Spring 2019

Bayesian compartmental models for zoonotic visceral leishmaniasis in the Americas

Marie Veronica Ozanne
University of Iowa

BAYESIAN COMPARTMENTAL MODELS FOR ZOONOTIC VISCERAL LEISHMANIASIS IN THE AMERICAS

by

Marie Veronica Ozanne

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Biostatistics
in the Graduate College of
The University of Iowa

May 2019

Thesis Supervisors: Professor Jacob J. Oleson
Assistant Professor Grant D. Brown

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Marie Veronica Ozanne

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Biostatistics at the May 2019 graduation.

Thesis committee: _____
Jacob J. Oleson, Thesis Supervisor

_____
Grant D. Brown, Thesis Supervisor

_____
Christine A. Petersen

_____
Mary E. Wilson

_____
Gideon Zamba

ACKNOWLEDGEMENTS

I consider myself fortunate to have been a part of the Department of Biostatistics at the University of Iowa, surrounded by vibrant and productive faculty, colleagues, and students. While many people have supported me throughout my graduate studies, I will to take this opportunity to thank those who made this dissertation possible.

First, I want to thank Dr. Gideon Zamba and Dr. Jacob Oleson for introducing me to Biostatistics through the Iowa Summer Institute in Biostatistics when I was an undergraduate. That experience changed the trajectory of my academic career and brought me where I am today. Next, I want to thank my advisors, Dr. Jacob Oleson and Dr. Grant Brown, for introducing me to an intriguing area of research, for challenging me to be a better statistician and communicator, and for always making time for me. I also want to thank my other PhD committee members, Dr. Petersen, Dr. Wilson, and Dr. Zamba, for sharing their time and expertise. Together, they have provided excellent guidance to ensure that content presented in this work is biologically relevant while remaining statistically sound. In addition, I want to thank the many other people who have contributed their data and expertise to this project, including Dr. Angela Toepp, Dr. Selma Jeronimo, Dr. Breanna Scorza, and members of the Petersen and Wilson laboratories. Finally, I want to thank my family and friends for all their support.

Research reported in this dissertation was supported by the Fogarty Interna-

tional Center of the National Institutes of Health under Award Number R01TW010500. The content is solely the author's responsibility and does not necessarily represent the official views of the National Institutes of Health.

ABSTRACT

Visceral leishmaniasis (VL) is a serious neglected tropical disease that is endemic in 98 countries and presents a significant public health risk. The epidemiology of VL is complex. In the Americas, it is a zoonotic disease that is caused by a parasite and transmitted among humans and dogs through the bite of an infected sand fly vector. The infection also can be transmitted vertically from mother to child during pregnancy. Infected individuals can be classified as asymptomatic or symptomatic; both classes can transmit infection. In part due to its complexity, VL transmission dynamics are not fully understood. Stochastic compartmental epidemic models are a powerful set of tools that can be used to study these transmission dynamics.

Past compartmental models for VL have been developed in a deterministic framework to accommodate complexity while remaining computationally tractable. In this work, we propose stochastic compartmental models for VL, which are simpler than their deterministic counterparts, but also have several advantages. Notably, this framework allows us to: (1) define a probability of infection transmission between two individuals, (2) obtain both parameter estimates and corresponding uncertainty measures, and (3) employ formal model comparisons.

In this dissertation, we develop both population level and individual level Bayesian compartmental models to study both vector and vertical VL transmission dynamics. As part of this model development, we introduce a compartmental model that allows for two infectious classes. We also derive source specific reproductive

numbers to quantify the contributions of different species and infectious classes to maintaining infection in a population. Finally, we propose a formal model comparison method for Bayesian models with high-dimensional discrete parameter spaces. These models, reproductive numbers, and model comparison method are explored in the context of simulations and real VL data from Brazil and the United States.

PUBLIC ABSTRACT

Visceral leishmaniasis (VL) is a potentially fatal neglected tropical disease that places hundreds of millions of people and millions of dogs at risk globally. In spite of its public health significance, the transmission dynamics of VL are not yet fully understood, in part due to the complexity of the infection process. Statistical models comprise a powerful set of tools that can be used to better understand this process and to study how different species and transmission routes contribute to it.

The first objective of this dissertation is to develop statistical models to study the transmission dynamics of visceral leishmaniasis in the Americas. We develop models to study the two main transmission modes: vector-borne as the result of a sand fly bite (humans and dogs) and *in utero* through transplacental transmission (dogs). Second, we derive reproductive numbers that allow us to quantify the importance of different classes of infectious individuals (e.g. infected dogs and humans) in maintaining infection in populations of interest. Third, we propose a formal model comparison method to determine which models are more appropriate to explain particular processes. We conclude with a discussion of how these methods enhance our ability to study VL and how we plan to expand on this work in the future.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

1.1 An Introduction to Visceral Leishmaniasis

Leishmaniasis is a complex and frequently fatal neglected tropical disease. It has several forms, the most serious of which is visceral. Visceral leishmaniasis (VL) is characterized by symptoms such as fever, weight loss, facial edema, hepatomegaly, and splenomegaly (Silva et al., 2001). In the Americas, VL is caused by the protozoan parasite *Leishmania infantum* (*L. infantum*) and it is zoonotic, meaning that it can be transmitted from other mammalian hosts to humans. This disease poses a threat to both humans and domesticated dogs, the latter of which serve as its primary animal reservoir. Leishmaniasis is endemic in 98 countries and three provinces, placing millions of humans and dogs at risk each year; in the Americas, 90% of cases occur in Brazil (Alvar et al., 2012; Desjeux, 2004; Toepp et al., 2017).

There are two main modes of transmission of *L. infantum*: vector-borne and vertical. While the former mode is known to present a major public health risk, the latter historically has been considered to be of minor importance in spreading infection. Recent work, however, by Petersen and Barr (2009) has indicated that vertical transmission is more important than previously was understood. In endemic areas in the Americas, like Brazil, infection primarily is spread among humans and dogs by the bite of an infected female *Lutzomyia longipalpis* sand fly vector (Desjeux, 2004; Petersen and Barr, 2009). In areas with vector transmission, vertical is a secondary

transmission mode. Vertical transmission is the process by which a pathogen is passed from mother to child in the time period before or after birth. While this transmission can occur via several different routes, *L. infantum* is transmitted transplacentally (da Silva et al., 2009). In some regions of the United States where vector transmission has been shown to be absent, VL is enzootic in some hound populations, where it is maintained in these canine populations through vertical transmission (Duprey et al., 2006; Petersen and Barr, 2009). This provides an ideal setting in which to study vertical transmission dynamics, since the rate of infection due to this transmission mode in areas with vector transmission is unknown due to the high probability of vector contact (da Silva et al., 2009; Mancianti et al., 1988).

To add to the complexity of VL, infectious individuals can be grouped into two infectious classes, asymptomatic and symptomatic, based on symptoms, diagnostic test results, and physical exam information. This means that there are four infectious groups: asymptomatic dogs, asymptomatic humans, symptomatic dogs, and symptomatic humans (Michel et al., 2011). While all of these groups might be able to transmit infection to uninfected vectors that bite them, a group's impact on maintaining infection in a population of interest depends on vector biting preference, the probability an uninfected sand fly becomes infected after biting a particular kind of infectious host, and the concentration of the infectious group in the population (Costa et al., 2002; Michel et al., 2011). It is important to account for these factors to adequately understand *L. infantum* transmission dynamics.

1.2 An Introduction to Compartmental Epidemic Models

Infectious diseases often have complex underlying processes which make modeling their dynamics challenging. One way to address this complexity is a compartmental epidemic model. The phrase "compartmental epidemic models" refers to a class of models that simplify true underlying infection processes by grouping the infection states of a pathogen of interest into compartments, as the name suggests. These models generally rely on two assumptions. First, individuals within the same compartment exhibit the same characteristics and infection behavior. Second, individuals in the population move through compartments in a particular order. These infection compartments are often discrete in nature.

One of the earliest contributions to compartmental epidemic modeling was made by Kermack and McKendrick (1927), who introduced discrete infection compartments in the context of a **S**usceptible, **I**nfectious, **R**emoved (SIR) model. Since then, numerous extensions have been proposed, including the **S**usceptible, **E**xposed, **I**nfectious, **R**emoved (SEIR) and carrier state models (Dietz, 1976; Keeling and Rohani, 2008; Li et al., 1999). These models can be approached either deterministically through differential equations or stochastically. The deterministic framework's most notable advantage is that, relative to stochastic compartmental models, analysis is more computationally tractable. Contrary to equivalent deterministic compartmental models, the stochastic models need to be relatively simple mathematically in order to be manageable (Andersson and Britton, 2000). When feasible, however, the stochastic framework has some notable advantages, as described by Andersson and

Britton (2000). First, one can define a probability of disease transmission between two individuals; this is a natural way to describe infectious disease spread because transmission does not occur deterministically. Second, the stochastic approach provides both estimates of our parameters and measures of uncertainty for those parameters. Third, stochastic techniques have formal methods for model comparison (Andersson and Britton, 2000).

Historically, stochastic compartmental models have not been applied to study either vertical or vector transmission dynamics of VL, likely due to the complexity of the infection process. However, deterministic models have been used for the last thirty years, mainly to study population-level transmission (Dye and Wolpert, 1988; Shimozako et al., 2017; Stauch et al., 2011). In spite of the absence of stochastic models developed for leishmaniasis, through judicious model specification, it is possible to study VL transmission dynamics through both main transmission modes (vertical and vector) while taking full advantage of the three stochastic model properties previously discussed. These models can be constructed at a population or an individual level, depending on the process of interest. From these models, we can also estimate the necessary quantities to calculate reproductive numbers for these infection processes, as detailed in Section 1.3.

1.3 An Introduction to Reproductive Numbers

A reproductive number is an important quantity in epidemiology that allows for quantification of the transmissibility of an infection in a population of interest. In

general, if a reproductive number is greater than one, this indicates that the infection will amplify in the population, and if it is less than one, the infection will die out. In its simplest form, the basic reproductive number is the total number of secondary cases resulting from a primary case in a totally susceptible population (Dietz, 1993). While appropriate in some applications, this definition can be limiting in two important ways. First, it requires a completely susceptible population. This assumption is seldom realistic and is inappropriate for the study of established infections in a population. Second, it remains constant over time, which is often too restrictive. For example, if there is seasonality to an infection, or if an intervention is introduced, a constant quantity cannot capture these changes; it needs to be updated under a number of specific conditions.

To address these limitations, researchers introduced more general reproductive numbers, including temporally varying and scaled (by the number of susceptible individuals) versions, known as "effective reproductive numbers" (Chowell et al., 2004; Lekone and Finkenstädt, 2006). Brown et al. (2016) introduced a more flexible estimate of the reproductive rate of an infection, the "empirically adjusted reproductive number". This formulation relies on the observation that to define a reproductive number in a real population, we only need to know the expected number of secondary infections produced by a single infected individual in a population of interest.

While the empirically adjusted reproductive number was developed in the context of a population-level stochastic spatial SEIR model with a single infectious source, it can be extended for application to an infection with multiple infectious

sources. To quantify the role of each source in maintaining infection in a population, we derive an **I**nfection **S**ource-specific **E**mpirically **A**djusted **R**eproductive **N**umber (ISEARN) in Chapter 3. Derivations rely on the assumption that an individual can be infected through only one source. In other words, if we have A, B, and C, which denote the events that an individual becomes infected with a particular pathogen as a result of source a, b, and c, then A, B, and C are mutually disjoint. Under this assumption, ISEARNs are additive, so they can be combined to recover an overall empirically adjusted reproductive number.

1.4 Bayesian Model Comparison Methods

As previously noted, stochastic techniques have formal methods for model comparison. Two metrics that are used for Bayesian model evaluation are Bayes factors and the Deviance Information Criterion (DIC) (Kass and Raftery, 1995; Spiegelhalter et al., 2002). DIC is one of the most popular metrics for comparing competing Bayesian models. Its form is similar to other information criteria: $DIC = D(\bar{\theta}) - 2p_D$, where $D(\bar{\theta})$ is the deviance of the average posterior draws and p_D is a penalty term that address model complexity (Spiegelhalter et al., 2002). While it tends to work well in many settings, it has been criticized for lack of a clear theoretical foundation (Plummer, 2008). In the compartmental modeling framework, there are a number of practical challenges associated with using DIC. First, the number of effective parameters in a compartmental model is large compared to the effective number of independent observations, which can be problematic for DIC (Plummer, 2008). Also,

compartmental likelihoods are not members of an exponential family; there has been criticism that DIC does not have a natural extension outside of this framework (Celeux et al., 2006). In addition, the form of the DIC requires a calculation of $\bar{\theta}$ to obtain a plug-in measure of fit, $\mathrm{D}(\bar{\theta})$. In a compartmental model with a large discrete parameter space, $\bar{\theta}$ is not guaranteed to lie within the parameter space (e.g. non-integer valued for binomial distribution), so some continuous approximation of this space is needed to calculate DIC. While this is feasible, it is not always obvious what continuous approximations will be most suitable.

Bayes factors, on the other hand, rest on sound theoretical footing and allow us to compare the probability that the observed data arose from one model relative to another. To use Bayes factors for model comparison, we require the use of proper priors for parameters that differ between two models of interest. If this requirement is satisfied, the remaining challenge to using Bayes factors for model comparison is calculation-based. The Bayes factor, BF_{12}, is defined by Kass and Raftery (1995) as

$$BF_{12} = \frac{P(\mathbf{Y}|\mathcal{M}_1)}{P(\mathbf{Y}|\mathcal{M}_2)} = \frac{P(\mathcal{M}_1|\mathbf{Y})}{P(\mathcal{M}_2|\mathbf{Y})} \cdot \frac{P(\mathcal{M}_2)}{P(\mathcal{M}_1)},$$

where $P(\mathbf{Y}|\mathcal{M}_i)$ is the probability of the observed data given Model i, $i \in \{1, 2\}$. These densities are obtained by integrating the joint posterior density, $P(\mathbf{Y}, \boldsymbol{\theta}_i|\mathcal{M}_i)$ over the parameter space, $\boldsymbol{\Theta}_i$. In all but the simplest cases where conjugate distributions are employed, this integration is intractable (Kass and Raftery, 1995), so one must turn to numerical methods to estimate these quantities. Considerable work has already been done to address this challenge. For example, Kass and Raftery (1995)

compare a variety of approximation methods. Laplace's method works well in situations where the likelihoods in question do not deviate strongly from normality. The Schwarz criterion relies on determining $\hat{\boldsymbol{\theta}}_i$, the maximum likelihood estimate (MLE) under Model i (Kass and Raftery, 1995). The authors also discuss using Monte Carlo integration to estimate these densities; this includes simple Monte Carlo, importance sampling, and adaptive Gaussian quadrature (the latter two methods improve efficiency). The most efficient procedure, adaptive Gaussian quadrature, is effective when the parameter space dimension of the model in question is modest (roughly < 9) (Kass and Raftery, 1995).

There is no reason that likelihoods associated with compartmental models should be close to normal. While there may be cases where normal approximations are adequate, part of the appeal of a compartmental model lies in its flexibility, so Laplace's method is not generally applicable. The high-dimensionality of the compartmental model parameter space presents problems both for the Schwarz criterion, because calculating the MLE is problematic, and for Monte Carlo integration techniques, because the dimension of the parameter space is too large. The dimensionality problem becomes even more extreme for spatio-temporal compartmental models. Approximate Bayesian computation (ABC) methods have been developed to fit large models with likelihoods that are difficult to evaluate (Toni et al., 2009). While we are able to evaluate likelihoods for the models presented in this dissertation, the framework for estimating Bayes factors for models fit using ABC gives us insight into constructing a method to estimate Bayes factors for models fit using MCMC where

at least one candidate models has a non-normal likelihood and a high-dimensional parameter space.

It is important to have methods to formally compare competing models for infectious disease dynamics in part because we can evaluate the impact of not accounting for some transmission modes or all infectious groups, for example. Specifically, in the context of VL, we can assess how excluding vertical transmission in an endemic area affects parameter estimates and reproductive numbers. We can also study the importance of distinguishing between asymptomatic and symptomatic individuals as opposed to grouping these individuals into a single infectious category. This is relevant because the former distinction requires more information to which we do not always have access. Formal model comparisons can provide insight into the cost to understanding transmission dynamics of not having that information.

1.5 Plan of Dissertation

In Chapter 2, we propose an individual level Bayesian compartmental model to study the dynamics of *L. infantum* vertical transmission in dogs in the absence of vector transmission. We also formulate reproductive numbers to quantify vertical transmissibility. We use this model to assess the role of maternal infection status during pregnancy in perpetuating *L. infantum* in the hunting hound population (in the absence of vector transmission) and investigate two hypotheses: pups born to diagnostically positive mothers (during pregnancy) (1) are more likely to progress to infection at some point in their lives, and (2) progress to infection more quickly than

those born to diagnostically negative mothers. We apply this model to surveillance data collected by the University of Iowa Petersen Lab from 2005-2018 in the United States.

In Chapters 3 and 4, we present population level Bayesian models to examine vector transmission dynamics of *L. infantum* in humans and canines in endemic areas. In Chapter 3, we formulate two models, a Bayesian multinomial logistic regression model and a stochastic SIR model, to estimate the proportions of individuals that fall into different infection categories because these population level infection state proportions are often difficult to observe, but are likely to be important determinantes of maintaining infection in an endemic area. These models are constructed for a single species to simplify the transmission process. We explore these models in the context of simulated data and historical human *L. infantum* transmission data from Brazil; canines are excluded and sand flies are assumed to acquire infection after biting an infected human with probability of one to simplify the process. Then in Chapter 4, we expand the compartmental model proposed in the previous section in two important ways. First, we incorporate humans, canines, and sand flies to study the entire process. Second, we separate infectious individuals into two classes, asymptomatic and symptomatic, and introduce a **S**usceptible, **A**symptomatic, s**Y**mptomatic, reco**V**ered, **R**emoved (SAYVR) model. These models are explored through simulation studies. For both proposed compartmental models, we present reproductive numbers. Specifically, we derive an **I**nfection **S**ource-specific **E**mpirically **A**djusted **R**eproductive **N**umber (ISEARN) to quantify the contributions of various infectious sources to

maintaining infection in the population.

In Chapter 5, we propose an ABC-inspired approach to approximate Bayes factors for models fit using MCMC. This method is especially useful when at least one candidate model does not have a normal likelihood and has a high-dimensional discrete parameter space. We employ this proposed technique to compare several models from this dissertation. First, we compare the multinomial logistic regression model and the stochastic SIR model performance from Section 3.1 for both simulations and real data analysis. Then, through simulation studies, we assess the importance of incorporating vertical transmission into a compartmental model in the presence of vector transmission and endemic infection. Finally, we evaluate the importance of separating infectious individuals into asymptomatic and symptomatic groups (rather than consolidating into one infectious group) in the context of simulations for visceral leishmaniasis.

In the last chapter, we evaluate the contributions made in Chapters 2-5. We also consider the limitations of these methods and propose future research directions.

CHAPTER 2
CANINE VISCERAL LEISHMANIASIS VERTICAL TRANSMISSION IN THE UNITED STATES

2.1 Vertical Transmission of Canine Visceral Leishmaniasis

In the United States, VL is enzootic in some hunting hound populations. Infection is maintained in the population through vertical transmission and vector transmission is absent (Petersen and Barr, 2009; Duprey et al., 2006; Schantz et al., 2005). In 1999, an outbreak of canine leishmaniasis was reported in a hound kennel in New York. The Centers for Disease Control (CDC) and Prevention, Division of Parasitic Diseases, began conducting widespread surveillance of kennels in 2000 (Petersen and Barr, 2009). The Petersen Lab at the University of Iowa are continuing surveillance efforts among participating kennels. This enables us to study the vertical transmission dynamics of VL, which have not been modeled widely and are not adequately understood.

Many infectious diseases can be transmitted both horizontally and vertically, including rubella, Chagas' disease, and AIDS (Busenberg and Cooke, 1993). Typically, vertical transmission is incorporated into a population compartmental model through a birth process, where some portion of the offspring are assumed to be infectious (in the case of an SIR model) or exposed (as in an SEIR model) at birth (Anderson and May, 1979; Li et al., 2001). This approach accounts for the vertical transmission contribution and is practical for analyses performed and conclusions drawn at the population level. Zou et al. (2017) account for transplacental transmis-

sion of VL in dogs in endemic areas in this fashion.

The objectives of the analyses in this chapter differ fundamentally from those that account for vertical transmission as described in the previous paragraph. First, there is no known vector transmission of VL in the United States, so we can study purely vertical transmission dynamics. In contrast, previous papers have focused on accounting for vertical transmission in the presence of horizontal transmission (Anderson and May, 1979; Li et al., 2001; Zou et al., 2017). Second, we are interested in understanding *L. infantum* infection progression after transplacental exposure, so we condition on the birth process in our model. Third, we have individual level data, so we can explicitly study how age and mother's status while pregnant (diagnostically positive or diagnostically negative) affect this process.

We introduce an individual-level Bayesian SIR model to study the infection progression of individuals that were exposed to infection *in utero*. At each time point, each individual has his/her own probability of transitioning from one infection state to another. Possible transitions are: $\mathcal{S} \rightarrow \mathcal{I}$, $\mathcal{I} \rightarrow \mathcal{R}$, and $\mathcal{S} \rightarrow \mathcal{R}$. Figure 2.1 provides a model schematic. This model generally can be used to study the effect of different levels of exposure to an infectious agent on progression over the course of an individual's life. For example, HIV can be transmitted vertically, and infected mothers can experience disease on a spectrum: acute, latent, and late (Abu-Raddad and Longini Jr, 2008). It can also be used to study transmission dynamics at an individual, rather than a population level, which is one way to avoid the assumption of a homogeneous contact process.

In this chapter, we use this SIR model to assess the role of maternal infection status during pregnancy in perpetuating *L. infantum* in the hunting hound population. Specifically, we use this model to explore the role that maternal infection status plays in perpetuating *L. infantum* infection in the population. To this end, we will investigate two hypotheses: pups born to mothers who are diagnostically positive at the time of pregnancy (1) are more likely to progress to infection at some point in their lives, and (2) progress to infection more quickly than those born to diagnostically negative mothers. We apply this model to data collected on hounds through an ongoing surveillance study conducted by the University of Iowa Petersen lab from 2005-2018 in the United States.

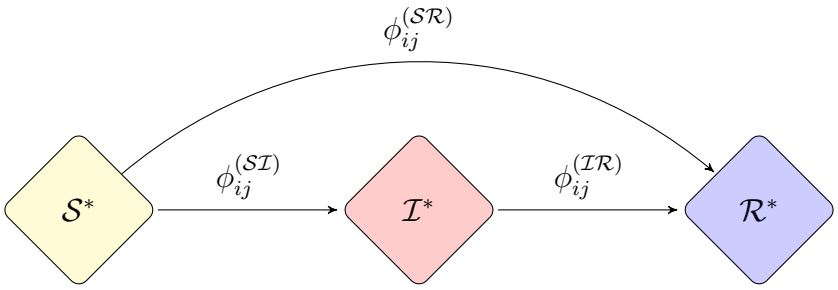

Figure 2.1: Schematic for the vertical SIR model. The vertically-derived infection progression process is at the individual level. At time j, each individual i has a different transition probability. The compartments are defined as follows: $\mathcal{S}^*$ is the set of vertically exposed, diagnostically negative individuals; $\mathcal{I}^*$ is the set of vertically exposed, diagnostically positive individuals; $\mathcal{R}^*$ is the set of vertically exposed, deceased individuals. The parameters $\phi_{ij}^{(\mathcal{SI})}$, $\phi_{ij}^{(\mathcal{SR})}$, and $\phi_{ij}^{(\mathcal{IR})}$ represent the transition probabilities for individual i at time j for $\mathcal{S} \to \mathcal{I}$, $\mathcal{S} \to \mathcal{R}$, and $\mathcal{I} \to \mathcal{R}$, respectively.

2.2 Canine Vertical Transmission Data

2.2.1 Data Description

Data were collected from 2005-2018 by the CDC and the Petersen Lab at the University of Iowa as part of an ongoing surveillance study from hunting hound populations predominantly located in the Midwestern United States. Although vertical transmission of *L. infantum* has been observed in multiple hunting hound breeds (Rosypal et al., 2005; Gaskin et al., 2002), we restrict our analyses to Foxhounds in this chapter because the infection process can vary by breed (Moreno and Alvar, 2002).

In this data set, we have 70 individual pups. For each pup, we have individual level data including year of birth, and results of at least one of three diagnostic tests for at least two time points: immunofluorescence anti-*Leishmania* antibody test (IFAT), quantitative polymerase-chain reaction (qPCR), and Dual-Path Platform (DPP®) canine visceral leishmaniasis (CVL). We also have at least one of these three test results for the mother of each pup during the year in which the pup was born. All dogs included in this data set were born to mothers that became positive on at least one test at some point in their lives, so all pups were exposed to *L. infantum* while *in utero*.

2.2.2 Infection State Classification

For each pup, both mother's status during pregnancy and pup's status at each time point are classified as either $\mathcal{S}$ (susceptible) or $\mathcal{I}$ (infected) based on the results

of up to three tests: IFAT, qPCR, and DPP® CVL, all of which have been compared to each other for consistency in a previous study (Larson et al., 2017). Positive test results are defined as follows. For IFAT, dilutions at or above 1/64 are considered positive. The qPCR tests gives a positive/negative result. For DPP® CVL, if the test gave a positive result in a 15 minute time frame, it was considered positive. We classify a dog into $\mathcal{S}$ if it has all negative test results. Note that while we use the term "susceptible" to be consistent with infectious disease modeling terminology, dogs in this category are exposed, but have parasite levels below the limits of detection for the administered tests. If at least one test is positive, we classify a dog into $\mathcal{I}$, as these tests correlate well with infection, as determined by parasite culture and physical exams (Larson et al., 2017). We refer to this as "diagnostically positive". Dogs are classified as removed ($\mathcal{R}$) when they die or are otherwise removed from the population. Pups' mothers are classified using the same criteria based on their test results during the year of pregnancy. Note that we are implicitly assuming that the sensitivity and specificity for each of these tests are both 1, which is not true in practice, so this is a simplifying assumption. Of the 70 dogs, 22 were born to $\mathcal{S}$ mothers and 48 were born to $\mathcal{I}$ mothers. Dogs are assumed to progress as shown in Figure 2.1.

