Epidemiological models of rabies in domestic dogs: dynamics and control

Hawthorne L. Beyer

This thesis is submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

University of Glasgow
Faculty of Biomedical and Life Sciences
Division of Ecology and Evolutionary Biology

March 2010

Abstract

Epidemiological models are frequently used to estimate basic parameters, evaluate alternative control strategies, and set levels for control measures such as vaccination, culling, or quarantine. However, inferences drawn from these models are sensitive to the assumptions upon which they are based. While many simple models provide qualitative insights into disease dynamics and control, they may not fully capture the mechanisms driving transmission dynamics and, therefore, may not be reasonable approximations of reality. This thesis examines how the predictions made by simple models are influenced by assumptions regarding the dispersion of the transition periods, alternative infection states, and transmission heterogeneity resulting from population structuring. More realistic models of rabies transmission dynamics among domestic dogs in Serengeti District (Tanzania) are developed and applied to the problem of assessing vaccination efficacy, and designing pulsed vaccination campaigns.

Several themes emerge from the discussion of the models. First, the characteristics of outbreaks can be strongly influenced by the dispersion of the incubation and infectious period distributions, which has important implications for parameter estimation, such as the estimation of the basic reproductive number, R_0. Similarly, alternative infection states, such as long incubation times, can substantially alter outbreak characteristics.

Second, we find that simple SEIR models fail to accurately capture important aspects of rabies disease outbreaks among domestic dog populations in northern Tanzania, and therefore may be a poor basis for assigning control targets in this system. More complex models that included the role of human intervention in limiting outbreak severity, or that included population structure, were able to reproduce the observed outbreak size distribution. We argue that there is greater support for the structured population model, and discuss the implications of the three models on the evaluation of vaccination efficacy.

Third, at a more regional scale, we build metapopulation models of rabies transmission among domestic dog sub-populations. We use a Bayesian framework to evaluate competing hypotheses about mechanisms driving transmission, and sources of reinfection external to the dog population. The distance between sub-populations, and the size of the sub-populations receiving and transmitting infection are identified as important components of transmission dynamics. We also find evidence for a relatively high rate of re-infection of these populations from

neighbouring inhabited districts, or from other species distributed throughout the study area, rather than from adjacent wildlife protected areas. We use the highest ranked models to quantify the efficacy of vaccination campaigns that took place between 2002-2007. This work demonstrates how a coarse, proximate sentinel of rabies infection is useful for making inferences about spatial disease dynamics and the efficacy of control measures.

Finally, we use these metapopulation models to evaluate alternative strategies of pulse vaccination in order to maximize the reduction in the occurrence of rabies. The strategies vary in both the way in which vaccine doses are allocated to sub-populations, and in the trade-off between the frequency and intensity of vaccination pulses. The most effective allocation strategy was based on a measure of the importance of sub-populations to disease dynamics, and it had 30-50% higher efficacy than the other strategies investigated. This work demonstrates the strong potential for the role of metapopulation models in optimizing disease control strategies.

Candidates declaration

I declare that the work recorded in this thesis is entirely my own, except where otherwise stated, and that it is also of my own composition. Much of the material included in this thesis has been produced in co-authorship with others, and my personal contribution to each chapter is as follows:

Chapter 2. In preparation for submission as: Beyer, H.L. and Haydon, D.T. Effects of the dispersion of transition period distributions on outbreak dynamics of SEIR models. *Theoretical Population Biology.* The idea arose from discussions between DTH and HLB. HLB carried out the modelling work and was senior author in writing the manuscript.

Chapter 4. In preparation for submission as: Beyer, H.L., Hampson, K., Lembo, T., Cleaveland, S., Kaare, M., and Haydon, D.T. Limiting determinants of outbreak size distributions: a case study of canine rabies. *Proceedings of the Royal Society of London B.* KH, TL, MK, SC: carried out the fieldwork, compiled the empirical data, conducted demographic surveys and assisted with the delivery of vaccination campaigns. All authors discussed the work. DTH and HLB designed the model. HLB fit the models and wrote the manuscript.

Chapter 5. In review as: Beyer, H.L., Hampson, K., Lembo, T., Cleaveland, S., Kaare, M., and Haydon, D.T. Metapopultion dynamics of rabies and the efficacy of vaccination. *Proceedings of the Royal Society of London B.* KH, TL, MK, SC: conducted fieldwork and demographic surveys, compiled empirical data, and assisted with vaccination campaigns. All authors discussed the work. DTH and HLB designed the model. HLB fit the models and wrote the manuscript.

Chapter 6. In preparation for submission as: Beyer, H.L., Hampson, K., Lembo, T., Cleaveland, S., Kaare, M., and Haydon, D.T. The implications of metapopulation dynamics on the design of rabies vaccination campaigns. *Vaccine.* KH, TL, MK, SC: carried out the fieldwork, compiled the empirical data, conducted demographic surveys and assisted with the delivery of vaccination campaigns. All authors discussed the work. DTH and HLB designed the model. HLB fit the models and wrote the manuscript.

I further declare that no part of this work has been submitted as part of any other degree.

Hawthorne L. Beyer
University of Glasgow
March 2010

Acknowledgements

First and foremost I offer my profound gratitude to Dan Haydon, who is an exemplary mentor. It was Dan who first ignited my interested in epidemiology, and with infinite patience taught me a great deal about modelling. He allowed me the freedom to follow my whims, but has always been available to work through the problems. His knowledge of the big-picture is particularly valuable, and his ability to see with great clarity exactly what the problems are with a manuscript is at the same time both extremely irritating and very helpful. I consider myself very fortunate to have had the opportunity to work with Dan, and have greatly enjoyed our various expeditions.

I owe particular thanks to Katie Hampson who has been incredibly generous with her hard-earned data, and has a lot of insight into disease ecology that I have benefited from. In particular, the detailed feedback she provides on manuscripts sets the gold-standard for coauthorship. Sarah Cleaveland also has been a generous collaborator and is one of the few scientists I know who I can unequivocally identify as having 'made a difference'. Trying to solve disease problems in developing countries has to be one of the most difficult problems we face. It impresses me no end that she is spearheading large scale disease eradication programmes in Tanzania, pulling together many important partners to make it happen.

Richard Reeve has been very helpful in demystifying mathematical problems. What particularly impresses me about Richard's ability is the speed with which he comprehends, then solves, problems. In the time it takes me to point at an equation and begin to grope for the words to voice my question Richard has understood both the subject matter and the problem, and usually has clear insight into a solution. I also thank Juan Manuel (Pajaro) Morales for his instruction in Bayesian analysis and WinBUGS, and for being such a good travelling companion in Ethiopia and Italy.

The generous financial support of the Leverhulme Trust is also gratefully acknowledged. This funding provided me with an exceptional opportunity. This grant also paid for the twelve processors that I have kept very busy for the last year. I would also like to thank the P.I.'s on that grant for their efforts in developing it: Dan Haydon, Sarah Cleaveland and Karen Laurenson.

There are many other people who have inspired, assisted, instructed, resuscitated, defibrillated. It would have been a much harder and less enjoyable experience had it not been for the siblings in my cohort: Elizabeth Masden, Douglas Kerlin, Anaid Diaz Palacios, Flavie Vial, Sunny Townsend and Meggan Craft.

Finally, I must thank my friends and family for being so patient, supportive, and understanding of how difficult it is to lead a balanced life while submersed in a Ph.D. Annabelle in particular has been the best of companions. Anne and John, Don and Georgette, Trevelyan and Fiona, and of course my two nieces, Erin and Isla, whose energetic attentions provided such welcome distraction from academia.

Dedicated to my nieces, Erin (Pipsqueak) and Isla (Smiles)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Managing infectious disease is one of the most challenging problems humans face (King *et al.*, 2006), from both human health, agricultural, and conservation perspectives. Rabies, for instance, exerts a major public health and economic burden as it is responsible for at least 55,000 deaths worldwide, and expenditure on treatment and control exceeds US$500 million per annum (Coleman *et al.*, 2004; Knobel *et al.*, 2005). Epidemiological models are fundamental tools for understanding disease dynamics, predicting outbreak severity, evaluating the efficacy of control interventions, and attempting to optimize the deployment of new control measures. Simple models, such as SEIR (susceptible, exposed, infectious, recovered / removed) compartment models, can be used to explore disease dynamics and control programmes in a qualitative manner, although they are often too simplistic to accurately capture the complexity of real epidemics. The degree to which a model is useful for designing and evaluating control measures depends on the extent to which the model is a reasonable approximation of reality. Models that do not fully capture the mechanisms driving transmission dynamics might underestimate the level of control needed to prevent major outbreaks occurring, or might result in inefficient allocation of limited resources by suggesting inappropriate control targets.

There are two aspects to this problem. First, it is important to understand how disease dynamics described by simple models are sensitive to the simplifying assumptions upon which those models are based. This includes, for instance, the

assumption that transition period distributions are exponentially distributed, that it is reasonable to use a four-compartment model (SEIR), and that populations are well-mixed. Second, models can be developed that do not make these assumptions, and, therefore, sacrifice a certain amount of analytical tractability and simplicity to become more realistic. These models can include, for instance, realistic transition period distributions, alternative infection states, and population structure. Of particular importance to the problem of rabies in East Africa are metapopulation models that capture some of the heterogeneity in transmission arising from the spatial structuring of the population.

1.2 Thesis organisation

This thesis has been compiled as a collection of 6 chapters in paper format, one of which appears as an appendix. As some of the chapters are based on similar datasets (but address different aspects of disease dynamics in that system) there is inevitably some repetition of information among chapters. As each chapter has its own introduction and discussion, I include only brief general introduction and general discussion chapters (labelled 1 and 7).

Chapter 2 reviews how the dispersion of incubation and infectious period distributions affects parameter estimation (R_0, β) and the characteristics of outbreaks. A common simplifying assumption in epidemiological models is that the incubation and infectious periods are exponentially distributed. This is often an unrealistic assumption as transition rates between epidemiological states change as a function of the time an individual has spent in a given state. As a result, the exponential distribution overestimates the frequency of durations that are much lower than or higher than the mean, so the variance of the distribution is high. Using more realistic, less dispersed distributions can profoundly alter parameter estimations and the characteristics of outbreaks such as persistence time (the time to fade-out), the critical time (time to the peak number of infectious individuals), number of transmissions (outbreak size), and the probability that an outbreak will be large.

Chapter 2 also reviews how stochastic and deterministic solutions to compart-

ment models differ. Deterministic solutions are sometimes criticised because they allow disease to persist at unrealistically low densities that would result in fade-out of the disease in real populations. When population counts are used instead of densities, this can result in disease persisting in a fraction of an individual (e.g. the atto-fox, 1×10^{-18} foxes; Mollison, 1991). Because fade-out does not occur in deterministic solutions, they often predict epidemic cycles driven by the recruitment of new susceptible individuals in the inter-epidemic troughs, thereby facilitating a new outbreak. Yet deterministic solutions are very efficient compared to stochastic simulations, and provide many useful qualitative results. Chapter 2, therefore, brings together several ideas that have been reported previously, but that are not often considered simultaneously. It is useful to review these concepts first as a foundation for the modelling that occurs in later chapters.

In addition to the assumption of exponentially distributed incubation and infectious periods, compartment models of rabies often also assume that an SEIR model is a reasonable representation of the infection process. However, there are three hypothesized alternative infection states that may be important to disease dynamics. First, some rabies infections are characterized by unusually long incubation times that may indicate two alternative incubation processes. Second, a carrier state has been hypothesized whereby an individual intermittently sheds live virus in saliva but without displaying typical clinical signs of the disease or suffering the increased mortality normally associated with infection (Fekadu, 1975). Finally, it is usually assumed that rabies is a fatal disease, but there is evidence that recovery from rabies infection is possible, especially in the earliest stages of infection.

Cleaveland & Dye (1995) incorporate these alternative infection states into compartment models and compare the behaviour of the endemic equilibrium among the models using deterministic methods. In Chapter 3, the four models of Cleaveland & Dye (1995) are generalized to include a variable number of incubation and infection stages, thereby facilitating the use of the method of stages (Cox & Miller, 1965) to accommodate realistic transition period distributions. Expressions for the basic reproductive number, R_0, that incorporate the alternative infection states and multiple stages for the states, are also presented for each of the models.

I review the pathological and empirical evidence for three alternative infection states, and quantify their effect on outbreak characteristics of rabies in domestic dogs using stochastic simulations of these compartment models. Further to Chapter 2, realistic distributions are fit to empirical data on domestic dog incubation and infectious period durations (Hampson *et al.*, 2009), and the consequence of assuming exponentially distributed transition times is also quantified.

Together, Chapters 2 and 3 are a review of the importance of model assumptions on quantifying outbreak dynamics, with particular emphasis on rabies. The next three chapters are more applied, and are based on data collected by my collaborators working in Tanzania (in particular Sarah Cleaveland, Katie Hampson, Tiziana Lembo and Magai Kaare). The focus of this work is the domestic dog population in Serengeti District (SD), northern Tanzania, which borders wildlife protected areas to the south and east (Serengeti National Park and the Ikorongo and Grumeti Game Reserves), and other inhabited districts to the north and west (Bunda, Musoma and Tarime Districts). The inhabitants of this district (approximately 175,000 people in 75 villages) live in primarily agro-pastoralist communities and use domestic dogs for guarding households and livestock. Rabies has been a problem in this region since 1979 (S. Cleaveland, *pers. comm.*).

Stochastic simulations of SEIR compartment models predict a bimodal distribution of outbreak sizes (Anderson & Watson, 1980). This dichotomy in outcomes is driven by stochastic fade-out of outbreaks, resulting in outbreaks that are small and short-lived (minor), or larger and longer-lasting (major). Anderson & Watson (1980) developed analytical approximations for the proportion of outbreaks that are minor and major as a function of R_0 and the number of stages used to represent the incubation and infectious period distributions. In the case of domestic dogs in Serengeti District, this approximation predicts that over 10% of outbreaks should be major. In fact, none of the 185 observed outbreaks were major. This indicates that simple SEIR models are not a good representation of disease dynamics in this system, which has important implications for the use of these models in evaluating intervention strategies and settings control targets. The major outbreaks are responsible for the majority of incidence in simulations, and therefore have the potential to

have a disproportionately large influence on estimates of control efficacy.

In Chapter 4 we explore two hypotheses that might account for the absence of major rabies outbreaks in the observed size distribution. First, human intervention shortly after the onset of cases may limit the severity of outbreaks. Although we model the influence of human intervention in general terms as a reduction in the transmission parameter soon after the start of an outbreak, this effect could result, for instance, from owners restricting the movement of dogs. Second, although homogeneous mixing is often assumed to be a reasonable assumption in small populations, transmission heterogeneity resulting from host population structure may limit outbreak size. We use Approximate Bayesian Computation (Toni *et al.*, 2009) to evaluate competing models that differ in the timing and strength of human intervention, or in population structure and coupling.

