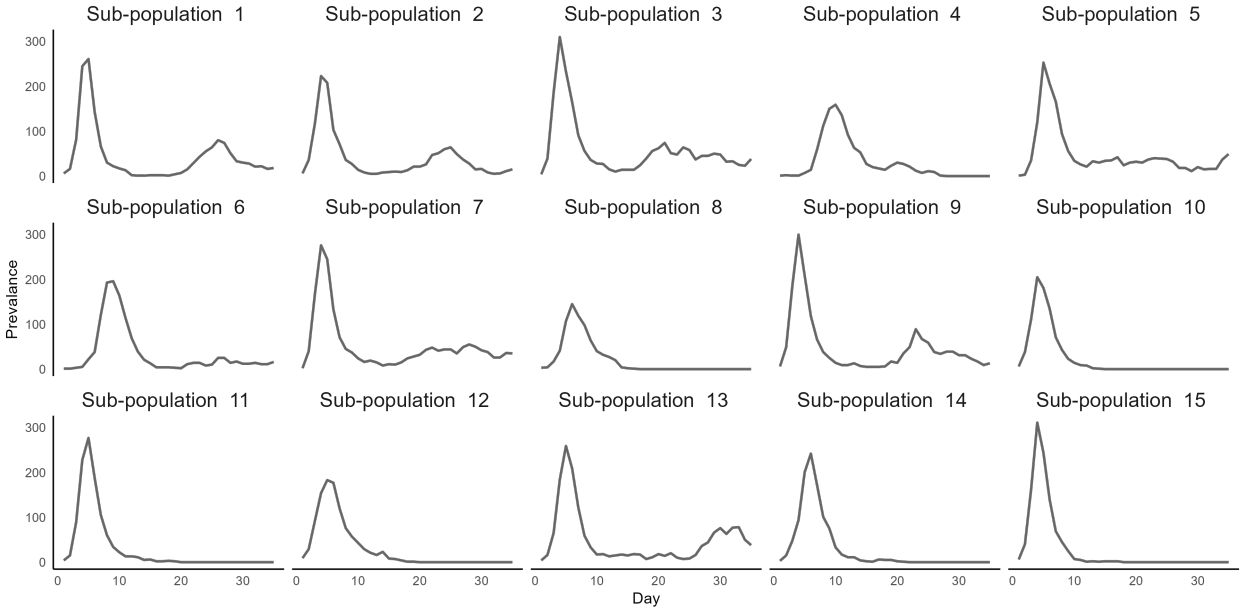

Figure 2: Synthetic data for fifteen sub-populations over 35 days.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

3.2 Estimation framework

We considered two frameworks to estimate the model parameters related to the transmission, recovery, and waning immunity rates: 1) estimation by considering each outbreak independently. 2) estimation by considering a hierarchical framework. Under each estimation framework, we used two different assumptions for prior distributions within the Bayesian framework.

Under the first assumption, we allow the waning immunity rate of an $SIRS$ model to be zero by using a less informative prior distribution for the waning immunity rate when the outbreaks are considered independently as well as at the hyper-parametric level. Under the second assumption, we used a strong informative prior at both these settings where waning immunity was constrained to be greater than 0.01. Tables 3 and 4 illustrate our choice of priors for the model parameters under both assumptions.

Table 3: Prior distributions for parameters when the outbreaks are considered independently

Parameter	**Assumption 1**	**Assumption 2**
β	Uniform (0.001,10)	Uniform (0.001,10)
γ	Uniform (0.00001,3)	Uniform(0.00001,3)
μ	Uniform (0,0.2)	TN (0.03, 0.1, 0.01, 0.2)

Table 4: Prior distributions for hyper-parameters under the hierarchical framework

Hyper parameter	**Assumption 1**	**Assumption 2**
Ψ_β	Uniform (0.001, 10)	Uniform (0.001, 10)
σ_β	Uniform (0, 2.5)	Uniform (0, 2.5)
Ψ_γ	Uniform (0.00001, 3)	Uniform (0.00001, 3)
σ_γ	Uniform (0, 1)	Uniform (0, 1)
Ψ_μ	Uniform (0, 0.2)	Uniform (0.01, 0.2)
σ_μ	Uniform (0, 0.15)	Uniform (0, 0.15)

We conducted parameter estimation under both assumptions when outbreaks were considered independently with the ABC-SMC algorithm of Toni, Welch, Strelkowa, Ipsen, and Stumpf (2009). See Supplementary Material S1.1 for the details of our calibration of the ABC-SMC algorithm. We also used the two-step algorithm of Alahakoon et al. (2022) to estimate the parameters under a stochastic hierarchical framework. See Supplementary Material S2.1 and S2.2 in relation to the calibration and diagnostics of this step. When estimating the hyper-parameters of the conditional prior distribution, the estimated correlations of the multivariate distribution were not substantial. Therefore, we used independent conditional prior distributions. See Supplementary Material S2 for further explanation.

