

Estimation of Invasive Pneumococcal Disease Dynamics Parameters and the Impact of Conjugate Vaccination in Australia

K. L. Sutton^{1,2}, H. T. Banks², and C. Castillo-Chávez¹

¹ Department of Mathematics & Statistics, Arizona State University, Tempe, AZ 85287-1804

² Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC 27695-8205

August 22, 2007

Abstract

Pneumococcal diseases, or infections from the etiological agent *Streptococcus pneumoniae*, have long been a major cause of morbidity and mortality worldwide. Recent advances in the development of vaccines for these infections have raised questions concerning their widespread and/or long-term use. In this work, we use surveillance data collected by the Australian National Notifiable Diseases Surveillance system to estimate parameters in a mathematical model of pneumococcal infection dynamics in a population with partial vaccination. The parameters obtained are of particular interest as they are not typically available in reported literature or measurable. The calibrated model is then used to assess the impact of the recent federally funded program which provides pneumococcal vaccines to large risk groups. The results presented here suggest the state of these infections may be changing in response to the programs, and warrants close quantitative monitoring.

1 Introduction

Infections caused by *Streptococcus pneumoniae*, or the pneumococcus, have long been a major cause of disease burden and mortality worldwide. Most affected are children under the ages of 5 in developing countries, where it is estimated that around 1 million die each year of pneumococcal pneumonia [33]. In more developed countries these diseases are prevalent in the young and the old, with the elderly population commonly being affected by invasive pneumococcal diseases (IPDs), commonly defined as pneumonia, meningitis, bacteremia and sepsis. These infections can result in case fatality, an outcome which occurs increasingly often in older age groups [1]. While children in more developed countries are typically afflicted by less serious infections such as otitis media, or the common ear

infection, it is one of the most common causes of childhood morbidity. In the US it is estimated that 90% of children experience at least one episode by 2 years of age ([20, 21]). Treatment efforts, usually consisting of antibiotic regimes, not only cost the US an estimated \$2 billion per year [23], but also result in the generation of antimicrobial resistance among common serotypes.

The pneumococcus is a gram-positive bacterium with over 90 serotypes/serogroups identified thus far. *S. pneumoniae* commonly colonizes the nasopharyngeal region of healthy adults, with estimates between 9-14 % of the population in the US and Spain. This prevalence varies depending on factors such as crowding, age-profiles, and previous exposure. This colonization is asymptomatic, occurs as a result of casual contacts with other colonized individuals, and occurs for shorter periods of time in healthy adults as opposed to that of young children [19, 22]. It is during this stage that the pneumococcus may successfully establish an infection or be cleared by one's immune response. Thus, it is important to note that this stage always precedes infection. Most individuals are able to clear these transient colonies, although some cannot without the aid of vaccination.

A polysaccharide vaccine protecting against 23 serotypes, which is relatively inexpensive and safe, has been available since 1983. However, this vaccine is ineffective in young children, who have not yet developed the ability to respond to polysaccharide antigens. In response to this problem, the conjugate vaccine, in which polysaccharides from 7 serotypes have been conjugated to a non-toxic protein, was developed and licensed in 2001. The implementation of this vaccine in the US has dramatically reduced cases of invasive disease in children, as well as decreased rates of carriage. In addition, infections have been reduced in those who are not vaccinated, suggesting a herd immunity effect. However, the long-term effects of these and other vaccines currently in development are unknown. Thus, theoretical population studies of the effects of vaccination on pneumococcal dynamics are of increasing interest. Of particular interest are theoretical studies that have been calibrated to a specific population and are therefore in a position to directly address questions such as the long-term effects of vaccines and the effects of potential control scenarios.

In this paper we discuss a mathematical model of pneumococcal infection dynamics with generic vaccination. Crucial model parameters, not typically available in relevant literature, are estimated using surveillance data collected by the Australian government [34]. An accurate picture of the disease dynamics in the Australian population is thereby described for January 2002 through December 2004. As of January 1, 2005, pneumococcal vaccination for all recommended risk groups was made publicly available and the number of infections has been dramatically reduced. The impact of this widespread vaccination is assessed based on the current data available, and the need for qualitative monitoring of this population is illustrated.

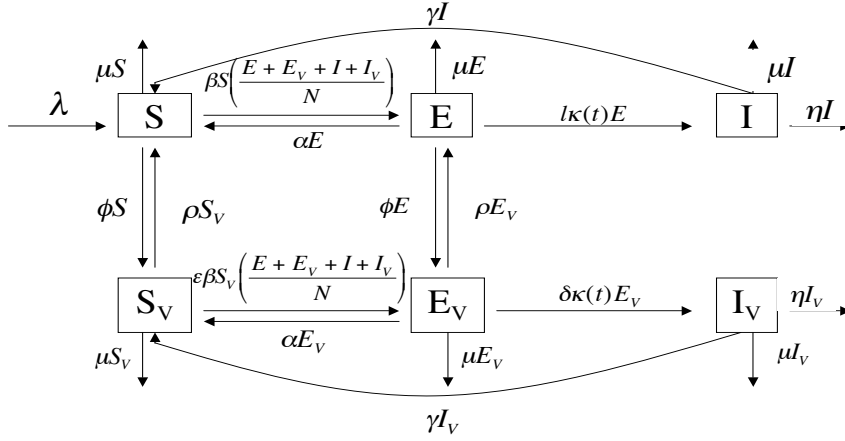


Figure 1: Pneumococcal infection dynamics with vaccination.

2 Model

We introduce a model of *Streptococcus pneumoniae* infection dynamics in which individuals are classified according to their epidemiological status. Individuals may be susceptible, represented by the S class, or asymptotically colonized by *S. pneumoniae* and in the E class. In the event that a colony of *S. pneumoniae* is able to successfully establish an infection, the individual is then considered infected or in the I class.

Individuals are recruited into the population at constant rate λ , with all newborns assumed to be susceptible. Susceptible individuals become asymptotically colonized through effective contacts, measured by per capita rate β , with other colonized individuals (E and E_V) and those with established pneumococcal infections (I and I_V). Thus, the transfer between these two compartments occurs at rate $\beta S \frac{E+E_V+I+I_V}{N}$ where N represents the total population and is calculated by the sum of all classes in the model, i.e., $N = S + S_V + E + E_V + I + I_V$ which is *not* assumed constant. Most individuals return to the susceptible class after colonization, i.e. clearing the bacteria, at per capita rate α . Infection occurs in only a small fraction of the colonized individuals, and is modeled by the rate $l\kappa E$, where $l < 1$ is a scaling factor for the proportion of the population that is ‘at risk’ and κ is the infection rate of this ‘at risk’ population. This linear term is a gross simplification, as this process is actually a function of many factors, namely the health of the individual’s immune system as well as their previous exposure to *S. pneumoniae* of similar serotype. Thus it is likely that a good way to model this transfer will require further study at the ‘within host’ level. Death due to pneumococcal infection occurs at per capita rate η , or the infection is cleared at rate γ . Natural death can occur at any stage at per capita rate μ .

Seasonality of invasive pneumococcal diseases has been observed and is incorporated here in the infection rate under the assumption that the fluctuations are due to seasonal

Variable/Parameter	Description	[Units]
$S(t)$	susceptible individuals	[individuals]
$E(t)$	asymptotically colonized individuals	[individuals]
$S_V(t)$	vaccinated susceptible individuals	[individuals]
$E_V(t)$	vaccinated colonized individuals	[individuals]
$I(t)$	individuals with infections	[individuals]
$I_V(t)$	vaccinated individuals with infections	[individuals]
λ	recruitment/birth rate	[individuals/month]
μ	natural death rate	[1/month]
β	effective contact rate	[1/month]
α	colonized recovery rate	[1/month]
l	proportion of susceptibles ‘at risk’	N/A
κ_0	mean infection rate	[1/month]
κ_1	multiple of seasonal deviation from mean infection rate	N/A
ω	frequency of seasonal oscillations in infection rate	N/A
τ	phase shift of infection rate	N/A
γ	infection recovery rate	[1/month]
η	disease-induced mortality rate	[1/month]
ϕ	effective vaccination rate	[1/month]
ρ	loss of vaccine-induced protection rate	[1/month]
ϵ	reduction of colonization events due to vaccination	N/A
δ	reduction of infection due to vaccination	N/A

Table 1: Descriptions and Units for Model Variable and Parameters.

changes in host susceptibility [6], [9]. Thus we ascribe the form to $\kappa(t)$ as

$$\kappa(t) = \kappa_0 (1 + \kappa_1 \cos[\omega(t - \tau)]) . \quad (1)$$

Seemingly in contrast to this assumption, it has been noted that the seasonal fluctuations of pneumococcal pneumonia correspond to those observed in influenza cases [15, 30]. However, while influenza fluctuations are thought to be a result of differences in effective contacts, it is reasonable to interpret the effect of influenza on pneumococcal disease dynamics as increasing host susceptibility. In fact, in a household longitudinal study of nasopharyngeal carriage rates over a 10 month period (including the months of highest IPD incidence), the prevalence of colonized individuals remained relatively constant [13]. In a study of frequent nasopharyngeal colonizers found seasonal changes but none attributable to *Streptococcus pneumoniae* [16]. Hendley, et al., found increased nasopharyngeal carriage due to *S. pneumoniae* among nasal swabs in children, but also demonstrated that these changes were due to increased nasal drainage during the colder months [12]. These studies support the implementation of a time-dependent infection rate and not a time-dependent effective contact rate.

We incorporate general pneumococcal immunization - that is, any type(s) of vaccine implemented may be considered simply by specifying parameter values. For this study, we consider a population vaccinated with either the 7-valent protein-polysaccharide conjugate vaccine (PCV7) or the 23-valent polysaccharide vaccine (PPV23). Upon effective vaccination, susceptible and colonized individuals enter the S_V and E_V classes, respectively, at per capita rate ϕ . Thus, only individuals who respond in some way progress to the vaccinated classes. Typically, adults over the age of 65 lose protection from the PPV23, which we model by the transfer of individuals in S_V and E_V to their corresponding unvaccinated classes at per capita rate ρ . In the case that only the PCV7 is being implemented, $\rho = 0$, since this is only recommended for children. However, PCV7 has noticeably conferred protection against colonization and thus this rate would be reduced by a factor $0 < \epsilon < 1$, ($1 - \epsilon$ would be the fraction by which colonization events are reduced). In contrast, PPV23 does not reduce susceptibility to colonization and $\epsilon = 1$ in that case. Upon immunization, not all individuals respond to all included serotypes, and even in that case, protection against infection is not complete. Therefore, effective vaccination is modeled as preventing a fraction $1 - \delta$ of infections. Thus, colonized individuals who have been vaccinated, E_V , progress to an infected state at rate $\delta\kappa E_V$. We assume that the duration of a given infection in vaccinated and unvaccinated individuals is roughly the same. However, the infection of vaccinated individuals are not assumed to indicate any loss of protection, and therefore, these individuals return to the S_V class upon recovery. A complete list of variables can be found in Table 1, and the model equations are given by

$$\frac{dS}{dt} = \lambda - \beta S \frac{E + E_V + I + I_V}{N} + \alpha E + \gamma I - \phi S + \rho S_V - \mu S \quad (2)$$

$$\frac{dE}{dt} = \beta S \frac{E + E_V + I + I_V}{N} - \alpha E - l\kappa(t)E - \phi E + \rho E_V - \mu E \quad (3)$$

$$\frac{dS_V}{dt} = \phi S - \epsilon \beta S_V \frac{E + E_V + I + I_V}{N} + \alpha E_V + \gamma I_V - \rho S_V - \mu S_V \quad (4)$$

$$\frac{dE_V}{dt} = \epsilon \beta S_V \frac{E + E_V + I + I_V}{N} - \alpha E_V + \phi E - \rho E_V - \delta \kappa(t)E_V - \mu E_V \quad (5)$$

$$\frac{dI}{dt} = l\kappa(t)E - (\gamma + \eta + \mu)I \quad (6)$$

$$\frac{dI}{dt} = \delta \kappa(t)E_V - (\gamma + \eta + \mu)I_V. \quad (7)$$

3 Methodology: Parameter Estimation

Here we describe the formulation of the inverse problem used to calibrate the model to the Australian population and then to assess the impact of widespread vaccination. We briefly describe the data set itself and the model quantities represented by the data in Section 3.1. Fixing certain parameters and initial conditions, as described in Sections 3.3 and 3.4, respectively, we use the data reported during the period January 2002 to December 2004 to estimate unknown model parameters, $\theta = (\beta, \kappa_0, \kappa_1, \delta)^T$. The estimation of these parameters via a weighted least squares algorithm is described in Section 3.2. The calibrated model along with data collected after January 2005 is then used to estimate vaccine parameters ϵ or δ .

