Modeling Dengue Transmission and Vaccination

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematics

---

by
Abhishek Pandey
May 2014

Accepted by:
Dr. Elena Dimitrova, Committee Chair
Dr. Jan Medlock
Dr. Anuj Mubayi
Dr. Peter Kiessler
Dr. Matthew Macauley

Abstract

Dengue is one of the most rapidly spreading mosquito-borne viral diseases in the world and inflicts significant health, economic and social burdens on populations. In this dissertation, I studied different aspects of modeling of dengue and vector-borne diseases in general. Among various dengue models that have appeared in literature, some explicitly model the mosquito population, while others model them implicitly. In spite of extensive use of both modeling approaches, little guidance exists for which type of model should be preferred. I developed a Bayesian approach that uses a Markov chain Monte Carlo (MCMC) method to fit disease models to epidemiological data and used it to explore how well these models explain observed incidence and to find good estimates for the epidemiological parameters for dengue. I fitted dengue hemorrhagic fever data from Thailand to both type of models and found using Akaike Information Criterion that explicitly incorporating the mosquito population may not be necessary in modeling dengue transmission. On comparing my estimates of the basic reproduction number, R_0, with other estimates in literature, I found a wide variability in R_0 estimates among studies. This variability in R_0 estimate for dengue transmission is not well understood. By fitting a simple dengue model to dengue incidence for varying R_0 values, I found a logarithmic type relationship between population immunity levels and R_0, which may be a reason for the variability in R_0 estimates. The result also highlighted the importance of finding better estimates of

population immunity level to help more accurately estimate R_0 and other epidemiological parameters for dengue. Driven by the seasonality in mosquito abundance and complex dynamics of dengue, introducing a vaccine may induce a transient period immediately after vaccine introduction where prevalence can spike higher than in the pre-vaccine period. These transient spikes could lead to doubts about the vaccination program among the public and decision makers, possibly impeding the vaccination program. Using simple dengue-transmission models, I found that large transient spikes in prevalence are robust phenomena that occur when vaccine efficacy and vaccine coverage is not either both very high or both very low. Despite the presence of these spikes, vaccination always reduced total number of infections in the 15 years after vaccine introduction. Therefore, policy makers should prepare for spikes in prevalence after vaccine introduction to mitigate the burden of these spikes and to accurately measure the effectiveness of the vaccine program.

Acknowledgments

First and foremost, I am indebted to my advisor, Dr. Jan Medlock, for his valuable guidance, caring, patience, and making my graduate life an enriching experience. His encouragement at each turn with his wisdom and insightful suggestions made research, writing, and teaching enjoyable. I could not have asked for a better mentor.

I owe my gratitude to Dr. Anuj Mubayi, for the guidance, encouragement and advice he has provided throughout my study. I thank him for serving on my committee and his continuing support.

Special thanks to my committee chair Dr. Elena Dimitrova for being so wonderful and supportive. I extend my heartfelt gratitude for all her assistance and warm encouragement.

My sincere appreciation to my committee members: Dr. Peter Kiessler and Dr. Matthew Macauley for their service and encouragement.

I also thank the Department of Mathematical Sciences, Clemson University and Dr. Jan Medlock for the financial support provided during my graduate study.

I am deeply grateful to my parents for their unwavering support, and to my friends Navamita, Gordon and Vivek for standing by me through thick and thin.

Finally, I dedicate my dissertation to my dearest didi (sister), Shuchi for her undying love and support for me.

Table of Contents

Title Page . . . . . **i**

Abstract . . . . . **ii**

Acknowledgments . . . . . **iv**

List of Tables . . . . . **vii**

List of Figures . . . . . **viii**

1 Introduction **1**

2 Parameter estimation using Bayeisan MCMC **9**
2.1 Baysian Markov chain Monte Carlo estimation 12
2.2 Examples 17
2.3 Summary 23

3 Comparing vector–host and SIR models for dengue transmission . **25**
3.1 Introduction 25
3.2 Methods 27
3.3 Results 35
3.4 Discussion 43

4 The role of population immunity for accurately estimating model parameters for dengue transmission **47**
4.1 Introduction 47
4.2 Model and Approach 49
4.3 Results 52
4.4 Discussion 58

5 The introduction of dengue vaccine may temporarily cause large spikes in prevalence **61**
5.1 Introduction 61
5.2 Methods and Approach 63

5.3 Results 65
5.4 Discussion 72

6 Conclusions 77

Appendices 80
A Stability analysis 81
B Steady-state solution 87
C Dengue vaccination models 89

Bibliography 95

List of Tables

2.1 Initial guesses used to estimate β and γ 18
2.2 Posterior summary of β and γ. 19
2.3 Initial guesses used to estimate β, γ, $S(0)$ and $I(0)$ 20
2.4 Posterior summary of β, γ, $S(0)$ and $I(0)$. 21

3.1 Posterior summary of parameter estimates. 39
3.2 Comparison of the vector–host and SIR models with and without fixed initial conditions. 43

4.1 Parameters of the dengue model. 52

C.1 Parameters of the dengue vaccination models. 91

List of Figures

1.1 Areas at risk for dengue transmission. 2

2.1 Estimates of β and γ vs. number of iterations. 19
2.2 Histogram for posterior densities of parameters β and γ. 20
2.3 Estimates of β, γ, $S(0)$ and $I(0)$ vs. number of iterations 21
2.4 Histogram for posterior densities of $\beta, \gamma, S(0)$ and $I(0)$. 22

3.1 Monthly DHF incidence in Thailand (1983-1997). 28
3.2 Monthly DHF incidence in Thailand (January 1984- March 1984). . . 29
3.3 Fits of the vector–host and SIR models. 37
3.4 Posterior parameter densities for the vector–host and SIR models. . . 38
3.5 PRCCs for the vector–host and SIR models 41
3.6 The maximum likelihood fits of models to the data. 42

4.1 Maximum of loglikelihood for R_0 from 1 to 20. 53
4.2 Model fit with optimal parameter estimates for $R_0 = 1.3$ and 17.6. . . 54
4.3 Variations in optimal parameter estimates as R_0 varies. 55
4.4 Maximum of loglikelihood for varying R_0 and $h_R(0)$ values. 56
4.5 Optimal estimates of $h_R(0)$ and h_R^* for varying R_0 values. 57

5.1 Diagram of the main model. 64
5.2 Simulated dengue prevalence after vaccine introduction ($p = 78\%$, $\phi = 30\%$). 66
5.3 Simulated dengue prevalence after vaccine introduction ($p = 90\%$, $\phi = 30\%$) 67
5.4 Effectiveness of the vaccination program at different levels of coverage. 69
5.5 Maximum prevalence over the 15 years after vaccine introduction for varying vaccine efficacy. 70
5.6 Maximum prevalence over the 15 years after vaccine introduction for varying time of vaccine introduction. 71
5.7 Transient prevalence spikes in simpler and more complex model . . . 73

Chapter 1

Introduction

Humankind has always been afflicted by infectious diseases, but epidemics were comparatively rare before the advent of human civilization. Once people stopped leading a nomadic life and began living in villages, then in towns and cities, pathogens that cause infectious diseases started spreading easily. People get exposed to these pathogens either by direct contact (through air, polluted water and food) or by indirect contact (through bloodsucking insect carriers of disease such as mosquitoes, fleas and lice).

Dengue is one of the most rapidly spreading mosquito-borne viral diseases in the world. It inflicts significant health, economic and social burden on populations. Worldwide, an estimated 2.5 billion people live in areas where dengue is an epidemic, of which approximately 975 million live in urban areas in tropical and sub-tropical countries in Southeast Asia, the Pacific and the Americas (Figure 1.1). Dengue has been recognized in over 100 countries and an estimated 50–100 million dengue infections occur annually (Guzmán and Kouri, 2002). Moreover, the global estimated number of disability-adjusted life years (DALYs) lost to dengue in 2001 was $528,000$.

Dengue virus (DEN) is a small, spherical and single-stranded RNA virus in

Figure 1.1: Areas at risk for dengue transmission, 2008 (WHO, 2009).

the genus *Flavivirus* in the family Flaviviridae, which also includes yellow fever virus and West Nile Virus. Dengue viruses are divided into four distinct classes, known as serotypes, referred to as DEN-1, DEN-2, DEN-3 and DEN-4. Infection with one serotype provides immunity to other viruses in that same serotype, but no long-term immunity to the other serotypes. Within each serotype, distinct genotypes have been identified which indicates the extensive genetic variability within the dengue serotypes. An individual infected with one of the four serotypes such that the individual have had no prior infection with any of the serotype is said to be infected with *primary infection*. Similarly, an infected individual is said to be infected with *secondary infection* if the individual have had prior infection with any one of the dengue serotypes. Persons living in areas where dengue is an endemic can be infected with three and probably four dengue serotypes during their lifetime (Gubler, 1998a).

The various serotypes of the dengue virus are transmitted to humans through the bites of *Aedes* mosquitoes. *Aedes aegypti* mosquitoes are the predominant vectors for dengue infection, however *Aedes albopictus* and other *Aedes* species are also able

to transmit dengue with varying degree of efficiency. The mosquitoes acquire the virus when they bite an infected human. The mosquitoes are capable of transmitting dengue if they bite another human immediately or after incubating the infection for eight to twelve days, which is known as the extrinsic incubation period. The mosquitoes remain infected for rest of their lives. Vertical transmission of virus from mother to offspring is thought to be rare in both mosquitoes and humans (Gubler, 1998b; Guzmán and Kouri, 2002; Sabin, 1952; Siler et al., 1926).

Once the dengue virus is inoculated into a human host, it incubates for a period of 4–10 days, the intrinsic incubation period. Following incubation, an infected person enters the acute phase of infection for about 5 days. The host recovers from the infection usually within 7–10 days. Infection with one type of dengue serotype provides lifelong protective immunity to the infecting serotype, and possibly partial short-lived protection from infection with other dengue serotypes (Sabin, 1952). The symptoms of disease vary greatly: mild fever, high fever with severe headache and joint pain, and internal hemorrhaging, circulatory failure and death. The cases are classified in order of increasing severity as dengue fever (DF), dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) (Guzmán and Kouri, 2002; Halstead, 2007). The severity of disease in an individual is determined by several factors such as age, ethnicity, previous dengue infection and possibly the presence of some chronic diseases (Guzmán and Kouri, 2002). Young children can be especially susceptible to dengue morbidity (Guzmán et al., 2002). Most patients who develop severe forms of dengue (DHF or DSS) have had prior infections with one or more dengue serotypes (Halstead, 2007). One explanation for this phenomenon is antibody-dependent enhancement (ADE), where the presence of antibodies to one dengue serotype enhance the replication of viruses from other serotypes, perhaps leading to increased susceptibility to infection or transmission once infected, in addition to increasing the risk

of severe disease (Halstead, 2007). Certain dengue genotypes, particularly those of DEN-2, are thought to be more virulent than others, since more number of cases of DHF have been associated with DEN-2 than with the other serotypes (Rico-Hesse, 2003).

Numerous approaches have been used to understand the epidemiology of dengue fever (DF) and dengue hemorrhagic fever (DHF). There are several ecological characteristics of the dengue virus (DENV) that have often been explored using mathematical models. Various mathematical models for dengue infection have appeared so far, and successfully helped us understand the different aspects of the disease. For example, a series of studies by Focks et al. (1993a,b, 1995, 2000) investigated the quantitative value of models using epidemiological data and simulations, while others focused purely on qualitative patterns motivated by ecological interests (Esteva and Mo Yang, 2005; Esteva and Vargas, 2000b; Ferguson et al., 1999a).

To ensure that a mathematical model captures the essential features of epidemics, it is imperative to validate the model by fitting it to observed data. Several estimation methods have been used in the literature to estimate parameters. Least-squares error fitting has been widely used for vector–host models (Chowell et al., 2007; Mubayi et al., 2010) and recently the expectation maximization (EM) algorithm has seen some use (Duncan and Gyöngy, 2006; Lavielle et al., 2011). In this dissertation, I present and use a Bayesian Markov chain Monte Carlo (MCMC) technique for estimation as it provides a huge amount of modeling flexibility. Compared with other estimation methods, Bayesian MCMC technique has advantage of giving a complete posterior distribution for parameters as posterior distributions that enables easy analysis of model parameters or function of parameters. Moreover being a procedure-based approach, it is easy to implement.

Among various models that have been used to study dengue transmission, some

explicitly model the mosquito population (e.g. Esteva and Vargas, 2000a; Medlock et al., 2009; Wearing and Rohani, 2006), while others model the mosquitoes implicitly in the transmission term (e.g. Adams et al., 2006; Cummings et al., 2005; Nagao and Koelle, 2008). In spite of extensive use of both modeling approaches, little guidance exists for which type of model should be preferred. In particular, I found no comparison in the literature of how well these models explain observed incidence. I studied the impact of these modeling assumptions on the dynamics of dengue to fit dengue hemorrhagic fever (DHF) data from Thailand to simple dengue models with and without explicitly modeling mosquitoes using Bayesian MCMC estimation. I found the parameter estimates obtained from both models consistent with previous studies. Most importantly, model selection found that the model with implicit mosquitoes was substantially better than the model with explicit mosquitoes for the DHF data from Thailand. Therefore, explicitly incorporating the mosquito population may not be necessary in modeling dengue transmission.

In mathematical modeling of diseases, the basic reproduction number (R_0) that is total number of secondary cases generated by one case during its infectious period in a completely susceptible population is an important measure as it provides a threshold to determine the amount of effort that is necessary to prevent an epidemic or to eliminate a disease from a population. The basic reproduction number for dengue transmission in Thailand has been estimated by various studies (e.g. Adams et al., 2006; Chao et al., 2012; Cummings et al., 2009; Nagao and Koelle, 2008). However; estimates of R_0 for dengue transmission vary considerably between these studies. These R_0 estimates vary from low values of around 1–3 to higher values of 10–12. The estimates of R_0 from my study of estimating parameters using Bayesian MCMC were in the range of 1–3. This variability in the estimates of R_0 is not well understood. Inaccurate estimates of R_0 may lead to incorrect assessment of disease risk and so it

is important to investigate this variability in R_0 estimates among studies. In order to understand the likely reasons for this variability, we fitted a simple dengue model to dengue incidence for varying values of reproduction number R_0 and found that different levels of population immunity may lead to wide variations in estimates of the basic reproduction number. More specifically, there seems to be logarithmic type relationship between the level of population immunity and the basic reproduction number, so estimating the level of population immunity and the basic reproduction number together may not provide accurate model parameters. Thus in order to fit dengue models to observed data from Thailand, it may be necessary to use appropriate estimates for the proportion of the population immune to infection at the beginning of the epidemic.

With the expanding geographic distribution and increased disease incidence in past several decades, the prevention and control of dengue infection has become very important. Unfortunately, tools available to prevent dengue infection are very limited. For many years, some viral diseases have been controlled using vaccines, however a dengue vaccine is not yet available. Therefore, in order to reduce or prevent dengue virus transmission, there is currently no alternative to vector control. The dengue vector control programs in most endemic countries have been frequently found insufficient, ineffective or both. The low success rate of vector control, the continuing spread of dengue and the increasing incidence of dengue call for a safe, effective and affordable vaccine. The ideal dengue vaccine should be affordable, free of side effects, and should induce life-long protection against infection with any of the four dengue serotypes (i.e. tetravalent) (Guzmán et al., 2010).

Recently, significant progress has been made in the development of vaccine candidates and several vaccine candidates are showing promise in clinical studies (Coller and Clements, 2011). A vaccine candidate from Sanofi Pasteur showed effi-

cacy of 30% and protection against 3 of the 4 serotypes of dengue (Halstead, 2012). It is expected that a licensed vaccine for dengue will be available in less than 10 years (Guzmán et al., 2010). Once vaccine is available, policy makers will need to develop suitable policies to allocate the vaccine. Mathematical models of dengue transmission predict complex temporal patterns in prevalence, driven by seasonal oscillations in mosquito abundance. In particular, vaccine introduction may induce a transient period immediately after vaccine introduction where prevalence can spike higher than in the pre-vaccine period. These spikes in prevalence could lead to doubts about the vaccination program among the public and even among decision makers, possibly impeding the vaccination program. Using simple dengue-transmission models, I found that large transient spikes in prevalence are robust phenomena that occur when vaccine coverage and vaccine efficacy are not either both very high or both very low. Despite the presence of transient spikes in prevalence, the models predict that vaccination does always reduce the total number of infections in the 15 years after vaccine introduction. I concluded that policy makers should prepare for spikes in prevalence after vaccine introduction to mitigate the burden of these spikes and to accurately measure the effectiveness of the vaccine program.

The chapters hereafter are organized as follows: In chapter 2, I present the Bayesian MCMC approach to estimate parameters of deterministic mathematical models. Using artificial incidence data, I show that the method can accurately estimate the parameter values in epidemic models. In chapter 3, I apply the Bayesian MCMC approach to vector–host and SIR-type models of dengue to estimate parameters using monthly DHF-case data from Thailand and perform model selection to chose a model that fits observed data more parsimoniously. In chapter 4, the variability in the estimates of the basic reproduction number for dengue transmission in Thailand is investigated using observed data from Thailand. The short- and long-

term impacts of vaccine introduction and in particular, possibility of large spikes in prevalence immediately after introduction of dengue vaccine is analyzed in chapter 5. Finally, I discuss general conclusions in chapter 6.

Chapter 2

Parameter estimation using Bayeisan MCMC

One of the main goals of the mathematical modeling of infectious diseases is to make reasonable predictions and develop control measures like vaccination and isolation. In order to achieve these goals, it is imperative that the corresponding mathematical model captures the essential features of the course of the disease outbreak. Thus, estimating unknown parameters of the models and their validation by checking whether they fit the observed data becomes very important. It is challenging to resolve this problem for a system of differential equations, which are the typical form for infectious disease models. Generally, there are no closed-form solutions and there are many unknown parameters. In addition, direct data on individual-level parameters such as transmission or susceptibility are extremely limited, due to an inability to feasibly conduct infection experiments, and instead must be estimated from indirect population-level data. Most studies either use point estimates of parameter values derived from clinical and laboratory experiments or from population-level data.

There are several methods that have been used to estimate the parameters of

mathematical models. The method that has been widely used for vector-borne disease models is least-squares error fitting techniques. Least-squares fitting minimizes the sum of squared residuals, a residual being the difference between an observed value and the fitted value provided by the model. Chowell et al. (2007) used least-squares error techniques as one of the methods to estimate the transmissibility of dengue fever during a 2002 epidemic in the Mexican state of Colima, using municipal epidemic data to evaluate the effect of spatial heterogeneity. Mubayi et al. (2010) computed initial estimates of parameters for Kala-azar (leishmaniasis) using least-squares fitting and monthly reported data from Indian state of Bihar for the years 2002 and 2005. They used these parameter estimates to compute, Kala-azar's reproduction numbers for the 21 most affected districts of Bihar and analyzed the impact of underreporting.

Recently, the expectation maximization (EM) algorithm has been used to estimate unknown parameters of disease models. The EM algorithm is an iterative scheme for obtaining maximum likelihood estimates. Duncan and Gyöngy (2006) used the EM algorithm to estimate the parameters of an age-structured model of smallpox and estimated R_0 using data from smallpox deaths in London over the period of 1708 to 1748. A variant of the EM algorithm, stochastic approximation expectation maximization was used by Lavielle et al. (2011) to estimate the parameters of a long-term HIV dynamic model.

A more sophisticated method to find good estimates for the unknown parameters of a model is to take a Bayesian approach that uses Markov chain Monte Carlo (MCMC) simulations to estimate the unknown parameters. In the Bayesian approach the parameter of interest is considered to be a quantity whose variation can be described by a probability distribution. Initially, this is a subjective distribution called the prior distribution, based on the experimenter's belief, which is formulated before the data are seen. A sample is then taken from a population indexed by the param-

eter of interest and the prior distribution is transformed using Bayes's Theorem into the posterior distribution.

Simple models allow the posterior distribution to be calculated explicitly, while other methods must be used for complex models. MCMC methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain with the desired stationary distribution. In Bayesian models, the MCMC method estimates the posterior probability distribution of parameters (Gelman and Rubin, 1996).

MCMC methods are used widely in many different areas of research and many disease modelers have used it to estimate the parameters of epidemic models. Cauchemez et al. (2004) used the Bayesian MCMC method on longitudinal data in order to estimate the main characteristics of infuenza transmission in households. Huang et al. (2006) used a hierarchical Bayesian approach to implement MCMC simulations to estimate the dynamic parameters of a HIV model proposed for characterizing long-term viral dynamics with antiretroviral therapy using longitudinal clinical data. Brownstein et al. (2004) estimated the parameters of human and nonhuman surveillance models for West Nile Virus using MCMC simulations and demonstrated that mosquito surveillance was a more accurate predictor of human risk than monitoring dead and infected wild birds for West Nile Virus. A Bayesian analysis of a dynamic model for the spread of the Usutu virus by MCMC improved the model fit and revealed the structure of interdependencies between model parameters (Reiczigel et al., 2010).

One of the reasons that makes Bayesian MCMC an attractive choice is that it is generally straightforward to implement and provides a huge amount of modeling flexibility. The method enables analysis of all of the model parameters and functions of the parameters. It also has advantage of not giving point estimates of parameters

but distributions for the parameters that capture uncertainty. Moreover, posterior summaries such as means, medians, maxima, minima, credible intervals, etc., can be easily obtained for individual parameters or for joint distributions of parameters. Also, if the data available is limited it reflects in the result by giving wider posterior distributions for the parameters.

2.1 Baysian Markov chain Monte Carlo estimation

The method begins with a mathematical model of disease and appropriate epidemiological data available for the disease. Bayesian inference is then performed, in which the information regarding the parameters of the model from previous studies is regarded as prior knowledge and is combined with the epidemiological data to update the information about the unknown parameters of the model. The Markov chain Monte Carlo method is used to update the parameter distributions.

Given a set of differential equations with parameters θ and the epidemiological data $D_i = D(t_i)$ at the discrete time points $\{t_1, t_2, ...\}$, the aim is to find a set of free parameters so that the model fits the data at those time points. Let $y(t_i \mid \theta)$ be the time series produced by the mathematical model at the same discrete time point, t_i's for which the data is available. An error function is assigned to this data and aim is to minimize the error. The most common error function is the mean square error, which can be written as:

$$E^2 = \sum_i \Big(D_i - y(t_i \mid \theta) \Big)^2 \tag{2.1}$$

Next, a likelihood function for the parameters of the model is constructed. The likelihood is the conditional probability of obtaining the data given the set of parameter

values. The likelihood function is derived by assigning some probability distribution to the error function. There are several ways to form the likelihood function, for example assigning the error function to have a binomial, normal or Poisson distribution. But in practice, a normal distribution works well, so we assumed that the error function (2.1) obeys a normal distribution with zero mean and standard deviation 1 and write likelihood function as

$$L(\theta) = \Pr(D \mid \theta) = \exp(-E^2) \tag{2.2}$$

In order to use Bayesian inference to estimate the parameters, a prior distribution for all the parameters is needed. If we have some prior information about the parameters, then it can be used to assign the prior distribution to the parameters. However, if we do not know any specific, definite information which can be used to assign the prior distribution to the parameters, then we can choose a prior which is flat. In this case the prior is called a noninformative prior. A noninformative prior expresses vague or general information about the parameters such as "the parameter is positive" or "the parameter is less than some limit". We called the prior distribution for the parameters $\Pr(\theta)$.