4 Results

Figure 3 shows the marginal posterior distributions for the sub-population-specific waning immunity rates when parameter estimation was carried out independently for the outbreaks. Despite the type of prior used, the posterior distributions of sub-populations that did not fade out had similar shapes. For the sub-populations that did fade out, the posterior distributions were positively skewed and were truncated at the lower bound (closer to zero) of the prior distribution. As expected, this observation indicated that when an epidemic fade-out was observed, the estimated posterior mode of waning immunity rate was closer to zero indicating that an SIR model is suitable even though data was generated from an $SIRS$ model with a waning immunity rate above zero. See Supplementary Material S1.2 for a comparison of posterior median and Highest Posterior Density (HPD) intervals (Chen, Shao, & Ibrahim, 2012) computed from *HDInterval* package in *R* (Meredith & Kruschke, 2020) and for visual diagnostics of other parameters.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

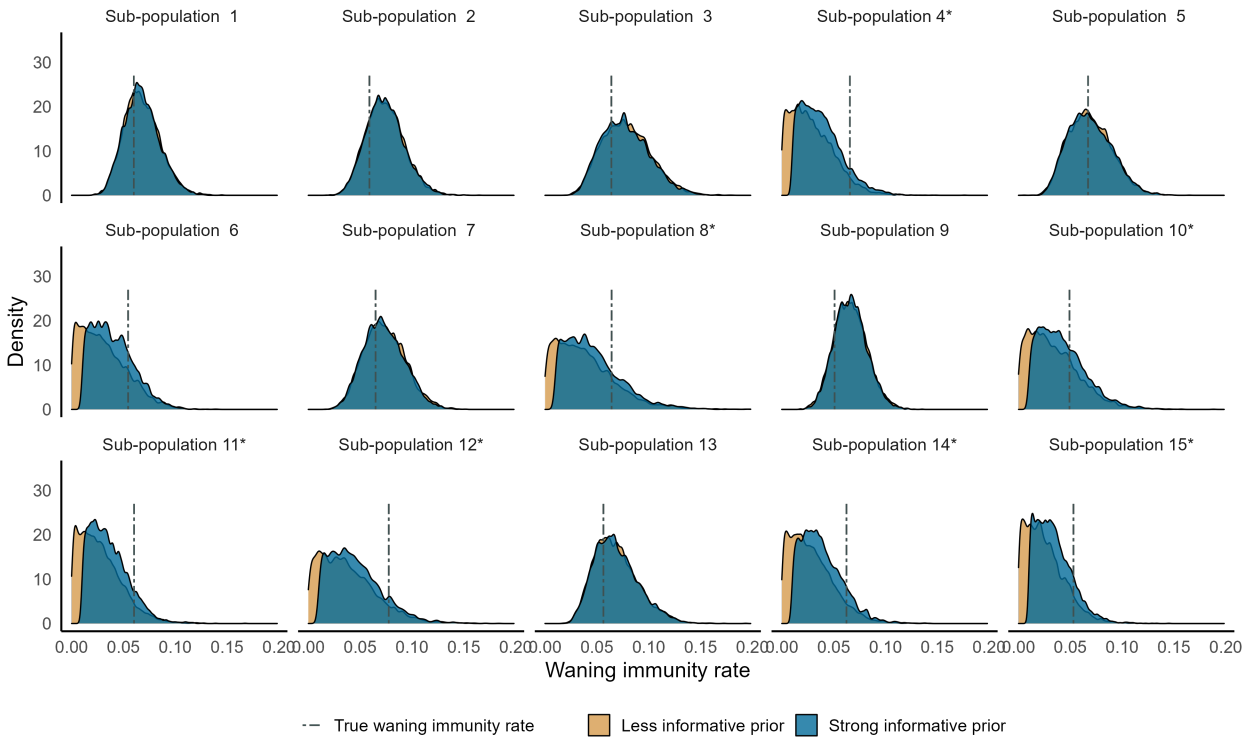

Figure 3: Independent estimation: Posterior distributions for waning immunity rate for the sub-populations under two assumptions. Asterisks represent sub-populations that experienced epidemic fade-out.

