THE PERTUSSIS EXAMPLE

AARON A. KING AND PEJ ROHANI

This document details the family of systems intended to model pertussis and in particular to focus on issues of reinfection and loss of immune protection to reinfection. We will describe the models, specify how we have used them to generate simulated epidemiological data, lay out the "bounds of our simulated ignorance" for the fitting exercises, and do a bit of preliminary investigation of the likelihood surface. In the simulation studies, we will focus on model selection and hypothesis testing. This is simplified somewhat by the fact that the candidate models are nested; the supermodel in this case is the SEIRR model.

Contents

1. The SEIRR equations

The SEIRR system is a simpler alternative to the multiple-compartment model of Wearing and Rohani (2009). The following key assumptions are made:

(1) Following recovery from infection, with probability $(1-\phi)$, individuals enter the R_1 class, where they are considered fully immune. With probability ϕ, recovering individuals reactivate full susceptibility and enter the S class. This captures so-called "polarising" immunity.
(2) Individuals in the R_1 class may lose immunity and become susceptible to immune boosting upon entry to the R_2 class, with rate $\alpha_1 R_1$.
(3) Within the R_2 class, exposure to infection is described by $\epsilon \lambda$, where ϵ modulates the force of infection in this class. Crucially, those exposed — with probability ξ — experience sub-clinical infections that boost immunity, moving affected individuals to the immune R_1 class ($R_2 \to R_1$).
(4) A fraction $(1-\xi)$ of exposed individuals in the R_2 class are assumed to develop full infection, i.e., become fully infectious ($R_2 \to E$).
(5) In the absence of immune boosting, individuals reactivate full susceptibilibty ($R_2 \to S$) after a mean period of $1/\alpha_2$ in the R_2 class.
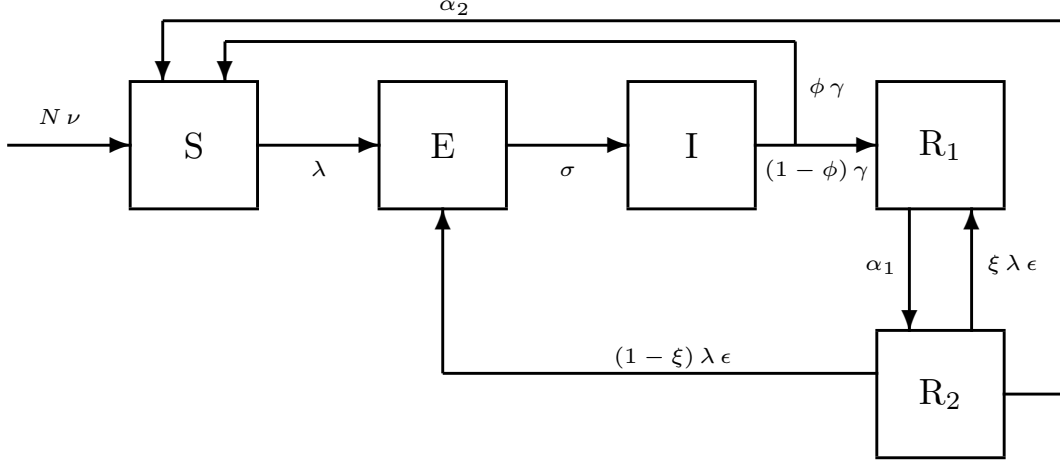
FIGURE 1. Diagram of the full SEIRR model, a simplified version of that studied by Wearing and Rohani (2009). When $\alpha_1 = \epsilon = \phi = 0$, this reduces to the SEIR model. When $\phi = \epsilon = 0$ and $\alpha_1 \gg \alpha_2$, this reduces to the SEIRS model. When $\phi = \epsilon = 0$ and $\alpha_1 = \alpha_2$, we obtain an SEIRS system where the waning of immunity is Gamma-distributed with a shape parameter of 2.