2.3 Model Specification

2.3.1 Likelihood

In this model, we employ a multinomial likelihood for our SIR model (Ozanne et al., 2019). Let Z_{ij}^{c} denote the infection state of the i^{th} dog at time j, where $i = 1, ..., N$, $j = 1, ..., T$, and $c \in \{\mathcal{S}, \mathcal{I}, \mathcal{R}\}$. The time index, j, represents the j^{th} year since surveillance began. The likelihood has the following structure:

$$\left(Z_{i,j+1}^{\mathcal{I}}, Z_{i,j+1}^{\mathcal{R}}, Z_{i,j+1}^{\mathcal{S}}\right) | Z_{ij}^{\mathcal{S}} \sim \text{Multinomial}\left(1, \boldsymbol{\phi}_{ij}^{(\mathcal{S}\cdot)}\right) \tag{2.1}$$

$$Z_{i,j+1}^{\mathcal{R}} | Z_{ij}^{\mathcal{I}} \sim \text{Bernoulli}\left(\boldsymbol{\phi}_{ij}^{(\mathcal{I}\mathcal{R})}\right) \tag{2.2}$$

where

$$\phi_{ij}^{(\mathcal{S}\mathcal{I})} = \exp\left\{\mathbf{x}_{ij}\boldsymbol{\beta}\right\}\left(1 + \exp\left\{\mathbf{x}_{ij}\boldsymbol{\beta}\right\} + \exp\left\{\mathbf{w}_{ij}\boldsymbol{\theta}\right\}\right)^{-1}, \tag{2.3}$$

$$\phi_{ij}^{(\mathcal{S}\mathcal{R})} = \exp\left\{\mathbf{w}_{ij}\boldsymbol{\theta}\right\}\left(1 + \exp\left\{\mathbf{x}_{ij}\boldsymbol{\beta}\right\} + \exp\left\{\mathbf{w}_{ij}\boldsymbol{\theta}\right\}\right)^{-1}, \tag{2.4}$$

$$\phi_{ij}^{(\mathcal{I}\mathcal{R})} = \left(1 + \exp\left\{\mathbf{z}_{ij}\boldsymbol{\xi}\right\}\right)^{-1}, \tag{2.5}$$

and

$$\boldsymbol{\phi}_{ij}^{(\mathcal{S}\cdot)} = \left(\phi_{ij}^{(\mathcal{S}\mathcal{I})}, \phi_{ij}^{(\mathcal{S}\mathcal{R})}, 1 - \phi_{ij}^{(\mathcal{S}\mathcal{I})} - \phi_{ij}^{(\mathcal{S}\mathcal{R})}\right).$$

The design matrices in Equations 2.3-2.5 should be defined to suit the problem of interest. For this application, they are specified in the following way:

$$\mathbf{x}_{ij} = \left[1, \mathbb{1}_{\{\text{Mother} \in \mathcal{S}\}_i}, Age_{ij} \cdot \mathbb{1}_{\{Age_{ij} > 2\}}, \left(\mathbb{1}_{\{\text{Mother} \in \mathcal{S}\}_i} * Age_{ij}\right) \cdot \mathbb{1}_{\{Age_{ij} > 2\}}\right],$$

$$\mathbf{w}_{ij} = \left[1, \mathbb{1}_{\{\text{Mother} \in \mathcal{S}\}_i}, Age_{ij} \cdot \mathbb{1}_{\{Age_{ij} > 5\}}, \left(\mathbb{1}_{\{\text{Mother} \in \mathcal{S}\}_i} * Age_{ij}\right) \cdot \mathbb{1}_{\{Age_{ij} > 5\}}\right],$$

$$\mathbf{z}_{ij} = \left[1, Age_{ij}\right].$$

Since dogs usually have litter mates, one might contend that the transition probabilities in Equations 2.1-2.5 for this application could incorporate a random litter term to capture this correlation. However, in this chapter, we are modeling infection progression in purebred Foxhounds that are members of only a few hunts, and there is considerable breeding among these hunts, so this is a fairly homogeneous population. Conditional on mother's status during pregnancy and pup age, the individuals can be considered independent.

2.3.2 Prior

We placed normal priors on $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and $\boldsymbol{\xi}$ that are centered at 0 and have a variance of 1. While this might not appear to be a vague prior on the surface, in this context it is vague. Small changes in these parameters correspond to big changes in the corresponding transition probabilities, so this choice of prior is general in this context.

2.4 Basic Reproductive Numbers for Vertical Transmission

When infection is transmitted transplacentally, an infectious female dog can only pass infection to her pups, so the number of potentially infectious contacts is equal to the number of pups in a litter. To calculate a reproductive number for vertical transmission from infected mothers, for instance, we determine the expected number of pups (born to infected mothers) that become infected at some point in their lives.

We simulate realizations of vertically-exposed pups to estimate reproductive

numbers in the following way. We sample with replacement 1000 times the posterior parameter estimates from fitting the vertical transmission SIR model to the vertical transmission data. For each $\ell \in \{1, 2, \ldots, 1000\}$, we consider $N^* = 70$ pups, starting at age 0, and distribute these pups between $\mathcal{S}$ mothers and $\mathcal{I}$ mothers according to a Binomial$(N^*, 1/2)$ distribution. We assume there are an equal number of $\mathcal{S}$ and $\mathcal{I}$ mothers (7 each). Using the posterior parameter estimates $(\beta, \theta, \xi)^\ell$, we propagate the 70 pups forward and count the number of dogs that ever enter the $\mathcal{I}$ category. Formally, let $Z_{ij}^{\mathcal{I}|\mathcal{S}}$ be 1 if individual i becomes newly seropositive ($\mathcal{I}$) at time j and 0 otherwise. Also, let $Z_i^M(m)$ be 1 if individual i's mother was in infection category m at the time of i's birth and 0 otherwise. Then, the expression in Equation 2.4 gives an empirical estimate of the reproductive number for mothers in $m \in \{S^*, I^*\}$:

$$\mathcal{R}_0^m = \frac{\sum_{j=1}^{J} \sum_{i \in N_m^*} Z_{ij}^{\mathcal{I}|\mathcal{S}}}{M(m)}. \tag{2.6}$$

These reproductive numbers should be interpreted in the usual way. If $\mathcal{R}_0^m$ is greater than one, then this indicates that infection should persist in the population as a result of infection type m, assuming that female dogs in this infectious category are still being bred.

2.5 Real Data Analysis

We implemented the individual-level SIR model (Section 2.3) with the data described in Section 2.2. The posterior medians and 95% credible intervals for the

parameters in the vertical transmission SIR model are shown in Table 2.1. The estimated intercepts are negative and relatively large, which corresponds to a small probability of transitioning out of the $\mathcal{S}$ category at very young ages. This agrees with observed transition behavior. The 95% credible interval for the coefficient for the indicator for mother's status (which equals 1 if the mother is diagnostically negative during pregnancy and 0 otherwise) is right-skewed; 85.5% of the sampled parameter estimates are less than zero. This suggests there is some protective effect associated with being born to a diagnostically negative mother. This corresponds to a greater risk for an $\mathcal{S} \rightarrow \mathcal{I}$ transition for pups born to $\mathcal{I}$ mothers than for those born to $\mathcal{S}$ mothers before age 3, as shown in Figure 2.2 (center panel). This relationship also can be observed in the coefficient estimate and credible interval for β_3, the interaction between pup age and mother's status (Table 2.1).

Table 2.1: Model Fit Summary - Canine Vertical Transmission Model

	Variable	**Median**	**95% Credible Interval**
β_0	Intercept	-2.020	(-2.473, -1.609)
β_1	Mother in $\mathcal{S}$	-0.438	(-1.311, 0.365)
β_2	(Pup Age)($\mathbb{1}_{\{Age>2\}}$)	0.244	(0.052, 0.426)
β_3	(Mother in $\mathcal{S}$*Pup Age)($\mathbb{1}_{\{Age>2\}}$)	0.230	(-0.067, 0.546)
θ_0	Intercept	-2.219	(-2.714, -1.780)
θ_1	Mother in $\mathcal{S}$	0.406	(-0.342, 1.122)
θ_2	(Pup Age)($\mathbb{1}_{\{Age>5\}}$)	0.032	(-0.322, 0.290)
θ_3	(Mother in $\mathcal{S}$*Pup Age)($\mathbb{1}_{\{Age>5\}}$)	-0.110	(-0.636, 0.374)
ξ_0	Intercept	-1.396	(-2.929, 0.079)
ξ_1	Pup Age	-0.449	(-1.031, -0.007)

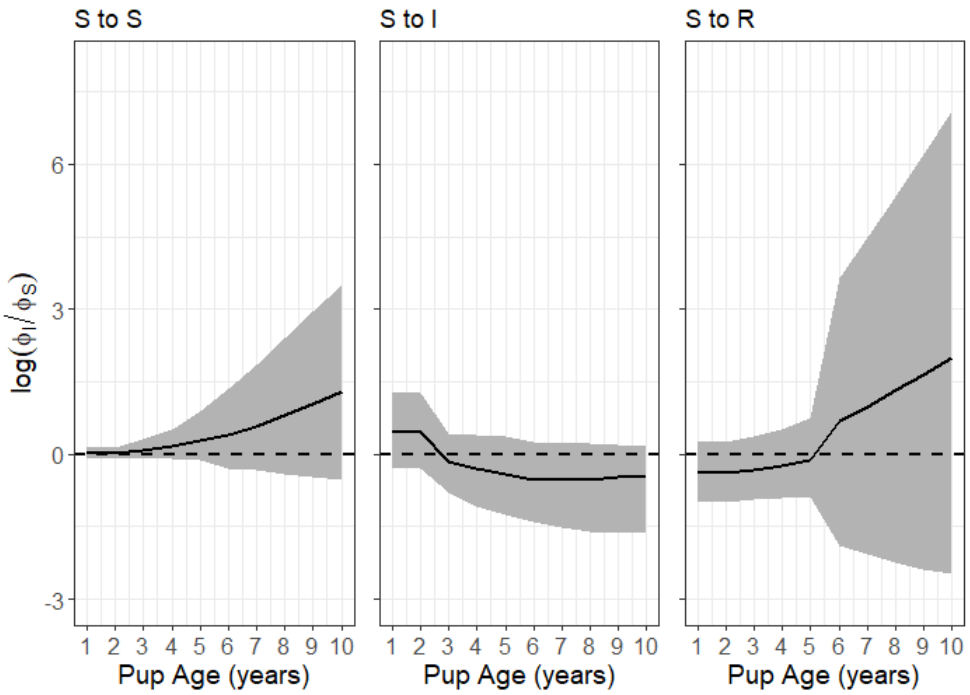

Figure 2.2: The median log relative risk (and 95% credible intervals) of staying in $\mathcal{S}$ (left), transitioning from $\mathcal{S} \to \mathcal{I}$ (center), and from $\mathcal{S} \to \mathcal{R}$ (right) for dogs born to mothers in $\mathcal{I}$ versus dogs born to mothers in $\mathcal{S}$.

In Figure 2.2, we examine the log relative risk of transitions from $\mathcal{S}$ (to $\mathcal{S}$, $\mathcal{I}$, and $\mathcal{R}$) as a function of pup age. There is considerable variability in this process; for all transitions, the 95% credible bands for the log relative risk cover zero. While pups born to $\mathcal{I}$ mothers are more likely to transition from $\mathcal{S}$ to $\mathcal{I}$ before age 3, they are less likely to transition after age 3. This suggests pups born to $\mathcal{I}$ mothers are likely to transition to $\mathcal{I}$ early in life. If we examine the data, these same dogs also die early. In contrast, those pups born to $\mathcal{I}$ mothers that do not transition before age 3 are likely healthier because they live longer and transition later (Figure 2.3, right).

Some of the results appear to be driven by a survivor bias.

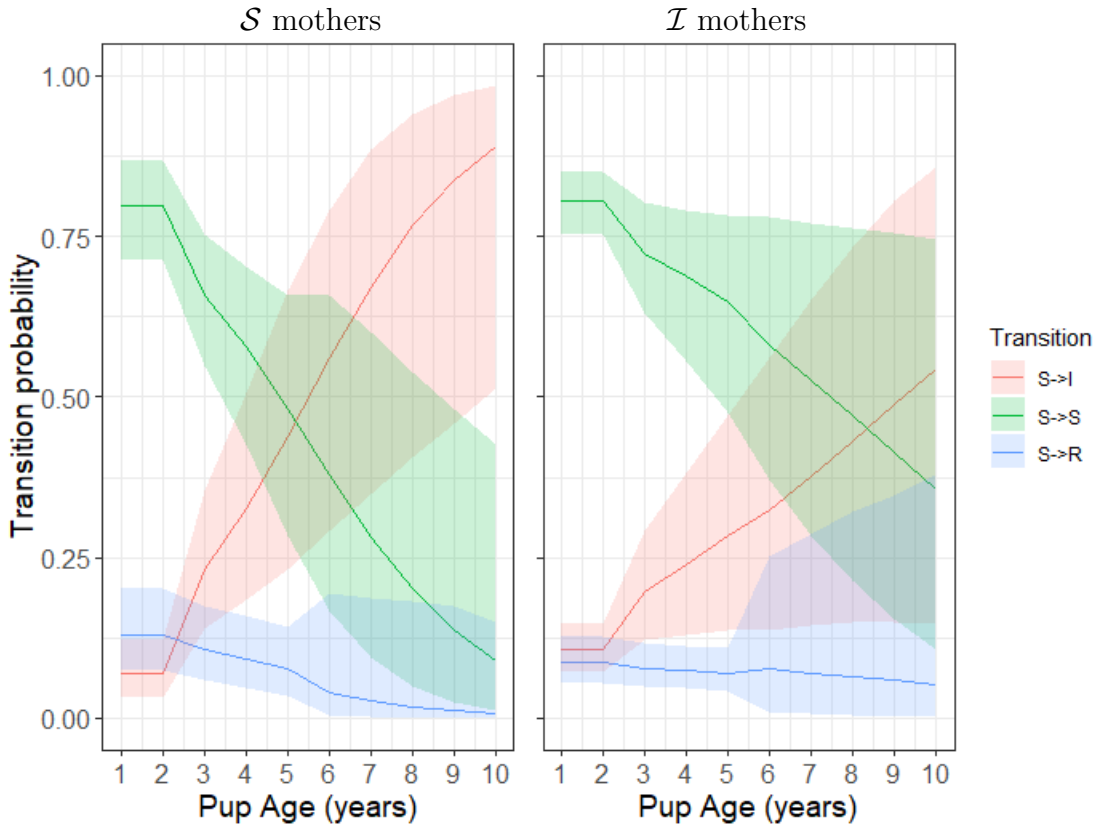

Figure 2.3: Transition probabilities from $\mathcal{S}$ for pups born to $\mathcal{S}$ mothers (left) and pups born to $\mathcal{I}$ mothers (right). Pups in the latter group are more likely to transition from $\mathcal{S}$ to $\mathcal{I}$ before the age of 3. After age 3, pups born to $\mathcal{S}$ mothers are more likely to transition to $\mathcal{I}$.

Reproductive numbers for vertical transmission are calculated for pups born to $\mathcal{S}$ mothers and pups born to $\mathcal{I}$ mothers as described in Section 2.4. The median reproductive number for the former group is 2.83, with a 95% credible interval of

(2.20, 3.47). The median reproductive number for the latter group is 3.37, with a 95% credible interval of (2.54, 4.04). The box plot in Figure 2.4 summarizes results for the reproductive numbers for the two groups. On average, $\mathcal{R}_0^{\mathcal{I}} > \mathcal{R}_0^{\mathcal{S}}$ for 86.3% of the posterior predictive data sets. This suggests that dogs born to $\mathcal{I}$ mothers are more likely to progress to infection at some point in their lives than those born to $\mathcal{S}$ mothers. Given the size of the available data set, this percentage is compelling.

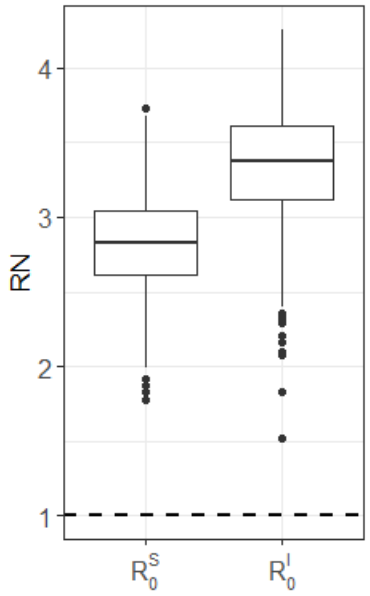

Figure 2.4: Box plots for reproductive numbers (RN) for vertical transmission for data. $\mathcal{R}_0^{\mathcal{S}}$ and $\mathcal{R}_0^{\mathcal{I}}$ correspond to the reproductive numbers for $\mathcal{S}$ moms and $\mathcal{I}$ moms, respectively. Either degree of maternal infection during pregnancy can increase the number of infected dogs from generation to generation. The 95% credible interval for $\mathcal{R}_0^{\mathcal{I}}/\mathcal{R}_0^{\mathcal{S}}$ is (0.878, 1.603). $\mathcal{R}_0^{\mathcal{I}} > \mathcal{R}_0^{\mathcal{S}}$ for 86.3% of the posterior predictive data sets.

2.6 Discussion

In this chapter, we developed an individual-level Bayesian SIR model for vertical transmission to address two hypotheses: pups born to diagnostically positive ($\mathcal{I}$) mothers at the time of pregnancy (1) are more likely to progress to infection at some point in their lives, and (2) progress to infection more quickly than those born to diagnostically negative ($\mathcal{S}$) mothers.

To address the first hypothesis, we calculated reproductive numbers for vertical transmission, $\mathcal{R}_0^{\mathcal{S}}$ and $\mathcal{R}_0^{\mathcal{I}}$ empirically and compared the distributions of these quantities for the two groups (Figure 2.4). We found that $\mathcal{R}_0^{\mathcal{I}} > \mathcal{R}_0^{\mathcal{S}}$ for 86.3% of the posterior predictive distributions. Given the small sample size, this result is compelling evidence in favor of (1). This suggests that with additional data, we will observe a clearer distinction between the two groups. Surveillance efforts are ongoing.

To address the second hypothesis, we examined the log RR for transitions out of the $\mathcal{S}$ category as a function of age. For young dogs (< 3 years), the median risk for the $\mathcal{S} \to \mathcal{I}$ transition is smaller for dogs born to diagnostically negative mothers. We do not, however, observe a significant difference in the RR for the $\mathcal{S} \to \mathcal{I}$ transition, but this is likely due to smaller sample size. In addition, some pups born to $\mathcal{I}$ mothers in the data set became diagnostically positive in their first year of life and then died in that same year. These dogs progressed to infection quickly. There were no pups born to $\mathcal{S}$ mothers in the data set that experienced such rapid progression. In contrast, other dogs born to $\mathcal{I}$ mothers lived longer and their immune systems apparently handled their parasite exposure more successfully than

those of their rapidly progressing counterparts.

CHAPTER 3
HUMAN ZOONOTIC VISCERAL LEISHMANIASIS TRANSMISSION IN BRAZIL

3.1 Vector Transmission of Visceral Leishmaniasis in Humans

While VL is a serious health problem for domesticated dogs, its impact on human health in endemic areas makes it a major public health concern. To better understand the public health impact, it is important to know the proportions of individuals in a particular population that fall into different states of *L. infantum* infection because these proportions can provide insights into the likelihood of expansion versus control of the infection. However, these population level infection state proportions are often difficult to observe, making estimation through statistical models a valuable tool.

One commonly applied family of statistical techniques for modeling infection outcomes in large populations is logistic regression. Binomial logistic regression has been used to model infectious diseases in a variety of contexts. For example, it has been used to differentiate infectious diseases based on clinical and laboratory characteristics to facilitate diagnosis. Chadwick et al. (2006) apply it to distinguish dengue fever from other infections in Singapore, and Bonsu and Harper (2004) apply it to differentiate between bacterial and viral meningitis. This method also has been used to identify the relative importance of prognostic factors in particular disease outcomes, like death (Werneck et al., 2003). Moreover, multinomial logistic regression, which is an extension of binomial logistic regression, has been utilized in dengue

hemorrhagic fever (DHF) research to determine if DHF and signs of clinical disease are associated with each of five serotypes (Fried et al., 2010). It also has been used to assess the association between demographic variables and *L. infantum* infection status in neighborhoods near to Natal, Brazil (Lima et al., 2012).

In addition to examining covariates that may be associated with a particular infection outcome, logistic regression techniques can be used to estimate the proportion of individuals that fall into each infection state over time. While these approaches are simple, powerful, and straightforward to implement, they have a number of practical limitations. For example, regression-type models can incorporate some temporal variability, but they do not naturally accommodate known infection dynamics like well-studied seasonality patterns, heterogeneity in infectivity, and infectious period duration (Franke et al., 2002; Morrison et al., 1995). While temporal covariates can be included, large quantities of data may be required to accurately estimate these effects.

Stochastic compartmental models comprise another class of models that can be used to estimate infection state proportions over time. This kind of model allows for greater flexibility in approximating infection prevalence because it easily accommodates known infection dynamics, and it improves the scientific interpretability of model results. This approach allows for inclusion of prior information of seasonal trends and the pattern of progression through different states of infection.

In this chapter, we will explore the utility of a multinomial logistic regression model and a Bayesian SIR model for estimating infection prevalence in humans over

time. The SIR model structure is illustrated in Figure 3.1, where individuals in the same infection state are assumed to transition to another state with the same probability. Possible transitions are $\mathcal{S} \to \mathcal{I}$ and $\mathcal{I} \to \mathcal{R}$. We explore both approaches to prevalence estimation through simulation studies and in the context of *L. infantum* infection in Brazil. We simplify the process by considering only humans and do not explicitely include dogs in the model specification.

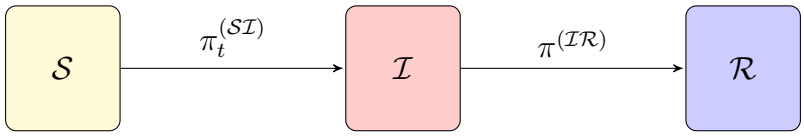

Figure 3.1: Schematic for the population level vector transmission SIR model. The parameters $\pi_t^{(\mathcal{SI})}$ and $\pi^{(\mathcal{IR})}$ refer to the probabilities of transitioning from $\mathcal{S} \to \mathcal{I}$ and $\mathcal{I} \to \mathcal{R}$, respectively.

3.2 Human Transmission Data

3.2.1 Data Description

This analysis is based on historical data; the original study was published by Lima et al. in 2012. During the months of January-February and June-July 2006, a study was conducted in Parnamirim, a city of 180,000 located next to Natal, to determine the covariates that were significantly associated with disease state. Households were chosen according to a random point process; 345 individuals were included in the study. Skin and serology tests were administered once to each study partici-

pant, and these test results were used to group people into infection status categories. These tests included an ELISA SLA serology test ($SLA^{+/-}$) and the leishmanial skin test ($LST^{+/-}$), which was administered intradermally and results were read 48-72 hours later. Of the 345 people, 314 had complete test result data. Missing data were usually due to inability to find the individual and read the LST results within 48-72 hours after test placement, or less often, due to insufficient amounts of serum to run serologic tests. No study participants were positive for acute or past VL, so this analysis focused on asymptomatic infection. In addition to the LST and serology test results, demographic information was also collected including age (dichotomized to ≤ 35 years and > 35 years), sex, and household location (urban, peri-urban, or rural) (Lima et al., 2012).

3.2.2 Infection State Classification

Individuals with LST^-, SLA^-, and VL^- test results are classified as susceptible ($\mathcal{S}$), as described in the introduction. Those with LST^-, SLA^+, VL^- results are assumed to have recently acquired infection and is classified as infectious ($\mathcal{I}$); this is the acute seropositive class described in the introduction. These individuals show no signs of infection, but are presumed to be able to transmit infection to the sand fly vector. Individuals with LST^+, $SLA^{+/-}$, VL^- are classified as removed ($\mathcal{R}$); this is the infected asymptomatic class previously described. Most individuals in this class develop a protective immune response that protects them from reinfection or from symptomatic disease; they develop a protective immune response and remain

LST^+ for life (Braz et al., 2014). Let $\mathcal{T} = \{1, ..., 28\}$ be the set of time points (weeks in a year) over which the endemic infection process is studied, and let $\mathcal{T}^* = \{4, 5, 6, 7, 8, 25, 26, 27, 28\}$ be the weeks at which an individual's blood was drawn, which we will assume accurately describes the time at which the individual occupies a particular infection state.

3.3 Multinomial Logistic Regression Model

3.3.1 Previous Analysis - Multinomial Regression Results

Lima et al. (2012) collected and analyzed the data. Their multinomial logistic regression model had four categories, but individuals who are LST^+ and VL^- can be classified in the removed (infected asymptomatic) category regardless of SLA status, so the previous analyses were rerun with three outcome categories (Jeronimo et al., 2004). Let π_{ijt} be the probability that the i^{th} individual belongs to the j^{th} category at time t, where $i \in \{1, 2, ..., 314\}$, $j \in \{\mathcal{S}, \mathcal{I}\}$ and $t \in \mathcal{T}^*$. The link function is

$$\eta_{ijt} = \log\left(\frac{\pi_{ijt}}{\pi_{i\mathcal{R}t}}\right) = \alpha_j + \mathbf{x}'_{it}\boldsymbol{\beta}_j, \tag{3.1}$$

where $\boldsymbol{\beta}_j$ includes coefficients for age, sex, household location, and week for the j^{th} disease category.

The multinomial logistic regression coefficient estimates in Table 3.1 were obtained using the *multinom* function from the `nnet` package in R (Venables and Ripley, 2002). The baseline category was the removed category.

The coefficient estimates and corresponding standard errors obtained are shown

in Table 3.1. Lima et al. (2012) did not include the month at which the tests were administered as a covariate. We include an expanded multinomial logistic regression model to accommodate the known seasonality of the infection and to facilitate comparison with estimates obtained using a Bayesian approach, which are shown in Table 3.4.

Table 3.1: Model Fit Summary - Frequentist Multinomial Logistic Regression

	$\mathcal{S}/\mathcal{R}$		$\mathcal{I}/\mathcal{R}$	
	Estimate	SE	Estimate	SE
Intercept	-0.100	0.317	-0.113	0.388
Area: Urban	0.092	0.335	0.295	0.373
Area: Periurban	0.404	0.344	-0.372	0.478
Sex: Male	-0.515	0.276	-0.840	0.370
Age $>$ 35	-0.811	0.266	-0.223	0.333
Week	0.045	0.015	-0.027	0.023

3.3.2 Bayesian Multinomial Regression Model

The Bayesian multinomial logistic regression model has the same form as detailed in the previous subsection, where π_{ij} is the probability that the i^{th} individual belongs to the j^{th} category, where $i \in \{1, 2, ..., 314\}$, $j \in \{\mathcal{S}, \mathcal{I}\}$, and the link function is the same as in Equation 1. Four covariates were included: age, sex, household location, and week (in the year) at which individuals' LST and serologic testing was done. Normal vague priors were put on the $\alpha_{\mathcal{R}}$ and $\boldsymbol{\beta}_j$ parameters; the mean was 0 in each case, and the variance was 100. The model was fit using the *MCMCmnl*

function from the `MCMCpack` package in R (Martin et al., 2011).