3.1 Description of Data

The number of cases of invasive pneumococcal diseases (clinically defined here as pneumonia, meningitis and bacteremia, or the presence of *S. pneumoniae* in normally sterile fluid) by month dating back to January 2001 are available on the Australian NNDS website [34]. However, it has been noted in annual reports that there were some reporting discrepancies in the first year of collection [32], which were resolved in subsequent years. Therefore, all of our parameter estimates are based on data collected beginning in January 2002, although information from December 2001 is used to calculate initial conditions (in our model $t = 0$ corresponds to December 1, 2001).

Annual reports [25, 26, 27, 24] on this collected data are published in the Australian journal *Communicable Diseases Intelligence*. The reports contain vaccination information if collected, and also case fatality estimates for the year. From the case fatality data we directly calculate the disease-induced mortality rate (η) as described in Section 3.3. The vaccination information from 2002, 2003, and 2004 provides another six data points with information on the efficacy of the implemented vaccine program. There is some slight

underreporting evident when the summaries in the annual reports are compared to the same data published on the NNDS website. Because of this discrepancy, proportions were used to extrapolate to the total number of cases in the data set for each year. This is described in further detail in the Appendix. While this is of course not exact, we are likely not committing errors of a magnitude that would influence the estimates of the parameters significantly. Moreover, the data is not assumed to be error-free and the extent to which this affects our estimates is taken into consideration in the error analysis described in Section 3.2.

Monthly case notifications $d_j^{(1)}$ are best represented as integrals of the new infection rates (including those in the vaccinated class) over each month, since they represent the number of cases reported during the month and do not provide any information on how long individuals remain in an infected state. Similarly, the annual observations $d_k^{(2)}, d_k^{(3)}$ indicate the cases recorded during the year occurring in either vaccinated or unvaccinated individuals. These data are integrals of the infection rate into the vaccinated or unvaccinated classes over the year. Thus, the observations and the model quantities are related by

- $d_j^{(1)} \sim \int_{t_j}^{t_{j+1}} [\kappa(s)E(s) + \delta\kappa(s)E_V(s)] ds$ for $j = 1, 2, \dots, 60$ (monthly cases [34]),
- $d_k^{(2)} \sim \int_{t_k}^{t_{k+1}} \kappa(s)E(s)ds$ for $k = 1, 2, 3$ (yearly unvaccinated cases [24, 25, 26]),
- $d_k^{(3)} \sim \int_{t_k}^{t_{k+1}} \delta\kappa(s)E_V(s)ds$ for $k = 1, 2, 3$ (yearly vaccinated cases [24, 25, 26]).

Let us define $Y^{(1)}(t_j, \theta)$, $Y^{(2)}(t_k, \theta)$ and $Y^{(3)}(t_k, \theta)$ to represent the integrals above, so that $d_j^{(1)} \sim Y^{(1)}(t_j, \theta)$, $d_k^{(2)} \sim Y^{(2)}(t_k, \theta)$, and $d_k^{(3)} \sim Y^{(3)}(t_k, \theta)$.

The data collected during the period January 2002 and December 2004 are used to estimate the parameters, β , κ_0 , κ_1 , and δ , and to directly calculate η . The monthly case notification data beginning in January 2005 are then used to examine the effects of increased conjugate vaccination on the population, by determining values of ϵ and δ for fixed vaccination rates (ϕ) that could have generated the number of cases observed.

3.2 Least Squares Estimation of Parameters

To begin, we use $n = 42$ data points ($n_1 = 36$ monthly cases and $n_2 = 6$ annual vaccinated or unvaccinated cases) to estimate four unknown parameters: $\theta = (\beta, \kappa_0, \kappa_1, \delta)^T$ ($\theta \in \Theta \subset \mathbb{R}^p$ where $p = 4$ and Θ is the feasible parameter space). As described above we have a vector system of three ‘types’ of longitudinal data. We use a statistical model to represent these observations, made up of the deterministic model and error accounting for discrepancies between the model predictions and the data. These include errors incurred when the measurements are reported, along with any stochastic behavior not accounted for by the deterministic model. We assume there exists a ‘true’ set θ_0 of parameters which generated the data, and our statistical model is then given by

$$d_j^{(1)} \equiv Y^{(1)}(t_j, \theta_0) + \epsilon_j^{(1)} \quad j = 1, \dots, 60. \quad (8)$$

$$d_k^{(2)} \equiv Y^{(2)}(t_k, \theta_0) + \epsilon_k^{(2)} \quad k = 1, 2, 3. \quad (9)$$

$$d_k^{(3)} \equiv Y^{(3)}(t_k, \theta_0) + \epsilon_k^{(3)} \quad k = 1, 2, 3. \quad (10)$$

The errors $(\epsilon_j^{(i)})$ in (8) - (10) for $i = 1, 2, 3$ in the above model are assumed to be random variables with means $E[\epsilon_j^{(i)}] = 0$ and constant variances $var(\epsilon_j^{(i)}) = \sigma_{0,i}^2$, where $\sigma_{0,i}$ is unknown. Here we have assumed that the size of the errors committed at each time for a given kind of “measurement” is constant and also does not depend on the magnitude of the measurement itself. We also assume that $\epsilon_j^{(i)}$ are independent and identically distributed (*i.i.d.*) random variables for each fixed i . Thus since each $d_j^{(i)}$ depends on the random variable $\epsilon_j^{(i)}$, it is a random variable itself with mean $E[d_j^{(i)}] = Y^{(i)}(t_j, \theta_0)$ and variance $var(d_j^{(i)}) = \sigma_{0,i}^2$, with the $d_j^{(i)}$ *i.i.d.* for fixed i . We further assume that $var(\epsilon_j^{(2)}) = var(\epsilon_j^{(3)}) = \sigma_{0,2}^2 = \sigma_{0,3}^2$ since these data arose from the same counting process.

We seek the set of parameters $\hat{\theta}$ that minimizes

$$J_{42}(\theta, \sigma_1^2, \sigma_2^2) = \frac{w_1}{\sigma_1^2} \sum_{j=1}^{n_1} \left| d_j^{(1)} - Y_j^{(1)} \right|^2 + \frac{w_2}{\sigma_2^2} \sum_{k=1}^{n_2/2} \left\{ \left| d_k^{(2)} - Y_k^{(2)} \right|^2 + \left| d_k^{(3)} - Y_k^{(3)} \right|^2 \right\}, \quad (11)$$

thereby minimizing the weighted differences between the model quantities and data. We remind the reader here that we have assumed $\sigma_2 = \sigma_3$. In the above cost function, $J_{42}(\theta, \sigma_1^2, \sigma_2^2)$, each residual has been weighted by the variance in the observations as well as a scaling factor to give equal importance to each type of data ($w_1 = 1$ and $w_2 = 9$), giving rise to a weighted least squares (WLS) estimator. The estimator $\hat{\theta}_{WLS}$ is also a random variable then, and its distribution is called the *sampling distribution*. This sampling distribution provides information about the uncertainty in the estimates $\hat{\theta}$ of the parameters obtained for a specific data set, which is actually one realization of $d^{(1)}, d^{(2)}, d^{(3)}$ where $d^{(1)} = \{d_j^{(1)}\}_{j=1}^{36}$, $d^{(2)} = \{d_k^{(2)}\}_{k=1}^3$, and $d^{(3)} = \{d_k^{(3)}\}_{k=1}^3$. For a summary of formulation involving ordinary and weighted least squares along with asymptotic distribution results as discussed below, see [2].

The parameters to be estimated are $\psi = (\theta, \sigma_1^2, \sigma_2^2)^T$ since the estimation of θ depends on the unknowns σ_1 and σ_2 . We employ an iterative approach, explained in the algorithm:

1. Guess initial estimates for $\hat{\sigma}_1^{(0)}, \hat{\sigma}_2^{(0)}$ where the superscript indexes the iteration.
2. Minimize the objective functional (11) for $\hat{\theta}^{(0)}$; $\hat{\theta}^{(0)} = \arg \min_{\theta \in \Theta} J_{42}(\theta, \hat{\sigma}_1^{(0)}, \hat{\sigma}_2^{(0)})$.

3. Calculate $\hat{\sigma}_1^{(1)}, \hat{\sigma}_2^{(1)}$ by the standard formulas

$$\hat{\sigma}_1^{(1)} = \left(\frac{1}{n_1 - p} \sum_{j=1}^{n_1} \left| d_j^{(1)} - Y_j^{(1)} \right|^2 \right)^{\frac{1}{2}}$$

$$\hat{\sigma}_2^{(1)} = \left(\frac{1}{n_2 - p} \left\{ \sum_{k=1}^{n_2/2} \left| d_k^{(2)} - Y_k^{(2)} \right|^2 + \sum_{k=1}^{n_2/2} \left| d_k^{(3)} - Y_k^{(3)} \right|^2 \right\} \right)^{\frac{1}{2}}.$$

4. Repeat steps 2 and 3 until $\left| \|\hat{\psi}^{(k)}\| - \|\hat{\psi}^{(k-1)}\| \right| \leq 10^{-q}$ where $\hat{\psi} = (\hat{\theta}, \hat{\sigma}_1, \hat{\sigma}_2)^T$ and q is the resolution desired for convergence.

Under reasonable smoothness and regularity assumptions, the standard nonlinear regression approximation theory for asymptotic (as $n \rightarrow \infty$) distributions can be invoked (outlined in [5, 7, 14] and Chapter 12 of [28]). Smoothness requirements are easily checked for this example and regularity requirements involve, among others, conditions on the manner in which observations are taken as sample size increases. This theory yields that the sampling distribution for $\hat{\theta}_{WLS}$ is approximately a p -multivariate Gaussian with mean $E[\hat{\theta}_{WLS}(d^{(1)}, d^{(2)}, d^{(3)})] = \theta_0$ and covariance matrix $cov[\hat{\theta}_{WLS}(d^{(1)}, d^{(2)}, d^{(3)})] = \Sigma_0 = \left[\sum_{i=1,2,3} \frac{1}{\sigma_{0,i}^2} \chi_i^T(\theta_0) \chi_i(\theta_0) \right]^{-1}$. Since $\theta_0, \sigma_{0,1}, \sigma_{0,2}, \sigma_{0,3}$ are unknown we must approximate them in Σ_0 and we obtain

$$\Sigma_0 \approx \hat{\Sigma} = \left[\sum_{i=1,2,3} \frac{1}{\hat{\sigma}_i^2} \chi_i^T(\hat{\theta}) \chi_i(\hat{\theta}) \right]^{-1}.$$

Here $\chi(\hat{\theta}) = \frac{\partial Y}{\partial \theta}$ is the $n \times p$ sensitivity matrix where Y is a column vector of model representation of the data given by $Y = (Y^{(1)}, Y^{(2)}, Y^{(3)})^T$. The sensitivity matrix is then defined by

$$\chi_i(\hat{\theta}) = \left[\frac{\partial C_i}{\partial \theta} \frac{\partial X}{\partial \theta} \right]_{\theta=\hat{\theta}} = \left[\frac{\partial C_i}{\partial \theta} z(\theta) \right]_{\theta=\hat{\theta}},$$

with the quantities C , X , and Z defined as follows. Let C_i be the linear operator such that $Y^{(i)} = C_i X$, X denote the state variables of the model, $X = (S, E, S_V, E_V, I, I_V)^T$, and let $G(X(t), t)$ denote the model equations so that $\frac{dX}{dt} = G(X(t), t)$. Then z is defined by $z = \frac{\partial X}{\partial \theta}$ and satisfies the sensitivity equations

$$\frac{dz}{dt} = \frac{\partial G}{\partial X} z + \frac{\partial G}{\partial \theta}. \quad (12)$$

Note that z depends on $\dot{X} = G(X(t), t)$, a nonlinear system of ODEs, and its solution X , and is therefore not available in closed form. We thus use the numerical solution of both X and z in the estimation of $\hat{\Sigma}$.