The mean field equations describing this the SEIRR system are given by:

$$\dot{S} = \nu\, N - \lambda(t)\, S - \mu\, S + \phi\, \gamma\, I + \alpha_2\, R_2 \tag{1}$$

$$\dot{E} = \lambda(t)\, S + (1 - \xi)\, \epsilon\, \lambda(t)\, R_2 - (\mu + \sigma)\, E \tag{2}$$

$$\dot{I} = \sigma\, E - (\mu + \gamma)\, I \tag{3}$$

$$\dot{R}_1 = (1 - \phi)\, \gamma\, I - (\mu + \alpha_1)\, R_1 + \xi\, \epsilon\, \lambda(t)\, R_2 \tag{4}$$

$$\dot{R}_2 = \alpha_1\, R_1 - (\mu + \alpha_2 + \epsilon\, \lambda(t))\ R_2 \tag{5}$$

In these equations, the force of infection, λ, is time dependent and given by

$$\lambda(t) = \frac{\beta(t)\, \zeta(t)\, I + \iota}{N}, \tag{6}$$

where ι represents the force of infection due to "immigration" or other effects extrinsic to the population, $\beta(t)$ is the time-dependent contact rate, which is assumed to follow the opening and closing of schools, and $\zeta(t)$ is white noise, which models environmental stochasticity. In particular, we assume that

$$\beta(t) = \begin{cases} \bar{\beta}\, (1 + b_1), & \text{when school is in,} \\ \bar{\beta}\, (1 - b_1), & \text{when school is out.} \end{cases} \tag{7}$$

The other parameters are constant; they are summarized in Table 1.

2. MEASUREMENT MODEL

It is assumed that individuals that are observed are thereby removed from the infectious class. In particular, the number of observed cases in a week is a noisy observation of the total number of E $\longrightarrow$ I transitions that occur in that week. Thus, if $N_{EI}(t)$ is the cumulative number of E $\to$ I transitions and $\Delta N_{EI}(t) = N_{EI}(t) - N_{EI}(t - 1/52)$ is the number of E $\to$ I transitions over the course of the preceding week, we assume that the reported cases, C_t follow a Gamma-Poisson or negative binomial distribution

$$C_t \sim \text{negbin}(\rho\, \Delta N_{EI}(t), \tau) \tag{8}$$

where $\text{negbin}(\mu, \tau)$ is the negative binomial distribution with mean μ and variance. $\mu + \tau\, \mu^2$.

3. Model selection and limiting cases

In the pertussis simulation study, we simulate the solution of a common problem: how can we assess the evidence for or against an important role for a particular mechanism? In the present instance, we are concerned with two mechanisms that have been hypothesized to play an important role in pertussis transmission: waning of immunity and boosting of immunity by natural exposure. We will assess the strength of evidence for each of these mechanisms using a model selection approach. Nested within the full SEIRR model are several submodels:

(1) When the intensity of the white noise ζ is zero, we lose the environmental stochasticity, i.e., the stochasticity is purely demographic.
(2) When $\alpha_1 = \phi = 0$, there is no loss of immunity and we have the SEIR system.
(3) When $\varepsilon = \phi = 0$, and both $\alpha_1, \alpha_2 > 0$, we have loss of immunity but no boosting: the system can be said to be of SEIRS form. If in addition $\alpha_1 = \alpha_2$, the waiting time to loss of immunity is Gamma distributed with shape parameter 2.

4. Model parameters

Table 1 explains the epidemiological meaning of model parameters along with whatever information is known about them.

Table 1. Model parameters, their interpretation, and plausible ranges.

Parameter	Interpretation	Information
ν	*per capita* birth rate	covariate
μ	*per capita* death rate	calculated
N	population size	covariate
$1/\sigma$	latent period	1 da $\leq 1/\sigma \leq$ 30 da
$1/\gamma$	infectious period	1 da $\leq 1/\gamma \leq$ 30 da
ι	annual influx of infectives	$\iota > 0$
$\bar{\beta}$	mean contact rate	$\bar{\beta} > 0$
b_1	amplitude of seasonality	$0 \leq b_1 \leq 1$
$1/\alpha_1$	mean period in R_1	1 da $\leq 1/\alpha_1 \leq 1/\mu$ yr
$1/\alpha_2$	mean period in R_2	1 da $\leq 1/\alpha_2 \leq 1/\mu$ yr
ξ	probability of immune boosting	$0 \leq \xi \leq 1$
ϕ	probability of not developing immunity	$0 \leq \phi \leq 1$
p	vaccination probability	$0 \leq p \leq 1$
ϵ	force of infection modulator	$0 \leq \epsilon \leq 1$
ρ	under-reporting rate	$0 < \rho \leq 1$
τ	reporting variability	$\tau > 0$
σ_ζ	environmental stochasticity	$\sigma_\zeta > 0$

