Modelling the long-term dynamics of pre-vaccination pertussis

Ganna Rozhnova1,2,* and Ana Nunes1

1Departamento de Física, Centro de Física da Matéria Condensada, Faculdade de Ciências da Universidade de Lisboa, 1649-003 Lisboa Codex, Portugal
2Kavli Institute for Theoretical Physics, University of California, Kohn Hall, Santa Barbara, CA 93106-4030, USA

The dynamics of strongly immunizing childhood infections is still not well understood. Although reports of successful modelling of several data records can be found in the previous literature, the key determinants of the observed temporal patterns have not yet been clearly identified. In particular, different models of immunity waning and degree of protection applied to disease- and vaccine-induced immunity have been debated in the previous literature on pertussis. Here, we study the effect of disease-acquired immunity on the long-term patterns of pertussis prevalence. We compare five minimal models, all of which are stochastic, seasonally forced, well-mixed models of infection, based on susceptible–infective–recovered dynamics in a closed population. These models reflect different assumptions about the immune response of naïve hosts, namely total permanent immunity, immunity waning, immunity waning together with immunity boosting, reinfection of recovered and repeat infection after partial immunity waning. The power spectra of the output prevalence time series characterize the long-term dynamics of the models. For epidemiological parameters consistent with published data, the power spectra show quantitative and even qualitative differences, which can be used to test their assumptions by comparison with ensembles of several-decades-long pre-vaccination data records. We illustrate this strategy on two publicly available historical datasets.

Keywords: pertussis; childhood infections; recurrent epidemics; stochastic fluctuations; power spectra

1. INTRODUCTION

Childhood infections remain a public health concern as well as a modelling challenge, despite many decades of control measures and theoretical efforts. Large-scale vaccination programmes against some of these diseases started in the 1940s–1960s in developed countries and led to marked reductions of incidence levels, but in developing countries they are still a cause of significant levels of infant mortality [1–5]. The resurgence of pertussis reported in developed countries after decades of high vaccine coverage [6–10], as well as a recent upsurge of measles in eastern and southern Africa, prompted a renewed interest in assessing the efficacy of control policies for these childhood infections.

Such an assessment must rely on simulations based on mathematical models [11–14], whose more recent versions highlight the importance of contact structure, stochasticity and seasonality as drivers of the observed temporal patterns of incidence [15–22]. Concurrently, analytical tools to deal with these ingredients have been developed [23–30].

Models of higher complexity involve many parameters, some of which are difficult to determine, leaving considerable room for data fitting [31–34]. On the other hand, one of the features of childhood infections is the variability exhibited by different data records, even within those that correspond to a single disease in comparable social environments [4,17,35]. Therefore, successful modelling of particular datasets reported in the literature [17,36,37] has not closed the discussion about the key determinants of the observed dynamics. Measles is an exception: a parsimonious, stochastic, susceptible–infective–recovered (SIR)-based model has been shown to adequately describe disease incidence in large urban populations [15,38,39].

At the opposite extreme, pertussis keeps defying mathematical modelling, as illustrated by recent efforts in different directions [10,37,40–43]. In particular, several proposals explore the hypothesis about disease-induced and vaccine-induced immunity [10,35,43], relying on assumptions about vaccine uptake and efficacy for the purpose of comparison with real data. However, we are still lacking a sound uncontested model for pre-vaccination pertussis.

Here, we have sought to contribute to the goal of using pre-vaccination data records to obtain information about the properties of disease-induced
immunity. The paper focuses on exploring the influence of naive hosts’ immune response in the long-term patterns of pertussis prevalence as given by the averaged power spectra of simulated time series corresponding to several decades. The power spectrum of the stochastic, seasonally forced, well-mixed SIR model is compared with those of four modifications of the model that reflect different assumptions about disease-induced immunity. These four different assumptions have been proposed in the literature in the framework of deterministic models that have all been shown to be compatible with available data. The five variants we consider deliberately avoid all the complications related to contact structure and spatial spread, as well as vaccine coverage and efficacy and waning of vaccine-induced immunity, in order to keep a relatively low number of free parameters in the model.