As we did not observe a substantial difference between the marginal posterior distributions under the different assumptions, we chose to analyse the posterior distributions under the second assumption. Hence, we compared these estimates with independent and hierarchical frameworks. Figure 4 illustrates the marginal posterior distributions of the waning immunity, μ (plot (A)), and transmission, β (plot (B)), rates. Particularly for the waning immunity rates of the sub-populations that displayed epidemic fade-outs, there is a striking difference between the posterior distributions obtained from the two frameworks. The shapes of these distributions changed from highly positively skewed (independent analysis) to slightly negatively skewed (hierarchical analysis) distributions that converged around the parameter values.

The right panel of Figure 4 illustrates the extent of improvement of parameter estimates under a hierarchical analysis in comparison to an independent analysis. For this we used the Region of Practical Equivalence (ROPE) criterion (J. Kruschke, 2014; J. K. Kruschke, 2013, 2018) and the posterior modes of both μ and β. We used the ROPE criterion (see Supplementary Material S2.6) to identify the percentage of the 95% Highest Posterior Density (HPD) intervals of μ and β that were included inside the ROPE. For the waning immunity rates of the sub-populations that displayed epidemic fade-outs (plot (C)), the median increase is 38%. The corresponding increase for those that did not fade out (plot (D)) is 20.5%. The median of the posterior modes for μ increased by 0.034 (plot (E)) compared to the independent analysis for sub-populations that observed fade-outs. However, for those that did not observe fade-outs, the variability of posterior modes diminished (plot (F)). Overall, under a hierarchical analysis, estimates of the μ improved; and the improvement is higher for sub-populations that displayed fade-outs.

For the ROPE percentages of β, the results are similar to those of μ. The median increase is 22% (plot (G)) and 15% (plot (H)) for sub-populations with fade-outs and non fade-outs respectively. The variability of the posterior modes of sub-populations that displayed both fade-outs and non-fade-outs diminished (plots (I) and (J)) and we did not observe a shift between the medians of the posterior modes. For the recovery rates, γ, we made similar observations as for the transmission rates. See Supplementary Material S2.6.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