5. The pertussis model implemented

The model above is implemented in the package and can be loaded with the command

```
> data(pertussis.sim)
```

The object thereby loaded contains a named length-8 list of pomp objects:

```
> names(pertussis.sim)
```

```
[1] "SEIR.small"  "SEIR.big"    "SEIRS.small" "SEIRS.big"    "SEIRR.small"
[6] "SEIRR.big"   "full.small"  "full.big"
```

There are three submodels (SEIR, SEIRS, SEIRR), representing a model with permanent immunity (SEIR), temporary immunity (SEIRS), and temporary immunity with boosting (SEIRR). The model labeled `full` is SEIRR but with polarizing immunity ($\phi > 0$) and environmental stochasticity $\sigma_\zeta > 0$. For each of these, there are 2 simulations, one at each of two population sizes (5×10^5 and 5×10^6). The parameter values used in the simulations are given in Table 2. The simulations are performed using an Euler-multinomial approximation to the continuous-time Markov chain implicitly defined by Fig. 1. Details on the implementation of this algorithm are given in He et al. (2010, Box 1).

TABLE 2. Parameter values used for the simulated data.

parameter	R name	SEIR	SEIRS	SEIRR	full
ν	birthrate	0.02	0.02	0.02	0.02
μ	deathrate	0.02	0.02	0.02	0.02
$\hat{\beta}$	mean.beta	450.00	150.00	150.00	150.00
b_1	ampl.beta	0.15	0.15	0.15	0.15
ι	imports	10.00	10.00	10.00	10.00
σ	sigma	46.00	46.00	46.00	46.00
γ	gamma	26.00	26.00	26.00	26.00
α_1	alpha	0.00	0.10	0.10	0.10
α_2/α_1	alpha.ratio	1.00	1.00	1.00	1.00
ρ	report.prob	0.30	0.10	0.11	0.10
ξ	boost.prob	0.00	0.00	0.75	0.75
ϕ	polar.prob	0.00	0.00	0.00	0.10
ϵ	foi.mod	0.00	0.00	0.50	0.50
σ_ζ	noise.sigma	0.00	0.00	0.00	0.01
τ	tau	0.01	0.01	0.01	0.01

Plots of the data are shown in Fig. 2.

We can get a benchmark for likelihood-based fitting methods by computing the true likelihood at the true parameter values. To do this, we run the `pfilter` particle filtering code.

Table 3 shows these likelihoods.

TABLE 3. Estimated log likelihood at the true parameters for the simulated pertussis data. To obtain these, 10 particle filtering runs, each with 10000 particles, were used. The column labeled "se" gives the standard error of the Monte Carlo likelihood calculation. The computation took 360 CPU mins on inexpensive processors.

dataset	loglik	se
SEIR.small	-3890.1	0.34
SEIR.big	-5788.9	0.43
SEIRS.small	-3828.7	0.21
SEIRS.big	-5704.5	0.18
SEIRR.small	-3795.5	0.17
SEIRR.big	-5712.1	0.19
full.small	-3828.8	0.22
full.big	-5828.3	0.26

TABLE 4. Design of the pertussis simulation study. Studies designated "alt" are simpler versions to be undertaken only if the corresponding hard version proves too difficult. The studies which use `full` as the truth are to be undertaken only if the first set prove too easy.