3.4 Bayesian SIR Human Transmission Model

A Bayesian approach where the latent infection states are estimated using a compartmental model is an alternative to multinomial logistic regression. Such an approach allows for the inclusion of prior information about seasonal trends and the pattern of progression through different states of *L. infantum* infection.

3.4.1 Data Model

Let $\mathbf{Y}_j = (S_j^*, I_j^*, R_j^*)^T$ be a vector of counts of individuals in the population that are in $\mathcal{S}$, $\mathcal{I}$, and $\mathcal{R}$ at time j, where $N = 180,000$, the population of Parnamirim. For $j \in \mathcal{T}^* \subset \mathcal{T}$,

$$\mathbf{Y}_j \sim \text{Multinomial}\left(m_j, \boldsymbol{\pi}_j = \left(\frac{S_j}{N}, \frac{I_j}{N}, \frac{R_j}{N}\right)\right), \tag{3.2}$$

where m_j is the number of individuals observed at time j, and S_j, I_j, and R_j are the estimated counts of individuals in $\mathcal{S}$, $\mathcal{I}$, and $\mathcal{R}$, at time j, respectively. These values are unobserved - they are estimated as part of the process model.

Individual-level characteristics such as age, sex, and household location play a role in whether an individual becomes infected (Lima et al., 2012). The composition of the population in an area of interest can influence the proportion of the population that are in $\mathcal{I}$ at a particular time. The data model in Equation 3.2 can be amended to reflect the effect of the individual-level covariates.

We assume that the process model, descibed in the following section, reflects the infection process for some homogeneous, baseline population. Let $\mathbf{Z}_j$ be the design

matrix for individual-level effects at time j and let $\boldsymbol{\gamma}$ be the vector of corresponding coefficients. Then, consider the data model in Equation 3.3,

$$\mathbf{Y}_j \sim \text{Multinomial}\left(m_j, \boldsymbol{\pi}_j = \left(\frac{S_j}{N}, \frac{I_j^{\text{aug}}}{N}, \frac{R_j}{N}\right)\right), \tag{3.3}$$

where

$$I_j^{\text{aug}} = I_j + \frac{1}{m_j}\mathbf{Z}_j'\boldsymbol{\gamma}.$$

We assume that $\boldsymbol{\gamma} \sim MVN\left(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma\right)$, so $\mathbf{Z}_j'\boldsymbol{\gamma}/m_j$, the average contribution of the individual covariates at time j, can either increase or decrease I_j, depending on the composition of the sample.

3.4.2 Process Model

Let S_t be the number of individuals in $\mathcal{S}$ at time t, I_t be the number of individuals in $\mathcal{I}$ at time t, and R_t be the number of individuals in $\mathcal{R}$ at time t. The compartment sizes at each time $t \in \mathcal{T}$ are deterministic functions of the compartment sizes at the previous time point and the transition compartments (the number of individuals that transition) at the current time point. They are:

$$\begin{aligned} S_t &= S_{t-1} - I_t^{(\mathcal{SI})}, \\ I_t &= I_{t-1} + I_t^{(\mathcal{SI})} - R_t^{(\mathcal{IR})}, \\ R_t &= R_{t-1} + R_t^{(\mathcal{IR})}. \end{aligned}$$

There are two transition compartments, $I_t^{(\mathcal{SI})}$ and $R_t^{(\mathcal{IR})}$. The former denotes the number of individuals that are transitioning from $\mathcal{S}$ to $\mathcal{I}$ at time t. The latter

denotes the number of individuals that are transitioning from $\mathcal{I}$ to $\mathcal{R}$ at time t. These compartments are modeled using a chain binomial structure (Equations 3.4 and 3.5).

$$I_t^{(\mathcal{SI})} \sim \text{Binomial}\left(S_{t-1}, \pi_t^{(\mathcal{SI})}\right), \tag{3.4}$$

$$R_t^{(\mathcal{IR})} \sim \text{Binomial}\left(I_{t-1}, \pi^{(\mathcal{IR})}\right). \tag{3.5}$$

Note that for all endemic sites, we have some initial compartment sizes, S_0, I_0, and R_0. At each iteration, these initial values are allowed to vary. Survey data indicates that the expected proportion of susceptibles is 51%, the expected proportion of infected is 1.9%, and the expected proportion of removed is 35%, as defined above, in the region of interest. Another 12% are estimated to fall into categories defined by positive VL test results, which are not present in this data set (Jeronimo et al., 2004). Therefore, to obtain $\pi^{(S_0)}$, $\pi^{(I_0)}$, and $\pi^{(R_0)}$, the three probabilities are normalized to sum to 1. The initial compartments have the following distributions:

$$S_0 \sim \text{Binomial}\left(N, \pi^{(S_0)}\right), \tag{3.6}$$

$$I_0 \sim \text{Binomial}\left(N, \pi^{(I_0)}\right), \tag{3.7}$$

$$R_0 = N - (S_0 + I_0). \tag{3.8}$$

3.4.3 Parameter Model

We expect $\pi_t^{(\mathcal{SI})}$, the transition probability from $\mathcal{S}$ to $\mathcal{I}$, to be influenced by seasonality, interventions, and other environmental factors. In keeping with the work of Lekone and Finkenstädt and Brown et al., we assume a homogeneous Poisson contact process; the number of bites at time t, $K_t \sim \text{Poisson}(\lambda_t)$ (Lekone and Finkenstädt,

2006; Brown et al., 2016). Let $Inf(t)$ be the event that a human becomes infected at time t and let $\overline{Inf(t)}$ be the complement. Then, the probability of a human being infected at time t is

$$P\left(Inf(t)\right) = 1 - P\left(\overline{Inf(t)}\right),$$

where

$$\begin{aligned}
P(\overline{Inf(t)}) &= \mathbb{E}\left[\overline{Inf(t)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\overline{Inf(t)}|K_t = k_t\right)\right] \\
&= \mathbb{E}\left[\left(1 - \left(\frac{I_{t-1}}{N}\right)p^h\right)^{k_t}\right] \\
&= \sum_{k=0}^{\infty}\left(1 - \left(\frac{I_{t-1}}{N}\right)p^h\right)^{k_t}\frac{e^{-\lambda_t}\lambda_t^{k_t}}{k_t!} \\
&= \exp\left\{-\left(\frac{I_{t-1}}{N}\right)(\lambda_t p^h)\right\}.
\end{aligned}$$

Note that the probability that a human becomes infected as a result of a bite, p^h, and the bite rate, λ_t, are usually unidentifiable, so $\lambda_t p^h$ is replaced by $e^{X_t\boldsymbol{\beta}}$ and X_t contains the information about seasonality or any relevant interventions at time t (Brown et al., 2016; Lekone and Finkenstädt, 2006). Then, the probability of a human being infected at time t is

$$\pi_t^{(\mathcal{SI})} = 1 - \exp\left\{-\frac{I_{t-1}}{N}e^{X_t'\boldsymbol{\beta}}\right\}. \tag{3.9}$$

The prior on $\pi^{(\mathcal{IR})}$, the transition probability from $\mathcal{I}$ to $\mathcal{R}$, is assumed to be

$$\pi^{(\mathcal{IR})} \sim \text{Beta}\left(\alpha^{(\mathcal{IR})}, \beta^{(\mathcal{IR})}\right). \tag{3.10}$$

There is some information available on exposure time for VL that can be used to define an informative prior for $\pi^{(\mathcal{IR})}$. Levy and Yiengst (1948) claim that individuals with VL progress to symptomatic disease within 10 days to 18 months after exposure. Because it is difficult to know the time of exposure in populations living in endemic regions, this estimate is not substantiated by objective observational studies, but rather is derived from other descriptions of disease characteristics. Accidental laboratory infections with species causing human cutaneous leishmaniasis give information on the period between a single known exposure and disease manifestations. Even though this exposure route is not natural and the species of *Leishmania* differs from *L. infantum*, this may be the best source to discern known exposure and infection status. These individuals have a median exposure time of 8-12 weeks (Herwaldt, 2001). We assume that an immune response decision, i.e. whether an exposed individual will progress to asymptomatic or symptomatic infection, is made 2-4 weeks prior to immune response detection. Therefore, we expect that an individual may transition from $\mathcal{I}$ to $\mathcal{R}$ within 4-6 weeks after infection. We assume that this 4-6 week time period can be described by two geometric distributions, one with a median (time to transition) of 4 and one with a median (time to transition) of 6. These correspond to a Geometric(0.15) and a Geometric(0.10) distribution, respectively. Then, we want to use these two success probabilties, 0.10 and 0.15, to determine the prior parameters, $\alpha^{(\mathcal{IR})}$ and $\beta^{(\mathcal{IR})}$ (Equation 3.4.3). To do this, we consider the two equations, $P(\pi^{(\mathcal{IR})} < 0.05) = 0.10$ and $P(\pi^{(\mathcal{IR})} > 0.95) = 0.15$, and solve to find the $\alpha^{(\mathcal{IR})}$ and $\beta^{(\mathcal{IR})}$ values that satisfy them. The prior distribution can be made more diffuse by looking at a larger interval

to establish $\alpha^{(\mathcal{IR})}$ and $\beta^{(\mathcal{IR})}$. In order to get meaningful results for the compartmental model, however, it is important to establish biologically plausible priors, so the prior still needs to be centered in the same place.

3.5 Empirically Adjusted Reproductive Number

The basic reproductive number, $\mathcal{R}_0$, quantifies the expected number of secondary infections that result from a single infectious individual under the assumption of a totally susceptible population. As previously noted, in many applications, the assumption of a totally susceptible population is not appropriate for our application. Many authors have provided adaptations to this basic reproduction number; in this chapter, we employ the empirically adjusted reproductive number approach proposed by Brown et al. (2016).

Brown et al. (2016) propose a general empirically adjusted reproductive number for spatio-temporal data. In our application, we have several simplifications. Let $I_k(t)$ be the event that individual k becomes infected at time t. Then, the expected number of infections is given in Equation 3.11,

$$\mathbb{E}\left[\sum_{k=0}^{N} I_k(t)\right] = S_t \cdot P\left(I_k(t) | k \in \mathcal{S}\right), \tag{3.11}$$

and the average number of infections per infectious individual is obtained by dividing the expectation from Equation 3.11 by the number of infectious individuals at time t: $(S_t/I_t) \cdot P\left(I_k(t) | k \in \mathcal{S}\right)$. At each time, this average is a single time, single location analogue of the 'next generation matrix', which is given in Equation 3.12:

$$G(t) = \frac{S_t \cdot P\left(I_k(t) | k \in \mathcal{S}\right)}{I_t} = \left(\frac{S_t}{I_t}\right) \pi_t^{(\mathcal{SI})}. \tag{3.12}$$

Then, Equation 3.13 gives the empirically adjusted reproductive number at time t:

$$\mathcal{R}^{(EA)}(t) = \sum_{w=t}^{t_\infty} G(w) \cdot \left(1 - \pi^{(\mathcal{IR})}\right)^t. \tag{3.13}$$

At each time t, we need to compute an infinite sum. As Brown et al. (2016) note, the pathogen lifespan weighting term, $(1-\pi^{(\mathcal{IR})})^t$, quickly and monotonically approaches 0, so we compute the summation over a finite number of weeks, after which subsequent terms make a negligible contribution to the sum.

3.6 Simulation Studies

3.6.1 Simulation Objective

The compartmental modeling approach proposed here naturally accommodates disease dynamics, while the multinomial logistic regression approach does not naturally accommodate such information. In this section, we will compare the estimation performance of these models for two scenarios, a dynamic epidemic and a non-dynamic epidemic. We will show that the compartmental modeling approach is conducive to either scenario. In contrast, the multinomial logistic regression approach cannot attain the same level of accuracy as the compartmental model in the case of a dynamic epidemic for the same number of observations.

3.6.2 Simulation Setup

Counts for each of the three disease states, $\mathcal{S}$, $\mathcal{I}$, and $\mathcal{R}$ were simulated assuming two models. The first model was a multinomial model where the proportions of individuals in each infection state were assumed to be constant over time. The second

model was a multinomial model where the proportions of individuals in each infection state at each time point were determined by the proportions in the population, as described in Equation 3.2. The design matrix, $\mathbf{X}$, had an intercept and a seasonality component, which was a Gaussian kernel centered at Week 15, with a standard deviation of 3. This design matrix is appropriate for our application, which we want our simulation to mirror, but other design matrices should be considered when infections with different dynamics are being modeled. The two simulated data sets were fit using the multinomial logistic regression model and the compartmental model, as previously described, respectively. The estimated infection state proportions for each of the four data-model combinations were compared to the known proportions.

3.6.3 Simulation Results

Two simulated data sets were generated, as described in the previous section. For each simulated data set, 35 observations were simulated each week for 28 weeks. The sample size was chosen because it is the average number of observations made each week in the *L. infantum* infection data set we study in this paper. The data sets then were fit assuming multinomial logistic regression and assuming the compartmental model structure, resulting in four model fits. Distributions for posterior predictive probabilities were generated for each of the four model-data combinations and compared to the true proportions for each category at each time.

Figure 3.2 shows the resulting posterior predictive probability distributions for each of the two models using the multinomial-generated data with constant propor-

tions across time. The multinomial logistic regression model captures the constant proportions well; distributions are concentrated around the truth. The compartmental model comes close to capturing the truth; the bulk of the distribution is generally at or close to the true proportion for each of the compartments. The parameter estimates and 95% credible intervals for each of the two models in Figure 3.2 are shown in Table 3.2. Note that the effect of week is not significantly different from zero in either model, as we expect given the data generating mechanism.

Figure 3.3 shows the resulting posterior predictive probability distributions for each of the two models using the compartmental model-generated data. The multinomial logistic regression model misses the nuances of the shape of the data, particularly in the $\mathcal{I}$ category, but also in the early weeks for both the $\mathcal{S}$ and $\mathcal{R}$ categories. In contrast, the SIR model captures the simulated data well. The corresponding parameter estimates and 95% credible intervals are shown in Table 3.3. Here we see that time is significant in the multinomial logistic regression model.

Table 3.2: Model Fit Summaries - Multinomial Simulated Data

Multinomial Logistic Regression

	$\mathcal{S}/\mathcal{R}$		$\mathcal{I}/\mathcal{R}$	
	Estimate	95% Credible Interval	Estimate	95% Credible Interval
Intercept	0.447	(0.178, 0.710)	-3.051	(-4.114, -2.148)
Week	0.000	(-0.016, 0.017)	-0.004	(-0.065, 0.055)

Compartmental Model

	Estimate	95% Credible Interval
Intercept	-1.599	(-2.105, -1.104)
Week	-1.076	(-3.390, 1.340)
$\pi^{(\mathcal{IR})}$	0.122	(0.096, 0.152)

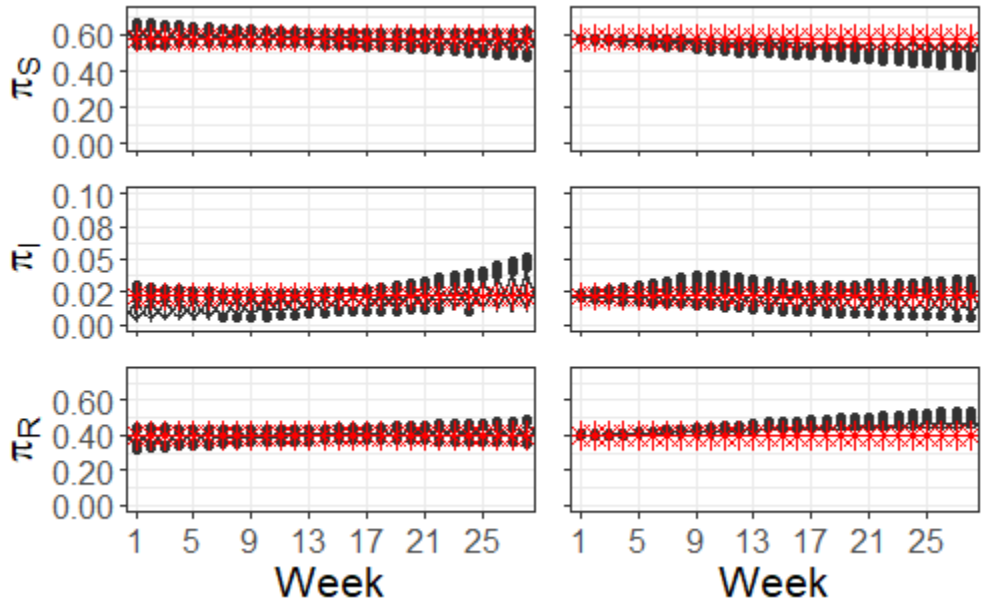

Figure 3.2: Posterior predictive probabilities resulting from fitting multinomial-generated data with constant proportions across time using multinomial logistic regression (left) and the compartmental model (right). Known proportions are shown as red stars. The multinomial logistic regression model results in tight intervals around the truth, whereas the compartmental modeling approach gets close to the truth, but under-predicts and over-predicts in the $\mathcal{S}$ and $\mathcal{R}$ classes, respectively. Distributions were based on 5000 simulations at each week.

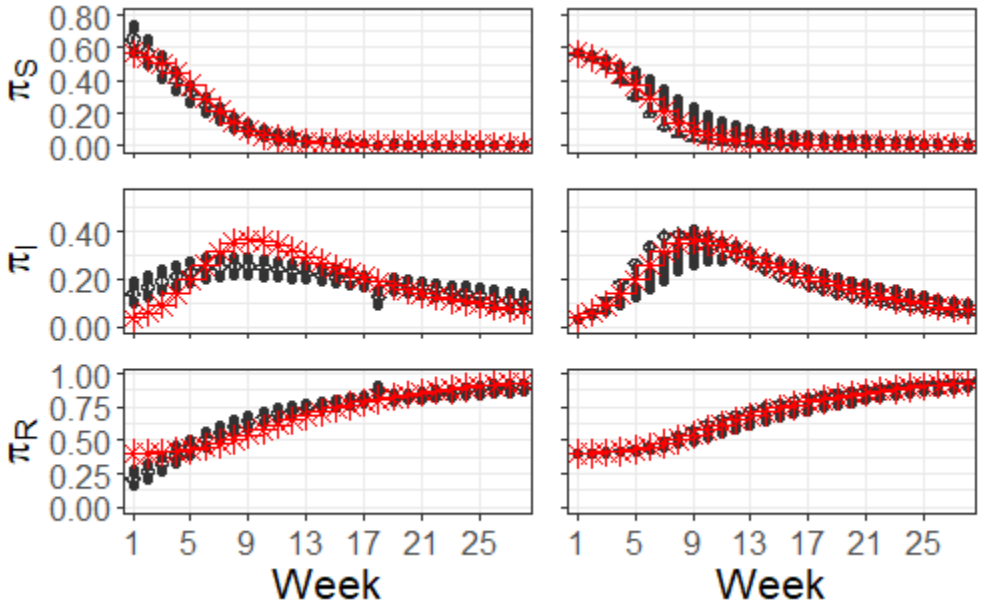

Figure 3.3: Posterior predictive probabilities resulting from fitting multinomial-generated data with a compartmental model structure on the proportions across time using multinomial logistic regression (left) and the compartmental model (right). Known proportions are shown as red stars. The multinomial logistic regression model missed the initial compartment sizes for $\mathcal{I}$ and $\mathcal{R}$ and the infected count peak. In contrast, the compartmental model captures the changes over time well for all compartments.

Table 3.3: Model Fit Summaries - SIR Simulated Data

Multinomial Logistic Regression

	$\mathcal{S}/\mathcal{R}$		$\mathcal{I}/\mathcal{R}$	
	Estimate	95% Credible Interval	Estimate	95% Credible Interval
Intercept	1.459	(1.040, 1.889)	-0.389	(-0.734, -0.038)
Week	-0.353	(-0.419, -0.294)	-0.055	(-0.077, -0.033)

Compartmental Model

	Estimate	95% Credible Interval
Intercept	0.324	(0.253, 0.376)
Week	-1.409	(-3.348, 0.609)
$\pi^{(\mathcal{IR})}$	0.101	(0.089, 0.115)

3.7 Real Data Analysis

For both models, the Raftery and Lewis diagnostic test was used to inform the number of iterations to be used for burn-in when using the Geweke diagnostic, which was used as a measure of convergence (Geweke, 1992; Raftery and Lewis, 1992). Both diagnostic tests were implemented using the `coda` package in R (Plummer et al., 2006).

The multinomial logistic regression model was run for 150,000 iterations; every 10th iteration was saved to address minor auto-correlation. A burn-in of 120 saved iterations was used for all parameters based on the Raftery and Lewis diagnostic test. All twelve parameters had test statistics from the Geweke diagnostic test with absolute values less than 1.96, so there was no evidence of lack of convergence for this model. Posterior parameter estimates are included in Table 3.4. Note, sampling methods like Hamiltonian Monte Carlo, which are designed to reduce autocorrelation

by more efficiently exploring the parameter space, can be employed in this scenario (Neal, 2011). However, since the autocorrelation is not severe, and there is no evidence of lack of convergence, we did not employ any of these sampling schemes.

The compartmental model was run for two million iterations to ensure convergence; every 500th iteration was saved because auto-correlation was high in this model. Note, sampling methods like Hamiltonian Monte Carlo are difficult to apply in this setting due to the discrete nature of compartment membership counts. Such an approach would require the introduction of continuous likelihood approximations, which may not be a good fit. Four separate MCMC chains were run in parallel. A burn-in of 334 saved iterations was used. The resulting Geweke test statistics for β_0 and β_1 were -0.2458 and 1.2989, respectively, so there was no significant evidence against convergence. The test statistics for the γ's were -0.5996, -0.6083, -0.6276, 0.6098, respectively, so similarly there was no significant evidence against convergence for these parameters. The test statistic for $\pi^{(\mathcal{IR})}$ was -0.9536. These parameters form a sufficient basis for simulation and drive epidemic behavior. Therefore, in the high dimensional context of compartmental models, these results are indicative of overall model convergence.

An added advantage of using the SIR framework to model this infection data is that we can easily calculate point estimates and credible intervals for the empirically-adjusted reproductive number, $\mathcal{R}^{(EA)}$, at each time point, t. These estimates and credible intervals are shown in Figure 3.5. In the first 4 weeks, $\mathcal{R}^{(EA)}$ is greater than 1, indicating that on average, an infectious individual infects more than 1 person.

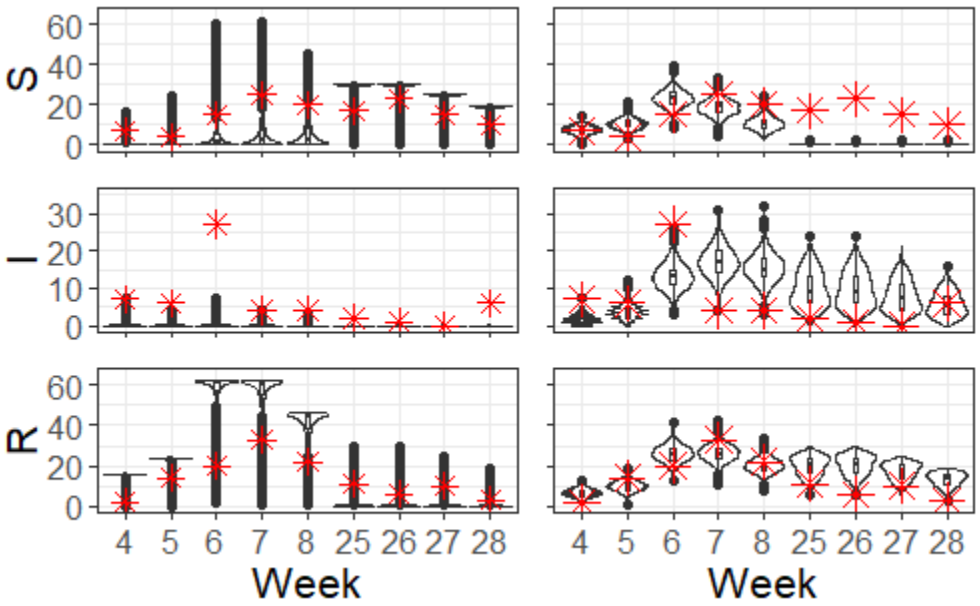

Figure 3.4: Comparison of posterior predictive counts from multinomial logistic regression (left) and the compartmental model (right) to observed counts (red stars) for each compartment at each week. At each week, 5000 counts were generated from the posterior predictive distribution; the distributions are represented using violin plots. The compartmental model allows for more variability in the posterior predictive counts and as a consequence does a better job of capturing the observed counts, particularly for the infectious state, where the counts are generally small.

Table 3.4: Model Fit Summary - Bayesian Multinomial Logistic Regression

	$\mathcal{S}/\mathcal{R}$		$\mathcal{I}/\mathcal{R}$	
	Estimate	95% Credible Interval	Estimate	95% Credible Interval
Intercept	-0.104	(-0.999, 0.961)	-0.112	(-1.642, 0.727)
Area: Urban	0.093	(-0.573, 0.760)	0.303	(-0.425, 1.038)
Area: Periurban	0.410	(-0.274, 1.093)	-0.387	(-1.376, 0.542)
Sex: Male	-0.523	(-1.077, 0.022)	-0.870	(-1.626, -0.160)
Age $>$ 35	-0.827	(-1.364, -0.304)	-0.225	(-0.886, 0.438)
Week	0.046	(0.016, 0.076)	-0.029	(-0.077, 0.016)

After 4 weeks, this number drops below 1. This change is due to the seasonality of the infection; it does not indicate that the infection will no longer be present in the population.

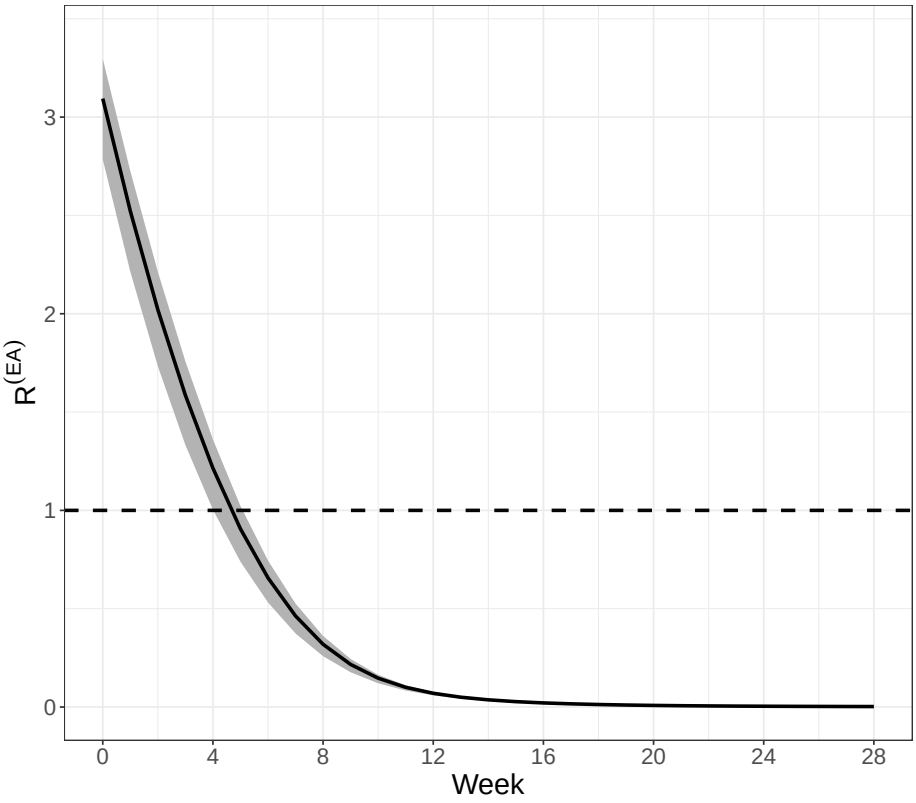

Figure 3.5: Empirically-adjusted reproductive number over time. Symmetric 95% credible intervals are in grey.