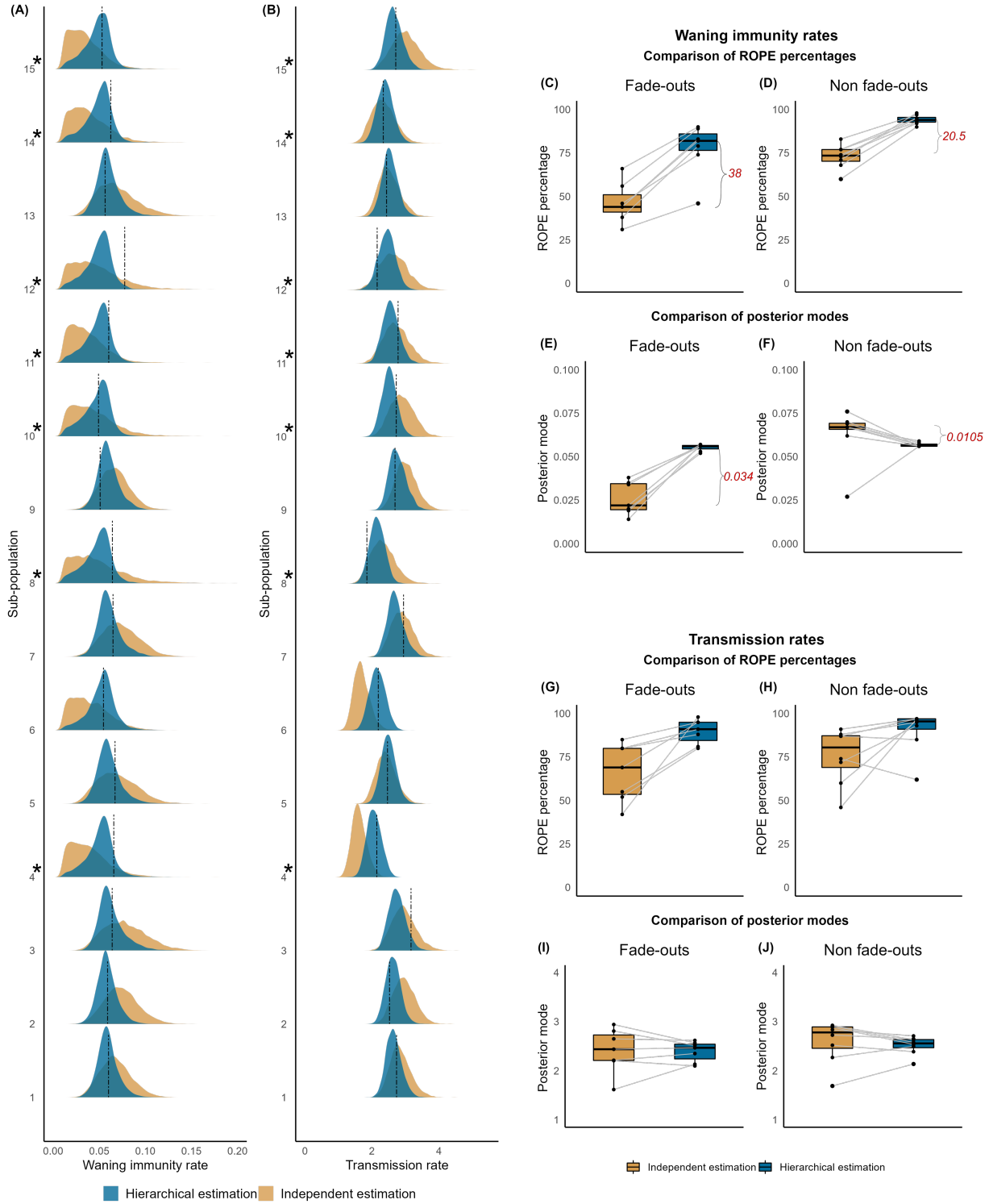

Figure 4: Hierarchical vs. independent estimation under Assumption 2:
Left panel : Marginal posterior distributions for waning immunity (**plot (A)**) and transmission rates (**plot (B)**). Asterisks represent sub-populations that experienced epidemic fade-out.
Top right panel : Paired comparison of ROPE percentages (**plot (C)**) and posterior modes (**plot (D)**) with independent and hierarchical estimation frameworks for the waning immunity rate when fade-outs and non fade-outs are observed.
Bottom right panel : Paired comparison of ROPE percentages (**plot (E)**) and posterior modes (**plot (F)**) with independent and hierarchical estimation frameworks for transmission rate when fade-outs and non fade-outs are observed.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Overall, in comparison to an independent analysis, a hierarchical framework improved the estimation under both assumptions. We did not observe a notable distinction between the corresponding posterior distributions under the two assumptions. This can be further clarified by a comparison of posterior medians and HPD intervals. See Supplementary Materialdix S2.5 for these details and for a visual comparison of posterior distributions of all the parameters under both assumptions.

Figure 5 shows the posterior distributions of the hyper parameters when a hierarchical analysis was used for the sub-populations under the second assumption (see Supplementary Material S2.7 for a comparison with both assumptions). Table 5 shows the posterior medians and HPD intervals for the hyper-parameters under both assumptions. The posterior distributions of the hyper-parameters for transmission rate and recovery rate had no visual distinction when different prior distributions for the waning immunity rate were used. Under both assumptions, the posterior medians for Ψ_β, σ_β Ψ_γ, and σ_γ were closer to the parameter values. Under a strong informative prior, the estimate of Ψ_μ improved slightly (posterior median 0.0524) in comparison to that under a less informative prior (posterior median 0.0493).