In Chapter 5 we develop metapopulation models to explore how heterogeneity in transmission dynamics resulting from spatial structure in a host population at a regional level drives disease dynamics. A Bayesian framework is used to evaluate competing metapopulation models of rabies transmission among domestic dog populations in Serengeti District, northern Tanzania. Because of the difficulty of collecting epidemiological data in this region, a proximate indicator of disease, medical records of animal-bite injuries, is used to infer the occurrence of suspected rabid dog cases in one month intervals. Hence, the metapopulation models are similar to stochastic patch-occupancy models. State-space models are used to explore the implications of different levels of reporting probability on model parameter estimates.

This is not a closed system and the metapopulation models include an external source of reinfection. We hypothesize that this source arises from neighbouring inhabited districts to the north and west, the protected areas (e.g. Serengeti National Park) to the south and east, or from other hosts that are distributed throughout the district. We use model selection approaches to rank the relative likelihood of the three reinfection sources and the rate at which reinfection of the dog population occurs. Finally, we use the top ranked models to quantify the efficacy of pulsed vaccination campaigns that took place between 2002-2007.

An obvious application of these metapopulation models is to address whether they can be used to improve the efficacy of intervention measures. In Chapter 6, therefore, we evaluate alternative strategies of pulse vaccination in order to maximize the reduction in the occurrence of rabies. The strategies vary according to the manner in which vaccine is allocated, and the trade-off between frequency and intensity of pulses.

Chapter 2

Effects of the dispersion of transition period distributions on outbreak dynamics of SEIR models

2.1 Abstract

In epidemiological modelling, the dispersion of the incubation and infectious period distributions have important consequences on the estimation of the basic reproductive number, R_0, and disease dynamics in large populations at the endemic equilibrium. However, the affect of the dispersion of these distributions on outbreak dynamics in smaller populations is less well explored. Here we use stochastic simulations of outbreaks to quantify the effect of the dispersion of the incubation and infectious periods on persistence time, critical time and outbreak size. We find that as the dispersion of the infectious period decreases, persistence time and the critical time are reduced. Less dispersed incubation periods result in a slight decrease in persistence times, and a slight increase in critical times. This effect becomes more pronounced as the duration of the incubation period increases relative to the infectious period. Outbreak size is insensitive to the dispersion of the distributions. The dichotomy in outcomes of simulated outbreaks, which are either small and brief (minor) or large and longer-lived (major), results in a bimodal distribution of outbreak characteristics. Less dispersed infectious periods increase the probability an outbreak will be major, particularly in systems where R_0 is small (< 6) but larger than 1. Deterministic solutions of these models describe the characteristics of out-

breaks conditional upon a major outbreak occurring, and therefore fail to capture important features of outbreaks (stochastic fade-out and persistence time). Deterministic and stochastic approaches can provide qualitatively similar results but differ in their quantitative predictions, which may be important when designing control measures. Importantly, outbreak dynamics (persistence and critical times, and the proportion of outbreaks that are major) are strongly influenced by the dispersion of the incubation and infectious period distributions using either of these approaches. We discuss the implications of this work to the design and evaluation of control measures.

Keywords: compartment models; stochastic simulation; method of stages; incubation; infectious; SEIR; dispersion

2.2 Introduction

Mathematical epidemiological models are increasingly being used to identify appropriate management responses to infectious disease outbreaks (Matthews *et al.*, 2003), inform public policy on disease management in the event of future outbreaks (Ferguson *et al.*, 2003; Haydon *et al.*, 2004), and design and evaluate control strategies (Haydon *et al.*, 1997; Keeling *et al.*, 2001, 2003; Haydon *et al.*, 2006; Tildesley *et al.*, 2006; Feng *et al.*, 2007). One common simplifying assumption in these models is that the probability of an event occurring (e.g. recovery of an infectious individual) is constant through time, and waiting times between events are therefore exponentially distributed. This is often an unrealistic assumption: transition rates between epidemiological states change as a function of the time an individual has spent in a given state. For instance, the chance of recovery from a non-fatal infection is usually low immediately following infection, and increases through time. The mathematically convenient exponential distribution overestimates the frequency of durations that are much shorter or longer than the mean (Lloyd, 2001c), thus the dispersion of this distribution is unrealistically large for many diseases.

Disease dynamics are sensitive to the distributions used to model the transition

periods (Anderson & Watson, 1980; Lloyd, 1996; Keeling & Grenfell, 1998; Andersson & Britton, 2000; Lloyd, 2001b,c). In SIR (susceptible, infectious, recovered) models, less dispersed infectious period distributions (IPDs) result in decreased stability of the endemic equilibrium (Lloyd, 1996, 2001c), decreased persistence time (time to fade-out) of the disease in the population (Andersson & Britton, 2000; Lloyd, 2001c), and outbreaks that take off faster and have a higher peak number of cases (Wearing *et al.*, 2005). In SEIR models, the addition of the exposure/incubation period (E), representing a time when individuals are infected but not yet infectious, adds a delay into the system that increases persistence time and critical time (the time to the peak number of infectious cases). Analytical approximations suggest that in large populations less dispersed incubation period distributions (EPDs) reduce persistence time, but that the effect of the dispersion of the IPD depends on the relative length of the incubation and infectious periods (Andersson & Britton, 2000). If the incubation period is short relative to the infectious period, SEIR models are well approximated by SIR models and using less dispersed IPDs will also decrease long term persistence time and stability (Lloyd, 2001c; Andersson & Britton, 2000). However, if the incubation period is much longer than the infectious period the opposite behaviour is observed: persistence times and model stability increase when less dispersed distributions are used (Lloyd, 2001c). In SEIR models, analytical approximations based on the assumption of large populations suggest less dispersed IPDs result in a slight decrease in the mean and variance of the size of major outbreaks, but the dispersion of the EPD has no affect on the size distribution (Anderson & Watson, 1980).

As a result of these effects, the dispersion of the EPD and IPD can have important consequences for the estimation of the basic reproductive number, R_0, based on the initial epidemic growth rate (λ) or from trajectory matching (Wearing *et al.*, 2005). Models using over-dispersed EPDs and IPDs result in an underestimation and overestimation of R_0 respectively (Anderson & Watson, 1980; Lloyd, 2001a; Wearing *et al.*, 2005), with the dispersion of the infectious period having a relatively smaller effect on R_0 (Figure 2.1). The bias resulting from the use of over-dispersed distributions can be substantial (Wearing *et al.*, 2005) in epidemics where R_0 is high

(approximately greater than 6), but bias is small for lower values of R_0 (Figure 2.1).

Inferences regarding the design and efficacy of control measures may, therefore, be sensitive to the distributions used (Wearing *et al.*, 2005; Feng *et al.*, 2007) depending on how parameters have been estimated. If R_0 is estimated using trajectory matching of an SEIR model, and the same models are used to evaluate the effectiveness of control strategies, then using over-dispersed distributions may not result in an important bias in the inferences. However, if R_0 is estimated directly from empirical data, for instance using contact tracing, then using over-dispersed distributions may substantially bias estimates of efficacy and the level of control measures needed to prevent major outbreaks. While underestimating the efficacy of control measures may only result in an inefficient allocation of resources, overestimating efficacy may result in inadequate protection of a population from disease outbreaks. Thus, the dispersion of the incubation and infectious periods can have an important influence on both the estimation of R_0 and the evaluation of control measures.

Although the influence of the dispersion of the incubation and infectious periods on R_0 is well understood (Anderson & Watson, 1980; Lloyd, 2001a; Wearing *et al.*, 2005), there has been less work examining the effects of the dispersion of the incubation and infectious periods on outbreak dynamics above and beyond the influence on R_0. Here we use stochastic simulations of SEIR compartment models to quantify the effect of less dispersed distributions on outbreak dynamics relative to an SEIR model with exponentially distributed incubation and infectious periods. We compare the stochastic simulations to deterministic solutions of these models, and discuss the implications of the assumption of exponentially distributed periods on the evaluation of disease control strategies. In contrast to the effect of these distributions on the estimation of R_0, the dispersion of the infectious period is the dominant of the two effects on the characteristics of outbreaks, but the dispersion of the incubation period also can be important when the duration of the incubation period is long relative to that of the infectious period.

Here, we assume that R_0 has been estimated directly from empirical data (e.g. using contact tracing), and not by means of trajectory matching using an SEIR model. The estimate of R_0 is therefore independent of any particular SEIR model. We

demonstrate that the assumption of exponentially distributed infectious periods can result in a substantial underestimation of the proportion of outbreaks that are major (sensu Anderson & Watson, 1980) versus those that fade-out almost immediately, implying a subsequent overestimation in the efficacy of control strategies.

2.3 Methods

We use a SEIR compartment model to investigate outbreak dynamics of a non-fatal disease in a small, entirely susceptible, well mixed population into which a single infectious individual is introduced. One way of incorporating more realistic event time distributions into stochastic simulations of epidemiological models is the method of stages (Cox & Miller, 1965; Anderson & Watson, 1980; Lloyd, 1996, 2001b), in which the incubation and infectious periods (of mean duration $1/\sigma$ and $1/\alpha$ respectively) are divided into m and n discrete, exponentially distributed stages respectively. The incubation and infectious periods are therefore the sum of m and n independent exponential random variables, each having a mean $1/m\sigma$ and $1/n\alpha$ respectively (Anderson & Watson, 1980). Overall, the distribution of time spent in the incubation and infectious states is gamma distributed, whereby the shape parameter corresponds to m or n stages, and the scale parameter is to $1/m\sigma$ or $1/n\alpha$, respectively. As the number of stages increases, the overall mean duration remains the same but the dispersion of the distribution decreases (Figure 2.2).

The model dynamics are determined by the following equations that govern the rates of change between the four epidemiological states (susceptible, exposed / incubation, infectious, and recovered):

$$dS/dt = bN - \beta SI - dS \tag{2.1}$$

$$dE_1/dt = \beta SI - (m\sigma + d)E_1 \tag{2.2}$$

$$dE_j/dt = m\sigma E_{j-1} - (m\sigma + d)E_j, \quad (j = 2, ..., m) \tag{2.3}$$

$$dI_1/dt = m\sigma E_m - (n\alpha + d)I_1 \tag{2.4}$$

$$dI_j/dt = n\alpha I_{j-1} - (n\alpha + d)I_j, \quad (j = 2, ..., n) \tag{2.5}$$

$$dR/dt = n\alpha I_n - dR \tag{2.6}$$

where $E = \sum_{i=1}^{n} E_i$, $I = \sum_{i=1}^{n} I_i$, $N = S + E + I + R$, and βSI represents density dependent transmission, which we assume is a reasonable assumption for modelling outbreaks in small populations, although several other transmission models are possible (reviewed in McCallum *et al.*, 2001). E_j and I_j are the *j*th stage of the incubation and infectious periods respectively. The model includes demographic processes (birth rate, b, and natural death rate, d). Initial conditions were $S = 1000$, $E_1 = 0$, $I_1 = 1$. The mean infectious period was 10 days, but to quantify outbreak dynamics for different ratios of incubation and infectious periods, the incubation period was 5, 10 or 20 days, corresponding to ratios of 1:2, 1:1, and 2:1.

As our goal was to quantify the effect of the dispersion of the EPD and IPD while controlling for its influence on R_0, the transmission parameter β was adjusted so that R_0 was constant (1.5) among models, regardless of the number of stages, using (Feng *et al.*, 2007):

$$R_0 = \left(\frac{m\sigma}{m\sigma + d}\right)^m \frac{\beta N}{n\alpha + d} \sum_{j=0}^{n-1} \left(\frac{n\alpha}{n\alpha + d}\right)^j \tag{2.7}$$

This expression relates R_0 to β while correcting for the natural death of individuals during the incubation or infectious periods and in the absence of natural death this simplifies to $R_0 = \beta N/\alpha$. Although we present the more general forms of the equations that include demographic parameters, the birth and death rate were set to 0 in our simulations to close the system and facilitate comparison of the stochastic and deterministic solutions.

The dynamical properties of these models were evaluated using 100,000 stochastic simulations of each model, with every combination of 1, 2, 5, 10 and 20 incubation and infectious stages. Simulations were based on a continuous time Gillespie algorithm (Gillespie, 1976), and were run until fade-out of the disease in the population. Each simulation was characterized by the outbreak size, the persistence time, and the critical time.

Numerical simulation was used to quantify the deterministic solutions to these models using the "odesolve" package in R (Appendix A; R Development Core Team, 2009). Deterministic methods do not predict fade-out, even in a closed system, so fade-out was assumed to occur when the number of susceptible individuals dropped below 1. Critical time could be identified precisely, and the outbreak size was calculated by subtracting the number of susceptible individuals remaining at the end of the simulation time from the initial number.

There is a dichotomy in the outcome of stochastic simulations of outbreaks, which may be small and brief (minor) or large and long-lived (major). Each simulated outbreak was categorized as minor or major based on the bimodal distribution of outbreak sizes (for the population sizes modelled here, an outbreak was considered major if there were more than 200 cases). Anderson & Watson (1980) provide an analytical solution to the problem of predicting the probability that an outbreak will be major for an SEIR model with gamma distributed periods. The approximation of the probability of a major outbreak is $1 - \pi^{\psi}$ where π is the smaller root of

$$\pi \left[1 + \frac{R_0}{m}(1 - \pi)\right]^m = 1 \tag{2.8}$$

and ψ is a function of the initial conditions (Anderson & Watson, 1980):

$$\psi = E + 1/m \sum_{j=1}^{m} (m - j + 1) I_j. \tag{2.9}$$

In our analysis, in which we assume an outbreak begins with a single, newly infectious individual entering a population, $\psi = 1$.

When $m = 1$, Equation 2.8 simplifies to the solution $pi = 1 - 1/R_0$ (Whittle, 1955; Anderson & May, 1991), which can be intuitively explained as follows. If we define

x as the proportion of the population that is susceptible and x' as the proportion of the population that is susceptible at the endemic equilibrium, then when $x = 1$, $xR_0 = R_0$, and when $x = x'$, $x'R_0 = 1$. Thus, $x' = 1/R_0$ and if the proportion of the population that is susceptible is reduced by $1/R_0$ by vaccinating $1 - 1/R_0$ of the population, then the effective reproductive number is 1 and the probability of a large outbreak approaches 0. The term m in Equation 2.8 is an adjustment to account for multiple stages (Anderson & Watson, 1980). As m increases, corresponding to a reduction in the variance of the duration of the infectious period distribution, the solution to π decreases, and the probability of a major outbreak therefore increases.

2.4 Results

The dichotomy in outcomes of simulated outbreaks (e.g. Figure 2.3a) resulted in a bimodal distribution of outbreak sizes, persistence times and critical times (Figure 2.3b). The probability of disease fade-out before a major outbreak occurs is an important feature of stochastically modelled outbreak dynamics. However, as fade-out is not possible in deterministic solutions they reflect the behaviour of the system conditional upon a major outbreak occurring. To facilitate comparison of stochastic and deterministic solutions we therefore characterize the proportion of outbreaks that are minor and major, but present only the mean characteristics of major outbreaks for the stochastic models.