Finally, the posterior distribution of the unknown parameters is calculated. The posterior distribution is the conditional distribution of the parameter values given the data. By Bayes's Theorem,

$$\Pr(\theta \mid D) = \frac{\Pr(D \mid \theta)\Pr(\theta)}{\Pr(D)}, \tag{2.3}$$

where $P(D \mid \theta)$ is the likelihood function $L(\theta)$ from (2.2), $\Pr(\theta \mid D)$ is the posterior distribution, $\Pr(\theta)$ is the prior distribution and $\Pr(D)$ is called the evidence, which

is an integral of the likelihood over the prior distribution of the parameters:

$$\Pr(D) = \int \Pr(D \mid \theta) \Pr(\theta) \, \mathrm{d}\theta. \tag{2.4}$$

The posterior distribution can be estimated by calculating the expression (2.3). Unfortunately, $\Pr(D)$ is computationally difficult if not impossible to calculate for all but the simplest likelihood as there is not enough information and there are two many possible parameter values.

However, we do know that the posterior distribution is proportional to the likelihood times the prior:

$$\Pr(\theta \mid D) \propto \Pr(D \mid \theta) \Pr(\theta). \tag{2.5}$$

Using Bayes's Theorem (2.5), a Markov chain is formed which asymptotically converges to the posterior distribution by using a simple Metropolis Algorithm. The Metropolis Algorithm is an iterative procedure that uses an acceptance–rejection rule to converge to the required distribution (Gelman et al., 2004). The algorithm is:

1. Start with some initial guess for the parameter values. A starting point θ^0 is chosen randomly from the prior distribution $\Pr(\theta)$.

2. For each iteration $n = 1, 2, 3, ...$

   (a) A new proposed set of parameter values is generated by sampling θ^* from the proposal distribution $J(\theta^* \mid \theta^{n-1})$. The proposal distribution $J(\theta^* \mid \theta^{n-1})$ must be symmetric, i.e. $J(\theta^* \mid \theta^{n-1}) = J(\theta^{n-1} \mid \theta^*)$, for this algorithm.

(b) Using the likelihood function, the ratio of the posterior estimates

$$r = \min\left\{\frac{\Pr(\theta^* \mid D)\Pr(\theta^*)}{\Pr(\theta^{n-1} \mid D)\Pr(\theta^{n-1})}, 1\right\} \tag{2.6}$$

is calculated.

(c) A random uniform number (α) between 0 and 1 is generated. Then the parameter values for this iteration are

$$\theta^n = \begin{cases} \theta^* & \text{if } \alpha < r, \\ \theta^{n-1} & \text{otherwise.} \end{cases} \tag{2.7}$$

This algorithm must be run for enough iterations for the parameter values to converge to the posterior distribution. There are several convergence diagnostics that can be employed to detect whether the chain has converged (Cowles and Carlin, 1996). We used the Gelman–Rubin test (Gelman and Rubin, 1992) for the convergence diagnostic of our simulations, which is based on multiple independent simulated chains. The variances within each chain are compared to the variances between the chains: large deviation between these two variances indicates non-convergence.

The posterior is insensitive to the choice of proposal density, $J(\theta^* \mid \theta^{n-1})$, but the number of iterations until the chain converges may be heavily affected. It is difficult to choose an efficient proposal distribution, but normal distributions have been found to be useful in many problems (Gelman et al., 2004). We used a multivariate normal distribution with mean θ^{n-1} and covariance $\lambda^2\Sigma$ as our proposal distribution,

$$J\left(\theta^* \mid \theta^{n-1}\right) \sim N\left(\theta^{n-1}, \lambda^2\Sigma\right). \tag{2.8}$$

For the multivariate normal proposal distribution, for optimal convergence, proposals

should be accepted at a rate of 0.44 in one dimension and 0.23 in higher dimensions (Gelman et al., 2004). To achieve this, we used a variant of the Metropolis Algorithm that updates the covariance matrix Σ and the scaling factor λ of the proposal distribution after every 500 iterations. We initially chose the covariance to be the $d \times d$ identity matrix ($\Sigma_0 = I$) and the initial scaling factor to be $\lambda_0 = 2.4/\sqrt{d}$, where d is the number of parameters being estimated. After every 500 iterations, the covariance matrix was updated by

$$\Sigma_k = p\Sigma_{k-1} + (1-p)\Sigma^* \tag{2.9}$$

where Σ^* is the covariance of the last 500 parameter values and $p = 0.25$ is the weight given to the old covariance matrix. Similarly, the scaling factor was updated using the Robbins–Monro algorithm (Robbins and Monro, 1951),

$$\lambda_k = \lambda_{k-1} \exp\left(\frac{\alpha^* - \hat{\alpha}}{k}\right), \tag{2.10}$$

where α^* is the acceptance rate for the last 500 iterations and the target acceptance rate is

$$\hat{\alpha} = \begin{cases} 0.44 & \text{for } d = 1, \\ 0.23 & \text{for } d > 1. \end{cases} \tag{2.11}$$

The adaptive algorithm we used has two phases: first, the adaptive phase, which was run until the Gelman–Rubin convergence test passed, and then the fixed phase, where one of chains was chosen randomly and using its last estimates of the parameters along with the updated covariance matrix and scaling factor as starting point, the algorithm was run for next m iterations to sample from the posterior distribution without updating the covariance matrix and the scaling factor. We used the samples from the only last phase for the final inferences.

2.2 Examples

We provide two simple examples of parameter estimation of a simple SIR model using Bayesian MCMC. To show that the method works, we used synthetic data generated for some particular choice of parameter values and compared the estimates from Bayesian MCMC with these values.

A simple SIR model is given by

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= -\beta IS, \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \beta IS - \gamma I, \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma I, \end{aligned} \tag{2.12}$$

where β is contact rate and γ is recovery rate and S, I and R are numbers of susceptibles, infectious and recovered in the population (Anderson and May, 1992) In our first example we estimated the contact rate and recovery rate of this SIR model using the synthetic data. The initial conditions of the model were also estimated along with the parameters in the second example.

2.2.1 Estimating contact rate and recovery rate

We estimated the contact rate β and the recovery rate γ in the aforementioned SIR model. In general, the epidemiological data is available as the number of infectious individuals at different points in time, so we generated a synthetic data set for the number of infectious individuals by solving the SIR model using particular values of parameters β and γ. We chose $\beta = 0.5$ per person per day, $\gamma = 0.3333$ per day and initial conditions $(S(0), I(0), R(0)) = (1, 0.000001, 0)$ to generate the synthetic data for infectious individuals at 180 daily data points using the ODE solver ODE45

Parameter	True Value	Initial guesses
β	0.5	$0.8, 0.4, 0.6, 0.5$
γ	0.3333	$0.2, 0.6, 0.4, 0.5$

Table 2.1: Initial guesses used to estimate β and γ

in Matlab. Now we used our Bayesian MCMC algorithm to estimate the values of contact rate β and recovery rate γ. At each step of our simulations, we generated the simulation data for infectious individuals at the same data points for which we have obtained the synthetic data using same initial conditions and the current estimates of parameter values to form the error function. Following the steps of the method described in section 2.1, the error function (2.1)) was formed . Assuming lack of any prior information about the parameters we wished to estimate, we chose noninformative priors for β and γ. We specified uniform distributions ($U(0, 10)$) for both the parameters. Table 2.1 provides the parameter values used to generate synthetic data as well as initial guesses used for parameter values to begin the simulations.

We ran four parallel MCMC chains and used Gelman–Rubin test along with an upper bound of 5000 iterations for stopping criteria. Once the convergence criteria was met, the estimates of parameters for each MCMC chain against the number of iteration was plotted (Figure 2.1).

All chains converged to distributions around the true values of the parameters. Finally, to make inferences about our result, second phase of the adaptive algorithm was run for next 1000 iterations (Figure 2.2). We used the posterior density to summarize our output in terms of means, medians and credible intervals (Table 2.2). The mean and median estimates were very close to the true values $(0.5000, 0.3333)$, with the mean being slightly closer than the median in this example. The credible intervals included the true parameter values and were relatively narrow, suggesting

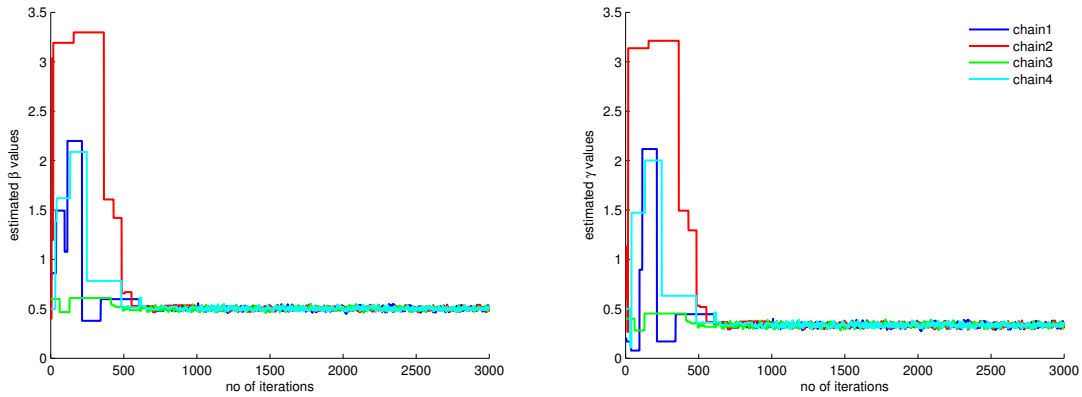

Figure 2.1: Estimates of β and γ vs. number of iterations.

Parameter	True Value	Mean	Median	90% Credible Interval
β	0.5	0.5002	0.5003	$(0.4897, 0.5110)$
γ	0.3333	0.3333	0.3338	$(0.3201, 0.3464)$

Table 2.2: Posterior summary of β and γ. See also Figure 2.2.

the data was sufficient for good estimation of the parameters.

2.2.2 Estimating initial conditions along with parameters

In general, we do not know the initial conditions that should be used for solving the mathematical models, so we must also estimate the initial conditions as well. Here we estimated the initial conditions for the SIR model along with the contact rate β and the recovery rate γ. The three equations in our SIR model resulted into three initial conditions, that is, one for each initial number of susceptibles $S(0)$, initial number of infectious $I(0)$ and initial number of recovered $R(0)$. In the beginning of the epidemic, we assumed that the number of recovered $R(0)$ must be 0, so is not needed to be estimated. We followed the same steps as before to estimate β, γ, $S(0)$ and $I(0)$ in order to fit the synthetic data generated in the previous example to the

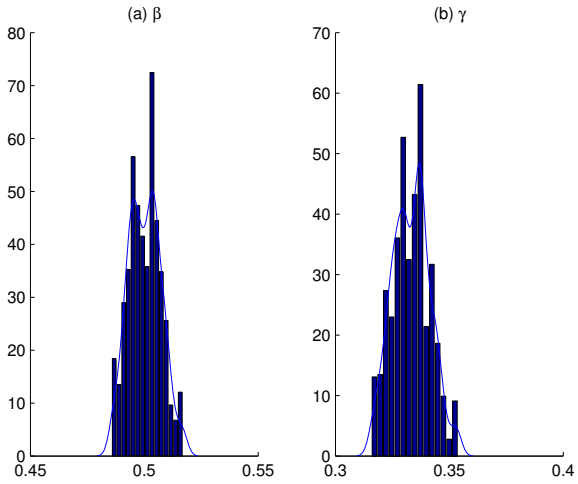

Figure 2.2: Histogram for posterior densities of parameters β and γ. See also Table 2.2.

Parameter	True Value	Initial guesses
β	0.5	$0.8, 0.4, 0.6, 0.5$
μ	0.3333	$0.2, 0.6, 0.4, 0.5$
S_0	1	$0.3, 0.7, 0.4, 0$
I_0	0.000001	$0.00000124, 0.00000226, 0.000000083, 0.000000152$

Table 2.3: Initial guesses used to estimate β, γ, $S(0)$ and $I(0)$

model. We began with over-dispersed initial guesses (Table 2.3) and assigned uniform priors to the parameters and initial conditions.

We ran four parallel MCMC chains and used Gelman–Rubin test along with an upper bound of $20,000$ iterations for stopping criteria. Once the convergence criteria was met, the estimates of parameters and initial conditions for each MCMC chain was plotted (Figure 2.3). The second phase of the adaptive algorithm was run for next 2000 iterations to make inferences about the results. The posterior density was used to summarize the output (Figure 2.3) in terms of means, medians and credible intervals (Table 2.4).

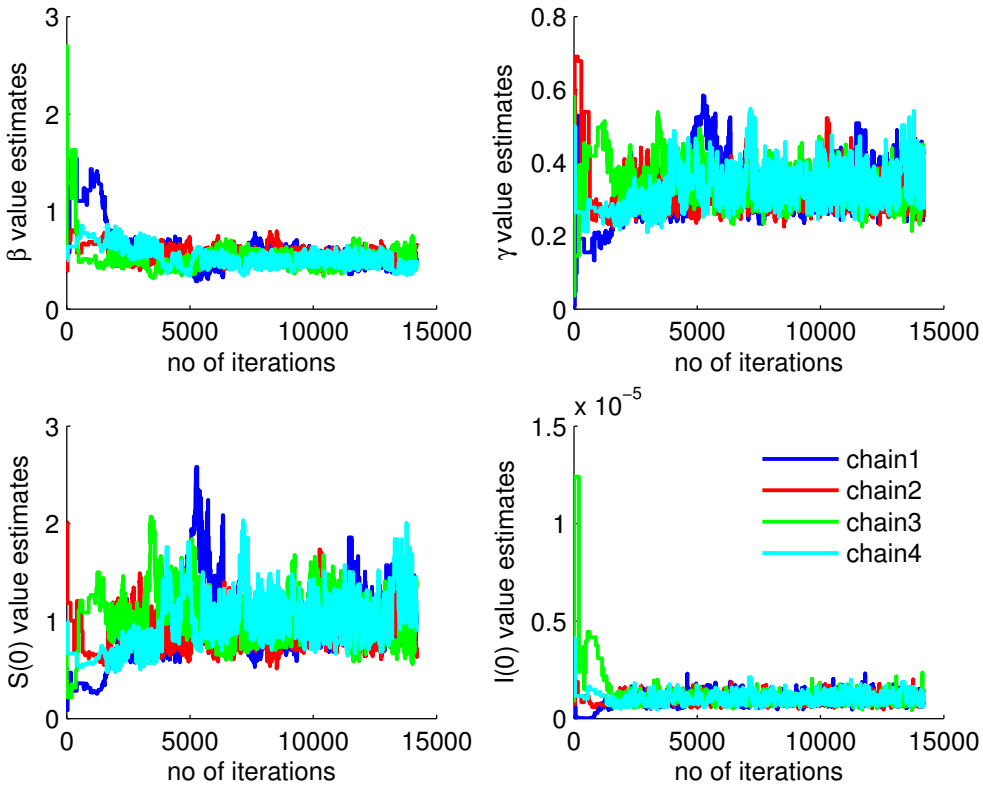

Figure 2.3: Estimates of β, γ, $S(0)$ and $I(0)$ vs. number of iterations

Parameter	True Value	Mean	Median	90% Credible Interval
β	0.5	0.5334	0.5240	$(0.4373, 0.6502)$
γ	0.3333	0.3260	0.3242	$(0.2714, 0.4019)$
S_0	1	0.9297	0.9363	$(0.6970, , 1.2940)$
I_0	0.000001	0.00000118	0.00000121	$(0.00000081, 0.00000162)$

Table 2.4: Posterior summary of β, γ, $S(0)$ and $I(0)$. See also Figure 2.4.

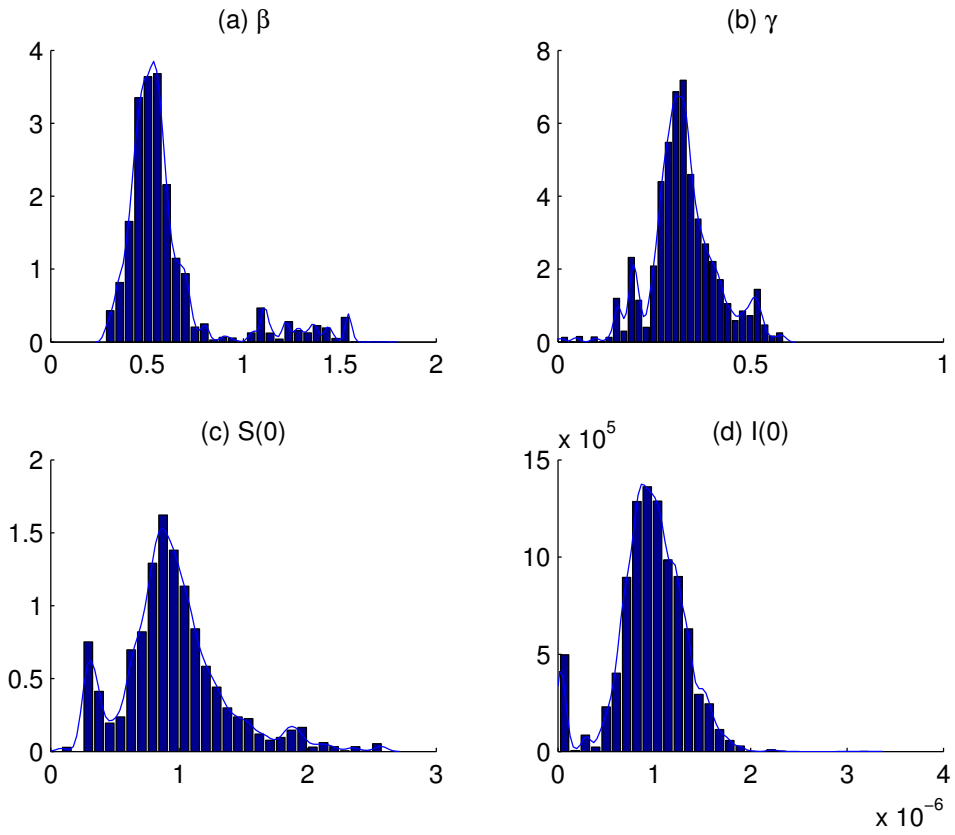

Figure 2.4: Histogram for posterior densities of $\beta, \gamma, S(0)$ and $I(0)$. See also Table 2.4

The mean and median estimates for the parameters β and γ were close to their true values. The mean for β was slightly closer to its true value than the median, whereas the median was slightly closer to the true value of γ than the mean. The mean and median estimates for the initial conditions were very close to the true values $(1, 0.000001)$. All the credible intervals included the true values, but the credible intervals of parameters were wider compared to the ones obtained when only parameters were estimated.

2.3 Summary

We developed and presented a method of Bayesian MCMC estimation for fitting disease models to observed data. In this method, prior information from previous studies regarding the model parameters is combined with epidemiological data to update the information about the model parameters. The connection between the observed data and the parameters is made by a likelihood function. The prior distribution of model parameters are then transformed in the posterior distribution by Bayes's Theorem. In Bayesian models, slight deviations from simple models renders the models intractable analytically. So, we used a MCMC method based on Metropolis algorithm to generate posterior distribution for the model parameters. To improve the speed of convergence to the posterior distribution, a variant of Metropolis algorithm that updates the covariance matrix and scaling factor of jumping distribution after every fixed number of iterations was used. In our simulations, Gelman–Rubin test determined when the convergence was met and was used as stopping criteria.

We demonstrated the use of the Bayesian MCMC estimation by applying it on a simple SIR-type model. We generated artificial incidence data by running the model with fixed parameter values and initial conditions. We performed Bayesian

MCMC estimation on the model and the artificial data to find estimates for the model parameters and initial conditions. The mean and median estimates for the parameters and initial conditions were very accurate and the credible intervals captured the uncertainties in the estimates well.

Parameter estimation using epidemiological data is one of the key aspects of disease modeling. We showed that Bayesian MCMC estimation is a novel method to estimate parameters of disease models and differential equation models in general. The method is easy to implement and has advantage of giving complete posterior distribution for parameters that enables easy analysis of model parameters or function of parameters.

Chapter 3

Comparing vector–host and SIR models for dengue transmission

3.1 Introduction

Dengue infection is one of the leading causes of illness in the tropics and subtropics, where it inflicts substantial health, economic and social burdens (Center for Disease Control and Prevention, Accessed January 14, 2013). Humans are infected with dengue viruses by the bite of an infective female mosquito *Aedes aegypti*, the principal vector of dengue. Once a person gets bitten by an infective mosquito, the virus undergoes an incubation period of about 4 to 7 days, after which the person enters the acute phase of infection. The acute phase can be as short as 2 days and as long as 10 days. If other female *A. aegypti* mosquitoes bite the ill person during this acute phase, those mosquitoes may become infected and subsequently begin the transmission cycle anew. Dengue infection is generally characterized by a sudden onset of fever and other nonspecific signs and symptoms, including frontal headache, body aches, nausea and vomiting (Gubler, 1998a). Symptoms range from mild fever

to high fever with severe headache and joint pain, and even to internal hemorrhaging, circulatory failure and death. Cases are classified, in order of increasing severity as dengue fever, dengue hemorrhagic fever (DHF) and dengue shock syndrome (Medlock et al., 2009). Dengue has been recognized in over 100 countries and an estimated 50–100 million cases of dengue fever and several hundred thousand DHF cases occur yearly, depending on epidemic activity (Guzmán and Kouri, 2002). Particularly, in Thailand, dengue disease incidence has increased from 9 per 100 000 in 1958 to 189 per 100 000 in 1998, with the largest reported incidence of 325 per 100 000 in 1987, making dengue a severe public health problem in Thailand (Nisalak et al., 2003).

Several mathematical models have been proposed to investigate dengue epidemiology, some of which explicitly model the mosquito population (e.g. Esteva and Vargas, 2000a; Medlock et al., 2009; Wearing and Rohani, 2006), while others implicitly model it in the transmission term (e.g. Adams et al., 2006; Cummings et al., 2005; Nagao and Koelle, 2008). Although both kinds of models have been extensively used for dengue, little guidance exists for which type of model should be preferred. In particular, there has been no comparison of how well these models explain observed incidence. In this study, we considered simple dengue models with and without explicitly modeling mosquitoes, fit both models to DHF incidence data, and used model selection to compare the models.

Fitting models to data validates the model as well as provides estimates of unknown model parameters. There are some examples in the literature where dengue models have been fit to data. Chowell et al. (Chowell et al., 2007) estimated the transmissibility of dengue during a 2002 epidemic in the Mexican state of Colima using municipal epidemic data to evaluate the effect of spatial heterogeneity. Ferguson et al. (Ferguson et al., 1999b) used longitudinal incidence of serious dengue disease from Thailand and estimated the basic reproductive number R_0 to gain insight into

the transmission dynamics and epidemiology of dengue. We fit a simple vector–host dengue model as well as an SIR-type dengue model and obtain estimates of unknown parameters like recovery rate, probability of severe form of disease, mosquito mortality rate, etc.

