Applications of Mathematical Modeling to Infectious Disease Dynamics in Developing Countries.

by

Christopher Castorena

Program in Computational Biology and Bioinformatics
Duke University

Date:_____

Approved:

_____
Katharina Koelle, Supervisor

_____
Michael Reed

_____
William Wilson

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Program in Computational Biology and Bioinformatics in the
Graduate School of Duke University

2013

ABSTRACT

Applications of Mathematical Modeling to

Infectious Disease Dynamics in Developing

Countries.

by

Christopher Castorena

Program in Computational Biology and Bioinformatics
Duke University

Date:_____

Approved:

_____
Katharina Koelle, Supervisor

_____
Michael Reed

_____
William Wilson

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree of Master of Science
in the Program in Computational Biology and Bioinformatics in the Graduate
School of Duke University

2013

Abstract

Mathematical modeling has proven to be an essential tool for the development of control strategies and in distinguishing driving factors in disease dynamics. A key determinant of a given model's potential to aid in such measures is the availability of data to parameterize the model. For developing countries in particular, data are often sparse and difficult to collect. It is therefore important to understand the types of data that are necessary for a modeling project to be successful. In this thesis I analyze the value of particular types of data for a set of infections. The first project analyzes the importance of considering age-specific mixing patterns in vaccine preventable infections in which disease severity varies with age. The second project uses a simulated data set to explore the plausibility of recovering the parameters of an epidemiological model from a time series data set of dengue haemorrhagic fever reports.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Abbreviations

CRS	congenital rubella syndrome
AAOI	average age of infection.
DHF	dengue haemorrhagic fever.
pMCMC	particle Markov Chain Monte Carlo.
MKL	Intel Math Kernel Library

Acknowledgements

I would first like to thank my advisor, Katia Koelle for her input and guidance throughout my time in her lab. She is a great researcher and has seemingly endless knowledge about disease modeling and ecology. I've learned a tremendous amount since joining her lab and I'll always be grateful for that. My committee has been very supportive throughout my time here. Mike Reed alway had his door open to me to chat about anything, which I took advantage of more than once. Will Wilson's always been a good friend to me anytime I walk down the hall and it used to be a treat to get to pet his giant lab when he brought her in. Randy Kramer has given me a good deal of his time and has been a very good person to talk to as I figure out my future career path. I'd also like to thank Rotem Ben-Shachar and Stacy Scholle for useful input in the draft of my thesis. A portion of this work was generously supported off of NSF grant number DMS-0943760.

1

Introduction

Mathematical models have proven to be useful tools for the development of infectious disease control strategies and for teasing apart the important underlying factors that drive disease dynamics. Even very simple models can make surprisingly accurate predictions. For example, an incredibly simple rubella model published by Knox in 1985 (Knox, 1985) did a fairly accurate job of predicting what would happen when Greece implemented a low coverage Rubella vaccination policy and saw a large outbreak of Congenital Rubella Syndrome (CRS) in the early 1990s (Panagiotopoulos et al., 1999). This model related the proportion of infants who acquire CRS to the vaccination rate as a simple exponential function of the force of infection. While simple models such as this can give important qualitative insights there is also a need for more quantitative results. For the rubella example, simple models can tell us that intermediate vaccination values can lead to bad outcomes, but a policy maker needs to know whether vaccinating a particular proportion of a population will increase or decrease the occurrence of CRS in the population and whether this changes from one population to another. As one moves from simple models that make mostly qualitative predictions to models can potentially make the quantitative

predictions needed to address current public health problems, model parametrization becomes both more important and more difficult. The available data often limit the types of models that can be accurately fit.

Much of the need for mathematical modeling of infectious diseases is currently occurring in developing countries where data are often noisy and sparse. Two of the most important questions that must be addressed when conducting modeling in these situations are "What data should I collect in the future?" and "What can be done with the data that I have?" For this thesis, I present two projects, each addressing one of these questions. The first project addresses the first question in the context of vaccination policies for permanently immunizing diseases by using an age-structured SIR[1] model to illustrate the importance of collecting data on age structured mixing patterns in developing countries. The second project addresses the second question by demonstrating the potential for fitting simple dengue models to relatively short time series, similar to a set of time series that are currently available for the different provinces of Thailand.

[1] SIR models group individuals into susceptible, infected and recovered compartments. Depending on the complexity of the model, there can be multiple subgroups within each of the three compartments.

2

The impact of age-assortative mixing on the value of vaccination policies

2.1 Introduction

Vaccination policies aim to reduce the burden of infectious diseases on populations through the immunization of susceptible individuals. These policies reduce incidence through the direct effect of preventing infection and the indirect effect of establishing herd immunity. Herd immunity is well known to facilitate disease elimination and, in cases where elimination is not possible, to reduce the force of infection and increase the average age of infection (Anderson and May, 1991, 1990). Infection severity for many diseases depends at least partly on the age of the host; well-orchestrated vaccination policies therefore take into account their dual effect on disease incidence reduction and the age distribution of cases. Policies that fail to do so can have unanticipated, counterproductive consequences. For example, a low coverage rubella vaccination policy in Greece may have led to an increase in the number of congenital rubella syndrome cases by shifting the average age of infection to childbearing years (Panagiotopoulos et al., 1999).

An important factor that is known to affect disease dynamics is the pattern of social mixing in a population. These mixing patterns arise from many different aspects of societal structure including school attendance, workplace interactions and family structure. Some generalities in these mixing patterns are well established empirically. For example, people tend to preferentially interact with others of similar age (Mossong et al., 2008; Horby et al., 2011) and mixing rates between children and parents are invariantly high (Mossong et al., 2008) (Figure 2.1 A). However, appreciable country- and region-specific differences are also apparent. For example, while the POLYMOD study (Mossong et al., 2008) indicates that mixing patterns are remarkably similar across European countries, an analogous study in Vietnam found lower tendencies for age-assortative mixing (Horby et al., 2011) (Figure 2.1 B). Similar to the types of epidemiological models used to anticipate the age distribution consequences of vaccination policies, models incorporating age-structured mixing patterns have been instrumental in understanding and predicting patterns of disease incidence and spread. For example, age-structured interactions have been invoked to explain long-term changes in pertussis prevalence (Rohani et al., 2010). Age-structured mixing patterns have also been used to evaluate the probability that a newly emerging virus will spread through a naive host population (Nishiura et al., 2011), and in the case of establishment, to evaluate the optimal allocation of vaccines (Shim et al., 2011; Medlock and Galvani, 2009).

In this chapter, I consider how vaccination policies, in the context of age-assortative mixing patterns, differentially affect the impact of disease incidence in a population. Specifically, following previous approaches that couple economic models with epidemiological models (e.g. (Althouse et al., 2010)), I assess the direct and indirect effects of a vaccination policy on a typical individual in a population under a range of different mixing patterns. While the epidemiological model is used to anticipate incidence levels and age distribution patterns under a proposed vaccination policy

and under a given mixing pattern, the economic model combines these predictions with information on age-dependent disease severity costs to quantify the direct and indirect economic consequences of the vaccination policy. I combine these two effects to assess the value of a vaccination policy.

2.2 Methods

The Epidemiological Model

I consider a dynamic epidemiological model with individuals classified as susceptible (S), infected (I), or recovered and immune to reinfection (R) that includes age structure and assumes permanent immunity following either a primary infection or vaccination. The epidemiological model is mathematically formulated by the following set of partial differential equations:

$$\begin{aligned}
\frac{\partial S(a,t)}{\partial t} + \frac{\partial S(a,t)}{\partial a} &= -\lambda(a,t)S(a,t) \\
\frac{\partial I(a,t)}{\partial t} + \frac{\partial I(a,t)}{\partial a} &= \lambda(a,t)S(a,t) - \gamma I(a,t) \\
\frac{\partial R(a,t)}{\partial t} + \frac{\partial R(a,t)}{\partial a} &= \gamma I(a,t)
\end{aligned} \tag{2.1}$$

with boundary conditions: $S(0,t) = (1\text{-}\rho)\mu N$, $I(0,t) = 0$, and $R(0,t) = \rho\mu N$. The variables $S(a,t)$, $I(a,t)$, and $R(a,t)$ are the number of susceptible, infected and recovered individuals of age a at time t, respectively. The parameters of the model are the infant vaccination rate ρ, the recovery rate γ, and the birth rate μ. I assume a constant population size, N, with a life span of 75 years, such that all births into age class 0 are matched by deaths of individuals at the age of 75. $\lambda(a,t)$ is the age-specific force of infection, described in greater detail below. I numerically

approximate the dynamics using the escalator boxcar train method (De Roos et al., 1992). This approach divides the population into discrete age bins of width Δa, such that the dynamics are modelled as a system of ordinary differential equations instead of the partial differential equations shown above. Aging is modelled by shifting individuals from the i^{th} age bin to the $(i+1)^{th}$ age bin every Δa time units. Here I use an age bin width (Δa) of three months.

Modelling the age-specific force of infection, $\lambda(a,t)$, requires an assumption on the mixing patterns in the population. Because the level of age-assortative mixing has been shown to be a useful metric for comparing mixing patterns across different populations (Farrington et al., 2009; Nold, 1980; Newman, 2003; Iozzi et al., 2010), I implement mixing using a variation of the preferred mixing matrix (Jacquez et al., 1988). For simplicity I therefore assume that all individuals have the same number of contacts, m, per unit time. I partition the population into age groups and define the degree of assortative mixing as the percent of an individual's contacts that are with others in the same age-group. I use a parameter $\theta \in [0,1]$ to control this percentage and assume that the remaining contacts are uniformly distributed between the other age groups. Mathematically, this contact rate matrix $\mathbf{C}$ is given by:

$$\mathbf{C}[i,j] = \begin{cases} \theta m & \text{if } i = j \\ \frac{(1-\theta)m}{\eta - 1} & \text{if i} \neq \text{j} \end{cases} \tag{2.2}$$

where $i, j \in \{1...\eta\}$ and η is the number of age groups. Figures 2.1(C-E) show contact matrices for the values of θ used in our simulations. In transforming the contact matrix, $\mathbf{C}$ into a transmission matrix $\mathbf{M}$, I use an age-independent parameter ϕ to model the probability that a contact between a susceptible and an infected individual leads to a new infection, that is $\mathbf{M} = \phi\mathbf{C}$. This is a common assumption in models using contact rate matrices (Horby et al., 2011; Rohani et al., 2010; Wallinga et al.,

2006) and has been demonstrated (at least in the case of pertussis) to be consistent with empirical data (Rohani et al., 2010). With this formulation of the model the basic reproduction number, R_0, is approximately $\frac{m\phi}{\gamma}$, which is independent of the degree of assortative mixing, θ. This allows R_0 to be calibrated for a specific disease using only ϕ, γ and m. In this formulation, the force of infection experienced by a susceptible individual of age a at time t is given by:

$$\lambda(a,t) = \Sigma_{j=1}^{\eta} \mathbf{M}_{kj} \frac{I_j(t)}{N_j(t)} \tag{2.3}$$

where k is the index of the age group to which the individual of age a belongs, $I_j(t)$ is the number of infected individuals in age group j at time t, and $N_j(t)$ is the total number of individuals in age group j at time t.