In stochastic models, outbreak size in major outbreaks was insensitive to the dispersion of the incubation and infectious periods, or the duration of the incubation period (Figure 2.4a-c). The stochastic and deterministic estimates of the size of major outbreaks was similar (Figure 2.5a, b), an agreement that improved as the dispersion of the IPD decreased (Figure 2.5a).

Persistence times (Figure 2.4d-f) and critical times (Figure 2.4g-i) decreased as the dispersion of the IPD decreased. There was an approximately 15% difference in persistence and critical times between the models with 1 and 20 infectious period stages. Decreasing the dispersion of the EPD resulted in a slight decrease in persistence times, and a slight increase in critical times. The strength of this effect was

stronger as the duration of the incubation period increased relative to that of the infectious period (Figure 2.4f, i).

The deterministic estimates were only approximately similar to the stochastic means for persistence (Figure 2.5c, d) and critical times (Figure 2.5e, f). In both cases the deterministic estimate for a model with exponentially distributed infectious periods was considerably higher than the mean of the stochastic simulations, and as the dispersion of the infectious period decreased the associated reduction in persistence and critical times was larger for the deterministic estimate. Also, the dispersion of the incubation period had small but opposite effects on persistence times in the stochastic and deterministic solutions (Figure 2.5c, d): as the dispersion of the EPD decreased persistence times slightly decreased and increased respectively.

The probability that an outbreak would be major was profoundly influenced by the dispersion of the IPD. It increased as the dispersion of the IPD decreased but was insensitive to the dispersion of the EPD or the duration of the incubation period (Figure 2.4j-l). There was close agreement between the stochastic and analytical estimates (data not shown). To examine the generality of the influence of the dispersion of the IPD on the probability an outbreak will be major, simulations were run for a range of values of R_0 (0.5-10), whereby the value of β was adjusted to reflect R_0 based on equation 2.7. Two models were used: both with an exponentially distributed EPD, but one with an exponentially distributed IPD, and the other with 20 stages used to represent the IPD. For values of R_0 greater than 1, the assumption of exponentially distributed infectious periods results in a lower probability that an outbreak will be major across a wide range of values of R_0 (Figure 2.6). When R_0 is less than or equal to 1 there is no discernible effect of the dispersion of the IPD because immediate fade-out occurs in the majority of simulations.

2.5 Discussion

For a given value of R_0, modelling outbreaks using less dispersed IPDs results in an increased probability that an outbreak will be major, and reduced disease persistence and critical times in these major outbreaks. One explanation for the

effect of less dispersed IPD is that individuals have a greater chance of passing on infection before they are removed from the population (Keeling & Grenfell, 1998; Lloyd, 2001c). With exponentially distributed infectious periods many individuals are infectious for less than the mean of the distribution and are therefore less likely to pass on infection, while a small number are infectious much longer than the mean. Although these latter individuals are more likely to transmit infection, the resulting secondary cases also tend to be infectious for short periods of time. In contrast, when less dispersed distributions are used, there is less variation among individuals in the number of transmission events. Exponential distributions thus tend to result in slower, 'smouldering' epidemics compared to the faster, more explosive epidemics that burn-out more rapidly when less dispersed distributions are used. This is clearly reflected in the patterns of persistence and critical times we observed (see Figure 2.4). Although these effects relate to outbreak behaviour, long-term persistence of the disease is also likely to be reduced because less dispersed IPDs destabilize the model at the endemic equilibrium, increasing variability and therefore result in a greater chance of stochastic fade-out (Lloyd, 2001b,c).

The effect of the EPD is to add a delay into the system that resulted in epidemics that take longer to peak and then fade-out. Increasing the duration of the incubation period relative to the infectious period resulted in a substantial increase in persistence and critical times, and increased the strength of the effect of the dispersion of the EPD on persistence and critical times. For diseases with incubation periods that are long relative to the infectious period, the dispersion of the EPD can have an even greater affect on critical times than the dispersion of the IPD (data not shown). Furthermore, the reduction in critical and persistence times resulting from less dispersed IPD's diminishes as the relative duration of the EPD increases, and will switch to a positive effect if the relative duration of the EPD is long enough. The effect of the dispersion of the IPD is, therefore, not monotonic with respect to critical and persistence times (Andersson & Britton, 2000). Thus, the dispersion of both the EPD and IPD has important implications for parameter estimation using simulation-based techniques such as trajectory matching (Wearing *et al.*, 2005) or approximate Bayesian computation (Toni *et al.*, 2009).

In SEIR models that do not include demographic processes the dispersion of the EPD does not influence the size of major outbreaks or the probability that an outbreak will be major. Transmission is driven by the duration of the infectious period and the contact rate among individuals, neither of which is directly influenced by the duration or dispersion of the incubation period. However, the EPD can influence outbreak dynamics in models that include demographic processes if the natural death rate is high (some incubating individuals are removed from the population before becoming infectious), or if the duration of the EPD allows recruitment of new susceptible individuals into the population prior to the individual becoming infectious. The former effect reduces the size of outbreaks, and the latter effect increases the size of outbreaks.

The bimodality of outcomes is an important feature of outbreaks that can be overlooked in deterministic solutions of SEIR models because they describe the system conditional upon a major outbreak having occurred. This is important from a disease control perspective because the reduction in probability of a major outbreak occurring as R_0 is reduced by control measures is a key outcome. When resources for the control of disease are limited and eradication is therefore not plausible, control measures can be deployed to reduce the risk of a major outbreak knowing that small outbreaks that fade-out quickly can be tolerated (Haydon *et al.*, 2006). However, the dispersion of the IPD has a strong influence on the probability that an outbreak will be major for values of R_0 that are low but greater than 1 (e.g. approximately $1 < R_0 \leq 6$, depending on the specifics of the model). An exponentially distributed IPD underestimates the probability an outbreak will be major, and control targets based on such models may, therefore, be overly optimistic.

At higher values of R_0 (e.g. > 6) the dispersion of the IPD will have a less pronounced effect on the proportion of major outbreaks simply because the majority of outbreaks will be major even for the exponential model. However, when R_0 is high the dispersion of the IPD could still have a strong influence on estimates of control efficacy because the expected reduction in major outbreaks resulting from control measures will be sensitive to the dispersion of the IPD. Furthermore, higher values of R_0 are expected to amplify the influence of the dispersion of the EPD

and IPD on persistence and critical times, which could be important when using trajectory matching to estimate parameters.

The dispersion of the incubation and infectious periods have important consequences for the estimation of R_0 and the quantitative characteristics of outbreaks based on either stochastic simulations or deterministic solutions of compartment models. When R_0 is estimated based on the initial epidemic growth rate, models using over-dispersed EPDs and IPDs result in an underestimation and overestimation of R_0 respectively, with the influence of the dispersion of the EPD being the larger of the two effects (Anderson & Watson, 1980; Lloyd, 1996, 2001a; Wearing *et al.*, 2005). Because R_0 is often used to deduce the value of the transmission parameter β, underestimating R_0 results in an underestimate of β. Deterministic solutions or stochastic simulations based on that estimate will therefore underestimate the severity of an outbreak, which is strongly influenced by the magnitude of β. This bias is small when the estimate of R_0 based on the exponential model is small (less than 5).

Including demographic parameters in the deterministic models can result in epidemic cycles that may dampen through time depending on the formulation of the model. One criticism of such models is that the density of infected individuals in the epidemic troughs is so low (e.g. 10^{-18} animals/km^2; Fowler, 2000) that stochastic fade-out of the disease would be inevitable (Mollison, 1991; Keeling & Grenfell, 1997; Fowler, 2000). In our stochastic simulations fade-out occurred among all simulations, in part because there is no recruitment of susceptible individuals in our population, but also because of stochastic extinction of many outbreaks at an early stage. Deterministic models can provide qualitative insights into outbreak dynamics, but fail to capture the important characteristics of outbreaks.

Compartment models can be used to quantify the relative efficacy of different control measures and to identify targets of the numbers of individuals to receive treatment or control (e.g. Anderson, 1986; Keeling *et al.*, 2003; Haydon *et al.*, 2006; Feng *et al.*, 2007). The mathematically convenient but biologically unrealistic assumption of exponentially distributed infectious periods can result in substantial overestimation of persistence and critical times, and underestimation of the pro-

bability of a major outbreak occurring. Assessments of control measures based on exponentially distributed infectious periods may therefore overestimate control efficacy and underestimate control measure targets needed to prevent or eliminate outbreaks. These conclusions are consistent with Feng *et al.* (2007) who show that estimates of the efficacy of control measures (combinations of quarantine and isolation) are sensitive to the distributions used, and that the relative rankings of efficiency of these measures (and therefore the choice of control strategy) can change when less dispersed distributions are used. Similarly, Wearing *et al.* (2005) warn that using exponentially distributed incubation and infectious periods can bias estimates of R_0, possibly resulting in overly optimistic estimates of the efficacy of disease control measures.

The properties of epidemiological models are sensitive to the distributions used, therefore the use of more realistic distributions should be adopted as standard practice. Given that gamma distributed event times can be simulated using the method of stages (Cox & Miller, 1965; Lloyd, 2001b), and that a gamma distribution with an integer shape parameter (a necessary precursor to using the method of stages) can be fit to empirical incubation and infectious period data, it is straightforward to build more realistic distributions into compartment models.

2.6 Acknowledgements

Financial support for this work was provided by The Leverhulme Trust.

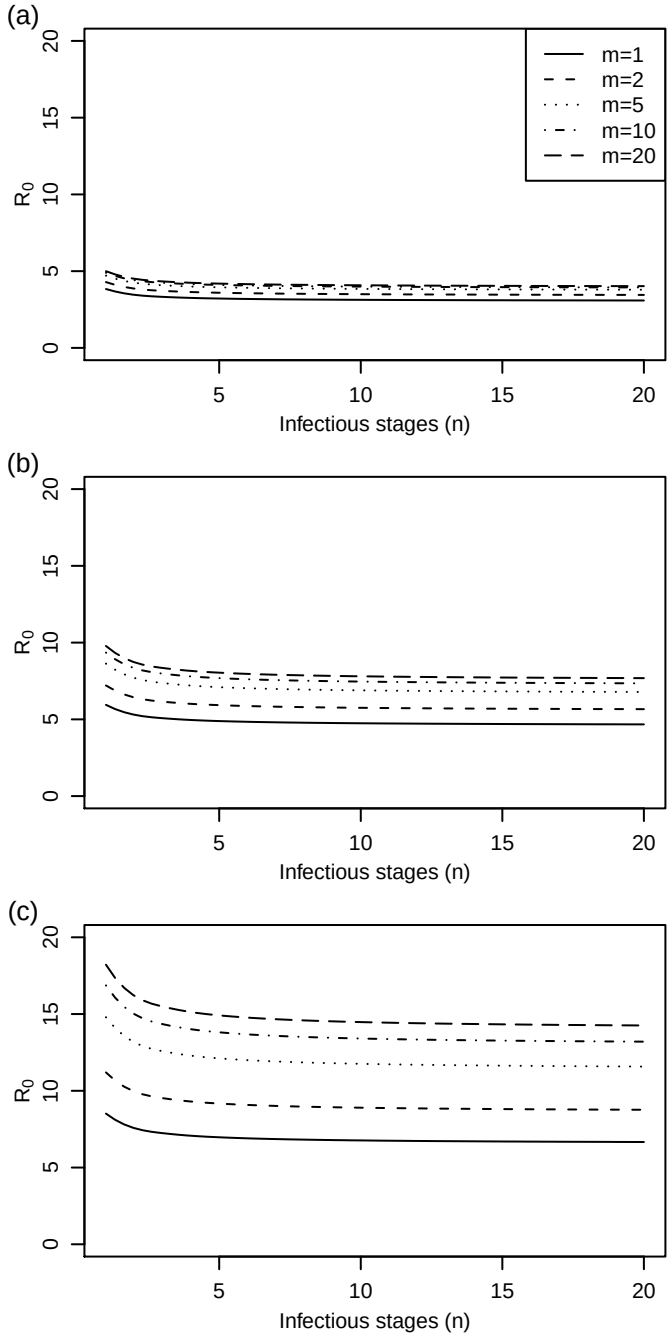

Figure 2.1: The effect of the dispersion of the incubation and infectious periods on the estimation of R_0 from the epidemic growth rate (λ), whereby $R_0 = (\lambda(\lambda(\sigma m)^{-1} + 1)^m)/(\alpha(1 - (\lambda(\alpha n)^{-1} + 1)^{-n}))$. As the number of stages used to model the incubation period (m) and infectious period (n) increases, the dispersion of the distribution decreases. One stage corresponds to the exponential distribution. The initial growth rate (λ) is 50, 75, and 100 yr^{-1} in plots a-c respectively, and the mean duration of the incubation and infectious period is 7 days in all models. The dispersion of the infectious period ($1/\alpha$) has a small effect relative to the dispersion of the incubation period ($1/\sigma$). Bias resulting from the assumption of exponentially distributed periods is proportional to R_0, thus when R_0 is small (< 4) this bias is negligible.

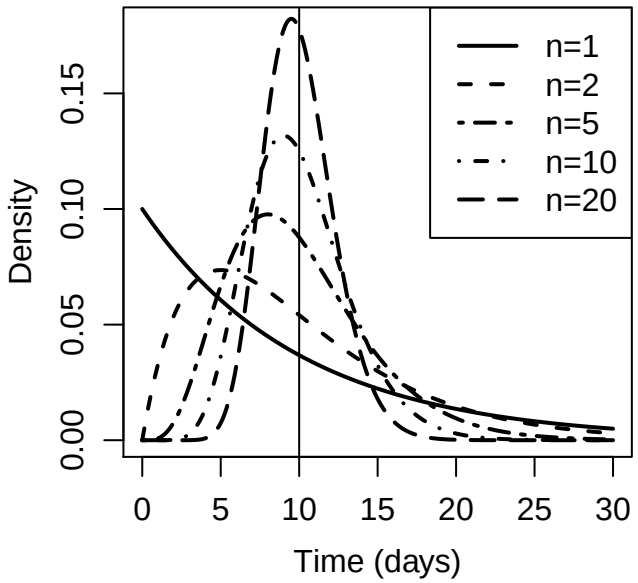

Figure 2.2: Probability density of gamma distributed infectious period times. When the shape parameter of the gamma distribution, n, equals 1, the gamma distribution simplifies to an exponential distribution (solid black line). When $n > 1$, this corresponds to compartment models containing n stages (see Methods section), and results in distributions with lower variance that may be more realistic representations of infectious period times. The vertical line identifies the common mean for all distributions (10 days).