The goal of the present study is to understand the impact of some modeling assumptions on quantifying estimates of epidemiological metrics for dengue. We applied Bayesian Markov chain Monte Carlo (MCMC) estimation on a simple vector–host dengue model as well as an SIR-type dengue model to estimate model parameters using monthly DHF incidence data in Thailand for January 1984 to March 1985. The Bayesian MCMC techniques that we used in this study have been commonly used to estimate model parameters of infectious diseases (Brownstein et al., 2004; Cauchemez et al., 2004; Huang et al., 2006; Reiczigel et al., 2010). We use the posterior distribution of the model parameters obtained from Bayesian MCMC to perform uncertainty and sensitivity analysis of basic reproductive number R_0 and thereafter, use model selection on a set of vector–host and SIR models to find a model which agrees with the data most parsimoniously.

3.2 Methods

We built two mathematical models of dengue transmission, one in which the mosquitoes are explicitly tracked and another without explicit mosquito populations. We then used Bayesian MCMC to fit DHF data from Thailand to these two models. In this section, we outline the data source, models and methods and refer to more detailed descriptions in Appendix A.

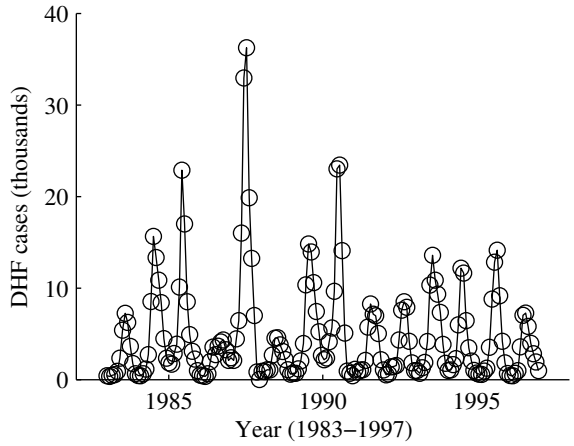

Figure 3.1: Monthly dengue hemorrhagic fever (DHF) incidence in Thailand from 1983 to 1997.

3.2.1 Data Source

The Thailand Ministry of Public Health have been recording the number of DHF cases since 1972. Cases are diagnosed using criteria established by the World Health Organization. We obtained the monthly incidence of DHF for Thailand from 1983 to 1997 (Figure 3.1) (Johns Hopkins Center for Immunization Research, Accessed January 14, 2013). We chose one epidemic, from January 1984 to March 1985 (Figure 3.2), to fit the dengue models: this particular epidemic was chosen as a clear, representative example among this data. More specifically, we used the cumulative monthly number of DHF cases for the period January 1984 to March 1985. Cumulative incidence is generally smoother than the original incidence data and thus easier to fit and it also easily handles delayed reporting on holidays and weekends.

3.2.2 Vector–host model

The Ross–Macdonald model, originally developed for malaria, is a standard mathematical model for vector-borne pathogens that tracks infections in both humans

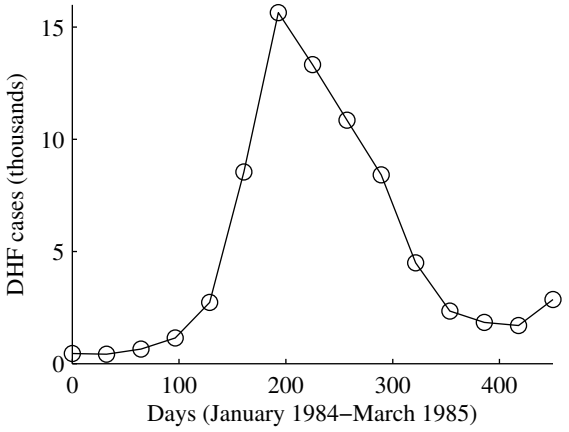

Figure 3.2: Monthly DHF incidence in Thailand from January 1984 to March 1985.

and mosquitoes (Macdonald, 1957). Following this framework, we built a vector–host model for dengue consisting of three human host compartments, susceptible (the number of susceptible humans is H_S), infectious (H_I) and recovered (H_R), and two mosquito compartments, susceptible (V_S) and infectious (V_I). Mosquitoes do not recover from infection. The model is the system of differential equations

$$\begin{aligned}
\frac{\mathrm{d}H_S}{\mathrm{d}t} &= B_H - mc\beta_H \frac{V_I}{V} H_S - \mu_H H_S, \\
\frac{\mathrm{d}H_I}{\mathrm{d}t} &= mc\beta_H \frac{V_I}{V} H_S - \gamma_H H_I - \mu_H H_I, \\
\frac{\mathrm{d}H_R}{\mathrm{d}t} &= \gamma_H H_I - \mu_H H_R, \\
\frac{\mathrm{d}V_S}{\mathrm{d}t} &= B_V - c\beta_V \frac{H_I}{H} V_S - \mu_V V_S, \\
\frac{\mathrm{d}V_I}{\mathrm{d}t} &= c\beta_V \frac{H_I}{H} V_S - \mu_V V_I,
\end{aligned} \tag{3.1}$$

where $H = H_S + H_I + H_R$ and $V = V_S + V_I$ are the human and mosquito population sizes, respectively. A susceptible human gets infected with force of infection $mc\beta_H \frac{V_I}{V}$, where m is number of mosquitoes per person, c is mean rate of bites per mosquito and β_H is the mosquito-to-human transmission probability per bite. Infectious people

recover at rate γ_H. The force of infection for mosquitoes is $c\beta_V \frac{H_I}{H}$, where β_V is the human-to-mosquito transmission probability. For simplicity, we ignored disease-induced mortality in both humans and mosquitoes, which is small (Burattini et al., 2008). Because we only fit the model to an epidemic lasting about a year, we assumed the human population was constant size by using the birth rate $B_H = \mu_H H$. We also assumed the mosquito population was constant size ($B_V = \mu_V V$), neglecting seasonal fluctuations for simplicity.

Standard mathematical analysis of the model (Appendix A.1) shows that the basic reproductive number, the number of new human infections caused by a single infected human in an otherwise completely susceptible population, is

$$R_0 = \frac{mc^2 \beta_H \beta_V}{\mu_V(\mu_H + \gamma_H)}. \tag{3.2}$$

In addition, there are two equilibrium points, the disease-free equilibrium and the endemic equilibrium. An equilibrium point is asymptotically stable if nearby orbits converge to it as time increases, and it is globally asymptotically stable if all orbits, not just those nearby, converge to the equilibrium (Alligood et al., 1996). For $R_0 > 1$, the disease-free equilibrium is unstable and the endemic equilibrium is locally asymptotically stable. The disease-free equilibrium is globally asymptotically stable when $R_0 \leq 1$ (and the endemic equilibrium is out of the relevant state space, having H_I and V_I negative, and unstable).

To simplify the parameter estimation, rather than fitting human mortality rate along with the other parameters, we fixed $\mu_H = 1/69$ y^{-1} based on the average human duration of life in Thailand in 1984 of about 69 years (World Bank, Accessed January 14, 2013). The remaining unknown parameters are the human recovery rate (γ_H), the mosquito mortality rate (μ_V), the probability of DHF (p), the mosquito

biting rate (c), the number of mosquitoes per person (m), the mosquito-to-human transmission probability (β_H) and the human-to-mosquito transmission probability (β_V). The biting rate, c, always appears in the model multiplied with either β_H or β_V. Similarly, m always appears multiplied with β_H. Therefore, only 2 of these 4 parameters can be separately estimated, which we chose to be $\beta_{aH} = mc\beta_H$ and $\beta_{aV} = c\beta_V$. In addition, the initial proportion of humans recovered in the host population ($h_R(0) = H_R(0)/H$) as well as initial proportion of mosquitoes infected in the vector population ($v_I(0) = V_I(0)/V$) are unknown and must be determined. Thus, we estimated a total of 5 unknown parameters and 2 initial conditions for the vector–host model. We used the incidence data for January 1984 and Thailand's population in year 1984 to calculate initial conditions for initial proportion of hosts infected, i.e. $h_I(0) = H_I(0)/H(0)$, where $H_I(0) = 454$ and $H(0) = 46\,806\,000$. Since both the human and mosquito populations are constant, initial proportions of susceptible humans and mosquitoes were calculated using the other initial conditions, i.e. $h_S(0) = 1 - h_I(0) - h_R(0)$ and $v_S(0) = 1 - v_I(0)$.

3.2.3 SIR model

Dengue transmission has been extensively modeled using SIR-type models, which only explicitly track human infections (e.g. Adams et al., 2006; Cummings et al., 2005; Nagao and Koelle, 2008). These SIR models are simpler than vector–host models, making analysis and parameter estimation easier. SIR models for dengue have typically been constructed directly (e.g. Cummings et al., 2005). Alternately, an SIR model can be derived from a vector–host model by assuming that infection dynamics in the vector are fast compared to those of the host, a quasi-equilibrium approximation (Keeling and Rohani, 2011).

We used a standard SIR model,

$$\begin{aligned}
\frac{\mathrm{d}H_S}{\mathrm{d}t} &= B_H - \beta\frac{H_I}{H}H_S - \mu_H H_S, \\
\frac{\mathrm{d}H_I}{\mathrm{d}t} &= \beta\frac{H_I}{H}H_S - \gamma_H H_I - \mu_H H_I, \\
\frac{\mathrm{d}H_R}{\mathrm{d}t} &= \gamma_H H_I - \mu_H H_R,
\end{aligned} \tag{3.3}$$

where $H = H_S + H_I + H_R$ is the human population size. Again, we kept the population size constant by setting the birth rate to $B_H = \mu_H H$. A susceptible person gets infected with force of infection $\beta\frac{H_I}{H}$, where β is the composite human-to-human transmission rate. Comparing the equilibria of the vector–host model and the SIR model (Appendix A.2) provides β in terms of the parameters of the vector–host model:

$$\beta \approx \frac{mc^2\beta_H\beta_V}{\mu_V}. \tag{3.4}$$

SIR model (3.3) is a standard mathematical model for *directly* transmitted pathogens like influenza and has been thoroughly analyzed (e.g. Hethcote, 2000). The basic reproductive number is

$$R_0 = \frac{\beta}{\mu_H + \gamma_H}. \tag{3.5}$$

As with vector–host model (3.1), there are two equilibrium points, the disease-free equilibrium and the endemic equilibrium: for $R_0 > 1$ the disease-free equilibrium is unstable and the endemic equilibrium is globally stable, while the disease-free equilibrium is globally asymptotically stable for $R_0 \leq 1$ (with the endemic equilibrium having $H_I < 0$ and being unstable).

The unknown parameters are the transmission rate (β), the recovery rate (γ_H) and the probability of DHF (p), along with the initial proportion of humans

recovered ($h_R(0) = H_R(0)/H$). As in the vector–host model, we used the fixed value for the human mortality rate $\mu_H = 1/69$ y^{-1} to simplify the parameter estimation. Again, the initial proportion of infected humans is given by $h_I(0) = H_I(0)/H(0)$ with $H_I(0) = 454$ and $H(0) = 46\,808\,000$. Like the vector–host model, the other initial condition is $h_S(0) = 1 - h_I(0) - h_R(0)$.

3.2.4 Bayesian Markov chain Monte Carlo estimation

To estimate the unknown parameters, we used a Bayesian MCMC technique. Bayesian inference uses prior information of the model parameters from previous studies, which is then combined with new data to generate estimates in the form of a probability distribution for the parameters. More precisely, for parameters θ and data D, with the prior parameter distribution $\Pr(\theta)$ and likelihood function $\Pr(D \mid \theta)$, the posterior parameter distribution $\Pr(\theta \mid D)$ is given by Bayes's Theorem:

$$\Pr(\theta \mid D) = \frac{\Pr(D \mid \theta)\Pr(\theta)}{\Pr(D)} \tag{3.6}$$

or, alternately,

$$\Pr(\theta \mid D) \propto \Pr(D \mid \theta)\Pr(\theta). \tag{3.7}$$

Because there are no general closed-form solutions, MCMC or other methods must be used to generate approximate samples from the posterior parameter distribution $\Pr(\theta \mid D)$.

The connection between the data and the parameters is made by the likelihood function $L(\theta) = \Pr(D \mid \theta)$, which is the conditional probability of obtaining the data (D) for the given parameter values (θ). Therefore, $L(\theta)$ needs to be maximized to obtain best-fit parameter set. In our case, the likelihood function is derived from

the vector–host and SIR models, the solutions to which provide estimates of the DHF monthly incidence data. We added a compartment to each model to calculate the cumulative number of DHF infections (H_C). We assumed that a fraction p of infections were diagnosed as DHF, with p constant in time. We added differential equations for the H_C compartment,

$$\frac{\mathrm{d}H_C}{\mathrm{d}t} = pmc\beta_H \frac{V_I}{V} H_S, \tag{3.8}$$

for the vector–host model, and

$$\frac{\mathrm{d}H_C}{\mathrm{d}t} = p\beta \frac{H_I}{H} H_S, \tag{3.9}$$

for the SIR model, which are precisely the rates of new infections multiplied by p. The "ode15s" function in Matlab was used to numerically solve the vector-host model (3.1) & (3.8) and the SIR model (3.3) & (3.9). These numerical solutions give the predicted monthly cumulative DHF incidence, $y_i = H_C(t_i)/H$, where $t_i = 0, 30, 60, ...$ days. Using the least-squares error between the cumulative DHF data D_i and the model prediction,

$$E^2 = \sum_{i=1}^{15} \Big(D_i - y_i(\theta) \Big)^2, \tag{3.10}$$

we assumed the errors were Gaussian, giving the likelihood function

$$L(\theta) = \Pr(D \mid \theta) = \exp\left(-E^2\right). \tag{3.11}$$

For the prior parameter distributions, we assigned wide uniform distributions, with ranges chosen to represent our general understanding about where the parameter values may lie. In the absence of any information on parameters estimates, we used

least-squares fitting to find best-guess estimates of parameters. Estimates of γ_H, μ_V and μ_H from the literature ($\gamma_H = 1/7\ d^{-1}$, $\mu_V = 1/14\ d^{-1}$ and $\mu_H = 1/69\ y^{-1}$) were used and the vector–host model was fitted to the data using least squares in Berkeley Madonna to find initial point estimates $\beta_{aH} = 0.002$, $\beta_{aV} = 1.8$ and $p = 0.04$. We used these initial point estimates to form uniform priors for these parameters such that their point estimates lie inside the range of priors. For the transmission term β of the SIR model, we simply choose a very wide uniform prior. Where parameters were common to both models, both models used the same prior (Table 3.1).

To generate the posterior parameter distribution, we used an MCMC method based on the Metropolis algorithm using a Gaussian jumping distribution with an adaptive covariance matrix. For each model, we simulated 4 independent MCMC chains and used the Gelman–Rubin test to determine when the chains had converged to the stationary distribution, i.e. the parameter posterior distribution. The Gelman–Rubin test signals convergence when the variance between independent chains is similar to the variance within the chains. (See section 2.1 for more details.) Once the Gelman–Rubin test passed, we continued sampling from one of the chains for 10 000 more iterations without updating the covariance matrix, saving every 5th iterate as the posterior parameter distribution.

3.3 Results

We estimated 7 total parameters for the vector–host model and 4 total parameters for the SIR model by Bayesian MCMC using the cumulative DHF incidence data. Both models with their maximum-likelihood (ML) parameter estimates fit the data well (Figure 3.3), with the vector–host model fitting slightly better. (More on model fitting and model selection below.)

The estimates of the human recovery rate were similar for both models (Figure 3.4(d) and Table 3.1). The average duration of human dengue infection is between 2 and 7 days approximately, with ML estimates of about 2 to 3 days. The initial proportion of humans recovered ($h_R(0)$) was estimated to be small in both models, indicating that the human populations were almost entirely susceptible when the outbreak started.

Estimates of the probability of DHF differed somewhat between models: ML of around 3 DHF cases per 1000 infections from the vector–host model and around 14 DHF cases per 1000 infections from the SIR model. The vector–host model includes several parameters not present in the SIR model. From the vector–host model, the range of average lifespan of mosquitoes ($1/\mu_V$) was found to be approximately 13 to 26 days, with ML estimate of about 15 days. The initial proportion of mosquitoes infected was very small (ML of about 0.5%), so that the outbreak had just started in the mosquitoes as well as the humans.

The transmission rates are not common between the models, but comparison of equilibria of both the models allowed us to compare the composite transmission rate β from the SIR model with $\beta = \beta_{aH}\beta_{aV}/\mu_V$ for the vector–host model (Figure 3.4(c)). Although, the ML estimates of β from both models are similar, the distribution from the vector–host model has more weight at higher values of β than the distribution from the SIR model: e.g. the median estimates are 0.4882 and 0.3243 respectively.

The basic reproductive number (R_0), the expected number of secondary cases produced by a single infection in a completely susceptible population, was calculated using equations (3.2) and (3.5) for the respective models, for each MCMC parameter sample (Figure 3.4(i)). For all parameter samples, $R_0 > 1$ as expected since the data show an epidemic, but the R_0 values from the vector–host model (ML: 1.57) are higher than from the SIR model (ML: 1.10).

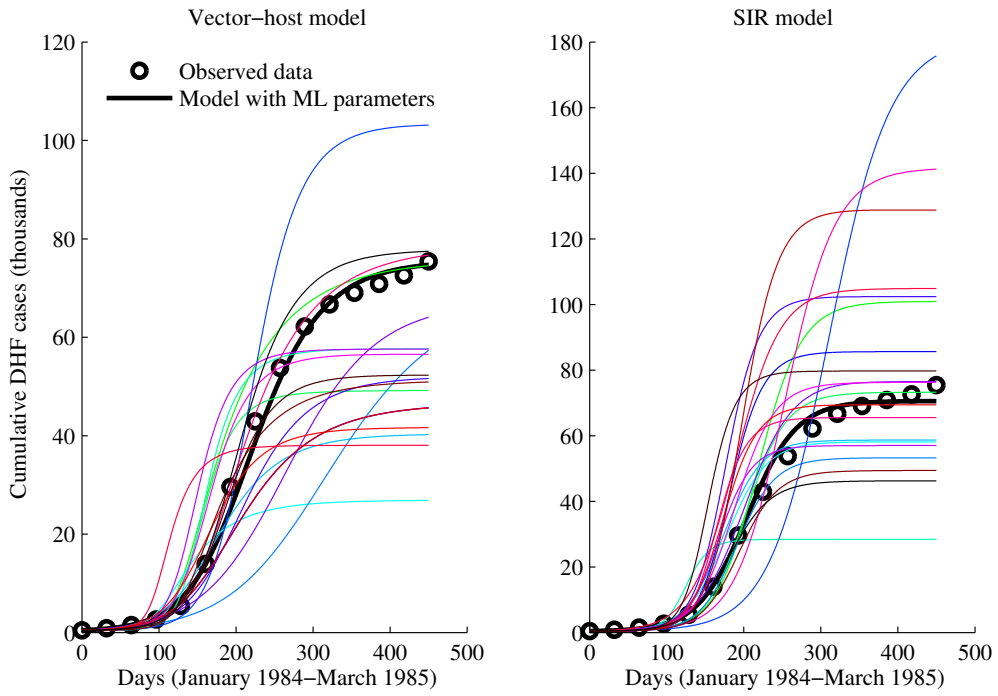

Figure 3.3: Fits of the vector–host and SIR models. Shown are the cumulative DHF cases from the data (black circles), and from the models with the maximum–likelihood parameter estimates (thick black curves) and 20 samples from the posterior parameter distribution (thin color curves).

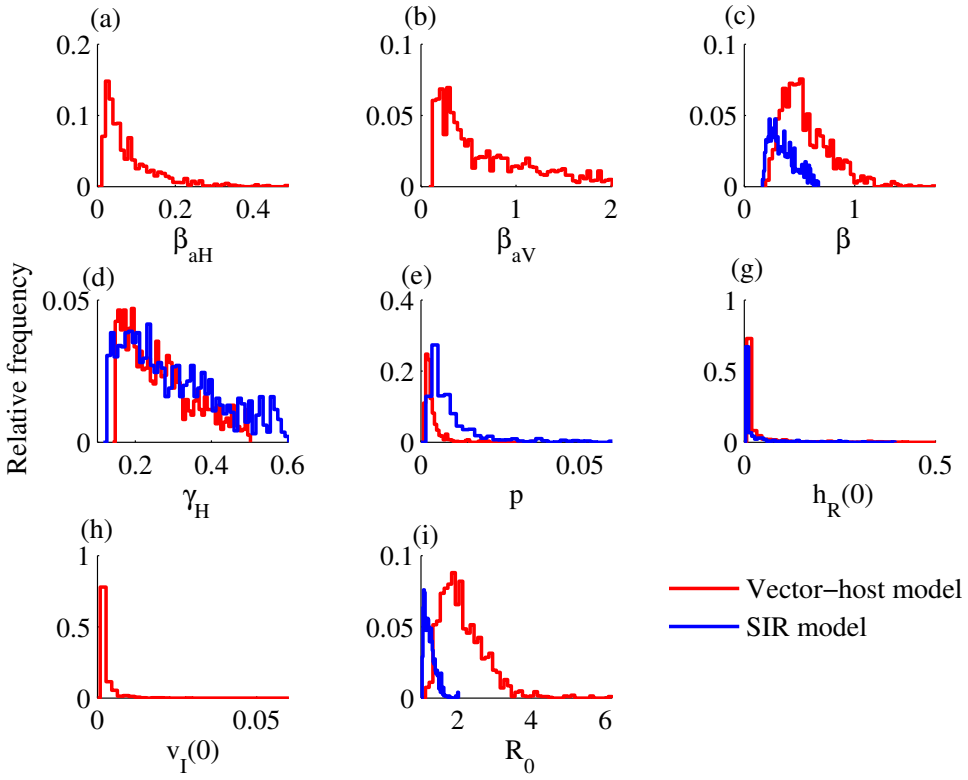

Figure 3.4: Posterior parameter densities for the vector–host and SIR models. (For the vector–host model, $\beta = \beta_{ah}\beta_{av}/\mu_v$.)