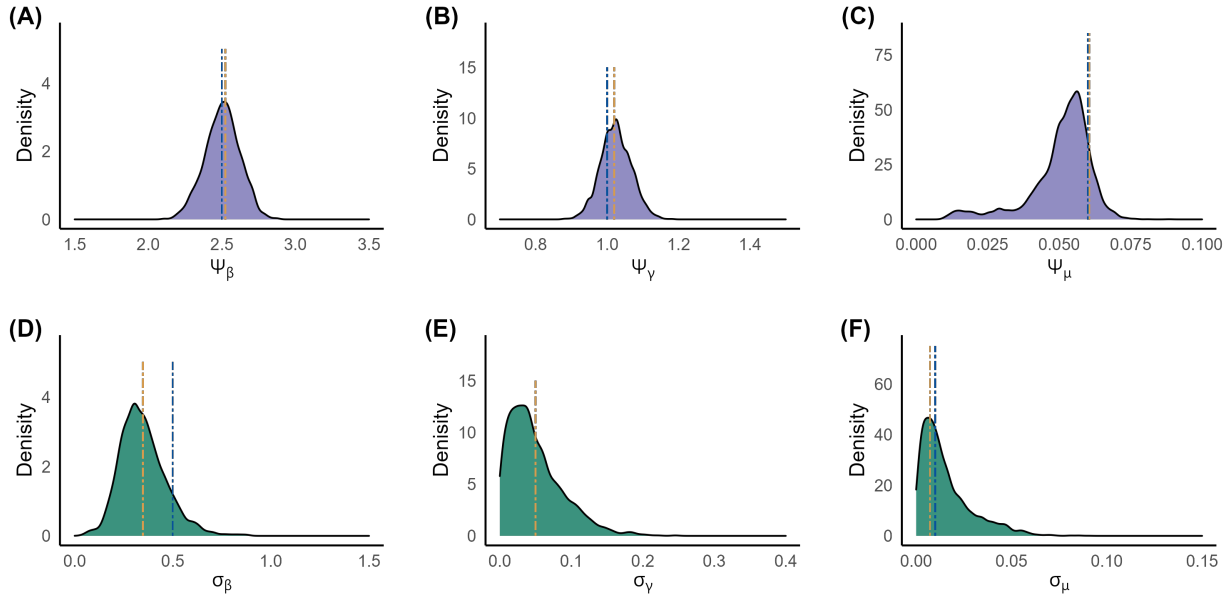

Figure 5: Marginal posterior distributions for the hyper-parameters for β (**A**), γ (**B**) and μ (**C**) under a strong informative prior. **Blue dashed line** : Parameter value. **Orange dashed line :** Mean of the sub-population specific parameters.

Table 5: Summary statistics for the posterior distributions of the hyper-parameters

Hyper parameter	**Assumption 1 (Less informative prior)**		**Assumption 2 (Strong informative prior)**	
	Posterior median	**HPD interval**	**Posterior median**	**HPD interval**
Ψ_β	2.5144	(2.2409, 2.7656)	2.5573	(2.2777, 2.7360)
σ_β	0.3956	(0.1333, 0.5983)	0.3888	(0.1483, 0.6043)
Ψ_γ	1.0198	(0.9280, 1.0957)	1.0302	(0.9412, 1.1088)
σ_γ	0.0539	(0.0003, 0.1245)	0.0506	(0.0007, 0.1240)
Ψ_μ	0.0493	(0.0085, 0.0662)	0.0524	(0.0249, 0.07020)
σ_μ	0.0143	(0.0002, 0.0509)	0.0095	(0.0000, 0.0442)

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

4.1 Estimates under incomplete time-series data

We draw the reader's attention to the disease dynamics of sub-population 6. This outbreak had not experienced an epidemic fade-out in 35 days (as the prevalence had not reached zero), nor had it displayed a distinct second wave. This was reflected in the posterior distribution of the waning immunity rate in that the shape of the distribution was positively skewed and $\mu = 0$ was allowed. That is, SIR-like dynamics were not excluded. This observation motivated us to study whether a hierarchical framework is still able to identify the presence of waning immunity within sub-populations when only a part of the time-series is observed and if an independent estimation can give insights as to whether an outbreak will eventually fade-out or not.

Accordingly, we generated two additional parameter sets from the probability distributions that we used to generate our synthetic data for 15 sub-populations. We assumed that these two parameter sets generated outbreaks in two new sub-populations (sub-populations 16 and 17) where only a part of the time-series was observed. The transmission rates for the two sub-populations were 2.6355 and 2.0364, the recovery rates were 1.0169 and 0.9352, and the waning immunity rates were 0.0642 and 0.0443 respectively. We generated sample paths up to 35 days. We ensured that one of the sub-populations displayed a second wave and the other, an epidemic fade-out. We then assumed that the observed time-series data in both sub-populations were up to 15 days only. Plot (A) of Figure 6 display the observed time-series in black and the complete data in two-dashed lines.