We show that the stochastic versions of the SIR model and its four variants have significantly different properties, which translate into quantitatively different prevalence and incidence power spectra. This opens the possibility of using the stochastic properties of long, well-resolved data records to constrain these and other variants of the model, with the power spectra of the pertussis incidence time series in large urban centres in the pre-vaccination era as the target long-term dynamics that the model should reproduce. We illustrate this strategy by applying it to two publicly available historical data records for pre-vaccination pertussis incidence.

2. MODELS AND METHODS

2.1. Models

We consider five seasonally forced, stochastic compartmental models, summarized in figure 1, as candidates for the description of pre-vaccination pertussis. Their deterministic counterparts are described in the electronic supplementary material.

In all cases, the population includes three classes of individuals, the susceptible \((S S)\), the infected \((I I)\) and the recovered \((R R)\). Also, in all five cases, there is replenishment through births of naive susceptibles. The birth rate is kept constant and equals the death rate \(\mu\).

2.1.1. SIR model

Model \((a)\) is the standard SIR model: recovery confers permanent immunity, and all infections are first infections. Infected individuals recover at a constant rate \(\delta\), and susceptible individuals are infected at a rate \(\lambda = \beta(t)I/N\), where \(N\) is the population size, \(I\) is the number of infected individuals and \(\beta(t)\) is a periodic function of period 1 year that represents the variable contact rate associated with the school terms. To study the long-term temporal patterns of the fluctuations of this stochastic process, the particular form of \(\beta(t)\) is not crucial, and we shall take \(\beta(t) = \beta_0(1 + \beta_1 \cos 2\pi t)\).

2.1.2. SIRS model

Model \((b)\) is the standard SIRS model. It differs from the SIR model in that recovery does not confer permanent immunity. Instead, immunity wanes at a constant rate \(\gamma\). The fraction of primary infections is given approximately by \(\mu/[\gamma(1 - s^* - \hat{r}^*) + \mu]\), where \(s^*\) and \(\hat{r}^*\) are the equilibrium values for the susceptibles and infectives in the deterministic counterpart of the model (see the electronic supplementary material for details). Although there is no direct correspondence between primary infections and age groups in the model, it is reasonable to admit that a fraction of the infectives close to the fraction of non-primary infections, given by \(\gamma(1 - s^* - \hat{r}^*)/[\gamma(1 - s^* - \hat{r}^*) + \mu]\), are not school children, and that \(\beta_1\) should be reduced proportionally. This correction could be refined, but, as we shall see in §3, the properties of the stochastic fluctuations in this model are largely insensitive to the value of \(\beta_1\).

2.1.3. SIRWS model

Model \((c)\) is a SIRS-type model allowing for immune boosting of recovered individuals upon re-exposure to infection. Infectives recover at rate \(\delta\) and susceptibles \(S\) get infected at rate \(\lambda_1\), as in the SIR model. However, in addition to the \(S\) and \(I\) classes, there are two classes of recovered individuals denoted by \(R\) and \(W\). The immunity of the former is not permanent and they move to the waning class \(W\) at a constant rate \(2\gamma\). The individuals in class \(W\) undergo two possible transitions: further immunity loss at rate \(2\gamma\) to become susceptible \(S\), or immunity boosting upon contact.

Figure 1. Five compartmental models for pertussis. Models \((a)\) and \((b)\) are the seasonally forced SIR and SIRS models, respectively, while models \((c - e)\) are extensions of the SIR(S) paradigm. Model \((c)\) allows for boosting of immunity proportional to the force of infection, while model \((d)\) allows for reinfection of recovered. Model \((e)\) accounts for a loss of infection-derived immunity and subsequent reinfection. Transitions corresponding to the birth and death of individuals were omitted for simplicity. The symbols associated with each transition are defined in the text.
with infectious individuals to return to the recovered class $R$ at a rate $\kappa \lambda_r = \kappa \beta_s I / N$, where $\kappa$ is the immunity boosting coefficient. We call this scheme the immune boosting model and denote it by SIRWS.

An age-structured version of this model with vaccination has been used to explain the recent re-emergence of pertussis cases, despite high vaccine coverage in Massachusetts, USA [10].