Parameter	Prior	Model	Posterior		
			ML	**Median**	**90% CI**
$\beta_{aH}\ (\mathrm{d}^{-1})$ Mosquito-to-human transmission rate	$U(0,1)$	VH	0.0686	0.0521	$(0.0146, 0.2241)$
$\beta_{aV}\ (\mathrm{d}^{-1})$ Human-to-mosquito transmission rate	$U(0.1,2)$	VH	0.4307	0.4867	$(0.1299, 1.6821)$
$\beta\ (\mathrm{d}^{-1})$ Composite transmission rate	$U(0,10)$	VH	0.4881	0.4882	$(0.2782, 0.9364)$
		SIR	0.5718	0.3243	$(0.1931, 0.5805)$
$\gamma_H\ (\mathrm{d}^{-1})$ Human recovery rate	$U(0.1,0.6)$	VH	0.3104	0.2480	$(0.1521, 0.4440)$
		SIR	0.5211	0.2650	$(0.1347, 0.5315)$
p Probability of DHF	$U(0,0.1)$	VH	0.0028	0.0022	$(0.0010, 0.0086)$
		SIR	0.0137	0.0057	$(0.0018, 0.0354)$
$\mu_H\ (\mathrm{y}^{-1})$ Human mortality rate	$1/69$	VH	—	—	—
		SIR	—	—	—
$\mu_V\ (\mathrm{d}^{-1})$ Mosquito mortality rate	$U(0.01,0.1)$	VH	0.0605	0.0531	$(0.0378, 0.0781)$
$h_R(0)$ Initial humans recovered	$U(0,1)$	VH	0.0067	0.0020	$(0.0000, 0.1320)$
		SIR	0.0332	0.0019	$(0.0000, 0.1363)$
$v_I(0)$ Initial mosquitoes infected	$U(0,1)$	VH	0.0009	0.0005	$(0.0000, 0.0056)$
R_0 Basic reproductive number	—	VH	1.5724	1.9733	$(1.3556, 3.2059)$
		SIR	1.0972	1.1989	$(1.0523, 1.5243)$

Table 3.1: Posterior summary of parameter estimates. For simplicity, μ_h was not estimated. R_0 is not a parameter, but rather a function of the other parameters. $U(a,b)$ is the uniform distribution between a and b. For parameters common to both models, the same prior was used for both models. "ML" is maximum-likelihood estimate; "CI" is credible interval; "VH" is the vector–host model. Parameter units are given in parentheses.

Because R_0 is an important metric for an infectious disease, we performed uncertainty and sensitivity analysis of R_0 for both models using partial rank correlation coefficients (PRCC). The PRCC measures the independent effect of each input parameter on R_0, assuming the parameters to be independent (Blower and Dowlatabadi, 1994). The ordering of these PRCCs directly corresponds to the level of statistical influence, the impact that uncertainty in the estimate of a parameter has on the variability of R_0 (Sanchez and Blower, 1997). We used the "prcc" function of the R library epiR (Stevenson, Accessed March 10, 2013).

For both models, all of the parameters were significantly different from 0 (p-value $< 2.5 \times 10^{-135}$). For the vector–host model, all parameters except $h_R(0)$ and $v_I(0)$ were most influential in determining the magnitude of R_0 ($|\text{PRCC}| > 0.5$), while only β and γ_H for the SIR model were most influential on the magnitude of R_0. A positive PRCC value indicates that an increase in that parameter leads to an increase in R_0, while a negative value shows that increasing that parameter decreases R_0. For the parameters that appear explicitly in the R_0 equations (3.2) and (3.5), the signs of the PRCCs were as expected. Of the remaining parameters, p and $v_I(0)$ have a negative influence on R_0, while $h_R(0)$ has a positive influence on R_0.

Parameter estimates for both models suggest that the initial proportion of humans recovered and the initial proportion of vectors infectious are very small. As a result, we tried fitting both models by fixing $h_R(0) = 0$ and $v_I(0) = 2h_I(0)$ and estimating the other parameters in order to decrease the complexity of the models (Figure 3.6). We fit the vector–host model by fixing $h_R(0)$ only, fixing $v_I(0)$ only and fixing both $h_I(0)$ and $v_I(0)$. Similarly we fit the SIR model by fixing $h_R(0)$.

We used the Akaike Information Criterion (AIC) to compare the competing 6 models (Table 3.2). The AIC is a measure of the relative goodness of fit of a statistical model, balancing fit with number of parameters, finding the simplest model

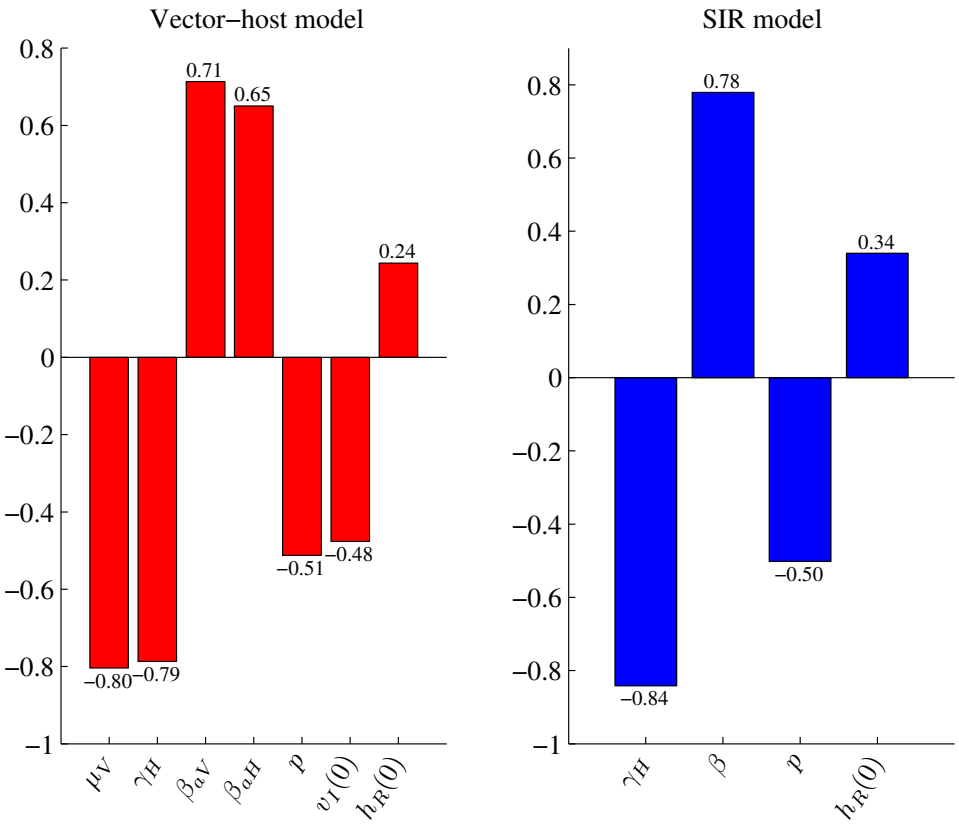

Figure 3.5: PRCCs for the effect of each parameter on R_0 for the vector–host and SIR models.

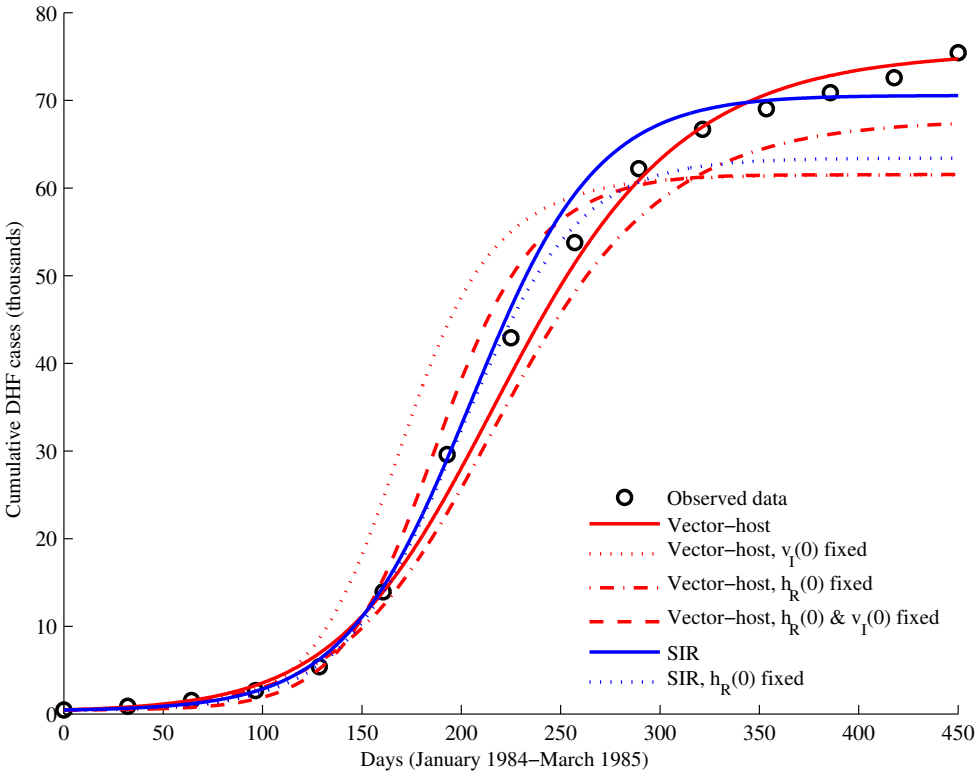

Figure 3.6: The maximum likelihood fits of models to the data.

Model	df	Log-likelihood	AIC	ΔAIC	Akaike weight
Vector–host	7	−0.0779	14.1559	7.7212	0.0130
Vector–host, $v_I(0)$ fixed	6	−0.6847	13.3695	6.9348	0.0193
Vector–host, $h_R(0)$ fixed	6	−0.1173	12.2346	5.7999	0.0341
Vector–host, $h_R(0)$ & $v_I(0)$ fixed	5	−0.5835	11.1669	4.7322	0.0581
SIR	4	−0.1020	8.2039	1.7692	0.2558
SIR, $h_R(0)$ fixed	3	−0.2173	6.4347	0	0.6196

Table 3.2: Comparison of the vector–host and SIR models with and without fixed initial conditions. "df" is degrees of freedom, i.e. number of parameters.

that best approximates the true, but unknown mechanisms generating the data. The SIR model with fixed $h_R(0)$ had the minimum AIC value, implying this model was the best among the models. The difference in AIC between the best model and the others (ΔAIC) gave "considerably less support" for all the vector–host models and "substantial support" for both SIR models (Burnham and Anderson, 2002). Alternatively, Akaike weights provide the probability that a model is the best among the set of candidate models. The Akaike weight for the SIR model with fixed $h_R(0)$ gave 62% probability of it being the better model whereas the SIR model where $h_R(0)$ is also estimated was 26% likely to be the better model. There was only a 12% probability that any of the vector–host models was best.

3.4 Discussion

The fitting of dengue incidence data from Thailand to simple vector–host and SIR model provided estimates of model parameters. The estimates of human recovery rate from both the models suggest a recovery period of 2 to 7 days, which is consistent with the estimates used in previous studies (Cummings et al., 2005; Gubler, 1998a). The estimates of the probability of DHF from the vector–host model and the SIR

model are that about 3 and 14 out of 1000 infections develop into DHF, respectively for the two models. Based on the annual number of dengue infections and DHF cases (Gubler, 1998a), 5 out of 1000 infections develop into DHF.

The ML estimate of the basic reproductive number (R_0) for the SIR model is 30% smaller than the estimate for the vector–host model. This is driven by the recovery rate (γ_H) being estimated as 68% larger in the SIR model. The MLE probability of DHF (p) is 4.9 times larger for the SIR model. The two models—one with high R_0 and low p, the other with low R_0 and high p—both fit the data well. The PRCC result showing a negative influence of p on R_0 confirms the relationship between these two parameters.

Dengue had been causing annual outbreaks in Thailand for some time prior to the 1984 epidemic (Nisalak et al., 2003). Despite this, our estimates of the initial proportion of people immune ($h_R(0)$) from both models are very small. A high birth rate (United Nations, Accessed March 25, 2013) and the reemergence of dengue serotypes 3 and 4 (Nisalak et al., 2003) could explain this low immunity. In addition, mosquito seasonality may be important to explain the monthly variation in dengue incidence (Wearing and Rohani, 2006), and keeping the mosquito population constant for simplicity in our model could have contributed towards small estimates of $h_R(0)$.

The vector–host model fits the data slightly better than the SIR model, but the fewer number of parameters results in the SIR model being strongly selected by the AIC. Alternative measures for model selection like the Bayesian Information Criterion and the Deviance Information Criterion more strongly penalize the number of parameters than AIC, so we expect the result of model selection to remain unchanged. This suggests that incorporating mosquito populations explicitly in dengue models may not be necessary to estimate incidence.

We believe that for any vector-borne pathogen, explicitly including vector

populations may generally be unnecessary to model prevalence or incidence in human or other primary host. We expect that models with and without explicit vectors will fit primary-host data about equally well and then the fewer parameters of the model without explicit vectors will result in it being preferred by formal model selection. Other factors like seasonality in mosquito abundance may be crucial to fit some long-term data (e.g. Figure 3.1), which could result in explicit-vector models fitting the data significantly better than implicit-vector models. In addition, explicit-vector models are necessary when interventions are targeted at the disease vector, e.g. insecticide or genetically modified mosquitoes. When the desired model output is the effectiveness or cost-effectiveness of an intervention that acts on the primary host, our result suggests that implicit-vector models are likely to be sufficient.

The composite transmission parameter (β) for the vector–host model was obtained from the equilibria of the two models and may not be a good approximation for our comparison of the dynamics of the models. This may explain the difference in the estimates of the composite transmission parameter (β) between the two models. This is reinforced by the fact that the SIR model fits the observed data well, but not for the same β values as the vector–host model. Thus, in addition to being preferred by model selection, use of the SIR model is justified when only the equilibrium values are of interest.

We chose to use DHF cases because the data was available monthly, while we are only aware of annually reported DF cases (Nisalak et al., 2003). Moreover, a person infected with DHF is more likely to visit hospital due to the severity of the disease, and so more likely to be diagnosed and reported. Therefore, data on reported DHF cases may be more accurate to actual DHF cases than DF data is to DF cases.

We used a Bayesian MCMC technique for estimation, though other estimation methods have also been used in the literature. In particular, least-squares error

fitting is popular (e.g. Mubayi et al., 2010) and the expectation maximization (EM) algorithm has seen some use (Duncan and Gyöngy, 2006; Lavielle et al., 2011). We choose Bayesian MCMC approach as it provides a huge amount of modeling flexibility and enables analysis of all the model parameters or functions of parameters. It also has advantage of providing a complete distribution for parameters as posterior distributions instead of point estimates. Moreover, posterior summaries such as mean, medians, maximum likelihoods, maximum, minimum and credible intervals are easy to obtain as well.

In this study, we fitted dengue incidence data from Thailand to vector–host and SIR models and obtained estimates of model parameters including average duration of dengue infection in humans, lifespan of mosquitoes and the probability of the severe form of disease. The parameter estimates were consistent with existing published values and PRCC values showed that all the parameters except initial conditions have significant influence on the magnitude of the basic reproduction number R_0. Both the vector–host model as well as the SIR model fit the incidence data well, however AIC model selection found the SIR model with fixed $h_R(0)$ to be substantially better than the vector–host model, implying that incorporating mosquito population explicitly in a dengue model may not be necessary to explain the incidence data from Thailand.

Chapter 4

The role of population immunity for accurately estimating model parameters for dengue transmission

4.1 Introduction

The basic reproduction number (R_0) is defined as the number of secondary cases that one case would generate on average over the course of its infectious period in a completely susceptible population (Heffernan et al., 2005). R_0 is a useful metric to determine whether or not an infectious disease can spread through a population. The outbreaks of epidemics and the persistence of endemic levels of an infectious disease are associated with $R_0 > 1$ (Heffernan et al., 2005). More specifically, when $R_0 > 1$ each infectious case generates more than one new infectious case and so infection would be able to invade the susceptible population, whereas for $R_0 < 1$,

each infectious case generates less than one new infectious case and therefore infection would be cleared from the population. This threshold allows one to determine the amount of effort that is necessary to prevent an epidemic or to eliminate a disease from a population. The estimation of R_0 has proved to be very critical in understanding the outbreak and spread of infectious diseases, for example severe acute respiratory syndrome (SARS) (Choi and Pak, 2003), West Nile virus (Wonham et al., 2004), foot and mouth disease (Ferguson et al., 1999c), dengue (Luz et al., 2003) and malaria (Hagmann et al., 2003). The value of R_0 varies considerably for different infectious diseases, as well as in different populations, as it depends on the duration of infectious period, the probability of transmission of infection per contact and the number of new susceptible individuals contacted per unit time. As a result, it is important to estimate R_0 for a given disease in a particular population.

Dengue, a mosquito-borne viral disease of tropical and sub-tropical regions around the world, poses a substantial health burden on the population in Thailand (Clark et al., 2005). Dengue fever and dengue hemorrhagic fever (DHF) are among the leading causes of hospitalization of children in Thailand and southeast Asia generally (Clark et al., 2005). Several studies have estimated the basic reproduction number and used it to understand the spread and dynamics of dengue transmission in Thailand (Adams et al., 2006; Chao et al., 2012; Cummings et al., 2009; Nagao and Koelle, 2008). The estimates of the reproduction number of dengue in Thailand vary considerably between these studies, from values of around 1–3 (Adams et al., 2006; Chao et al., 2012) to 6.7 (Cummings et al., 2009) to 10–12 (Nagao and Koelle, 2008). More recently, we fitted standard dengue models to monthly observed DHF cases in Thailand in 1984 using Bayesian MCMC and found R_0 estimates in the range 1–3 (Pandey et al., 2013). This variability in R_0 estimates among different studies is not well understood and may possibly be attributed to the complex dynamics of

dengue due to interactions of its 4 different serotypes (Adams et al., 2006; Chao et al., 2012; Cummings et al., 2009; Nagao and Koelle, 2008). In addition, different methods of estimating R_0 may contribute to the variability (Heffernan et al., 2005; Li et al., 2011).

In the present study, we fit a simple SIR-type deterministic model to monthly DHF cases in Thailand in 1984 to investigate the variability in the basic reproduction number for dengue. We found that that our model fits the observed data evenly for varying values of the basic reproduction number as other parameters vary. Specifically, there was a logarithmic type relationship between the level of initial population immunity and the basic reproduction number that suggests difficulty in estimating the level of population immunity and the basic reproduction number together from same data source. Thus, in order to estimate the model parameters accurately, either the basic reproduction number or the proportion of the population immune at the beginning of the epidemic must be estimated from separate data.

4.2 Model and Approach

We used a standard SIR-type deterministic dengue model (Hethcote, 2000). As we are only interested in infections in humans, we did not model mosquito population explicitly (Pandey et al., 2013). The model consists of three compartments, one each for the number of people susceptible (H_S), infectious (H_I) and recovered (H_R). The model is the system of differential equations

$$\begin{aligned}
\frac{\mathrm{d}H_S}{\mathrm{d}t} &= B_H - \beta\frac{H_I}{H}H_S - \mu_H H_S, \\
\frac{\mathrm{d}H_I}{\mathrm{d}t} &= \beta\frac{H_I}{H}H_S - \gamma_H H_I - \mu_H H_I, \\
\frac{\mathrm{d}H_R}{\mathrm{d}t} &= \gamma_H H_I - \mu_H H_R,
\end{aligned} \tag{4.1}$$

where $H = H_S + H_I + H_R$ is the human population size. A susceptible person gets infected with force of infection $\beta\frac{H_I}{H}$, where β is the composite human-to-human transmission rate (Pandey et al., 2013). The average duration of infection is $1/\gamma_H$ and $1/\mu_H$ is the average lifespan of people in Thailand. Since we only fit the model to an epidemic lasting a year, we assumed the population size is constant by setting the birth rate equal to the death rate, $B_H = \mu_H H$. The basic reproduction number is

$$R_0 = \frac{\beta}{\mu_H + \gamma_H}. \tag{4.2}$$

We obtained the monthly incidence of DHF for Thailand from 1983 to 1997 (Johns Hopkins Center for Immunization Research, Accessed January 14, 2013). We choose the 1984 epidemic as a clear, representative example among this data. Since cumulative incidence is generally smoother and thus easier to fit, we used the cumulative monthly number of DHF cases for the period January 1984 to December 1984. We added a compartment to our model (4.1) to calculate the cumulative number of DHF infections (H_C). We assumed a fraction p of infections were diagnosed as DHF, with p constant in time. The differential equation for the H_C compartment is

$$\frac{\mathrm{d}H_C}{\mathrm{d}t} = p\beta\frac{H_I}{H}H_S, \tag{4.3}$$

which is precisely the rate of new infections multiplied by p.

The unknown parameters are the transmission rate (β), the recovery rate (γ_H) and the probability of DHF (p), along with the initial proportion of people immune ($h_R(0) = H_R(0)/H$). As we are interested in the variability in reproduction number R_0, we reparametrized our model using (4.2) to avoid estimating β directly. For simplicity, we fixed $\mu_H = 1/69$ y^{-1} based on the average human duration of life

in Thailand in 1984 of about 69 years (World Bank, Accessed January 14, 2013). We used the first month of the incidence data (January 1984) to calculate initial conditions for the initial population infected, $H_I(0) = 454$. Since the population is constant, the initial susceptible population was calculated as $H_S(0) = H(0) - H_I(0) - H_R(0)$ where $H(0) = 46\,806\,000$ is Thailand's population in 1984.

We used the "ode15s" function in Matlab to numerically solve the model (4.1) and (4.3) to generate the predicted monthly cumulative DHF incidence, $y_i = H_C(t_i)/H$, where $t_i = 0, 30, 60, ...$ days. Using the least-squared error between the cumulative DHF data D_i and the model prediction $y_i(\theta)$ for the input parameter values θ,

$$E^2 = \sum_{i=1}^{12} \Big(D_i - y_i(\theta) \Big)^2, \tag{4.4}$$

we assumed the errors were Gaussian, giving the loglikelihood function

$$\text{Loglik}(\theta) = \log \left(\Pr(D \mid \theta) \right) = -E^2, \tag{4.5}$$

which is log of the conditional probability of obtaining the data (D) for the given parameter values (θ).

For our simulations, we varied R_0 from 1 to 20 and sought optimal estimates for rest of the unknown parameters (γ_H, p and $h_R(0)$) that maximize the likelihood function (4.5). For a given value of R_0, we used the Matlab function "fminsearchbnd", which first transforms the problem into an unconstrained problem and then uses Nelder–Mead smilex direct search to minimize the negative of the loglikelihood function over the rest of parameters while remaining within the wide bounds that were chosen for rest of the parameters based on on our past work (Pandey et al., 2013) to represent our general understanding about where the parameter values lie (Table 4.1).

Parameter	Definition	Value
B_H	birth rate (y^{-1})	$1/69$
μ_H	per capita mortality rate (y^{-1})	$1/69$
γ_H	recovery rate (d^{-1})	$(0.125, 1)$
p	proportion of dengue infections that become DHF	$(0, 0.1)$
$h_R(0)$	initial proportion of people immune	$(0, 1)$
R_0	basic reproduction number	$(1, 20)$

Table 4.1: Parameters of the dengue model.

For each R_0 value, we ran our simulation from 11 random initial conditions to ensure that algorithm did not get stuck at local minima and saved the parameter values that resulted in the best loglikelihood value. For some values of R_0, the method found suboptimal local maxima, so we used Matlab's function "smooth" to implement a 3-point moving average to smooth our results in order to highlight the overall trend.

4.3 Results

For values of the basic reproduction number from 1 to 20, we found optimal estimates of the remaining model parameters: recovery rate (γ_H), proportion of dengue infections that develop DHF cases (p) and the initial proportion of humans immune ($h_R(0)$).