3.8 Discussion

In this chapter, we proposed two models, a multinomial logistic regression model and a SIR model, to estimate the proportions of individuals that fall into

different infection states. We explored these models both in the context of simulation studies and of human *L. infantum* infection in northeastern Brazil. We also adapted the empirically-adjusted reproductive number introduced by Brown et al. (2016) to obtain estimates and credible intervals for a reproductive number for VL.

When the simulated data arose from a multinomial likelihood, informal model comparisons suggested that the multinomial logistic regression model and the SIR model perform comparably. In contrast, when the simulated data arose from a compartmental model, informal model comparisons suggested that the SIR model outperformed the multinomial logistic regression model. We expect some variation due to seasonality, interventions, or natural host-vector-pathogen dynamics. The compartmental modeling approach assumed that this variability exists and offers a convenient framework for including it in the model (Lekone and Finkenstädt, 2006). Formal model comparisons can be found in Chapter 5.

Analysis of data from the *L. infantum* endemic region supported the conclusion that the compartmental model is a more reasonable modeling choice for this dynamic endemic site when the objective is to estimate the proportion of individuals in the population in each infection state over time. The compartmental model captured the observed counts better than the multinomial logistic regression model. The framework of the SIR model also allowed us to easily compute estimates for the empirically-adjusted reproductive number at each time point, along with credible intervals. As shown in Figure 3.5, $\mathcal{R}^{(EA)}$ was larger than 1 in the first four weeks of the year and dropped below 1 after the fourth week.

Multinomial logistic regression had two implementation advantages over the compartmental modeling approach. First, there are several R packages that implement the model, both in the frequentist and Bayesian frameworks. Second, the model runs quickly for reasonably small sample sizes. In spite of these advantages, the multinomial logistic regression did not perform as well as the compartmental model when the objective was to estimate the proportion of individuals in various states of infection across time. In Chapter 5, we will show that there is decisive evidence in favor of the SIR model. We also quantified a reproductive number for the infection.

This compartmental modeling approach may be applicable to other infectious diseases besides VL that have known seasonality, established/experimental interventions, or other host-vector-pathogen dynamics. Infectious diseases like Lyme disease and dengue hemorrhagic fever may benefit from analysis using this approach because of seasonality, for example (Piesman et al., 1987; Schwartz et al., 2008). However, this modeling approach may be ill-suited to infectious diseases that may have many potential sources of infection, like sepsis, because it is difficult to determine what form the transition probabilities should take. Furthermore, if we do not have good information about the biology/dynamics of a particular disease or infection, the compartmental model may not provide sensible results.

CHAPTER 4
MULTIPLE SPECIES TRANSMISSION MODELS

4.1 The SAYVR Model

Various compartmental epidemic models have been developed to accommodate a range of infection dynamics. In addition to the SIR model employed in Chapters 2 and 3, more complex models include the SEIR model and the carrier state model. Both the carrier state model and the SEIR model include multiple infected compartments, but these compartments are subject to restrictions, which may limit application to infectious diseases like leishmaniasis (Keeling and Rohani, 2008). In the SEIR model, individuals that occupy the "exposed" (E) category are infected but not yet able to transmit infection. In the carrier state model, some previously infectious individuals do not fully recover, but rather transition to a carrier state where they do not display symptoms and are less infectious. These carriers can transition back to an infectious (symptomatic) state. Other infectious individuals fully recover. Neither of these models, however, accommodate a disease process where infected individuals can occupy an infectious asymptomatic class before potentially transitioning to an infectious symptomatic class.

Within the deterministic compartmental framework, a number of models have been proposed that separate asymptomatic and symptomatic infectious individuals specifically for the study of VL. Rock et al. (2015) provide a thorough literature review of these mathematical models. Le Rutte et al. (2017) and Zou et al. (2017)

make additional contributions to this set of models. While these models can be implemented through differential equations, they are generally more complex than what we can include in a stochastic compartmental model.

In this chapter, we introduce a Bayesian SAYVR model which, like the deterministic models described above, has both asymptomatic ($\mathcal{A}$) and symptomatic ($\mathcal{Y}$) compartments, but takes advantage of the stochastic framework, as described in Section 1.2. This model also includes separate recovered ($\mathcal{V}$) and removed ($\mathcal{R}$) categories to accommodate *L. infantum* infection since symptomatic humans either recover from clinical disease or succumb to it and die (Alvar et al., 2012; Martins-Melo et al., 2014). Since culling is used as a control measure for dogs in Brazil, we assume that the probability of transitioning into that compartment is zero (for dogs) in this chapter. If this model were applied to study leishmaniasis in Europe where dogs are treated and allowed to recover, however, we would include a non-zero transition probability to $\mathcal{V}$ for dogs, too. We incorporate the contributions of each host species and the sand fly vector into the $\mathcal{S} \to \mathcal{A}$ transition probability. The model is summarized in Figure 4.1, where we allow a number of transitions and incorporate the interactions among dogs, sand flies, and humans. Since we do not have human or canine data to fit this model, the main purpose of this chapter is introduce a stochastic compartmental model that can be used for future analyses.

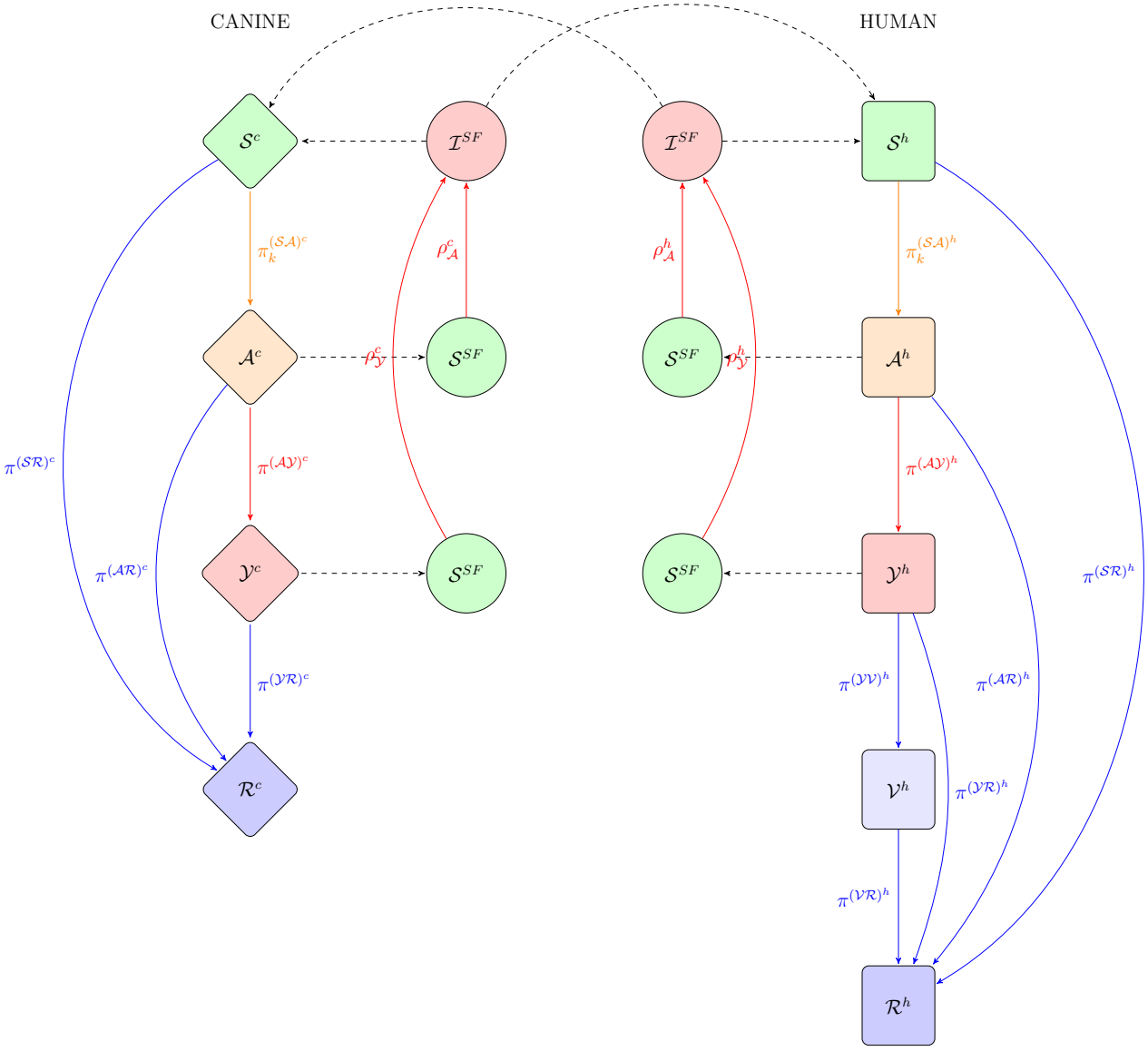

Figure 4.1: Schematic for SAYVR compartmental models for dogs (diamonds), for humans (squares), and the sand flies that connect the processes (circles). Solid arrows denote transitions between compartments within a species; dashed lines denote infection transmission involving the sand fly vector.

4.2 Model Specification

4.2.1 Sand Fly Data

As part of an effort to understand sand fly feeding behaviors, a lab-based study was performed at the University of Iowa by the Petersen Lab. Sand flies were fed on sedated dogs with varying VL clinical status for thirty minutes, after which those female sand flies containing a visible blood meal were separated and incubated for 48 days. Then, they were frozen individually in 100 μL of DNA lysis buffer. Using a Qiagen Puregene Gentra kit, total DNA was isolated from individual blood fed flies and eluted with 30 μL of buffer. Next, qPCR was performed. A sand fly was considered infected if it contained greater than one parasite equivalent. In total, 292 sand flies were fed on dogs undergoing xenodiagnosis. Of those 292 flies, 251 were fed on asymptomatic dogs and 41 were fed on symptomatic dogs.

4.2.2 Data Model

The data described in 4.2.1 give us information about the probabilities that sand flies become infected after feeding on asymptomatic or symptomatic canine hosts. This information has to be obtained in a laboratory setting so that we can distinguish between flies that have fed on asymptomatic versus symptomatic dogs. To incorporate this information into our population compartmental model, we assume that if sand flies fed on these two kinds of hosts in a natural setting in Brazil, then they would become infected with these probabilities. We also assume that these probabilities are constant as a function of time. Let $Z_{\mathcal{A}}^{(\mathrm{SF})}$ and $Z_{\mathcal{Y}}^{(\mathrm{SF})}$ denote the number

of sand flies that were found to be infected when they were fed on asymptomatic and symptomatic dogs, respectively. The contributions to the likelihood are

$$Z_{\mathcal{A}}^{(\mathrm{SF})} \sim \text{Binomial}\left(N_{\mathcal{A}}^{(\mathrm{SF})}, \rho_{\mathcal{A}}\right), \tag{4.1}$$

$$Z_{\mathcal{Y}}^{(\mathrm{SF})} \sim \text{Binomial}\left(N_{\mathcal{Y}}^{(\mathrm{SF})}, \rho_{\mathcal{Y}}\right), \tag{4.2}$$

where $N_{\mathcal{A}}^{(\mathrm{SF})} = 251$, $N_{\mathcal{Y}}^{(\mathrm{SF})} = 41$, $\rho_{\mathcal{A}} = 155/251$, and $\rho_{\mathcal{Y}} = 37/41$.

4.2.3 Process Model

For each host species, r, and at each month, t, B_t^r individuals are born. We assume that this birth process follows the distribution:

$$B_t^r \sim \text{Poisson}\left(\nu_t^r\right). \tag{4.3}$$

Of the B_t^r individuals born at time t, we assume that for dogs ($r = c$), some portion of those individuals are born asymptomatic. While this is an oversimplification of the vertical transmission process, it is a standard way of incorporating vertical transmission into a population level compartmental epidemic model (Anderson and May, 1979; Li et al., 2001; Zou et al., 2017). The number of dogs that are born asymptomatic, $B_t^{(\mathcal{A})^c}$, is assumed to follow:

$$B^{(\mathcal{A})_t^c} \sim \text{Binomial}\left(B_t^c, \left(\frac{A_{t-1}^c + Y_{t-1}^c}{N_{t-1}^c}\right)\gamma^c\right). \tag{4.4}$$

The probability of an asymptomatic birth depends on the proportion of infected (asymptomatic or symptomatic) dogs in the population and γ^c, the probability that a pup exposed to an infected dam in utero becomes infected.

The compartment totals at each time t depend on the birth processes previously described, the compartment sizes at the previous time point, and on the transition compartments at the current time point. Note that for humans ($r = h$), there are five compartments, whereas for dogs ($r = c$), there are only four. Dogs are not allowed to recover, as previously mentioned. For a species in r, they have the following form:

$$
\begin{aligned}
S_t^r &= S_{t-1}^r + B_t^{(\mathcal{S})^r} - A_t^{(\mathcal{SA})^r} - R_t^{(\mathcal{SR})^r};\\
A_t^r &= A_{t-1}^r + B_t^{(\mathcal{A})^r} + A_t^{(\mathcal{SA})^r} - Y_t^{(\mathcal{AY})^r} - R_t^{(\mathcal{AR})^r};\\
Y_t^r &= Y_{t-1}^r + Y_t^{(\mathcal{AY})^r} - V_t^{(\mathcal{YV})^r} - R_t^{(\mathcal{YR})^r};\\
V_t^r &= V_{t-1}^r + V_t^{(\mathcal{YV})^r} - R_t^{(\mathcal{VR})^r};\\
R_t^r &= R_{t-1}^r + R_t^{(\mathcal{SR})^r} + R_t^{(\mathcal{AR})^r} + R_t^{(\mathcal{YR})^r} + R_t^{(\mathcal{VR})^r}.
\end{aligned}
$$

There are up to seven transition compartments. The notation conveys the number of individuals that are transitioning at time t, the previously occupied compartment, and the compartment to which the individuals are transitioning. For example, $A_t^{(\mathcal{SA})^r}$ denotes the number of individuals that are transitioning from $\mathcal{S}$ to $\mathcal{A}$ at time t. The transition compartments are modeled according to a chain binomial structure, as

shown in Equations 4.5-4.9. For dogs ($r = c$), $V_t^{(\mathcal{YV})^c} = R_t^{(\mathcal{VR})^c} = 0 \ \forall t$.

$$A_t^{(\mathcal{SA})^r} \sim \text{Binomial}\left(S_{t-1}^r, \pi_t^{(\mathcal{SA})^r}\right); \tag{4.5}$$

$$R_t^{(\mathcal{SR})^r} \sim \text{Binomial}\left(S_{t-1}^r - A_t^{(\mathcal{SA})^r}, \pi^{(\mathcal{SR})^r}\right); \tag{4.6}$$

$$Y_t^{(\mathcal{AY})^r} \sim \text{Binomial}\left(A_{t-1}^r, \pi^{(\mathcal{AY})^r}\right); \tag{4.7}$$

$$R_t^{(\mathcal{AR})^r} \sim \text{Binomial}\left(A_{t-1}^r - Y_t^{(\mathcal{AY})^r}, \pi^{(\mathcal{AR})^r}\right); \tag{4.8}$$

$$V_t^{(\mathcal{YV})^r} \sim \text{Binomial}\left(Y_{t-1}^r, \pi^{(\mathcal{YV})^r}\right); \tag{4.9}$$

$$R_t^{(\mathcal{YR})^r} \sim \text{Binomial}\left(Y_{t-1}^r - V_t^{(\mathcal{YV})^r}, \pi^{(\mathcal{YR})^r}\right); \tag{4.10}$$

$$R_t^{(\mathcal{VR})^r} \sim \text{Binomial}\left(V_{t-1}^r, \pi^{(\mathcal{VR})^r}\right). \tag{4.11}$$

As usual for endemic sites, we have initial compartment sizes, S_0^r, A_0^r, Y_0^r, and V_0^r for each host species r. These initial compartment counts are assumed to follow binomial distributions, as shown in Equations 4.12-4.15. Note that $R_0^r = 0$ for both host species because individuals in this category are deceased and $V_0^c = 0$ since dogs are not allowed to recover per national policy in Brazil (Nunes et al., 2008).

$$S_0^r \sim \text{Binomial}\left(N_0^r, \pi^{(S_0)}\right) \tag{4.12}$$

$$A_0^r \sim \text{Binomial}\left(N_0^r - S_0^r, \pi^{(A_0)}\right) \tag{4.13}$$

$$Y_0^h \sim \text{Binomial}\left(N_0^h - S_0^h - A_0^h, \pi^{(Y_0)}\right); \ Y_0^c = N_0^c - S_0^c - A_0^c \tag{4.14}$$

$$V_0^r = N_0^r - S_0^r - A_0^r - Y_0^r \tag{4.15}$$

4.2.4 Parameter Model

We expect $\pi_t^{(\mathcal{SA})^r}$, the probability of transitioning from $\mathcal{S}$ to $\mathcal{A}$ for each host species r, to depend on the proportion of infectious sand flies in the population. This

proportion should depend on three things: (1) the proportions of asymptomatic dogs, asymptomatic humans, symptomatic dogs, and symptomatic humans at a particular time t, (2) the probability that an uninfected sand fly becomes infected after contact with an infected (asymptomatic or symptomatic) dog or human ($\rho^r_{\mathcal{A}},\ \rho^r_{\mathcal{Y}}\ \in (0,1)$), and (3) sand fly feeding preference ($\alpha \in (0,1)$). This gives the expression:

$$\begin{aligned}
\delta_t &= \alpha \left[\rho^h_{\mathcal{A}} \left(\frac{A^h_t}{N^h_t} \right) + \rho^h_Y \left(\frac{Y^h_t}{N^h_t} \right) \right] + (1-\alpha) \left[\hat{\rho}^c_{\mathcal{A}} \left(\frac{A^c_t}{N^c_t} \right) + \hat{\rho}^c_Y \left(\frac{Y^c_t}{N^c_t} \right) \right], \\
&= \delta^h_t + \delta^c_t, \\
&= \alpha \cdot \rho^h_{\mathcal{A}} \left(\frac{A^h_t}{N^h_t} \right) + \alpha \cdot \rho^h_Y \left(\frac{Y^h_t}{N^h_t} \right) + (1-\alpha) \cdot \hat{\rho}^c_{\mathcal{A}} \left(\frac{A^c_t}{N^c_t} \right) + (1-\alpha) \cdot \hat{\rho}^c_Y \left(\frac{Y^c_t}{N^c_t} \right), \\
&= \delta^{h(\mathcal{A})}_t + \delta^{h(\mathcal{Y})}_t + \delta^{c(\mathcal{A})}_t + \delta^{c(\mathcal{Y})}_t,
\end{aligned}$$

where $\hat{\rho}^c_{\mathcal{A}} = X^{(\mathrm{SF})}_{\mathcal{A}} / N^{(\mathrm{SF})}_{\mathcal{A}}$ and $\hat{\rho}^c_{\mathcal{Y}} = X^{(\mathrm{SF})}_{\mathcal{Y}} / N^{(\mathrm{SF})}_{\mathcal{Y}}$ from our data model.

The typical lifespan for *Lutzomyia longipalpis* sand flies is 10 days or less in their natural habitat, so flies will be born (uninfected) and die during each month, the time scale on which we are modeling the infection process (Diaz-Albiter et al., 2009). To reconcile these two time scales, we account for flies that are infectious at the beginning of a month and those that are not infectious, but can become infectious during that month and then bite a susceptible dog. The $\mathcal{S} \to \mathcal{A}$ transition probability is a first order Markov Process,

$$\pi^{(\mathcal{SA})^r}(t) = 1 - \exp\left\{ -\left[\delta_t + (1-\delta_t)\delta_t \right] e^{X_t \boldsymbol{\beta}^r} \right\}. \tag{4.16}$$

We derive this transition probability under the same homogeneous Poisson contact process assumption as used in Section 3.4.3, where individuals in r are bitten

by flies at some rate λ_t and become infected with probability p^r. As before, let $Inf^r(t)$ be the event that an individual from host species r becomes infected at time t and $\overline{Inf^r(t)}$ be the complement. Then, the probability of the event $Inf^r(t)$ is

$$P\left(Inf^r(t)\right) = 1 - P\left(\overline{Inf^r(t)}\right),$$

where

$$\begin{aligned}
P\left[\overline{Inf^r(t)}\right] &= \mathbb{E}\left[\overline{Inf^r(t)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\overline{Inf^r(t)}|K=k\right]\right] \\
&= \sum_{k=0}^{\infty}\left[1-\left(\delta_t p^r + (1-\delta_t)\,\delta_t p^r\right)\right]^k \left[\frac{\exp\left\{-\lambda_t\right\}\lambda_t^k}{k!}\right] \\
&= \exp\left\{\left[1-\left(\delta_t p^r + (1-\delta_t)\,\delta_t p^r\right)\right]\lambda_t\right\}\exp\left\{-\lambda_t\right\} \\
&\quad \times \underbrace{\sum_{k=0}^{\infty}\frac{e^{-\left[1-\left(\delta_t p^r+(1-\delta_t)\delta_t p^r\right)\right]\lambda_t}\left\{\left[1-\left(\delta_t p^r+(1-\delta_t)\delta_t p^r\right)\right]\lambda_t\right\}^k}{k!}}_{1} \\
&= \exp\left\{-\delta_t\left(\lambda_t p^r\right) - (1-\delta_t)\delta_t\left(\lambda_t p^r\right)\right\} \\
&= \exp\left\{-\left[\delta_t + (1-\delta_t)\delta_t\right]\left(\lambda_t p^r\right)\right\} \\
&= \exp\left\{-\left[\delta_t + (1-\delta_t)\delta_t\right]e^{X_t\boldsymbol{\beta}^r}\right\}.
\end{aligned}$$

Note that we let $\log(\lambda_t p^r) = X_t'\boldsymbol{\beta}_t$ because λ_t and p^r are generally not identifiable, as discussed in Section 3.4.3.

The remaining transition probabilities are biological because they depend on immune responses, for example. They are not, however, tied to the seasonality of transmission, so they do not depend on t in this model. We assume that they follow beta distributions, as shown in Equations 4.17-4.20. The canine transition probabilties $\pi^{(\mathcal{V}\mathcal{V})^c}$ and $\pi^{(\mathcal{V}\mathcal{R})^c}$ are set to zero because dogs do not enter the recovered

compartment due to culling. To ensure biological plausibility, we suggest centering the priors for each of these six transition probabilities at a biologically reasonable value and setting the scale parameter, e.g. $\beta^{(\mathcal{SR})^r} = 1$.

$$\pi^{(\mathcal{SR})^r} \sim \text{Beta}\left(\alpha^{(\mathcal{SR})^r}, \beta^{(\mathcal{SR})^r}\right); \tag{4.17}$$

$$\pi^{(\mathcal{AY})^r} \sim \text{Beta}\left(\alpha^{(\mathcal{AY})^r}, \beta^{(\mathcal{AY})^r}\right); \tag{4.18}$$

$$\pi^{(\mathcal{AR})^r} \sim \text{Beta}\left(\alpha^{(\mathcal{AR})^r}, \beta^{(\mathcal{AR})^r}\right); \tag{4.19}$$

$$\pi^{(\mathcal{YV})^r} \sim \text{Beta}\left(\alpha^{(\mathcal{YV})^r}, \beta^{(\mathcal{YV})^r}\right); \tag{4.20}$$

$$\pi^{(\mathcal{YR})^r} \sim \text{Beta}\left(\alpha^{(\mathcal{YR})^r}, \beta^{(\mathcal{YR})^r}\right); \tag{4.21}$$

$$\pi^{(\mathcal{VR})^r} \sim \text{Beta}\left(\alpha^{(\mathcal{VR})^r}, \beta^{(\mathcal{VR})^r}\right). \tag{4.22}$$

4.3 Infection Source-specific Empirically Adjusted Reproductive Numbers

A susceptible human or dog can derive *L. infantum* infection from one of several sources: asymptomatic humans, symptomatic humans, asymptomatic dogs, or symptomatic dogs. To better understand the contributions of each of these sources to maintaining infection in a particular host population, it is useful to calculate a reproductive number that corresponds to each source. To this end, let $I_i^r(t, s_b)$ be the event that individual i in host species r becomes infected as a result of being bitten by a sand fly that derived its infection from an individual in $b \in \mathcal{B}^* = \left\{A^h, Y^h, A^c, Y^c\right\}$. Then, the expected number of secondary infections in species r attributable to an

individual in b is

$$\mathbb{E}\left[\sum_{i=0}^{N_t^r}\left(I_i^r(t,s_b)\right)\right] = S^r(t)\cdot P\left(I_i^r(t,s_b)|i\in\mathcal{S}^r\right)$$
$$= S^r(t)\cdot\pi^{(\mathcal{SA})^r}(t,s_b).$$

Furthermore, the average number of infections in host species r per infectious individual in b is

$$G^r(t,s_b) = \frac{S^r(t)\cdot\pi^{(\mathcal{SA})^r}(t,s_b)}{b(t)}. \tag{4.23}$$

The source specific transition probabilities, $\pi^{(\mathcal{SA})^r}(t,s_b)$, are derived in Appendix C. This derivation relies on the assumption that these infectious sources are mutually disjoint - an individual can become infected as a result of only one infectious source.