Truth	Hypothesis	Parameters to estimate	Parameters fixed
SEIRR.small	SEIR	$\bar{\beta}, b_1, \rho$	$\alpha_1 = \alpha_2 = \varepsilon = \phi = 0$
SEIRR.small	SEIRS	$\bar{\beta}, b_1, \rho, \alpha_1, \alpha_2$	$\varepsilon = \phi = 0$
SEIRR.small	SEIRS (alt)	$\bar{\beta}, b_1, \rho, \alpha_1$	$\alpha_2 = \alpha_1, \varepsilon = \phi = 0$
SEIRR.small	SEIRR	$\bar{\beta}, b_1, \rho, \alpha_1, \alpha_2, \varepsilon, \xi$	$\phi = 0$
SEIRR.small	SEIRR (alt)	$\bar{\beta}, b_1, \rho, \alpha_1, \varepsilon$	$\alpha_2 = \alpha_1, \xi = \phi = 0$
full.small	SEIR	$\bar{\beta}, b_1, \rho$	$\alpha_1 = \alpha_2 = \varepsilon = \phi = 0$
full.small	SEIRS	$\bar{\beta}, b_1, \rho, \alpha_1, \alpha_2$	$\varepsilon = \phi = 0$
full.small	SEIRS (alt)	$\bar{\beta}, b_1, \rho, \alpha_1$	$\alpha_2 = \alpha_1, \varepsilon = \phi = 0$
full.small	SEIRR	$\bar{\beta}, b_1, \rho, \alpha_1, \alpha_2, \varepsilon, \xi$	$\phi = 0$
full.small	SEIRR (alt)	$\bar{\beta}, b_1, \rho, \alpha_1, \varepsilon$	$\alpha_2 = \alpha_1, \xi = \phi = 0$

6. DESIGN OF THE STUDY

Our objective is to simulate scientific inference focused on the question of waning and boosting of immunity. Specifically, we imagine we have a dataset and we wish to know (1) whether we can distinguish the signature of waning of immunity in these data, and, if we can do so, (2) whether we can distinguish that of boosting of immunity in the same data. We will therefore fix upon a single data set, `SEIRR.small`, as the focus of our study. We will then fit each of the SEIR, SEIRS, and SEIRR models to these data and use model selection techniques of one sort or another to identify the best model. To keep the problem tractable, we will assume that we know certain parameters and can therefore treat them as fixed. These include the latent period $(1/\sigma)$, the infectious period $(1/\gamma)$, the rate of imported infections (ι), the birth and death rates (ν, μ), the population size N, and the measurement model overdispersion parameter τ. We will suppose that no vaccination occurs $(p = 0)$ and that all infections result in immunity $(\phi = 0)$. In addition, we will initially assume that the process noise is purely demographic: $\sigma_\zeta = 0$. Thus, for every model, we must fit the three parameters $\bar{\beta}$, b_1, and ρ. For the SEIR model, these are the only parameters that must be fit, since we assume $\alpha_1 = \alpha_2 = \varepsilon = 0$. For the SEIRS model, we must additionally fit α_1 and α_2, though we assume $\varepsilon = 0$. Finally, in the SEIRR model, we must fit ε and ξ in addition to all the others, for a total of seven parameters. This is summarized in Table 4.

In the event that it proves overly time-consuming to effectively fit the five parameters of the SEIRS model, we might allow ourselves to assume that $\alpha_1 = \alpha_2$. In any case, it is likely that the relative values of these two parameters will be poorly identified. Likewise, if it proves really troublesome to fit six or seven parameters in the SEIRR model, we might allow ourselves to assume that $\xi = 0$. This would have the interesting effect of introducing model misspecification into the problem and give us the opportunity to explore and comment on that issue as well.

Prescribed ignorance. At the outset, we do not know where the MLEs for the SEIR and SEIRS models will be beyond the obvious facts that all the parameters are positive and ρ and ξ are probabilities. Based on other research, we have some reason to expect to uncover identifiability issues between the reporting rate ρ and the other parameters, especially $\bar{\beta}$. In addition, waning of immunity acts to increase the overall mean prevalance and should thus be reflected in high $\bar{\beta}$ estimates for the SEIR model. As mentioned above, we expect identifiability problems to crop up between α_1 and α_2 as well. For these reasons, it may initially be best for optimization approaches to initialize optimizers over a wide range and Bayesian approaches to use very flat priors. Once we can form a clearer notion of where the MLEs exist, it may be worthwhile to revisit this issue.

References

He, D., E. L. Ionides, and A. A. King. 2010. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. Journal of the Royal Society Interface, **7**:271–283.