We found good fits for varying values of R_0 (Figure 4.1). With increasing values of R_0 the likelihood increases as well, but the model fits appear similar (Figure 4.2). As R_0 varied from 1 to 20, the optimal estimates for recovery rate (γ_H) decreased from around 0.6 and stabilized around 0.3 (i.e., the average duration of dengue infection varied from around 2 days to 3 days) (Figure 4.3a). Similarly, the optimal estimates for proportion of dengue infections that develop DHF cases (p) varied from 0.02 to 0.08 (Figure 4.3b). Equally good fits (similar likelihoods) are achieved for higher R_0 values

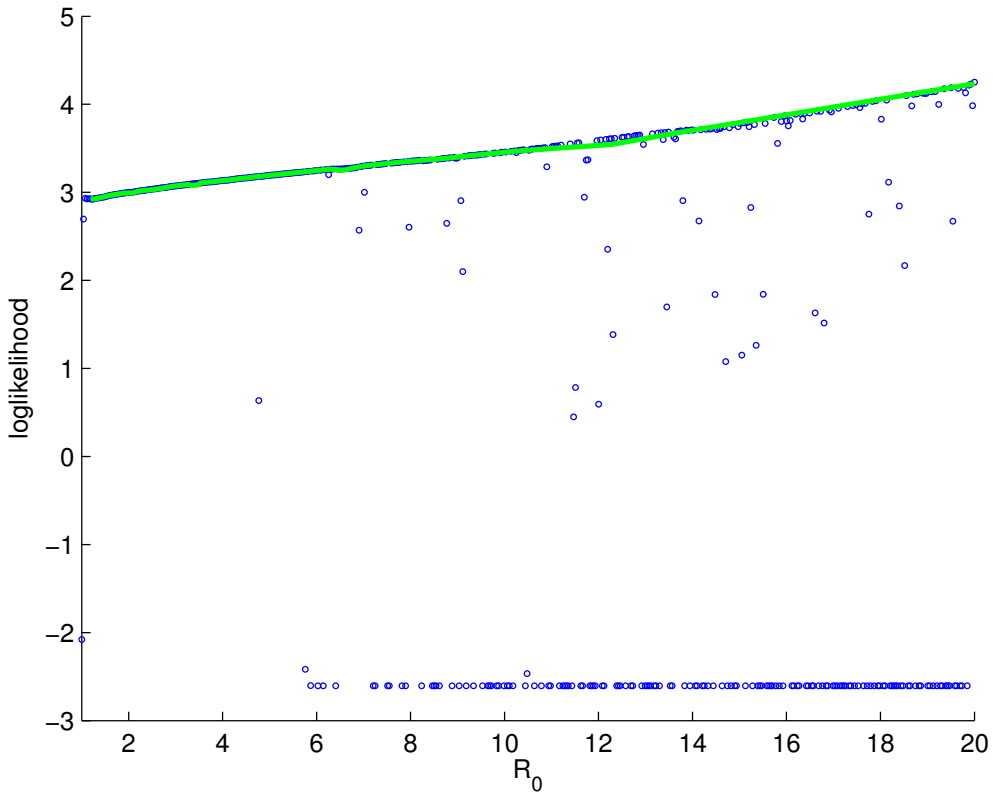

Figure 4.1: Maximum of loglikelihood for R_0 from 1 to 20. The maximum of loglikelihood corresponds to the maximum likelihood. A 5-point moving-average method was used for smoothing.

at relatively smaller recovery rate (γ_H) estimates and relatively larger probability of DHF (p) estimates, but the variation in the estimates is not large and is consistent with existing estimates (Cummings et al., 2005; Gubler, 1998a).

The estimates of the initial proportion of people immune $h_R(0)$ increased as R_0 increased (Figure 4.3c). Moreover, the variation in $h_R(0)$ was very large as $h_R(0)$ increased from values of around 0.16 to 0.96. Immunity in around 15% of population at the beginning of the epidemic resulted in R_0 of 1 to 2, whereas immunity in more than 90% of population at the beginning of the epidemic resulted in large reproduction

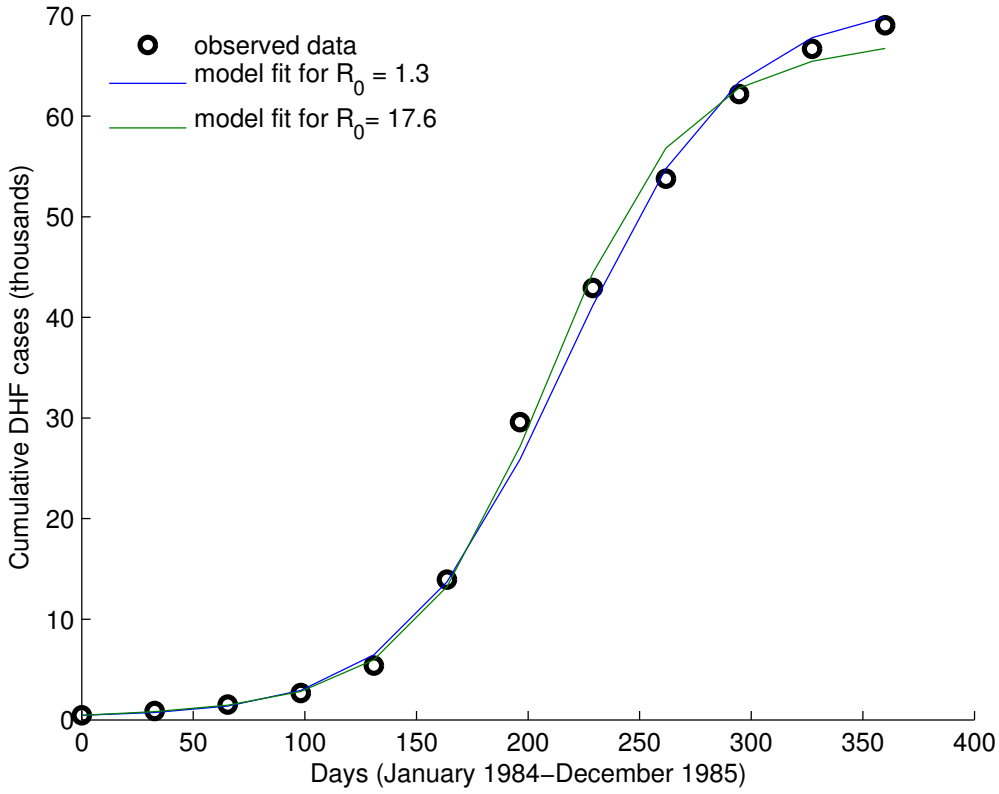

Figure 4.2: Model fit with optimal parameter estimates for $R_0 = 1.3$ and 17.6. Black dots are cumulative Thailand DHF incidence data. Blue and green lines represent the model fits corresponding to $R_0 = 1.3$ and 17.6 respectively.

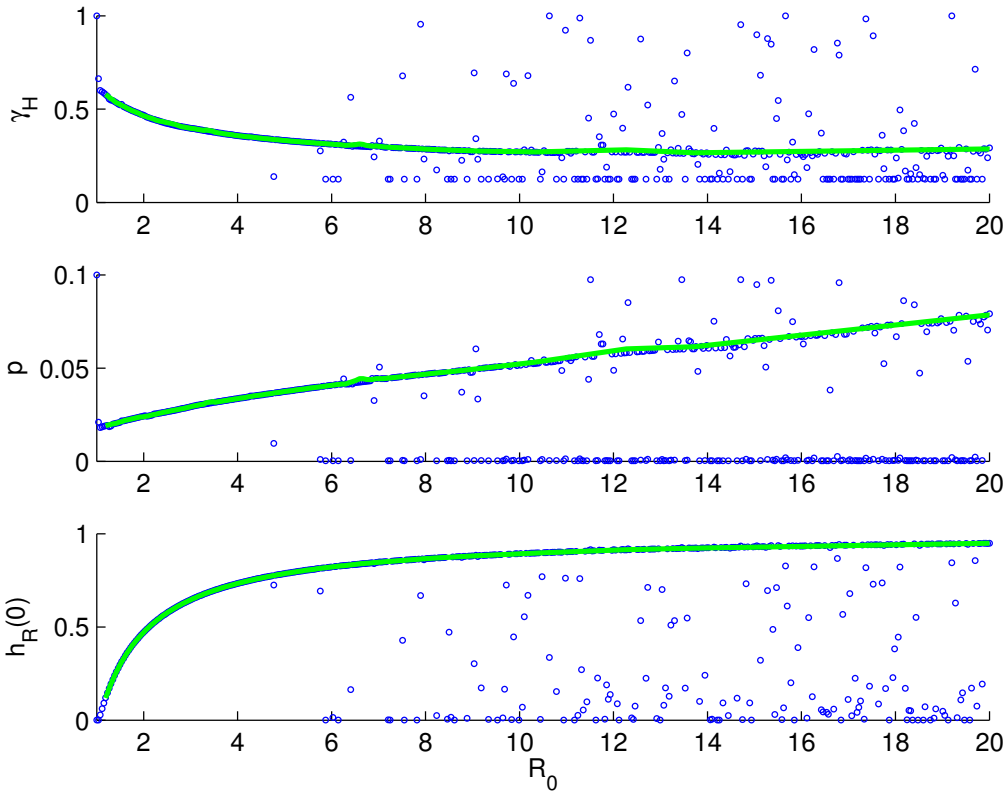

Figure 4.3: Variations in optimal parameter estimates as the basic reproduction number R_0 varies. A 5-point moving-average method was used for smoothing.

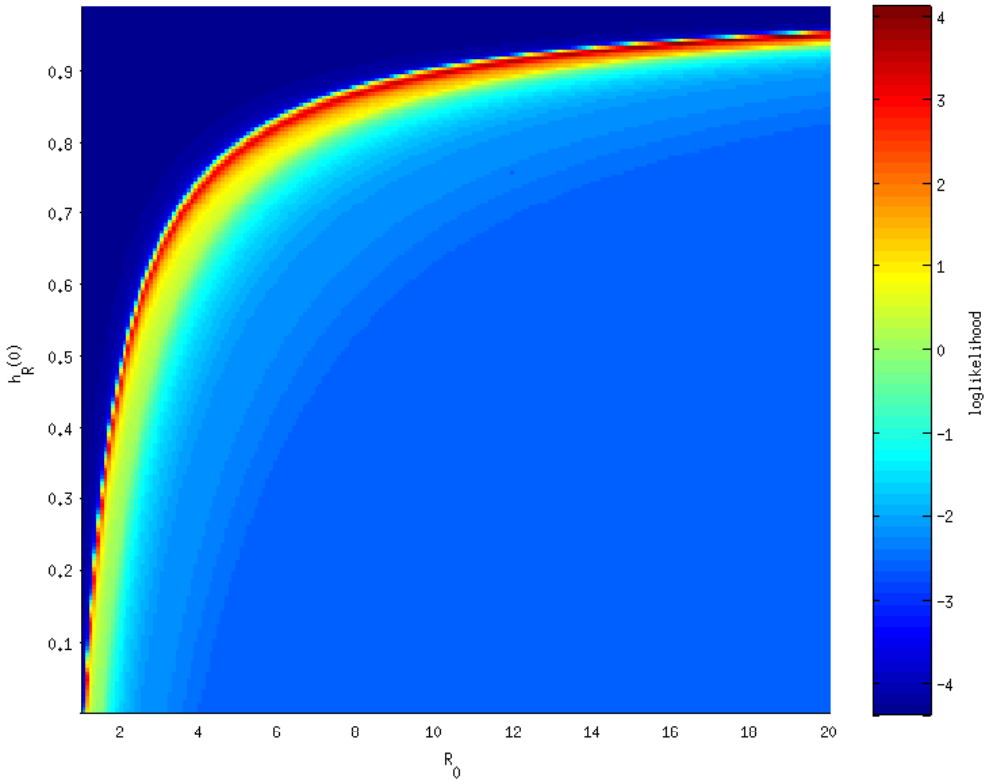

Figure 4.4: Maximum of loglikelihood for varying R_0 and $h_R(0)$ values.

numbers ($R_0 > 11$). This relationship between the level of population immunity at the beginning of the epidemic and the basic reproduction number was clear when we found the maximum likelihood for varying R_0 and the proportion of humans immune $h_R(0)$ together and maximized likelihood over the remaining parameters (γ_H and p) (Figure 4.4).

This connection between R_0 and $h_R(0)$ can be further explored by assuming that dengue incidence has reached a steady-state. We found a relationship between the initial proportion of humans recovered $h_R(0)$ and the cumulative number of infections after the end of one year (details in Appendix B), The balance between birth,

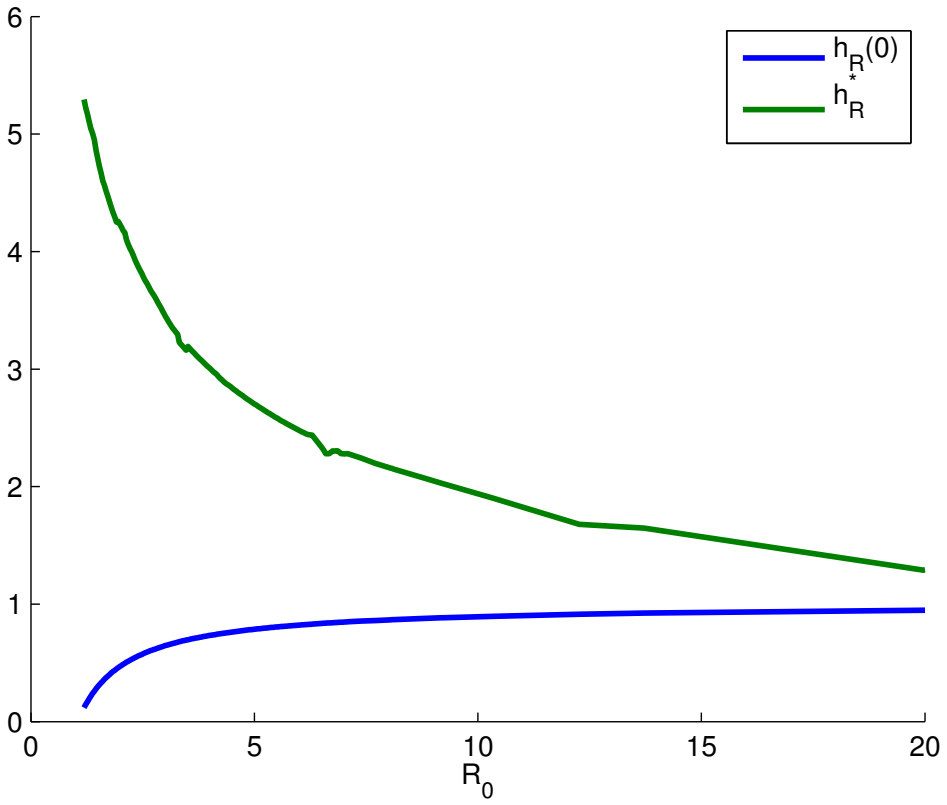

Figure 4.5: Optimal estimates of $h_R(0)$ and h_R^* for varying R_0 values.

death and infection gives

$$h_R^* = \frac{(H_C(1) - H_C(0))\gamma_H}{(\mu_H + \gamma_H)p\mu_H H} \tag{4.6}$$

where h_R^* is the proportion of people immune when system is at steady-state. For varying basic reproduction number, we plotted the optimal estimates of the initial proportion of humans immune ($h_R(0)$) at the beginning of 1984 and h_R^* calculated using the optimal estimates of the rest of the model parameters, human population ($H = 46\,806\,000$) in 1984 and fixed $\mu_H = 1/69\ y^{-1}$. We found that system is not in steady–state for the observed data for R_0 between 1 to 20 (Figure 4.5).

4.4 Discussion

We found that equally good fits of our standard dengue model to 1984 monthly DHF incidence data from Thailand can be achieved for wide range of R_0 values. As R_0 varied, the other parameters also varied to generate similar maximum likelihoods. The largest variation in optimal parameter values was found in the initial proportion of people immune ($h_R(0)$). $h_R(0)$ increased from low values of around 0.16 for R_0 in the range of 1 to 2 to high values of about 0.96 for $R_0 > 11$. The relationship between $h_R(0)$ and R_0 means that when estimating parameters using observed data, low levels of the initial population immunity are consistent with lower values for R_0, whereas high levels of population immunity are consistent with higher values for R_0. The role of the level of initial population immunity in determining estimates for the basic reproduction number seems justified as for high levels of the population immunity, the disease needs to be more infectious and thus will have higher reproduction number in order to generate the same number of infections in the the population, compared to when the level of the initial population immunity is low.

Our results highlight the importance of finding the immunity level in population in order to estimate the rest of the parameters. In other words, choosing inaccurate estimates for the level of initial population immunity to estimate R_0 may fit the data well, but may not give correct values. In order to estimate model parameters accurately, it is vital that we use an estimate from another source for the level of initial population immunity or for the basic reproduction number. It is difficult to know the level of population immunity for any pathogen and 4 different dengue serotypes circulating in Thailand make it even more challenging. Specific data that classifies dengue incidence by serotype along with the knowledge of dengue serotypes circulating in the population at present and in prior years may be used to find reason-

able estimates for the level of population immunity at the beginning of the epidemic.

The optimal estimates of the duration of infection ($1/\gamma_H$) and the proportion of dengue infections that develop DHF cases (p) increased with increasing values of R_0. The variation in both of these parameters was not large, as the optimal estimates for the duration of infection varied between about 2 to 3 days and the variation in the number of DHF cases per 100 dengue infections was between 2 and 8.

Based on the assumption that the epidemic has already reached steady-state and using the optimal estimates for initial population immunity, we found that for the 1984 epidemic in Thailand, the population immunity levels were not consistent with the basic reproduction number between 1 to 20 meaning either the system is not in steady-state for the observed data or the it has a R_0 higher than 20. We believe the system has not reached the steady-state for the DHF incidence data in Thailand.

As the transmission parameter (β) depends on the biting rates of mosquitoes, and probabilities of transmission, it is difficult to estimate a range of values for this transmission parameter. Thus, we avoided estimating β by reparametrizing the model in terms of R_0. For our simulations, we varied R_0 from 1 to 20 to ensure a wide range that includes values for dengue transmission that have been used or estimated in the literature.

The monthly DHF incidence data from Thailand in 1984 was used for this study. However, our results do not change considerably when we chose epidemic data from three different years (1985, 1987 and 1990). The likelihood in our study was chosen on the assumptions that the errors follow a Gaussian distribution, but we believe that we would get similar results if a different likelihood was chosen.

We used a standard SIR-type dengue model to investigate the reasons behind the wide variability in estimates of the basic reproduction number in Thailand. We found that the initial population immunity plays a vital role in determining the

estimate for R_0. Assuming a small initial population immunity results in lower estimates for R_0, whereas large initial population immunity gives higher values for R_0. This relationship between the initial population immunity and the basic reproduction number may be a reason for wide range of R_0 estimates found in studies. Moreover, finding better estimates of population immunity level would help to more accurately estimate R_0 and other epidemiological parameters for dengue.

Chapter 5

The introduction of dengue vaccine may temporarily cause large spikes in prevalence

5.1 Introduction

Dengue is an RNA virus in the family Flaviviridae. There are four serotypes of dengue virus, each of which induces a specific antigenic response in humans. Infection with any of the four serotypes can cause disease ranging from dengue fever (DF) to dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS), in order of increasing severity (Guzmán and Kouri, 2002). Infection with a serotype provides life-long immunity to that serotype and an increased risk for the severe forms of disease (DHF and DSS) during subsequent infection with a different serotype (Nagao and Koelle, 2008). Dengue infection has been recognized in over 100 countries: DF and DHF are important public-health problems, especially in the tropics and subtropics, where nearly 2.5 billion people are at risk of infection (Guzmán and Kouri, 2002). An

estimated 50 million cases of DF occur every year, including $500,000$ hospitalizations for DHF (Guzmán et al., 2010).

Despite the expansion in geographic range of the virus over past four decades (Beatty et al., 2012), there is no dengue vaccine licensed for use presently. Dengue vaccine development is an area of active research: significant advances have occurred in recent years and several vaccine candidates are showing promise in clinical studies (Coller and Clements, 2011). The most advanced vaccine candidate showed efficacy of 30% and protection against 3 of the 4 serotypes of dengue (Halstead, 2012). With several dengue vaccine candidates progressing through clinical trials, a licensed dengue vaccine is expected to be available in less than 10 years (Guzmán et al., 2010).

Once the vaccine becomes available, policy makers will have to decide how to best allocate it. As a result, it is imperative to carefully examine the effects of vaccine-allocation policies. Mathematical models of dengue transmission predict complex temporal patterns in prevalence, driven by seasonality in mosquito abundance (Reich et al., 2013), and may include a transient period immediately after vaccine introduction where prevalence can spike higher than in the pre-vaccine period. An increase in infections, however brief, can raise doubts about the vaccination program, while the longer-term outcome may be highly favorable. A temporary increase in infections might lead to public doubts and refusal to use the vaccine or even policy makers' ending the vaccination program altogether. Moreover, these spikes can also pose serious problems by overwhelming resources like available hospital beds. Thus, investigating short- and long-term effects of the vaccination introduction is important.

In the present study, we found that simple dengue transmission models frequently predict large transient spikes in prevalence in the years after vaccination is begun. These presence of these spikes was highly sensitive to the level of vaccination, the efficacy of the vaccine, and the timing of the vaccination program. Despite the

presence or absence of transient spikes, vaccination reduced dengue infections when averaged over a the first 15 years after vaccine introduction and when averaged over the very long term.

5.2 Methods and Approach

Here we briefly describe the mathematical models for dengue transmission that we used in this study. See Appendices C.1–C.3 for detailed descriptions of the models.

We developed standard SIR-type deterministic dengue models to explore the short- and long-term effects of vaccine introduction. In the models, the population was divided into unvaccinated and vaccinated people, and then each of these was further divided by infection history (Figure 5.1). In the main model, a person with no prior dengue infection (state $S1$) can have a primary infection from any serotype of dengue ($I1$), recovery from which provides him with life-long immunity to that serotype ($S2$). He can then acquire infection from any of the remaining three serotypes, and enter the secondary-infection class ($I2$). Third or fourth infection from dengue is very rare (Halstead, 2003), so we assumed that an individual recovering from secondary infection becomes immune to all serotypes (R). Dengue mortality in humans is at most 1–2% (Rajapakse, 2011) and thus ignored in our model for simplicity. As we are only interested in infections in humans, we did not model mosquito population explicitly (Pandey et al., 2013). We used model parameters consistent with the literature, along with a small level of seasonal forcing of transmission to capture seasonal oscillations in mosquito population size (Nisalak et al., 2003).