As we estimated the waning immunity rates of the 15 sub-populations to be above zero, we now have evidence to suggest that waning of immunity for the infectious disease is possible. Therefore, we conducted parameter estimation for the two new sub-populations under Assumption 2, that is, under a strong informative prior. We estimated the parameters by considering the outbreaks independently as well as under a hierarchical framework. We carried out the latter framework by considering data from all the 17 outbreaks; that is, 15 sub-populations with 35-day time-series data and 2 sub-populations with 15-day time-series data.

Plots (B) and (C) of Figure 6 illustrate the posterior distributions under the two estimation frameworks for the two new outbreaks for μ and β respectively. The parameter estimates of both sub-populations under the hierarchical analysis improved with a notable improvement for the waning immunity rate of sub-population 17. The posterior distribution of μ of sub-population 17 changed from highly positively skewed to slightly negatively skewed and converged around the parameter value when the estimation changed from an independent to a hierarchical framework. This analysis is consistent with our study of 15 sub-populations where epidemic fade-outs were observed.

Using the methods by Ballard et al. (2016), we calculated the epidemic fade-out probabilities, 0.3003 and 0.7199, given the true parameters for sub-populations 16 and 17 respectively. Plot (D) of Figure 6 illustrate the estimated distributions of probability of epidemic fade-out for the two sub-populations when partial (up-to 15 days) and complete (up-to 35 days) time-series data are considered (see Supplementary Material S2.8 for details). When we carried out independent estimation, the distributions (with incomplete and complete data) of the epidemic fade-out probability slightly converged towards the true value. However, under a hierarchical framework, these distributions (with incomplete and complete data) converged towards 0.5; the probability of epidemic fade-out at the population level for all the 17 sub-populations.