The Economic Model

I now specify an economic model to assess the consequences of different levels of age-assortative mixing on the value of a vaccination policy. To begin, I can evaluate the expected cost of infection over an unvaccinated individual's lifetime using the equation:

$$K(\theta,\rho) = \int_0^T k(a) f(a \mid \theta,\rho) da \tag{2.4}$$

where the function $f(a|\theta,\rho)$ specifies the probability that an individual becomes infected at age a, given a particular level of assortative mixing (θ) and vaccination rate (ρ), and T is the life span of an individual. The function $k(a)$ provides the

expected cost of infection at age a, which includes costs associated with hospital care, lost productivity and permanent disability (Takahashi et al., 2011; Al-Awaidy et al., 2006). I use the catalytic model (Muench, 1991) to write $f(a|\theta,\rho)$ as a function of the equilibrium age-specific force of infection for a particular level of assortative mixing and vaccination coverage, $\lambda(a \mid \theta,\rho)$:

$$f(a|\theta,\rho) = \lambda(a \mid \theta,\rho)e^{-\int_0^a \lambda(x|\theta,\rho)dx} \tag{2.5}$$

This expected lifetime cost of infection can then be used to calculate the expected value of the proposed vaccination policy to a typical individual in the population. The expected cost of the disease with no vaccination policy is $K(\theta,0)$. Under a vaccination policy with coverage level ρ, a vaccinated individual has an expected cost of zero. The effect on this individual, which I call the direct effect, is the difference between his expected cost under this policy and his expected cost under a no vaccination policy. Under this same policy, the expected lifetime cost for an unvaccinated individual is $K(\theta,\rho)$. The effect of the policy on this individual, which I call the indirect effect of the policy, is again the difference between the expected cost under a no vaccination policy and the expected cost under the considered policy. Combining these two effects, the expected value of a vaccination policy on a typical individual is given by:

$$V(\theta,\rho) = \underbrace{\rho K(\theta,0)}_{\text{direct effect}} + \underbrace{(1\text{-}\rho)(K(\theta,0)\text{-}K(\theta,\rho))}_{\text{indirect effect}} \tag{2.6}$$

where the vaccination rate ρ is used to weight the direct and indirect effects of vaccination. I use this framework to assess the value of a vaccination policy under different age-assortative mixing patterns and under four distinct disease cost curves. I first analyse the base case of age-independent costs. The remaining cost curves

assume that severity peaks in either young, intermediate or elderly age groups.

2.3 Results

Effects of vaccination and age-assortative mixing on the age distribution of infection

Increases in population vaccination rates are well known to decrease the probability that an unvaccinated individual becomes infected during his lifetime. Consistent with this expectation, simulations of the dynamic epidemiological model described above show lower lifetime infection probabilities at higher vaccination rates and also indicate that these lifetime infection probabilities are independent of the degree of age-assortative mixing (Figure 2.2 A), depending solely on the value of the basic reproduction number, R_0.

Increases in vaccination rates are also known to shift the average age of infection (hereafter, AAOI) to older age groups (Anderson and May, 1991; Panagiotopoulos et al., 1999; Keeling and Rohani, 2008). Again, consistent with this expectation, our simulation results show increases in the AAOI with increases in vaccination rates (Figure 2.2 B), regardless of the degree of assortative mixing. Interestingly, for any given level of vaccination, these simulations also show that the AAOI is lower at higher levels of age-assortative mixing (Figure 2.2 B). This is because, for diseases with permanent immunity, young age groups have the highest proportion of infected individuals. Higher levels of assortative mixing therefore result in higher forces of infection for young individuals by increasing the proportion of a young person's contacts that are with infected individuals. This in turn lowers the AAOI.

Effects of vaccination and age-assortative mixing on the expected lifetime cost of infection

The expected cost of disease over an unvaccinated individual's lifetime (equation 2.4) can now be calculated by incorporating the simulation results of the epidemiological

model. I first consider the simplest case: when the cost of infection is independent of an individual's age (Figure 2.3 A). In this case, the expected cost reduces to the product of the age-independent cost of infection, k, and the probability that an individual becomes infected over his lifespan. The latter I have shown always decreases with increases in vaccination rates and is independent of the level of age-assortative mixing (Figure 2.2 A). When costs of infection are independent of age, the extent of age-assortative mixing therefore does not influence the expected cost of disease over an unvaccinated individual's life (Figure 2.4 A) and higher vaccination rates always act to decrease the expected cost. This latter result is consistent with vaccination yielding a positive externality, in which unvaccinated individuals benefit from the herd immunity generated by vaccinated individuals.

I now consider a cost curve for which infections at a younger age incur a greater cost (Figure 2.3 B). Under this cost curve, the expected lifetime cost to an unvaccinated individual always decreases with increases in vaccination rates (Figure 2.4 B). This is because an increase in vaccination rate decreases the lifetime infection probability of an unvaccinated individual as well as decreases the probability that an unvaccinated individual who does become infected does so at a young age. These simulations also show that the expected lifetime cost is higher at higher levels of assortative mixing, for any given vaccination rate (Figure 2.4 B). This is because the AAOI is lower at higher levels of age-assortative mixing (Figure 2.2 B).

When infections at an older age incur a greater cost (Figure 2.3 C), the expected lifetime cost to an unvaccinated individual increases with an increase in the vaccination rate until close to the disease eradication threshold of 1-$\frac{1}{R_0}$, at which point the expected lifetime cost then decreases. This pattern arises from vaccination affecting the expected cost in two ways. First, vaccination decreases the lifetime probability of infection (Figure 2.2 A), which lowers the expected cost. Second, it increases the average age of infection if an unvaccinated individual does become infected (Figure

2.2 B), which increases the expected cost. Under our parameterization, the lifetime probability of infection changes very little for low to intermediate vaccination rates (Figure 2.2 A). In contrast, there is a relatively large change in the AAOI over this same range of vaccination rates (Figure 2.2 B). This leads to an increasing expected cost over this range of vaccination rates. As the vaccination level approaches the eradication threshold, however, there is a sharp drop in the lifetime probability of infection (Figure 2.2 A), which leads to a decreased expected cost at higher vaccination levels. Figure 2.4 C also shows that, for a given vaccination rate, the expected lifetime cost of infection is lower at higher levels of assortative mixing. This is because higher levels of age-assortative mixing result in lower AAOI (Figure 2.2 B).

The fourth class of diseases I consider are diseases that incur the highest cost of infection at intermediate ages (Figure 2.3 D). Examples include rubella virus, which can be passed from mother to fetus during pregnancy and cause congenital rubella syndrome (CRS) in infants(Anderson and May, 1991), and mumps, which can lead to infertility if adolescent boys become infected(Anderson et al., 1987). Figure 2.4 D shows the expected lifetime cost of infection for an unvaccinated individual under this cost curve. In this case, intermediate values of vaccination again lead to the highest expected cost of infection, a pattern that has been remarked upon previously in the CRS literature (Knox, 1985; Panagiotopoulos et al., 1999; Morice et al., 2003). Under our parameterization, the expected lifetime cost is also lower at higher levels of assortative mixing. As before, this is because a higher level of assortative mixing leads to younger AAOI and fewer infections during intermediate ages (Figures 2.2 B and 2.3 D).

Effects of age-assortative mixing on the economic value of a vaccination policy

I now consider how a vaccination policy differentially affects vaccinated and unvaccinated individuals and how this impacts the overall value of a vaccination policy for a typical individual (equation 2.6). Figure 2.5 A shows the economic value of a vaccination policy for vaccinated and unvaccinated individuals, independently, in the case of the age-independent cost curve shown in Figure 2.3 A. The direct value of the vaccination policy, as defined in equation equation 2.6, is always independent of vaccination level. Under this cost curve, the direct value of the policy is also not sensitive to the level of age-assortative mixing (Figure 2.5 A). The indirect value of the vaccination policy, defined in equation 2.6, increases with vaccination level, showing the benefits of herd immunity to unvaccinated individuals. Because the expected lifetime disease cost to an unvaccinated individual is in this case not sensitive to the level of age-assortative mixing (Figure 2.4 A), the indirect value of vaccination is also insensitive to θ. Figure 2.6 A combines the direct and indirect values of vaccination, showing that the economic value of a vaccination policy for a typical individual increases with vaccination level in the case of the age-independent cost curve, and that the value is not sensitive to the level of age-assortative mixing.

For infections with higher severity at young ages (Figure 2.3 B), the direct value of vaccination again is not affected by vaccination levels. However, the direct value of vaccination is observed to be higher at higher levels of assortative mixing (Figure 2.5 B). This increase occurs because a higher level of assortative mixing shifts the age distribution of infection to younger ages (Figure 2.2 B), which makes protection against infection more valuable. The indirect value of vaccination also increases with vaccination levels. For a fixed vaccination level, the value of vaccination is higher at higher vaccination levels. Figure 2.6 B shows the overall expected value of vaccina-

tion for the young cost curve. In this case the value also increases with vaccination level. At any given vaccination level, the expected value of vaccination is higher at higher levels of assortative mixing because higher levels of assortative mixing result in younger, more severe infections.

The cost curve with highest severity in the elderly is the first curve I analyse to show the potential for negative effects of a vaccination. Figure 2.5 C shows that vaccination can have a small direct value on vaccinated individuals relative to the negative indirect effect on unvaccinated individuals. In the absence of vaccination, there is a very low probability that an elderly individual is susceptible; vaccination therefore only has a significant direct effect on the small number of individuals who would have experienced infection at this age. At any given vaccination level, higher levels of assortative mixing lower the direct value of vaccination. The indirect value of vaccination, always negative in this case, is of a smaller magnitude at higher levels of assortative mixing (Figure 2.5 C). These two patterns can be understood considering the relationship between assortative mixing and AAOI at a fixed vaccination level (Figure 2.2 B). A higher level of assortative mixing has a lower AAOI, resulting in a lower average severity of infection. In the case of the elderly cost curve, the direct value of vaccination is therefore lower at higher levels of assortative mixing. The negative indirect value of vaccination is smaller in magnitude because an unvaccinated individual on average becomes infected at a younger age when assortative mixing is high, which leads to a less severe infection. Figure 2.6 C shows the expected value of vaccination for a typical individual under the elderly cost curve. In this case the expected value is negative until the vaccination level approaches the eradication threshold. This is because there is almost no direct value of vaccination but there is a highly negative indirect effect of vaccination. For low to intermediate vaccination levels, the expected value is less negative with higher levels of assortative mixing because the negative indirect effect dominates over the

direct effect. As the vaccination level approaches the eradication threshold however, the direct effect of vaccination, which is higher at lower levels of assortative mixing, begins to dominate and there is a switch in the order of the curves in Figure 2.6 C. These results suggest that a vaccination policy for diseases that have a strong negative impact on the elderly and minimal effect on other age groups will have to attain a coverage level very close to the eradication threshold before the direct positive effect on vaccinated individuals is able to counterbalance the negative indirect effect on unvaccinated individuals. These results further suggest that even if coverage does approach the eradication threshold the value of the policy will be minimal.