The standard errors for $\hat{\theta} = (\hat{\beta}, \hat{\kappa}_0, \hat{\kappa}_1, \hat{\delta})^T$ can be computed from the diagonal entries of the covariance matrix, $SE(\hat{\theta}_k) = \sqrt{\hat{\Sigma}_{kk}}$. These standard errors do not reveal how close the estimates are to the ‘true’ values for the parameters. But rather, they are an indicator of the reliability of the estimation procedure we have used to obtain them. Small standard errors can provide confidence in the values obtained for the parameter estimate, but do not indicate that these values are ‘correct’ in any way. On the other hand, if the standard errors are relatively large, our ability to estimate that parameter accurately is suspect. The standard error in that case would not suggest reasonable confidence in the corresponding estimate.

Once the model has been calibrated to the Australian population, i.e., parameter estimates found for $\hat{\beta}, \hat{\kappa}_0, \hat{\kappa}_1, \hat{\delta}$ using the data from 2002-2004, we proceed by estimating vaccination parameters that would quantify the impact of the widespread vaccination program beginning in 2005. To this end, we use the monthly case notification data from January 2005 through December 2005 to estimate $\theta = \epsilon$ or $\theta = \delta$. The algorithm is simpler in the sense that we are now using one type of data and therefore can use an ordinary least squares (OLS) approach to find the parameters $\hat{\theta} = \hat{\epsilon}$ or $\hat{\theta} = \hat{\delta}$ that minimizes the objective function

$$J_{12}(\hat{\theta}) = \sum_{j=37}^{60} \left| d_j^{(1)} - Y_j^{(1)} \right|^2.$$

Since this expression does not involve σ_4 , an iterative process is not required here and we can first determine the values of the parameter estimates before using the standard formula to arrive at an estimate for the observational variance

$$\hat{\sigma}_4^2 = \frac{1}{n_3 - p} \sum_{j=37}^{60} \left| d_j^{(1)} - Y_j^{(1)} \right|^2,$$

where $n_3 = 12$ and now $p = 1$.

We can apply the asymptotic theory of distributions again and see that the sampling distribution of $\hat{\theta}_{OLS}(d^{(1)})$ is approximately a Gaussian with mean $E[\hat{\theta}(d^{(1)})] = \theta_0$ ($\theta_0 = \epsilon_0$ or $\theta_0 = \delta_0$) and covariance matrix $cov[\hat{\theta}(d^{(1)})] = \Sigma_0 = \sigma_{0,4}^2 [\chi^T(\theta_0)\chi(\theta_0)]^{-1}$. Again we estimate Σ_0 by $\hat{\Sigma}(\hat{\theta}, \hat{\sigma}_4)$. Further computations are analogous to those described above.

3.3 Parameters Calculated from Literature

Many of the model parameters are available from various sources of scientific literature. In this section we briefly discuss which parameters can be reasonably estimated from reported values. Details, if omitted here, of these calculations along with their explanation can be

Parameter	Value	Description	Sources
α	1.3555	colonized recovery rate [1/month]	[3, 38]
ρ	0.02741	loss of vaccine-induced protection rate [1/month]	[1, 38, 25]
ω	$\frac{\pi}{6}$	infection rate oscillation frequency	[34]
τ	8	infection rate oscillation phase shift	[34]
l	0.05	fraction of susceptibles ‘at risk’	N/A
ϕ	0.0007579	effective vaccination rate [1/month]	[25, 38]
λ	25,000	recruitment/birth rate [people/month]	[37]
μ	0.0003	natural death rate [1/month]	[36]
ϵ	0.953 or [0.94,1]	reduction of colonization events due to vaccination	[4, 8, 17, 18, 38]
γ	$[\frac{1}{2}, 2]$	infection recovery rate [1/month]	N/A
η	0.0787	disease-induced mortality rate [1/month]	[25, 26, 24]

Table 2: Estimates of “Known” Model Parameters.

found in the Appendix. The information here was gathered from primary scientific research papers, *S. pneumoniae* reviews, or from the website of the Australian Bureau of Statistics. All resources used for the calculation of each parameter are cited next to their values in Table 2

Parameters α , ρ , ω , and τ are readily calculated from literature. The length of asymptomatic colonization, or carriage, has been described for children and adults in a review by Bridy-Pappas et al. [3]. This information together with the age-sex profile of the Australian population [38] was used to calculate a population average for α . The age-sex profile was also used to calculate the population average for the rate of loss of protection from vaccination, which occurs about three years after PPV23 vaccination of adults over 65 years of age.

The proportion of the population who is considered to have a higher risk of progressing to infection is not easily quantified, so we have chosen a value that seems reasonable. The magnitude of this parameter seems to primarily affect the value for the vaccine efficacy, δ , with smaller values of l corresponding to smaller values of δ . Thus, we have chosen $l = 0.05$ simply because this value consistently results in $0 < \delta < 1$. Therefore, our estimates of these parameters are not going to be absolute, but rather allows for the discussion of

relative changes in the values of $\hat{\delta}$.

Two parameters governing the oscillatory behavior of the infection rate were calculated directly from observing the infection patterns in the monthly data. It was noted that the infections peak annually, and thus the frequency $\omega = \frac{2\pi}{6}$, and that they peak in July, so the phase shift is then $\tau = 8$.

The rate of vaccination can be estimated by a description of the recommended target groups and an estimate of the actual vaccine coverage attained report in [25] and [27]. The first of these reports contains estimates for coverage as of 2001, and the second discusses coverage as of 2004 and estimated increases during the year 2005. These numbers reported do not account for the individuals who do not respond to the vaccine, and therefore this is not an *effective* vaccination rate. However, in vaccine efficacy studies, this occurs in less than 5% of those who have received either of the vaccines considered here. Therefore we do not distinguish between an effective or practical vaccination rate.

Initial estimates for the parameters governing the overall population growth can be obtained from the Australian Bureau of Statistics website. They have reported there the annual number of births and deaths for the years 1995-2005 along with the estimated resident population estimates. However, these parameters do not produce solutions that agree well with the population data as discussed further in the Appendix, and are therefore not used in the estimation of model parameters.

The reduction in colonization events has been measured in previous vaccination trials and in other populations after the implementation of the PCV7 vaccine. These estimates were used to calculate an estimate for this parameter ϵ during the period of January 2002 - December 2004. However, in populations where the use of the conjugate vaccine has been widespread, significant reductions in colonization have been observed, even in unvaccinated persons. While this parameter in our model literally quantifies the direct reduction in vaccinated persons, a reduction in this value when the model is fitted to the set of data beginning in January 2005 would indicate that either the protection from colonization is greater than thought, or herd immunity could be occurring. On the other hand, a large value of ϵ would indicate that serotype replacement has occurred to such an extent as to increase the number of colonized individuals. This particular event would be disconcerting as it would implicate the increased threat of serotype replacement in invasive infections, a threat that the scientific and medical communities are aware of, although no documentation has been found to support or dispell this fear.

In the literature, there is minimal information concerning the recovery rates for pneumococcal infections, and certainly no mention of population averages for this parameter. However, we can reason intuitively that it is not common to recover from these serious infections in under 2 weeks. Similarly, we can guess that it is probably rare to remain ill for more than 2 months. Thus, a broad reasonable range for the recovery rate is $\frac{1}{2} \leq \gamma \leq 2$. Since the data we have obtained contains virtually no information about this parameter (discussed in Section 4) we numerically explore estimates of the other model parameters for fixed values of γ in this range in Section 5.

The disease-induced mortality rate can be calculated directly from the annual reports of

case fatalities under the assumption that the recovery rate, $\gamma \gg 1/\text{year}$. This assumption would allow us to justify the approximation $\int_j^{j+12} \kappa(s)(E(s) + \delta E_V(s))ds \approx \int_j^{j+12} I(s)ds$. Under this assumption η is calculated as

$$\eta = \frac{1}{3} * \left(\sum_{j=1,13,25} \frac{\int_{t=j}^{j+12} \eta I(s)ds}{\int_j^{j+12} I(s)ds} \right) \approx 0.0787. \quad (13)$$

3.4 Calculation of Initial Conditions

Initial conditions for the model are a snapshot of the Australian population in December 2001, $N(t_0) = S(t_0) + E(t_0) + S_V(t_0) + E_V(t_0) + I(t_0) = 19,529,274$ [35]. The estimates discussed here are by no means exact and therefore sensitivity of estimated model parameters to these initial conditions are later considered in Section 5. The number of cases of IPD reported for December 2001 is 38 [34], but the vaccination status of these cases are unknown, and we initially take $I(t_0) = 33$ and $I_V(t_0) = 5$.

Many studies have estimated carriage prevalence rates in various populations, especially in light of the effects that conjugate vaccination may have on this aspect of pneumococcal diseases. Here we use studies that have focused on the Australian population (both aboriginal and non-aboriginal) in an effort to quantify the proportion of the population that will be colonized at any given time. In a review of the relevant literature, we have found carriage rates in Australian aboriginal children to be 89% ([10]), 90% ([29]), and 49% ([31]), which results in an average of 76%. Non-aboriginal children in Australia were found to be colonized at rates of 29% ([11]), 43% ([29]), and 25% ([31]), or an average of 32.3%. Nasopharyngeal colonization is not as commonly quantified in adults, so the data here is more sparse. Carriage rates in aboriginal adults are around 34% ([10]), and in a study of UK households, the adult population there (assumed to have characteristics similar to Australian non-aboriginal adults) had about an 8% colonization prevalence ([13]). These numbers, in conjunction with the age and sex profiles presented in [38] and [39], allow for the estimation of the colonization prevalence, which we determine to be around 10-15%.

The population for whom pneumococcal vaccination is recommended consists of adults including and above the age of 65 years, and indigenous adults over 50 years of age. Indigenous children 0-5 years of age in Central Australia and all indigenous children 0-2 years of age were also recommended to be vaccinated. The vaccine coverage as of 2002 is discussed in [25], and while it seems to be well quantified in adults ($\approx 26.5\%$), the coverage attained during the year 2001 among targeted children is not mentioned. To begin, we assume that also 26.5% of the children have been vaccinated. Later in the model calibration section, we will discuss any errors that this could have introduced into our estimates for model parameters by using initial conditions where 0 and 100% coverage of targeted children is assumed. By this reasoning, we arrive at the estimates for the model initial conditions as shown in Table 3.

Variable	Individuals
$S(t_0)$	16,576,528
$E(t_0)$	2,453,326
$S_V(t_0)$	435,002
$E_V(t_0)$	64,380
$I(t_0)$	33
$I_V(t_0)$	5

Table 3: Initial conditions of model state variables.