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

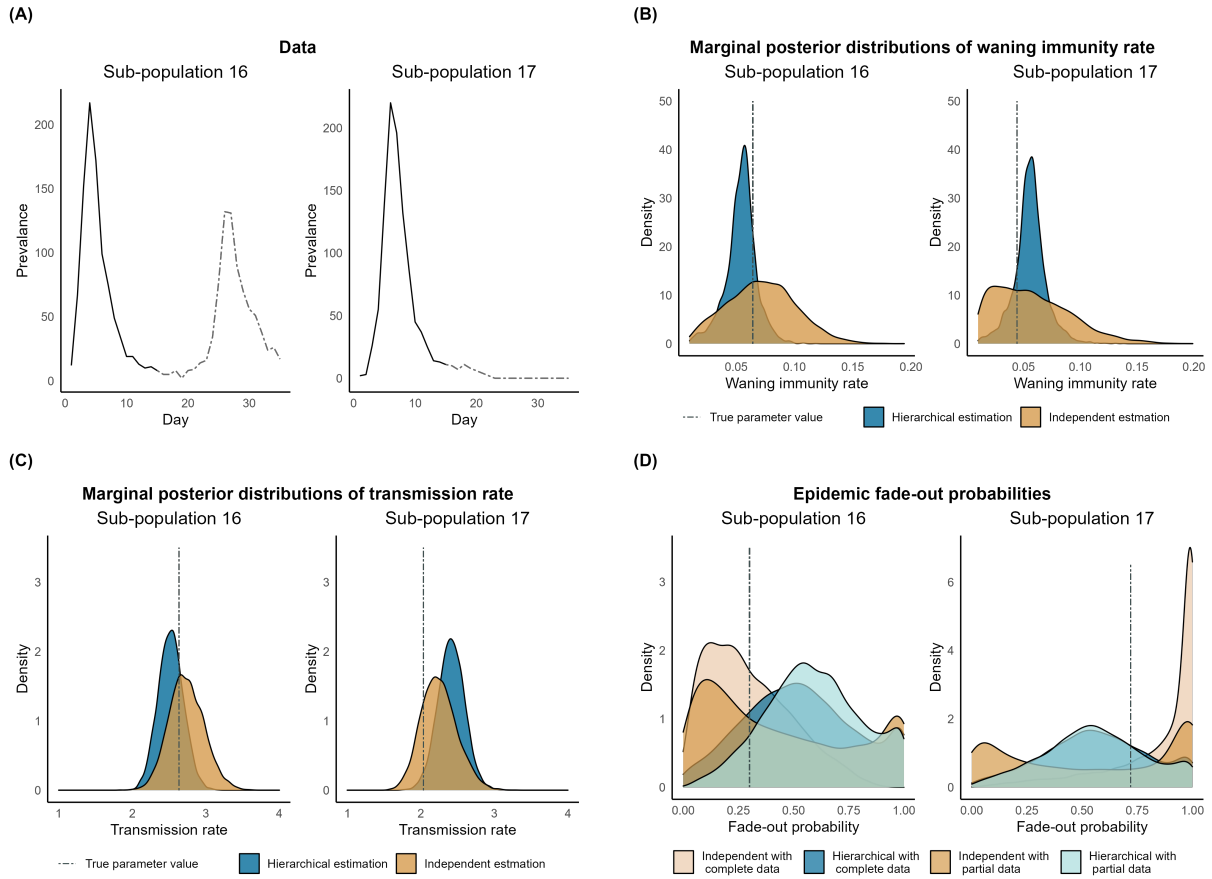

Figure 6: **Plot (A):** Data observed (in black line) and expected (in grey dashed) for sub-populations 16 and 17.
Plot (B): Posterior distributions for μ for the sub-populations 16 and 17 under parameter estimation independently and under a hierarchical framework.
Plot (C): Posterior distributions for β for the sub-populations 16 and 17 under parameter estimation independently and under a hierarchical framework.
Plot (D) : Probability of epidemic fade-out distributions from the estimated parameters. The horizontal line represent the probability fade-out calculated with the true parameter values using the methods by Ballard et al. (2016).

5 Discussion

We have studied a hypothetical infectious disease where waning immunity exists but biological evidence is limited. We have shown that when multiple outbreaks take place in multiple communities, parameter estimates generally improve when estimation is carried out within a hierarchical framework in comparison to when the outbreaks are studied independently. Application of the parameter estimation framework introduced by Alahakoon et al. (2022) yields improved estimates.

Epidemic fade-out is a combined result of characteristics of the sub-populations and stochastic effects at low prevalence levels. Therefore, it is possible to observe epidemic fade-out or multiple waves in different sub-populations. We have shown that when an epidemic fade-out is observed in a sub-population where multiple waves are possible, the parameter(s) that indicate the possibility of re-infection/ multiple waves are incorrectly estimated when the outbreak is studied independently.

On the other hand, when a hierarchical framework is used to study multiple outbreaks, the population distribution (the conditional prior) creates a dependence structure across the model parameters of the sub-populations. This dependence structure enables information sharing among all the sub-populations and

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

aids in improving the estimates for re-infection (and other model parameters) even when an epidemic fade-out is observed. We have shown that the waning immunity rate parameter can be correctly estimated without a significant impact on the choice of the prior distribution within a hierarchical framework.

Furthermore, we have shown that even under incomplete data conditions of some sub-populations, the waning immunity rate can be estimated when a hierarchical framework is used. Interestingly though, in such data conditions, studying the outbreaks independently may still give important insights as to whether the outbreak will eventually fade-out or not. In such situations, we suggest the use and comparison with both estimation frameworks.

Another possible application occurs within surveillance frameworks where data accumulation is expected. As an example, when the first few outbreaks of an emerging infectious disease are observed with epidemic fade-outs and then it is possible to have observed non-fade-outs, a hierarchical framework can aid in identifying the presence of waning immunity among the communities that observed fade-outs. Similarly, this framework can also be applied to other types of data, for example malaria parasite dynamics in individuals (Cao et al., 2019).

This study and the study of Alahakoon et al. (2022) have considered using hierarchical frameworks only within $SIRS$ model structures. In this study, we have extended the framework of Alahakoon et al. (2022) to study multiple parameters. It will be helpful to study the effects of epidemic fade-out within hierarchical frameworks with other complex model structures and actual data. This estimation framework is also applicable to other similar compartmental-type stochastic epidemic models. Our future work includes the application of this estimation framework to real data.