In the case where infection at intermediate ages has the highest severity (Figure 2.5 D), the direct value of a vaccination policy to a vaccinated individual is higher at lower levels of assortative mixing. This is because the expected lifetime cost of disease in the absence of a vaccination policy is higher at lower levels of assortative mixing (see $\rho = 0$ in Figure 2.4 D). At low vaccination values, the magnitude of the negative indirect value of vaccination is smaller at higher levels of assortative mixing because higher levels of assortative mixing drive the AAOI lower, away from the peak of the cost curve. At higher vaccination levels, however, the relationship is reversed. This is because at high vaccination levels, higher levels of assortative mixing shift the AAOI lower, this time towards the peak of the cost curve (Figure 2.3 D). In this case, the overall expected value of vaccination is negative for low vaccination values and then passes a threshold vaccination rate after which a typical individual benefits from the vaccination policy (Figure 2.6 D). This threshold depends on the degree of assorative mixing, with higher vaccination rates needed to pass the threshold value at higher levels of assortative mixing.

2.4 Discussion

Using a simple age-structured epidemiological model I have investigated the role of assortative mixing in modulating the age-profile of infections under an infant vaccination policy. I have shown that high levels of assortative mixing result in a younger AAOI. I then use an economic model to show that this difference in the age-distribution of infection can have important consequences for the value of vaccination policies.

For diseases like measles, which have highest severity in the very young, a vaccination strategy will always be more valuable in populations having higher levels of assortative mixing (Figure 2.6 B) because in the absence of vaccination, these populations will tend to have younger, more severe infections (Figure 2.2 B).

For diseases that have highest severity for intermediate-aged infections, simulation results indicate that vaccination policies will need to surpass a threshold coverage level before they have a net positive effect on the population. While this is well established in the literature (Knox, 1985; Panagiotopoulos et al., 1999), the current analysis indicates that this threshold value depends on the level of assortative mixing in the population. This is especially relevant for developing countries which are considering rubella elimination campaigns and others where vaccines have been introduced onto the private market (Robertson et al., 1997) because there is a concern that this introduction may lead to an increase in the number of cases of congenital rubella syndrome, similar to what was seen in Greece in the early 1990s (Panagiotopoulos et al., 1999). A recent survey study in Vietnam indicated that mixing patterns in developing countries may have a significantly lower level of assortative mixing than those in developed populations. If this is the case, the results presented here indicate that a developing country may not need to attain as high a vaccination level in order to see positive benefits from a rubella vaccination policy.

In the model presented, a high premium was put on simplicity. Other important factors that were not considered here include the effects of mixing between parents and children, age-specific differences in the total number of daily contacts, and maternally derived immunity. The availability of self-reported contact data has in many ways solved the past problem of making ad-hoc assumptions on assumed contact patterns in disease models for developed countries. However, with a notable exception in Vietnam (Horby et al., 2011), these data do not exist for developing countries. There is therefore still a need for assumptions to be made on mixing patterns when modelling the impact of control strategies in developing countries. The results presented here highlight the need to check the robustness of modelling results to uncertainties in mixing patterns, particularly when analysing the expected impact of a control strategy.

The framework developed here can easily be extended to incorporate additional complexities or to look at different disease systems and control methods. For example, this framework could easily be modified to explore the possible effects of the introduction of the MMR vaccine in a developing country. The MMR vaccine has been introduced in the private market in many developing countries (Robertson et al., 1997), where it could lead to an increase in the number of cases of congenital rubella syndrome and male infertility caused by mumps if high coverage levels are not quickly attained. Preventing this increase in disease severity while also taking advantage of the efficiency of combined vaccines will require a model that is able to integrate all three diseases into one framework that takes into account the age-specific severity profiles of each disease.

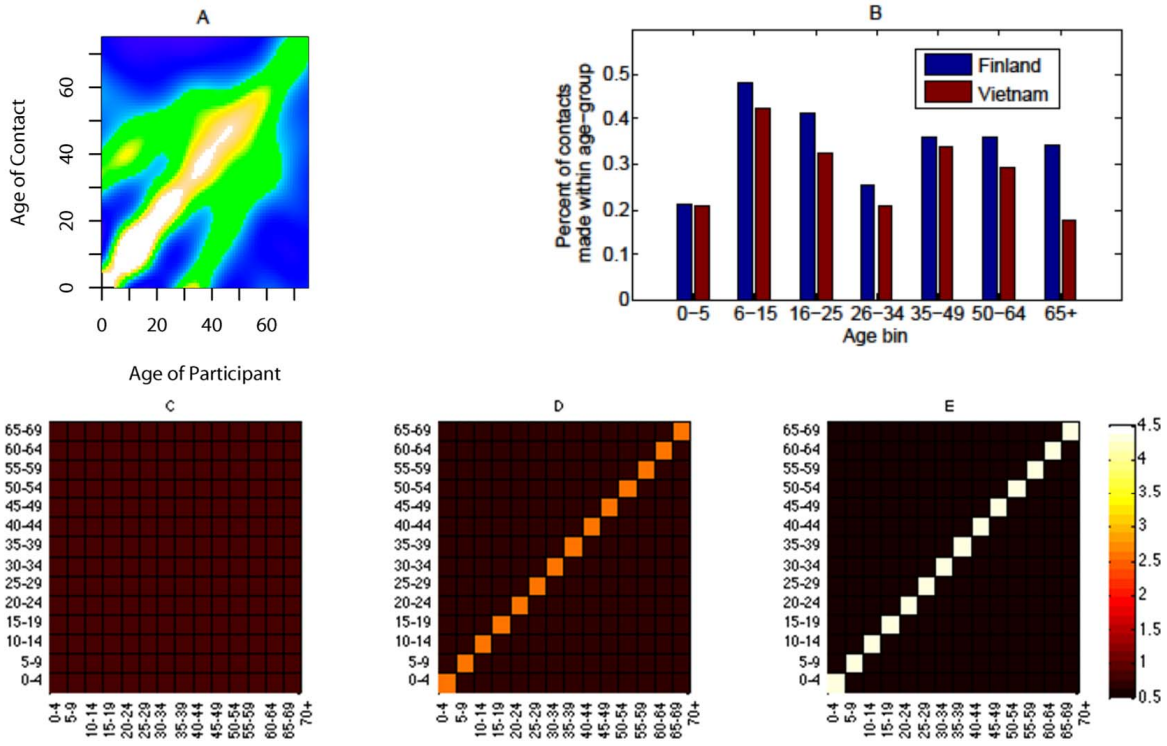

FIGURE 2.1: **Age-assortative mixing patterns.** (A) An estimated contact rate matrix, reproduced from [4]. The matrix was estimated using survey data from the European POLYMOD study and smoothed as described in (Mossong et al., 2008). White indicates high contact rates, green intermediate contact rates, and blue low contact rates, relative to the country-specific contact intensity. (B) Estimates of the percent of within age-group contacts for Finland (Mossong et al., 2008) and Vietnam (Horby et al., 2011). Data were kindly provided by the authors of these papers. (C-E) Contact rate matrices used in the epidemiological model simulations. Individuals were binned into five year age groups. The level of assortative mixing, θ, is defined as the proportion of an individual's contacts that are with other individuals in his/her age group. With age groups of five years and a host lifespan of 75 years, I consider values of θ representing homogenous mixing ($\theta = \frac{1}{15}$, subplot C), an intermediate level of assortative mixing ($\theta = \frac{3}{15}$, subplot D) and a high level of assortative mixing ($\theta = \frac{5}{15}$, subplot E).

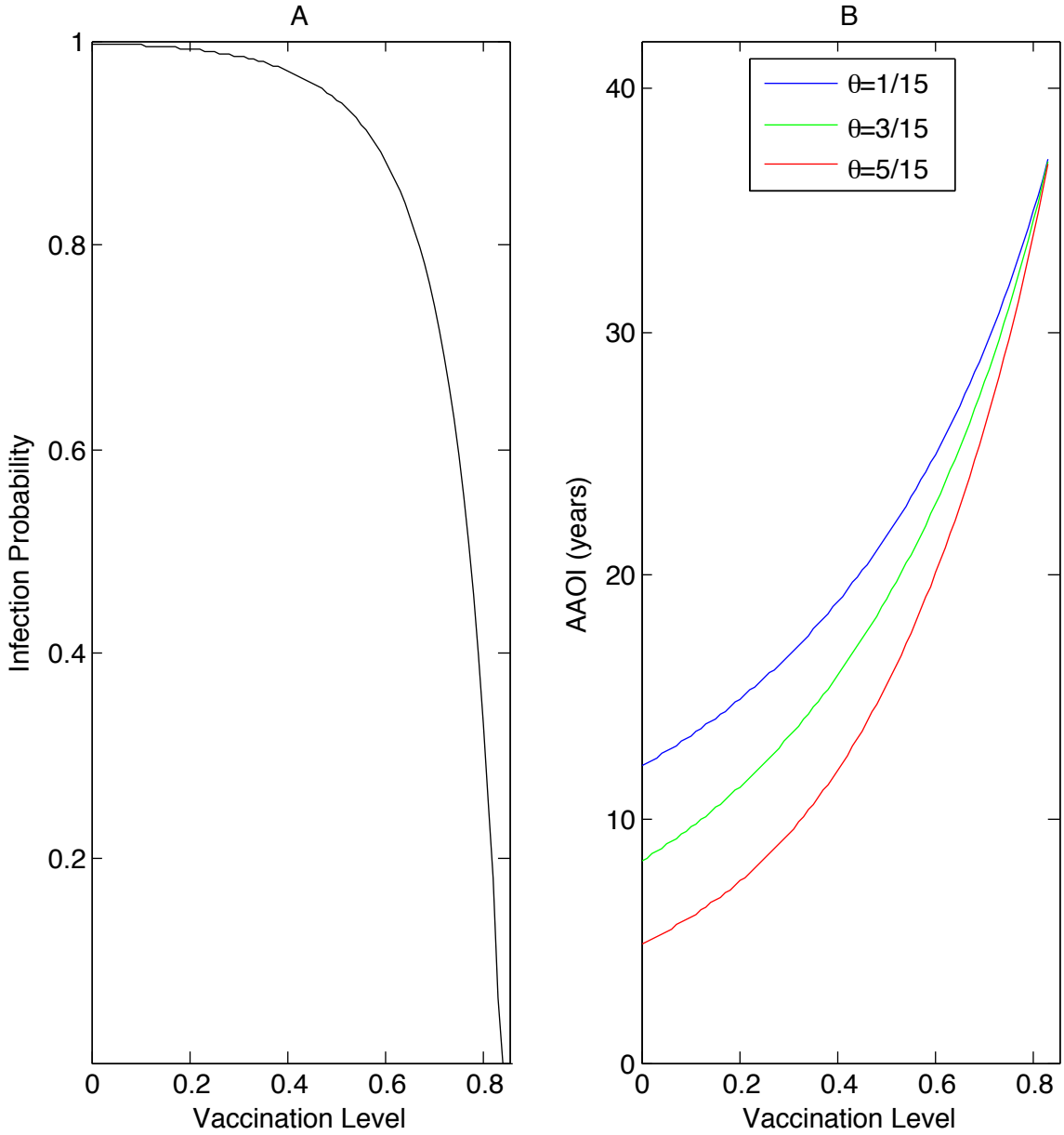

FIGURE 2.2: **Simulation results from the epidemiological model.** (A) The lifetime probability of infection as a fraction of vaccination level, ρ. (B) The average age of infection(AAOI) as a fraction of vaccination level,ρ under a range of vaccination rates for three different levels of assortative mixing ($\theta = \frac{1}{15}$, $\frac{3}{15}$, and $\frac{5}{15}$). The epidemiological model was parameterized with $m = 13$ contacts per day, infectious period $\frac{1}{\gamma} = 14$ days, and infectivity parameter $\phi = \frac{1}{30}$, yielding a value of R_0 of approximately 6.