2.1.4. SIRIS model

Model $(d)$, proposed by Aguas et al. [43], sets a scenario based on the SIRS model with a moderate rate $\gamma$ of immunity loss in which recovered individuals are immune to severe disease but susceptible with reduced susceptibility to mild forms of the disease. The classes of individuals infected with severe and mild infections are denoted by $I_1$ and $I_2$, respectively. Recovery from both forms of the disease takes place at the same rate $\delta$. Infectiousness of mild infections is reduced by a factor $\eta \in [0,1]$, and susceptibility of recovered individuals to mild infections is also reduced by a factor $\sigma \in [0,1]$. Moreover, because mild infections typically occur in adults who are not affected by seasonal forcing, the force of infection is therefore taken to be $\lambda = (\beta(t) I_1 + \beta_s \eta I_2) / N$ for susceptible individuals $S$ and $\lambda = \beta_s (I_1 + \eta I_2) / N$ for recovered individuals $R$. We call this scheme the reinfection model and denote it by SIRIS.

The main feature of this model is that it exhibits a reinfection threshold [44], a value of infectiousness above which total disease incidence rises by one or more orders of magnitude, owing to high incidence levels of mild infections. These mild infections are sub-clinical and so the disease incidence/prevalence is associated with the class $I_2$.

2.1.5. SIRSI model

Finally, model $(e)$, proposed in the study of Wearing & Rohani [40], assumes an immune response that is a combination of $(b)$ and $(d)$, in the sense that recovered individuals are fully immune to the disease, but they lose this immunity at a certain rate $\gamma$ to become susceptible to repeat infections, although with reduced susceptibility. The two classes of susceptible individuals are denoted by $S_1$ for the naive susceptibles and $S_2$ for the susceptibles generated by immunity waning, and $\sigma \in [0,1]$ is the reduced susceptibility factor for repeat infections. The classes of individuals infected with first and repeat infections are denoted by $I_1$ and $I_2$, respectively. Recovery from both classes takes place at the same rate $\delta$. The class of repeatedly infected individuals contributes with reduced infectiousness to the pool of infectives responsible for disease transmission, and $\eta \in [0,1]$ is the reduced infectiousness factor. Repeat infections play a similar role in this model to that of mild infections in the SIRIS model, and the force of infection is taken to be $\lambda = (\beta(t) I_1 + \beta_s \eta I_2) / N$ for $S_1$ and $\lambda = \beta_s (I_1 + \eta I_2) / N$ for $S_2$, because repeat infections typically occur in adults whose contacts are not subject to school-term forcing. Similar to the SIRIS model, the disease incidence/prevalence is identified with the class $I_1$. We call this scheme the repeat infection model and denote it by SIRSI.
new cases in a month interval is similar for the two data-sets, indicating a ratio of effective populations close to 1. Differences in surveillance coverage and/or in reporting efficiency could explain this. For London, estimates of pertussis reporting efficiency in the 5–25% range can be found in the literature [46]. The fact that the effective population for Ontario is approximately of the same size indicates that a higher reporting efficiency compensates for the smaller population.

It is known from several case notification datasets that the interepidemic periods for pertussis in the pre-vaccination era show significant multi-annual structure. The multi-annual periods range from 2 to 3 years, and annual outbreaks are also observed [17,40,47]. The power spectra computed from the data confirm these observations (figures 3b and 4c). In the spectra plots, the shaded region marks the range of frequencies, $f \in [1/3, 1/2]$ yr$^{-1}$, corresponding to the interepidemperiods $1/f \in [2,3]$ years. Because of limited length and resolution in time, the spectra have a short range and a low resolution in frequency that affects mostly the annual peak. The widths of the multi-annual peaks are comparable. For Ontario, the power spectra of the full time series and of the detrended and undetrended partial datasets show the same dominant frequency components. In what follows, we shall consider only the power spectrum of the undetrended data.

Figure 3. (a) New weekly cases of pertussis in (Greater) London before vaccination. The dashed horizontal line is the average number of weekly cases. (b) Spectrum of the time series. (Online version in colour.)