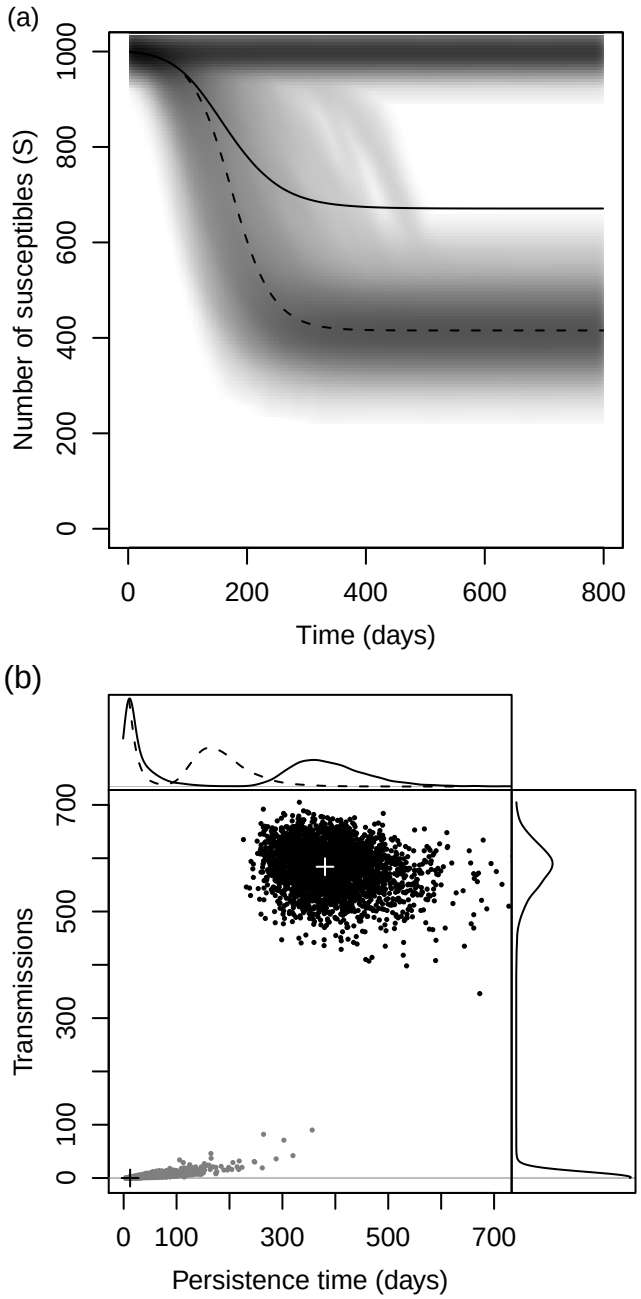

Figure 2.3: Illustration of the bimodal nature of simulated epidemics based on 100,000 stochastic simulations of an SEIR model. (a) Number of susceptible individuals through simulation time. Note the high density bands above and below the mean (black line) that correspond to minor and major epidemics respectively. The dashed line represents the deterministic solution of the model. (b) Main plot: relationship between disease persistence time (days) and outbreak size for minor (grey dots) and major (black dots) epidemics, with the median of these distributions identified by the black and white crosses respectively. The distribution of persistence times (top, solid line), critical times (top, dashed line) and outbreak sizes (right) are bimodal.

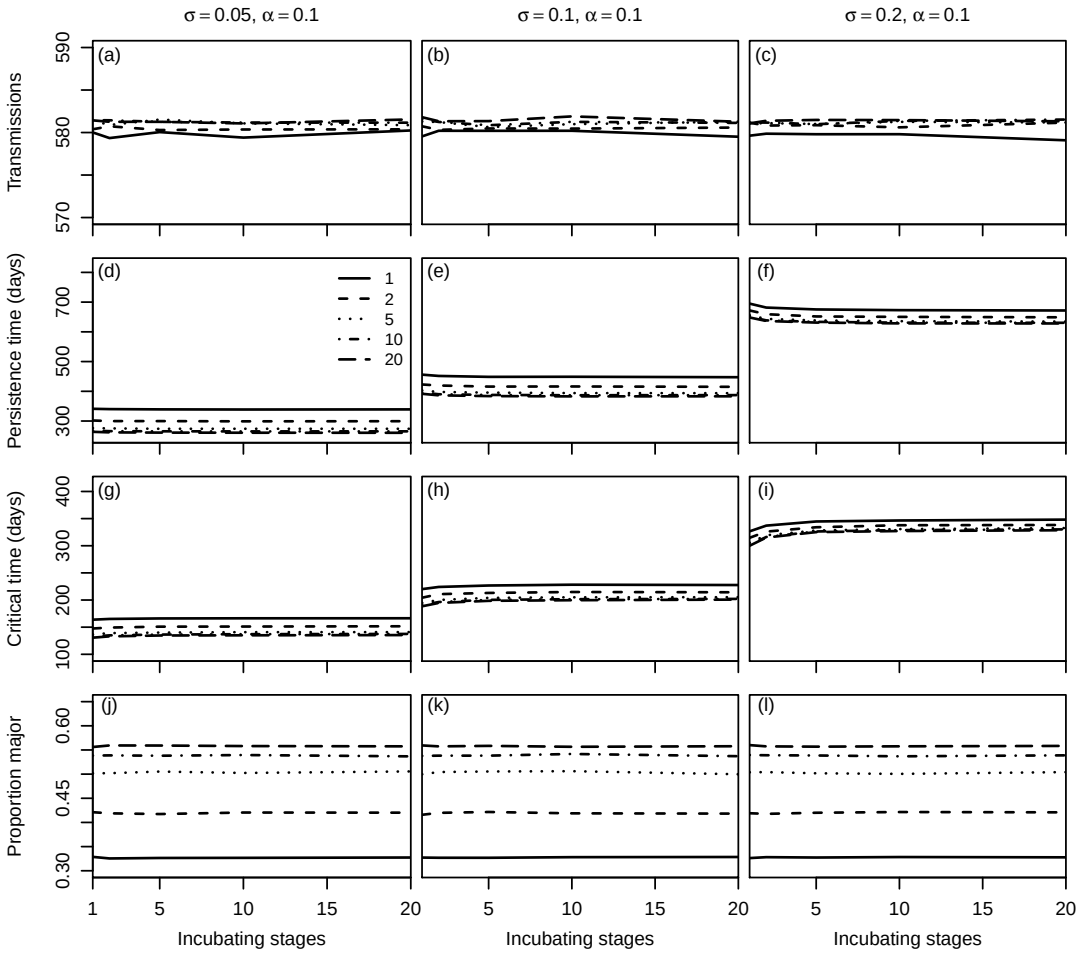

Figure 2.4: The effect of the dispersion of the incubation and infectious periods on the characteristics of major outbreaks, based on stochastic simulations of three SEIR models (columns of plots) that differ only in the duration of the incubation period, which is half, equal to, and double the duration of the infectious period (10 days). For each of these models, different numbers of stages were used to represent the incubation period (x-axis) and infectious period (different line styles; see legend), thereby determining the dispersion of these distributions. Outbreak size (a-c) was insensitive to the dispersion of the distributions or to the relative durations of the incubation and infectious periods. Less dispersed infectious period distributions resulted in shorter persistence times (d-f) and critical times (g-i), and increased the probability that an outbreak would be major (j-l). Less dispersed incubation period distributions resulted in a slight decrease in persistence times and a slight increase in critical times, and the strength of this effect increased as the duration of the incubation period increased. Increasing the duration of the incubation period had no effect on the proportion of outbreaks that are large. Refer to Methods section for details of the model.

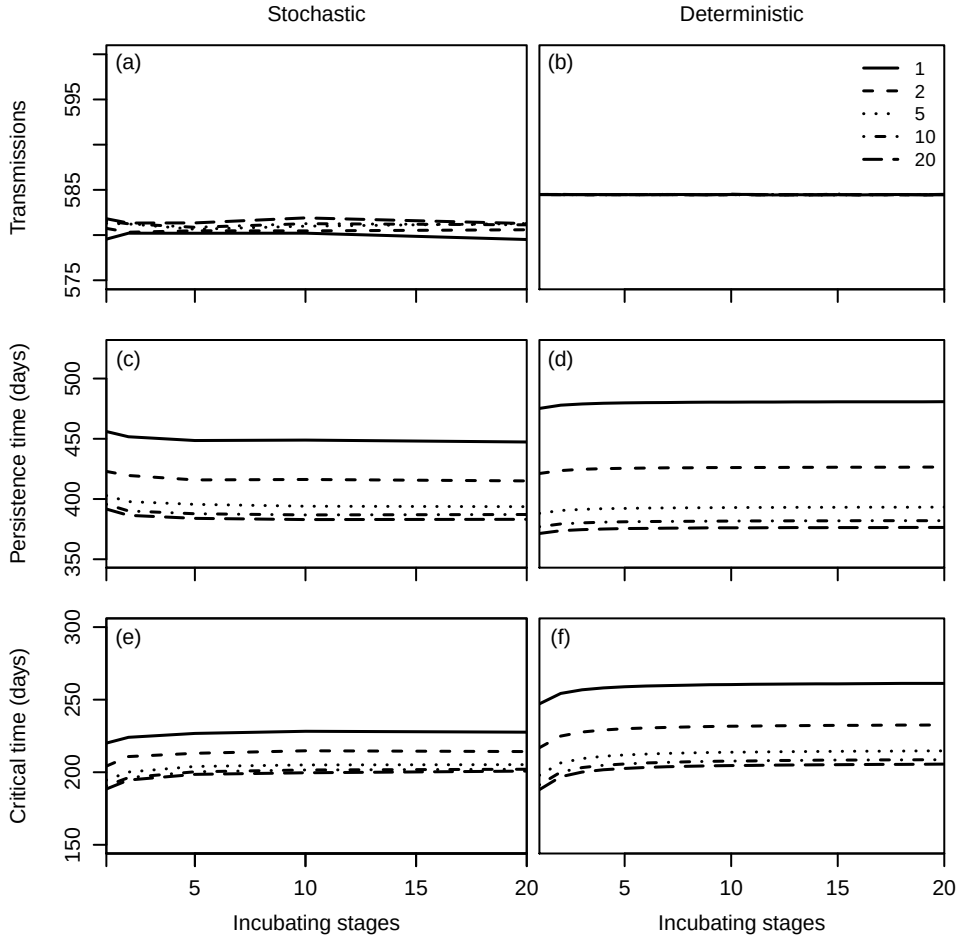

Figure 2.5: Comparison of stochastic and deterministic solutions to an SEIR model with an incubation and infection period of 10 days, and with different numbers of stages used to represent the incubation period (x axis) and infectious period (different line styles; see legend), thereby determining the dispersion of these distributions. The effect of the dispersion of the incubation and infectious periods is quantified with respect to three characteristics of major outbreaks: outbreak size (a, b), persistence time (c, d), and critical time (e, f). Refer to Methods section for details of the model.

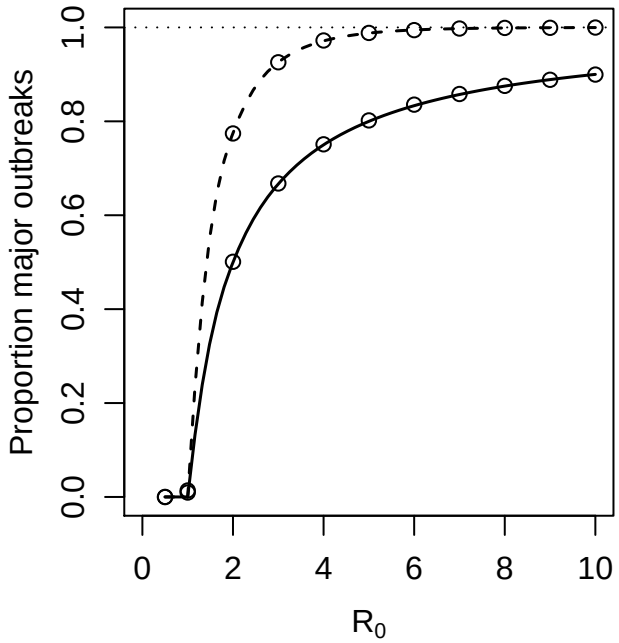

Figure 2.6: The proportion of outbreaks that are major outbreaks as a function of R_0 for a SEIR model with a mean incubation and infectious period of 10 days, and starting conditions of one new infectious individual introduced into a population of 1000 susceptible individuals (see Methods section for details of model structure). The transmission parameter, β, was adjusted to reflect different levels of R_0. The lines represent predictions based on the analytical approximation of Anderson & Watson (1980). The solid line corresponds to the exponential model, and the dashed line is a model with a 20 stage infectious period and a one stage incubation period (the probability an outbreak will be major is not related to the dispersion of the incubation period). The predictions are validated by 50,000 stochastic simulations of the models at various levels of R_0 (open circles), whereby the strong bimodality in outbreak sizes was used to classify each outbreak as minor or major (major outbreaks had at least 200 cases). Relative to the exponential model, less dispersed infectious period distributions increase the probability that an outbreak will be major.

Chapter 3

The importance of realistic distributions and alternative infection states in models of rabies outbreaks

3.1 Abstract

Epidemiological models are increasingly used to identify appropriate management responses to infectious disease outbreaks, inform public policy on disease management in the event of future outbreaks, and design and evaluate control strategies. However, inferences drawn from these models are sensitive to model structure and the simplifying assumptions upon which the model is based. We review the pathological evidence for, and quantify the effect of, hypothesized alternative rabies infection states (long incubation periods, carrier individuals, and recovery and subsequent immunity from infection) on outbreak dynamics described by the outbreak size, persistence time, the peak number of infectious cases, and the critical time. We also examine how realistic infectious period distributions and host population demography affect outbreak dynamics. Alternative infection states and host demography (growth rate) had the strongest effect on outbreak dynamics and, therefore, could have a profound influence on parameter estimation or the estimation of the basic reproductive number.

Keywords: compartment models; stochastic simulation; method of stages; Lyssavirus; long incubation period; carrier state

3.2 Introduction

Rabies virus (RV; genus *Lyssavirus*) is a neuropathogen causing an acute encephalitis that is usually fatal to mammalian hosts (Rupprecht *et al.*, 2002). Rabies exerts a major public health and economic burden: it is responsible for at least 55,000 deaths worldwide (predominantly in Africa and Asia), and expenditure on treatment and control exceeds US$500 million per annum (Coleman *et al.*, 2004; Knobel *et al.*, 2005). RV is endemic to all continents with the exception of Antarctica, with the domestic dog being the primary reservoir in Africa and Asia (Rupprecht *et al.*, 2002; Nel & Markotter, 2007). RV is a multi-host pathogen that infects a wide range of mammals (Hanlon *et al.*, 2007) and is therefore also an important threat to animal populations of conservation concern (Woodroffe, 2001; Haydon *et al.*, 2006; Randall *et al.*, 2006; Cleaveland *et al.*, 2007), but it can be effectively controlled or eliminated by vaccinating hosts (Eisinger & Thulke, 2008; Lembo *et al.*, 2010).

Epidemiological models are frequently used to estimate basic parameters (Anderson & May, 1991), evaluate alternative control strategies (Haydon *et al.*, 1997; Ferguson *et al.*, 2003; Keeling *et al.*, 2003; Haydon *et al.*, 2004; Feng *et al.*, 2009), and set levels for control measures such as vaccination (Coleman & Dye, 1996; Kitala *et al.*, 2002; Haydon *et al.*, 2006), culling (Matthews *et al.*, 2003), or quarantine/isolation (Feng *et al.*, 2007). However, inferences drawn from these models are sensitive to the assumptions upon which they are based. One common approach is to use compartment models that classify the population into discrete epidemiological states representing susceptible, exposed / incubation, infectious, and removed (or recovered) individuals (SEIR models). These models assume that these discrete states are adequate approximations of continuous state changes, that there is homogeneous mixing within the population, and that transition times between states are exponentially distributed.