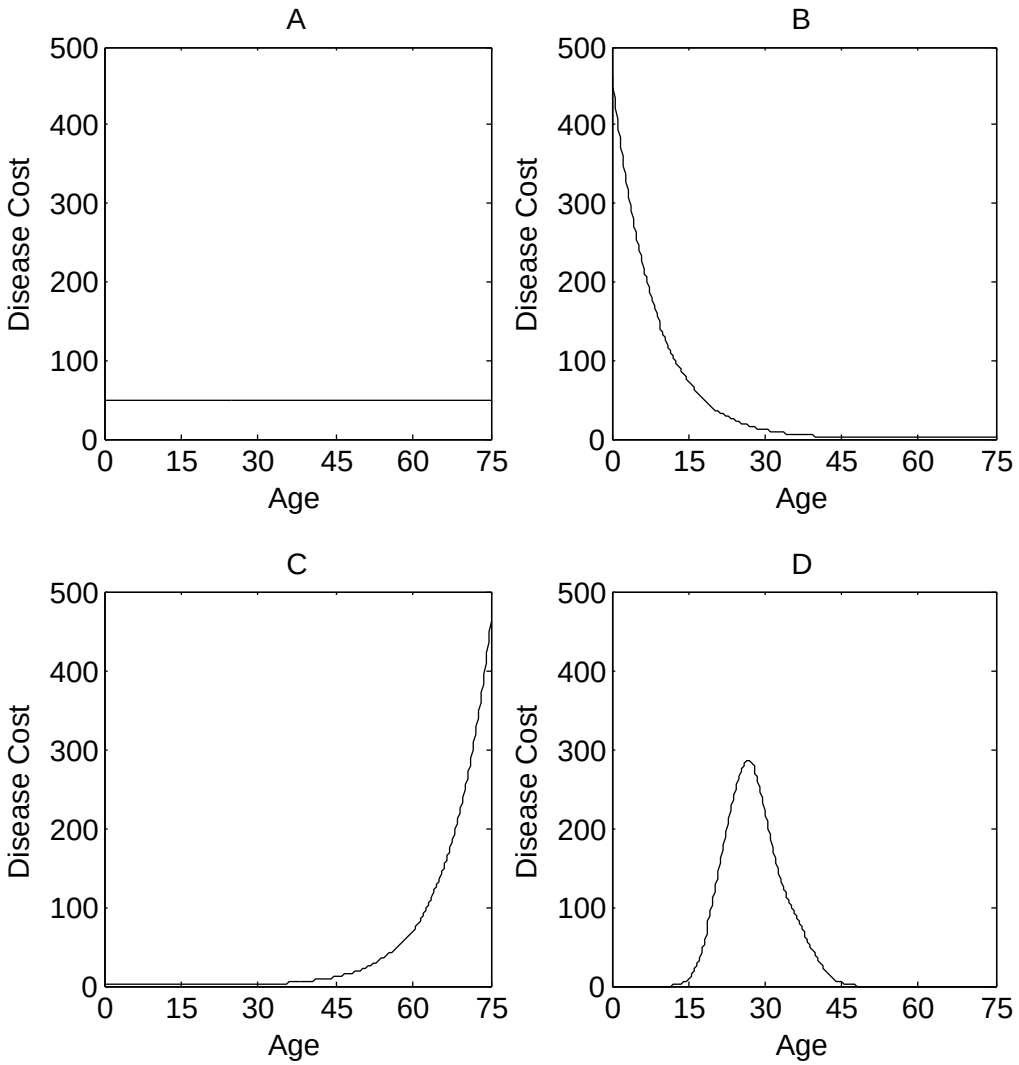

FIGURE 2.3: **Disease cost curves.** (A) An age-independent disease cost curve, with cost $k(a)$ set to 50. (B) A disease cost curve with higher infection costs in the young. The curve $k(a)$ is proportional to $exp(\text{-}\frac{1}{8}a)$ (C) A disease cost curve with higher infection costs in the elderly. The curve $k(a)$ is proportional to $exp(\text{-}\frac{1}{8}(75\text{-}a))$ (D) A disease cost curve peaking at intermediate ages. The cost curve is proportional to a smoothed fecundity curve for Vietnam and therefore can represent this country's rubella cost curve. All curves are scaled to have a mean cost of fifty across ages.

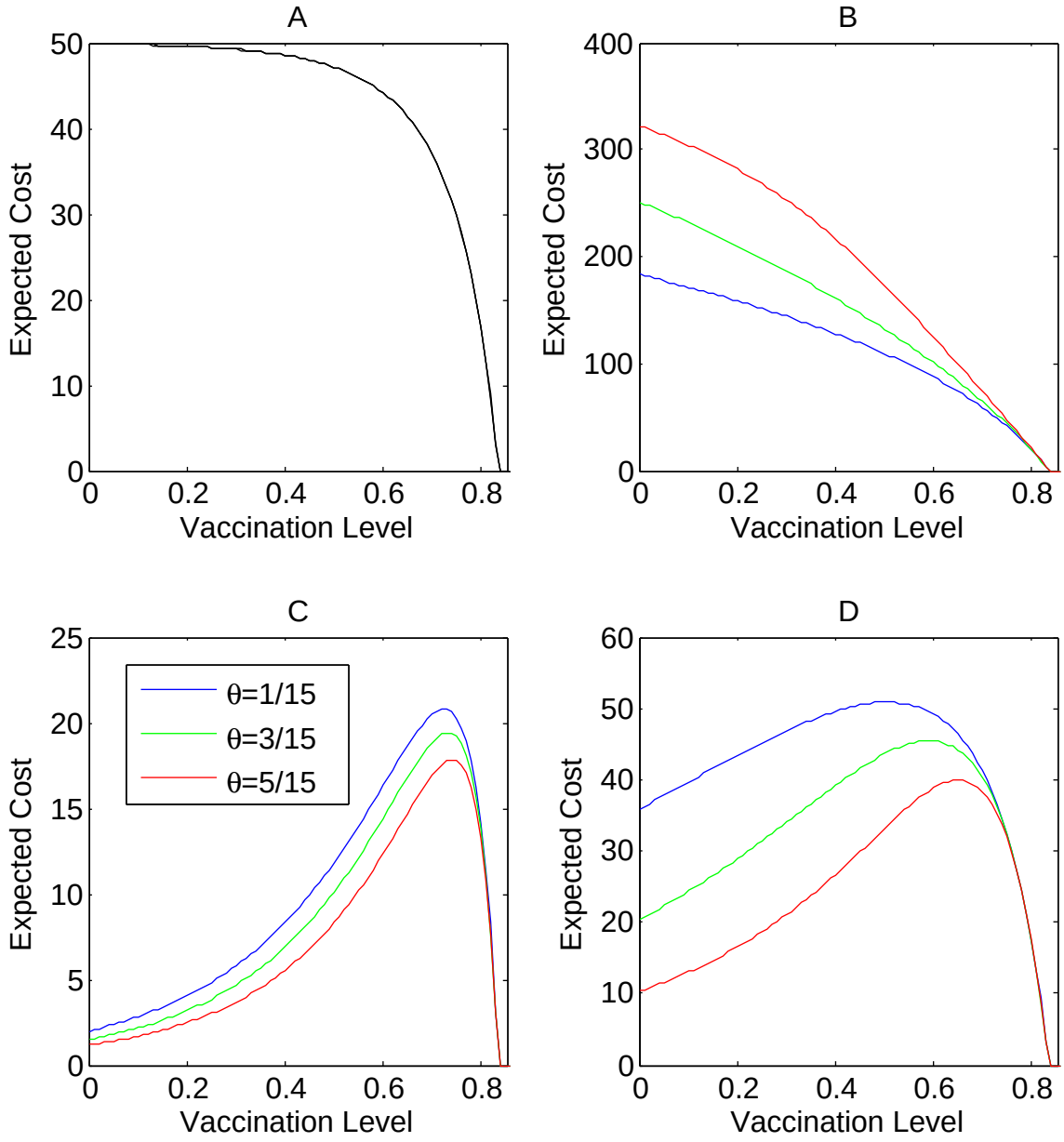

FIGURE 2.4: **Expected lifetime cost of infection to an unvaccinated individual, $K(\theta, \rho)$.** (A) The expected lifetime cost of infection for the age-independent cost curve (Figure 3A). (B) The expected lifetime cost of infection for the cost curve with highest severity in the young (Figure 3B). (C) The expected lifetime cost of infection for the cost curve with highest severity in the elderly (Figure 3C). (D) The expected lifetime cost of infection for the infections with highest severity at intermediate ages (Figure 3D).