We modeled the vaccine as having efficacy ϕ at preventing infection. The vaccination program was modeled as having two possible components: one component vaccinates a proportion p of infants, and the other component vaccinates people in

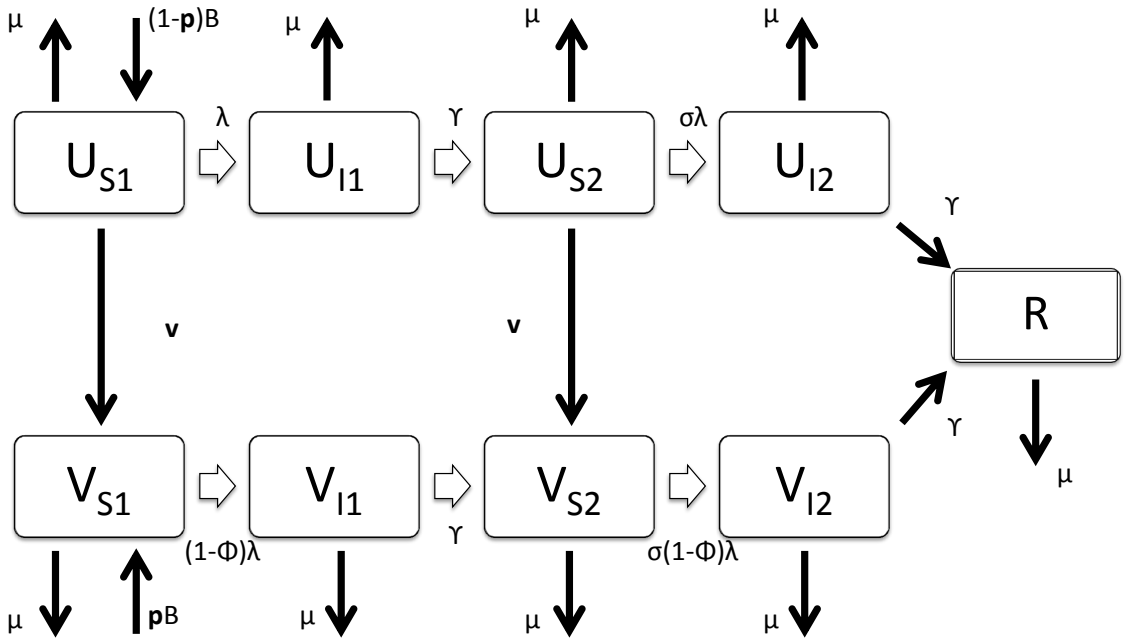

Figure 5.1: Diagram of the main model. U denotes unvaccinated people, while V denotes vaccinated. $S1$ and $S2$ are people susceptible to primary and secondary dengue infection, respectively. $I1$ and $I2$ are people infected with primary and secondary infection, respectively. R is people recovered from secondary infection and immune to further infections. See Table C.1 for definition of the other symbols.

the general population at rate v. For simplicity, for the main results we used only vaccination of infants ($v = 0$).

To simulate vaccine introduction, we computed a solution to the model with no vaccine ($p = 0$) from arbitrary initial conditions until it converged to regular periodic oscillations (a stable limit cycle, in mathematical terms (Strogatz, 1994)). From a new initial point on this periodic solution, we then computed the solution to the model with vaccine introduced ($p > 0$).

5.3 Results

The model prevalence converged to regular periodic oscillations, with or without vaccination (e.g. Figure 5.2). Without vaccination, prevalence oscillates with a period of 2 years. The perturbation caused by the introduction of vaccine results in a transient period where the prevalence has not yet converged to periodic oscillations. During this transient period, large spikes can occur, to levels above those present before the vaccine was introduced, and this transient period may last many years.

For example, vaccinating 78% of infants at 30% vaccine efficacy results in an initial period of about 60 years when prevalence can spike more than twice as high as before vaccine introduction (Figure 5.2). After the transient period, prevalence converges to annual oscillations with a smaller maximum than in the 2-year oscillations prior to vaccine introduction (Figure 5.2(C)). In contrast, vaccinating instead 90% of infants results in a transient period of about 70 years, where prevalence can spike more than three times as high as before vaccine introduction, and prevalence then settles down to 3-year oscillations with a higher maximum than before vaccine introduction (Figure 5.3).

These two examples show that the effectiveness of a vaccination program may differ depending on exactly what is evaluated. Effectiveness may consider individual points in time (e.g. the height of the spikes in prevalence) or periods of time (e.g. total number of infections in a fixed period). In addition, effectiveness may be evaluated over the period just after vaccine introduction or, as is more convenient from a modeling perspective, the period after the prevalence has converged to periodic oscillations. To capture the transient period over many simulations, we considered the transient period to be the first 15 years after vaccine introduction. To quantify the severity of transient spikes at any point in time, we calculated the maximum prevalence during

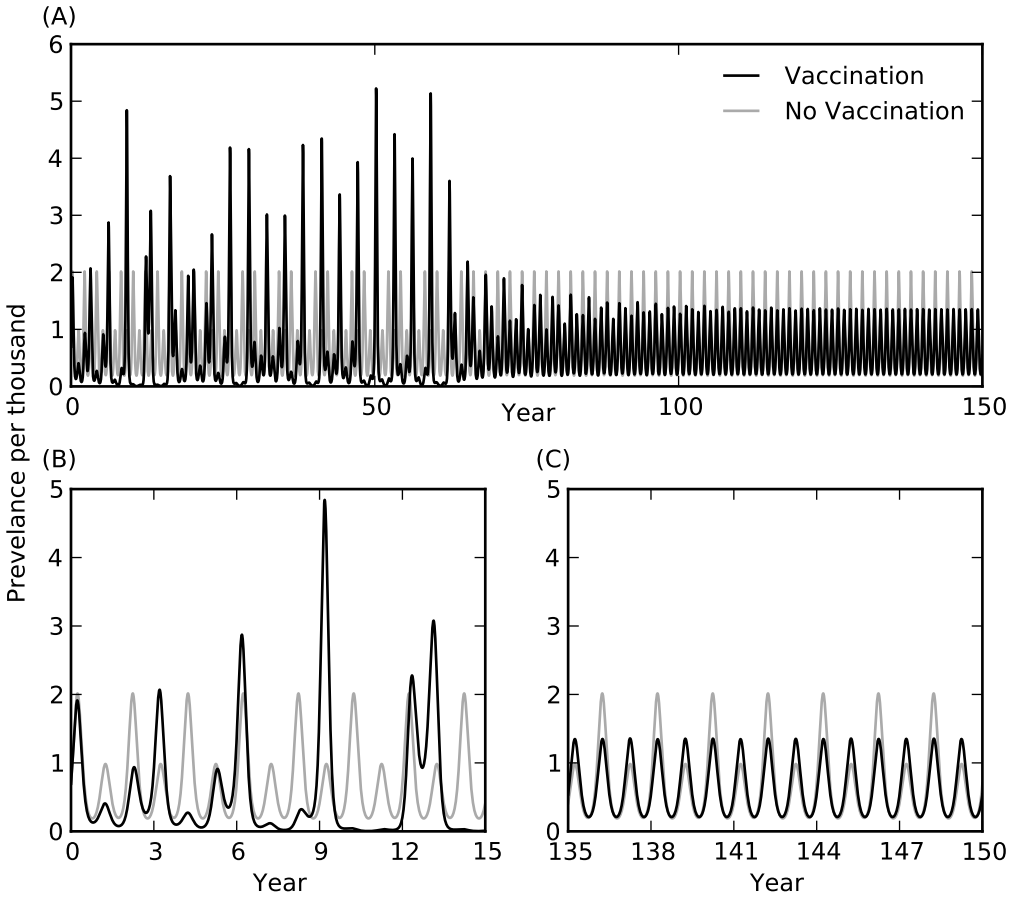

Figure 5.2: Simulated dengue prevalence after vaccine introduction. Starting at year $t = 0$, $p = 78\%$ of infants are vaccinated with a vaccine of efficacy $\phi = 30\%$. The black curves are prevalence after vaccine introduction, while the gray curves are the prevalence had vaccine not been introduced. (A) shows prevalence for 150 years after vaccine introduction, while (B) & (C) show only the first and last 15 years of this period. See subsection C.1 for model & parameter definitions.

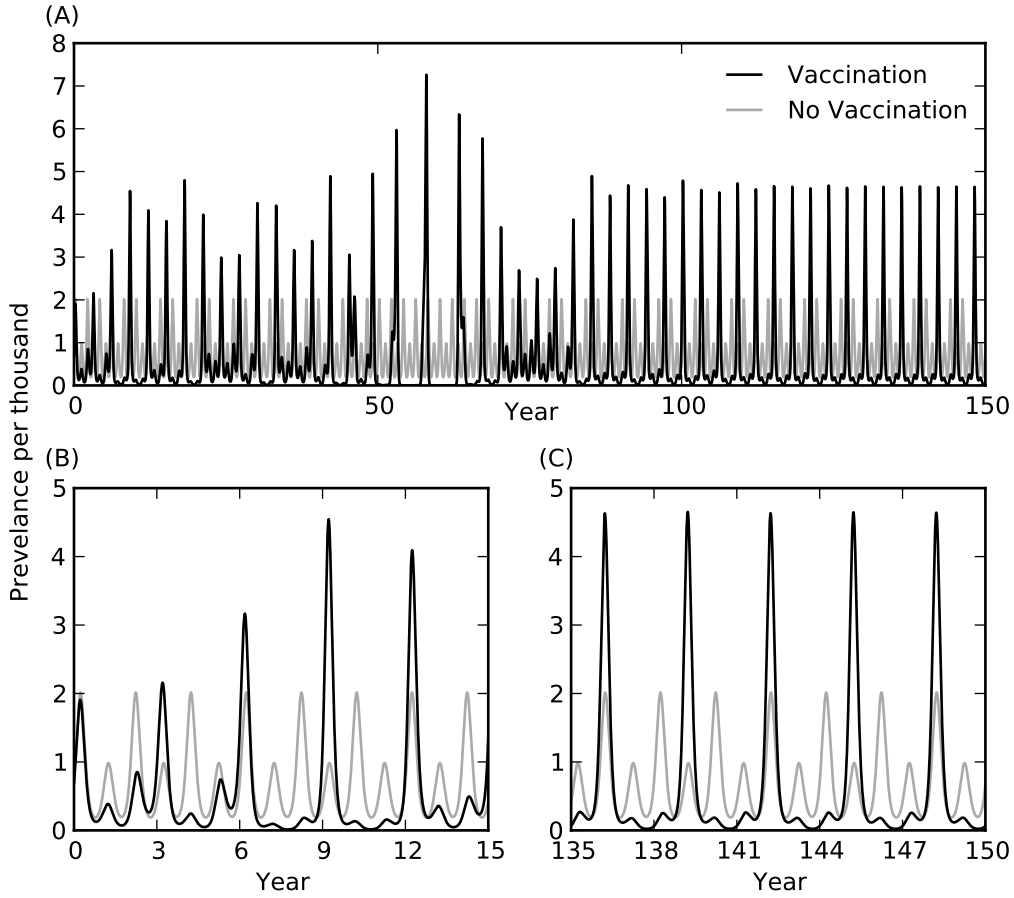

Figure 5.3: Simulated dengue prevalence after vaccine introduction. The model and parameter values are as in Figure 5.2, but with $p = 90\%$ of infants vaccinated.

first 15 years after introduction of the vaccine. To examine the effectiveness over a period of time, we also calculated the total number of infections during the first 15 years after vaccine introduction. For the period after prevalence has converged to periodic oscillations, we calculated both the maximum prevalence and the mean prevalence per year.

To explore the transient spikes in prevalence, we varied vaccine coverage (p), for low ($\phi = 30\%$) and high ($\phi = 70\%$) vaccine efficacy (Figure 5.4(A)). Large transient spikes (i.e. above the pre-vaccine maximum of about 2 per thousand) were present after vaccine introduction for both levels of vaccine efficacy. Moreover, for the higher vaccine efficacy, large transient spikes appeared at lower vaccine coverage.

By varying vaccine efficacy, we found that large transient spikes occur particularly when vaccine efficacy is neither very low nor very high (Figure 5.5). For 80% vaccine coverage, when vaccine efficacy is below 30%, the perturbation caused by vaccine introduction is insufficient to generate transient spikes. On the other hand, when vaccine efficacy is higher than 97% at 80% vaccine coverage, vaccination quickly reduces prevalence without large transient spikes. Similarly, with 40% vaccine coverage, there are no large transient spikes below 58% or above 96% vaccine efficacy.

Since prevalence oscillates with a period of 2 years without vaccination, we varied the time of vaccine introduction over the 2-year period (Figure 5.6). The presence or absence of large transient spikes after vaccine introduction is highly sensitive to the time of vaccine introduction.

Despite the presence of large transient spikes in prevalence, the total number of infections over the first 15 years was lower with vaccine than without in all of the simulations we performed (Figure 5.4(B)). However, we cannot rule out that the initial burden may be higher in some cases, especially over shorter time spans than 15 years. Moreover, higher vaccine efficacy yields higher drop in the initial burden at

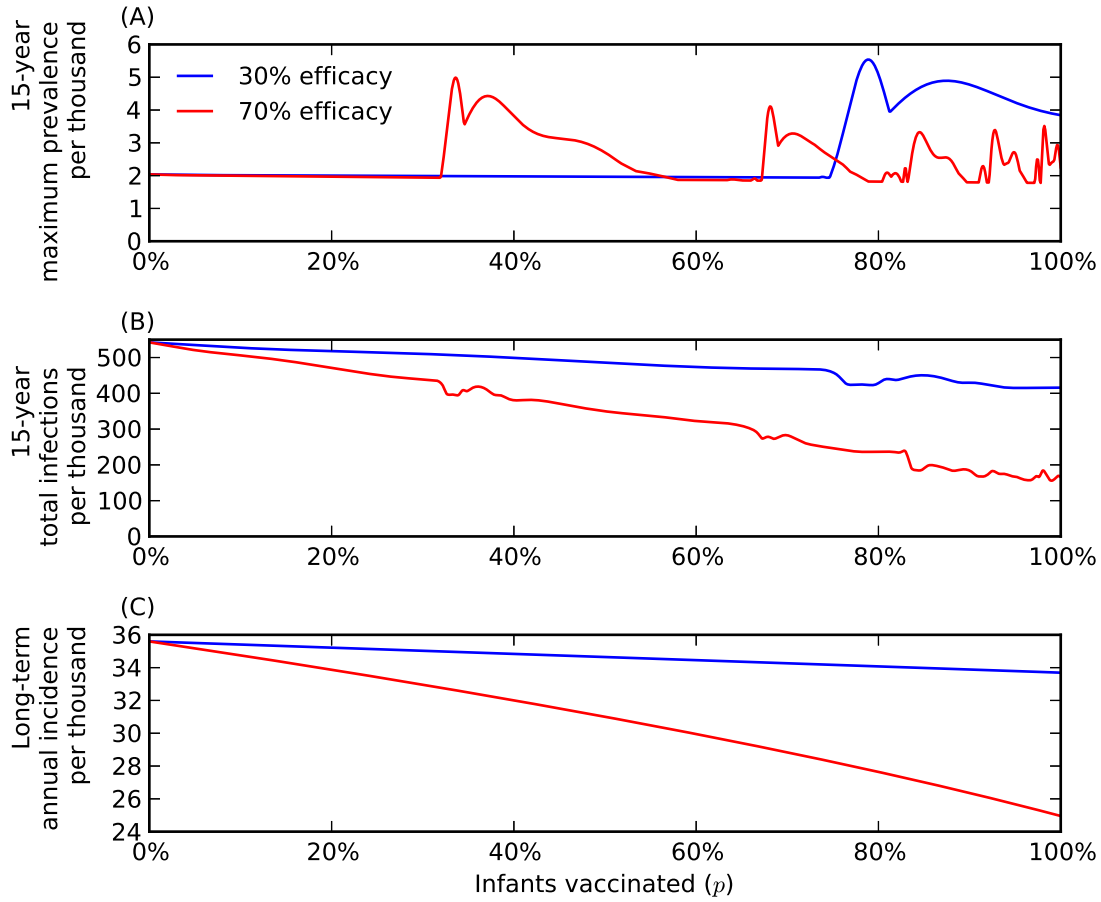

Figure 5.4: Effectiveness of the vaccination program at different levels of coverage. (A) Maximum prevalence in the first 15 years after vaccine introduction. (B) Total number of infections over the first 15 years after vaccine introduction. (C) Long-term mean annual incidence.

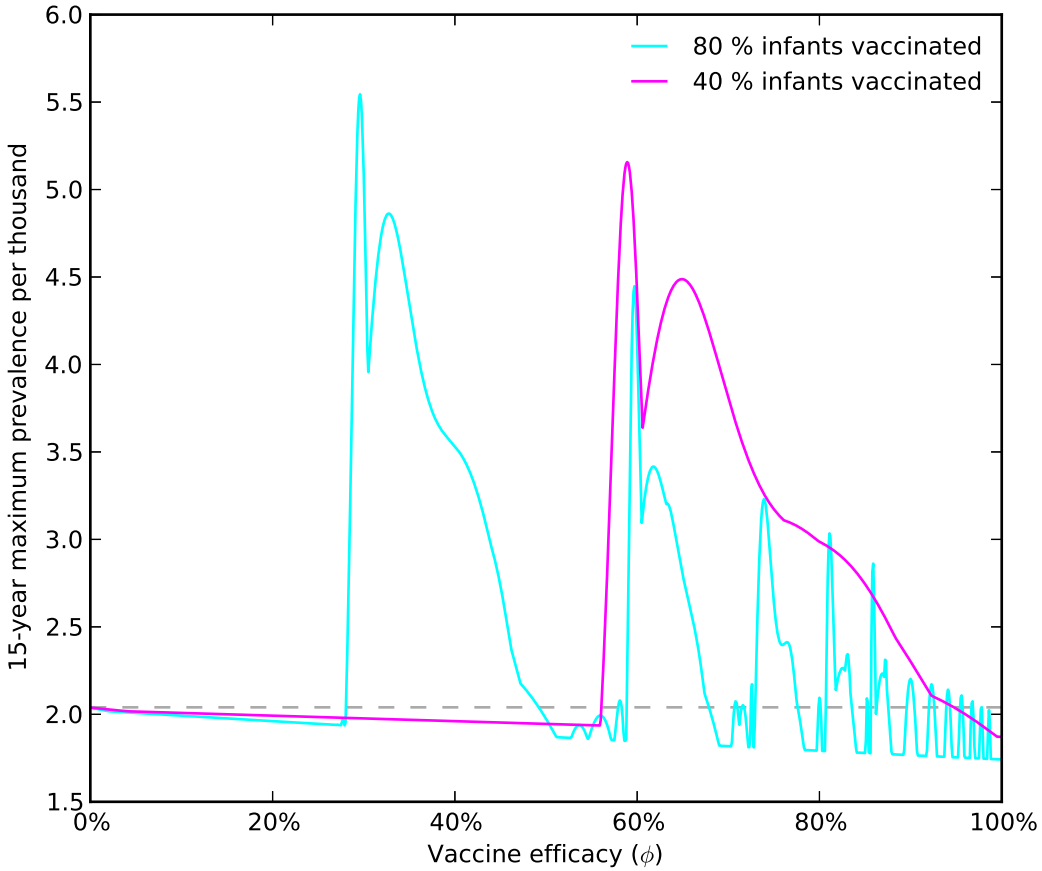

Figure 5.5: Maximum prevalence over the 15 years after vaccine introduction for varying vaccine efficacy. $p = 80\%$ (cyan) and $p = 40\%$ (magenta) of infants are vaccinated. The dashed gray line is the maximum prevalence prior to vaccine introduction.

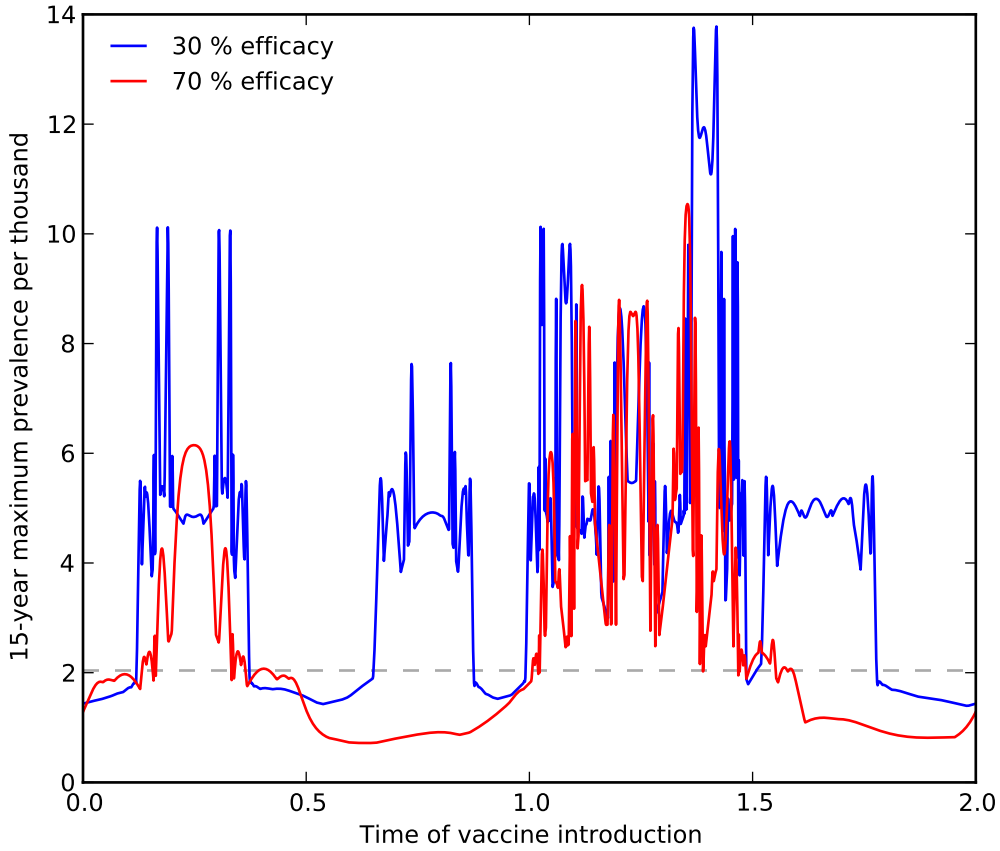

Figure 5.6: Maximum prevalence over the 15 years after vaccine introduction for varying time of vaccine introduction. $p = 80\%$ of infants are vaccinated with a vaccine of efficacy $\phi = 30\%$ (blue) and $\phi = 70\%$ (red). The time that the vaccination program begins was varied from the beginning ($t = 0$) to the end ($t = 2$) of the 2-year cycle in prevalence that exists when there is no vaccine. The dashed gray line is the maximum prevalence prior to vaccine introduction.

constant vaccine rates. Thus, although the presence of large transient spikes may pose immediate problems by overwhelming the resources, over a 15-year period, vaccination reduces infections.

After the transient period caused by vaccine introduction, prevalence converges again to a regular periodic oscillation, but the oscillation may have a different period and the maximum prevalence may be lower or higher than prior to vaccine introduction (e.g. Figures 5.2(C) & 5.3(C)). In all of our simulations, we found that in the long-term, the mean annual incidence decreased as vaccine coverage increased (Figure 5.4(C)). Although some vaccine introductions did lead to long-term prevalence with a higher maximum, their period was longer, so that the mean over many years was always lower than without vaccine. Moreover, as expected, long-term mean annual incidence was lower for higher vaccine efficacy.