Although rabies is often modelled using an SEIR framework, three alternative infection states have been hypothesized that may have an important affect on outbreak dynamics. First, that the incubation period can be long, sometimes lasting years (Charlton *et al.*, 1997; Tepsumethanon *et al.*, 2004; Johnson *et al.*, 2008). Rare, long incubation periods may facilitate long-term disease persistence by allowing

time for the susceptible individuals in a population to increase following an outbreak, thereby triggering a new outbreak when the infected individual eventually becomes infectious. Second, RV infections are often assumed to be invariably fatal. While this is generally true when the infection has spread to the central nervous system (CNS), rabies is highly immunogenic and infection may be cleared by an immune response prior to CNS infection (Hooper, 2005). This response could also result in improved immunity to any subsequent exposure to RV. Third, it has been hypothesized that a carrier state is possible whereby an individual intermittently sheds live virus in saliva but without displaying typical clinical signs or suffering the increased mortality normally associated with the infectious stage of the disease (Fekadu, 1975).

Cleaveland & Dye (1995) investigated the influence of these alternative infection states on the long-term endemic equilibrium of disease using deterministic models. Relative to the simple SEIR model, long incubations times and immunity had little impact on the period of the epidemic cycles predicted by these models, but carriers increased the period from approximately 10 to 16 years. All three alternative infection states increased the minimum of the number of infected dogs in the epidemic troughs, implying that long incubation, carriers, and immunity might all reduce the probability of stochastic disease fadeout.

The affect of these alternative infection states on outbreak dynamics, however, is poorly understood. Here, we review the pathological evidence of these alternative infectious stages, and quantify their influence on outbreak dynamics relative to the simple SEIR model using stochastic simulations. Inferences regarding disease dynamics are also sensitive to the dispersion of the incubation (EPD) and infectious (IPD) period distributions (Keeling & Grenfell, 1998; Andersson & Britton, 2000; Lloyd, 2001c; Wearing *et al.*, 2005; Feng *et al.*, 2007, Chapter 2). We, therefore, also quantify the affect of using realistic distributions on outbreak characteristics relative to the exponential model.

3.2.1 Rabies pathology and epidemiology

The primary infection mechanism is transmission of virus in the saliva of an infectious animal to an uninfected animal, usually by means of biting. Although other modes of transmission are possible (e.g. ingestion of infected material, aerosol transmission, exchange of saliva via licking mucous membranes or an open wound) they are considered rare and ineffectual compared to bite transmission (Hanlon *et al.*, 2007; Rupprecht *et al.*, 2002). Following inoculation, the virus enters cells, replicates, and either spreads to adjacent cells or is released into the blood. Although RV is neurotropic and direct entry into the peripheral nervous system is possible (Shankar *et al.*, 1991), it may take a variable amount of time for the virus to first encounter a neuronal cell. There are also differences among strains in the time required to invade the nervous system (Nel & Markotter, 2007).

Once in a neurone the virus moves rapidly through the peripheral nervous system (50-100 mm/day; Tsiang *et al.*, 1991) to the CNS. Infection of organs and other non-nervous tissue subsequently occurs by means of centrifugal dissemination throughout the peripheral nervous system from the CNS (Murphy, 1977; Jackson *et al.*, 1999). Infection of the salivary gland in this way facilitates onward transmission of the virus. Replication of the virus in the brain results in the neurological changes commonly associated with rabies, including increased aggression, high pain tolerance, increased movement rates, gradual paralysis, and hydrophobia (Hanlon *et al.*, 2007). However, there is considerable variability in the range of neurological signs that may be a function of damage to different regions of the brain.

Although RV is highly immunogenic and can be cleared from a host by a normal viral immune response (Hooper, 2005), there are several ways in which the immune system can be evaded. First, while replicating within a cell, and when moving directly between adjacent cells, the virons are not exposed to virus neutralising antibodies (VNA). Thus, the virus may be able to persist in a localised group of infected cells despite an immune response. Furthermore, low concentrations of virus may not trigger an immune response. Second, the nervous system is an immune privileged site as the blood brain barrier prevents or limits the passage of VNA and lymphocytes (Nel & Markotter, 2007). Thus, following CNS invasion, it is less likely

an infection will be cleared, although the permeability of the blood brain barrier and the pathogenicity of the RV strain in the host are important factors in determining this (Baloul & Lafon, 2003; Wang *et al.*, 2005; Roy & Hooper, 2008).

The duration of the incubation stage is highly variable, usually ranging from 2 weeks to several months or even years (Charlton *et al.*, 1997; Tepsumethanon *et al.*, 2004; Johnson *et al.*, 2008). Long incubation periods may be facilitated by intramuscular inoculation whereby RV invades muscle cells, persists and replicates within those cells for long periods of time, and only eventually spreads to the peripheral nervous system at which time the usual pathogenesis resumes (Baer & Cleary, 1972; Charlton *et al.*, 1997). Charlton *et al.* (1997) provide experimental evidence that muscle tissue is the site of delay of progression of infection, and that a limited immune response may follow intramuscular infection. The duration of the infectious period is less variable as the neurological effects typically lead to death within a few days.

Thus, recovery from early infection, immunity, and long incubation periods are repeatedly reported aspects of RV infection. However, the existence of a carrier state has also been hypothesized (Fekadu, 1975; Fekadu *et al.*, 1981; Fekadu, 1991) whereby live virus is shed intermittently over long periods of time in the saliva of apparently healthy individuals that display no clinical signs. Because the animal does not suffer the increased mortality of a typical infection, and being asymptomatic is undetected by humans that might otherwise destroy it, carrier individuals have the potential to infect numerous other individuals over a long period.

Neither serum VNA nor viral RNA in the saliva is evidence of the carrier state as both conditions can arise as the result of a RV infection cleared by an immune response (Zhang *et al.*, 2008). Furthermore, live virus can be shed in saliva before the onset of clinical signs (Fekadu *et al.*, 1982; Rupprecht *et al.*, 2002), so short-term observations of live virus in the absence of clinical signs is not proof of a carrier individual either. The sole indication of the carrier state is the demonstration of live virus in the saliva over long periods of time, although the virus may only be detected intermittently.

The carrier state has only been rarely documented (Fekadu, 1975; Fekadu *et al.*,

1981; Fekadu, 1991), and only in a small number of individuals. Fekadu *et al.* (1981) describe a domestic dog that recovered from rabies infection following experimental intramuscular inoculation, but that shed low concentrations of live virus 42 and 169 days after recovery. The presence of live virus was established when it resulted in fatal rabies infections in mice inoculated intra-cerebrally. Two further studies provide limited evidence of a carrier state. Aghomo & Rupprecht (1990) isolated live virus from 4 of 1500 saliva samples from apparently healthy, unvaccinated domestic dogs distributed over a broad area in southern Nigeria. However, as these dogs were not subsequently monitored it is not known whether they developed clinical signs shortly after sampling, or conversely, whether there was long-term shedding of virus. Furthermore, 4 in 1500 samples is consistent with the incidence of rabies in this region, which is endemic in the domestic dog population. Thus, this study provides weak evidence of a carrier state. East *et al.* (2001) claimed that viral RNA (detected using RT-PCR) in saliva samples and positive VNA titres indicated a carrier state among hyaenas (*Crocuta crocuta*). However, as they did not isolate live virus from saliva we argue this study also provides weak evidence of a carrier state. More recently, Zhang *et al.* (2008) found no evidence of a carrier state among 153 domestic dogs in China.

Furthermore, pathologically, it is not clear how such a state could arise. There is little evidence to suggest that a salivary gland infection is possible without a simultaneous infection in the nervous system. Charlton *et al.* (1983) demonstrated experimentally that spread among salivary gland cells occurs via neuronal connections, with limited direct cell to cell transmission, and that neural networks are necessary for widespread infection of salivary glands. This implies that in carrier individuals infection is either cleared from the CNS after it has infected the salivary glands, where it persists, or that the CNS infection is at a low level and fails to cause fatal encephalitis or clinical signs. It is perhaps possible that the carrier state may only occur with a specific strain of rabies in a specific host, that is localized in one region.

The apparently intermittent shedding of virus reported by Fekadu *et al.* (1981) is consistent with the mechanism proposed by Charlton *et al.* (1997) to account for long

incubation periods in muscle fibres: within cells, where it is protected from VNA, the virus may replicate for long periods of time, only eventually being released following the disintegration of the cell. This suggests a possible alternative explanation of the presence of live virus in animals that have recovered from infection. Virus may persist in some infected cells for a considerable length of time until the cells dies, when low concentrations of live virus might be detected.

3.3 Methods

One way of introducing more realistic event time distributions into stochastic models is the method of stages (Cox & Miller, 1965; Lloyd, 2001c, Chapter 2), in which the incubation and infectious periods (of mean duration σ and α respectively) are divided into m and n discrete, exponentially distributed stages respectively. The incubation and infectious periods are therefore the sum of m and n independent exponential random variables, each having a mean $1/m\sigma$ and $1/n\alpha$ respectively (Anderson & Watson, 1980). Overall, the distribution of time spent in the incubation and infectious states is gamma distributed. As the number of stages increases, the overall mean duration remains the same but the dispersion of the distribution decreases.

We adapted the four compartment models of rabies proposed by Cleaveland & Dye (1995), allowing the incubation and infectious periods to be modelled using a variable number of stages. All of the models include demographic processes (birth rate, b, and natural death rate, d). The first model is a simple SEIR (susceptible, exposed / incubation, infectious, removed) model:

$$dS/dt = bN - \beta SI - dS \tag{3.1}$$

$$dE_1/dt = \beta SI - (m\sigma + d)E_1 \tag{3.2}$$

$$dE_j/dt = m\sigma E_{j-1} - (m\sigma + d)E_j, \quad (j = 2, ..., m) \tag{3.3}$$

$$dI_1/dt = m\sigma E_m - (n\alpha + d)I_1 \tag{3.4}$$

$$dI_j/dt = n\alpha I_{j-1} - (n\alpha + d)I_j, \quad (j = 2, ..., n) \tag{3.5}$$

$$dR/dt = (n\alpha + d)I_j \tag{3.6}$$

where $N = S + E + I$, $E = \sum_{i=1}^{m} E_i$, $I = \sum_{i=1}^{n} I_i$, βSI represents density dependent transmission, and $1/\sigma$ and $1/\alpha$ are the mean duration of the incubation and infectious periods respectively. E_j and I_j are the *j*th stage of the incubation and infectious periods respectively.

The second model adds an alternative, longer incubation period (L) of mean duration $1/\sigma_L$ that is occupied by ϕ_L proportion of infected animals. Thus, equations 3.2-3.5 are replaced with:

$$dE_1/dt = (1 - \phi_L)\beta SI - (m\sigma + d)E_1 \tag{3.7}$$

$$dE_j/dt = m\sigma E_{j-1} - (m\sigma + d)E_j, \quad (j = 2, ..., m) \tag{3.8}$$

$$dL_1/dt = \phi_L \beta SI - (p\sigma_L + d)L_1 \tag{3.9}$$

$$dL_j/dt = p\sigma L_{j-1} - (p\sigma_L + d)L_j, \quad (j = 2, ..., p) \tag{3.10}$$

$$dI_1/dt = m\sigma E_m + p\sigma_L L_p - (n\alpha + d)I_1 \tag{3.11}$$

$$dI_j/dt = n\alpha I_{j-1} - (n\alpha + d)I_j, \quad (j = 2, ..., n) \tag{3.12}$$

where p is the number of stages in the long incubation period (here, $p = m$ in model 2).

In the third model a small proportion of animals (ϕ_C) enter a carrier state (C) whereby they are able to transmit disease but do not suffer from clinical effects and therefore are only removed from the population at the same rate at which animals naturally die (d). However, compared to normally infectious individuals they are less effective at transmitting disease (parameter η) because they shed virus intermittently and do not suffer the behavioural changes such as increased aggression that can facilitate transmission. Thus, equations 3.1-3.6 are replaced with:

$$dS/dt = bN - \beta S(I + \eta C) - dS \tag{3.13}$$

$$dE_1/dt = \beta S(I + \eta C) - (m\sigma + d)E_1 \tag{3.14}$$

$$dE_j/dt = m\sigma E_{j-1} - (m\sigma + d)E_j, \quad (j = 2, ..., m) \tag{3.15}$$

$$dI_1/dt = (1 - \phi_C)m\sigma E_m - (n\alpha + d)I_1 \tag{3.16}$$

$$dI_j/dt = n\alpha I_{j-1} - (n\alpha + d)I_j, \quad (j = 2, ..., n) \tag{3.17}$$

$$dC/dt = \phi_C m\sigma E_m - dC \tag{3.18}$$

$$dR/dt = (n\alpha + d)I_j + dC \tag{3.19}$$

In the fourth model a proportion of animals (ϕ_R) recover from infection and enter a new immune class (M). Unlike non-fatal pathogens, however, recovery from rabies infection occurs prior to the infectious period. Here, equations 3.4-3.6 are replaced by:

$$dI_1/dt = (1 - \phi_R)m\sigma E_m - (n\alpha + d)I_1 \tag{3.20}$$

$$dI_j/dt = n\alpha I_{j-1} - (n\alpha + d)I_j, \quad (j = 2, ..., n) \tag{3.21}$$

$$dM/dt = \phi_R m\sigma E_m - dM \tag{3.22}$$

$$dR/dt = (n\alpha + d)I_j + dM \tag{3.23}$$

For this model $N = S + E + I + M$.

We explored these models in the context of rabies outbreaks among domestic dogs in a small, well mixed population. The demographic and epidemiological parameters were estimated from dog surveys and contact-tracing in Serengeti District (SD), northern Tanzania (Hampson *et al.*, 2009). The birth rate (0.538 dogs/yr) was slightly higher than the natural death rate (0.45 dogs/yr), therefore in the absence of disease the population will, on average, increase through time. The population growth rate was the difference between the birth and death rate ($r = b - d$). In Model 2, 5% of infected animals were characterized by long incubation periods of mean duration 140 days. In Model 3, 0.1% of infectious animals were carriers that

suffered no increased rate of mortality due to disease, but that were only 0.1 times as effective as transmitting disease as normal infectious individuals. In Model 4, 20% of infected individuals recovered and became immune. Initial conditions were $S = 500$, $E_1 = 0$, $I_1 = 1$, but in the case of Model 3 the initial single infectious individual was stochastically assigned to the alternative carrier state with probability ϕ_C. The population size of 500 was selected because it corresponds to typical population sizes of domestic dogs in SD villages, and allows us assess qualitative differences among models.