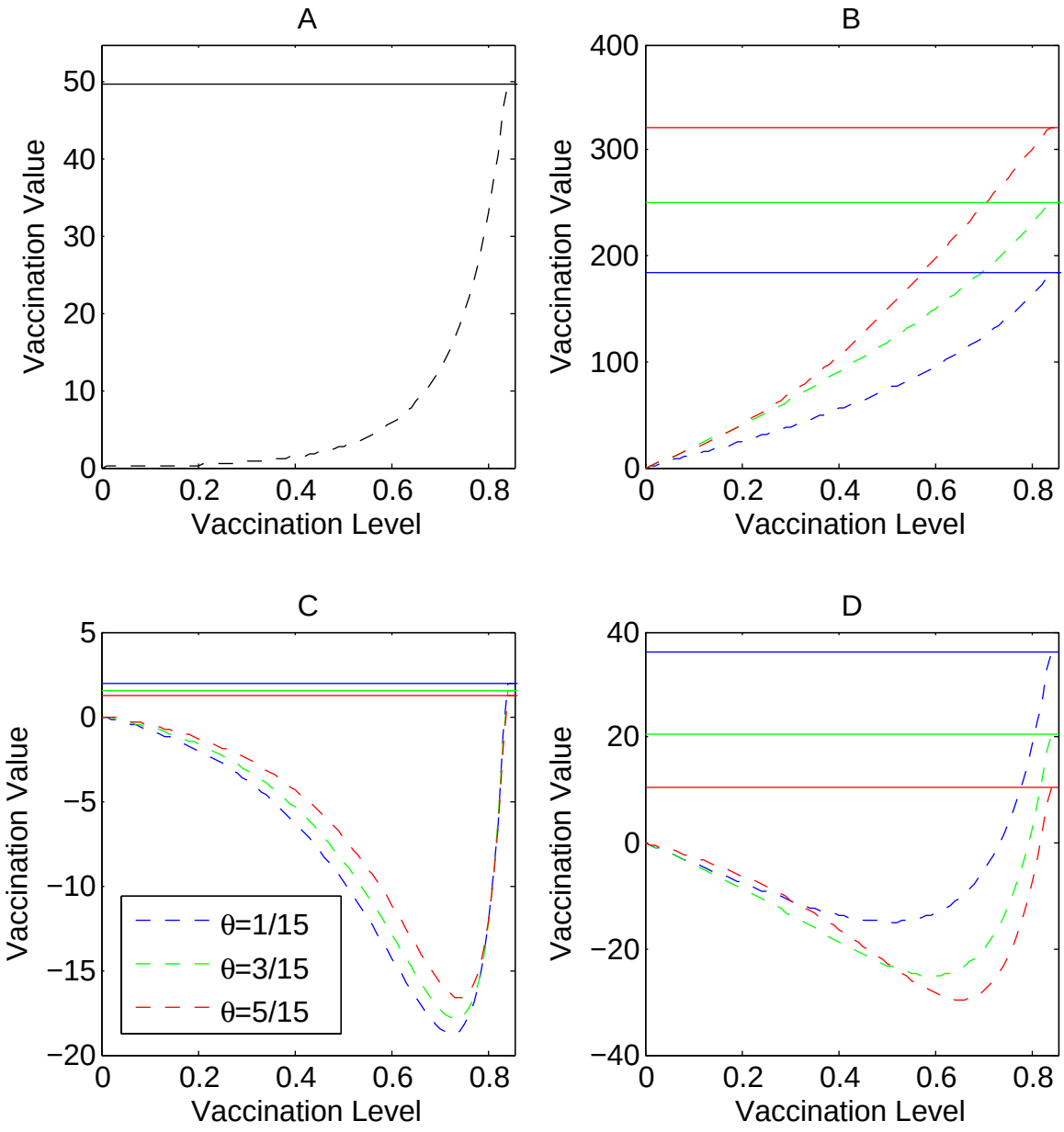

FIGURE 2.5: **The direct and indirect value of vaccination.** Solid lines show the direct value and dashed lines show the indirect value when disease severity is (A) independent of age, (B) highest in the young, (C) highest in the elderly and (D) highest for intermediate ages for different levels of assortative mixing. In all cases the direct value converges to the indirect value as the vaccination level approaches the eradication threshold. Grey lines in (C,D) illustrate where the value of vaccination is zero.

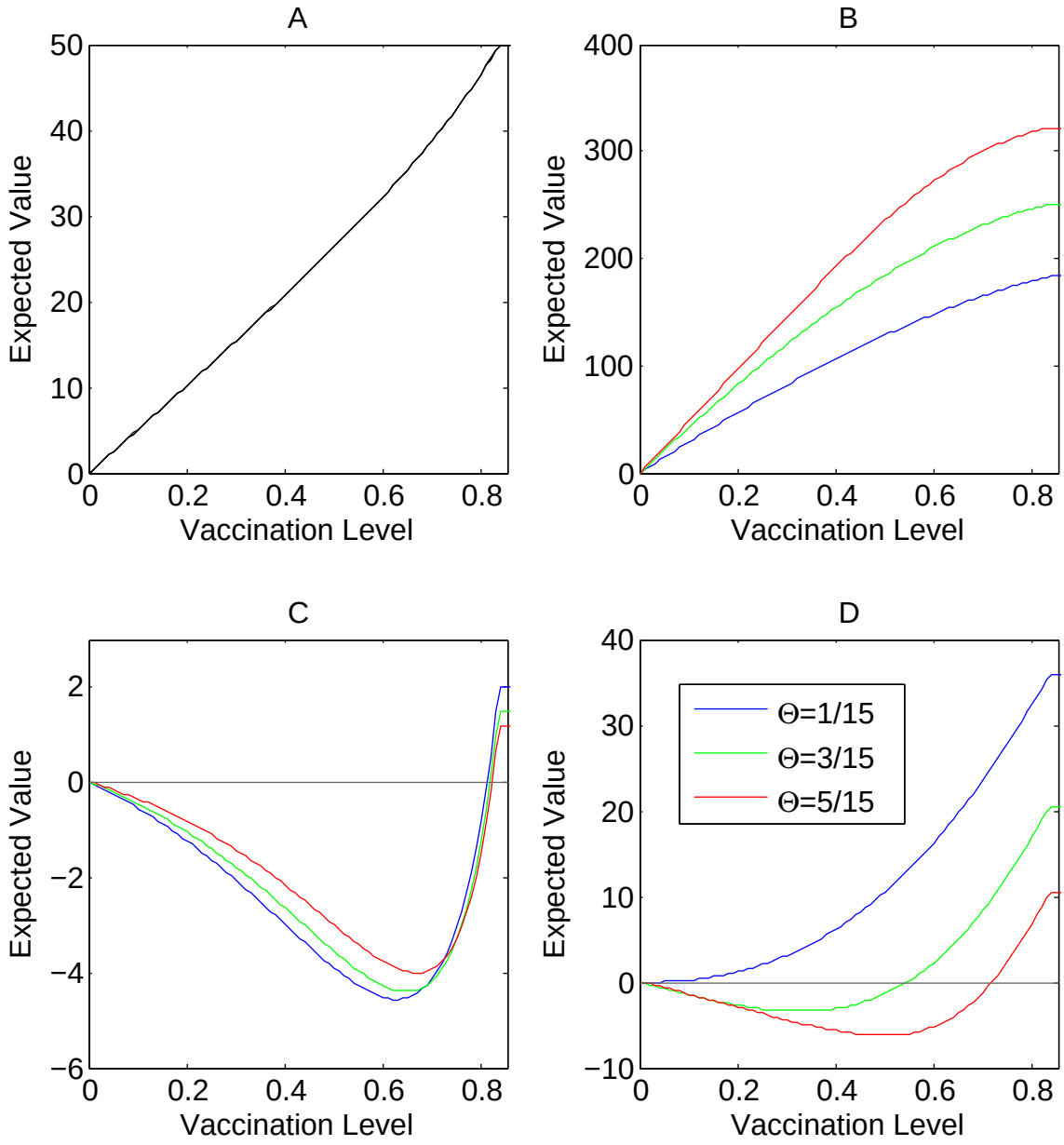

FIGURE 2.6: **The expected value of vaccination for a typical individual in the population** For different levels of assortative mixing, subplots show the expected value of vaccination when disease severity is (A) independant of age, (B) highest in the young, (C) highest in the elderly and (D) highest for intermediate ages. Grey lines indicate where the value of vaccination is 0.

3

Bayesian Parameter Inference of a Four Serotype Dengue Model Using Monthly Dengue Hemorrhagic Fever Case Report Data

3.1 Introduction

Dengue is a vector-borne virus that affects 390 million people every year (Bhatt et al., 2013). Dengue virus is categorized into four distinct serotypes. Infection with any one serotype is thought to grant the infected person permanent immunity to that serotype, but may lead to enhanced severity for infection with one of the other three serotypes. This enhanced severity is thought to arise through a process known as antibody dependant enhancement (ADE) in which heterologous antibody present from the primary infection binds to the virus but cannot neutralize it. This antibody-virus complex then binds to $fc\gamma$ receptors present on monocytes. These monocytes, which are not easily infected in the absence of antibody, are then infected by the virus which leads to increased viral replication and more severe manifestations of the disease (Whitehead et al., 2007). However, the majority of infections are either asymptomatic or mild (Burke et al., 1988) (Endy et al., 2002).

There is a vast literature on dengue modelling and which factors of this complex infection are important to include for various modeling aims (Johansson et al., 2011). Some models include the vector population explicitly (Focks et al., 1995) (Keeling and Rohani, 2008) while others do not (Nagao and Koelle, 2008) (Wearing and Rohani, 2006) (Adams and Boots, 2006) (Coutinho et al., 2006). Other factors that are sometimes included in dengue models include seasonality (Wearing and Rohani, 2006), enhancement of secondary infections (Adams and Boots, 2006) (Billings et al., 2007) (Medlock and Galvani, 2009) and temporary cross protection between serotypes (Wearing and Rohani, 2006) (Nagao and Koelle, 2008). While many of these models reproduce dynamics that are qualitatively similar to those observed empirically, little has been done in the area of formal model fitting and comparison. This is an important problem both for ecologists who wish to quantitatively compare different hypothesis about factors driving disease dynamics and for policy makers who need to choose and fit models for making predictions on the possible impacts of proposed control strategies.

Fitting mechanistic epidemiological models to data is a difficult problem. Many parameters cannot be directly measured and must be inferred from their effects on the dynamics of case report data. Even some of the simplest disease models are highly non-linear and case report data are often very noisy and only available for a limited number of years. These problems are particularly important for dengue models. The four serotypes of dengue often lead to high dimensional models and stiff systems of equations[1]. This is a problem because for non-linear models one of the best (and only) inference techniques available involves simulating hundreds of thousands of years worth of data from the model each time a likelihood value needs to be estimated (Doucet et al., 2001) (Andrieu and Doucet, 2010) (Ionides et al.,

[1] Here the term "stiff" is used to mean that the time step over which it is safe to assume that the rates of each type of event in the system remains constant is small.

2006). The high dimensionality means that each step is expensive and the stiffness means that small step sizes are needed during simulation. Depending on whether a frequentist (Ionides et al., 2006) or Bayesian approach (Andrieu and Doucet, 2010) is taken, hundreds of thousands of likelihood evaluations may be needed for each model that is being fit and compared. In addition to these difficulties with the underlying model, case report data for dengue are often only available for dengue hemorrhagic fever cases, without serotye classification. This makes inference difficult because the data that are available to fit the model is mostly comprised of a small fraction of secondary infections and it is not clear whether the unknown parameters have enough of an effect on these dynamics to make inference possible. The only study to date using these techniques to look at parameter inferrability in the context of multi pathogen models looked at fitting a two serotype model to a forty year long mock time series data set where the serotype of each case was known (Shrestha et al., 2011). The study also assumed that the epidemiological parameters of the model were known and fit only a subset of the unknown parameters. This project looks at the feasibility of inferring all of the unknown parameters from a simple four serotype dengue model on a mock time series data set. This mock data set is comprised of monthly reported DHF cases over a twenty year time span which is similar to what is seen in real data sets. This project frames the inference problem in a Bayesian context and uses particle Markov Chain Monte Carlo for inference.

3.2 Methods

3.2.1 State space model

Fitting mechanistic disease models to time series data is generally done in the context of state space models which are comprised of an observation model and an underlying process model that describes the true underlying state of the population. The process model used for this project is a four sertoype SIR model. This process

model is coupled to a time series of mock monthly DHF incidence using a binomial observation process model described below.