The model also showed large transient spikes in prevalence when we modeled vaccination programs in the whole population rather than just infants ($p = 0$ & $v > 0$) and vaccination programs that combine the two ($p > 0$ & $v > 0$). We also tested simpler and more-complex dengue models. The simpler model assumed that there are no secondary infections, so that and individuals who recover from their first infections move directly to recovered class (subsection C.2). Our more-complex model included a period of short-term cross-protection after primary infection (Wearing and Rohani, 2006) (subsection C.3). Both models exhibited the potential for large transient spikes (Figure 5.7).

5.4 Discussion

Using mathematical models of dengue transmission, we found that vaccine introduction may lead to a transient period when infection prevalences spike higher

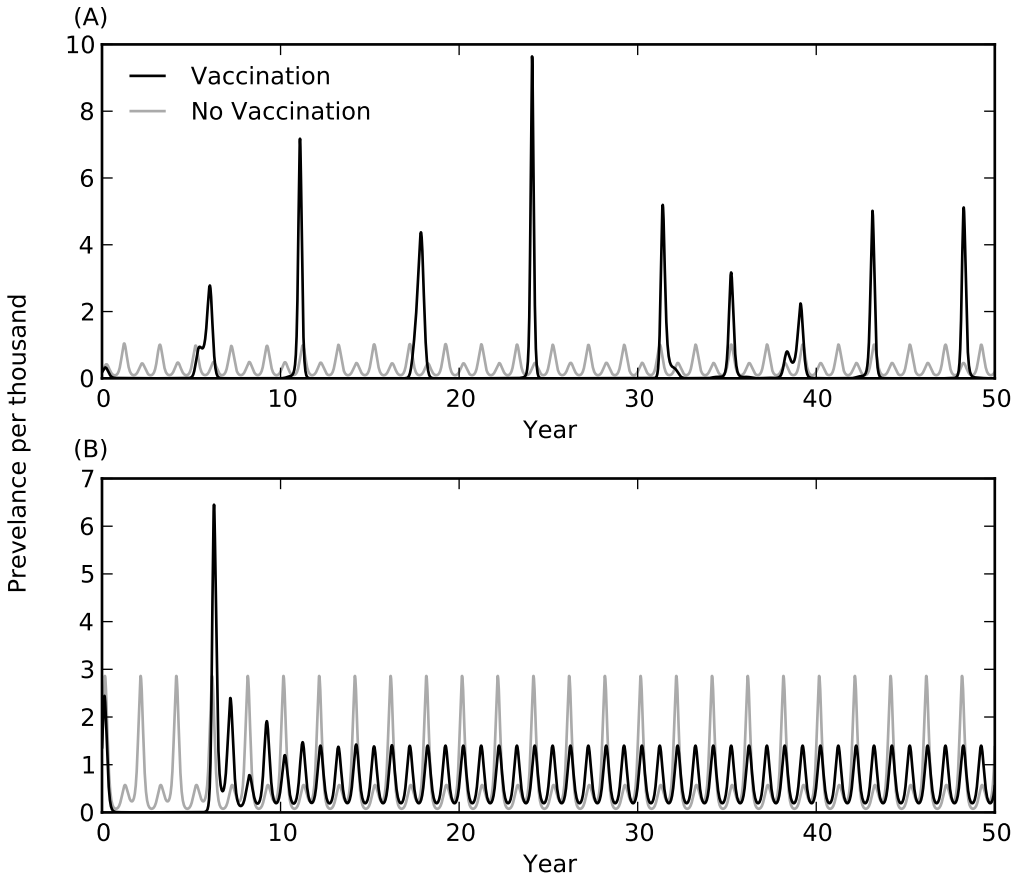

Figure 5.7: Transient prevalence spikes in simpler (A) and more complex (B) dengue models. (A) $p = 80\%$ of infants are vaccinated and the remainder of the population is vaccinated at a per-capita rate of $v = 0.6$ per year. See subsection C.2 for model & parameter definitions. (B) $p = 60\%$ of infants are vaccinated and the remainder of the population is vaccinated at a per-capita rate of $v = 0.9$ per year. Vaccine efficacy is $\phi = 70\%$ in both models. See subsection C.3 for model & parameter definitions.

than in the pre-vaccine period. These spikes in infection prevalence may pose serious problems by overwhelming health resources like hospital beds, as well as create doubt about the efficacy of the vaccination program. In our models, the presence of large spikes required that vaccine coverage and vaccination efficacy were not both very low or both very high. The occurrence of large transient spikes for lower vaccine coverage and higher vaccine efficacy suggests that a sufficiently large perturbation to the system is required in order for large transient spikes. These perturbations are generated by the combination of vaccine coverage and vaccine efficacy. When vaccine coverage and efficacy are both very high, the vaccination program reduces prevalence so quickly that no spikes appear.

We also found that the presence or absence of large transient spikes was highly sensitive to the time when the vaccine program is first begun. Indeed, the results are so sensitive to introduction time, and likely also to changes in parameter values or model structure, that we do not believe that model results can be used to minimize the chance of large transient spikes.

Despite the presence of large transient spikes in prevalence, the total number of infections over the first 15 years after vaccine introduction was always less than in the 15 years prior to vaccine introduction. The decrease in infections was higher for higher efficacy as well as for higher vaccine coverage. Likewise, in the long term, the mean number of infection per year was always smaller than during the pre-vaccine period and decreases with increasing vaccine coverage and vaccine efficacy.

Vaccination reduces the susceptibility of the population, but the short-term interaction of the change in susceptibility with the seasonal forcing of the mosquito population causes complex results, including large transient spikes. When averaged over longer times, the impact is as expected: both the number of infections in the 15-years following vaccine introduction and the long-term mean annual were always

found to be lower after vaccine introduction than before.

Large transient spikes after vaccine introduction were frequent for our main model, along with simpler and more-complex models that we also tested. In general, we believe that the occurrence of large transient spikes in response to perturbation is a robust phenomenon of seasonally forced epidemic models. We are unaware of any theoretical research on the short-time transient behavior of seasonally forced epidemic models, however there is a rich literature on the long-time behavior of such models (Grenfell et al., 2001; Keeling and Grenfell, 1997; King and Schaffer, 1999; King et al., 1996).

In our models, we have used generic parameter values rather than those for a specific location. In particular, different parameter values may change the period and maximum amplitude of the pre-vaccine oscillation (King and Schaffer, 1999; Nagao and Koelle, 2008). However, because of the robustness of the appearance of large transient spikes in prevalence, we expect that these spikes would continue to appear for different parameter values. Similarly, we expect large transient spikes would also be present 1.) if the vaccination program were modeled as starting gradually rather than instantaneously, 2.) if the human population were growing rather than remaining constant size and 3.) if the vaccine were introduced at a state other than regular periodic oscillation.

We used mathematical models to evaluate the short- and long-term effects of introducing a dengue vaccine. We found that vaccine introduction may lead to a transient period when infection prevalences spike higher than in the pre-vaccine period. We believe such transient spikes are robust to changes in parameters and model structure, and thus must be accounted for in planning vaccination programs because they may overwhelm health resources. Despite the presence of transient spikes, the vaccination program is likely to be effective at reducing the total number of

infections during the first few years after introduction, as well as decreasing infections in the long term. Policy makers should be prepared for transient spikes to mitigate their burden and to accurately understand the effectiveness of the vaccine program.

Chapter 6

Conclusions

In this dissertation, I studied different aspects of the mathematical modeling of dengue transmission and vector–borne disease transmission in general. I developed a novel method of Bayesian MCMC estimation for fitting disease models to observed data. The Bayesian MCMC estimation is easy to implement and has the advantage of giving a complete distribution for the parameters instead of point estimates, which enables easy analysis of model parameters and functions.

Two distinct modeling approaches that have been extensively used to explain dengue transmission were compared in order to explore how well dengue models under different modeling assumptions explain observed incidence. I fitted the monthly DHF cases of the year 1984 in Thailand to simple dengue models with and without explicitly modeling mosquitoes using Bayesian MCMC estimation and compared their goodness of fit by applying AIC model selection. Both models agreed with the data equally well, so the model with implicit modeling of mosquitoes being simpler, was strongly selected by AIC. Thus, incorporating mosquito population explicitly in dengue models may not be necessary to explain the incidence data. The fitting of dengue transmission data from Thailand to dengue models also provided estimates of model parameters like

duration of dengue infection in human and probability of DHF that were consistent with their corresponding estimates from the literature.

Based on results from my fitting of dengue models to observed incidence, as well as other studies in literature, I found significant variability in the estimates of the basic reproduction number (R_0) for dengue transmission in Thailand. I investigated this variability in R_0 estimates among studies by fitting a standard dengue model to dengue incidence for varying values of R_0 and found a logarithmic type relationship between the level of population immunity and the basic reproduction number. Particularly, choosing low levels of population immunity resulted in small R_0 estimates between 1 to 3, while larger R_0 estimates (> 11) were obtained for high levels of population immunity. The variability in rest of the parameters of the model for varying R_0 values was not very large and their estimates were consistent with those in the literature. I concluded that in order to estimate model parameters, particularly the basic reproduction number, for dengue in Thailand accurately, it is necessary to use appropriate estimates for the population immunity obtained from different data.

In consideration of a new dengue vaccine being expected within next 10 years, I considered the short- and long-term impacts of vaccine introduction in a population. Our particular focus was on the possibility of large transient spikes in prevalence immediately after vaccine introduction, driven by the perturbation to disease dynamics due to vaccine introduction and seasonality in mosquito abundance. Using a simple dengue transmission model, I found large transient spikes in prevalence to be a robust phenomenon that occurs as long as vaccine efficacy and coverage are not both very high or very low. Although these spikes could lead to doubts about the vaccine program in the public and decision makers and may overburden health resources like hospital beds, the models predicted that vaccination does always reduce the total number of infections in the 15 years after vaccine introduction. Thus policy makers

should consider the possibility of transient spikes while making vaccine program and to mitigate the burden of large spikes in prevalence and to accurately measure the effectiveness of the vaccine program.

I outlined and demonstrated the method of Bayesian MCMC estimation to estimate parameters of disease models. To explore how well dengue models under different modeling assumptions explain observed incidence and to find good model parameter estimates, I fitted the monthly DHF cases in Thailand to simple dengue models with and without explicitly modeling mosquitoes using Bayesian MCMC estimation. Both models fitted the data well and so the simpler model was selected strongly by AIC model selection. Thus I concluded that incorporating mosquito population explicitly in dengue models may not be necessary to explain the incidence data. I investigated the wide variability in the basic reproduction number's estimates for dengue transmission in Thailand among studies by fitting a SIR-type dengue model to dengue incidence for varying values of R_0. I found a logarithmic type relationship between the level of population immunity and the basic reproduction number and concluded that in order to estimate model parameters, particularly the basic reproduction number accurately, reasonable estimate of population immunity from another source is necessary. Finally, using a simple dengue transmission model, I found that introducing dengue vaccine may temporarily cause large spikes in prevalence when vaccine efficacy and coverage are not both very high or very low. Despite the presence of these spikes, vaccination always reduced the total number of infections over a period of 15 years after vaccine introduction, suggesting that vaccine is likely to be effective. As these spikes can pose immediate problems by overburdening the resources, policy makers should consider the possibility of transient spikes and prepare to mitigate the burden of these spikes while making vaccination program.

Appendices

Appendix A Stability analysis

A.1 Stability analysis of the vector–host model

For the vector–host model (3.1), scaling the state variables by their respective population sizes, to the proportions $h_S = H_S/H$, $h_I = H_I/H$, $h_R = H_R/H$, $v_S = V_S/V$ and $v_I = V_I/V$, gives the system of differential equations

$$\begin{aligned}
\frac{\mathrm{d}h_S}{\mathrm{d}t} &= \mu_H - \beta_{aH} v_I h_S - \mu_H h_S, \\
\frac{\mathrm{d}h_I}{\mathrm{d}t} &= \beta_{aH} v_I h_S - \gamma_H h_I - \mu_H h_I, \\
\frac{\mathrm{d}h_R}{\mathrm{d}t} &= \gamma_H h_I - \mu_H h_R, \\
\frac{\mathrm{d}v_S}{\mathrm{d}t} &= \mu_V - \beta_{aV} h_I v_S - \mu_V v_S, \\
\frac{\mathrm{d}v_I}{\mathrm{d}t} &= \beta_{aV} h_I v_S - \mu_V v_I.
\end{aligned} \tag{1}$$

Using the next-generation method (Diekmann et al., 2010), the basic reproductive number is

$$R_0 = \frac{\beta_{aH}\beta_{aV}}{\mu_V(\mu_H + \gamma_H)}. \tag{2}$$

Because $h_S + h_I + h_R = 1$ and $v_S + v_I = 1$, the reduced system

$$\begin{aligned}
\frac{\mathrm{d}h_S}{\mathrm{d}t} &= \mu_H - \beta_{aH} v_I h_S - \mu_H h_S, \\
\frac{\mathrm{d}h_I}{\mathrm{d}t} &= \beta_{aH} v_I h_S - \gamma_H h_I - \mu_H h_I, \\
\frac{\mathrm{d}v_I}{\mathrm{d}t} &= \beta_{aV} h_I v_S - \mu_V v_I,
\end{aligned} \tag{3}$$

is equivalent to the full system (1). This system is defined on the domain

$$\Omega = \{(h_S, h_I, v_I) : 0 \leq v_I \leq 1, 0 \leq h_S, 0 \leq h_I, h_S + h_I \leq 1\}. \tag{4}$$

A simple check shows that the vector field defined by model (3) on the boundary of Ω does not point to the exterior of Ω, so Ω is positively invariant under the flow induced by system (3). This guarantees that the model numbers of humans and mosquitoes in the various epidemiological compartments never become negative, which is an obvious biological constraint.

The equilibrium points of system (3) are

$$E_0 = (1, 0, 0) \qquad \text{and} \qquad E_e = (h_S^*, h_I^*, v_I^*), \tag{5}$$

where

$$h_S^* = \frac{\delta + M}{\delta + MR_0}, \qquad h_I^* = \frac{R_0 - 1}{\delta + MR_0}, \qquad v_I^* = \frac{\delta(R_0 - 1)}{(\delta + M)R_0}, \tag{6}$$

with

$$\delta = \frac{\beta_{aV}}{\mu_V} \qquad \text{and} \qquad M = \frac{\mu_H + \gamma_H}{\mu_H}. \tag{7}$$

E_0 is the disease-free equilibrium and E_e is the endemic equilibrium. For $R_0 < 1$, E_0 is the only equilibrium in Ω but the endemic equilibrium E_e also lies in Ω for $R_0 \geq 1$. The local stability of the equilibrium points is governed by the Jacobian matrix

$$\mathrm{D}F = \begin{bmatrix} -\beta_{aH} v_I - \mu_H & 0 & -\beta_{aH} h_S \\ \beta_{aH} v_I & -(\mu_H + \gamma_H) & \beta_{aH} h_S \\ 0 & \beta_{aV} - \beta_{aV} v_I & -\beta_{aV} h_I - \mu_V \end{bmatrix}. \tag{8}$$

A.1.1 Disease-free equilibrium

The Jacobian matrix (8) at E_0 is

$$\mathrm{D}F(E_0) = \begin{bmatrix} -\mu_H & 0 & -\beta_{aH} \\ 0 & -(\mu_H + \gamma_H) & \beta_{aH} \\ 0 & \beta_{aV} & -\mu_V \end{bmatrix}, \tag{9}$$

which has eigenvalues

$$-\mu_H \quad \text{and} \quad \frac{-(\mu_H + \gamma_H + \mu_V) \pm \sqrt{(\mu_H + \gamma_H + \mu_V)^2 - 4\mu_V(\mu_H + \gamma_H)(1 - R_0)}}{2}. \tag{10}$$

All of the eigenvalues have negative real part for $R_0 < 1$ and so E_0 is locally asymptotically stable for $R_0 < 1$.

To show global stability of E_0, we consider the Lyapunov function on interior of Ω

$$\Lambda = \frac{\beta_{aH}}{\mu_V} v_I + h_I \tag{11}$$

which has orbital derivative

$$\begin{aligned} \frac{\mathrm{d}\Lambda}{\mathrm{d}t} &= \frac{\beta_{aH}}{\mu_V} \frac{\mathrm{d}v_I}{\mathrm{d}t} + \frac{\mathrm{d}h_I}{\mathrm{d}t} \\ &= -\beta_{aH}(1 - h_S)v_I - (\mu_H + \gamma_H)[1 - R_0(1 - v_I)]h_I. \end{aligned} \tag{12}$$

For $R_0 \leq 1$, the orbital derivative $\frac{\mathrm{d}\Lambda}{\mathrm{d}t} \leq 0$ in Ω and the subset of Ω where $\frac{\mathrm{d}\Lambda}{\mathrm{d}t} = 0$ is given by

$$(1 - h_S)v_I = 0 \tag{13}$$

and

$$\begin{aligned} h_I &= 0 \quad \text{if } R_0 < 1, \\ h_I v_I &= 0 \quad \text{if } R_0 = 1. \end{aligned} \tag{14}$$

Thus $\{E_0\}$ is the only invariant set contained in $\frac{\mathrm{d}\Lambda}{\mathrm{d}t} = 0$. Also, the interior of Ω is bounded. Therefore, E_0 is locally stable and all trajectories starting in Ω approach E_0 as $t \to +\infty$ (Hale, 1969, p. 317, Corollary 1.1). This establishes the global asymptotic stability of E_0 for $R_0 \leq 1$.

For $R_0 > 1$, the eigenvalue

$$\frac{-(\mu_H + \gamma_H + \mu_V) + \sqrt{(\mu_H + \gamma_H + \mu_V)^2 - 4\mu_V(\mu_H + \gamma_H)(1 - R_0)}}{2} > 0, \tag{15}$$

so E_0 is unstable.

A.1.2 Endemic equilibrium

As R_0 increases through 1, the disease-free equilibrium E_0 becomes unstable and the endemic equilibrium E_e moves from outside to inside Ω. The Jacobian matrix at E_e is

$$\mathrm{D}F(E_e) = \begin{bmatrix} -\mu_H \frac{\delta + MR_0}{\delta + M} & 0 & -\frac{\mu_H M R_0}{\delta} \frac{\delta + M}{\delta + MR_0} \\ \frac{\mu_H M(R_0 - 1)}{\delta + M} & -\mu_H M & \frac{\mu_H M R_0}{\delta} \frac{\delta + M}{\delta + MR_0} \\ 0 & \frac{\mu_V \delta}{R_0} \frac{\delta + MR_0}{\delta + M} & -\mu_V R_0 \frac{\delta + M}{\delta + MR_0} \end{bmatrix}. \tag{16}$$

The characteristic polynomial of matrix (16) is

$$p(\lambda) = \lambda^3 + A\lambda^2 + B\lambda + C, \tag{17}$$

where

$$
\begin{aligned}
A &= \mu_H \frac{\delta + MR_0}{\delta + M} + \mu_H M + \mu_V R_0 \frac{\delta + M}{\delta + MR_0} \\
B &= \mu_H^2 M \frac{\delta + MR_0}{\delta + M} + \mu_V \mu_H R_0 + \frac{\mu_H \mu_V M(R_0 - 1)\delta}{\delta + MR_0} \\
C &= \mu_H^2 \mu_V M(R_0 - 1).
\end{aligned}
\tag{18}
$$

For $R_0 > 1$, the coefficients A, B, and C are positive and

$$
AB > \mu_H^2 \mu_V M R_0 > C, \tag{19}
$$

so the characteristic polynomial (17) satisfies the Routh–Hurwithz conditions (Brauer and Castillo-Chávez, 2001). Therefore, E_e is locally asymptotically stable.

A.2 Comparing equilibria of the vector–host and SIR model

The endemic equilibrium (6) for the vector–host model has

$$
\begin{aligned}
h_S^* &= \frac{\delta + M}{\delta + MR_0} = \frac{1 + \frac{\mu_H}{\beta_{aH}} R_0}{R_0 + \frac{\mu_H}{\beta_{aH}} R_0}, \\
h_I^* &= \frac{R_0 - 1}{\delta + MR_0} = (R_0 - 1) \frac{\mu_H}{(\mu_H + \gamma_H)\left(R_0 + \frac{\mu_H}{\beta_{aH}} R_0\right)}.
\end{aligned}
\tag{20}
$$

If

$$
\frac{\mu_H}{\beta_{aH}} R_0 \ll 1, \tag{21}
$$

as is true of our ML estimates, then since $R_0 > 1$,

$$
\begin{aligned}
h_S^* &\approx \frac{1}{R_0}, \\
h_I^* &\approx (R_0 - 1) \frac{\mu_H}{(\mu_H + \gamma_H) R_0} = (R_0 - 1) \frac{\mu_H}{\beta},
\end{aligned}
\tag{22}
$$

where

$$\beta = \frac{\beta_{aV}\beta_{aH}}{\mu_V}, \tag{23}$$

which is exactly the endemic equilibrium of the SIR model (3.3).

Appendix B Steady-state solution

Rewriting the model (4.1) in proportion form by scaling the state variables by their respective population sizes, to the proportions $h_S = H_S/H$, $h_I = H_I/H$ and $h_R = H_R/H$, gives

$$\begin{aligned} \frac{\mathrm{d}h_S}{\mathrm{d}t} &= \mu_H - \beta h_S h_I - \mu_H h_S, \\ \frac{\mathrm{d}h_I}{\mathrm{d}t} &= \beta h_S h_I - (\gamma_H + \mu_H) h_I \\ \frac{\mathrm{d}h_R}{\mathrm{d}t} &= \gamma_H h_I - \mu_H h_R, \end{aligned} \tag{24}$$

The endemic steady-state solution of model (24) is

$$h_S^* = \frac{\gamma_H + \mu_H}{\beta}, \qquad h_I^* = \frac{\mu_H(\beta - \gamma_H - \mu_H)}{\beta(\mu_H + \gamma_H)}, \qquad h_R^* = \frac{\gamma_H(\beta - \gamma_H - \mu_H)}{\beta(\mu_H + \gamma_H)}. \tag{25}$$

The endemic steady-state solution can be re-written as

$$h_S^* = \frac{1}{R_0}, \qquad h_I^* = \frac{\mu_H(R_0 - 1)}{R_0(\mu_H + \gamma_H)}, \qquad h_R^* = \frac{\gamma_H(R_0 - 1)}{R_0(\mu_H + \gamma_H)}. \tag{26}$$

where $R_0 = \beta/(\gamma_H + \mu_H)$ is the basic reproduction number.

The cumulative class h_C in proportion form is

$$\frac{\mathrm{d}h_C}{\mathrm{d}t} = p\beta h_S h_I, \tag{27}$$

and when system is in steady-state, the force of infection is $h_S = h_S^*$ and $h_I = h_I^*$. So, integrating (27) when system is in steady-state gives, $h_C(t) = p\beta h_S^* h_I^* t + C$ and

by plugging $t = 1$ year, we get

$$h_C(1) = \frac{p\mu_H(R_0 - 1)}{R_0} + h_C(0). \tag{28}$$

and thus

$$\frac{R_0 - 1}{R_0} = \frac{h_C(1) - h_C(0)}{p\mu_H} \tag{29}$$

Substituting (29) in the expression for h_R^* in (26), we get a relation between the population immunity and the cumulative DHF after one year:

$$h_R^* = \frac{(H_C(1) - H_C(0))\gamma_H}{(\mu_H + \gamma_H)p\mu_H H}. \tag{30}$$

where $H_C(0)$ is cumulative DHF incidence in the beginning of epidemic and $H_C(0)$ is cumulative DHF incidence over one year. H is total population.