Samples of the duration of the incubation and infectious periods ($n = 296$ and 237 respectively) were derived from case histories of suspected rabid domestic dogs (Hampson *et al.*, 2009). There were two problems to overcome when fitting a gamma distribution to estimates of the duration of the incubation and infectious periods. First, most estimates were made to the nearest day. This discretisation had a strong influence on the fitted parameter values, particularly the integer shape parameter that represents the number of stages (m or n). Second, some durations were recorded to the nearest week, or over a range of days. For durations that were estimated to the nearest day we assumed a ± 0.5 d error, and for all other point estimate values we assumed $\pm 20\%$ error. This reflects the assumption that the resolution at which the duration is recorded is related to the accuracy of the estimate because recent events (recorded in units of days) are likely to be recalled with greater accuracy than events that occurred weeks or months earlier. We therefore used a constant error for the durations recorded in days, and an error that was proportional to the duration for those durations recorded in weeks or months. For each duration estimate a new value was sampled from a uniform distribution defined by these error limits, and the gamma distribution was fit to this sample using maximum likelihood. This was repeated 1000 times. The maximum likelihood value for the shape parameter was the same for all iterations ($m = 1$ and $n = 3$ for the incubation and infectious periods respectively). The mean duration of the incubation and infectious periods was 22.5 and 3.12 days respectively (Figure 3.1).

It is often not feasible to estimate the value of the transmission parameter (β) empirically because of the difficulty of observing contact between infectious and

susceptible individuals that results in disease transmission. This parameter can, however, be deduced from an estimate of the basic reproductive number (R_0) and a transmission model, which is often based on the assumption of either density or frequency dependent transmission (McCallum *et al.*, 2001). R_0, defined as the average number of secondary cases that are expected to arise from the introduction of a newly infectious individual into an entirely susceptible population (Anderson & May, 1991), was estimated to be approximately $1.1 - 1.2$ among domestic dogs in Serengeti District (Hampson *et al.*, 2009). In this system transmission appears to be neither strictly frequency nor density dependent (Hampson *et al.*, 2009), but for simulating outbreaks in small populations we assume that density dependent transmission is a reasonable simplification.

Under density dependent transmission the rate at which new cases are generated increases in the early stages of the outbreak as the number of infectious cases increases and the number of susceptible individuals remains high. But the depletion of susceptible individuals as an outbreak progresses then results in a reduction in the rate at which new cases arise. Because transmission frequency varies as a function of host density, R_0 will increase as host density increases. It may, therefore, be inappropriate to use a value of R_0 estimated at one density and apply it to a population at a different density. Our goal, however, was to quantify the relative effects of the dispersion of the transition period distributions and of alternative models of infection states on the characteristics of outbreaks while controlling for R_0. The transmission parameter β was therefore adjusted so that R_0 was constant (1.19; Hampson *et al.*, 2009) among all models based on these expressions for R_0:

$$R_0 = \left(\frac{m\sigma}{m\sigma + d}\right)^m \frac{\beta N}{n\alpha + d} \sum_{j=0}^{n-1} \left(\frac{n\alpha}{n\alpha + d}\right)^j \tag{3.24}$$

$$R_0 = \left[(1-\phi_L)\left(\frac{m\sigma}{m\sigma + d}\right)^m + \phi_L \left(\frac{p\sigma_L}{p\sigma_L + d}\right)^p\right] \frac{\beta N}{n\alpha + d} \sum_{j=0}^{n-1} \left(\frac{n\alpha}{n\alpha + d}\right)^j \tag{3.25}$$

$$R_0 = \left(\frac{m\sigma}{m\sigma + d}\right)^m \beta N \left[\frac{\eta\phi_C}{d} + \frac{1-\phi_C}{n\alpha + d} \sum_{j=0}^{n-1} \left(\frac{n\alpha}{n\alpha + d}\right)^j\right] \tag{3.26}$$

$$R_0 = \left(\frac{m\sigma}{m\sigma + d}\right)^m \frac{\beta N(1-\phi_R)}{n\alpha + d} \sum_{j=0}^{n-1} \left(\frac{n\alpha}{n\alpha + d}\right)^j \tag{3.27}$$

for Models 1-4 respectively. These expressions relate R_0 to β while correcting for the natural death of individuals during the incubation or infectious periods. In practice, β varied little among simulations because the natural death rate is long relative to the duration of the incubation and infectious periods. The expression for R_0 for model 1 is based on Feng *et al.* (2007), which was used as a basis for deriving the other three expressions.

The dynamical properties of these models, for both exponential and gamma distributed IPD's, were evaluated using 50,000 stochastic simulations of each model, in a population of 500 susceptible dogs into which a single infectious dog was introduced. Simulations were based on a continuous time Gillespie algorithm (Gillespie, 1976), and were run until fade-out of the disease in the population. Each simulation was characterized by the outbreak size (the total number of cases), the persistence time (time to fade-out), the peak number of infectious cases, and the critical time (the time of the peak of the number of infectious cases). To further quantify the interaction between demography and outbreak dynamics we simulate the model with a gamma distributed IPD at levels of population growth rate ($-0.3 < r < 0.3$ in increments of 0.1) by adjusting the birth rate parameter.

There is a dichotomy in the outcome of stochastic simulations of outbreaks, which may be small and brief (minor) or large and long-lived (major). For larger values of R_0 (e.g. $R_0 > 3$) the bimodal distribution of outbreak sizes do not overlap and it is therefore straightforward to reliably classify outbreaks as minor or major. When R_0 is close to 1, however, the distributions of minor and major outbreak sizes

overlap, making it more difficult to distinguish the two. Here, we define major outbreaks as having more than 122 cases (this cut-off was established using k-means clustering (MacQueen, 1967) of the bimodal frequency distribution of outbreak sizes).

3.4 Results

The characteristics of simulated outbreaks are summarized in Table 3.1. The proportion of outbreaks that were major was similar for models 1, 2 and 4, but approximately 13% lower for the model that included carrier individuals (model 3). Relative to models using exponential IPDs, using gamma IPDs increased the proportion of major outbreaks by 43-49% for all four models.

The mean outbreak size differed little among the four models, with an approximately 7% difference between the smallest (model 1) and largest (models 2 and 3) sizes, and similar outbreak sizes using either exponential or gamma IPDs. However, the frequency of extreme values relative to the simple SEIR model (model 1) was higher for the carrier model (model 3), and to a lesser extent the long incubation time model (model 2), leading to an increase in the variance, skewness and 97.5% quantile values of the outbreak size distribution (Figure 3.2a).

Relative to the simple SEIR model (model 1), the addition of long incubation times (model 2) or carrier individuals (model 3) substantially increased the mean, variance and extreme values in persistence and critical times (Table 3.1). For instance, the mean persistence time was 42% longer for model 2 (Figure 3.2b), and the 97.5% quantile of persistence times was 4.1 yr for model 2 compared to 2.7 yr for model 1. Allowing some animals to recover with subsequent immunity (model 4) slightly reduced persistence and critical times. The predominant effect of using gamma IPDs was to increase the mean persistence and critical times by approximately 8-14% for all four models. The peak number of cases was similar among the four models, but gamma IPDs resulted in a slight decrease in peak cases relative to models using exponential IPDs.

The demographic growth rate (r) had a strong effect on outbreak characteristics.

For all models, the proportion of outbreaks that were major, and the size, persistence times, and peak number of infectious cases of those major outbreaks increased as the population growth rate increased (Figure 3.3a-d). Negative population growth rates resulted in smaller, shorter outbreaks with fewer peak cases of infectious individuals, and reduced the proportion of outbreaks that were major. Increasingly positive population growth rates resulted in non-linear (approximately exponential) increases in persistence and critical times, and outbreak sizes. The greatest discrepancy among the four models was in persistence times, which were consistently 50% higher for models 2 and 3.

3.5 Discussion

The hypothesized alternative long incubation state is supported by both pathological and empirical (e.g. case history) evidence (Charlton *et al.*, 1997; Tepsumethanon *et al.*, 2004; Johnson *et al.*, 2008). However, although long incubation times for rabies are possible, even lasting years, it is difficult to quantify the frequency and duration of these long incubation periods. There is strong sampling bias against long incubation times as experimental studies are unable to monitor individuals indefinitely so observations are censored, and case history reconstructions are less likely to detect transmission events that occurred a long time prior to the appearance of signs. It is not clear, therefore, whether it is more appropriate to model incubation times as a single distribution with high variance (a fat tail), or as two separate distributions representing different pathological processes (direct infection of the peripheral nervous system, or a period of intramuscular incubation that precedes infection of the nervous system). Using a single distribution to represent the incubation period may underestimate the frequency and duration of individuals with long incubation times, and therefore underestimate outbreak persistence times, and outbreak sizes in populations with positive growth rates.

The hypothesized alternative carrier state is unsupported by pathological and empirical evidence (Charlton *et al.*, 1983; Zhang *et al.*, 2008), with the possible exception of a small number of apparently exceptional cases (e.g. Fekadu *et al.*, 1981;

Aghomo & Rupprecht, 1990). If carrier individuals do exist, they may arise so rarely that they are inconsequential from an epidemiological modelling and disease control perspective. Although the inclusion this state can result in substantial changes to disease dynamics, without direct evidence of the existence of the carrier state and an estimate of the frequency of incidence in a host population it is difficult to justify including this state in epidemiological models.

There is experimental evidence that recovery from rabies infection is possible, especially prior to infection of the CNS, and that vaccination provides some immunity to future infection. However, different RV strains can vary considerably in pathogenicity and the degree to which they trigger an immune response in hosts. Thus, the degree of immunity may vary depending on the strain, rather than being a binary immune/susceptible state as modelled here. Also, little is know about how immunity might wane with time, and the rate at which recovered individuals become susceptible again. Further work is needed in this area. However, as natural recovery from exposure to rabies had little effect on outbreak characteristics, even at a 20% recovery rate, it may be of limited significance in epidemiological models.

Outbreak dynamics were strongly influenced by host demography (population growth rate). When estimating parameters or R_0 using trajectory matching methods it is therefore important to take into account the demographic context in which outbreaks occur. Negative growth rates had a dampening effect on outbreak severity, while positive growth rates magnified outbreak severity. Even though R_0 was constant for all our simulations, the rate at which susceptible individuals are recruited to the population resulted in widely different outcomes. This implies that poor estimates of recruitment, or assuming that demography is not important, could result in substantial bias to parameter estimates.

Compared to persistence time, outbreak size data may offer limited resolution for evaluating competing models of infectious states. The differences in outbreak sizes among the alternative infection state models were small (less than 7% relative to the simple SEIR model) whereas the differences in mean persistence times were much larger (up to 42%). Furthermore, there was more variation among models with respect to persistence times than to outbreak sizes (Figure 3.2a, b). However,

although persistence time data would potentially provide greater opportunity to distinguish among competing models, persistence times are difficult to quantify in the field and are sensitive to detection of the first and last cases in an outbreak.

There was 100% fade-out of the disease among all simulations for all models, indicating that none of the models are consistent with long term persistence of infection in the absence of reintroduction from an external (unmodelled) source. However, the model that included a separate class of long incubators (model 2) had substantially longer persistence times and more extreme values than the simple SEIR model (model 1), indicating that this model was more consistent with long-term persistence of rabies. Long incubation times allow the population of susceptible individuals to increase (in cases where there is positive population growth) before the incubation individual becomes infectious. In populations where the recruitment rate of susceptible individuals is high enough, this could potentially fuel cycles of outbreaks and recovery, thereby facilitating long-term persistence of the disease in a population.

The use of realistic distributions (a gamma distributed IPD in this case) increased the probability that an outbreak would be major, but resulted in outbreak sizes that were similar to those predicted by the model with an exponentially distributed IPD. Gamma distributed IPD's also increased the length of persistence and critical times, but these effect sizes were approximately one-quarter the size of the differences among the four infection state models. Previous work has emphasized the importance of the dispersion of the EPD and IPD on estimates of R_0 (Anderson & Watson, 1980; Lloyd, 2001a; Wearing *et al.*, 2005). We argue, however, that for some diseases alternative infection states and population growth rates may have an even more important influence on outbreak dynamics and therefore on the estimation of R_0. This is particularly true for pathogens where R_0 is small (e.g. < 3) and the effects of the dispersion of the incubation and infectious distributions are relatively inconsequential (Chapter 1).

We have examined outbreak dynamics of these models using stochastic simulation, and Cleaveland & Dye (1995) have examined the behaviour of the endemic equilibrium using deterministic approaches. Although there is limited scope for

examining outbreak dynamics using deterministic approaches, using stochastic simulations to explore the behaviour of the endemic equilibrium could provide useful insights into the critical community size that is required to facilitate long-term persistence, and the effect, if any, of alternative infection states on epidemic cycles.

It is important to consider the validity of the simplifying assumptions when interpreting the results of epidemiological models. The mathematically convenient but biologically unrealistic assumption of exponentially distributed infectious periods may introduce important error into control measure targets (Feng *et al.*, 2007). Stochastic simulation outcomes are sensitive to the distributions used, therefore the use of more realistic distributions should be adopted as standard practice. Given that gamma distributed event times can be simulated using the method of stages (Cox & Miller, 1965), and that a gamma distribution with an integer shape parameter (a necessary precursor to using the method of stages) can be fit to empirical incubation and infectious period data, it is straightforward to build realistic distributions into stochastic compartment models.

3.6 Acknowledgements

Financial support for this work was provided by The Leverhulme Trust.

	Exponential models				Gamma models			
	1	2	3	4	1	2	3	4
Proportion major	0.144	0.143	0.125	0.144	0.210	0.213	0.181	0.206
Outbreak size								
mean	239	251	252	244	240	257	252	247
s.d.	60.1	65.7	86.4	63.5	57.5	64.4	83.1	61.8
97.5% quantile	356	389	482	368	354	391	461	371
skewness	0.156	0.433	1.60	0.210	0.164	0.456	1.58	0.257
Persistence time (days)								
mean	571	808	734	536	633	877	792	597
s.d.	168	281	557	166	193	295	543	184
97.5% quantile	984	1493	2350	942	1102	1595	2420	1040
skewness	1.05	1.28	4.06	1.17	1.13	1.22	3.82	1.14
Critical time (days)								
mean	247	291	260	223	277	332	292	255
s.d.	137	196	173	134	154	214	173	149
97.5% quantile	584	806	664	565	666	895	712	622
skewness	1.34	1.90	2.70	1.63	1.41	1.90	1.94	1.40
Peak infectious cases								
mean	9.16	8.88	8.94	8.42	8.77	8.57	8.58	8.06
s.d.	2.17	2.03	2.08	1.96	1.93	1.90	1.91	1.77
97.5% quantile	14	14	14	13	13	13	13	12
skewness	0.702	0.750	0.808	0.755	0.666	0.762	0.850	0.675

Table 3.1: Characteristics of simulated rabies outbreaks based on 50,000 stochastic simulations of four compartment models of rabies using exponentially distributed incubation and infectious periods ("Exponential models") and more realistic gamma distributions ("Gamma models"). The four models correspond to a standard SEIR model (model 1), an SEIR model with an alternative long incubation state (model 2), an SEIR model that includes carriers (model 3), and an SEIR model that allows for recovery and immunity (model 4). The proportion of outbreaks that are major is shown in the first row of values. Subsequent summary statistics are based on major outbreaks only. Outbreak dynamics are summarized by the mean, standard deviation (s.d.), 97.5% quantile values, and skewness of the outbreak size, persistence time (days), critical time (days) and peak number of infectious cases.