Figure 4. (a,b) New monthly cases of pertussis in the Canadian province of Ontario before vaccination (a) and the detrended time series (b). The dashed horizontal line in (b) is the average number of monthly cases. (c) The solid black line is the spectrum of the full time series shown in (b), the dashed-dotted and the dashed lines are the spectra calculated from the partial time series. (Online version in colour.)
Table 1. The epidemiological description of the parameters and the range of their values for pertussis.

<table>
<thead>
<tr>
<th>Epidemiological meaning</th>
<th>Notation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per capita birth/death rate</td>
<td>( \mu )</td>
<td>( 1/50 \text{ yr}^{-1} )</td>
</tr>
<tr>
<td>Average lifespan</td>
<td>( 1/\mu )</td>
<td>50 years</td>
</tr>
<tr>
<td>Rate of recovery from infection</td>
<td>( \delta )</td>
<td>( 365/22 \text{ yr}^{-1} )</td>
</tr>
<tr>
<td>Average infectious period</td>
<td>( 1/\delta )</td>
<td>22 days</td>
</tr>
<tr>
<td>Basic reproductive ratio</td>
<td>( R_0 )</td>
<td>282.39 \text{ yr}^{-1}</td>
</tr>
<tr>
<td>Average contact rate</td>
<td>( \beta_1 )</td>
<td>( [0.01, 0.1] \text{ yr}^{-1} )</td>
</tr>
<tr>
<td>Amplitude of seasonal forcing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate of loss of naturally acquired immunity</td>
<td>( \gamma )</td>
<td>( [1/40, 1/10] \text{ yr}^{-1} )</td>
</tr>
<tr>
<td>Average infection-derived immunity period</td>
<td>( \eta )</td>
<td>([10, 40] \text{ years} )</td>
</tr>
<tr>
<td>Relative infectiousness of repeat to primary (SIRSI)</td>
<td>( \sigma )</td>
<td>([0, 1] )</td>
</tr>
<tr>
<td>Relative susceptibility of repeat to primary (SIRSI)</td>
<td>( \kappa )</td>
<td>([0, 2.20] )</td>
</tr>
</tbody>
</table>

For the SIRS and SIRSI models, the whole range of possible \( R_0 \) values estimated from the average age at first infection is explored in the electronic supplementary material.

### 2.3. Parameter values

The values of the demographic and epidemiological parameters that are well established in independent data sources \([17,45]\) are kept fixed. These are the birth/death rate \( \mu \) (or the average lifespan \( 1/\mu \)), the rate of recovery from infection \( \delta \) (or average infectious period \( 1/\delta \)) and the average contact rate \( \beta_1 \). The value of the last corresponds to the basic reproductive ratio \( R_0 = \beta_1/(\delta + \mu) \) reported in the literature for the two datasets considered in this study \([17]\). If the reported value of \( R_0 \) comes from average age at infection data, together with the assumption of SIR dynamics, then it has to be corrected for models with temporary immunity. In the electronic supplementary material, we explore the whole range of possible \( R_0 \) values for the SIRS and SIRSI models and show that the analysis of §3 below is not affected by this correction. We also give for the SIRWS and SIRIS models the dependence of \( R_0 \) computed from average age at infection data on the model parameters.

For the remaining parameters, we either use the estimates found in the literature for a particular model or explore a range of possible values if such estimates are absent. For the SIRS model, we take \( 1/\gamma \) to be equal to 20–40 years. This is above the accepted range of duration of naturally acquired immunity for pertussis \([48]\), but the large upper limit allows for a comparison with the SIR model. On the other hand, as we shall see, taking \( 1/\gamma \) smaller than 20 years would worsen the model’s predictions. For the SIRSI model, we take \( 1/\gamma = 20 \text{ years} \) \([43]\) and the relative infectiousness and relative susceptibility of mild infections \( \eta, \sigma \in [0, 1] \). This analysis can be extrapolated to other values of \( 1/\gamma \), using the fact that this model has the SIRS model as one of its limiting cases. For the SIRWS model, we take the value of the boosting coefficient \( \kappa = 20 \) and \( 1/\gamma = 10 \text{ years} \) considered in Lavine et al. \([10]\) and, in addition, we study the behaviour of the model for \( 1/\gamma = 40 \text{ years} \) and \( \kappa \in [0.2, 10] \). For the SIRSI model, we set \( 1/\gamma = 20 \text{ years} \) and the relative infectiousness of and relative susceptibility to repeat infections \( \eta, \sigma \in [0, 1] \) \([40]\). We discuss in the electronic supplementary material the behaviour of this model for other values of the duration of immunity.