4 Algorithm Testing

As routine good practice *before* use with experimental data, we explore the strength and weaknesses of the algorithm with simulated “data” from model quantities which we believe the surveillance data represents. Through this examination we are able to test the convergence of the parameter estimates $\hat{\theta}$ to the known values θ_0 . In this way we are able to identify which parameters we can estimate from our data set. Further, we are able to explore how the reliability of the algorithm, and hence that of the estimates, may change as error is introduced in the “observations”. This testing may also suggest a poor interpretation of the surveillance data if the algorithm gives drastically different results when using the observed data as compared to the generated data. Thus this procedure provides us with a means to avoid common pitfalls during the estimation process.

We generate a set of data $Y^{(1)}, Y^{(2)}, Y^{(3)}$ that corresponds to what we have on hand, described in Section 3.1 with a fixed set of parameters $\psi_0 = (\theta_0, \sigma_{0.1}, \sigma_{0.2})^T$. We initiate the iterative weighted least squares estimation process (Section 3.2) with an initial guess $\hat{\psi}$ that is slightly perturbed from the ψ_0 . If the estimation procedure is working well, the $\hat{\psi}$ should converge to the ‘true’ set rather quickly. We look to the standard errors as one indication of the ability of the algorithm to estimate that particular parameter using the data set.

In reality, no data set is free of observation error, and we would like to see how well our algorithm performs when the data contains some noise. To this end, we add error by sampling from a normal distribution with mean 0 and constant variance ($\epsilon_j^{(i)} \sim \mathcal{N}(0, \sigma_i^2)$) as per the assumptions of our statistical model (equations (8) - (10)). The magnitude of this variance determines how much noise is added. We can expect that 99% of the time, numbers generated from this distribution will be within the interval $[-3\sigma_i, 3\sigma_i]$. Thus we take

$$d_j^{(i)} \sim Y_j^{(i)} + \epsilon_j^{(i)}$$

where $\epsilon_j^{(i)} \sim \sigma_i * \mathcal{N}(0, 1) = \mathcal{N}(0, \sigma_i^2)$. We add noise that is scaled to our observations, but with variance that is constant with respect to time. For $k\%$ noise then, we take

ψ	ψ_0	initial $\hat{\psi}$	$\hat{\psi}_{(0)}$	$SE(\hat{\theta}_{(0)})$	$\hat{\psi}_{(1)}$	$SE(\hat{\theta}_{(1)})$
β	1.5	1.4	1.5	$2e^{-9}$	1.4984	$1e^{-3}$
κ_0	$1.4e^{-3}$	$1.35e^{-3}$	$1.4e^{-3}$	$1.6e^{-11}$	$1.4113e^{-3}$	$8.4e^{-6}$
κ_1	0.55	0.6	0.55	$4e^{-9}$	0.55219	$2e^{-3}$
δ	0.5	0.45	0.5	$1.5e^{-8}$	0.49296	$6e^{-3}$
$\sigma_{1,(0)}$	0	$1e^{-4}$	$2.527e^{-6}$			
$\sigma_{2,(0)}$	0	1	$2.368e^{-5}$			
$\sigma_{1,(1)}$	1.6788	2			1.4305	
$\sigma_{2,(1)}$	10.0727	13			9.8923	
ψ	ψ_0	initial $\hat{\psi}$	$\hat{\psi}_{(5)}$	$SE(\hat{\theta}_{(5)})$	$\hat{\psi}_{(10)}$	$SE(\hat{\theta}_{(10)})$
β	1.5	1.4	1.4974	$6e^{-3}$	1.5268	0.02
κ_0	$1.4e^{-3}$	$1.35e^{-3}$	$1.3885e^{-3}$	$5e^{-5}$	$1.2438e^{-3}$	$1e^{-4}$
κ_1	0.55	0.6	0.5223	0.01	0.52670	0.03
δ	0.5	0.45	0.5468	0.04	0.46690	0.1
$\sigma_{1,(5)}$	8.3939	10	8.5024			
$\sigma_{2,(5)}$	50.3635	60	66.1831			
$\sigma_{1,(10)}$	16.7878	20			16.830	
$\sigma_{2,(10)}$	100.727	120			190.196	

Table 4: Algorithm testing for parameter estimation of $\theta = (\beta, \kappa_0, \kappa_1, \delta)^T$. Model was fit to generated data with $k = 0, 1, 5, 10\%$ noise where subscripts in $\psi_{(k)}$ denote then the level of noise in the simulated data.

$\sigma_i = \frac{k}{100} * \text{average}_j(d_j^{(i)})$, so k scales the constant variance relative to the magnitude of the observations.

The results from the testing of the algorithm's ability to estimate the parameters $\hat{\psi} = (\hat{\beta}, \hat{\kappa}_0, \hat{\kappa}_1, \hat{\delta}, \hat{\sigma}_1, \hat{\sigma}_2)^T$ with $k = 0, 1, 5, 10\%$ noise are displayed in Table 4 and Figure 2. The estimation procedure appears to be reliable for all parameters since the values for the parameters are close to their true value and the standard errors $SE(\hat{\theta}_k)$ are considerably less than their corresponding parameter value $\hat{\theta}_k$. Initially, the standard error of all parameters are many orders of magnitude less than their parameter value, but the standard errors increase with added noise, indicating that our ability may depend on the amount of observational error in the data. Still the standard errors are sufficiently small to suggest confidence in parameter estimates.

As discussed in Section 3.3, there is relatively little information on recovery rates from or duration of invasive pneumococcal diseases, so our ability to estimate this rate (γ) is of interest. However, when the parameters to be estimated are expanded to include γ , its estimated value does not coverage to its true value (shown in Table 5). Although the fit to clean data is not discouraging, we see that with only 1% noise the standard error $SE(\hat{\gamma})$ for the recovery rate increases dramatically, and we cannot justify estimating this

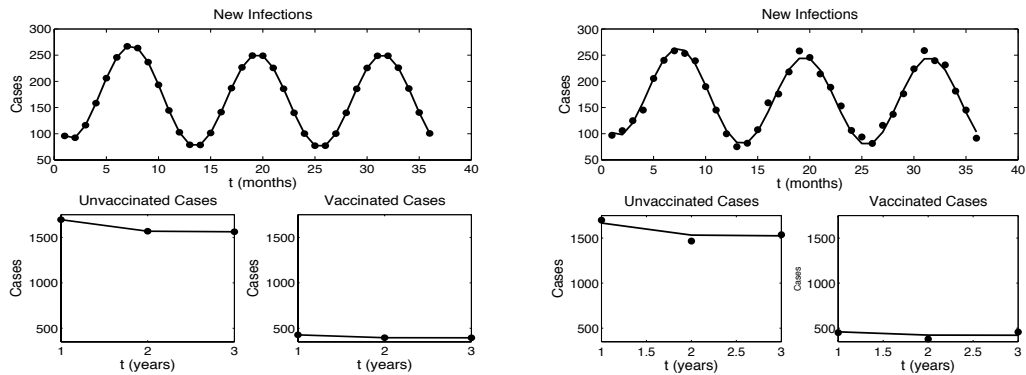


Figure 2: Best fit model solutions to generated data with 0% noise (left) and 5% noise (right). On each side, the top panels represent the monthly cases of IPD, and the bottom shows the annual data with the unvaccinated cases on the left side and vaccinated cases shown on the right.

parameter. While we cannot reliably determine a value for this parameter with data of the type used in this paper, it is nevertheless important to investigate any effect large changes in this value could have on our results. This is done in Section 5.

The lack of information concerning the parameter l , the proportion of the susceptible population which is particularly likely to develop infections, motivates us to attempt to estimate this parameter with our data set. The results of this are shown in Table 6, and suggests that the data available does not enable us to reliably estimate this parameter since the estimated value $\hat{\delta}$, is significantly different from the true value δ_0 , and the standard errors of both δ and l are quite large. Notably, the parameter most affected is δ , suggesting that it is reasonable to fix l to a value that restricts δ to a reasonable range.

ψ	ψ_0	initial $\hat{\psi}$	$\hat{\psi}_{(0)}$	$SE(\hat{\theta}_{(0)})$	$\hat{\psi}_{(1)}$	$SE(\hat{\theta}_{(1)})$
β	1.5	1.4	1.5	$5e^{-6}$	1.4984	0.004
κ_0	$1.4e^{-3}$	$1.35e^{-3}$	$1.4046e^{-3}$	$2e^{-7}$	$1.4106e^{-3}$	$4e^{-5}$
κ_1	0.55	0.6	0.54999	$6e^{-6}$	0.55203	0.006
δ	0.5	0.45	0.49181	$3e^{-4}$	0.49309	0.02
γ	1	1.2	1.1646	0.05	1.1487	39
$\sigma_{1,(0)}$	0	$1e^{-4}$	0.00163			
$\sigma_{2,(0)}$	0	1	13.048			
$\sigma_{1,(1)}$	2.1547	2			1.45003	
$\sigma_{2,(1)}$	12.928	13			14.0705	

Table 5: Algorithm test for the estimation of $\theta = (\beta, \kappa_0, \kappa_1, \delta, \gamma)^T$ from 0 and 1% noisy data.

ψ	ψ_0	initial $\hat{\psi}$	$\hat{\psi}_{(0)}$	$SE(\hat{\theta}_{(0)})$
β	1.5	1.4	1.5	$5e^{-6}$
κ_0	$1.4e^{-3}$	$1.35e^{-4}$	$1.2908e^{-3}$	$8e^{-4}$
κ_1	0.55	0.6	0.54999	$3e^{-6}$
δ	0.5	0.45	0.51065	0.3
l	0.05	0.052	0.05503	0.03
$\sigma_{1,(0)}$	0	$1e^{-4}$	0.0015966	
$\sigma_{2,(0)}$	0	1	57.9842	

Table 6: Algorithm test for the ability to estimate $\theta = (\beta, \kappa_0, \kappa_1, \delta l)^T$

ψ	$\hat{\psi}$	$SE(\hat{\theta})$
β	1.52175	0.02
κ_0	1.3656^{-3}	$1.3e^{-4}$
κ_1	0.56444	0.04
δ	0.7197	0.06
σ_1	28.924	
σ_2	86.386	

Table 7: Model calibration to Australian IPD data from 2002-2004; estimation of $\hat{\psi} = (\hat{\beta}, \hat{\kappa}_0, \hat{\kappa}_1, \hat{\delta}, \hat{\sigma}_1, \hat{\sigma}_2)^T$.

5 Model Calibration

Using the iterative weighted least squares procedure described in Section 3.2, we find a best fit solution to the Australian IPD data (Figure 3) and estimates for the model parameters, shown in Table 7. The model solution and data agree well, and parameter values are on the scale of our initial guesses, although their values differ slightly to minimize the cost function in the functional (11). Further, the standard errors are relatively small and suggests that the estimates obtained here are reliable (as we would have expected based on the results in Section 4). These parameters are displayed several times throughout this section for ease of comparison and will be displayed in **boldface**.

The data fits in Figure 3 reveal that the model solution with the parameters shown in Table 7 fits the Australian surveillance data from 2002-2004, with the top panel showing the fit to the monthly case notification data, the bottom left panel the unvaccinated cases reported annually, and the bottom right the annual vaccinated cases. The demographic parameters here are $\lambda = 25,000$ and $\mu = 0.0003$ instead of those reported on the Australian Bureau of Statistics website, since those parameters seemed to underestimate the resident population of Australia as discussed in further detail in the Appendix. For the sake of consistency of the model to the Australian population, we use these values throughout the rest of the paper. The initial conditions and vaccination rate used to generate this solution are calculated under the assumption of 26.5% vaccine coverage of both the adult and child target groups. This is fairly accurate for the adults, but no data is available on the coverage for the younger target group at that time. We next explore how the calibrated parameter values might change in the event that vaccination rate and initial conditions have been miscalculated.