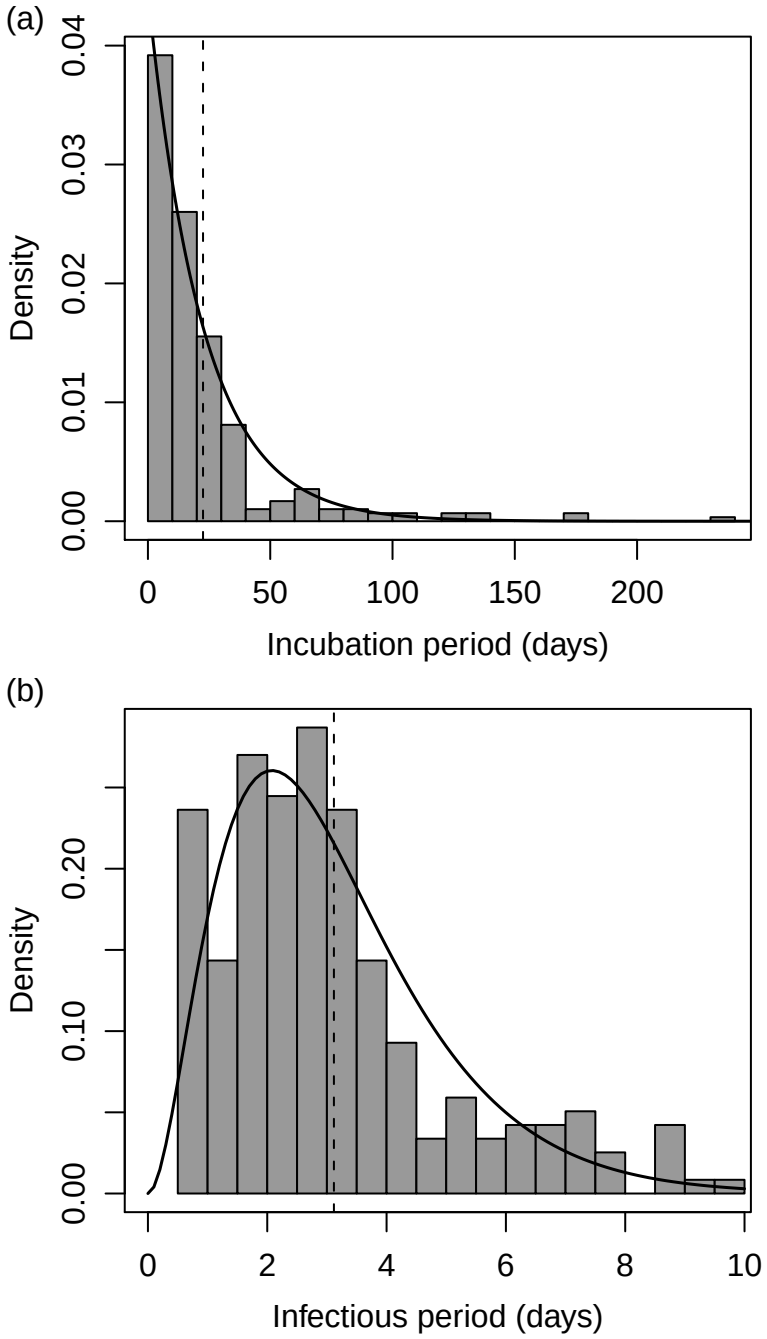

Figure 3.1: Distributions of the durations of incubation (a) and infectious (b) periods of rabid dogs in Serengeti District, Tanzania. Solid lines represent the maximum likelihood fits of gamma distributions with integer shape parameters, and the dashed line indicates the mean of the distributions. The mean duration of the incubation and infectious periods was 22.5 days (s.d. 22.7 d) and 3.12 days (s.d. 1.83 d) respectively.

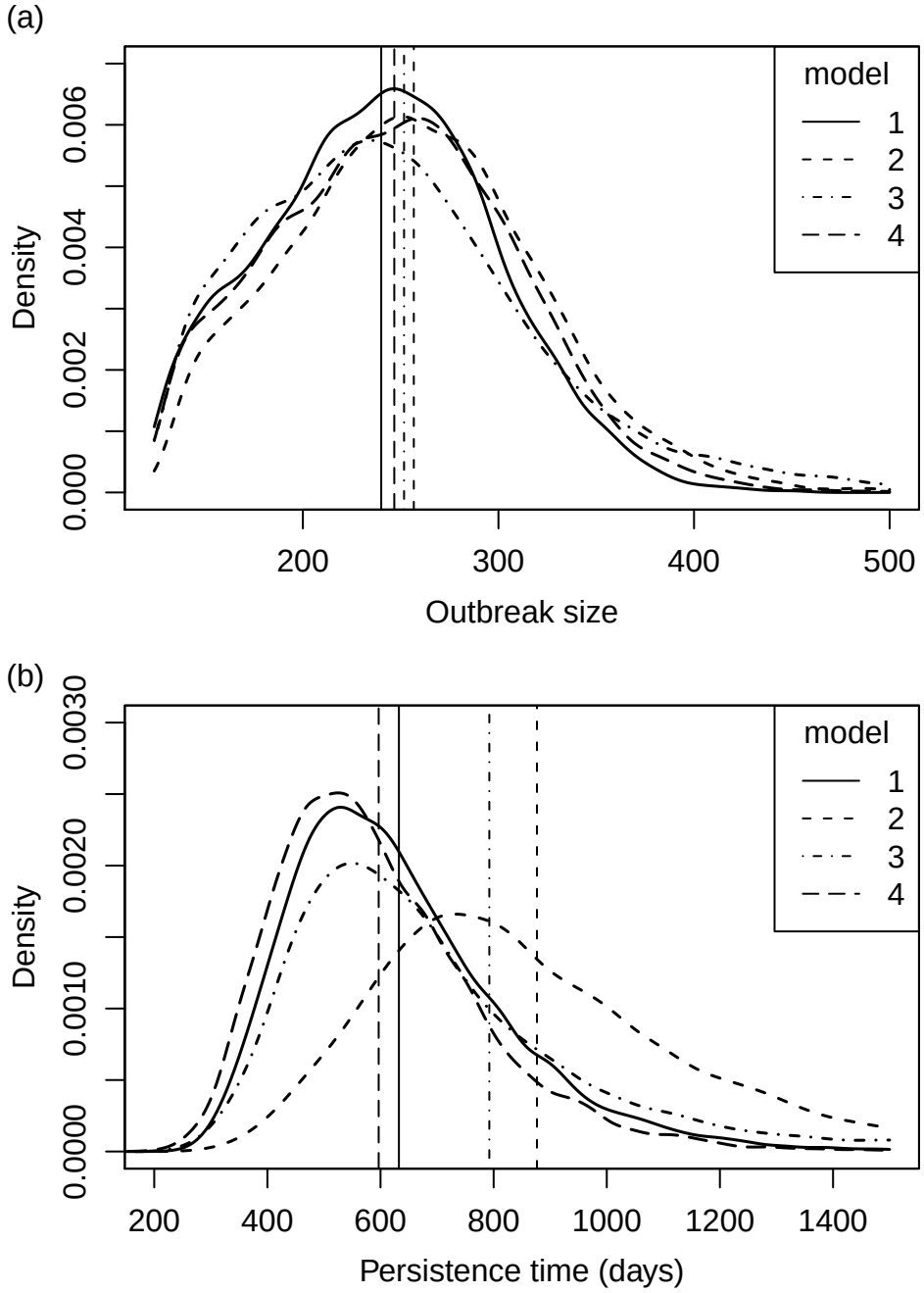

Figure 3.2: Frequency distributions of outbreak size (a) and persistence times (b) of four compartment models of rabies (represented by different line styles in each plot). The four models correspond to a standard SEIR model (model 1), a model with an alternative long incubation state (model 2), a model that includes carrier individuals (model 3), and a model that allows for recovery and immunity (model 4). Vertical lines represent the mean of each distribution. Outbreak dynamics were quantified using 50,000 stochastic simulations of the models with an initial population of 500 susceptible animals into which a single infectious individual is introduced (see Methods for details).

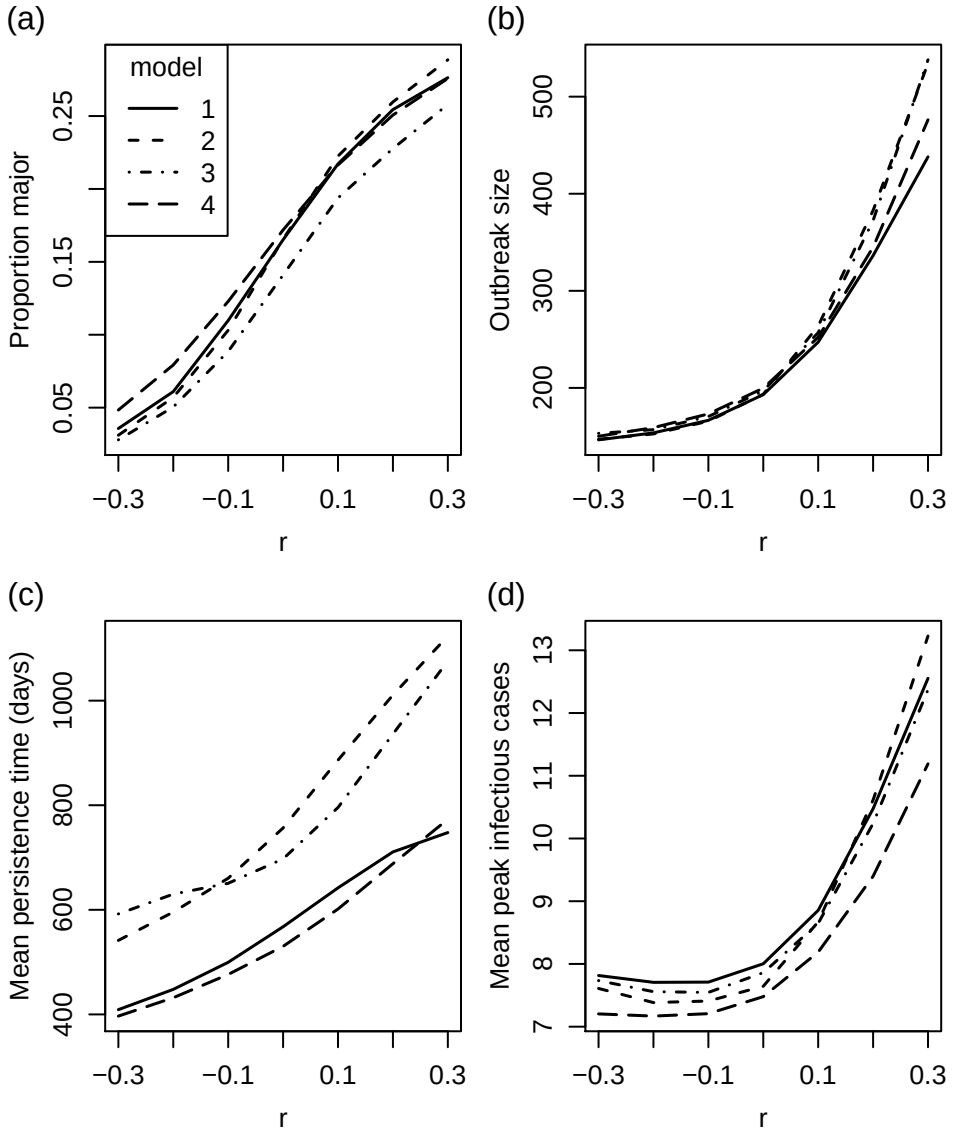

Figure 3.3: The relationship between the host population growth rate and outbreak characteristics of four compartment models of rabies (see Methods for details). The change in outbreak dynamics as a function of population growth rate is summarized by the the proportion of outbreaks that are major (a), the mean outbreak size of major outbreaks (b), the mean persistence time of major outbreaks (c), and the mean peak number of infectious cases for major outbreaks (d).

Chapter 4

Limiting determinants of outbreak size distributions: a case study of canine rabies

> "Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know"
>
> US Defence Secretary Donald Rumsfeld
> Department of Defense, 12 February 2002

4.1 Abstract

For epidemiological models to be useful in designing and evaluating disease control measures (e.g. vaccination) they must be a reasonable approximation of reality. We contrast the observed distribution of rabies outbreak sizes among domestic dogs in Tanzania to predictions from simple compartment models, which predict a bi-modal distribution of outbreak sizes. The large, long-lived (major) outbreaks were absent from observed data but accounted for 84% of simulated rabies incidence, an important discrepancy. We hypothesize outbreak severity may be limited by human intervention reducing transmission rates subsequent to the start of an outbreak, or transmission heterogeneity resulting from host population structure. We use Approximate Bayesian Computation to evaluate competing models that differ in the timing and strength of human intervention, or in population structure and coupling. Both mechanisms reproduced the observed outbreak size distribution, but for the

intervention model this was conditional on a 98% reduction in transmission soon after the onset of cases, which is unrealistic in practice. The highest ranked structure model had numerous small groups, with transmission rates 15 times higher within- versus between-groups. We conclude that even in small populations structure is an important driver of outbreak dynamics, implying the common assumption of homogeneous mixing may not be valid. Local population structuring limits the spread of infection and the size of outbreaks. Including populations structuring into models is likely to be important for accurately evaluating the efficacy of interventions.

Keywords: rabies; outbreak dynamics; vaccination; sequential Monte Carlo; Approximate Bayesian Computation

4.2 Introduction

Epidemiological models are frequently used to estimate basic parameters (Anderson & May, 1991), evaluate alternative control strategies (Haydon *et al.*, 1997; Ferguson *et al.*, 2003; Keeling *et al.*, 2003; Haydon *et al.*, 2004; Feng *et al.*, 2009), and set levels for control measures such as vaccination (Coleman & Dye, 1996; Kitala *et al.*, 2002; Haydon *et al.*, 2006), culling (Matthews *et al.*, 2003), or quarantine/isolation (Feng *et al.*, 2007). Simple models, such as SEIR (susceptible, exposed, infectious, recovered) compartment models, can be used to explore disease dynamics and control programmes in a qualitative manner, although they are often too simplistic to accurately capture the complexity of real epidemics. The degree to which a model is useful for designing and evaluating control measures depends on the extent to which the model is a reasonable approximation of reality. Models that do not fully capture the mechanisms driving transmission dynamics might underestimate the level of control needed to prevent major outbreaks occurring, or might result in inefficient allocation of limited resources by suggesting inappropriate control targets.