The quantity in Equation 4.23 can be thought of as an entry in the "next generation matrix" (Allen and van den Driessche, 2008; Brown et al., 2016), where the columns correspond to the original source of the infection carried by the sand fly (asymptomatic dog, asymptomatic human, sympomatic dog, or symptomatic human) and the rows denote the infection states to which previously susceptible individuals can transition as a result of being bitten by an infected sand fly. The next generation matrix is

$$\mathbf{G}^r(t) = \begin{array}{c} \\ \mathcal{A} \\ \mathcal{Y} \end{array}\!\!\begin{array}{c} \begin{array}{cccc} \mathcal{A}^c & \mathcal{A}^h & \mathcal{Y}^c & \mathcal{Y}^h \end{array} \\ \left|\begin{array}{cccc} \frac{\pi^{(\mathcal{SA})^r}(t,s_{\mathcal{A}^c})S^r(t)}{A^c(t)} & \frac{\pi^{(\mathcal{SA})^r}(t,s_{\mathcal{A}^h})S^r(t)}{A^h(t)} & \frac{\pi^{(\mathcal{SA})^r}(t,s_{\mathcal{Y}^c})S^r(t)}{Y^c(t)} & \frac{\pi^{(\mathcal{SA})^r}(t,s_{\mathcal{Y}^h})S^r(t)}{Y^h(t)} \\ 0 & 0 & 0 & 0 \end{array}\right| \end{array}.$$

To calculate the ISEARN, we need to account for the probability that an infectious individual remains in an infectious state at each time t. Let $T_i^r(t)$ be the event

that individual i in host species r is infectious at time t, $T_i^r(t, A)$ be the event that individual i in host species r is asymptomatic at time t, and $T_i^r(t, Y)$ be the event that individual i in host species r is symptomatic at time t. The probability an infectious

individual in host species r remains infectious at time t is

$$
\begin{aligned}
P\left[T_i^r(t)|T_i^r(t-1)\right] &= P\left[T_i^r(t,A)\cup T_i^r(t,Y)|T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right] \\
&= P\left[T_i^r(t,A)|T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right] \\
&+ P\left[T_i^r(t,Y)|T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right] \text{ since } T_i^r(t,A)\cap T_i^r(t,Y)=\emptyset \\
&= \frac{P\left[T_i^r(t,A)\cap\{T_i^r(t-1,A)\cup T_i^r(t-1,Y)\}\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]} \\
&+ \frac{P\left[T_i^r(t,Y)\cap\{T_i^r(t-1,A)\cup T_i^r(t-1,Y)\}\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]} \\
&= \frac{P\left[\{T_i^r(t,A)\cap T_i^r(t-1,A)\}\cup\{T_i^r(t,A)\cap T_i^r(t-1,Y)\}\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]} \\
&+ \frac{P\left[\{T_i^r(t,Y)\cap T_i^r(t-1,A)\}\cup\{T_i^r(t,Y)\cap T_i^r(t-1,Y)\}\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]} \\
&= \frac{P\left[T_i^r(t,A)\cap T_i^r(t-1,A)\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]} + \frac{\overbrace{P\left[T_i^r(t,A)\cap T_i^r(t-1,Y)\right]}^{0}}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]} \\
&+ \frac{P\left[T_i^r(t,Y)\cap T_i^r(t-1,A)\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]} + \frac{P\left[T_i^r(t,Y)\cap T_i^r(t-1,Y)\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]} \\
&= \left(\frac{P\left[T_i^r(t,A)\cap T_i^r(t-1,A)\right]}{P\left[T_i^r(t-1,A)\right]}\right)\left(\frac{P\left[T_i^r(t-1,A)\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]}\right) \\
&+ \left(\frac{P\left[T_i^r(t,Y)\cap T_i^r(t-1,A)\right]}{P\left[T_i^r(t-1,A)\right]}\right)\left(\frac{P\left[T_i^r(t-1,A)\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]}\right) \\
&+ \left(\frac{P\left[T_i^r(t,Y)\cap T_i^r(t-1,Y)\right]}{P\left[T_i^r(t-1,Y)\right]}\right)\left(\frac{P\left[T_i^r(t-1,Y)\right]}{P\left[T_i^r(t-1,A)\cup T_i^r(t-1,Y)\right]}\right) \\
&= \pi^{(\mathcal{AA})^r}\left(\frac{A_{t-1}^r/N_{t-1}^r}{(A_{t-1}^r+Y_{t-1}^r)/N_{t-1}^r}\right) + \pi^{(\mathcal{AY})^r}\left(\frac{A_{t-1}^r/N_{t-1}^r}{(A_{t-1}^r+Y_{t-1}^r)/N_{t-1}^r}\right) \\
&+ \pi^{(\mathcal{YY})^r}\left(\frac{Y_{t-1}^r/N_{t-1}^r}{(A_{t-1}^r+Y_{t-1}^r)/N_{t-1}^r}\right) \\
&= \left(1-\pi^{(\mathcal{AR})^r}\right)\left(\frac{A_{t-1}^r}{A_{t-1}^r+Y_{t-1}^r}\right) + \left(1-\pi^{(\mathcal{YV})^r}-\pi^{(\mathcal{YR})^r}\right)\left(\frac{Y_{t-1}^r}{A_{t-1}^r+Y_{t-1}^r}\right) \\
&\equiv \pi_t^{(\mathcal{II})^r}.
\end{aligned}
$$

Using this term, and calculating the total weighted average number of expected in-

fections over time to generalize the result to the pathogen's lifetime, we arrive at the ISEARN,

$$\mathcal{R}^{(\mathcal{EA})^r}(t, s_b) = \sum_{w=t}^{t_\infty} G^r(w, s_b) \left(\frac{b(w)}{\sum_{b \in \mathcal{B}^*} b(w)} \right) \prod_{k=1}^{t} \pi_k^{(\mathcal{II})^r}. \tag{4.24}$$

If any infectious category is unoccupied at time t, the corresponding entry in $\mathbf{G}^r(t)$ is defined to be 0, and the reproductive number for that source will also be 0.

Equation 4.24 allows us to calculate the expected portion of (vector-transmitted) infections attributable to infectious source b. For example, $\mathcal{R}^{(\mathcal{EA})}(t, s_{A^h})$ is the expected portion of new vector-borne infections attributable to asymptomatic humans at time t. We can interpret the ISEARN similarly to other reproductive numbers. If $\mathcal{R}^{(\mathcal{EA})}(t, s_b)$ is greater than one, then the parasite is expected to further colonize the population as a result of vector transmission from infectious source $b \in \mathcal{B}^*$. These quantities give insight into the contributions of various infectious sources in maintaining infection in a population.

If $\mathcal{R}^{(\mathcal{EA})}(t, s_b) < 1 \ \forall b \in \mathcal{B}^*$, no one source results in further parasite colonization in population r, but multiple sources may maintain infection in that population. We can calculate an overall reproductive number to capture the total expected number of new infections due to vector exposure from all sources. As a consequence of the assumption that the set of events $\{I^r(t, s_b)\}$ are mutually disjoint, we have the property that $\sum_{b \in \mathcal{B}} \pi^{(\mathcal{SA})^r}(t, s_b) = \pi^{(\mathcal{SA})^r}(t)$. This leads to Equation 4.25,

$$\mathcal{R}_{add}^{(\mathcal{EA})}(t) = \sum_{b \in \mathcal{B}^*} \sum_{w=t}^{t_\infty} G(w, s_b) \left(\frac{b(w)}{\sum_{b \in \mathcal{B}^*} b(w)} \right) \prod_{k=1}^{t} \pi_k^{(\mathcal{II})^r}, \tag{4.25}$$

where the contributions for each species are weighted by the proportion of infectious

individuals for which they account. This is consistent with Brown et al. (2016) - empirically-adjusted reproductive numbers are additive across spatial locations. If $\mathcal{R}_{add}^{(\mathcal{EA})}$ is greater than one, infection will grow in the population due to vector transmission from all sources. If it is less than one, the opposite is true.

4.4 Simulation Studies

4.4.1 Simulation Objective

The SAYVR model proposed in Section 4.2 is intended for future application to VL data to be collected in Brazil over the next two to three years. Thus, the primary objective of this simulation study is to validate the proposed model and show that the underlying parameters in the process can be recovered from simulated data. The simulated data sets employed in this section reflect the overarching structure of VL progression, and we have aspired to choose parameter values that match the biological process. However, there is not yet prior data to inform all of the parameters, so some parameter choices may not match the true biological progression. Since the objective of this section is model validation, not inference, these choices are less important.

4.4.2 Simulation Setup

To perform this simulation study, we generate 70 processes according to the SAYVR model specified in Section 4.2. The basis parameter values used in these simulations are summarized in Table 4.1. While some of these parameters were chosen based on VL literature (see Table 4.1 table notes), others were chosen out of convenience, so we do not expect these simulated processes to exactly reflect a real

L. infantum infection progression process. In this study, we treat $\mathbf{Y}^{(\mathcal{AY})^r}$, $\mathbf{R}^{(\mathcal{AR})^r}$, $\mathbf{V}^{(\mathcal{YV})^r}$, and $\mathbf{R}^{(\mathcal{YR})^r}$ as observed. Note that in the context of a real study, we could obtain data that would inform these transition compartments from surveillance studies, treatments, and death records in the case of humans, and from surveillance studies and culling programs in the case of dogs. We then fit the SAYVR model to the remaining parameters using MCMC. The general algorithm is available in Appendix B.3. For each data set, we run the algorithm with three chains for 100,000 iterations each and save every 100th iteration to reduce autocorrelation. We assess convergence using the Gelman-Rubin diagnostic, which was implemented through the `R` package `coda` (Gelman and Rubin, 1992; Plummer et al., 2006).

4.4.3 Simulation Results

Simulated infection processes were generated according to the model described in Section 4.2. Under this model, it was possible to generate processes with transition compartments that have zero counts for all times. Since the associated transition probabilities cannot be estimated in this situation, this is not considered a valid infection process to which to fit our model. For all other simulated processes, the SAYVR model was fit to the simulated data using MCMC. To improve mixing and thus speed up convergence, the acceptance ratios for basis parameters fit using Metropolis Hastings ($\boldsymbol{\beta}$, in this case) were checked every 100th iteration. If the acceptance ratio was greater than 0.8, 0.05 was added to the proposal standard deviation. In future analyses, one could also establish a rule for decreasing the proposal standard deviation

Table 4.1: SAVYR model parameter simulation values.

	Simulation Value	
Parameter	**Human**	**Dog**
ϕ^r	$(14.16/1000)/12$[1]	$0.0011(365/12)$[2]
ν_t^r	$\phi^h \cdot N_t^h$	$\phi^c \cdot N_t^c$
N_0^r	200,000	100,000
$\pi^{(S_0)^r}$	0.51[3]	0.50
$\pi^{(A_0)^r}$	0.35[3]	0.30
$\pi^{(Y_0)^r}$	0.06	0.20
$\pi^{(V_0)^r}$	0.08	—
α	0.30	0.30
γ	—	0.25
$\rho_{\mathcal{A}}$	0.0001	$155/251$[4]
$\rho_{\mathcal{Y}}$	0.113[5]	$37/41$[4]
β_0	-4	-4
β_1	-10	-10
$\pi^{(\mathcal{SR})^r}$	1.00×10^{-5}	1.00×10^{-5}
$\pi^{(\mathcal{AY})^r}$	3.83×10^{-6}[6]	5.00×10^{-3}
$\pi^{(\mathcal{AR})^r}$	1.00×10^{-4}	1.00×10^{-3}
$\pi^{(\mathcal{YR})^r}$	2.00×10^{-3}	2.00×10^{-1}
$\pi^{(\mathcal{YV})^r}$	5.00×10^{-2}	—
$\pi^{(\mathcal{VR})^r}$	1.00×10^{-3}	—

[1] World Bank, crude human birthrate in Brazil, 2016, on month scale;

[2] Daily canine birthrate, on month scale (Courtenay et al., 2002);

[3] Jeronimo et al. (2004);

[4] Petersen Lab - University of Iowa;

[5] $0.005(42/83) + 0.995(12/81 + 32/201 + 42/83)$ (Deane and Deane, 1955; Sherlock, 1996; Molina et al., 1999; Costa et al., 2000); convex combination of HIV positive proportion;

[6] Lima et al. (2018)

if the acceptance ratio was too low (e.g. less than 0.2). In this case, it would be prudent to increase or decrease the proposal standard deviation by a multiplicative factor to ensure than this proposal quantity is always greater than zero. The cut-off for the Gelman-Rubin diagnostic was 1.2.

We were able to recover the underlying infection processes through the model fits for these valid infection processes. The medians and associated confidence intervals for the basis parameters, averaged across the various simulated data set model fits, are in Table 4.2. Although the medians are shrunken slightly towards zero, the 95% credible intervals generally cover the true parameter values.

Table 4.2: Model Fit Summary - SAYVR Simulated Data

	True value	**Median**	**95% Credible Interval**
β_0	-4.00×10^0	-4.00×10^0	$(-7.87 \times 10^0, -0.08 \times 10^0)$
β_1	-10.00×10^0	-9.97×10^0	$(-13.90 \times 10^0, -6.04 \times 10^0)$
$\pi^{(\mathcal{SR})^c}$	1.00×10^{-5}	9.26×10^{-6}	$(6.85 \times 10^{-7}, 2.50 \times 10^{-5})$
$\pi^{(\mathcal{SR})^h}$	1.00×10^{-5}	9.17×10^{-6}	$(2.11 \times 10^{-6}, 4.35 \times 10^{-5})$
$\pi^{(\mathcal{AY})^c}$	5.00×10^{-3}	4.81×10^{-3}	$(4.47 \times 10^{-3}, 5.19 \times 10^{-3})$
$\pi^{(\mathcal{AY})^h}$	3.83×10^{-6}	4.53×10^{-6}	$(8.64 \times 10^{-7}, 1.26 \times 10^{-5})$
$\pi^{(\mathcal{AR})^c}$	1.00×10^{-3}	9.41×10^{-4}	$(8.15 \times 10^{-4}, 1.09 \times 10^{-3})$
$\pi^{(\mathcal{AR})^h}$	1.00×10^{-4}	9.25×10^{-5}	$(6.48 \times 10^{-5}, 1.23 \times 10^{-4})$
$\pi^{(\mathcal{YR})^c}$	2.00×10^{-1}	1.65×10^{-1}	$(1.63 \times 10^{-1}, 1.68 \times 10^{-1})$
$\pi^{(\mathcal{YR})^h}$	2.00×10^{-3}	1.70×10^{-3}	$(1.38 \times 10^{-3}, 2.00 \times 10^{-3})$
$\pi^{(\mathcal{YV})^h}$	5.00×10^{-2}	4.46×10^{-2}	$(4.31 \times 10^{-2}, 4.61 \times 10^{-2})$
$\pi^{(\mathcal{VR})^h}$	1.00×10^{-3}	8.91×10^{-4}	$(7.24 \times 10^{-4}, 1.05 \times 10^{-3})$

4.5 Discussion

In this chapter, we proposed a Bayesian SAYVR model, which accommodates two infectious classes, asymptomatic ($\mathcal{A}$) and symptomatic ($\mathcal{Y}$), and two removal classes, recovered ($\mathcal{V}$) and removed ($\mathcal{R}$; death in this context). We developed models for both humans and dogs, and include the contributions of each of these host species and the sand fly vector into the probability of transitioning from $\mathcal{S}$ to $\mathcal{A}$, $\pi_t^{(\mathcal{SA})}$. The inclusion of two distinct infectious categories allowed us to evaluate the impact of asymptomatic individuals, which may go unnoticed, on infection propogation, for example. Two separate removal categories gave us the flexibility to consider this process over a much longer time period, since people can either recover from the disease or succumb to it (Belo et al., 2014). We also derived the ISEARN, which extended the empirically-adjusted reproductive number proposed by Brown et al. (2016) in two ways. First, we calculated reproductive numbers for each infectious source, which can lend insight into the impact of different sources on maintaining infection in the population. Second, we accounted for the fact that individuals can occupy either of two infectious categories. Both the proposed model and ISEARN are intended for future application to VL data.

While we did not have access to VL data for this dissertation chapter, we conducted a simulation study to ensure that the proposed SAYVR model could be fit using MCMC and that the underlying infection process could be recovered. In fitting these models, it was useful to update proposals for basis parameters that were fit using Metropolis Hastings periodically if the acceptance ratio was either too high

or too low. This was especially important because these models are large, so we can expect that in practice when fitting this model to real data, the computational time needed to achieve convergence could be long, so we would not want to have to refit the model to adjust proposals to improve mixing.

When these models are fit to VL data, there are several things to consider. First, it may be prudent to fix the initial compartments at reasonable values because mixing can be poor for these quantities. This also cuts down on computation time. Second, careful consideration should be put into prior choices for basis parameters. Generally, these parameters need to be centered at biologically plausible values and sensible prior distributions can be determined from external information to ensure that the resulting infection process makes sense (Ozanne et al., 2019). These well-considered prior choices also can be used to generate reasonable starting values for the algorithm, although more work will be needed to rigorously establish how to choose starting values.

CHAPTER 5
AN ABC-INSPIRED APPROACH TO COMPARING BAYESIAN MODELS USING BAYES FACTORS

5.1 Introduction

In practical settings, we often consider a set of candidate models for a particular phenomenon. Then, we choose from this set the model that "best" represents the observed data. Formal model comparison methods allow us to formalize what we mean by "best" and generally strike a balance between model fit and complexity. This allows us to choose a model that, relative to the other models under consideration, explains the data well without overfitting.

A Bayes factor is a well-established, theoretically sound method to compare two candidate Bayesian models. Bayes factors require the use of proper priors for parameters that differ between the two models of interest. This restriction is satisfied by the models examined in this chapter. The Bayes factor, BF_{12}, is calculated as a ratio of probabilities:

$$BF_{12} = \frac{P(\mathbf{Y}|\mathcal{M}_1)}{P(\mathbf{Y}|\mathcal{M}_2)} = \frac{P(\mathcal{M}_1|\mathbf{Y})}{P(\mathcal{M}_2|\mathbf{Y})} \cdot \frac{P(\mathcal{M}_2)}{P(\mathcal{M}_1)},$$

where $P(\mathbf{Y}|\mathcal{M}_i)$ is the probability of the observed data given Model i, $i \in \{1, 2\}$ (Kass and Raftery, 1995). These densities are obtained by integrating the joint posterior density, $P(\mathbf{Y}, \boldsymbol{\theta}_i|\mathcal{M}_i)$ over the parameter space, $\boldsymbol{\Theta}_i$.

The principal challenge to using a Bayes factor for model comparison is practical in nature. This integration is intractable in all but the simplest cases where conjugate distributions are employed (Kass and Raftery, 1995), so numerical meth-

ods must be employed to estimate these Bayes factors. While considerable work has already been done to address this computational challenge, existing methods are not generally applicable to compartmental models due to their non-normal likelihoods and large discrete parameter spaces (Kass and Raftery, 1995).

5.2 Methods

Approximate Bayesian computation (ABC) methods have been developed to fit large models that have likelihoods that are difficult or impossible to evaluate (Toni et al., 2009). While we are able to evaluate likelihoods for fitting our models of interest in this chapter, the framework for estimating Bayes factors for models fit using ABC gives us insight into constructing a method for estimating them for models fit using MCMC where at least one of the candidate models has a non-normal likelihood and a high-dimensional parameter space.

When employing ABC, we decompose the problem into basis parameters, $\boldsymbol{\theta}$, and data, $\mathbf{Y}$. A simple ABC algorithm is based on rejection sampling. First, we specify a prior distribution, $\pi(\boldsymbol{\theta})$. Then, for $j = 1, ..., K$, we draw samples $\boldsymbol{\theta}^{(j)}$ from $\pi(\boldsymbol{\theta})$. Based on $\boldsymbol{\theta}^{(j)}$, we simulate a data set, $\mathbf{Y}^{(j)}$. If $\rho\left(\mathbf{Y}^{(j)}, \mathbf{Y}\right) > \epsilon$, then we reject $\boldsymbol{\theta}^{(j)}$, where $\rho(\cdot)$ is some distance metric. Note that the tolerance, ϵ, is specified in advance and quantifies how "close" to the data we require our sample to be to accept the generating parameters, $\boldsymbol{\theta}^{(j)}$. These steps are repeated until a set number of points, N, have been accepted (Beaumont et al., 2002; Beaumont, 2010).

The main drawback to this rejection-ABC algorithm is that it is inefficient.

Since parameter values are sampled from the prior, we ignore any additional information about the parameters that we could have learned from the data. This means that most of the samples we draw from the prior will result in data sets that are not similar to the observed data. One way to overcome this inefficiency is through sequential Monte Carlo (SMC). This gives rise to the SMC-ABC algorithm, which is characterized by weighted resampling of the points we have already drawn and by iteratively reducing the tolerance, ϵ. As detailed by Toni et al. (2009), to fit a model using SMC-ABC, $\left\{\boldsymbol{\theta}^{(1)}, ..., \boldsymbol{\theta}^{(N^*)}\right\}$ are sampled from the prior distribution and then propogated through a series of intermediate distributions $\pi(\boldsymbol{\theta}|\rho(y, y^*) \leq \epsilon_k)$ for $k = 1, ..., K^* - 1$. The algorithm ceases when the sample from the intermediate distribution represents a sample from the target distribution $\pi(\theta|\rho(y, y^*) \leq \epsilon_{K^*})$ (Toni et al., 2009).

If we wish to compare two competing ABC models, Bayes factors can easily be estimated. We can also use the logic behind Bayes factor calculations for ABC models to develop an estimation method for MCMC models. The simplest method for calculating Bayes factors relies on rejection-ABC, which we have noted is inefficient (Beaumont et al., 2002). Therefore, we do not want to sample from the prior for our MCMC models. We might also consider an algorithm like rejection-ABC, where we sample $\boldsymbol{\theta}^{(j)}$ from the posterior distribution, $\pi(\boldsymbol{\theta}|\mathbf{Y})$, to generate $\mathbf{Y}^{(j)}$, rather than from the prior. While simulating data from the posterior MCMC estimates is likely to be more efficient, it is also likely to be biased. To improve efficiency and overcome bias, we can borrow from the logic behind SMC-ABC and use importance sampling

(Beaumont, 2010). Specifically, the draws from the set of basis parameters, $\boldsymbol{\theta}^{(j)}$, from the posterior, will be weighted to correct for the use of a biased distribution to ensure that the Bayes factor will be unbiased.

To this end, we propose the following algorithm for estimating $P(\mathbf{Y}|\mathcal{M}_i)$. Note that this procedure assumes that the model priors, $P(\mathcal{M}_1)$ and $P(\mathcal{M}_2)$ are equal. For each model and for $j = 1, ..., K$,

1. Sample $\boldsymbol{\theta}^{(j)}$ from the set of MCMC samples and generate posterior predictive data sets $\mathbf{Y}^{(j)}$.
2. For MCMC samples $\boldsymbol{\theta}^{(j)}$, find an approximating density, $G(\cdot)$ (with appropriate support) and evaluate it for the samples, i.e. $G(\boldsymbol{\theta}^{(j)})$. This provides an empirical posterior marginal density estimate.
3. Compute importance weights $w_j \propto \pi(\boldsymbol{\theta}^{(j)})/G(\boldsymbol{\theta}^{(j)})$ and normalize them (such that $\sum_{j=1}^{K} w_j = 1$).
4. Calculate the sum $\sum_{j=1}^{K} \mathbf{1}_{\left\{\rho(\mathbf{Y}^{(j)}-\mathbf{Y})<\epsilon\right\}} w_j$, where $\rho(\mathbf{Y}^{(j)} - \mathbf{Y})$ is a measure of the distance between $\mathbf{Y}^{(j)}$ and $\mathbf{Y}$, the observed data.

There is no generally agreed-upon, best distance metric to use in this scenario. In this dissertation, we present results using the distance function, $\rho(\cdot)$, to be the Euclidean distance function $\sum_{c\in\{\mathcal{S},\mathcal{I},\mathcal{R}\}} \sum_{t=1}^{T} \left(y_{ct}^{(j)} - y_{ct}\right)^2$, which is a natural choice. Ozanne et al. (2019) do some work to assess the sensitivity to this choice. Note, however, an appropriate distance metric might be application-specific.

We now have to discuss how to choose the tolerance, ϵ, from Step 4 of this algorithm. In a fully ABC approach, this tolerance is already set. In rejection-ABC,

it is prespecified, for example, and in SMC-ABC, we use the smallest tolerance, ϵ_{K^*}, that was used to fit the model. Since the model fitting process is inherently different for models fit using MCMC, however, it does not depend on some ϵ, so we need to use a different quantity to inform our choice of tolerance.

When we use the Metropolis-Hastings algorithm for Markov chain simulation, we accept proposals for each parameter with probabilities that can be tuned such that we will adequately explore the parameter space; this is the well-known acceptance ratio (AR). We propose using these acceptance ratios to inform our choice of ϵ when the candidate models of interest have been fit using MCMC rather than ABC. Let $\{AR_{i1}, ...AR_{im}\}$ be the set of acceptance ratios from Model i ($i = 1, 2$), and let $\widehat{AR}_i$ be the median acceptance ratio for Model i. Choosing the median allows us to minimize the effects on tolerance choice from models fit using Metropolis-within-Gibbs (the acceptance ratio for a Gibbs sampler is 1). We wish to choose the strictest tolerance possible, so we use the following algorithm to determine ϵ.

1. Choose ϵ_i such that $\sum_{j=1}^{K} \mathbf{1}_{\left\{\rho(\mathbf{Y}_i^{(j)} - \mathbf{Y}) < \epsilon_i\right\}}/K = \min\left\{\widehat{AR}_1, \widehat{AR}_2\right\}$. Note that ϵ_1 and ϵ_2 very likely will be different.

2. From $\{\epsilon_1, \epsilon_2\}$ obtained in Step 1, choose $\epsilon = \min\{\epsilon_1, \epsilon_2\}$.

3. If for this choice of ϵ, $\sum_{j=1}^{K} \mathbf{1}_{\left\{\rho(\mathbf{Y}_i^{(j)} - \mathbf{Y}) < \epsilon\right\}}/K < 0.001$ for one of the models, choose a tolerance for that model such that $\sum_{j=1}^{K} \mathbf{1}_{\left\{\rho(\mathbf{Y}_i^{(j)} - \mathbf{Y}) < \epsilon\right\}}/K$ is as close to 0.001 as possible. While this threshhold is somewhat arbitrary, we set it to provide guidance for choosing ϵ when the acceptance of posterior predictive data sets is low for one model.

In the following sections, we employ this algorithm for several model comparisons. First, in Section 5.3, we formally investigate the utility of employing a compartmental model versus a multinomial logistic regression model to estimate infection state proportions in a population. These model comparisons are performed both in the context of simulation studies and real VL data. Then, in Sections 5.4-5.5, we assess the importance of accounting for separate asymptomatic and symptomatic infection categories versus a single infectious category and for including vertical transmission in the presence of vector transmission, respectively.

5.3 Model Comparisons: Multinomial Logistic Regression versus Compartmental Modeling

5.3.1 Model Comparison Objective

In Chapter 3, two models were discussed to estimate population *L. infantum* infection state proportions for humans in Brazil. The multinomial logistic regression model is familiar to infectious disease epidemiologists and is easily implemented through existing `R` packages, but it does not naturally accommodate known infection dynamics. In contrast, the compartmental model is less familiar and more involved to implement, but it accommodates infection dynamic information and is easier to interpret. The objective of this section is to formally assess whether the extra complexity of the compartmental model is needed to estimate population infection state proportions.