The effects of changes in the vaccination rate ϕ on the estimated parameters are given in Table 8. When ϕ is decreased by about 15% there are no notable changes in the values from the original value of ϕ (displayed in boldface), and only δ changes by about 80% if ϕ is increased by an order of magnitude. Since we are not attempting to estimate δ accurately and are instead looking for relative changes in this value based on data before

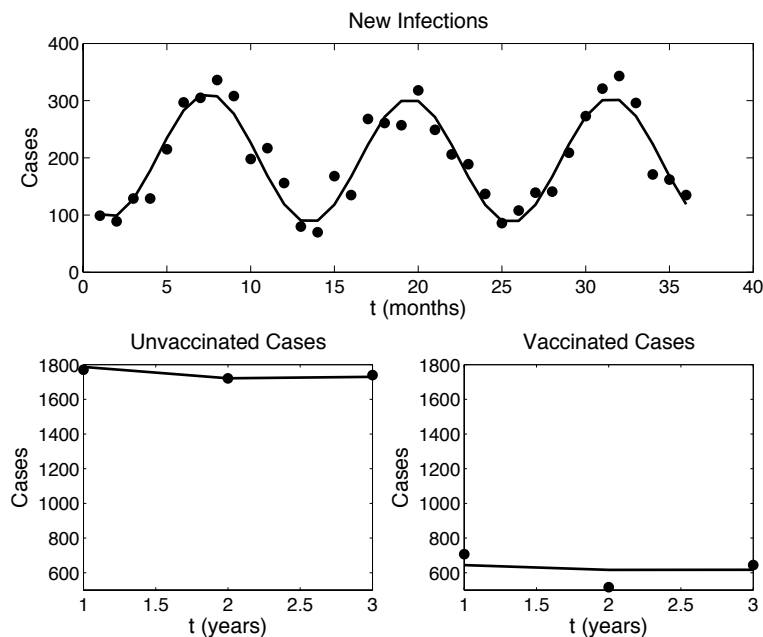


Figure 3: Best fit solution to Australian IPD data with parameters shown in Table 7. Top panel: monthly cases; bottom left panel: annual unvaccinated cases; bottom right panel: annual vaccinated cases.

and after 2005, there is no cause for concern. Additionally, it is not likely that we have erred this greatly in the calculation of the vaccination rate. In the annual report of 2005, it is stated that 51% of adults over the age of 65 had been vaccinated, which would be reflected by a vaccination rate of $\phi \approx 0.001$. However, the model and annual data with this vaccination rate no longer agreed, even in their general trends, as is evident by the large calculated observational variance σ_2 . Thus we conclude that the current value of the vaccination rate, for this set of initial conditions, appears adequate.

In the event that our assumption that 26.5% of targeted children to be vaccinated is incorrect, the initial conditions along with the vaccination rate would be erroneous. In Table 9 we investigate the effect that this could possibly have on the estimated values of the parameters. Results of all three inverse problem calculations appear to be essentially identical with only minor differences in the parameter values. These results suggest that possible errors in the calculation of ϕ and the initial conditions are inconsequential in the estimation of the desired parameters.

Another possible source of error in our calibration may be the recovery rate, γ , on which there is relatively little information. However, we are able to reasonably deduce a range over which this parameter might vary. To investigate the effects of large changes in γ on our parameters we proceed by fixing $\gamma = \frac{1}{2}$ and $\gamma = 2$. The corresponding results

ψ	$\hat{\psi}(\phi = 6e^{-4})$	$SE(\hat{\theta})$	$\hat{\psi}(\phi \approx 7e^{-4})$	$SE(\hat{\theta})$
β	1.525	0.02	1.52175	0.02
κ_0	$1.341e^{-3}$	$1.3e^{-4}$	$1.366e^{-3}$	$1.3e^{-4}$
κ_1	0.5649	0.04	0.56444	0.04
δ	0.7760	0.06	0.71957	0.06
σ_1	28.967		28.924	
σ_2	81.939		86.386	
ψ	$\hat{\psi}(\phi \approx 0.001)$	$SE(\hat{\theta})$	$\hat{\psi}(\phi = 0.007)$	$SE(\hat{\theta})$
β	1.5217	0.02	1.5144	0.03
κ_0	$1.4439e^{-3}$	$1.5e^{-4}$	$1.5966e^{-3}$	$1.8e^{-4}$
κ_1	0.5638	0.04	0.5633	0.04
δ	0.5295	0.06	0.1482	0.04
σ_1	29.0914		29.5498	
σ_2	134.98		289.1339	

Table 8: Effect of vaccination rate ϕ on calibrated model parameters $\hat{\psi} = (\hat{\beta}, \hat{\kappa}_0, \hat{\kappa}_1, \hat{\delta}, \hat{\sigma}_1, \hat{\sigma}_2)^T$.

ψ	$\hat{\psi}_{(0)}$	$SE(\hat{\theta}_{(0)})$	$\hat{\psi}_{(26.5)}$	$SE(\hat{\theta}_{(26.5)})$	$\hat{\psi}_{(75)}$	$SE(\hat{\theta}_{(75)})$
β	1.5217	0.02	1.52175	0.02	1.5218	0.02
κ_0	$1.3650e^{-3}$	$1.3e^{-4}$	$1.3656e^{-3}$	$1.3e^{-4}$	$1.3667e^{-3}$	$1.3e^{-4}$
κ_1	0.56444	0.04	0.56444	0.04	0.56444	0.04
δ	0.73269	0.06	0.71957	0.06	0.69663	0.06
σ_1	28.924		28.924		28.924	
σ_2	86.3906		86.386		86.386	

Table 9: Effect of varied vaccine coverage of children on the estimation of parameters $\hat{\psi} = (\hat{\beta}, \hat{\kappa}_0, \hat{\kappa}_1, \hat{\delta}, \hat{\sigma}_1, \hat{\sigma}_2)^T$. The subscripts denote the percent vaccination coverage of targeted children used in the calculation of the ICs and ϕ .

ψ	$\hat{\psi} (\gamma = \frac{1}{2})$	$SE(\hat{\theta})$	$\hat{\psi} (\gamma = 1)$	$SE(\hat{\theta})$	$\hat{\psi} (\gamma = 2)$	$SE(\hat{\theta})$
β	1.52159	0.02	1.52175	0.02	1.5218	0.02
κ_0	1.3660^{-3}	$1.3e^{-4}$	$1.3656e^{-3}$	$1.3e^{-4}$	$1.3655e^{-3}$	$1.3e^{-4}$
κ_1	0.56467	0.04	0.56444	0.04	0.56444	0.04
δ	0.71965	0.06	0.71957	0.06	0.71951	0.06
σ_1	28.924		28.924		28.924	
σ_2	86.377		86.386		86.388	

Table 10: Model parameters using Australian IPD data from 2002-2004 where the value of the recovery rate γ is shown in parentheses.

are displayed in Table 10. Again, the estimates of all parameters change minimally and we conclude that even large changes in the recovery rate γ do not affect our estimates.

The vaccination status of some of the cases were classified as ‘partial’ for some individuals who had not received the recommended doses for their age group. Recall that this is not relevant for elderly receiving the polysaccharide vaccine, due to this being given as a single dose. In the previous fits ‘partial’ vaccinated cases were considered fully vaccinated, resulting in an overestimation of the value for δ . Parameters were estimated from data in which ‘partial’ vaccinated cases were considered unvaccinated. This is an underestimation of the cases due to vaccine failure and therefore also an underestimation of δ (and an overestimation of the vaccine efficacy $1 - \delta$). The parameters from both of these fits are displayed in Table 11.

With the exception of that for δ , the estimates for all of the parameters change very little. This agrees with our intuition since the only difference in the data sets used to estimate these parameters is the annual data, which contains mostly information on δ and relatively little on the other parameters. Thus, this numerical experiment has provided us with a range of where a feasible value for δ during the time period 2002-2004 likely lies, i.e., $\delta \in [0.35, 0.72]$.

To test the assumptions of the statistical model that we have chosen to represent our data, we plot the residuals between the model and observations as a function of the model, that is, $|d_j^{(1)} - Y_j^{(1)}|$ vs $Y_j^{(1)}$ (Figure 4). The lack of a clear relationship between these two quantities would indicate that our assumptions are reasonable and the errors of each observation does not depend on the model values. However, we see six groups of points, which can be explained by the oscillatory pattern of the infections. In the top panel we have plotted just one half of the period of the infection rate and see a completely random pattern, indicating no relationship among these quantities. When we extend this time period for another half of a period, thus plotting an entire period in the middle panel, we see that there are two points in each group of points. Thus, the pattern observed is driven by the seasonality of the infections and not by any incorrect assumptions. On the contrary, only a pattern in the dependent variable (the residuals) would suggest that

ψ	$\hat{\psi}$ (partial \equiv full)	$SE(\hat{\theta})$	$\hat{\psi}$ (partial \equiv unvacc)	$SE(\hat{\theta})$
β	1.52175	0.02	1.5268	0.02
κ_0	1.3656^{-3}	$1.3e^{-4}$	$1.5442e^{-3}$	$1.7e^{-4}$
κ_1	0.56444	0.04	0.56221	0.04
δ	0.71957	0.06	0.35145	0.07
σ_1	28.924		28.822	
σ_2	86.386		132.753	

Table 11: Model parameters using Australian IPD data from 2002-2004. The second column contains parameter estimates from data with partial vaccinated cases considered as unvaccinated, and the fourth column considers partial vaccination equivalent to full vaccinated.

β	1.52475
κ_0	$1.4549e^{-3}$
κ_1	0.56332
δ	0.53551

Table 12: Model parameters calibrated to Australian IPD surveillance data during 2002-2004.

incorrect assumptions have been made. This analysis suggests that it is reasonable to assume constant variance among observations of the same type, providing support for the statistical model underlying the parameter estimation procedure.

A similar plot of the residuals from the best fit solution to data in which partial vaccination is considered equivalent to unvaccination is not shown here. The difference between these two data sets lies in the annual data, and so the monthly data is unchanged. Although the vaccine efficacy parameter δ is drastically different when estimated from each data set, this does not appear to affect the model fit to monthly data, as can be seen from Figure 5. In fact, when the difference between the two monthly solutions are plotted the solution is in the range of $[-1.5, 2]$ and varies around zero.

In this section, we have tested various sources of error as well as assumptions of our statistical model used in the estimation. We have confirmed that the estimation of these parameters for the purposes of calibrating this model to the Australian invasive pneumococcal disease data has been successful. The calibrated parameters are then taken as an average of the two sets displayed in Table 11 and are shown here in Table 12.

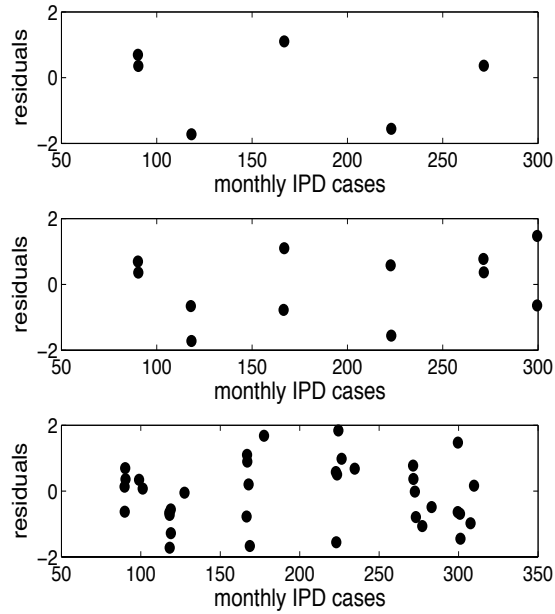


Figure 4: Residuals as a function of model variables. Top panel is over the period January 2003 through June 2003, middle panel is for January 2003 through December 2003, and bottom panel is for all three years

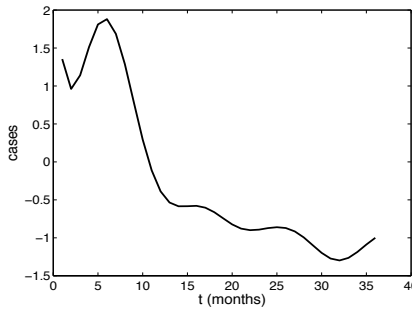


Figure 5: Difference in monthly data fits where partial vaccinated cases were considered vaccinated subtracted from fits to data with partially vaccinated considered as unvaccinated.