For illustration purposes, we use only three values of the amplitude of seasonal forcing \( \beta_1 = 0, 0.05, 0.1 \) and discuss the predictions of the models for intermediate values of \( \beta_1 \) whenever necessary. A summary of the parameter values is given in table 1.

### 2.4. Stochastic simulations

To study the behaviour of the five candidates for the description of pre-vaccination pertussis, we use stochastic simulations of the processes described in figure 1. The simulations are based on a modification of Gillespie’s algorithm \([49]\), which accounts for the explicit time dependence in the contact rate \([50,51]\). In §§3.1–3.5, each simulation run starts from a random initial condition and the prevalence of the disease is recorded with a time step of 0.05 years for 450 years after 50 years of transient period. From each simulation run, the power spectrum of the prevalence time series is computed with the use of the discrete Fourier transform. The final spectra are averages of \( 10^5 \) simulations. The population size used in the simulations in §§3.1–3.5 is \( 5 \times 10^5 \). Changing population size and/or computing the incidence in a given interval instead of the prevalence of the disease may change somewhat the amplitude of the multi-annual and annual peaks but does not influence their position.

### 3. RESULTS

We investigate the dynamics of the stochastic, seasonally forced, SIR, SIRS, SIRWS, SIRIS and SIRSI models by comparing the power spectra of long time series of the five models in the relevant parameter ranges. We first describe the performance of the SIR model and then assess whether each of its four extensions improves or worsens the results. In all cases, the spectra correspond to the fluctuations around the equilibrium of the unforced deterministic versions of each model \((\beta_1 = 0)\) or around the associated period 1 year.
stable attractor ($\beta_1 > 0$). Numerical and simulational results for the parameter ranges we have explored never showed evidence of other attractors. A direct quantitative comparison of these power spectra with the spectra shown in figures 3 and 4 based on the amplitude of the multi-annual peaks and their precise location is undermined by the low resolution in frequency and poor statistics of the two datasets, which correspond to much shorter periods than those that can be used in the numerical simulations. That is why our first criterion in making the comparison between the models’ predictions and the spectra computed from the data will be the position of the multi-annual peak. In order to be compatible with data records for pre-vaccination pertussis, a model’s spectrum should have the multi-annual peak located in the range of frequencies $[1/3, 1/2] \text{ yr}^{-1}$ (figures 3 and 4). This region is the shaded region shown in all plots. All simulation spectrum plots are shown in lin-log scale, and a dashed line indicates the dominant frequency of the stochastic multi-annual peak.

### 3.1. SIR model

The model’s power spectrum computed from stochastic simulations for different values of the forcing amplitude $\beta_1$ is shown in figure 5. The forced spectra exhibit two peaks: a dominant annual peak due to the deterministic limit cycle and a broad subdominant multi-annual stochastic peak similar to that of the unforced case. The contribution of annual epidemics in the time series increases as $\beta_1$ increases, resulting in an enhanced annual peak. The multi-annual stochastic peak reflects the presence of noisy oscillations in the incidence time series with that dominant frequency. The mechanisms by which internal noise excites these resonant fluctuations has been treated in several papers [26,27,30,41]. They can be described analytically using van Kampen’s expansion of the master equation of the underlying stochastic process around the attractor of the deterministic system [52].

For the SIR model, the dominant period of stochastic fluctuations is 2.7 years. The stochastic peak lies in the shaded region, and its frequency and shape are largely independent of $\beta_1$. Therefore, in this study, the dominant frequency of the main stochastic peak in the seasonally forced models can be computed from the unforced model whose analysis is easier (see the electronic supplementary material for details on the analytical computation of the power spectra). In §§3.2–3.5, we will discuss how the SIRS, SIRWS, SIR and SIRSI extensions of the SIR model are reflected in the position and amplitude of the stochastic peak.