For the SIR model, individuals are grouped by how many infections they have had. It is assumed that each serotype has equal prevalence in the population, that is, each serotype makes up $\frac{1}{4}$ of the total number of infected individuals. Those in group S_i have recovered from i strains and are susceptible to 4-i strains. The force of infection is $\frac{4\text{-}i}{4}$ times the full force of infection, $\beta(t)S_0I_{tot}$. Individuals in category I_i are experiencing their i^{th} infection. The model also keeps track of the cumulative number of secondary infections which is needed for the observation model described below. Recovering from a strain grants permanent immunity to that strain. Secondary infections are assumed to lead to DHF with probability ρ. Noise is incorporated into the transmission parameter, $\beta(t)$, using external parameter noise (Keeling and Rohani, 2008) and the model is simulated using the Euler-Maruyama algorithm (Kloeden and Platen, 1992). The stochastic differential equations governing this system are given by:

$$
\begin{aligned}
\dot{S}_0 &= \mu N - \beta(t) I_{tot}(t) \frac{S_0(t)}{N} - \mu S_0(t) \\
\dot{I}_1 &= \beta(t) I_{tot}(t) \frac{S_0(t)}{N} - (\nu + \mu) I_1(t) \\
\dot{S}_1 &= \nu I_1(t) - \frac{3}{4} \beta(t) I_{tot}(t) \frac{S_1(t)}{N} - \mu S_1(t) \\
\dot{I}_2 &= \frac{3}{4} \beta(t) I_{tot}(t) \frac{S_1(t)}{N} - (\nu + \mu) I_2(t) \\
\dot{S}_2 &= \nu I_2(t) - \frac{2}{4} \beta(t) I_{tot}(t) \frac{S_2(t)}{N} - \mu S_2(t) \\
\dot{I}_3 &= \frac{2}{4} \beta(t) I_{tot}(t) \frac{S_2(t)}{N} - (\nu + \mu) I_3(t) \\
\dot{S}_3 &= \nu I_3(t) - \frac{1}{4} \beta(t) I_{tot}(t) \frac{S_3(t)}{N} - \mu S_3(t) \\
\dot{I}_4 &= \frac{1}{4} \beta(t) I_{tot}(t) \frac{S_3(t)}{N} - (\nu + \mu) I_4(t) \\
\dot{S}_4 &= \nu I_4(t) - \mu S_4(t) \\
\dot{C}_2 &= \nu I_2(t)
\end{aligned}
\tag{3.1}
$$

where:

$I_{tot}(t) = I_1(t) + I_2(t) + I_3(t) + I_4(t) + m$

$\beta(t) = \beta_m(1 + \epsilon sin(\frac{2*\pi(t+\phi)}{365}))(1 + \xi)$

$\xi \sim N(0, \frac{\psi}{\sqrt{dt}})$

The observation model assumes that the number of cases of DHF during month t, denoted Y_t, is binomially distributed where the number of trials is equal to the number of secondary cases of dengue that occurred over the month, C_2^t, with probability of DHF manifestation ρ:

$$Y_t \sim Binomial(C_2^t, \rho) \tag{3.2}$$

Because observations are made at monthly intervals, the epidemiological model above is integrated using Euler-Maruyama for a one month time interval between observations.

3.2.2 Parameter inference

The model described above was used to simulated 23 years of monthly DHF incidence using parameter values in Table 3.1. Figure 3.2 shows this time series along with monthly DHF reports in Bangkok from 1981-2005. Both time series exhibit similar magnitudes and mixtures of annual and multi-annual periodicity. The parameters from the model that do not have empirical estimates from the literature are β_m, m,ϵ,ϕ,ψ and ρ. These parameters are inferred from the mock time series using particle Markov Chain Monte Carlo (pMCMC). This is a Bayesian approach which replaces the exact evaluation of the likelihood function in the standard MCMC algorithm with a particle filter estimate (Andrieu and Doucet, 2010). This approach allows one to use MCMC for parameter inference on non-linear, non-Gaussian state space models such as the one described above. See supplemental section in this chapter for a mathematical description of the pMCMC algorithm and some techniques for increasing the speed of the algorithm. Uniform distributions were used for all priors. In this analysis, initial conditions are treated as nuisance parameters and integrated out of the particle-filter likelihood estimates (details are given in the particle filter description in the supplemental section). Four independent chains were run. Each chain was seeded with a different starting parameter vector which was perturbed from the true value used to simulate the data.

3.3 Results

pMCMC along with an importance sampling prior distribution on initial conditions described in the supplemental section was able to quickly recover the parameters of interest from the mock simulated data, as well as the initial conditions of the system. Figure 3.2 shows the estimated posterior distributions of the initial conditions which are sampled during the particle filter likelihood estimate in the particle MCMC algorithm (See Supplemental Materials section). Each component of the initial condition vector is well captured by these distributions. The cost of estimating the initial conditions in this manner comes in the form of increased computation (more particles are needed for a given level of precision for the likelihood estimates) and higher variance in the posterior distributions of the parameter values. Figure 3.3 shows the estimated posterior distributions of the parameters where samples are pooled across chains. The true parameter values used to simulate the mock data were well captured by each of these distributions.

3.4 Discussion

Using a simple, four-serotype dengue model, I have investigated the plausibility of recovering the epidemiological parameters of a simple, four serotype dengue model from a mock time series of monthly DHF which shares many similarities with an empirical time series from Bangkok. Using particle Markov Chain Monte Carlo, I was able to recover the unknown parameters of the model as well as accurate estimates of the initial state of the population. While previous work focused on inferring a subset of unknown parameters while treating many of the epidemiological parameters as known (Shrestha et al., 2011), this work points towards the plausibility of fitting all unknown parameters for dengue models to relatively short, highly aggregated time series. The mock time series used here shares many of the features observed in the

empirical time series from Bangkok, however, this model will need to be modified if it is to be used to fit this time series from Bangkok. Perhaps the most important modification to incorporate will be relaxing the assumption that R_0 remains constant over the course of the time series. As with all simulation based inference techniques, one of the main hurdles to comparing model variations and exploring parameter space is computational cost. Even with programming the particle filter in a low-level language and parallelizing it on multiple processors, these analyses still took over a month to run. Computational costs ruled out exploring many modifications to this model that may have allowed the empirical data to be fit. These modifications include increasing the number of compartments to account for temporary-immunity following infection and including demographic noise which is computationally very expensive. This project explored several techniques for taking advantage of multiple processor computing systems, highly optimized compilers and efficient linear algebra libraries. The speed-up resulting from these techniques will continue to improve as new technologies come onto the market. Graphics processing units, along with associated numerical libraries have the potential to drastically speed up this type of computation. The many integrated core architecture being developed by Intel could soon make computing systems with tens or hundreds of cores affordable even for local desktop computers. All of these technologies may make the type of analysis done here possible for a much wider range of models very soon. This phenomena of computational techniques becoming available and needing to wait for computing capabilities to catch up happens quite often in scientific computing. Examples of this include numerical integration, Gibbs sampling and Bayesian statistics in general. This project has demonstrated the potential for all unknown parameters in a dengue model to be inferred from even a modest amount of data and outlines several techniques for speeding up simulation inference techniques. It remains to be seen whether improved computing resources and advances in parallel computing tech-

niques will allow for the exploration of models that do a good job of fitting empirical data and rigorous comparison amongst these models.

3.5 Supplementary Materials

Particle Markov Chain Monte Carlo

Particle Markov Chain Monte Carlo (pMCMC) is an extension of Markov Chain Monte Carlo which uses a particle filter to estimate the likelihood value (Andrieu and Doucet, 2010). Let y denote the vector of observation values, θ_i denote i^{th} state of the chain and IC_i denote the i^{th} sample of the initial conditions vector. The prior distribution on the parameter vector is denoted by $\pi(\cdot)$ The pMCMC algorithm proceeds as follows:

1. Initialize the starting parameter vector θ_0. Run a particle filter to obtain an estimate of the likelihood of θ_0, $\hat{P}(y|\theta_0)$ and a sample from the posterior distribution of initial conditions, IC_0.

2. for iteration $i \geqslant 1$

   (a) sample $\theta^* \sim q(\cdot|\theta_{i\text{-}1})$.

   (b) run a particle filter to obtain an estimate of the likelihood of θ^*, $\hat{P}(y|\theta^*)$ and a sample from the estimated posterior distribution of initial conditions conditioned on θ^*, IC^*.

   (c) With probability $min(\frac{\pi(\theta^*)\hat{P}(y|\theta^*)q(\theta_{i\text{-}1}|\theta^*)}{\pi(\theta_{i\text{-}1})\hat{P}(y|\theta_{i\text{-}1})q(\theta^*|\theta_{i\text{-}1}))}, 1)$ set $\theta_i = \theta^*$ and $IC_i = IC^*$. Otherwise, set $\theta_i = \theta_{i\text{-}1}$ and $IC_i = IC_{i\text{-}1}$.

3. Take $\{\theta_i\}$ to be a sample from $P(\theta|y)$ and $\{IC_i\}$ to be a sample from $P(IC|y)$.

A single component Metropolis-Hastings algorithm (Gilks et al., 1996) was used where at each iteration, a parameter was chosen at random to be perturbed using Gaussian noise. R_0 and ρ were blocked into a single component because of their high negative correlation. This block was sampled according to a bi-variate normal distribution with negative covariance, which lead to significantly better mixing of these two parameters.

Particle Filtering Algorithm

A particle filter is a statistical tool for inferring distributions of unknown values in a state-space model. Here it is used to estimate the posterior density value of a parameter set, $P(\theta|y)$ and also to sample from the posterior distribution of initial conditions conditioned on the parameter vector and the data, y. Following (Doucet et al., 2001) the algorithm proceeds as follows:

1. Initialization, t=0
    - For $i = 1, ..., N$, sample $\tilde{x}_0^{(i)} \sim \pi(x_0)$ and set $t = 1$
2. Importance sampling step
    - For $i = 1, ..., N$, sample $\tilde{x}_t^{(i)} \sim p(x_t|x_{t-1}^{(i)})$ and set $\tilde{x}_{0:t}^{(i)} = (x_{0:t-1}^{(i)}, \tilde{x}_t^{(i)})$
    - For $i = 1, ...N$, evaluate the importance weights $\tilde{w}_t^{(i)} = g(y_t|\tilde{x}_t^{(i)})$ Where $g(\cdot|\cdot)$ is the density function from the observation model.
    - Let $\tilde{W}_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}$ denote the normalized importance weights.
3. Selection step

- Resample with replacement N particles $(x_{0:t}^{(i)}; i = 1...N)$ from the set $(\tilde{x}_{0:t}^{(i)}; i = 1, ..., N)$ according to the normalized importance weights.
- set t=t+1 and go to step 2

An estimate of the likelihood of θ given y is given by:

$\hat{p}_\theta = \prod_{t=0}^{T} \hat{p}_\theta(y_t)$

where

$\hat{p}_\theta(y_t) = \frac{1}{N} \sum_{k=1}^{N} w_t^{(k)}$

and a sample from the distribution of initial conditions given y can be obtained by drawing an index i from discrete distribution $\{\tilde{W}_T^{(j)}\}$ and take $x_0^{(i)}$ as a sample from this distribution (Doucet et al., 2001). For the prior distribution on initial states, $\pi(\cdot)$, a 1,000 year trajectory was simulated and N points were then sampled from the 1,000 states from this simulation that had the correct phase, ϕ. These N samples are then re-sampled according to $g(y_0|x_0)$.

3.5.1 Techniques for reducing runtime

The pMCMC method is very computationally expensive. In this section, I illustrate a number of techniques that were used in this project to speed up the runtime. About 50% of the runtime was spent generating the Gaussian random variables used in the simulation of the stochastic differential equations (equations 3.1). The Intel Math Kernel Library's Vector Statistical has a very efficient function for generating large numbers of Gaussian random variables. Using this function resulted in the generation of Gaussian random variables taking up a non-significant amount of runtime. The library also has similar functions for evaluating trigonometric functions which are commonly used to incorporate seasonality into models and the exponential function which is frequently used for evaluating the likelihood function of the observation

model.