6 Effect of Increased Vaccination

As of January 1, 2005, pneumococcal vaccination was offered free of charge to all adults over the age of 65, indigenous adults over 50, and children under the age of 2. Free

ϕ	0.002253	0.007984	0.0139651
$\hat{\epsilon}$	0.116	0.176	0.206
$SE(\hat{\epsilon})$	0.3	0.4	0.4
Unvacc	1630	1470	1327
Vacc	113	287	442

Table 13: Estimation of ϵ , the direct reduction of colonization due to vaccination, from 2005 Australian IPD surveillance data. The observed unvaccinated and vaccinated cases are 1211 and 537, respectively. The three vaccination rates used are based on 0, 75, 100% increases in vaccinated adults.

vaccination of adults of these ages had been offered in some provinces such as Victoria since 1998, and for some children, but it was not until 2005 that vaccination was funded nationally. In this section, we use the model with parameters estimated from the 2002-2004 data together with the 2005-2006 data to investigate the impact of the vaccination program on the Australian population.

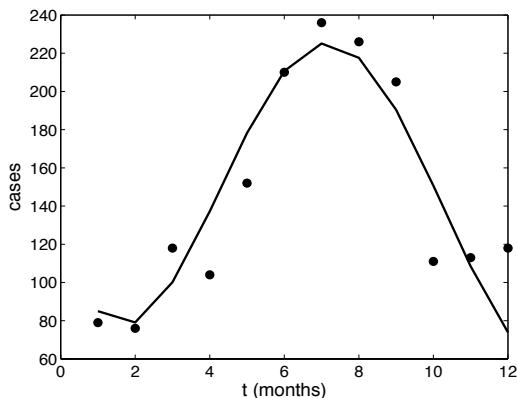


Figure 6: Best fit solution to 2005 Australian surveillance data assuming a vaccination rate of $\phi = 0.007984$.

We seek to find vaccination parameters that minimize the difference between the model and observations via least squares estimation (see Section 3.2). Thus, we are looking for values of ϵ and δ that could have resulted in the observed number of cases. We do not estimate the vaccination rate ϕ since there is sufficient information for us to determine a realistic range for this parameter in the annual report concerning the data collected during the year 2005 ([27]). It is reported here that 91% of children under the age of 2 were vaccinated during the year 2005. Neglecting the loss of protection during that time period, especially since this is unobserved in children and most children were previously

	$l = 0.05$	$l = 0.04$		$l = 0.05$	$l = 0.04$
$\hat{\epsilon}$	0.176	0.353	$\hat{\delta}$	$3.22e^{-14}$	0.1202
$SE(\hat{\epsilon})$	0.4	0.96	$SE(\hat{\delta})$	2	2.3
Unvacc	1470	1231	Unvacc	1771	1417
Vacc	287	491	Vacc	$7.58e^{-11}$	283

Table 14: Estimation of $\hat{\epsilon}$ and $\hat{\delta}$ with $l = 0.05$ and $l = 0.04$.

unvaccinated, we calculate a vaccination rate of children. The proportion of adults 65 and over who had been vaccinated as of 2004 is reported here as 51% and we calculate a range of ϕ that would have resulted from zero to 100% vaccinated. Additionally, the number of cases that occurred in vaccinated or unvaccinated individuals is contained in the 2005 annual report (again, this information was reported in 70% of cases and we have applied these proportions to characterize the remaining cases in the data set). While this only represents one annual data point and hence is not appropriate for least squares estimation, it does provide additional information against which we can verify the validity of our fitted solutions.

We first consider changes in ϵ that could have resulted in the observed decreases in infections during the year 2005 with estimates displayed in Table 13 and a sample fit shown in Figure 6. Initial conditions for these fits have been taken as state variable quantities when the model has been run with calibrated parameters to the time step corresponding to January 1, 2005. All estimates for ϵ are drastically lower (by about 90%) than the feasible lower bound for this parameter, which is 0.93875. This parameter is strictly interpreted here as representing the direct reduction in colonization in vaccinated individuals, an effect which has only been observed in protein conjugate vaccination. A substantially lower value, like those estimated here, could suggest herd immunity, an effect in which the vaccination of a portion of the population provides protection for the unvaccinated portion as well. In this case, herd immunity could be interpreted as a decrease in infections in unvaccinated individuals who are colonized at a lower rate due to the decrease in colonized vaccinated individuals. This suggests that the unvaccinated cases would be overestimated by our model and the vaccinated cases underestimated if herd immunity is occurring which is precisely what occurs across all three vaccination rates used. The standard errors are notably large, and do not provide additional confidence in our estimates. However, the annual report did note that there were reduced infections in unvaccinated groups, and thus, it is likely that herd immunity is occurring. However, other changes in the population could have occurred which would have also produced a reduction in infections.

The PCV7 vaccine is thought to be much more effective in children than the PPV23 is in the individuals for whom it is recommended, and the large increase of PCV7 vaccine could have resulted in improved protection of vaccinated individuals, notable at even the population level. With this in mind, we have held $\epsilon = 0.93875$ constant and estimated

$\hat{\delta}$ as $\hat{\delta} = 3.22e^{-14}$. The standard error $SE(\hat{\delta}) = 2$, suggests that we will not be able to get a reliable estimate for δ from this data with the parameters used. Also, the number of vaccinated individuals estimated was $7.58e^{-11}$, which is, of course, unrealistic. We conclude that it is not likely that a sufficiently large decrease in δ could have accounted for the reduction of infections observed in 2005.

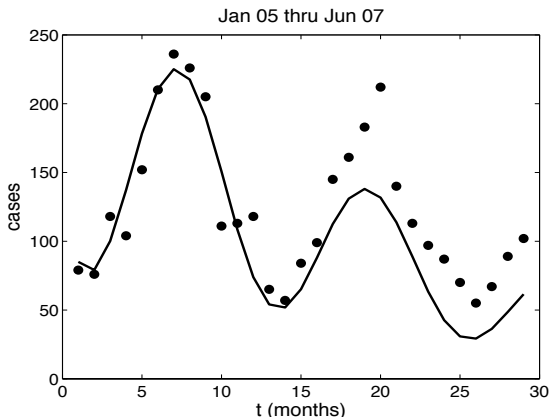


Figure 7: Best fit solution from 2005 Australian surveillance data run out to June 2007.

With the increased vaccination of recommended groups it is likely that the parameter, l , would decrease. This parameter, which scales for the relative risk of the individuals in the unvaccinated population to that of the vaccinated population, is not easily quantified and we were unable to calculate any reasonable estimate of it. We do not attempt to estimate this parameter here, instead, we note the effect of a 20% decrease in l on estimates of ϵ and δ (Table 14). Most notable in this experiment is the substantial change observed in the estimate of δ , which could indicate that a smaller herd immunity effect is occurring and also the population vaccine efficacy has changed when PCV7 is implemented. This indicates that some measurement of the relative risk to pneumococcal infections is of interest to those wishing to understand the effects of the vaccination programs.

Another aspect of immunization by PCV7 is that the long-term effects on a population are essentially unknown, so the effects of the new vaccination program should be evaluated as time progresses. If we overlay a best fit solution of the model with the data up to June 2007, we see a general underestimation of cases after 2005 (Figure 7), with the effect exaggerated in 2007 as compared with 2006. The data suggests that the landscape of pneumococcal disease dynamics is changing on a yearly basis under this vaccination program. If the changes could be explained via vaccination rates alone, we would likely see an overestimation of the data by the model, since the vaccination rate, ϕ , has likely decreased after the first year of the federally funded program. Since the opposite is observed, it appears that the values of ϵ or δ (or possibly both) have increased over these time periods. However, we are not able to estimate these parameters reliably with the

available data, as their standard errors are an order of magnitude larger than their values and thus are not shown here.

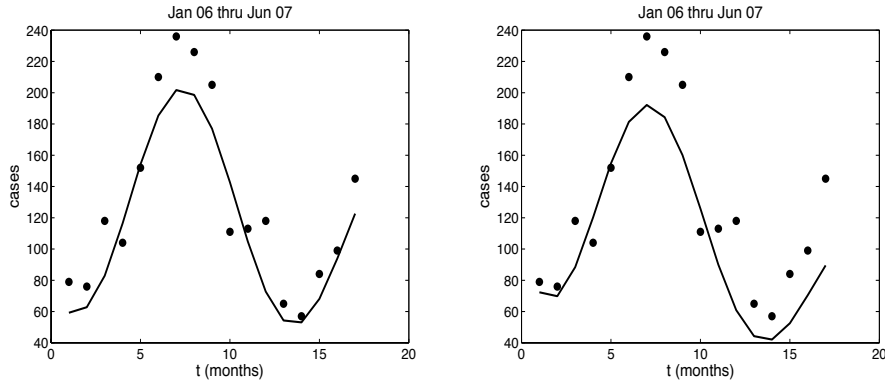


Figure 8: Model overlayed with Australian surveillance data during January 2006 thru June 2007. Left: $\epsilon = 0.7$, $\delta = 0.53551$; Right: $\epsilon = 0.353$, $\delta = 1$.

Although we are not able to reliably estimate ϵ or δ over the 2006-2007 time period, we see by inspection better agreement between the model and data when ϵ has been increased from 0.353 to 0.7 (Figure 8 left) or δ has been increased from 0.53551 to 1 (Figure 8 right). Here, the cases are still underestimated considerably, and it is likely that even greater increases in these parameters would have produced the data. Increases in ϵ could indicate serotype replacement by non-vaccine serotypes, a possibility which authorities are aware of, but no conclusive evidence has yet been found. Similarly, an increase in δ would indicate that the vaccines are, in general, less effective in the population than in previous times. Again, this would suggest either an evolutionary change in the pneumococcus, or the host response to the vaccines. In either case, a reevaluation of the vaccination programs would be necessary. These results are presented not to support or dispute specific changes associated with the Australian population or the pneumococci prevalent there, but rather, that changes appear to be occurring at a rapid pace. These studies suggest further close and quantitative study of the vaccination programs.

7 Discussion and Concluding Remarks

In this work, we have calibrated a mathematical model describing the dynamics of pneumococcal infections in the Australian population. We have estimated via a weighted least squares procedure pertinent model parameters of interest, but not typically measurable or reported in literature. All standard errors for these parameters provide confidence in the values obtained, even when initial assumptions have been relaxed as discussed in Section 5. These values, however, depend on the relative risk of unvaccinated to vaccinated indi-

viduals, and is not available in literature. However, the parameter values reported here allow for the study of changes in the effects of the immunization programs before and after 2005.

Once pneumococcal vaccination was made publicly available to large recommended risk groups, a notable decrease in infections was observed. Our analysis suggests that this reduction is not solely accounted for by an increase in vaccination, but also a decrease in vaccinated colonized individuals, and even appears to suggest the presence of herd immunity in 2005. The potential impact of this reduction of colonization due to PCV7 is very promising from a public health perspective, but serotype replacement has been observed in some populations in the colonization stage, and its effects are uncertain. It is possible that widespread and long-term vaccination by PCV7 could provide a selective pressure for the increased invasiveness of non-vaccine-included pneumococcal serotypes.