### 3.2. SIRS model

In the SIRS model, the rate of immunity waning $\gamma$ is a free parameter. In the case of lifelong immunity, $\gamma = 0$, the SIRS model reduces to the SIR model considered in the previous section. Figure 6 shows the model’s spectrum for different values of the forcing amplitude $\beta_1$ and two typical values of $\gamma$. In the unforced SIRS model, $\beta_1 = 0$, the spectrum has a pronounced stochastic peak. However, it is situated outside the region of interest both for $1/\gamma = 40$ years and for $1/\gamma = 20$ years. For fixed $\gamma$ and increasing $\beta_1$, the frequency of the stochastic peak is slightly shifted towards the target region, but it remains outside this region even for large values of $\gamma$ and, at the same time, the power associated with the annual peak becomes very large. Note that the dominant frequency of the stochastic peak would be even further away from the shaded region if $\beta_1$ were decreased to take into account that non-primary infections do not undergo seasonal forcing. We conclude that the SIRS model’s spectrum is compatible with the London and Ontario data only in the limit of $\gamma \rightarrow 0$ when it approaches the SIR model. The performance of the SIRS model for other values of $R_0$ is discussed in the electronic supplementary material.

### 3.3. SIRIS model

Next, we analyse the power spectra from simulations of the SIRIS model. This model has three free parameters apart from $\beta_1$: the rate of immunity loss $\gamma$, and the relative infectiousness $\eta$ and relative susceptibility $\sigma$. For the limiting case of $\eta = 0$ and $\sigma = 0$, the model reduces to the SIRS model whose multi-annual peaks are located outside the shaded region. The results for fixed $\beta_1 = 0.05$, $1/\gamma = 20$ years and different values of $\eta$ and $\sigma$ are shown in figure 7. The spectrum for small values of $\eta$ and $\sigma$ is reminiscent of the SIRS model. A comparison of the spectra in figure 7a–c shows that the stochastic and deterministic peaks are the highest for the smallest values of both $\eta$ and $\sigma$. This is because it is easier to generate incidence oscillations in this parameter range. Increasing $\eta$ and $\sigma$ results in a flattening of the spectrum occurring through a gradual disappearance of, first, stochastic and, then, deterministic peaks. Moreover, varying the amplitude of seasonal forcing does not change this picture. For example, for the smallest values of $\eta$ and $\sigma$ used in figure 7, we observe that the position and the shape of the stochastic...

Figure 5. The average power spectra from simulations of the SIR model for different amplitudes of seasonal forcing, $\beta_1$. Parameters: green line, $\beta_1 = 0$; orange line, 0.05; black line, 0.1. (Online version in colour.)
peaks agree perfectly for $\beta_1 \in [0, 0.1]$ and that they are always outside the region of interest (this result is given in the electronic supplementary material). The parameter region with higher values of $\eta$, $h$, and $s$ of mild to severe infections is even less interesting because of the flattening of the spectrum. Thus for all relevant ranges of $h$, $s$, and $b_1$, the stochastic multi-annual peak in this model is situated outside the region of interest except when it approaches the limit of the SIR model.

3.4. SIRWS model

Here, we show the power spectra computed from simulations of the SIRWS model for boosting coefficient $\kappa = 20$ and different values of $\gamma$ and $\beta_1$ (figure 8). In the absence of seasonality, $\beta_1 = 0$, the spectra have a dominant multi-annual peak situated in the target region for the whole range of $1/\gamma \in [10, 40]$ years. When $\beta_1$ increases, a deterministic annual peak appears in the spectrum, while the dominant frequency of the multi-annual peak stays unchanged. Similar to the SIR model, the higher the amplitude of seasonal forcing $\beta_1$, the higher the annual peak. However, unlike in that model, this phenomenon is accompanied by an overall complication of the structure of the spectrum for higher values of $\beta_1$. For example, the power spectrum for $1/\gamma = 10$ years and $\beta_1 = 0.1$ (black line in figure 8c) has several secondary multi-annual peaks, one of which is situated near but outside the shaded region. Thus, the SIRWS model predicts a rather broad range of frequencies for the stochastic fluctuations.