Another way to significantly speed up a particle filter is to simultaneously step all particles using matrix algebra. If we let the N denote the number of particles and n denote the dimension of the process model, then instead of individually stepping N particles at each step of the filter, the system can be viewed as an N by n matrix, say M. Now, for instance, when calculating the force of infection for our model, instead of needing to nN multiplications to calculate the number of deaths in our model, M can be copied into a buffer matrix and scaled by μ which saves a significant amount of time because scaling a matrix is faster than individually multiplying each element. The same thing can be done when calculating the number of recoveries in the model. The calculation of $\beta(t)$ can now be significantly sped up by generating a vector of $N(1, \frac{\phi}{\sqrt{dt}})$ normal random variables using the function described above. This is then scaled by $\beta_m(1 + \epsilon sin(\frac{2*\pi(t+\phi)}{365}))$ which now only needs to be calculated once instead of N times. For this model, this process of stepping the particles simultaneously led to a 30-40 percent decrease in run time.

The last modification that was used to speed up the particle filter was parallelizing the stepping process described above. If we let $\mathcal{N}$ denote the number of processors available, at each step of the particle filter, $\lfloor \frac{N}{\mathcal{N}} \rfloor$ particles can be farmed out to each processor with one of the processors getting the remaining $N \mod \mathcal{N}$ particles. This can lead to a decrease in run-time that's linear in $\mathcal{N}$ depending on how large the communication costs between processors are relative to the cost of stepping a particle.

Table 3.1: Parameter values used to simulate mock data

Variable Name	Description	Value
N	Population Size	5 million
μ	birth/death rate	$1/50$ year^{-1}
ν	recovery rate	$1/5$ day^{-1} (Chan and Johansson, 2012)
β_m	Transmission rate	1 days^{-1}
m	Import rate	2
ϵ	Magnitude of seasonal forcing	0.1
ϕ	Phase	0.5
ψ	Magnitude of environmental noise	0.08
ρ	P(DHF | Secondary infection)	0.0338 (Sangkawibha et al., 1984)

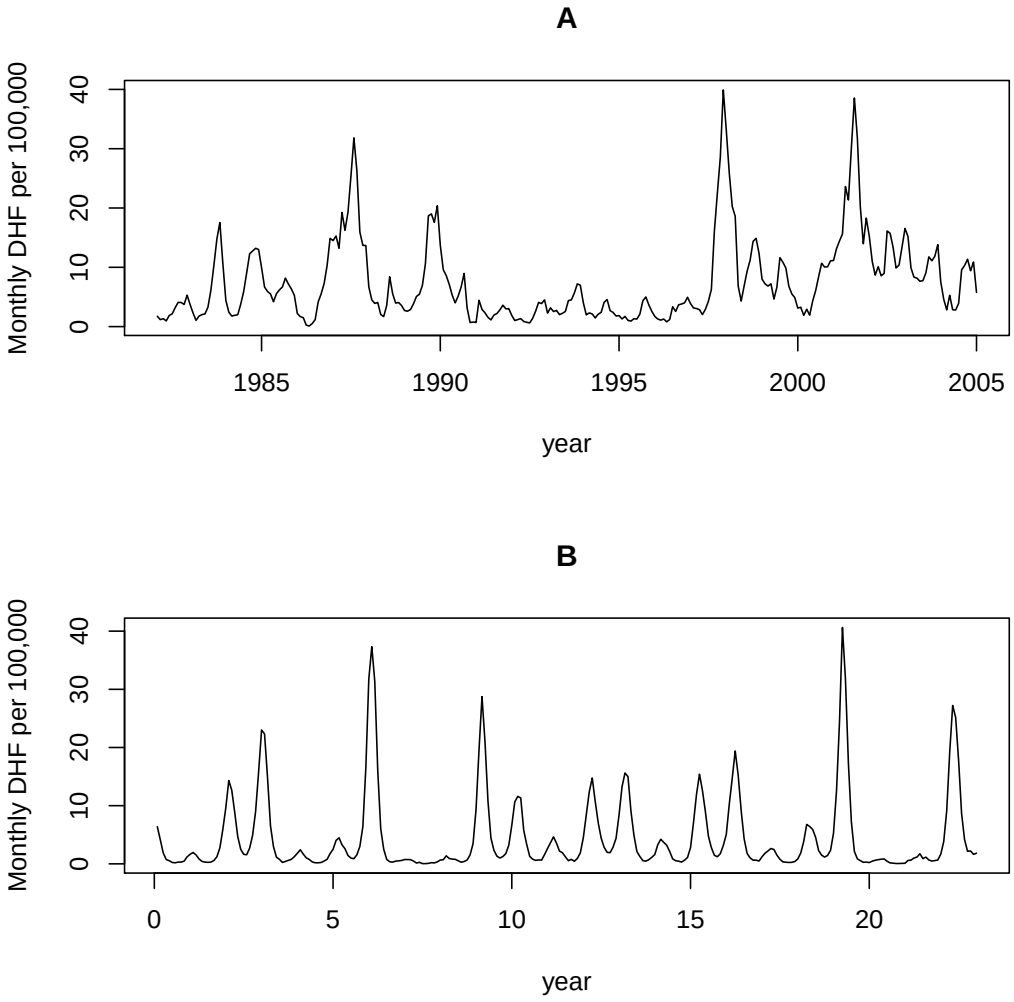

FIGURE 3.1: **Mock data used for parameter inference.** (A) Reported monthly DHF from Bangkok over the period 1981-2005. (B) Mock data simulated using (equation 3.1) and parameters values from Table 3.1.

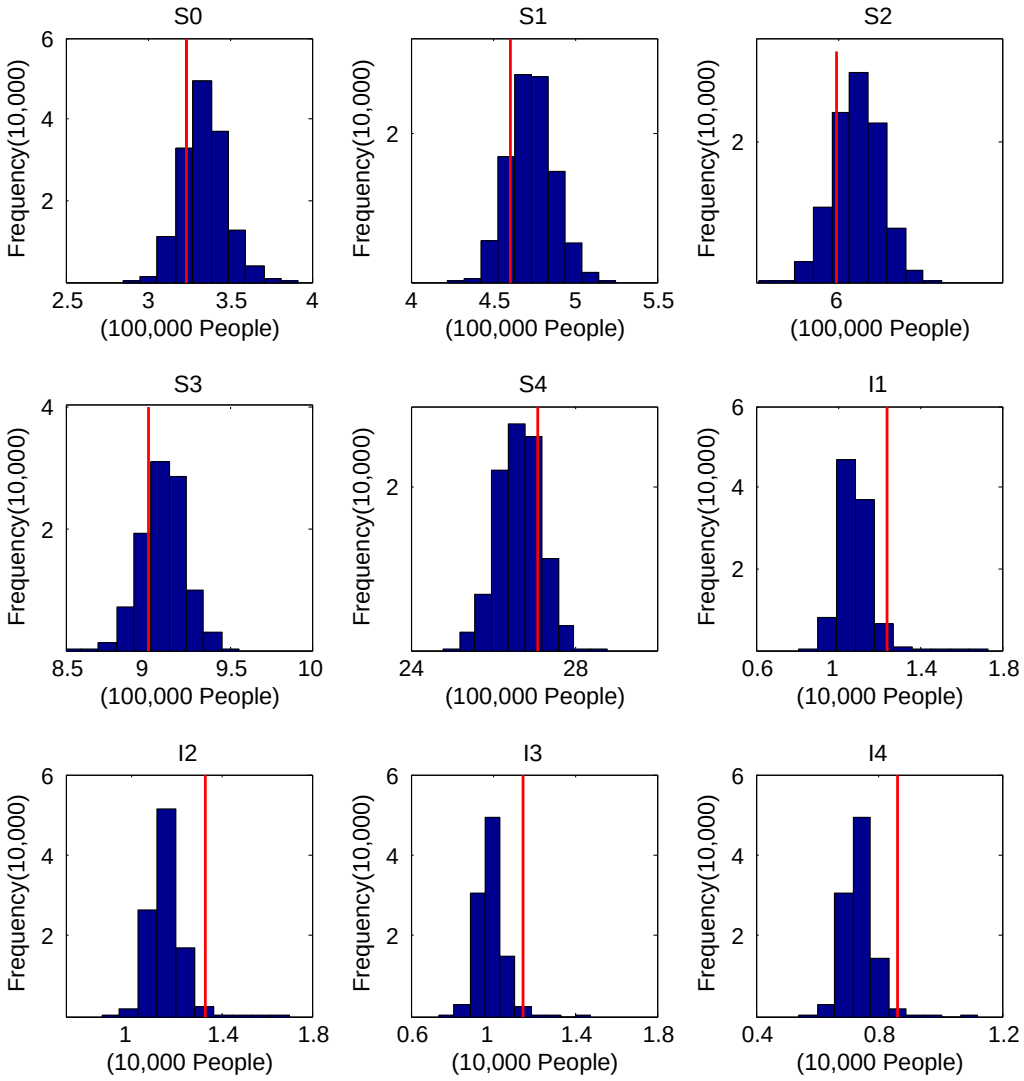

FIGURE 3.2: **Posterior distributions of initial conditions** Estimated posterior distributions for S_0,S_1,S_2,S_3,S_4,I_1,I_2,I_3,I_4 respectively. These samples were obtained using composition sampling to integrate over parameter space.

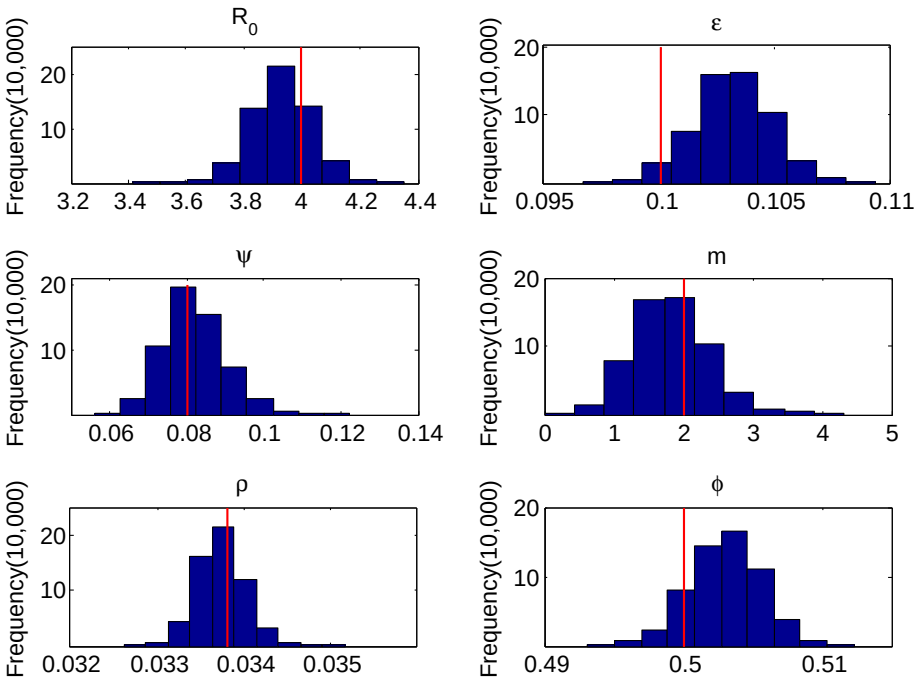

FIGURE 3.3: **Parameter posterior distribution estimates** The above plots show the pMCMC posterior density estimates for Ro, ϵ, ψ, m, ρ and ϕ respectively. Red lines show the true parameter value used to generate the mock data.