Wearing, H. J. and P. Rohani. 2009. Estimating the duration of pertussis immunity using epidemiological signatures. PLoS Pathogens, **5**:e1000647.
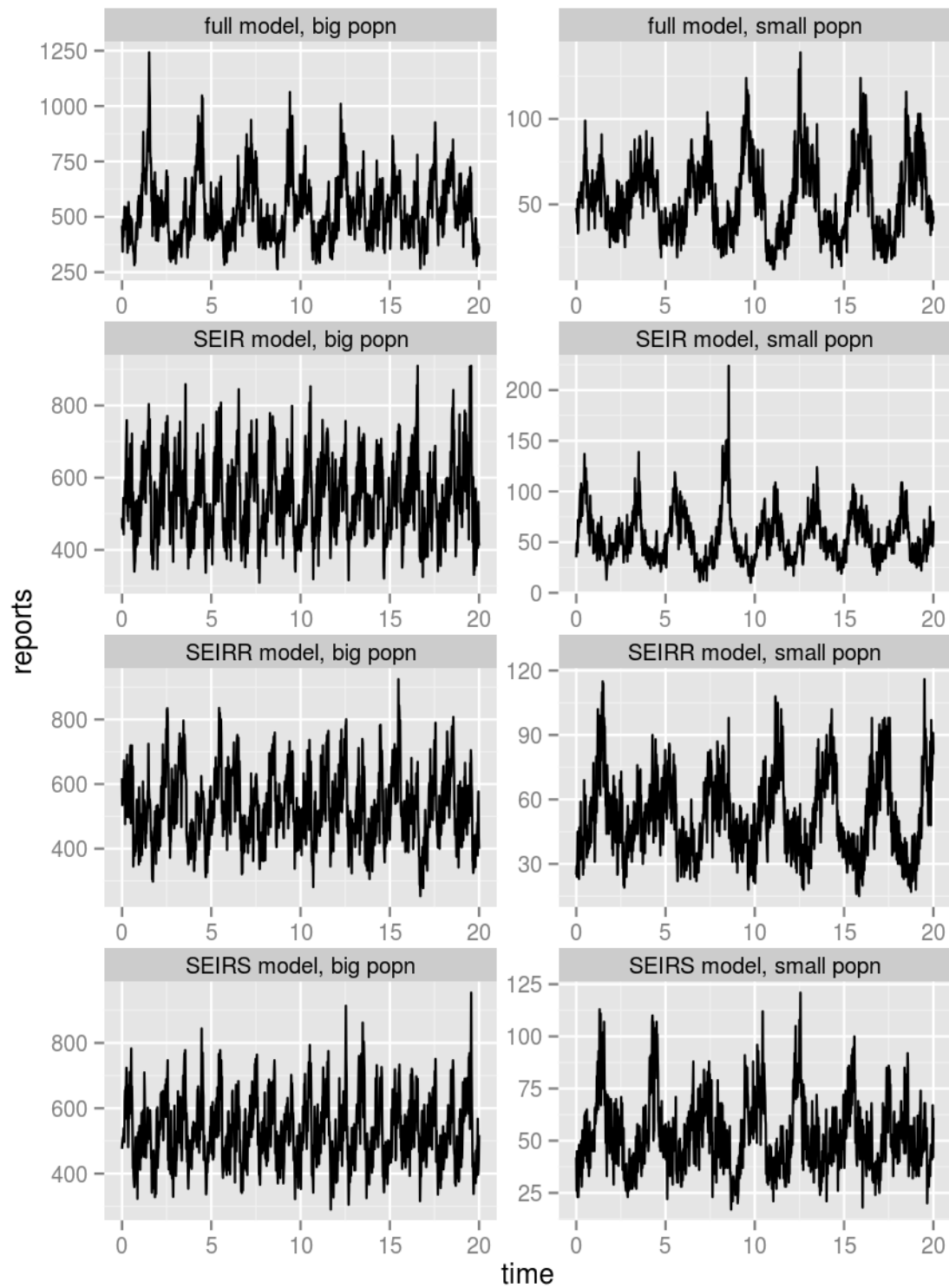
FIGURE 2. Plots of the simulated pertussis data. Population size 5×10^5 (black); 5×10^6 (red). Top: the full data sets. Bottom: the first 20 yr of simulated data.