Appendix C Dengue vaccination models

C.1 Main model

We developed a simple deterministic model for dengue to explore the short- and long-term effects of vaccine introduction. The model consists of four unvaccinated (U), four vaccinated (V) and one recovered (R) compartment, the variables for which represent the number of people in that compartment over time (Figure 5.1). An individual with no prior dengue infection (subscript S) can have a primary infection from any serotype of dengue ($I1$), recovery from which provides him with life-long immunity to that serotype ($S2$). He can then acquire infection from any of the remaining three serotypes, and enter the secondary-infection class ($I2$). Third or fourth infection from dengue is very rare (Halstead, 2003), so we assume that an individual recovering from secondary infection becomes immune to all serotypes (R). Individuals in both primary- and secondary-infection class recover at rate $\gamma = 7/365\,\mathrm{y}^{-1}$ so that the mean infectious period is 7 days (Gubler, 1998a). The parameters B and μ are natural birth and death rates in humans respectively, which were chosen to be $B = \mu = 1/50\,\mathrm{y}^{-1}$ so that the mean human lifespan is 50 years and the population size is constant. Dengue-induced mortality in humans is around 1–2% (Rajapakse, 2011) and thus ignored in our model for simplicity. We were only interested in infections in humans, so we did not model mosquito population explicitly (Pandey et al., 2013). Infection occurs when a susceptible individual comes in contact with individuals from any of the infectious classes through mosquito bites at the rate given by the force of infection

$$\lambda(t) = [1 + \epsilon \cos(2\pi t)] \frac{\beta I(t)}{N}, \tag{31}$$

where the total number of people currently infected is

$$I(t) = U_{I1}(t) + U_{I2}(t) + V_{I1}(t) + V_{I2}(t), \tag{32}$$

N is the total human population size and β is a composite transmission parameter (Pandey et al., 2013), taken to be $\beta = 400\,\mathrm{y}^{-1}$ (Cummings et al., 2005). Seasonality in the mosquito population was captured by the cosine term in the force of infection, where t is units of years, making transmission most intense at the beginning of a year (i.e. when t is near an integer $0, 1, 2, ...$) and least intense in the middle of a year (t near $0.5, 1.5, 2.5, ...$). We used a small amplitude of seasonal forcing, $\epsilon = 0.1$. The force of infection for secondary infection was reduced by the factor $\sigma = \frac{3}{4}$ since they are only susceptible to 3 of the 4 serotypes.

The vaccination program with two components was modeled: a proportion p of newborns are vaccinated and rest of the susceptible population is vaccinated at rate v. The vaccine of efficacy ϕ was modeled as reducing the force of infection by the factor $1 - \phi$.

Parameter	Definition	Values
β	transmission parameter (y^{-1})	400
ϵ	seasonal-forcing amplitude	0.1
B	constant birth rate (y^{-1})	20/1000
μ	natural mortality rate (y^{-1})	20/1000
γ	recovery rate (y^{-1})	365/7
σ	susceptibility reduction for secondary infection	3/4
p	proportion of newborns vaccinated	(0–1)
v	vaccination rate for rest of population	(≥ 0)
ϕ	vaccine efficacy	(0–1)

Table C.1: Parameters of the dengue vaccination models.

The model equations are

$$\begin{aligned}
\frac{\mathrm{d}U_{S1}}{\mathrm{d}t} &= (1-p)BN - (\lambda + v + \mu)U_{S1}, \\
\frac{\mathrm{d}U_{I1}}{\mathrm{d}t} &= \lambda U_{S1} - (\gamma + \mu)U_{I1}, \\
\frac{\mathrm{d}U_{S2}}{\mathrm{d}t} &= \gamma U_{I1} - (\sigma\lambda + v + \mu)U_{S2}, \\
\frac{\mathrm{d}U_{I2}}{\mathrm{d}t} &= \sigma\lambda U_{S2} - (\gamma + \mu)U_{I2}, \\
\frac{\mathrm{d}V_{S1}}{\mathrm{d}t} &= pBN + vU_{S1} - [(1-\phi)\lambda + \mu]V_{S1}, \\
\frac{\mathrm{d}V_{I1}}{\mathrm{d}t} &= (1-\phi)\lambda V_{S1} - (\gamma + \mu)V_{I1}, \\
\frac{\mathrm{d}V_{S2}}{\mathrm{d}t} &= \gamma V_{I1} + vU_{S2} - [\sigma(1-\phi)\lambda + \mu]U_{S2}, \\
\frac{\mathrm{d}V_{I2}}{\mathrm{d}t} &= \sigma(1-\phi)\lambda V_{S2} - (\gamma + \mu)V_{I2}, \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma(U_{I2} + V_{I2}) - \mu R.
\end{aligned} \tag{33}$$

The parameters of the model are also shown in Table C.1.

For the pre-vaccine state, using parameter values in Table C.1, setting both vaccination rates p and v to 0, setting the initial conditions for unvaccinated susceptible individuals with no prior infection, unvaccinated infected individuals with no

prior infection and permanently recovered individuals to be 0.09, 0.01 and 0.9, with all other initial conditions to be zero, we simulated the dengue model (33) until it reached a limit cycle, a regular periodic oscillation. With the parameter values used, the model (33) converged to a 2-year cycle (Figure 5.2 and 5.3). We defined $t = 0$ to be just before the year with the larger peak in prevalence (e.g. see Figure 5.2(A)). In most of the simulations, vaccination was begun at $t_v = 0$, but we also varied the start time $t_v \in [0, 2)$ (Figure 5.6).

To measure effectiveness of the vaccination program, we used 15-year maximum prevalence

$$\max_{t\in[t_v,t_v+15]} I(t), \tag{34}$$

15-year total infections

$$\gamma \int_{t_v}^{t_v+15} I(t)\ \mathrm{d}t \tag{35}$$

and long-term annual incidence

$$\frac{\gamma}{P} \int_{t_c}^{t_c+P} I(t)\ \mathrm{d}t, \tag{36}$$

where t_c is a time after the system has converged to the post-vaccine limit cycle and P is the period of that limit cycle.

C.2 Simpler model

For our simpler dengue model, we assumed that there is no secondary infection so that an individual after recovering from a dengue infection is immune to all

serotypes. This reduces our original dengue model (33) to

$$\begin{aligned}
\frac{\mathrm{d}U_{S1}}{\mathrm{d}t} &= (1-p)BN - (\lambda + v + \mu)U_{S1}, \\
\frac{\mathrm{d}U_{I1}}{\mathrm{d}t} &= \lambda U_{S1} - (\gamma + \mu)U_{I1}, \\
\frac{\mathrm{d}V_{S1}}{\mathrm{d}t} &= pBN + vU_{S1} - [(1-\phi)\lambda + \mu]V_{S1}, \\
\frac{\mathrm{d}V_{I1}}{\mathrm{d}t} &= (1-\phi)\lambda V_{S1} - (\gamma + \mu)V_{I1}, \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma(U_{I1} + V_{I1}) - \mu R,
\end{aligned} \tag{37}$$

with the same force of infection (31), but with total number infected

$$I(t) = U_{I1}(t) + V_{I1}(t). \tag{38}$$

C.3 More complex model

A more complex model compared to the model (33) was obtained by adding the hypothesis of short-term cross-protection after primary infection. Wearing and Rohani (2006) hypothesized that after recovering from primary infection, an individual has short-term cross-protection to the other serotypes for about 9 months, after which he becomes susceptible to the remaining three serotypes. We added two classes of people temporarily immune to all serotypes, one for unvaccinated (U_T) and one for vaccinated (V_T) people, to our existing model. After primary infection, people now enter these temporary immune classes, and before moving on to be susceptible

to secondary infection. The model equations for this new model are

$$\begin{aligned}
\frac{\mathrm{d}U_{S1}}{\mathrm{d}t} &= (1-p)BN - (\lambda + v + \mu)U_{S1}, \\
\frac{\mathrm{d}U_{I1}}{\mathrm{d}t} &= \lambda U_{S1} - (\gamma + \mu)U_{I1}, \\
\frac{\mathrm{d}U_{T}}{\mathrm{d}t} &= \gamma U_{I1} - (\rho + \mu)U_{T}, \\
\frac{\mathrm{d}U_{S2}}{\mathrm{d}t} &= \rho U_{T} - (\sigma\lambda + v + \mu)U_{S2}, \\
\frac{\mathrm{d}U_{I2}}{\mathrm{d}t} &= \sigma\lambda U_{S2} - (\gamma + \mu)U_{I2}, \\
\frac{\mathrm{d}V_{S1}}{\mathrm{d}t} &= pBN + vU_{S1} - [(1-\phi)\lambda + \mu]V_{S1}, \\
\frac{\mathrm{d}V_{I1}}{\mathrm{d}t} &= (1-\phi)\lambda V_{S1} - (\gamma + \mu)V_{I1}, \\
\frac{\mathrm{d}V_{T}}{\mathrm{d}t} &= \gamma V_{I1} - (\rho + \mu)V_{T}, \\
\frac{\mathrm{d}V_{S2}}{\mathrm{d}t} &= \rho V_{T} + vU_{S2} - [\sigma(1-\phi)\lambda + \mu]U_{S2}, \\
\frac{\mathrm{d}V_{I2}}{\mathrm{d}t} &= \sigma(1-\phi)\lambda V_{S2} - (\gamma + \mu)V_{I2}, \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma(U_{I2} + V_{I2}) - \mu R,
\end{aligned} \tag{39}$$

where $\frac{1}{\rho} = 0.75\,\mathrm{y}$ is the average duration of short-term cross protection against all serotypes. With the same force of infection and total number infected as before, (31) & (32).

Bibliography

B. Adams, E.C. Holmes, C. Zhang, M.P. Mammen Jr., S. Nimmannitya, S. Kalayanarooj, and M. Boots. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. *Procedings of the National Academy of Sciences USA*, 103(38):14234–14239, 2006.

K.T. Alligood, T.D. Sauer, and J.A. Yorke. *Chaos: An Introduction to Dynamical Systems*. Springer, New York, 1996.

R.M. Anderson and R.M. May. *Infectious Diseases of Humans: Dynamics and Control*, volume 26. Wiley Online Library, 1992.

M. Beatty, M.F. Boni, S. Brown, R. Buathong, D. Burke, L. Coudeville, D.A. Cummings, R. Edelman, J. Farrar, D.A. Focks, et al. Assessing the potential of a candidate dengue vaccine with mathematical modeling. *PLoS Neglected Tropical Diseases*, 6(3):e1450, 2012.

S.M. Blower and H. Dowlatabadi. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *International Statistical Review*, 62(2):229–243, 1994.

F. Brauer and C. Castillo-Chávez. *Mathematical Models in Population Biology and Epidemiology*. Springer, New York, 2001.

J.S. Brownstein, T.R. Holford, and D. Fish. Enhancing West Nile virus surveillance, United States. *Emerging Infectious Diseases*, 10(6):1129–1133, 2004.

M.N. Burattini, M. Chen, A. Chow, F.A.B. Coutinho, K.T. Goh, L.F. Lopez, S. Ma, and E. Massad. Modelling the control strategies against dengue in Singapore. *Epidemiology And Infection*, 136(3):309–319, 2008.

K.P. Burnham and D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, 2nd edition, 2002.

S. Cauchemez, F. Carrat, C. Viboud, A.J. Valleron, and P.Y. Boelle. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in Medicine*, 23(22):3469–3487, 2004.

Center for Disease Control and Prevention. *Dengue*, Accessed January 14, 2013. URL `http://www.cdc.gov/dengue/`.

D.L. Chao, S.B. Halstead, M.E. Halloran, and I.M. Longini, Jr. Controlling dengue with vaccines in Thailand. *PLoS Neglected Tropical Diseases*, 6(10):e1876, 2012.

B.C.K. Choi and A.W.P. Pak. A simple approximate mathematical model to predict the number of severe acute respiratory syndrome cases and deaths. *Journal of Epidemiology and Community Health*, 57(10):831–835, 2003.

G. Chowell, P. Diaz-Dueñas, J.C. Miller, A. Alcazar-Velazco, J.M. Hyman, P.W. Fenimore, and C. Castillo-Chavez. Estimation of the reproduction number of dengue fever from spatial epidemic data. *Mathematical Biosciences*, 208(2):571–589, 2007.

D.V. Clark, M.P. Mammen, A. Nisalak, V. Puthimethee, and T.P. Endy. Economic impact of dengue fever/dengue hemorrhagic fever in thailand at the family and population levels. *The American Journal of Tropical Medicine and Hygiene*, 72(6): 786–791, 2005.

B.G. Coller and D.E. Clements. Dengue vaccines: progress and challenges. *Current Opinion in Immunology*, 23(3):391–398, 2011.

M.K. Cowles and B.P. Carlin. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, pages 883–904, 1996.

D.A.T. Cummings, I.B. Schwartz, L.B. Shaw L. Billings, and D.S. Burke. Dynamic effects of antibody-dependent enhancement on the fitness of viruses. *Proceedings of the National Academy of Sciences USA*, 102(42):15259–15264, 2005.

D.A.T Cummings, S. Iamsirithaworn, J.T. Lessler, A. McDermott, R. Prasanthong, A. Nisalak, R.G. Jarman, D.S. Burke, and R.V. Gibbons. The impact of the demographic transition on dengue in Thailand: insights from a statistical analysis and mathematical modeling. *PLoS Medicine*, 6(9):e1000139, 2009.

O. Diekmann, J.A.P. Heesterbeek, and M.G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of Royal Society Interface*, 7(47):873–885, 2010.

S. Duncan and M. Gyöngy. Using the EM algorithm to estimate the disease parameters for smallpox in 17th Century London. In *Proceedings of the 2006 IEEE International Conference on Control Applications*, pages 3312–3317, Madison, Wisconsin, 2006. Omnipress.

L. Esteva and H. Mo Yang. Mathematical model to assess the control of *Aedes aegypti* mosquitoes by the sterile insect technique. *Mathematical Biosciences*, 198 (2):132–147, 2005.

L. Esteva and C. Vargas. Influence of vertical and mechanical transmission on the dynamics of dengue disease. *Mathematical Biosciences*, 167(1):51–64, 2000a.

L. Esteva and C. Vargas. Influence of vertical and mechanical transmission on the dynamics of dengue disease. *Mathematical Biosciences*, 167(1):51–64, 2000b.

N. Ferguson, R. Anderson, and S. Gupta. The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proceedings of the National Academy of Sciences, USA*, 96(2):790, 1999a.

N.M. Ferguson, C.A. Donnelly, and R.M. Anderson. Transmission dynamics and epidemiology of dengue: insights from age–stratified sero–prevalence surveys. *Philosophical Transactions B*, 354(1384):757–768, 1999b.

N.M. Ferguson, C.A. Donnelly, M.E.J. Woolhouse, and R.M. Anderson. Estimation of the basic reproduction number of BSE: the intensity of transmission in british cattle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1414):23–32, 1999c.

D.A. Focks, D.G. Haile, E. Daniels, and G.A. Mount. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development. *Journal of Medical Entomology*, 30(6):1003–1017, 1993a.

D.A. Focks, D.G. Haile, E. Daniels, and G.A. Mount. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): simulation results and validation. *Journal of Medical Entomology*, 30(6):1018–1028, 1993b.

D.A. Focks, E. Daniels, D.G. Haile, and J.E. Keesling. 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(5):489, 1995.

D.A. Focks, R.J. Brenner, J. Hayes, and E. Daniels. Transmission thresholds for dengue in terms of *Aedes aegypti* pupae per person with discussion of their utility in source reduction efforts. *The American Journal of Tropical Medicine and Hygiene*, 62(1):11, 2000.

A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

A. Gelman and D.B. Rubin. Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research*, 5(4):339–355, 1996.

A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, 2nd edition, 2004.

B.T. Grenfell, O.N. Bjørnstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.

D.J. Gubler. Dengue and dengue hemorrhagic fever. *Clinical Microbiology Reviews*, 11(3):480–496, 1998a.

D.J. Gubler. Resurgent vector-borne diseases as a global health problem. *Emerging Infectious Diseases*, 4(3):442–450, 1998b.

M.G. Guzmán and G. Kouri. Dengue: an update. *Lancet Infectious Diseases*, 2(1): 33–42, 2002.

M.G. Guzmán, G. Kouri, J. Bravo, L. Valdes, S. Vasquez, and S.B. Halstead. Effect of age on outcome of secondary dengue 2 infections. *International Journal of Infectious Diseases*, 6(2):118–124, 2002.

M.G. Guzmán, S.B. Halstead, H. Artsob, P. Buchy, J. Farrar, D.J. Gubler, E. Hunsperger, A. Kroeger, H.S. Margolis, E. Martínez, et al. Dengue: a continuing global threat. *Nature Reviews Microbiology*, 8:S7–S16, 2010.

R. Hagmann, J.D. Charlwood, V. Gil, C. Ferreira, V. Do Rosário, and T.A. Smith. Malaria and its possible control on the island of príncipe. *Malaria Journal*, 2(1): 15, 2003.

J.K. Hale. *Ordinary Differential Equations*. Pure and Applied Mathematics, John Wiley & Sons, New York, 1969.

S.B. Halstead. Neutralization and antibody-dependent enhancement of dengue viruses. *Advances in Virus Research*, 60:421–467, 2003.

S.B. Halstead. Dengue. *The Lancet*, 370(9599):1644–1652, 2007.

S.B. Halstead. Dengue vaccine development: a 75% solution? *The Lancet*, 380(9853): 1535–1536, 2012.

J.M. Heffernan, R.J. Smith, and L.M. Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293, 2005.

H. W. Hethcote. The mathematics of infectious disease. *SIAM Review*, 42(4):599–653, 2000.

Y. Huang, D. Liu, and H. Wu. Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics*, 62(2):413–423, 2006.

Johns Hopkins Center for Immunization Research. *Monthly incidence of Dengue Hemorrhagic Fever, 1983–1997, in 72 provinces of Thailand*, Accessed January 14, 2013. URL `http://web.archive.org/web/20041216074102/http://www.jhsph.edu/cir/dengue.html`.

M.J. Keeling and B.T. Grenfell. Disease extinction and community size: modeling the persistence of measles. *Science*, 275(5296):65–67, 1997.

M.J. Keeling and P. Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, Princeton, 2011.

A.A. King and W.M. Schaffer. The rainbow bridge: Hamiltonian limits and resonance in predator-prey dynamics. *Journal of Mathematical Biology*, 39(5):439–469, 1999.

A.A. King, W.M. Schaffer, C. Gordon, J. Treat, and M. Kot. Weakly dissipative predator–prey systems. *Bulletin of Mathematical Biology*, 58(5):835–859, 1996.

M. Lavielle, A. Samson, A. Karina Fermin, and F. Mentré. Maximum likelihood estimation of long-term HIV dynamic models and antiviral response. *Biometrics*, 67(1):250–259, 2011.

J. Li, D. Blakeley, and R. Smith? The failure of 0. *Computational and mathematical methods in medicine*, 2011:17 pages, 2011.

P.M. Luz, C.T. Codeço, E. Massad, and C.J. Struchiner. Uncertainties regarding dengue modeling in Rio de Janeiro, Brazil. *Memórias do Instituto Oswaldo Cruz*, 98(7):871–878, 2003.

G. Macdonald. *The Epidemiology and Control of Malaria*. Oxford Univeristy Press, New York, 1957.

J. Medlock, P.M. Luz, C.J. Struchiner, and A.P. Galvani. The impact of transgenic mosquitoes on dengue virulence to humans and mosquitoes. *The American Naturalist*, 174(4):565–577, 2009.

A. Mubayi, C. Castillo-Chavez, G. Chowell, C. Kribs-Zaleta, N. Ali Siddiqui, and N. Kumar. Transmission dynamics and underreporting of Kala-Azar in the Indian state of Bihar. *Journal of Theoretical Biology*, 262(1):177–185, 2010.

Y. Nagao and K. Koelle. Decreases in dengue transmission may act to increase the incidence of dengue hemorrhagic fever. *Proceedings of National Academy of Sciences USA*, 105(6):2238–2243, 2008.

A. Nisalak, T.P. Endy, S. Nimmannitya, S. Kalayanarooj, R.M. Scott, D.S. Burke, C.H. Hoke, B.L. Innis, D.W. Vaughn, et al. Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999. *The American Journal of Tropical Medicine and Hygiene*, 68(2):191–202, 2003.

A. Pandey, A. Mubayi, and J. Medlock. Comparing vector–host and SIR models for dengue transmission. *Mathematical Biosciences*, 246(2):252–259, 2013.

S. Rajapakse. Dengue shock. *Journal of Emergencies, Trauma and Shock*, 4(1):120, 2011.

N.G. Reich, S. Shrestha, A.A. King, P. Rohani, J. Lessler, S. Kalayanarooj, I. Yoon, R.V. Gibbons, D.S. Burke, and D.A.T. Cummings. Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *Journal of The Royal Society Interface*, 10(86):20130414, 2013.

J. Reiczigel, K. Brugger, F. Rubel, N. Solymosi, and Z. Lang. Bayesian analysis of a dynamical model for the spread of the Usutu virus. *Stochastic Environmental Research and Risk Assessment*, 24(3):455–462, 2010.

R. Rico-Hesse. Microevolution and virulence of dengue viruses. *Advances in Virus Research*, 59:315–341, 2003.

H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pages 400–407, 1951.

A.B. Sabin. Research on dengue during World War II. *The American Journal of Tropical Medicine and Hygiene*, 1(1):30–50, 1952.

M.A. Sanchez and S.M. Blower. Uncertainty and sensitivity analysis of the basic reproductive rate: tuberculosis as an example. *American Journal of Epidemiology*, 145(12):1127–1137, 1997.

J.F. Siler, M.W. Hall, and A.P. Hitchens. Dengue: its history, epidemiology, mechanism of transmission, etiology, clinical manifestations, immunity, and prevention. *Philippine Journal of Science*, 29(20):1–304, 1926.

M. Stevenson. *epiR: An R Package for the Analysis of Epidemiological Data*, Accessed March 10, 2013. URL `http://cran.r-project.org/web/packages/epiR/index.html`.

S.H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, Colorado, 1994.

United Nations. *World Fertility Data 2012*, Accessed March 25, 2013. URL `http://www.un.org/esa/population/publications/WFD2012/MainFrame.html`.

H.J. Wearing and P. Rohani. Ecological and immunological determinants of dengue epidemics. *Proceedings of the National Academy of Sciences USA*, 103(31):11802–11807, 2006.

WHO. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control*. World Health Organization, 2009. ISBN 9241547871.

M.J. Wonham, T. de Camino-Beck, and M.A. Lewis. An epidemiological model for West Nile virus: invasion analysis and control applications. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1538):501–507, 2004.

World Bank. *Life Expectancy at Birth, Total (Years)*, Accessed January 14, 2013. URL `http://data.worldbank.org/indicator/SP.DYN.LE00.IN/countries/TH-4E-XT?page=5&display=default`.