We have shown that vaccination parameters which accurately describe the infections for 2005 underestimate the cases for 2006 and 2007. In contrast, changing parameters to reduce the effect of the vaccines provides better agreement between the model and observed data. These results do not conclusively support any one effect generating this behavior but clearly suggest that changes are occurring within the Australian population, such as susceptibility to infection, and possibly the evolution of the pneumococci themselves endemic in that area. At any rate, the analysis here suggests that the high rate of PCV7 immunization will likely result in a drastic change of the landscape of pneumococcal disease dynamics, and requires close monitoring and quantitative study by public health officials.

8 Acknowledgements

The authors would like to express thanks to Alun Lloyd and Eunha Shim for helpful discussions and insight. This research was supported by the National Institute of Allergy and Infectious Diseases under grant 9R01AI071915-05, and a scholarship distributed by the Achievement Rewards for College Scientists Foundation which has been generously donated by the Ralph Matteucci Endowment.

References

- [1] A. S. Artz, W. B. Ershler, and D. L. Longo. Pneumococcal vaccination and revaccination of older adults. *Clinical Microbiology Reviews*, 16(2):308–318, Apr 2003.
- [2] H. T. Banks, M. Davidian, and J. Samuels Jr. An inverse problem statistical methodology summary. Technical Report CRSC-TR07-14, Center for Research in Scientific Computation, North Carolina State University, August 2007.
- [3] A. Bridy-Pappas, M. Margolis, K. Center, and Isaacman. Streptococcus pneumoniae: description of the pathogen, disease epidemiology, treatment, and prevention. *Pharmacotherapy*, 25(9):1193–1212, 2005.

- [4] R. Cohen, C. Levy, F. de La Rocque, N. Gelbert, A. Wollner, B. Fritzell, R. Bonnet, E. Tetelboum, and E. Varon. Impact of pneumococcal conjugate vaccine and of reduction of antibiotic use on nasopharyngeal carriage of nonsusceptible pneumococci in children with acute otitis media. *Pediatric Infectious Diseases Journal*, 25:1001–1007, 2006.
- [5] M. Davidian and D. M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall/CRC, Boca Raton, 1995.
- [6] S. F. Dowell, C. G. Whitney, C. Wright, C. E. Rose Jr., and A. Schuchat. Seasonal patterns of invasive pneumococcal disease. *Emerging Infectious Diseases*, 9(5):573–579, Mar 2003.
- [7] A. R. Gallant. *Nonlinear Statistical Models. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*,. John Wiley & Sons, Inc., New York, 1987.
- [8] F. Ghaffar, T. Barton, J. Lozano, L. S. Muniz, P. Hicks, V. Gan, N. Ahmad, and G. H. McCracken Jr. Effect of the 7-valent pneumococcal conjugate vaccine on nasopharyngeal colonization by *Streptococcus pneumoniae* in the first 2 years of life. *Clinical Infectious Diseases*, 39:930–938, 2004.
- [9] M. A. Gordon, A. L. Walsh, M. Chaponda, S. D., M. Mbwini, M. E. Molyneux, and S. B. Gordon. Bacteraemia and mortality among adult medical admissions in Malawi - predominance of non-typhi Salmonellae and *Streptococcus pneumoniae*. *Journal of Infection*, 42:44–49, 2001.
- [10] D. Hansman, S. Morris, M. Gregory, and B. McDonald. Pneumococcal carriage amongst Australian aborigines in Alice Springs, Northern Territory. *Journal of Hygiene*, 95(3):677–684, Dec 1985.
- [11] D. Hansman and S. Morris. Pneumococcal carriage amongst children in Adelaide, South Australia. *Epidemiology and Infection*, 101(2):411–417, Oct 1988.
- [12] J. O. Hendley, F. G. Hayden, and B. Winther. Weekly point prevalence of *Streptococcus pneumoniae*, *Hemophilus influenzae* and *Moraxella catarrhalis* in the upper airways of normal young children: effect of respiratory illness and season. *acta pathologica, microbiologica, et immunologica Scandinavica*, 113(3):213–220, Mar 2005.
- [13] M. Hussain, A. Melegaro, R. G. Pebody, R. George, W. J. Edmunds, R. Talukdar, S. A. Martin, A. Efstratiou, and E. Miller. A longitudinal household study of *Streptococcus pneumoniae* nasopharyngeal carriage in a UK setting. *Epidemiology and Infection*, 133:891–898, 2005.
- [14] R. I. Jennrich. Asymptotic properties of non-linear least squares estimators. *Ann Math Statist*, 40:633–643, 1969.

- [15] P. Kalima, F. X. Emmanuel, and T. Riordan. Epidemiology of *Streptococcus pneumoniae* infections at the Edinburgh City Hospital: 1980-95. *Epidemiology and Infection*, 122(2):251–257, Apr 1999.
- [16] P. Marchisio, S. Gironi, S. Esposito, G. C. Schito, S. Mannelli, N. Principi, and A. P. C. Group. Seasonal variations in nasopharyngeal carriage of respiratory pathogens in healthy Italian children attending day-care centres or schools. *Journal of Medical Microbiology*, 50:1095–1099, 2001.
- [17] E. V. Millar, K. L. O’Brien, J. P. Watt, M. A. Bronsdon, J. Dallas, C. G. Whitney, R. Reid, and M. Santosham. Effect of community-wide conjugate pneumococcal vaccine use in infancy on nasopharyngeal carriage through 3 years of age: a cross-sectional study in a high-risk population. *Clinical Infectious Diseases*, 43:8–15, 2006.
- [18] M. R. Moore, T. B. Hyde, T. W. Hennessy, D. J. Parks, A. L. Reasonover, M. Harker-Jones, J. Gove, D. L. Bruden, K. Rudolph, A. Parkinson, J. C. Butler, and A. Schuchat. Impact of a conjugate vaccine on community-wide carriage of nonsusceptible *Streptococcus pneumoniae* in Alaska. *Journal of Infectious Diseases*, 190:2031–2038, 2004.
- [19] D. M. Musher. *Streptococcus pneumoniae*. In G. E. Mandell, J. E. Bennett, and R. Dolin, editors, *Mandell, Douglas, and Bennett’s principles and practice of infectious diseases*, pages 2128–2144. Churchill Livingstone, Philadelphia, PA, 5th edition, 2000.
- [20] M. J. Owen, C. D. Baldwin, P. R. Swank, A. K. Pannu, D. L. Johnson, and V. M. Howie. Relation of infant feeding practices, cigarette smoking exposure, and group child care to the onset and duration of otitis media with effusion in the first two years of life. *Journal of Pediatrics*, 123:702–711, 1993.
- [21] J. L. Paradise, H. E. Rockette, D. K. Colborn, B. S. Bernard, C. G. Smith, M. Kurs-Laskey, and J. E. Janosky. Otitis media in 2253 Pittsburgh-area infants: prevalence and risk factors during the first two years of life. *Pediatrics*, 99(3):318–333, 1997.
- [22] G. Peter and J. O. Klein. *Streptococcus pneumoniae*. In S. S. Long, L. K. Pickering, and C. G. Prober, editors, *Principles and practice of pediatric infectious diseases*, pages 739–746. Churchill Livingstone, Philadelphia, PA, 2nd edition, 2002.
- [23] K. A. Robinson, W. Baughman, G. Rothrock, N. L. Barrett, M. Pass, C. Lexau, B. Damaske, K. Stefonek, B. Barnes, J. Patterson, E. R. Zell, A. Schuchat, and C. G. Whitney. Epidemiology of invasive *Streptococcus pneumoniae* infections in the United States, 1995-1998: Opportunities for prevention in the conjugate vaccine era. *Journal of American Medical Association*, 285(13):1729–1735, 2001.

- [24] P. W. Roche, V. L. Krause, M. Bartlett, D. Coleman, H. Cook, C. Davis, J. E. Fielding, R. Holland, C. Giele, R. Gilmour, and R. Kampen. Invasive pneumococcal disease in Australia, 2004. *Communicable Diseases Intelligence*, 30(1):80–92, 2006.
- [25] P. Roche, V. Krause, R. Andrews, L. Carter, D. Coleman, H. Cook, M. Counahan, C. Giele, R. Gilmore, S. Hart, and R. Pugh. Invasive pneumococcal disease in Australia, 2002. *Communicable Diseases Intelligence*, 27(4):466–477, 2003.
- [26] P. Roche, V. Krause, M. Bartlett, D. Coleman, H. Cook, M. Counahan, C. Davis, L. Del Fabbro, C. Geile, R. Gilmore, R. Kampen, and M. Young. Invasive pneumococcal disease in Australia, 2003. *Communicable Diseases Intelligence*, 28(4):441–454, 2004.
- [27] P. Roche, V. Krause, and H. Cook. Invasive pneumococcal disease in Australia, 2005. *Communicable Diseases Intelligence*, 31:86–100, 2007.
- [28] G. A. F. Seber and C. J. Wild. *Nonlinear Regression*. John Wiley & Sons, Inc., New York, 1989.
- [29] E. Stubbs, K. Hare, C. Wilson, P. Morris, and A. J. Leach. *Streptococcus pneumoniae* and noncapsular *Haemophilus influenzae* nasal carriage and hand contamination in children: A comparison of two populations at risk of otitis media. *Pediatric Infectious Disease Journal*, 24(5):423–428, 2005.
- [30] T. R. Talbot, K. A. Poehling, T. V. Hartert, P. G. Arbogast, N. B. Halasa, K. M. Edwards, W. Schaffner, A. S. Craig, and M. R. Griffin. Seasonality of invasive pneumococcal disease: Temporal relation to documented influenza and respiratory syncytial viral circulation. *American Journal of Medicine*, 118:285–291, 2005.
- [31] K. Watson, K. Carville, J. Bowman, P. Jacoby, T. V. Riley, A. J. Leach, and D. Lehmann. Upper respiratory tract bacterial carriage in aboriginal and non-aboriginal children in a semi-arid area of western Australia. *Pediatric Infectious Disease Journal*, 25(9):782–790, Sep 2006.
- [32] M. Watson, K. Bayley, J. M. Bell, G. W. Gilbert, G. Hogg, A. D. Keil, V. Krause, D. Murphy, P. Roche, H. V. Smith, M. G. Stewart, J. Stylianopoulos, and J. Turnidge. Laboratory surveillance of invasive pneumococcal disease in Australia in 2001 to 2002-implications for vaccine serotype coverage. *Communicable Diseases Intelligence*, 27(4):478–487, 2003.
- [33] WHO. Pneumococcal vaccines. *WHO Weekly Epidemiological Record*, 74(23):177–183., 1999.
- [34] National Notifiable Diseases Surveillance System. Communicable Diseases Australia, <http://www9.health.gov.au/cda/Source/CDA-index.cfm>.

- [35] Australian Bureau of Statistics. Australian Demographic Statistics: Table 4. Estimated Resident Population, States and Territories (Number), <http://www.abs.gov.au/ausstats/abs@.nsf/0e5fa1cc95cd093c4a2568110007852b/6949409dc8b8fb92ca256bc60001b3d1!OpenDocument>.
- [36] Australian Bureau of Statistics. Deaths, Australia, 2005, <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/C67A858BA00CB846CA2568A9001393C6>.
- [37] Australian Bureau of Statistics. Births, Australia, 2005, <http://www.abs.gov.au/AUSSTATS/abs@.nsf/ProductsbyCatalogue/FF9E15176D6887D8CA2568A9001393B2?OpenDocument>.
- [38] Australian Government: Department of Health and Ageing. Health and Ageing Factbook 2006. Chapter 1; Section: Ageing Population., <http://www.health.gov.au/internet/wcms/publishing.nsf/content/factbook2006-1~factbook2006-ch1-introduction~chapter1-sect-3>.
- [39] Australian Government: Department of Health. Health and Ageing Factbook 2006. Chapter 6; Section: Age Profile, <http://www.health.gov.au/internet/wcms/publishing.nsf/Content/Factbook2006-1~factbook2006-ch6-introduction~chapter6-sect-2>.