Complementary results for $k \neq 20$ are given in the electronic supplementary material. For $\kappa = 0.2, 1, 10$ and $1/\gamma = 40$, the stochastic multi-annual peaks
of all power spectra lie in the shaded region. The same happens for $k = 10$ and $1/g = 10$ years but not for $k = 0.2$. For $\kappa \to 0$ and $\gamma \gg \mu$, the SIRWS model exhibits a SIRS-like behaviour, and thus the multi-annual peaks of the SIRWS model are outside the region of interest.

### 3.5. SIRSI model

Finally, we present the results for the SIRSI model. Figure 9 shows the spectra for different values of $\beta_1 \in [0, 0.1]$, $\sigma, \eta \in (0, 1)$ and $1/\gamma = 20$ years. As for SIR, the simulation spectra of the SIRSI model have a high stochastic multi-annual peak in addition to the annual peak. The amplitude of the deterministic peaks increases as $\beta_1$ increases while the stochastic peaks remain almost unchanged. The increase, however, is smaller in the SIRSI model than in the SIR model. This means that, notwithstanding the similarity of their simulation spectra, the contribution of annual epidemics is smaller in the former than in the latter, corresponding to a qualitative difference in the time series of the two models.

For the three sets of parameter values of figure 9, the dominant stochastic multi-annual peaks lie in the shaded region, and their frequencies and shapes are largely independent of $\beta_1$. To explore the consequences of varying $\eta$ and $\sigma$ for fixed $\gamma$, we considered in the interval (0,1) nine equally spaced values and constructed a grid of 81 points in the $(\eta, \sigma)$ space. As the position and the shape of the stochastic peak in this model is independent of $\beta_1$, we can calculate the stochastic peak’s frequency from the unforced model (see the electronic supplementary material for more details). For example, for $1/\gamma = 20$ years used in figure 9, the number of spectra for the SIRSI model with the
stochastic peak within the shaded region is 62 (for \(1/\gamma \in [10, 40]\) years, this number varies from 43 to 78). We thus conclude that the SIRSI model’s spectrum illustrated in figure 9 is robust with respect to variation of all free parameters, exhibiting in the accepted parameter ranges some variability in the amplitude of the stochastic peak. The robustness of these results with respect to variations of \(R_0\) too is illustrated in the electronic supplementary material.

3.6. Comparison with real data

The results for the power spectra of stochastic simulations show that, out of the four extensions of the SIR model under consideration, only SIRWS and SIRSI have frequency values for the stochastic multiannual peak that preserve or improve the compatibility with data for pre-vaccination pertussis of the SIR model. Here we compare the other features of spectra of the SIR model and of the SIRWS and SIRSI extensions with the spectra for the London and Ontario datasets shown in figures 3 and 4.

We determine the population size \(N\) for each model, as a first step towards achieving this. As discussed before, both datasets have a similar average rate of new cases per month and thus correspond to similar effective populations. However, the deterministic counterparts of the models, and consequently the stochastic models too, predict different densities of infectives in the steady state. To resolve this, for each of the three models, we make simulations with the same length and sampling time as those of the London or Ontario datasets and record the incidence of the disease in the sampling interval. The transient period in all simulations is taken to be 50 years, and each simulation starts from a random initial condition. We calibrate the effective population size \(N\) for a given model by imposing that the incidence averaged over 10\(^3\) different simulation runs is the same as the time averaged incidence in the corresponding dataset, given by the dashed line in figures 3 and 4. In figure 10, we show typical incidence time series for \(\beta_1 = 0.05\) and the effective population size \(N\) for each of the three models from which individual power spectra are computed.

The second step is to determine for each model the level of seasonality that corresponds to the dataset.
We estimate the value of the amplitude of the seasonal forcing $b_1$ as the value for which the amplitude of the annual deterministic peak in the power spectrum of incidence time series averaged over $10^3$ different simulation runs is equal to the amplitude of the annual peak in the spectrum computed from the data.