6 Acknowledgements

Unless otherwise mentioned, computations were done in MATLAB or R across 32 clusters. All the computations were carried out by the use of the Nectar Research Cloud (project Infectious Diseases), a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy (NCRIS). All the plots were generated with ggplot2 (Wickham, 2016) in R. The codes are publicly available (see Supplementary Material for details).

7 Funding

Punya Alahakoon is supported by a Melbourne Research Scholarship from the University of Melbourne. P.G. Taylor would like to acknowledge the support of the Australian Research Council via the Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS).

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

References

Alahakoon, P., McCaw, J. M., & Taylor, P. G. (2022). Estimation of the probability of epidemic fade-out from multiple outbreak data. *Epidemics*, 100539.

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: dynamics and control*. Oxford university press.

Ballard, P., Bean, N., & Ross, J. (2016). The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography. *Journal of Theoretical Biology*, *393*, 170–178.

Bartlett, M. S. (1960). The critical community size for measles in the United States. *Journal of the Royal Statistical Society: Series A (General)*, *123*(1), 37–44.

Camacho, A., Ballesteros, S., Graham, A. L., Carrat, F., Ratmann, O., & Cazelles, B. (2011). Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1725), 3635–3643.

Camacho, A., & Cazelles, B. (2013). Does homologous reinfection drive multiple-wave influenza outbreaks? Accounting for immunodynamics in epidemiological models. *Epidemics*, *5*(4), 187–196.

Cao, P., Collins, K. A., Zaloumis, S., Wattanakul, T., Tarning, J., Simpson, J. A., ... McCaw, J. M. (2019). Modeling the dynamics of plasmodium falciparum gametocytes in humans during malaria infection. *Elife*, *8*, e49058.

Chen, M.-H., Shao, Q.-M., & Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.

Doob, J. L. (1945). Markoff chains–denumerable case. *Transactions of the American Mathematical Society*, *58*(3), 455–473.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, *81*(25), 2340–2361.

Heffernan, J., & Keeling, M. (2009). Implications of vaccination and waning immunity. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1664), 2071–2080.

Keeling, M., & Rohani, P. (2011). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press. Retrieved from `https://books.google.com.au/books?id=LxzILSuKDhUC`

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, *115*(772), 700–721.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with r, jags, and stan*. Academic Press.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in methods and practices in psychological science*, *1*(2), 270–280.

Lavine, J. S., Bjornstad, O. N., & Antia, R. (2021). Immunological characteristics govern the transition of COVID-19 to endemicity. *Science*, *371*(6530), 741–745.

Lloyd, A. L. (2004). Estimating variability in models for recurrent epidemics: assessing the use of moment closure techniques. *Theoretical Population Biology*, *65*(1), 49–65.

Lloyd-Smith, J. O., Cross, P. C., Briggs, C. J., Daugherty, M., Getz, W. M., Latto, J., ... Swei, A. (2005). Should we expect population thresholds for wildlife disease? *Trends in Ecology & Evolution*, *20*(9), 511–519.

Mathews, J. D., McCaw, C. T., McVernon, J., McBryde, E. S., & McCaw, J. M. (2007). A biological model for influenza transmission: pandemic planning implications of asymptomatic infection and immunity. *PLoS One*, *2*(11), e1220.

Meerson, B., & Sasorov, P. V. (2009). WKB theory of epidemic fade-out in stochastic populations. *Physical Review E*, *80*(4), 041130.

Meredith, M., & Kruschke, J. (2020). HDInterval: Highest (Posterior) Density Intervals [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=HDInterval` (R package version 0.2.2)

Mooi, F. R., Van Der Maas, N. A., & De Melker, H. E. (2014). Pertussis resurgence: waning immunity and pathogen adaptation–two sides of the same coin. *Epidemiology & Infection*, *142*(4), 685–694.

Telenti, A., Arvin, A., Corey, L., Corti, D., Diamond, M. S., García-Sastre, A., ... Virgin, H. W. (2021). After the pandemic: perspectives on the future trajectory of COVID-19. *Nature*, *596*(7873), 495–504.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society*

medRxiv preprint doi: https://doi.org/10.1101/2022.09.14.22279950; this version posted September 17, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Interface, *6*(31), 187–202.

van Herwaarden, O. A. (1997). Stochastic epidemics: the probability of extinction of an infectious disease at the end of a major outbreak. *Journal of Mathematical Biology*, *35*(7), 793–813.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`