Appendix

Data

In the annual reports of the pneumococcal surveillance data there is a slight record-keeping discrepancy that is evident when comparing the annual number of total infections reported on the NNDS website and in the papers. There is no explanation for this in the reports, and thus we have no reason to assume that there are known errors in the monthly data presented online. Here we describe the manner in which the vaccinated/unvaccinated cases from the annual reports were converted for use in the least squares estimation.

The total number of cases for which vaccination information is reported (tot_v), along with the number of cases considered fully vaccinated ($full$), the cases which occurred in partially vaccinated individuals ($part$) and those which occurred in unvaccinated individuals (un). Unless otherwise noted, partially vaccinated cases were considered equivalent to fully vaccinated cases, and the vaccination data was then calculated by the formulas

$$\begin{aligned}
 Unvacc &= \frac{un}{tot_v} * total \\
 Vacc &= \frac{part + full}{tot_v} * total
 \end{aligned}
 \tag{14}$$

where $total$ refers to the total annual cases reported on the NNDS website.

Parameter Calculation

In this section we describe in detail the calculation of the parameters which are not estimated in the paper.

- α : The average length of carriage is 4 months in infants (up to 24 months old), and 2-4 weeks in adults [3]. The age-sex profile of the Australian population was found in the Health and Ageing Factbook 2006 [38], with children ages 0-4 comprising 6.2% of the population. Since the length of time in this compartment adversely affects (increases) the number of colonization events, an overestimate of this parameter would result in an overestimate in colonized individuals and thus, infections. Also, it would likely be more reasonable that there is some continuity in the colonization period as a function of age, so while this average overestimates the period for children 2-4 yrs old, it likely underestimates the period for children 5-10 yrs old. Therefore, it is reasonable to estimate α as $\alpha = \frac{1}{4} * 0.062 + \frac{1}{0.7} * 0.938 = 1.3555$.
- ρ : Loss of protection from vaccination only occurs in adults ≥ 65 , and antibody levels reach pre-vaccination levels in ≈ 3 yrs (36 months) [1]. The proportion of the elderly individuals, both non-indigenous and indigenous, in the target group of vaccinated individuals is $\frac{1}{2,548,936} * (40,112 + 2,475,623)$ based on demographic information as of 2001, found in [38]. ρ is then $\rho = \frac{1}{36} * \frac{(40,112+2,475,623)}{2,548,936} = 0.02741$. For a description of the target population for vaccination see the discussion for the calculation of the initial conditions in Section 3.4.
- ϕ : The vaccination rate is estimated by assuming that the vaccinated and unvaccinated portions of the population have reached a quasi-steady state in that no notable exchange is occurring between the two kinds of classes. That is, no significant surge of vaccination or loss of protection change is reported in the annual summaries or observed in the annual data. If the system is completely free of infection, the steady state, or disease-free equilibrium is given by

$$\begin{pmatrix} S^0 \\ E^0 \\ S_V^0 \\ E_V^0 \\ I^0 \\ I_V^0 \end{pmatrix} = \begin{pmatrix} \frac{\lambda}{\mu} \frac{\rho+\mu}{\phi+\rho+\mu} \\ 0 \\ \frac{\lambda}{\mu} \frac{\phi}{\phi+\rho+\mu} \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The proportions of the population in S^0 or S_V^0 were calculated based on the vaccine coverage estimates used in calculating the initial conditions, and we then solved for the vaccination rate ϕ that would give those proportions. In this way, we have

calculated vaccination rates corresponding to 0, 26.5 and 100% vaccine coverage of children shown below with the percentage of children vaccinated denoted in superscripts

$$\begin{aligned}\phi^{(0)} &= 7.206e^{-4} \\ \phi^{(26.5)} &= 7.3375e^{-4} \\ \phi^{(100)} &= 7.579e^{-4}.\end{aligned}\tag{15}$$

When the model is run with these initial conditions and vaccination rates, very little transfer between the vaccinated and unvaccinated classes is noticeable. This is by no means accurate and in the model calibration we have checked the effects of miscalculating the initial conditions, as well as various vaccination rates for a fixed set of initial conditions. In both cases, it appears that our initial guesses for initial conditions and vaccination rates are sufficient.

- λ : The number of births per year for the years 1995-2005 is found on the Australian Bureau of Statistics website at [37]. An estimate for λ can be obtained by averaging the annual births for Australia over the years 2002-2004 and dividing this by 12. Thus, $\lambda = 21,011$ births/month.
- μ : The number of deaths per year for the years 1995-2005 is found on the Australian Bureau of Statistics website [36], along with the annual estimated resident population. The IPD case fatalities were subtracted from the deaths for the years 2002-2004, and this average was then divided by the average resident population for those years. Thus,

$$\mu = \frac{\text{deaths} - \text{IPD_deaths}}{\text{total_popn}} \approx 0.000551.$$

- ϵ : Reduced susceptibility to colonization has been observed in children vaccinated by PCV7 in numerous studies. Three such studies' results showed vaccinated individuals susceptibilities being reduced to 80.2% [4], 86% [8], and 60.2% (average $\approx 75.5\%$) as compared to unvaccinated children. Two studies, however, showed no significant decrease in carriage in two high-risk populations ([17], [18]). Notably, Australian aborigines are considered to be a very high-risk population, with some of the largest prevalence rates observed anywhere in the world. If we assume the carriage in aboriginal children could be reduced as well, and that children will make up at most 25% of the vaccinated population, then we would arrive at a lower bound for ϵ as $\epsilon \approx 0.25 * 0.755 + 0.75 = 0.93875$. Thus a feasible range for ϵ is $\epsilon \in [0.93875, 1]$. We hold $\epsilon = 1$ fixed for the model calibration since the portion of vaccinated individuals who received PCV7 during 2002-2004 was small.

Population Data

Estimates of the resident population of Australia is collected quarterly, available on the website of the Australian Bureau of Statistics [35]. We use the estimate in December 2001 in the calculation of the initial conditions, and the estimates reported for March 2002 through June 2006 to refine our estimates of λ and μ from those calculated in Section 3.3. These data are then of the form

- $d_l^{(4)} \sim N(t_l) = S(t_l) + E(t_l) + S_V(t_l) + E_V(t_l) + I(t_l)$ for $l = 1, \dots, 12$ and $(t_1, \dots, t_{12}) = (3, 6, \dots, 36)$ (quarterly population estimates [35]).

The demographic parameters, λ and μ , are estimated from the population data only, represented by the statistical model below (16). Again, we assume that $\epsilon_l^{(4)}$ are *i.i.d.* for $l = 1, \dots, 12$ and $E[\epsilon_l^{(4)}] = 0$, $var(\epsilon_l^{(4)}) = \sigma_4^2$ for each l .

$$d_l^{(4)} = Y^{(4)}(t_l, \theta) + \epsilon_l^{(4)} = N(t_l) + \epsilon_l^{(4)} \quad l = 1, \dots, 12. \quad (16)$$

Since this is a scalar system, we can use an ordinary least squares approach to estimate the parameters $(\theta = (\lambda, \mu)^T)$ by minimizing the objective functional in (17)

$$J_{12}(\theta, \sigma_4) = \sum_{l=1}^{12} \left| d_l^{(4)} - Y_l^{(4)} \right|^2 \quad (17)$$

and the standard deviation of the observational errors can be computed directly from the formula

$$\hat{\sigma}_4^2 = \frac{1}{n_3 - p} \sum_{l=1}^{n_3} \left| d_l^{(4)} - Y_l^{(4)} \right|^2$$

where $n_3 = 12$ and $p = 2$.

The standard errors for $\hat{\theta} = (\hat{\lambda}, \hat{\mu})^T$ are computed analogously to the vector case where the covariance matrix is defined as

$$\hat{\Sigma} = \hat{\sigma}_4^2 \left[\chi_4^T(\hat{\theta}) \chi_4(\hat{\theta}) \right]^{-1}.$$

Table 15 contains the results of fixing the demographic parameters, λ and μ , or estimating them via OLS. The first two columns show the residual when the model is compared with the data and the parameters are fixed at the displayed values. In the other columns, standard errors are reported for parameters which have been estimated, and otherwise the parameter is fixed at the displayed value. When we estimate both parameters, they quickly reach the constraints placed on them, regardless of how large the restraints are set. This suggests that we may not be able to estimate both of these parameters simultaneously.

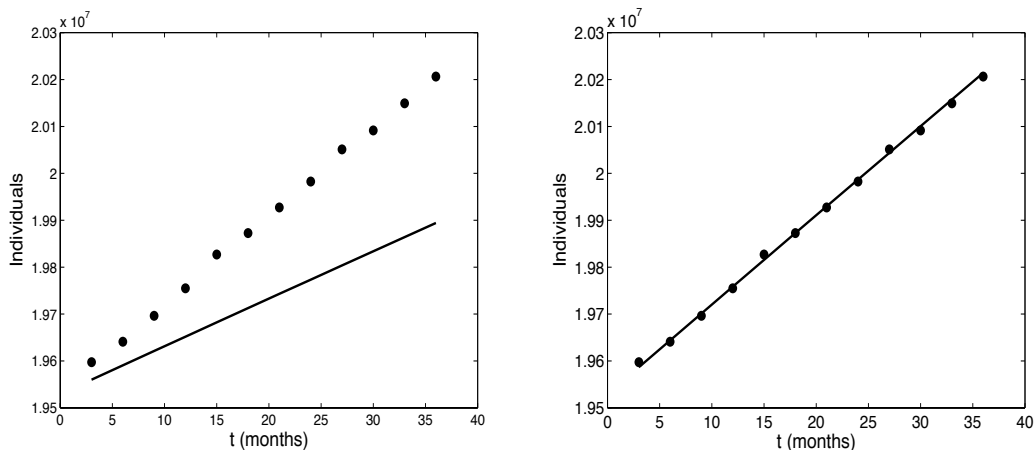


Figure 9: Total population of the model overlayed with estimated resident population data. Left: $\lambda = 21,011$ and $\mu = 0.000551$ calculated from ABS. Right: $\lambda = 25,000$ and $\mu = 0.0003$.

	θ	θ	Estimate	$SE(\hat{\theta})$	Estimate	$SE(\hat{\theta})$	Estimate	$SE(\hat{\theta})$
λ	21,011	25,000	35,000*	$1.8e^8$	24,886	$1.7e^8$	25,000	N/A
μ	0.000551	0.0003	0.001*	8.9	0.0003	N/A	$3.058e^{-4}$	8.7
$\hat{\sigma}$	N/A	N/A	$4.6156e^7$		$4.5323e^7$		$4.5278e^7$	
RSS	$4.5096e^{11}$	$5.7456e^8$	$4.6146e^8$		$4.9856e^8$		$4.9806e^8$	

Table 15: Estimation of demographic parameters, λ and μ , from quarterly Australian population estimates. * denote estimated values coinciding with the imposed limits on the feasible parameter space.

When estimating just one parameter at a time, we see that the value does not differ much from the initial guess, and from a comparison of the residual sum of squares, we see that not much is to be gained from estimating either λ or μ . However, the model population corresponds much closer to the observed data when $\lambda = 25,000$ and $\mu = 0.0003$, and since it is important that our model follows the population over a period of years, we take these as the parameters used in the study.