Using the values of $N$ and of $b_1$ determined in this way, we produce from the stochastic models time series mimicking the London or Ontario datasets, and average their power spectra over $10^3$ simulation runs. The power spectra from simulations (dashed red (black) lines) and the spectra computed from the datasets (black solid lines) are shown in figure 11. The power spectra obtained with the SIR model have stochastic multi-annual peaks with amplitudes that are similar to, though on average larger than, those of the datasets’ power spectra. For the SIRWS extension, the average amplitude of the stochastic peak is even larger, while the SIRSI extension gives a better agreement with this feature of the data with respect to the SIR model. The results of figure 11 thus suggest that the SIRSI model is the one that best reproduces the stochastic properties of both datasets. However, as shown by the four examples plotted in the figure for each case (dashed grey lines), the poor statistics of the short-term samplings translate into a considerable variability in the power spectra computed from each time series. The two historical data records considered here illustrate the principle that the incidence power spectra can be used to discriminate the three variants of the model, but a final conclusion on their performance would require the analysis of longer time series.

4. DISCUSSION AND CONCLUSIONS

Different assumptions about disease- and vaccine-induced immunity have been proposed in the recent literature to model the dynamics of pertussis in the presence of mass vaccination. Our motivation has been to study separately the influence of disease-acquired immunity only, in order to reduce the number of free parameters, and to consider explicitly the stochastic properties of the incidence time series as part of the model’s output, in order to look for further constraints.

We have used the basic SIR model and four variants that reflect different immune responses. We characterized the dynamic patterns of each of these five models through the average power spectra of an ensemble of long-term prevalence time series obtained from stochastic simulations. We have found that the power spectra of the five alternative models show quantitative and
even qualitative differences. One of them (SIRS) fails to reproduce a stochastic peak in the frequency range found in real data for pertussis, and another one (SIRIS) is too stiff to produce stochastic fluctuations with the observed amplitudes.

Assessing the performance of the other two variants, SIRWS and SIRSI, with respect to the SIR model requires a quantitative comparison of the simulated power spectra with real data; for that purpose, the effective population size and the forcing amplitude of each data record must be determined. We illustrated this procedure by considering two publicly available historical data records for pre-vaccination pertussis incidence. On the basis of these relatively short time series, we found that the output of the SIRSI model agrees with the main aspects of the phenomenology of the disease in a broad parameter range. It reproduces the multiannual interepidemic periods and the amplitude of fluctuations found in the data with a better agreement than SIR for the latter, while the SIRWS variant tends to produce stochastic fluctuations even larger than SIR.

Considerations about the recovery profile and its influence on the model’s behaviour concur in favouring models such as SIRSI, with a much flatter spectrum than SIRWS. For unforced models, it is known [29] that the effect of changing from an exponentially distributed recovery profile to a unimodal gamma distributed recovery profile with the same average is an enhancement of the fluctuations, together with a small displacement of the main stochastic peak towards higher frequencies. We have checked (results not shown) that this holds too for the SIRSI and SIRWS models in the parameter region explored in the previous section. Therefore, under more realistic recovery profiles, the output of the SIRWS model would be further away from the target power spectra, as would be that of the SIR model.

However, the short length of the time series used in our illustrative study prevents any definite conclusion about the performance of the SIR model alternatives. Indeed, the ensemble of the power spectra obtained from simulated time series of this length exhibits enough variability to accommodate the two samples of real data. An extension of this analysis to a larger set of data could help to further elucidate the dynamics of disease-acquired immunity, setting a solid ground for more complex models to deal with vaccine-acquired immunity.

The authors thank Gabriela Gomes for helpful discussions. Financial support from the Portuguese Foundation for Science and Technology (FCT) under contracts no. POCTI/ISFL/2/261, SFRH/BD/32164/2006 and SFRH/BDP/69137/2010 is gratefully acknowledged. G.R. was also supported by the Calouste Gulbenkian Foundation Program ‘Stimulus for Research’ and by the National Science Foundation under grant no. NSF PHY05-51164.

REFERENCES


J. R. Soc. Interface (2012)