5.3.2 Model Comparison Setup

First, we will compare the estimation performance for the multinomial logistic regression and SIR models in the context of simulations. The simulated data sets were generated according to the two models described in Section 3.6.2. The first likelihood was generated using a multinomial model where the proportions of individuals in each infection state were assumed to be constant over time. The second likelihood was generated using a multinomial model where the proportions of individuals in each infection state at each time point were determined by the proportions in the population, as described in Equation 3.2. Second, we will evaluate the estimation performance for the multinomial logistic regression and SIR models in the context of real data, as described in Section 3.2.

5.3.3 Model Comparison Results

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ refer to the multinomial logistic regression model and the SIR model, respectively. As discussed in the Section 5.3.2, data were simulated under two likelihoods. The first data generating mechanism considered was a multinomial likelihood with constant infection state proportions across time. In this case, the Bayes factor for $\mathcal{M}_1$ versus $\mathcal{M}_2$ was 1.55 (BF_{12}). According to the interpretation guidelines detailed by Toni et al. (2009) and adapted from Kass and Raftery (1995), this is very weak evidence in favor of $\mathcal{M}_1$, the multinomial logistic regression model. This result supports the assertion that the compartmental model is flexible enough to capture even this unrealistic scenario where the infection state proportions remain

constant over time. The other data generating mechanism considered was a multinomial likelihood with infection state proportions generated through an underlying SIR model. For this scenario, the Bayes factor for $\mathcal{M}_2$ versus $\mathcal{M}_1$ was greater than 150 (BF_{21}). This is very strong evidence in favor of the compartmental model (Kass and Raftery, 1995; Toni et al., 2009). Note that this is a more realistic scenario because we know that *L. infantum* transmission is seasonal, so infection state proportions should change over time.

In the context of real VL data, we obtained a similar result to that obtained for the second simulation. The Bayes factor for $\mathcal{M}_2$ versus $\mathcal{M}_1$ is larger than 150 (B_{21}). This result formally communicates the importance of using a compartmental model to study *L. infantum* transmission dynamics over the more familiar multinomial logistic regression model.

5.4 Model Comparisons: Asymptomatic/Symptomatic vs. Infected

5.4.1 Model Comparison Objective

In Section 5.3, we demonstrated the utility of using a compartmental modeling framework to study *L. infantum* infection. To employ this framework, we need to group individuals into compartments. In this section, we will focus specifically on the importance of designating certain infectious compartments. We can use a diagnostically positive status as a proxy for infection, as in Chapter 2, which gives us a single infectious group. If we have additional physical exam information, we can separate infectious individuals into asymptomatic and symptomatic groups. This information

is not necessarily available, however, particularly for dogs in endemic areas, where physical exams are not always possible. The objective of this section is to assess the impact of grouping asymptomatic and symptomatic individuals into a single infectious compartment because we lack sufficient information to separate them. This can inform the importance of developing other methods to separate asymptomatic and symptomatic individuals that does not require physical exam information and relies only on diagnostic test results. This is discussed as a future research direction in Section 6.2.

5.4.2 Model Comparison Setup

We know that individuals infected with *L. infantum* are either asymptomatic or symptomatic. In this simulation study, we generate an infection process according to the SAYVR model, as described in Section 4.2. As in the simulation studies conducted in the previous chapter, we treat the following transition compartments generated through that process as observed: $\mathbf{Y}^{(\mathcal{AY})^r}$, $\mathbf{R}^{(\mathcal{AR})^r}$, $\mathbf{V}^{(\mathcal{YV})^r}$, and $\mathbf{R}^{(\mathcal{YR})^r}$. Then, we fit these data using two models. The first model is the SAYVR model, and the second model is what we will refer to in this section as the SIVR (**S**usceptible, **I**nfectious, reco**V**ered, **R**emoved) model. This is similar to an SIR model, but with separate recovered and removed categories. The asymptomatic and symptomatic classes are merged into a single infectious class to mimic the case where we only have diagnostic test results and no additional physical exam information.

The resulting data model is as follows. Let $Z_{\mathcal{I}}^{(\mathrm{SF})}$ denote the number of sand

flies that were found to be infected after feeding on infectious (either asymptomatic or symptomatic) dogs. The likelihood is

$$Z_{\mathcal{I}}^{(\mathrm{SF})} \sim \text{Binomial}\left(N_{\mathcal{I}}^{(\mathrm{SF})}, \rho_{\mathcal{I}}\right), \tag{5.1}$$

where $N_{\mathcal{I}}^{(\mathrm{SF})} = 292$ and $\rho_{\mathcal{I}} = 192/292$.

The corresponding process model has the following components. The birth process is the same as in the SAYVR model, but now we allow a portion of the B_t^c births (Equation 4.3) at each time t to be born infectious and call this quantity $B_t^{(\mathcal{I})^c}$. This is just renaming $B_t^{(\mathcal{A})^c}$ (Equation 4.4) from the SAYVR model. Note that $B_t^{(\mathcal{S})^r}$, the number of individual born susceptible at time t, is simply $B_t^r - B_t^{(\mathcal{A})^r}$.

The compartment totals under the SIVR model have the following form:

$$\begin{aligned}
S_t^r &= S_{t-1}^r + B_t^{(\mathcal{S})^r} - I_t^{(\mathcal{SI})^r} - R_t^{(\mathcal{SR})^r};\\
I_t^r &= I_{t-1}^r + B_t^{(\mathcal{I})^r} + I_t^{(\mathcal{SI})^r} - V_t^{(\mathcal{IV})^r} - R_t^{(\mathcal{IR})^r}\\
V_t^r &= V_{t-1}^r + V_t^{(\mathcal{IV})^r} - R_t^{(\mathcal{VR})^r};\\
R_t^r &= R_{t-1}^r + R_t^{(\mathcal{SR})^r} + R_t^{(\mathcal{IR})^r} + R_t^{(\mathcal{VR})^r}.
\end{aligned}$$

As before, the transition compartments are modeled according to a chain binomial structure, as shown in the subsequent equations. For dogs, $V_t^{(\mathcal{IV})^c}$ and $R_t^{(\mathcal{VR})^c}$ are

both zero for all t.

$$
\begin{aligned}
I_t^{(\mathcal{SI})^r} &\sim \text{Binomial}\left(S_{t-1}^r, \pi_t^{(\mathcal{SI})^r}\right); \\
R_t^{(\mathcal{SR})^r} &\sim \text{Binomial}\left(S_{t-1}^r - I_t^{(\mathcal{SI})^r}, \pi_t^{(\mathcal{SR})^r}\right); \\
V_t^{(\mathcal{IV})^r} &\sim \text{Binomial}\left(I_{t-1}^r, \pi^{(\mathcal{IV})^r}\right); \\
R_t^{(\mathcal{IR})^r} &\sim \text{Binomial}\left(I_{t-1}^r - V_t^{(\mathcal{IV})^r}, \pi^{(\mathcal{IR})}\right); \\
R_t^{(\mathcal{VR})^r} &\sim \text{Binomial}\left(V_{t-1}^r, \pi^{(\mathcal{VR})^r}\right).
\end{aligned}
$$

We can calculate the initial compartment size I_0^r by adding A_0^r and Y_0^r from Equations 4.13-4.14. The other initial compartments, S_0^r and V_0^r, remain unchanged.

The associated transition probabilities for these transition compartments for the SIVR model are similar to those for the SAYVR model. The transition probability from $\mathcal{S}$ to $\mathcal{I}$, $\pi_t^{(\mathcal{SI})^r}$ is the same as $\pi_t^{(\mathcal{SA})^r}$ because we assume that individuals who become infected must first become asymptomatic and can later develop symptoms. In other words, susceptible individuals do not transition to symptomatic without first passing through the asymptomatic class. The other transition probabilities are assumed to follow beta distributions, as before:

$$
\begin{aligned}
\pi^{(\mathcal{SR})^r} &\sim \text{Beta}\left(\alpha^{(\mathcal{SR})^r}, \beta^{(\mathcal{SR})^r}\right); \\
\pi^{(\mathcal{IV})^r} &\sim \text{Beta}\left(\alpha^{(\mathcal{IV})^r}, \beta^{(\mathcal{IV})^r}\right); \\
\pi^{(\mathcal{IR})^r} &\sim \text{Beta}\left(\alpha^{(\mathcal{IR})^r}, \beta^{(\mathcal{IR})^r}\right); \\
\pi^{(\mathcal{VR})^r} &\sim \text{Beta}\left(\alpha^{(\mathcal{VR})^r}, \beta^{(\mathcal{VR})^r}\right).
\end{aligned}
$$

As before, for dogs, $\pi^{(\mathcal{IV})^c}$ and $\pi^{(\mathcal{VR})^c}$ are zero due to culling practices.

5.4.3 Model Comparison Results

For these results, we designate $\mathcal{M}_1$ as the SAYVR model and $\mathcal{M}_2$ as the SIVR model. Employing a Euclidean distance metric, we arrive at a Bayes factor for $\mathcal{M}_1$ versus $\mathcal{M}_2$ that was greater than 150 (BF_{12}). This indicates very strong evidence in favor of the SAYVR model (Kass and Raftery, 1995; Toni et al., 2009). This result suggests that it is important to separate infectious individuals into asymptomatic and symptomatic groups. Note that this Bayes factor was calculated for a simulation study, where we made specific assumptions about the available data for fitting these models, so the result is suggestive of the importance of distinguishing between these infectious classes. This study should be repeated for models fit to VL data in the future.

5.5 Model Comparisons: Importance of Including Vertical Transmission

5.5.1 Model Comparison Objective

In endemic areas, *L. infantum* is primarily transmitted through a sand fly vector, and vertical is a secondary transmission mode. Due to the high likelihood of vector contact, however, the rate of vertical transmission is unknown (Mancianti et al., 1988). The objective of this section is to study the importance of accounting for vertical transmission in the presence of widespread vector transmission in the context of a simulation study.

5.5.2 Model Comparison Setup

As in Section 5.4, we simulate an infection process according to the SAYVR model from Section 4.2 and treat the transition compartments $\mathbf{Y}^{(\mathcal{AY})^r}$, $\mathbf{R}^{(\mathcal{AR})^r}$, $\mathbf{V}^{(\mathcal{YV})^r}$, and $\mathbf{R}^{(\mathcal{YR})^r}$ as data. Then, we fit these data using two models. As in the previous section, the first model is the true SAYVR model with vertical transmission. The second model is the SAYVR model without vertical transmission in dogs. This means that we assume that all dogs are born susceptible, i.e. $B_t^c = B_t^{(\mathcal{S})^c}$.

5.5.3 Model Comparison Results

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ correspond to the SAYVR model with and without vertical transmission, respectively. The Bayes factor for $\mathcal{M}_1$ versus $\mathcal{M}_2$ is 3.11 (BF_{12}), which is very weak evidence in favor of the SAYVR model with vertical transmission (Kass and Raftery, 1995; Toni et al., 2009). This result supports the reality that in an endemic area with vector transmission of *L. infantum*, the secondary vertical transmission mode cannot be disentangled from the primary vector transmission mode (Petersen and Barr, 2009). While we do not obtain strong evidence that we should account for vertical transmission in this simulation, we might expect it to be more important to account for this transmission mode in other scenarios. For example, if a control strategy greatly reduced vector transmission, already asymptomatic or symptomatic dams could maintain infection in the population through vertical transmission. Then, if the control strategy failed for some reason and the vector recovered, it would be important to account for this vertically-maintained canine reservoir.

5.6 Discussion

In this chapter, we introduced a novel approach to estimating Bayes factors that was inspired by SMC-ABC and can be used to formally compare Bayesian models that have been fit using MCMC. We specifically discussed how to choose the tolerance, ϵ, in the algorithm using acceptance ratios from Metropolis Hastings. This method is especially useful for estimating Bayes factors when the models of interest have non-normal likelihoods and the parameter spaces are high-dimensional. Then we applied this method to formally compare several models.

In Section 5.3, we applied this method to compare two models for estimating infection state proportions: a multinomial logistic regression model and a SIR model. We demonstrated, both in the context of simulations and historic VL data, that the SIR model is a more reasonable data generating mechanism than the multinomial logistic regression model. This justifies the added complexity and time associated with implementing the compartmental model.

In Sections 5.4-5.5, we assessed the importance of accounting for (1) two infectious groups, asymptomatic and symptomatic, separately and (2) vertical transmission in dogs in the presence of vector transmission. In the context of simulated data, we found strong evidence that is important, when possible, to separate infectious individuals into an asymptomatic and a symptomatic compartment instead of combining them into a single infectious compartment. In contrast, we only found very weak evidence that vertical transmission should be included in the presence of vector transmission when no intervention has been applied to dramatically reduce

the latter. Once these models are fit to VL data, we can gain more insight into the importance of accounting for (1) and (2).

CHAPTER 6
DISCUSSION

In this dissertation, we have developed and applied both individual and population level Bayesian compartmental models to study the infection dynamics of a complex zoonotic disease with multiple transmission modes and host species. We also extended the empirically-adjusted reproductive number for application to an infection with multiple infection sources and infectious states. Finally, we introduced a novel method for estimating Bayes factors for formal model comparison when at least one candidate model has a non-normal likelihood and a high dimensional discrete parameter space. The methods presented in this work can be useful both for studying VL, for which they were developed, and for studying other infectious diseases with similar transmission modes or infectious state properties.

6.1 Modeling Contributions

6.1.1 Individual Level Models

In Chapter 2, we proposed an individual level compartmental model to study infection progression as a result of vertical transmission. Historically, VL compartmental models have been developed for endemic areas, so vertical transmission is included in a birth process and is secondary to vector transmission. Our proposed model is the first compartmental model to explicitly study the *L. infantum* vertical infection process. When we fit this model to canine vertical transmission data, we found evidence that pups born to diagnostically positive dams were more likely to

progress to diagnostically positive status at some point in their lives, as well as earlier in life, than their counterparts born to diagnostically negative dams. This individual level model also may be useful to study the impact of maternal infection status on offspring infection progression for other vertically-transmitted infections that fall on a spectrum, like HIV, hepatitis B, or hepatitis C, for example (Ponziani et al., 2018; Shinazi, 2000).

6.1.2 Population Level Models

In Chapters 3 and 4, we formulated population level compartmental models for VL. First we developed a Bayesian SIR model for studying VL transmission dynamics in humans in Brazil. This was a contribution to stochastic compartmental modeling for this disease, since such modeling has been deterministic historically.

We also extended the compartmental epidemic modeling framework through the SAYVR model. This model is flexible enough to accommodate two infectious classes without restricting one to be more infectious than the other. It also had both a recovered and a removed class. This made it applicable to diseases like VL, where humans with clinical disease can recover with treatment, but which also has a (human) case fatality rate of 8.1% in Brazil (Martins-Melo et al., 2014). Finally, in our formulation of the transmission probability, $\pi^{(\mathcal{SA})}$, we incorporated contributions of two host species, each with two infectious states, and vector feeding behavior. To quantify the latter, we leveraged information about host species infection densities, feeding preferences, and vector infection rates. This formulation takes advantage of

the information we will have from human and canine surveillance, as well as ongoing laboratory-based sand fly studies. It does not require additional field-based sand fly studies.

In addition to model development, we extended the empirically-adjusted reproductive number proposed by Brown et al. (2016) to the case where we have multiple infection sources and infectious states. Under the assumption that infection sources are mutually exclusive, we developed the ISEARN, which allowed us to quantify the contribution of each infectious source to maintaining infection the population. This reproductive number also accounts for the probability that an infectious individual continues to occupy one of two infectious states.

6.1.3 Model Comparison

In addition to the modeling and reproductive number contributions detailed above, we proposed a method for estimating Bayes factors for formal model comparison. We borrowed logic from Bayes factor estimation for models fit using SMC-ABC and adapted it for application to models fit using MCMC. This extended the set of existing Bayes factor estimation methods to accommodate Bayesian models with non-normal likelihoods and high-dimensional discrete parameter spaces that can still be fit using likelihood based methods.

6.2 Future Research Directions

There are a number of possible future research directions that we can explore to extend the work presented in this dissertation. First, the SAYVR model and

ISEARN proposed in Chapter 4 will be applied to VL data that currently is being collected in Brazil. This is provide important insight into the utility of these methods in practice. Another future line of research if to investigate how best to incorporate coinfections or imperfect diagnostic tests into a compartmental model *L. infantum* infection progression. Yet another research path is computational; it is important to fit these models more efficiently.

6.2.1 Coinfections and Infection Progression

Most immunocompetent individuals infected with *L. infantum* do not manifest clinical disease (van Griensven et al., 2014). Comorbid infections in those already infected with *L. infantum* can increase the chances of progressing to clinical disease. For example, humans infected with both HIV and *L. infantum* have higher rates of treatment failure, relapse, and death than their non-HIV infected counterparts (van Griensven et al., 2014). Similarly, dogs with tickborne co-infections progress to clinical disease at higher rates than those that are free of such pathogens (Toepp et al., 2019). Clearly, comorbid infections play an important role in *L. infantum* infection progression, so it is important to assess approaches to include them in future models. In the case of multiple coinfections, one way to do this would be to count up the number, establish meaningful categories based on this number, and include it as a covariate in an individual level compartmental model (Toepp et al., 2019).

6.2.2 Infection Compartment Designation

When we fit models to VL data in this dissertation, we assumed that the diagnostic tests used to group individuals into infection compartments were always accurate. We know, however, that the sensitivity and specificity of diagnostic tests are less than one. We can incorporate this information into a Bayesian hierarchical model, where test sensitivity and specificity inform the probability that an individual occupies a particular infection compartment.

We also based infection state classification on positive/negative test results. While this allowed us to identify infectious (diagnostically positive) individuals, we could not distinguish between asymptomatic and symptomatic based on this information. Recently, a reader has been added to the DPP CVL test that provides continuous values. This test is performed for dogs, so we can use this additional information to establish cutoffs to separate asymptomatic and symptomatic individuals in a canine model. This can be implemented using a classification tree, where the response is clinical status based on a physical exam. This classification rule can be especially helpful in Brazil, where it is not necessarily feasible to perform a physical exam for each dog.

6.2.3 Computation

Compartmental epidemic models tend to have high dimensional target densities and exhibit strong correlation. When these models are fit using MCMC, posterior parameter distribution draws are often highly correlated. Hamiltonian Monte Carlo

can be applied to reduce autocorrelation (Girolami and Calderhead, 2011), but it is difficult to apply for the models in this dissertation due to the discrete nature of the compartment membership counts. Thus, we can adapt Hamiltonian Monte Carlo to apply to discrete compartmental models.

We know that we can approximate a binomial distribution with size N and success probabilitiy p well with a normal distribution when $N \cdot p$ and $N \cdot (1 - p)$ are both sufficiently large. When these conditions are satisfied for the transition compartments in a compartmental model, then we can directly apply Hamiltonian Monte Carlo to sample from the normal approximations to the transition compartment distributions. In cases where these conditions are not met, however, a normal approximation is not valid, so we need to sample from the discrete distribution. We can apply reversible jump MCMC (Green, 1995) to this problem to jump between discrete and continuous parameter spaces according to some normal approximation rule. This approach may lead to more efficient sampling for these high dimensional, highly correlated compartmental models.

6.3 Conclusions

In this work, we developed Bayesian compartmental models, reproductive numbers, and a formal model comparison method to facilitate the study of *L. infantum* infection dynamics in the Americas. We proposed models to study two transmission modes, vertical and vector. We also expanded the compartmental modeling framework with the SAYVR model and developed the ISEARN to complement it.

Application of our individual level compartmental model to vertical transmission data in this dissertation allowed us to demonstrate a relationship between positive diagnostic status in pregnant dogs and faster infection progression in their vertically exposed offspring. We expect that application of our population level SAYVR model to future VL data will lend insight into transmission dynamics and control measures in endemic areas in Brazil.

APPENDIX A
FULL CONDITIONAL DISTRIBUTIONS

A.1 Canine Individual Level Vertical Transmission SIR Model

- $p(\boldsymbol{\beta}|\cdot) \propto f(\mathbf{Z}|\boldsymbol{\beta},\boldsymbol{\theta})p(\boldsymbol{\beta})$
- $p(\boldsymbol{\theta}|\cdot) \propto f(\mathbf{Z}|\boldsymbol{\theta},\boldsymbol{\beta})p(\boldsymbol{\theta})$
- $p(\boldsymbol{\xi}|\cdot) \propto f(\mathbf{Z}|\boldsymbol{\xi})p(\boldsymbol{\xi})$

A.2 Human Population Level Vector Transmission SIR Model

- $p\left(S_0|\cdot\right) \propto f\left(\mathbf{y}|\mathbf{S},\mathbf{I},\mathbf{R},\boldsymbol{\gamma}\right) p\left(\mathbf{I}^{(\mathcal{SI})}|\mathbf{S},\boldsymbol{\beta}\right) p\left(S_0|\pi^{(S_0)}\right)$
- $p\left(I_0|\cdot\right) \propto f\left(\mathbf{y}|\mathbf{S},\mathbf{I},\mathbf{R},\boldsymbol{\gamma}\right) p\left(\mathbf{R}^{(\mathcal{IR})}|\mathbf{I},\pi^{(\mathcal{IR})}\right) p(I_0|\pi^{(I_0)})$
- $p\left(\boldsymbol{\beta}|\cdot\right) \propto f\left(\mathbf{y}|\mathbf{S},\mathbf{I},\mathbf{R},\boldsymbol{\gamma}\right) p\left(\mathbf{I}^{(\mathcal{SI})}|\mathbf{S},\boldsymbol{\beta}\right) p\left(\boldsymbol{\beta}\right)$
- $p\left(\boldsymbol{\gamma}|\cdot\right) \propto f\left(\mathbf{y}|\mathbf{S},\mathbf{I},\mathbf{R},\boldsymbol{\gamma}\right) p\left(\boldsymbol{\gamma}\right)$
- $p\left(\mathbf{I}^{(\mathcal{SI})}|\cdot\right) \propto f\left(\mathbf{y}|\mathbf{S},\mathbf{I},\mathbf{R},\boldsymbol{\gamma}\right) p\left(\mathbf{I}^{(\mathcal{SI})}|\mathbf{S},\boldsymbol{\beta}\right)$
- $p\left(\mathbf{R}^{(\mathcal{IR})}|\cdot\right) \propto f\left(\mathbf{y}|\mathbf{S},\mathbf{I},\mathbf{R},\boldsymbol{\gamma}\right) p\left(\mathbf{R}^{(\mathcal{IR})}|\mathbf{I},\pi^{(\mathcal{IR})}\right)$
- $p\left(\pi^{(\mathcal{IR})}|\cdot\right) \propto f\left(\mathbf{R}^{(\mathcal{IR})}|\mathbf{I},\pi^{(\mathcal{IR})}\right) p\left(\pi^{(\mathcal{IR})}\right)$

  Note: $p\left(\pi^{(\mathcal{IR})}|\cdot\right) \propto \left[\pi^{(\mathcal{IR})}\right]^{\sum_{t\in\mathcal{T}} R_t^{(\mathcal{IR})}+\alpha^{(\mathcal{IR})}-1} \left[1-\pi^{(\mathcal{IR})}\right]^{\sum_{t\in\mathcal{T}}\left(I_{t-1}-R_t^{(\mathcal{IR})}\right)+\beta^{(\mathcal{IR})}-1}$

A.3 Three Species Population Level Vector/Vertical Transmission SAYVR Model

- Birth Processes:

$$p\left(\mathbf{B}^r|\cdot\right) \propto p\left(\mathbf{B}^r|\mathbf{S}^r,\mathbf{A}^r,\mathbf{Y}^r,\phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r,\mathbf{S}^r,\mathbf{A}^r,\mathbf{Y}^r,\gamma^r\right)^{\mathbb{1}_{\{r=c\}}}$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c,\mathbf{S}^h,\mathbf{A}^c,\mathbf{A}^h,\mathbf{Y}^c,\mathbf{Y}^h,\mathbf{V}^h,\boldsymbol{\beta},Z_{\mathcal{A}}^{(\mathrm{SF})},Z_{\mathcal{Y}}^{(\mathrm{SF})},N_{\mathcal{A}}^{(\mathrm{SF})},N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{R}^{(\mathcal{SR})^r}|\mathbf{S}^r,\mathbf{A}^{(\mathcal{SA})^r},\pi^{(\mathcal{SR})^r}\right)$$
$$\times \left[p\left(\mathbf{Y}^{(\mathcal{AY})^r}|\mathbf{A}^r,\pi^{(\mathcal{AY})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{AR})^r}|\mathbf{A}^r,\mathbf{Y}^{(\mathcal{AY})^r},\pi^{(\mathcal{AY})^r}\right)\right]^{\mathbb{1}_{\{r=c\}}}$$

$$p\left(\mathbf{B}\left(\mathcal{A}\right)^c|\cdot\right) \propto p\left(\mathbf{B}(\mathcal{A})^c|\mathbf{B}^c,\mathbf{S}^c,\mathbf{A}^c,\mathbf{Y}^c,\gamma^c\right)$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c,\mathbf{S}^h,\mathbf{A}^c,\mathbf{A}^h,\mathbf{Y}^c,\mathbf{Y}^h,\mathbf{V}^h,\boldsymbol{\beta},Z_{\mathcal{A}}^{(\mathrm{SF})},Z_{\mathcal{Y}}^{(\mathrm{SF})},N_{\mathcal{A}}^{(\mathrm{SF})},N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{Y}^{(\mathcal{AY})^c}|\mathbf{A}^c,\pi^{(\mathcal{AY})^c}\right) \times p\left(\mathbf{R}^{(\mathcal{AR})^c}|\mathbf{A}^c,\mathbf{Y}^{(\mathcal{AY})^c},\pi^{(\mathcal{AY})^c}\right)$$