4

Conclusions

The focus of this thesis was to illustrate how mathematical modeling can be used to explore what kinds of models can be fit and implemented for a given data set and what types of data will be useful to collect in the future. Each chapter addressed one of these problems in the context of a currently important question in the control and understanding of infectious diseases in developing countries. The first chapter addressed a problem that is currently of great interest to the World Health Organization, which is currently exploring different ways of implementing MMR (combined measles, mumps and rubella) vaccination campaigns in the developing world. Each of these diseases has highest severity in a different age range. The youngest individuals are affected the most by measles. Adolescent boys can develop fertility problems if they become infected with mumps and a rubella infection can be devastating to a child and mother if she become infected during pregnancy. This makes a vaccination campaign a complicated endeavor because it is more economical to combine the vaccines but it is not clear that each of these disease should be included in a campaign if high coverage levels cannot be attained quickly. It has already been seen in Europe that if high coverage levels are not attained for a disease such as rubella

or mumps, the country may be better off waiting to begin a vaccination campaign until higher coverage levels can be attained (Panagiotopoulos et al., 1999). This first chapter used a simple economic model to illustrate that there may be a threshold vaccination value for diseases such as rubella or mumps which must be surpassed in order to see the positive effects of a vaccination policy outweigh it's negative effects. This threshold was shown to be lower when the population has a lower level of age-assortative mixing. This is promising because there is evidence that developing countries have a lower level of age-assortative mixing and may therefore not need as high a vaccination level to avoid adverse vaccine effects that have been seen in a low coverage vaccination campaign in Greece (Panagiotopoulos et al., 1999). Vietnam is the only country for which this mixing data has been collected (Horby et al., 2011). This study suggests that this type of data collection should be done in more developing countries because it can be useful in developing vaccination strategies.

The second project took a more complicated disease, dengue, and explored the plausibility of model fitting using a modest data set similar to what would be seen in many developing countries. This results of this project suggest that even relatively short, low resolution data sets may contain enough signal to fit the parameters of a dengue model that are not directly measurable. This is a currently relative problem because modeling cannot be used to inform policy until the model parameters have been inferred. This project suggests that there is still a significant amount of work to be done in the area of simulation based inference if it is to be used to inform dengue control policy. As parallel computing systems become more affordable, it will become possible to explore a larger array of models and search for the best model for a given problem.

Each of these projects shows promising results for the diseases that they address and suggest work that still remains to be done. The first project shows that rubella and mumps vaccine coverage levels may not need to be as high as previously

thought to see positive net effects. The work that remains to be done is to estimate age-specific mixing information for more developing countries. The second project suggests that policy makers may not need as long or detailed a data set as previously believed in order to conduct model inference. The work that remains to be done for this project is to improve simulation based-inference techniques so that they are less computationally expensive and to improve the affordability of large-scale parallel computing systems.

Bibliography

Adams, B. and Boots, M. (2006), "Modelling the relationship between antibody-dependent enhancement and immunological distance with application to dengue," *Journal of Theoretical Biology*, 242, 337 – 346.

Al-Awaidy, S., Griffiths, U., Nwar, H., Bawikar, S., Al-Aisiri, M., Khandekar, R., Mohammad, A., and Robertson, S. (2006), "Costs of congenital rubella syndrome (CRS) in Oman: Evidence based on long-term follow-up of 43 children," *Vaccine*, 24, 6437 – 6445.

Althouse, B., Bergstrom, T., and Bergstrom, C. (2010), "A public choice framework for controlling transmissible and evolving diseases," *Proceedings of the National Academy of Sciences*, 107, 1696–1701.

Anderson, R. and May, R. (1990), "Immunisation and herd immunity," *The Lancet*, 335, 641 – 645.

Anderson, R. and May, R. (1991), *Infectious Diseases of Humans : Dynamics and Control*, Oxford University Press, Oxford.

Anderson, R., Crombie, J., and Grenfell, B. (1987), "The Epidemiology of Mumps in the UK: A Preliminary Study of Virus Transmission, Herd Immunity and the Potential Impact of Immunization," *Epidemiology and Infection*, 99, 65–84.

Andrieu, C. and Doucet, A. (2010), "Particle Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society*, 72, 269 – 342.

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J., and Hay, S. I. (2013), "The global distribution and burden of dengue," *Nature*, 496, 504–507.

Billings, L., Schwartz, I. B., Shaw, L. B., McCrary, M., Burke, D. S., and Cummings, D. A. (2007), "Instabilities in multiserotype disease models with antibody-dependent enhancement," *Journal of Theoretical Biology*, 246, 18 – 27.

Burke, D. S., Nisalak, A., Johnson, D. E., and Scott, R. M. (1988), "A Prospective Study of Dengue Infections in Bangkok," *The American Journal of Tropical Medicine and Hygiene*, 38, 172–180.

Chan, M. and Johansson, M. (2012), "The Incubation Periods of Dengue Viruses," *PLoS ONE*, 7, e50972.

Coutinho, F., Burattinia, M., Lopeza, L., and Massada, E. (2006), "Threshold Conditions for a Non-Autonomous Epidemic System Describing the Population Dynamics of Dengue," *Bulletin of Mathematical Biology*, 68, 2263–2282.

De Roos, A., Diekmann, O., and Metz, J. (1992), "Studying the Dynamics of Structured Population Models: A Versatile Technique and Its Application to Daphnia," *The American Naturalist*, 139, 123–147.

Doucet, A., de Freitas, N., and Gordon, N. (2001), *Sequential Monte Carlo Methods in Practice*, Springer.

Endy, T. P., Chunsuttiwat, S., Nisalak, A., Libraty, D. H., Green, S., Rothman, A. L., Vaughn, D. W., and Ennis, F. A. (2002), "Epidemiology of Inapparent and Symptomatic Acute Dengue Virus Infection: A Prospective Study of Primary School Children in Kamphaeng Phet, Thailand," *American Journal of Epidemiology*, 156, 40–51.

Farrington, C., Whitaker, H., Wallinga, J., and Manfredi, P. (2009), "Measures of Disassortativeness and their Application to Directly Transmitted Infections," *Biometrical Journal*, 51, 387–407.

Focks, D. A., Daniels, E., Haile, D. G., and Keesling, J. E. (1995), "A Simulation Model of the Epidemiology of Urban Dengue Fever: Literature Analysis, Model Development, Preliminary Validation, and Samples of Simulation Results," *The American Journal of Tropical Medicine and Hygiene*, 53, 489–506.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall.

Horby, P., Thai, P., Hens, N., Yen, N., Mai, Q., Thoang, D., Linh, N., Huong, N., Alexander, N., Edmunds, J., Duong, T., Fox, A., and Hien, N. (2011), "Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases," *PLoS One*, 6, e16965.

Ionides, E. L., Bret, C., and King, A. A. (2006), "Inference for nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, 103, 18438–18443.

Iozzi, F., Trusiano, F., Chinazzi, M., Billari, F., Zagheni, E., Merler, S., Ajelli, M., Del Fava, E., and Manfredi, P. (2010), "Little Italy: An Agent-Based Approach

to the Estimation of Contact Patterns- Fitting Predicted Matrices to Serological Data," *PLoS Comput Biol*, 6, e1001021.

Jacquez, J., Simon, C., Koopman, J., Sattenspiel, L., and Perry, T. (1988), "Modeling and analyzing HIV transmission: the effect of contact patterns," *Mathematical Biosciences*, 92, 119 – 199.

Johansson, M. A., Hombach, J., and Cummings, D. A. (2011), "Models of the impact of dengue vaccines: A review of current research and potential approaches," *Vaccine*, 29, 5860 – 5868.

Keeling, M. and Rohani, P. (2008), *Modeling infectious diseases in humans and animals*, Princeton University Press, Princeton.

Kloeden, P. and Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Springer.

Knox, E. (1985), "Theoretical Aspects of Rubella Vaccination Strategies," *Reviews of Infectious Diseases*, 7, S194–S197.

Medlock, J. and Galvani, A. (2009), "Optimizing Influenza Vaccine Distribution," *Science*, 325, 1705–1708.

Morice, A., Carvajal, X., León, M., Machado, V., Badilla, X., Reef, S., Lievano, F., Depetris, A., and Castillo-Solórzano, C. (2003), "Accelerated Rubella Control and Congenital Rubella Syndrome Prevention Strengthen Measles Eradication: The Costa Rican Experience," *Journal of Infectious Diseases*, 187, S158–S163.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., and Edmunds, J. (2008), "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases," *PLoS Med*, 5, e74.

Muench, H. (1991), *Catalytic models in epidemiology*, Harvard University Press, Cambridge,MA.

Nagao, Y. and Koelle, K. (2008), "Decreases in dengue transmission may act to increase the incidence of dengue hemorrhagic fever," *Proceedings of the National Academy of Sciences*, 105, 2238–2243.

Newman, M. (2003), "Mixing patterns in networks," *Phys. Rev. E*, 67, 026126.

Nishiura, H., Cook, A., and Cowling, B. (2011), "Assortativity and the Probability of Epidemic Extinction: A Case Study of Pandemic Influenza A (H1N1-2009)," *Interdiscip Perspect Infect Dis*, 2011.

Nold, A. (1980), "Heterogeneity in disease-transmission modeling," *Mathematical Biosciences*, 52, 227 – 240.

Panagiotopoulos, T., Berger, A., Antoniadou, l., and Valassi-Adam, E. (1999), "Increase in congenital rubella occurrence after immunisation in Greece: retrospective survey and systematic review," *BMJ*, 319, 1462–1467.

Robertson, S., Cutts, F., Samuel, R., and J, D. (1997), "Control of rubella and congenital rubella syndrome (CRS) in developing countries, part 2: Vaccination against rubella," *Bulletin of the World Health Organization*, 75, 69 – 80.

Rohani, P., Zhong, X., and King, A. (2010), "Contact Network Structure Explains the Changing Epidemiology of Pertussis," *Science*, 330, 982–985.

Sangkawibha, N., Rojanasuphot, S., Ahandrik, S., Viriyapongse, S., Jatanasen, S., Salitul, V., Phanthumachinda, B., and Halstead, S. B. (1984), "Risk factors in dengue shock syndrome: a prospective epidemiologic study in Rayong, Thailand. I. The 1980 outbreak." *American Journal of Epidemiology*, 120, 653–669.

Shim, E., Meyers, L., and Galvani, A. (2011), "Optimal H1N1 vaccination strategies based on self-interest versus group interest," *BMC Public Health*, 11, S4.

Shrestha, S., King, A., and Rohani, P. (2011), "Statistical inference for multi-pathogen systems," *PLoS Comput Biol*, 7.

Takahashi, K., Ohkusa, Y., and Kim, J. (2011), "The economic disease burden of measles in Japan and a benefit cost analysis of vaccination, a retrospective study," *BMC Health Services Research*, 11, 254.

Wallinga, J., Teunis, P., and Kretzschmar, M. (2006), "Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents," *American Journal of Epidemiology*, 164.

Wearing, H. J. and Rohani, P. (2006), "Ecological and immunological determinants of dengue epidemics," *Proceedings of the National Academy of Sciences*, 103, 11802–11807+.

Whitehead, S. S., Blaney, J. E., Durbin, A. P., and Murphy, B. R. (2007), "Prospects for a dengue virus vaccine," *Nature Reviews Microbiology*, 5, 518–528.