- Initial Compartments:

$$p\left(S_0^r|\cdot\right) \propto p\left(S_0^r|N_0^r,\pi^{(S_0)^r}\right) \times p\left(\mathbf{B}^r|\mathbf{S}^r,\mathbf{A}^r,\mathbf{Y}^r,\mathbf{V}^r,\phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r,\mathbf{S}^r,\mathbf{A}^r,\mathbf{Y}^r,\gamma^r\right)^{\mathbb{1}_{\{r=c\}}}$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c,\mathbf{S}^h,\mathbf{A}^c,\mathbf{A}^h,\mathbf{Y}^c,\mathbf{Y}^h,\mathbf{V}^h,\boldsymbol{\beta},Z_{\mathcal{A}}^{(\mathrm{SF})},Z_{\mathcal{Y}}^{(\mathrm{SF})},N_{\mathcal{A}}^{(\mathrm{SF})},N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{R}^{(\mathcal{SR})^r}|\mathbf{S}^r,\mathbf{A}^{(\mathcal{SA})^r},\pi^{(\mathcal{SR})^r}\right)$$

$$p\left(A_0^r|\cdot\right) \propto p\left(A_0^r|N_0^r,S_0^r,\pi^{(A_0)^r}\right)$$
$$\times p\left(\mathbf{B}^r|\mathbf{S}^r,\mathbf{A}^r,\mathbf{Y}^r,\mathbf{V}^r,\phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r,\mathbf{S}^r,\mathbf{A}^r,\mathbf{Y}^r,\gamma^r\right)^{\mathbb{1}_{\{r=c\}}}$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c,\mathbf{S}^h,\mathbf{A}^c,\mathbf{A}^h,\mathbf{Y}^c,\mathbf{Y}^h,\mathbf{V}^h,\boldsymbol{\beta},Z_{\mathcal{A}}^{(\mathrm{SF})},Z_{\mathcal{Y}}^{(\mathrm{SF})},N_{\mathcal{A}}^{(\mathrm{SF})},N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{Y}^{(\mathcal{AY})^r}|\mathbf{A}^r,\pi^{(\mathcal{AY})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{AR})^r}|\mathbf{A}^r,\mathbf{Y}^{(\mathcal{AY})^r},\pi^{(\mathcal{AR})^r}\right)$$

$$
\begin{aligned}
p\left(Y_0^r|\cdot\right) \propto\ & p\left(Y_0^r|N_0^r, S_0^r, A_0^r, \pi^{(Y_0)^r}\right) \\
& \times p\left(\mathbf{B}^r|\mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \mathbf{V}^r, \phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r, \mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \gamma^r\right)^{\mathbb{1}\{r=c\}} \\
& \times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right) \\
& \times \left[p\left(\mathbf{V}^{(\mathcal{YV})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YV})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \mathbf{V}^{(\mathcal{YV})^r}, \pi^{(\mathcal{YR})^r}\right)\right]^{\mathbb{1}\{r=h\}} \\
& \times p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YR})^r}\right)^{\mathbb{1}\{r=c\}} \\
& \times p\left(\mathbf{R}^{(\mathcal{SR})^r}|\mathbf{S}^r, \mathbf{A}^{(\mathcal{SA})^r}, \pi^{(\mathcal{SR})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{AR})^r}|\mathbf{Y}^r, \pi^{(\mathcal{AR})^r}\right)
\end{aligned}
$$

$$
\begin{aligned}
p\left(V_0^h|\cdot\right) \propto\ & p\left(V_0^h|N_0^h, A_0^h, Y_0^h, S_0^h, \pi^{(V_0)^h}\right) \\
& \times p\left(\mathbf{B}^r|\mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \mathbf{V}^r, \phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r, \mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \gamma^r\right)^{\mathbb{1}\{r=c\}} \\
& \times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right) \\
& \times \left[p\left(\mathbf{V}^{(\mathcal{YV})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YV})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \mathbf{V}^{(\mathcal{YV})^r}, \pi^{(\mathcal{YR})^r}\right)\right]^{\mathbb{1}\{r=h\}} \\
& \times p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YR})^r}\right)^{\mathbb{1}\{r=c\}} \\
& \times p\left(\mathbf{R}^{(\mathcal{SR})^r}|\mathbf{S}^r, \mathbf{A}^{(\mathcal{SA})^r}, \pi^{(\mathcal{SR})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{AR})^r}|\mathbf{Y}^r, \pi^{(\mathcal{AR})^r}\right) \\
& \times p\left(\mathbf{V}^{(\mathcal{YV})^h}|\mathbf{Y}^h, \pi^{(\mathbf{YV})^h}\right) \times p\left(\mathbf{R}^{(\mathcal{VR})^h}|\mathbf{V}^h, \pi^{(\mathcal{VR})^h}\right)
\end{aligned}
$$

- Transition Compartments:

$$
\begin{aligned}
p\left(\mathbf{A}^{(\mathcal{SA})^r}|\cdot\right) \propto\ & \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right) \\
& \times p\left(\mathbf{B}^r|\mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \mathbf{V}^r, \phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r, \mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \gamma^r\right)^{\mathbb{1}\{r=c\}} \\
& \times p\left(\mathbf{R}^{(\mathcal{SR})^r}|\mathbf{S}^r, \mathbf{A}^{(\mathcal{SA})^r}, \pi^{(\mathcal{SR})^r}\right) \\
& \times p\left(\mathbf{Y}^{(\mathcal{AY})^r}|\mathbf{A}^r, \pi^{(\mathcal{AY})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{AR})^r}|\mathbf{A}^r, \mathbf{Y}^{(\mathcal{AY})^r}, \pi^{(\mathcal{AR})^r}\right)
\end{aligned}
$$

$$p\left(\mathbf{R}^{(\mathcal{SR})^r}|\cdot\right) \propto p\left(\mathbf{R}^{(\mathcal{SR})^r}|\mathbf{S}^r, \mathbf{A}^{(\mathcal{SA})^r}, \pi^{(\mathcal{SR})^r}\right)$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{B}^r|\mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \mathbf{V}^r, \phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r, \mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \gamma^r\right)^{\mathbb{1}_{\{r=c\}}}$$

$$p\left(\mathbf{Y}^{(\mathcal{AY})^r}|\cdot\right) \propto p\left(\mathbf{Y}^{(\mathcal{AY})^r}|\mathbf{A}^r, \pi^{(\mathcal{AY})^r}\right)$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{B}^r|\mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \mathbf{V}^r, \phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r, \mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \gamma^r\right)^{\mathbb{1}_{\{r=c\}}}$$
$$\times \left[p\left(\mathbf{V}^{(\mathcal{YV})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YV})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \mathbf{V}^{(\mathcal{YV})^r}, \pi^{(\mathcal{YR})^r}\right)\right]^{\mathbb{1}_{\{r=h\}}}$$
$$\times p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YR})^r}\right)^{\mathbb{1}_{\{r=c\}}}$$

$$p\left(\mathbf{R}^{(\mathcal{AR})^r}|\cdot\right) \propto p\left(\mathbf{R}^{(\mathcal{AR})^r}|\mathbf{A}^r, \mathbf{Y}^{(\mathcal{AY})^r}, \pi^{(\mathcal{AR})^r}\right)$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{B}^r|\mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \mathbf{V}^r, \phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r, \mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \gamma^r\right)^{\mathbb{1}_{\{r=c\}}}$$
$$\times p\left(\mathbf{Y}^{(\mathcal{AY})^r}|\mathbf{A}^r, \pi^{(\mathcal{AY})^r}\right)$$

$$p\left(\mathbf{V}^{(\mathcal{YV})^h}|\cdot\right) \propto p\left(\mathbf{V}^{(\mathcal{YV})^h}|\mathbf{Y}^h, \pi^{(\mathcal{YV})^h}\right)$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{B}^h|\mathbf{S}^h, \mathbf{A}^h, \mathbf{Y}^h, \mathbf{V}^h, \phi^h\right)$$
$$\times p\left(\mathbf{R}^{(\mathcal{YR})^h}|\mathbf{Y}^h, \mathbf{V}^{(\mathcal{YV})^h}, \pi^{(\mathcal{YR})^h}\right)$$
$$\times p\left(\mathbf{R}^{(\mathcal{VR})^h}|\mathbf{V}^h, \pi^{(\mathcal{VR})^h}\right)$$

$$p\left(\mathbf{R}^{(\mathcal{YR})^r}|\cdot\right) \propto \left[p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \mathbf{V}^{(\mathcal{YV})^r}, \pi^{(\mathcal{YR})^r}\right) \times p\left(\mathbf{V}^{(\mathcal{YV})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YV})^r}\right)\right]^{\mathbb{1}\{r=h\}}$$
$$\times p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YR})^r}\right)^{\mathbb{1}\{r=c\}}$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{B}^r|\mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \mathbf{V}^r, \phi^r\right) \times p\left(\mathbf{B}(\mathcal{A})^r|\mathbf{B}^r, \mathbf{S}^r, \mathbf{A}^r, \mathbf{Y}^r, \gamma^r\right)^{\mathbb{1}\{r=c\}}$$
$$\times p\left(\mathbf{V}^{(\mathcal{YV})^r}|\mathbf{Y}^r, \pi^{(\mathcal{YV})^r}\right)^{\mathbb{1}\{r=h\}}$$

$$p\left(\mathbf{R}^{(\mathcal{VR})^h}|\cdot\right) \propto p\left(\mathbf{R}^{(\mathcal{VR})^h}|\mathbf{V}^h, \pi^{(\mathcal{VR})^h}\right)$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$
$$\times p\left(\mathbf{B}^h|\mathbf{S}^h, \mathbf{A}^h, \mathbf{Y}^h, \mathbf{V}^h, \phi^h\right)$$

- Transition Probabilities:

$$p\left(\boldsymbol{\beta}|\cdot\right) \propto p(\boldsymbol{\beta})$$
$$\times \prod_{r\in\{c,h\}} p\left(\mathbf{A}^{(\mathcal{SA})^r}|\mathbf{S}^c, \mathbf{S}^h, \mathbf{A}^c, \mathbf{A}^h, \mathbf{Y}^c, \mathbf{Y}^h, \mathbf{V}^h, \boldsymbol{\beta}, Z_{\mathcal{A}}^{(\mathrm{SF})}, Z_{\mathcal{Y}}^{(\mathrm{SF})}, N_{\mathcal{A}}^{(\mathrm{SF})}, N_{\mathcal{Y}}^{(\mathrm{SF})}\right)$$

$$p\left(\pi^{(\mathcal{SR})^r}|\cdot\right) \propto p\left(\pi^{(\mathcal{SR})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{SR})^r}|\mathbf{S}^r, \mathbf{A}^{(\mathcal{SA})^r}, \pi^{(\mathcal{SR})^r}\right)$$
$$\propto \left[\pi^{(\mathcal{SR})^r}\right]^{\alpha^{(\mathcal{SR})^r}-1} \left[1-\pi^{(\mathcal{SR})^r}\right]^{\beta^{(\mathcal{SR})^r}-1} \prod_{t\in\mathcal{T}} \left[\pi^{(\mathcal{SR})^r}\right]^{R_t^{(\mathcal{SR})^r}} \left[1-\pi^{(\mathcal{SR})^r}\right]^{S_{t-1}^r - A_t^{(\mathcal{SA})^r}}$$
$$\propto \left[\pi^{(\mathcal{SR})^r}\right]^{\sum_{t\in\mathcal{T}} R_t^{(\mathcal{SR})^r} + \alpha^{(\mathcal{SR})^r}-1} \left[1-\pi^{(\mathcal{SR})^r}\right]^{\sum_{t\in\mathcal{T}}\left(S_{t-1}^r - A_t^{(\mathcal{SA})^r}\right) + \beta^{(\mathcal{SR})^r}-1}$$

$$p\left(\pi^{(\mathcal{AY})^r}|\cdot\right) \propto p\left(\pi^{(\mathcal{AY})^r}\right) \times p\left(\mathbf{Y}^{(\mathcal{AY})^r}|\mathbf{A}^r, \pi^{(\mathcal{AY})^r}\right)$$
$$\propto \left[\pi^{(\mathcal{AY})^r}\right]^{\alpha^{(\mathcal{AY})^r}-1} \left[1-\pi^{(\mathcal{AY})^r}\right]^{\beta^{(\mathcal{AY})^r}-1} \prod_{t\in\mathcal{T}} \left[\pi^{(\mathcal{AY})^r}\right]^{Y_t^{(\mathcal{AY})^r}} \left[1-\pi^{(\mathcal{AY})^r}\right]^{A_{t-1}^r}$$
$$\propto \left[\pi^{(\mathcal{AY})^r}\right]^{\sum_{t\in\mathcal{T}} Y_t^{(\mathcal{AY})^r} + \alpha^{(\mathcal{AY})^r}-1} \left[1-\pi^{(\mathcal{AY})^r}\right]^{\sum_{t\in\mathcal{T}} A_{t-1}^r + \beta^{(\mathcal{AY})^r}-1}$$

$$
\begin{aligned}
p\left(\pi^{(\mathcal{AR})^r}|\cdot\right) &\propto p\left(\pi^{(\mathcal{AR})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{AR})^r}|\mathbf{A}^r, \mathbf{Y}^{(\mathcal{AY})^r}, \pi^{(\mathcal{AR})^r}\right) \\
&\propto \left[\pi^{(\mathcal{AR})^r}\right]^{\alpha^{(\mathcal{AR})^r}-1} \left[1-\pi^{(\mathcal{AR})^r}\right]^{\beta^{(\mathcal{AR})^r}-1} \prod_{t\in\mathcal{T}} \left[\pi^{(\mathcal{AR})^r}\right]^{R_t^{(\mathcal{AR})^r}} \left[1-\pi^{(\mathcal{AR})^r}\right]^{A_{t-1}^r - Y_t^{(\mathcal{AY})^r}} \\
&\propto \left[\pi^{(\mathcal{AR})^r}\right]^{\sum_{t\in\mathcal{T}} R_t^{(\mathcal{AR})^r}+\alpha^{(\mathcal{AR})^r}-1} \left[1-\pi^{(\mathcal{AR})^r}\right]^{\sum_{t\in\mathcal{T}}\left(A_{t-1}^r - Y_t^{(\mathcal{AY})^r}\right)+\beta^{(\mathcal{AR})^r}-1}
\end{aligned}
$$

$$
\begin{aligned}
p\left(\pi^{(\mathcal{YV})^h}|\cdot\right) &\propto p\left(\pi^{(\mathcal{YV})^h}\right) \times p\left(\mathbf{V}^{(\mathcal{YV})^h}|\mathbf{Y}^h, \pi^{(\mathcal{YV})^h}\right) \\
&\propto \left[\pi^{(\mathcal{YV})^h}\right]^{\alpha^{(\mathcal{YV})^h}-1} \left[1-\pi^{(\mathcal{YV})^h}\right]^{\beta^{(\mathcal{YV})^h}-1} \prod_{t\in\mathcal{T}} \left[\pi^{(\mathcal{YV})^h}\right]^{V_t^{(\mathcal{YV})^h}} \left[1-\pi^{(\mathcal{YV})^h}\right]^{Y_{t-1}^h} \\
&\propto \left[\pi^{(\mathcal{YV})^h}\right]^{\sum_{t\in\mathcal{T}} V_t^{(\mathcal{YV})^h}+\alpha^{(\mathcal{YV})^h}-1} \left[1-\pi^{(\mathcal{YV})^h}\right]^{\sum_{t\in\mathcal{T}} Y_{t-1}^h+\beta^{(\mathcal{YV})^h}-1}
\end{aligned}
$$

$$
\begin{aligned}
p\left(\pi^{(\mathcal{YR})^r}|\cdot\right) &\propto p\left(\pi^{(\mathcal{YR})^r}\right) \times p\left(\mathbf{R}^{(\mathcal{YR})^r}|\mathbf{Y}^r, \mathbf{V}^{(\mathcal{YV})^r}, \pi^{(\mathcal{YR})^r}\right) \\
&\propto \left[\pi^{(\mathcal{YR})^r}\right]^{\alpha^{(\mathcal{YR})^r}-1} \left[1-\pi^{(\mathcal{YR})^r}\right]^{\beta^{(\mathcal{YR})^r}-1} \prod_{t\in\mathcal{T}} \left[\pi^{(\mathcal{YR})^r}\right]^{R_t^{(\mathcal{YR})^r}} \left[1-\pi^{(\mathcal{YR})^r}\right]^{Y_{t-1}^r - V_t^{(\mathcal{YV})^r}} \\
&\propto \left[\pi^{(\mathcal{YR})^r}\right]^{\sum_{t\in\mathcal{T}} R_t^{(\mathcal{YR})^r}+\alpha^{(\mathcal{YR})^r}-1} \left[1-\pi^{(\mathcal{YR})^r}\right]^{\sum_{t\in\mathcal{T}}\left(Y_{t-1}^r - V_t^{(\mathcal{YV})^r}\right)+\beta^{(\mathcal{YR})^r}-1}
\end{aligned}
$$

$$
\begin{aligned}
p\left(\pi^{(\mathcal{VR})^h}|\cdot\right) &\propto p\left(\pi^{(\mathcal{VR})^h}\right) \times p\left(\mathbf{R}^{(\mathcal{VR})^h}|\mathbf{V}^h, \pi^{(\mathcal{VR})^h}\right) \\
&\propto \left[\pi^{(\mathcal{VR})^h}\right]^{\alpha^{(\mathcal{VR})^h}-1} \left[1-\pi^{(\mathcal{VR})^h}\right]^{\beta^{(\mathcal{VR})^h}-1} \prod_{t\in\mathcal{T}} \left[\pi^{(\mathcal{VR})^h}\right]^{R_t^{(\mathcal{VR})^h}} \left[1-\pi^{(\mathcal{VR})^h}\right]^{V_{t-1}^h} \\
&\propto \left[\pi^{(\mathcal{VR})^h}\right]^{\sum_{t\in\mathcal{T}} R_t^{(\mathcal{VR})^h}+\alpha^{(\mathcal{VR})^h}-1} \left[1-\pi^{(\mathcal{VR})^h}\right]^{\sum_{t\in\mathcal{T}} V_{t-1}^h+\beta^{(\mathcal{VR})^h}-1}
\end{aligned}
$$

APPENDIX B
MARKOV CHAIN MONTE CARLO ALGORITHMS

B.1 Canine Individual Level Vertical Transmission SIR Model

1. Initialize all parameters.
2. Update $\boldsymbol{\beta}$ using Metropolis-within-Gibbs.
3. Update $\boldsymbol{\theta}$ using Metropolis-within-Gibbs.
4. Update $\boldsymbol{\xi}$ using Metropolis-within-Gibbs.

B.2 Human Population Level Vector Transmission SIR Model

1. Initialize all values.
2. Update initial values S_0, I_0, and R_0.
3. Update $\boldsymbol{\beta}$ using Metropolis Hastings via a Gaussian random walk.
4. Calculate the deterministic vector $\boldsymbol{\pi}^{(\mathcal{SI})}$.
5. Update $\boldsymbol{\gamma}$ using Metropolis Hastings via a Gaussian random walk.
6. Update $\mathbf{I}^{(\mathcal{SI})}$ using Metropolis Hastings.
7. Calculate the deterministic vectors $\mathbf{S}, \mathbf{I}, \mathbf{R}$.
8. Calculate the deterministic vector $\boldsymbol{\pi}^{(\mathcal{SI})}$.
9. Update $\mathbf{R}^{(\mathcal{IR})}$ using Metropolis Hastings.
10. Calculate the deterministic vectors $\mathbf{I}, \mathbf{R}$.
11. Update $\pi^{(\mathcal{IR})}$ using Gibbs.

B.3 Three Species Population Level Vector/Vertical Transmission SAYVR Model

1. Initialize all values.
2. Update total births, $\mathbf{B}^c$ and $\mathbf{B}^h$, using Metropolis Hastings.
3. Update canine asymptomatic births, $\mathbf{B}(\mathcal{A})^c$, using Metropolis Hastings.
4. Update initial values, S_0^c, S_0^h, A_0^c, A_0^h, Y_0^c, Y_0^h, and V_0^h.
5. Update $\boldsymbol{\beta}$ using Metropolis Hastings via a Gaussian random walk.
6. Calculate the deterministic vector $\boldsymbol{\pi}^{(\mathcal{SA})}$.
7. Update $\mathbf{A}^{(\mathcal{SA})^c}$ and $\mathbf{A}^{(\mathcal{SA})^h}$ using Metropolis Hastings.
8. Update $\mathbf{R}^{(\mathcal{SR})^c}$ and $\mathbf{R}^{(\mathcal{SR})^h}$ using Metropolis Hastings.
9. Update $\mathbf{Y}^{(\mathcal{AY})^c}$ and $\mathbf{Y}^{(\mathcal{AY})^h}$ using Metropolis Hastings.
10. Update $\mathbf{R}^{(\mathcal{AR})^c}$ and $\mathbf{R}^{(\mathcal{AR})^h}$ using Metropolis Hastings.
11. Update $\mathbf{V}^{(\mathcal{YV})^h}$ using Metropolis Hastings.
12. Update $\mathbf{R}^{(\mathcal{YR})^c}$ and $\mathbf{R}^{(\mathcal{YR})^h}$ using Metropolis Hastings.
13. Update $\mathbf{R}^{(\mathcal{VR})^h}$ using Metropolis Hastings.
14. Update $\pi^{(\mathcal{SR})^c}$, $\pi^{(\mathcal{SR})^h}$, $\pi^{(\mathcal{AY})^c}$, $\pi^{(\mathcal{AY})^h}$, $\pi^{(\mathcal{AR})^c}$, $\pi^{(\mathcal{AR})^h}$, $\pi^{(\mathcal{YR})^c}$, $\pi^{(\mathcal{YR})^h}$, $\pi^{(\mathcal{YV})^h}$, and $\pi^{(\mathcal{VR})^h}$ using Gibbs.
15. Calculate deterministic infection compartments $\mathbf{S}^c$, $\mathbf{S}^h$, $\mathbf{A}^c$, $\mathbf{A}^h$, $\mathbf{Y}^c$, $\mathbf{Y}^h$, $\mathbf{V}^h$, $\mathbf{R}^c$, and $\mathbf{R}^h$.

APPENDIX C
SOURCE-SPECIFIC TRANSITION PROBABILITY DERIVATIONS

Define the following:

- $A(t)^r$ = event that individual in r becomes infected at time t due to any infection source; $!A(t)^r =$ not $A(t)^r$
- $A(t, s_h)^r$ = event that individual in r becomes infected at time t due to human infection source; $!A(t, s_h)^r =$ not $A(t, s_h)^r$
- $A(t, s_c)^r$ = event that individual in r becomes infected at time t due to canine infection source; $!A(t, s_c)^r =$ not $A(t, s_c)^r$
- $A(t, s_{Ah})^r$ = event that individual in r becomes infected at time t due to asymptomatic human infection source; $A(t, s_{Ah})^r =$ not $A(t, s_{Ah})^r$
- $A(t, s_{Ac})^r$ = event that individual in r becomes infected at time t due to asymptomatic canine infection source; $A(t, s_{Ac})^r =$ not $A(t, s_{Ac})^r$
- $A(t, s_{Yh})^r$ = event that individual in r becomes infected at time t due to symptomatic human infection source; $A(t, s_{Yh})^r =$ not $A(t, s_{Yh})^r$
- $A(t, s_{Yc})^r$ = event that individual in r becomes infected at time t due to symptomatic canine infection source; $A(t, s_{Yc})^r =$ not $A(t, s_{Yc})^r$

Assumptions:

- $\cap_{b \in \mathcal{B}^*} A(t, s_b) = \emptyset$
- $P\left[\cup_{b \in \mathcal{B}^*} A(t, s_b)\right] = \sum_{b \in \mathcal{B}^*} P\left[A(t, s_b)\right] = P[A(t, s_h)] + P[A(t, s_c)] = P[A(t)]$

Known probabilities:

- $\pi^{(\mathcal{SA})^r}(t) = P[A(t)^r] = 1 - \exp\left\{-\left[\delta_t - (1-\delta_t)\delta_t\right] e^{X_t\beta^r}\right\}$
- $\pi^{(\mathcal{SA})^r}_{h|!c}(t) = P[A(t,s_h)^r|!A(t,s_c)^r] = 1 - \exp\left\{-\left[\delta^h_t - (1-\delta_t)\delta^h_t\right] e^{X_t\beta^r}\right\}$
- $\pi^{(\mathcal{SA})^r}_{c|!h}(t) = P[A(t,s_c)^r|!A(t,s_h)^r] = 1 - \exp\left\{-\left[\delta^c_t - (1-\delta_t)\delta^c_t\right] e^{X_t\beta^r}\right\}$
- $\pi^{(\mathcal{SA})^r}_{Ah|!\{Yh,c\}}(t) = P[A(t,s_{Ah})^r|!\{A(t,s_{Yh})^r, A(t,s_c)^r\}] = 1-\exp\left\{-\left[\delta^{Ah}_t - (1-\delta_t)\delta^{Ah}_t\right] e^{X_t\beta^r}\right\}$
- $\pi^{(\mathcal{SA})^r}_{Yh|!\{Ah,c\}}(t) = P[A(t,s_{Yh})^r|!\{A(t,s_{Ah})^r, A(t,s_c)^r\}] = 1-\exp\left\{-\left[\delta^{Yh}_t - (1-\delta_t)\delta^{Yh}_t\right] e^{X_t\beta^r}\right\}$
- $\pi^{(\mathcal{SA})^r}_{Ac|!\{Yc,h\}}(t) = P[A(t,s_{Ac})^r|!\{A(t,s_{Yc})^r, A(t,s_h)^r\}] = 1-\exp\left\{-\left[\delta^{Ac}_t - (1-\delta_t)\delta^{Ac}_t\right] e^{X_t\beta^r}\right\}$
- $\pi^{(\mathcal{SA})^r}_{Yc|!\{Ac,h\}}(t) = P[A(t,s_{Yc})^r|!\{A(t,s_{Ac})^r, A(t,s_h)^r\}] = 1-\exp\left\{-\left[\delta^{Yc}_t - (1-\delta_t)\delta^{Yc}_t\right] e^{X_t\beta^r}\right\}$

Derivations (assume mutually disjoint events, r implied):

- Derive $P[A(t,s_c)]$:

$$
\begin{aligned}
1 &= P[!A(t,s_h)|A(t,s_c)] \\
&= \frac{P[A(t,s_c)|!A(t,s_h)]P[!A(t,s_h)]}{P[A(t,s_c)]} \\
\Rightarrow P[A(t,s_c)] &= P[A(t,s_c)|!A(t,s_h)]P[!A(t,s_h)] \\
&= P[A(t,s_c)|!A(t,s_h)]\left(1 - P[A(t,s_h)]\right) \\
&= \pi^{(\mathcal{SA})}_{c|!h}(t)\left(1 - P[A(t,s_h)]\right) \\
&= \pi^{(\mathcal{SA})}_{c|!h}(t)\left(1 - \{P[A(t)] - P[A(t,s_c)]\}\right) \\
&= \pi^{(\mathcal{SA})}_{c|!h}(t)\left(1 - \pi^{(\mathcal{SA})}(t) + P[A(t,s_c)]\right) \\
&= \frac{\pi^{(\mathcal{SA})}_{c|!h}(t)\left(1 - \pi^{(\mathcal{SA})}(t)\right)}{1 - \pi^{(\mathcal{SA})}_{c|!h}(t)}
\end{aligned}
$$

- Derive $P[A(t, s_h)]$:

$$
\begin{aligned}
\Rightarrow P[A(t, s_h)] &= \pi^{(\mathcal{SA})}(t) - P[A(t, s_c)] \\
&= \pi^{(\mathcal{SA})}(t) - \frac{\pi^{(\mathcal{SA})}_{c|!h}(t)\left(1 - \pi^{(\mathcal{SA})}(t)\right)}{1 - \pi^{(\mathcal{SA})}_{c|!h}(t)} \\
&= \frac{\pi^{(\mathcal{SA})}(t)\left(1 - \pi^{(\mathcal{SA})}_{c|!h}(t)\right) - \pi^{(\mathcal{SA})}_{c|!h}(t)\left(1 - \pi^{(\mathcal{SA})}(t)\right)}{1 - \pi^{(\mathcal{SA})}_{c|!h}(t)} \\
&= \frac{\pi^{(\mathcal{SA})}(t) - \pi^{(\mathcal{SA})}_{c|!h}(t)}{1 - \pi^{(\mathcal{SA})}_{c|!h}(t)}
\end{aligned}
$$

- Derive $P[A(t, s_{Ah})]$:

$$
\begin{aligned}
1 &= P[!A(t, s_{Yh}) \cap !A(t, s_c) | A(t, s_{Ah})] \\
&= \frac{P[A(t, s_{Ah}) | !A(t, s_{Yh}) \cap !A(t, s_c)] P[!A(t, s_{Yh}) \cap !A(t, s_c)]}{P[A(t, s_{Ah})]} \\
&= \frac{P[A(t, s_{Ah}) | !A(t, s_{Yh}) \cap !A(t, s_c)] P[!A(t, s_{Yh}) | !A(t, s_c)] P[!A(t, s_c)]}{P[A(t, s_{Ah})]} \\
&= \frac{\pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t) P[!A(t, s_{Yh}) | !A(t, s_c)] \left(1 - P[A(t, s_c)]\right)}{P[A(t, s_{Ah})]}
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow P[A(t,s_{Ah})] &= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)])\,P[!A(t,s_{Yh})|!A(t,s_c)] \\
&= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)])\,(1-P[A(t,s_{Yh})|!A(t,s_c)]) \\
&= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)]) \\
&\times (1-\{P[A(t,s_h)|!A(t,s_c)]-P[A(t,s_{Ah})|!A(t,s_c)]\}) \\
&= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)]) \\
&\times \left(1-\pi^{(\mathcal{SA})}_{h|!c}(t)+P[A(t,s_{Ah})|!A(t,s_c)]\right) \\
&= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)]) \\
&\times \left(1-\pi^{(\mathcal{SA})}_{h|!c}(t)+\frac{P[A(t,s_{Ah})\cap !A(t,s_c)]}{P[!A(t,s_c)]}\right) \\
&= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)]) \\
&\times \left(1-\pi^{(\mathcal{SA})}_{h|!c}(t)+\frac{P[A(t,s_{Ah})]-\overbrace{P[A(t,s_{Ah})\cap A(t,s_c)]}^{=0,\text{ disjoint events}}}{P[!A(t,s_c)]}\right) \\
&= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)]) \\
&\times \left(1-\pi^{(\mathcal{SA})}_{h|!c}(t)+\frac{P[A(t,s_{Ah})]}{1-P[A(t,s_c)]}\right) \\
&= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)])\left(1-\pi^{(\mathcal{SA})}_{h|!c}(t)\right) \\
&+ \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)])\left(\frac{P[A(t,s_{Ah})]}{1-P[A(t,s_c)]}\right) \\
&= \pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)])\left(1-\pi^{(\mathcal{SA})}_{h|!c}(t)\right)+\pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)P[A(t,s_{Ah})] \\
\Rightarrow P[A(t,s_{Ah})] &= \frac{\pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)\,(1-P[A(t,s_c)])\left(1-\pi^{(\mathcal{SA})}_{h|!c}(t)\right)}{1-\pi^{(\mathcal{SA})}_{Ah|!\{Yh,c\}}(t)}
\end{aligned}
$$

- Derive $P[A(t,s_{Yh})]$:

$$P[A(t,s_{Yh})] = P[A(t,s_h)] - P[A(t,s_{Ah})]$$

- Derive $P[A(t, s_{Ac})]$:

$$
\begin{aligned}
1 &= P[!A(t, s_{Yc}) \cap !A(t, s_h) | A(t, s_{Ac})] \\
&= \frac{P[A(t, s_{Ac}) | !A(t, s_{Yc}) \cap !A(t, s_h)] P[!A(t, s_{Yc}) \cap !A(t, s_h)]}{P[A(t, s_{Ac})]} \\
&= \frac{P[A(t, s_{Ac}) | !A(t, s_{Yc}) \cap !A(t, s_h)] P[!A(t, s_{Yc}) | !A(t, s_h)] P[!A(t, s_h)]}{P[A(t, s_{Ac})]} \\
&= \frac{\pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}} P[!\{A(t, s_{Yc}) | !A(t, s_h)\}] P[!A(t, s_h)\}]}{P[A(t, s_{Ac})]}
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow P[A(t, s_{Ac})] &= \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}} P[!A(t, s_{Yc}) | !A(t, s_h)] P[!A(t, s_h)] \\
&= \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}} \left(1 - P[A(t, s_h)]\right) \left(1 - P[A(t, s_{Yc}) | !A(t, s_h)]\right) \\
&= \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) \left(1 - P[A(t, s_h)]\right) \\
&\times \left(1 - \{P[A(t, s_c) | !A(t, s_h)] - P[A(t, s_{Ac}) | !A(t, s_h)]\}\right) \\
&= \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) \left(1 - P[A(t, s_h)]\right) \\
&\times \left(1 - \pi^{(\mathcal{SA})}_{c|!h}(t) + \frac{P[A(t, s_{Ac}) \cap !A(t, s_h)]}{P[!A(t, s_h)]}\right) \\
&= \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) \left(1 - P[A(t, s_h)]\right) \\
&\times \left(1 - \pi^{(\mathcal{SA})}_{c|!h}(t) + \frac{P[A(t, s_{Ac})] - \overbrace{P[A(t, s_{Ac}) \cap A(t, s_h)]}^{=0, \text{ disjoint events}}}{1 - P[A(t, s_h)]}\right) \\
&= \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) \left(1 - P[A(t, s_h)]\right) \\
&\times \left(1 - \pi^{(\mathcal{SA})}_{c|!h}(t) + \frac{P[A(t, s_{Ac})]}{1 - P[A(t, s_h)]}\right) \\
&= \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) \left(1 - P[A(t, s_h)]\right) \left(1 - \pi^{(\mathcal{SA})}_{c|!h}(t)\right) \\
&+ \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) \left(1 - P[A(t, s_h)]\right) \left(\frac{P[A(t, s_{Ac})]}{1 - P[A(t, s_h)]}\right) \\
&= \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) \left(1 - P[A(t, s_h)]\right) \left(1 - \pi^{(\mathcal{SA})}_{c|!h}(t)\right) + \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) P[A(t, s_{Ac})] \\
\Rightarrow P[A(t, s_{Ac})] &= \frac{\pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t) \left(1 - P[A(t, s_h)]\right) \left(1 - \pi^{(\mathcal{SA})}_{c|!h}(t)\right)}{1 - \pi^{(\mathcal{SA})}_{Ac|!\{Yc,h\}}(t)}
\end{aligned}
$$

- Derive $P[A(t, s_{Yc})]$:

$$P[A(t, s_{Yc})] = P[A(t, s_c)] - P[A(t, s_{Ac})]$$

REFERENCES

Abu-Raddad, L. and Longini Jr, I. (2008). No HIV stage is dominant in dirving the HIV epidemic in sub-Saharan Africa. *AIDS*, 22:1055–1061.

Agyingi, E., Ross, D., and Bathena, K. (2011). A model of the transmission dynamics of leishmaniasis. *Journal of Biological Systems*, 19:237–250.

Allen, L. J. and van den Driessche, P. (2008). The basic reproduction number in some discrete-time epidemic models. *Journal of Difference Equations and Applications*, 14:1127–1147.

Alvar, J., Veléz, I., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J., den Boer, M., and the WHO Leishmaniasis Control Team (2012). Leishmaniasis worldwide and global estimates of its incidence. *PLoS One*, 7(5):e35671.

Anderson, R. and May, R. (1979). Population biology of infectious diseases: Part i. *Nature*, 280(2):361–367.

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

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*. Springer-Verlag.

Beaumont, M., Cornuet, J., Marin, J., and et al. (2009). Adaptive approximate bayesian computation. *Biometrika*, 96(4):983–990.

Beaumont, M., Zhang, W., and Balding, D. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162:2025–2035.

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41:379–406.

Belo, V. S., Struchiner, C. J., Barbosa, D. S., Nascimento, B. W. L., Horta, M. A. P., da Silva, E. S., and Werneck, G. L. (2014). Risk factors for adverse prognosis and death in American visceral leishmaniasis: a meta-analysis. *PLoS Neglected Tropical Diseases*, 8(7):e2982.

Bonsu, B. K. and Harper, M. B. (2004). Differentiating acute bacterial meningitis from acute viral meningitis among children with cerebrospinal fluid pleosytosis: a multivariable regression model. *The Pediatric Infectious Disease Journal*, 23:511–517.

Braz, R., Nascimento, E., Martins, D., and et al. (2014). Asymptomatic leishmania infection: a new challenge for leishmania control. *Clinical Infectious Diseases*, 58:1424–1429.

Braz, R. F. S., Nascimento, E. T., Martins, D. R. A., Wilson, M. E., Pearson, R. D., Reed, S. G., and Jeronimo, S. M. B. (2002). The sensitivity and specificity of leishmania chagasi recombinant k39 antigen in the diagnosis of American visceral leishmaniasis and in differentiating active from subclinical infection. *The American Journal of Tropical Medicine and Hygiene*, 67(4):344–348.

Brown, G. D., Oleson, J. J., and Porter, A. T. (2016). An empirically adjusted approach to reproductive number estimation for stochastic compartmental models: A case study of two ebola outbreaks. *Biometrics*, 72:335–343.

Busenberg, S. and Cooke, K. (1993). *Vertically Transmitted Diseases. Models and Dynamics, Biomathematics 23*. Springer-Verlag, Berlin.

Cao, J., Wang, Y., Alofi, A., Al-Mazrooei, A., and Elaiw, A. (2015). Global stability of an epidemic model with carrier state in heterogeneous networks. *IMA Journal of Applied Mathematics*, 80:1025–1048.

Celeux, G., Forbes, F., Robert, C., and Titterington, M. (2006). Rejoinder to discussion of deviance information criteria for missing data models. *Bayesian Analysis*, 1:701–706.

Chadwick, D., Arch, B., Wilder-Smith, A., and Paton, N. (2006). Distinguishing dengue fever from other infections on the basis of simple clinical and laboratory features: application of logistic regression analysis. *Journal of Clinical Virology*, 35:147–153.

Chowell, G., Ammon, C. E., Hengartner, N. W., and Hyman, J. M. (2006). Estimation of the reproductive number of the Spanish flu epidemic in Geneva, Switzerland. *Vaccine*, 24(44-46):6747–6750.

Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., and Hyman, J. M. (2004). The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology*, 229:119–126.

Costa, C. H. N., Gomes, R. B. B., Silva, M. R. B., M, G. L., Ramos, P. K. S., Santos, R. S., Shaw, J. J., David, J. R., and Macguire, J. H. (2000). Competence of the human host as a reservoir for Leishmania chagasi. *The Journal of Infectious Diseases*, 182:997–1000.

Costa, C. H. N., Stewart, J. M., Gomes, R. B., Garcez, L. M., Ramos, P. K., Bozza, M., Satoskar, A., Dissanayake, S., Santos, R. S., Silva, M. R., Shaw, J. J., David, J. R., and Maguire, J. H. (2002). Asymptomatic human carriers of Leishmania chagasi. *Journal of Tropical Medicine and Hygiene*, 66(4):334–337.

Courtenay, O., Quinnell, R. J., Garcez, L. M., Shaw, J. J., and Dye, C. (2002). Infectiousness in a cohort of Brazilian dogs: why culling fails to control visceral leishmaniasis in areas of high transmission. *Journal of Infectious Diseases*, 186(9):1314–1320.

da Silva, S. M., Ribeiro, V. M., Ribeiro, R. R., Tafuri, W. L., Melo, M. N., and Michalick, M. S. M. (2009). First report of vertical transmission of leishmania (leishmania) infantum in a naturally infected bitch from Brazil. *Veterinary Parasitology*, 166:159–162.

Deane, L. M. and Deane, M. P. (1955). Preliminary observations on the comparative importance of man, the dog and the fox (lycalopex vetulus) as reservoirs of leishmania donovani, in an endemic area for calazar, Ceara. *O Hospital*, 48:79–98.

Desjeux, P. (2004). Leishmaniasis: current situation and perspectives. *Comparative Immunology, Microbiology and Infectious Diseases*, 27(5):305–318.

Diaz-Albiter, H., Mitford, R., Genta, F. A., Sant'Anna, M. R. V., and Dillon, R. J. (2009). Reactive oxygen species scavenging by catalase is important for female *Lutzomyia longipalpis* fecundity and mortality. *PLoS One*, 6(3):e17486.

Dietz, K. (1976). Infectious diseases under the influence of seasonal fluctuations. *Lect. Notes Biomath.*, 11:1–15.

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2:23–41.

Duprey, Z. H., Steurer, F. J., Rooney, J. A., Kirchhoff, L. V., Jackson, J. E., Rowton, E. D., and Schantz, P. M. (2006). Canine visceral leishmaniasis, united states and canada, 2000-2003. *Emerging Infectious Diseases*, 12(3):440–446.

Dye, C. and Wolpert, D. M. (1988). Earthquakes, influenza and cycles of indian kala-azar. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 82:843–850.

Franke, C., Ziller, M., Staubach, C., and et al. (2002). Impact of el niño/southern oscillation on visceral leishmaniasis, Brazil. *Emerging Infectious Diseases*, 8:914–917.

Fried, J. R., Gibbons, R. V., Kalayanarooj, S., Thomas, S. J., Srikiatkhachorn, A., Yoon, I., Jarman, R. G., Green, S., Rothman, A. L., and Cummings, D. A. T. (2010). Serotype-specific differences in the risk of dengue hemorrhagic fever: an analysis of data collected in Bangkok, Thailand from 1994 to 2006. *PLoS Neglected Tropical Diseases*, 4(3):e617.

Gaskin, A., P, S., Jackson, J., Birkenheuer, A., Tomlinson, L., Gramiccia, M., Levy, M., Steurer, F., Kollmar, E., Hegarty, B., Ahn, A., and Breitschwerdt, E. (2002). Visceral leishmaniasis in a New York foxhound kennel. *Journal of Veterinary Internal Medicine*, 16:34–44.

Gelman, A. and Rubin, D. B. (1992). Inference from interative simulation using multiple sequences. *Statistical Science*, 7(4):457–511.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, 4:169–193.

Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 73:123–214.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.

Hasibeder, G., Dye, C., and Carpenter, J. (1992). Mathematical modelling and theory for estimating the basic reproductive number of canine leishmaniasis. *Parasitology*, 105:43–53.

Herwaldt, B. L. (2001). Laboratory-acquired parasitic infections from accidental exposures. *Clinical Microbiology Reviews*, 14:659–688.

Hsu, S. B. and Hsieh, Y. H. (2008). On the role of asymptomatic infection in transmission dynamics of infectious diseases. *Bulletin of Mathematical Biology*, 70:134–155.

Jeronimo, S. M., Duggal, P., Braz, R., Cheng, C., Monteiro, G., Nascimento, E., Martins, D., Karplus, T., Ximenes, M., Oliveira, C., Pinheiro, V., Pereira, W., Peralta, J., Sousa, J., Medeiros, I., Pearson, R., Burns, T., Pugh, E., and Wilson, M. E. (2004). An emerging peri-urban pattern of infection with Leishmania chagasi, the protozoan causing visceral leishmaniasis in northeast Brazil. *Scandinavian Journal of Infectious Diseases*, 36(6-7):443–449.

Jeronimo, S. M., Oliveira, R., Mackay, S., Costa, R., Sweet, J., Nascimento, E., Luz, K., Fernandes, M., Jernigan, J., and Pearson, R. (1994). An urban outbreak of visceral leishmaniasis in Natal, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 88:386–388.

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773–795.

Keeling, M. and Rohani, P. (2008). *Modeling Infectious Disease in Humans and Animals*. Princeton and Oxford: Princeton University Press.

Kermack, W. and McKendrick, A. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society, London*, 115:700–721.

Larson, M., Toepp, A., Scott, B., Kurtz, M., Fowler, H., Esfandiari, J., Howard, R., Vallur, A., Duthie, M., and Petersen, C. (2017). Semi-quantitative measurement of asymptomatic L. infantum infection and symptomatic visceral leishmaniasis in dogs using Dual-Path platform CVL. *Applied Microbiology and Biotechnology*, 101(1):381–390.

Laurenti, M., Rossi, C., Ribeiro da Matta, V., Tomokane, T., Corbett, C., Secundino, N., Pimenta, P., and Marcondes, M. (2013). Asymptomatic dogs are highly competent to transmit *Leishmania* (*Leishmania*) *infantum chagasi* to the natural vector. *Veterinary Parasitology*, 196:296–300.

Le Rutte, E., Chapman, L., Coffeng, L., Jervis, S., Hasker, E., Dwivedi, S., Karthick, M., Das, A., Mahapatra, T., Chaudhuri, I., Boelaert, M., Medley, G., Srikantiah, S., Hollingsworth, T., and de Vlas, S. (2017). Elimination of visceral leishmaniasis in the Indian subcontinent: a comparison of predictions from three transmission models. *Epidemics*, 18:67–80.

Lekone, P. and Finkenstädt, B. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 62:1170–1177.

Levy, M. and Yiengst, M. (1948). Kala-azar: report of a case showing unusual leukocyte response and prolonged incubation period. *Journal of the American Medical Association*, 136:81–84.

Li, M., Smith, H., and Wang, L. (2001). Global dynamics of an SEIR model with vertical transmission. *SIAM Journal on Applied Mathematics*, 62:58–69.

Li, M. Y., Graef, J. R., Wang, L., and Karsai, J. (1999). Global dynamics of a SEIR model with varying total population size. *Mathematical Biosciences*, 160:191–213.

Lima, I. D., Lima, A. L. M., Mendes-Aguiar, C. d. O., Coutinho, J. F. V., Wilson, M. E., Pearson, R. D., Quieroz, J. W., and Jeronimo, S. M. B. (2018). Changing demographics of visceral leishmaniasis in northeast Brazil: lessons for the future. *PLoS Neglected Tropical Diseases*, 12(3):e0006164.

Lima, I. D., Queiroz, J. W., Lacerda, H. G., Queiroz, P. V. S., Pontes, N., Barbosa, J. D. A., Martins, D. R., Weirather, J. L., Pearson, R. D., Wilson, M. E., and Jeronimo, S. M. B. (2012). *Leishmania infantum chagasi* in northeastern Brazil: Asymptomatic infection at the urban perimeter. *American Journal of Tropical Medicine and Hygiene*, 86(1):99–107.

Mancianti, F., Gramiccia, M., Gradoni, L., and Pieri, S. (1988). Studies of canine leishmaniasis control. 1. evolution of infection of different clinical forms of canine leishmaniasis following antimonial treatment. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 82(4):566–567.

Martin, A. D., Quinn, K. M., and Park, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, 42:22.

Martins-Melo, F. R., da Silveira Lima, M., Ramos Jr, A. N., Alencar, C. H., and Heukelbach, J. (2014). Mortality and case fatality due to visceral leishmaniasis in Brazil: A nationwise analysis of epidemiology, trends and spatial patterns. *PLoS One*, 9(4):e93770.

Michel, G., Pomares, C., Ferrua, B., and Marty, P. (2011). Importance of worldwide asymptomatic carriers of Leishmania infantum (L. chagasi) in human. *Acta Tropica*, 119:69–75.

Molina, R., Lohse, J. M., Pulido, F., Laguna, R., Lopez-Velez, R., and Alvar, J. (1999). Infection of sand flies by humans coinfected with leishmania infantum and human immunodeficiency virus. *American Journal of Tropical Medicine and Hygiene*, 60:51–53.

Moreno, J. and Alvar, J. (2002). Canine leishmaniasis: epidemiological risk and the experimental model. *Trends in Parasitology*, 18(9):399–405.

Morrison, A., Ferro, C., Pardo, R., and et al. (1995). Seasonal abundance of lutzomyia longipalpis (dipteral psychodidae) at an endemic focus of visceral leishmaniasis in Colombia. *Journal of Medical Entomology*, 32:538–548.

Neal, R. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G., and Meng, X., editors, *Handbook of Markov Chain Monte Carlo*, chapter 5. Chapman & Hall.

Nunes, C. M., de Lima, V. M. F., de Paula, H. B., Perri, S. H. V., de Andrade, A. M., Dias, F. E. F., and Burattini, M. N. (2008). Dog culling and replacement in an area endemic for visceral leishmaniasis in Brazil. *Veterinary Parasitology*, 153(6):19–23.

Ozanne, M. V., Brown, G. D., Oleson, J. J., Lima, I. D., Queiroz, J. W., Jeronimo, S. M., Petersen, C. A., and Wilson, M. E. (2019). Bayesian compartmental model for an infectious disease with dynamic states of infection. *Journal of Applied Statistics*, 46(6):1043–1065.

Petersen, C. A. and Barr, S. (2009). Canine leishmaniasis in North America: emerging or newly recognized? *Veterinary Clinics of North America: Small Animal Practice*, 39:1065–1074 (vi).

Piesman, J., Mather, T. N., Dammin, G. J., Telford, III, S., Lastavica, C., and Spielman, A. (1987). Seasonal variation of transmission risk of lyme disease and human babesiosis. *American Journal of Epidemiology*, 126(6):1187.

Plummer, M. (2008). Penalized loss functions for bayesian model comparison. *Biostatistics*, 9:523–539.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6:7–11.

Ponziani, F. R., Miele, L., Tortora, A., Furnari, M., Bodini, G., Pompili, M., Gasbarrini, A., and Giannini, E. G. (2018). Treatment of early stage chronic hepatitis C virus infection. *Expert Review of Clinical Pharmacology*, 11(5):519–524.

Quinnell, R. J. and Courtenay, O. (2009). Transmission, reservoir hosts and control of zoonotic visceral leishmaniasis. *Parasitology*, 136:1915–1934.

Raftery, A. L. and Lewis, S. (1992). How many iterations in the Gibbs sampler? *Bayesian Statistics*, 4:763–773.

Rock, K. S., le Rutte, E. A., de Vlas, S. J., Adams, E. R., Medley, G. F., and Hollingsworth, T. D. (2015). Uniting mathematics and biology for control of visceral leishmaniasis. *Trends in Parasitology*, 31(6):251–259.

Rosypal, A. C., Troy, G. C., Zajac, A. M., Frank, G., and Lindsay, D. S. (2005). Transplacental transmission of a North American isolate of Leishmania infantum in an experimentally infected beagle. *Journal of Parasitology*, 91(4):970–972.

Rubin, D. B. (1984). Bayesianly justifiable and relevant requency calculations for the applied statistician. *The Annals of Statistics*, 12(4):1151–1172.

Schantz, P. M., Steurer, F. J., Duprey, Z. H., Kurpel, K. P., Barr, S. C., Jackson, J. E., Breitschwerdt, E. B., Levy, M. G., and Fox, J. C. (2005). Autochthonous visceral leishmaniasis in dogs in North America. *Journal of the American Veterinary Medical Association*, 226:1316–1322.

Schwartz, E., Weld, L., Wilder-Smith, A., von Sonnenburg, F., Keystone, J., Kain, K., Torresi, J., and Freedman, D. (2008). Seasonality, annual trends, and characteristics of dengue among ill returned travelers, 1997-2006. *Emerging Infectious Diseases*, 14(7):1081–1088.

Sherlock, I. A. (1996). Ecological interactions of visceral leishmaniasis in the State of Bahia, Brazil. *Memorias do Instituto Oswaldo Cruz*, 91:671–683.

Shimozako, H. J., Wu, J., and Eduardo, M. (2017). Mathematical modelling for zoonotic visceral leishmaniasis dynamics: A new analysis considering updated parameters and notified human Brazilian data. *Infectious Disease Modelling*, 2:143–160.

Shinazi, R. B. (2000). Horizontal versus vertical transmission of parasites in a stochastic spatial model. *Mathematical Biosciences*, 168(1):1–8.

Silva, E. S., Gontijo, C. M., Pacheco, R. S., Fiuza, V. O., and Brazil, R. P. (2001). Visceral leishmaniasis in the metropolitan region of Belo Horizonte, state of Minas Gerais, Brazil. *Memorias do Instituto Oswaldo Cruz*, 96(3):285–291.

Spiegelhalter, D., Best, N., Carlin, B., and et al. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64:583–639.

Stauch, A., Sarkar, R. R., Picado, A., Ostyn, B., Sundar, S., Rijal, S., Boelaert, M., Dujardin, J. C., and Duerr, H. P. (2011). Visceral leishmaniasis in the indian subcontinent: Modelling epidemiology and control. *PLoS Neglected Tropical Diseases*, 5:e1405.

Toepp, A., Schaut, R., Scott, B., Mathur, D., and Berens, A. (2017). Leishmania incidence and prevalence in U.S. hunting hounds maintained via vertical transmission. *Veterinary Parasitology: Regional Studies and Reports*, 10:75–81.

Toepp, A. J., Monteiro, G. R., Coutinho, J. F., Lima, A. L., Larson, M., Wilson, G., Grinnage-Pully, T., Bennett, C., Mahachi, K., Anderson, B., Ozanne, M. V., Anderson, M., Fowler, H., Parrish, M., Willardson, K., Saucier, J., Tyrrell, P., Palmer, Z., Buch, J., Chandrashekar, R., Brown, G. D., Oleson, J. J., Jeronimo, S. M., and Petersen, C. A. (2019). Comorbid infections induce progression of visceral leishmaniasis. *Parasites & Vectors*, 12.

Toni, T., D, W., Strelkowa, N., and et al. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6:187–202.

van Griensven, J., Carillo, E., López-Vélez, R., Lynen, L., and Moreno, J. (2014). Leishmaniasis in immunosuppressed individuals. *Clinical Microbiology and Infection*, 20(4):286–299.

Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S*. Springer, 4 edition.

Werneck, G. L., Batista, M. S. A., Gomes, J. R. B., Costa, D. L., and Costa, C. H. N. (2003). Prognostic factors for death from visceral leishmaniasis in Teresina, Brazil. *Infection*, 31(3):174–177.

Zou, L., Chen, J., and Ruan, S. (2017). Modeling and analyzing the transmission dynamics of visceral leishmaniasis. *Mathematical Biosciences & Engineering*, 14(5-6):1585–1604.

Zou, L., Zhang, W., and Ruan, S. (2010). Modeling the transmission dynamics and control of hepatitis B virus in China. *Journal of Theoretical Biology*, 262(2):330–338.