Here we critically examine how well simple compartment models describe outbreak dynamics of rabies in domestic dog populations in an east African agro-

pastoralist community. Stochastic simulations of SEIR compartment models result in a dichotomy in outcomes of outbreaks, which may be small and brief (minor) or large and long-lived (major), with relatively little probability of intermediate outbreak sizes. The distribution of outbreak sizes is therefore bimodal, with the relative proportion of minor and major outbreaks depending on the magnitude of the basic reproductive number, R_0, and the distribution of the infectious period (Anderson & Watson, 1980, Chapter 2). 'Major' outbreaks, therefore, refer to the outbreaks that take-off and typically result in infection of 50-100% of the population depending on the model. The impetus for this investigation is the observation that major rabies outbreaks appear to be absent from the host population (Figure 4.1a). Extensive empirical work (infectious case histories) in this system indicates the value of R_0 being around 1.1-1.2 (Hampson *et al.*, 2009). Therefore, based on the most simple SEIR models, we would expect more than 10% of outbreaks would be major (Anderson & Watson, 1980).

We quantify the expected distribution of outbreak sizes based on realistic incubation and infectious period distributions, and estimates of the proportion of the host population that is vaccinated. This model predicts that major outbreaks would occur in this system, which have not been observed in practice. We therefore evaluate two hypotheses that may account for this discrepancy. First, human intervention at the early stages of an outbreak could result in a reduction in transmission rates that would limit the severity of the outbreak. This corresponds to behaviour such as tying up dogs when rabies is known to be present, and killing suspected rabid animals. We develop competing models of the timing of the onset of human intervention and quantify the strength of effect that would be needed to reproduce the observed distribution of outbreak sizes. Second, heterogeneous mixing (structure in the dog population) may serve to limit outbreaks. Although the populations are quite small in the communities we study (mean: 288 dogs per village), the assumption of homogeneous mixing may not be appropriate. We develop competing models of population structure and quantify the relative rates of within and between group transmission needed to generate outbreak distributions similar to those observed. We then further investigate the potential impacts of these

alternative mechanisms by using an independent metric. Comparing vaccination efficacy between observed data and simulations, we identify population structure as an important factor constraining the spread of disease within these populations.

4.2.1 Estimation of R_0

The basic reproductive number is defined as the average number of secondary infections produced by the introduction of a single infectious individual into an entirely susceptible population (Anderson & May, 1991). A variety of approaches can be used to estimate R_0. One of the most direct methods is based on infection histories that document who infects whom and provides a distribution of the number of secondary cases resulting from each infectious case. In the early stages of an epidemic it is reasonable to assume that the population of susceptible individuals is constant, and the mean of this distribution is an empirical estimate of R_0. A related approach is to estimate the intrinsic growth rate of the infected population at the beginning of an outbreak. The expression that relates R_0 to this growth rate is model dependent (see Heffernan *et al.*, 2005). Both of these methods are sensitive to the stochastic variability typical of the early stages of infection, and to reporting inaccuracies (missed cases). They can also only be meaningfully applied to disease for which transmissions can be readily observed, or reconstructed (Heesterbeek & Dietz, 1996), precluding most airborne pathogens.

A more general approach is to use trajectory matching of simulated outbreaks to the observed outbreak (e.g. Wearing *et al.*, 2005). In the case of SEIR models this approach usually requires that the durations of the incubation and infectious periods have been estimated independently. The remaining model parameters (e.g. β) are estimated by minimising the difference between the observed and simulated epidemic trajectories (e.g. using least squares errors), and R_0 can then be derived from these parameters (e.g. $R_0 = \beta/\gamma$ in the case of a simple mass action model).

In populations comprised of discrete, disjoint classes the 'next generation method' (Diekmann *et al.*, 1990) can be used to estimate R_0, defined as the spectral radius of the next generation operator (Heffernan *et al.*, 2005). This technique can therefore be applied to populations with, for instance, age or spatial structure. One issue with

this approach is that it can be difficult to parameterize transmission rates among classes, and that the estimate of R_0 may be sensitive to the way in which continuous variables (e.g. age) have been descretized.

Other approaches include the reconstruction of epidemic trees (Haydon *et al.*, 2003; Hampson *et al.*, 2009), and inferences based on the final size of the epidemic or on data from equilibrium situations (Heesterbeek & Dietz, 1996; Heffernan *et al.*, 2005).

Most of these approaches to estimating R_0 suffer from the problem of qualifying whether a population is entirely susceptible. Immunity of some individuals may arise following vaccination, or naturally following recovery from infection. If a significant proportion of the population is immune the effective reproductive number (R_e) is estimated rather than R_0. There may be a complex relationship between R_e and R_0 as a function of density dependence in transmission rates, and how the immune individuals may dilute the susceptible individuals. Estimating the frequency of natural immunity in a population can be difficult and expensive.

4.3 Methods

4.3.1 Rabies outbreak sizes in domestic dogs

Serengeti District (SD) in northern Tanzania borders wildlife protected areas to the south and east (Serengeti National Park and Ikorongo and Grumeti Game Reserves) and inhabited districts to the north and west (Bunda, Musoma and Tarime). SD consists of 75 villages, inhabited by approximately 174,400 people (Population and Housing Census of Tanzania 2002) in agro-pastoralist communities that use domestic dogs for guarding households and livestock.

The incidence of suspected infectious rabid dogs in each village in each month (Figure 4.2) was quantified from contact tracing using medical records of patients with animal-bite injuries from hospitals and dispensaries, case reports from livestock offices, and community-based surveillance activities (Hampson *et al.*, 2009). The initial number of susceptible dogs in each village in January 2002 was estimated based on the human population size and the average number of dogs per household

in this region (Knobel *et al.*, 2008; Lembo *et al.*, 2008), and the numbers of dogs that were vaccinated. The number of susceptible and vaccinated dogs in subsequent months was modelled as a function of the birth and death rate, the number of dogs vaccinated during vaccination campaigns, and the rate at which vaccination coverage wanes (refer to Hampson *et al.*, 2009, for details).

Failure to observe all cases and the long incubation time of rabies makes it difficult to determine when outbreaks start and end, and which cases should be considered part of the same outbreak. Here we assume that cases within a village that are separated by three or more months of no detected cases constitute different outbreaks. To facilitate comparison of outbreaks sizes from villages with different population sizes, we convert the outbreak size to a proportion by dividing by the estimated number of susceptible dogs at the beginning of the outbreak (Figure 4.1a).

It is likely that not all cases of rabies were detected using these methods. However, only if detection rates are very low ($< 10\%$) would there be a risk of failing to detect large outbreaks that would fundamentally bias our analysis. We conservatively modelled the detection rate of cases to be low, at 50%, and test the sensitivity of our results to this assumption using detection rates of 40% and 60%. We suspect the true detection rate is higher than 50%, which would influence parameter estimates, but only reinforces the conclusion that major outbreaks are absent from the population. We argue that because rabies is a highly visible disease, and the host population is in close contact with the human population, and the local population is educated about this disease, it is unlikely that detection rates could be as low as 50%.

This approach is based on the assumption that movements of infectious dogs among villages are balanced within the course of a single outbreak (no net loss or gain from this movement), and that this movement does not alter the size of the outbreak. This first assumption was necessary in order to distinguish among outbreaks because it is not feasible to trace outbreaks that transition across multiple villages.

Estimates of the duration of the incubation and infectious periods were based on case-histories of infectious dogs obtained from contact tracing (Hampson *et al.*, 2009).

Interviews with local people were used to classify the cause of death as natural, killed by humans, or unknown. Thus we can estimate the reduction in infectious period resulting from human intervention, and the approximate proportion of infectious dogs that are killed.

4.3.2 Expected outbreak size distribution

We quantify the distribution of rabies outbreak sizes predicted by a SEIRV (susceptible, exposed / incubation, infectious, removed, vaccinated) compartment model of a small, well mixed population into which a single infectious individual is introduced. One way of incorporating more realistic event time distributions into stochastic simulations of epidemiological models is the method of stages (Cox & Miller, 1965; Anderson & Watson, 1980; Lloyd, 1996, 2001b, Chapter 1), in which the incubation and infectious periods (of mean duration $1/\sigma$ and $1/\alpha$ respectively) are divided into m and n discrete, exponentially distributed stages respectively. The incubation and infectious periods are therefore the sum of m and n independent exponential random variables, each having a mean $1/m\sigma$ and $1/n\alpha$ respectively (Anderson & Watson, 1980). Overall, the distribution of time spent in the incubation and infectious states is gamma distributed with a shape parameter corresponding to m or n stages. As the number of stages increases, the overall mean duration remains the same but the dispersion of the distribution decreases.

The model dynamics are determined by the following equations that govern the rates of change between the epidemiological states:

$$dS/dt = -\beta SI/N \tag{4.1}$$

$$dE_1/dt = \beta SI/N - m\sigma E_1 \tag{4.2}$$

$$dE_j/dt = m\sigma E_{j-1} - m\sigma E_j, \quad (j = 2, ..., m) \tag{4.3}$$

$$dI_1/dt = m\sigma E_m - n\alpha I_1 \tag{4.4}$$

$$dI_j/dt = n\alpha I_{j-1} - n\alpha I_j, \quad (j = 2, ..., n) \tag{4.5}$$

$$dR_j/dt = n\alpha I_j \tag{4.6}$$

where $N = S+E+I+V$, $E = \sum_{i=1}^{m} E_i$, and $I = \sum_{i=1}^{n} I_i$. E_j and I_j are the jth stage of

the exposed and infectious periods respectively. Frequency dependent transmission ($\beta SI/N$) is used here to allow for a dilution effect of vaccinated individuals, which differ in number among villages. We assume that the waning of vaccination and demographic processes are not important effects in single outbreaks. Therefore, birth and natural death are omitted, and no dynamics of the vaccinated individuals are explicitly included. We use a conservative estimate of $R_0 = 1.14$ (Hampson *et al.*, 2009) to parameterize the model ($\beta = 0.365$).

Most estimates of the duration of the incubation and infectious periods ($n = 296$ and 237 respectively) were made to the nearest day, but the rest were recorded to the nearest week, or over a range of days. This discretisation has a strong influence on the fitted parameter values, particularly the shape parameter (representing m and n). To resolve this, for durations that were estimated to the nearest day we assume a ± 0.5 d error, and for all other values we assume $\pm 20\%$ error. For each recorded duration, a continuous value was sampled from a uniform distribution defined by these error limits, and the gamma distribution was fit to this new sample using maximum likelihood. This was repeated 1000 times. In both cases the maximum likelihood value for the shape parameter was the same for all iterations ($m = 1$ and $n = 3$ for the incubation and infectious periods respectively). The rate parameter was calculated as the mean of the rate parameters for each of the iterations. The mean durations of the incubation ($1/\sigma$) and infectious periods ($1/\alpha$) were 22.5 and 3.12 days respectively (Figure 4.3a, b). This infectious period duration is the estimated mean for the population (Hampson *et al.*, 2009), and therefore includes cases where animals were killed by humans.

For each observed outbreak ($N = 185$) the number of susceptible and vaccinated dogs at the start of the outbreak was used to stochastically simulate outbreaks using a Gillespie algorithm (Gillespie, 1976). The resulting simulated outbreak size was used to estimate the detected outbreak size by sampling from a binomial distribution with a probability of 0.5, corresponding to the estimated detection probability. This process was repeated to generate a sample of 100 outbreak sizes for each observed outbreak. The number of simulated detected infectious dogs was expressed as a proportion of the number of susceptible dogs (Figure 4.1b) to facilitate comparison

with the observed data (Figure 4.1a).

4.3.3 Approximate Bayesian computation

We use approximate Bayesian computation (ABC; Beaumont *et al.*, 2002; Marjoram *et al.*, 2003; Sisson, 2007; Toni *et al.*, 2009) for parameter estimation and model comparison of our stochastic models. ABC has been suggested as an alternative to likelihood methods when likelihood functions are analytically or computationally intractable, whereby the calculation of the likelihood is replaced by a stochastic simulation procedure. We provide a general summary of the ABC method as it applies to all of the models we evaluate.

The simplest implementation of ABC is a rejection algorithm (Pritchard *et al.*, 1999). A set of candidate parameter values, θ^*, are drawn from initial sampling distributions defined by the investigator for each of the random variables in the model. These parameter values are used in the model to simulate a dataset (x^*) that can be compared to the observed dataset (x^0). The difference between the simulated and observed datasets ($\rho(x^0, x^*)$) is quantified using a vector of summary statistics, which are designed to distinguish important differences between these two realisations. If the difference is above a threshold, ϵ, then θ^* is rejected, otherwise it is accepted. This process is repeated, and the accepted values represent a sample from the posterior distribution. The premise of ABC is that the posterior distribution $\pi(\theta|x)$ can be approximated by $\pi(\theta|\rho(x^0, x^*) \leq \epsilon)$ (Toni *et al.*, 2009).

The rejection algorithm has been criticised as being inefficient when the prior and posterior distributions are very different (Marjoram *et al.*, 2003; Toni *et al.*, 2009), which makes this approach impractical if simulations are computationally costly. Marjoram *et al.* (2003) proposed an MCMC-based algorithm to resolve this problem. While it is more efficient (rejection rates are lower), the samples from the posterior distribution are serially correlated, and if the likelihood surface is complex and the proposal mechanism is poor the sampling chains can become stuck in areas of the state space (Sisson, 2007).

Here we use the ABC sequential Monte Carlo (SMC) algorithm (Toni *et al.*, 2009) in which a population of "particles" are sampled from the prior distribution and are

propagated through a series of intermediate distributions using stochastic simulations and an increasingly stringent vector of tolerances ($\epsilon_1, ..., \epsilon_P$) for each sequential population of particles ($1, ..., P$). The final population, θ_P, represents a sample from the conditional posterior distribution $\pi(\theta|\rho(x^0, x^*) \leq \epsilon_P)$. The algorithm is described as follows:

1. Define the initial sampling distributions π, the vector of tolerances $\epsilon_1 ... \epsilon_P$ such that $\epsilon_1 > ... > \epsilon_P \geq 0$, and set population indicator $p = 1$

2. Set particle number $n = 1$

3. If $p = 1$, independently sample proposed parameter vector θ^{**} from π. Otherwise, sample a particle θ^* from the previous population of particles θ_{p-1} with weights w_{p-1}, and perturb the particle using a kernel, K_p, to obtain θ^{**}. If $\pi(\theta^{**}) = 0$, repeat the perturbation until $\pi(\theta^{**}) > 0$

4. Use θ^{**} to simulate a dataset, and calculate the distance, ρ, between the observed and simulated datasets.

5. If $\rho \geq \epsilon_p$ then return to Step 3, otherwise accept the proposed particle ($\theta_p^{(n)} = \theta^{**}$) and set the weight for the particle: if $p = 1$ then $w_p^n = 1$, otherwise

$$w_p^{(n)} = \frac{\pi(\theta_p^{(n)})}{\sum_{j=1}^{N} w_{p-1}^{(j)} K_p(\theta_{p-1}^{(j)}, \theta_p^{(n)})} \tag{4.7}$$

6. If $n < N$, increment n by 1 and return to Step 3.

7. Normalize the weights to sum to 1

8. If $p < P$, increment p by 1, and return to Step 2.

The weights (Equation 4.7) are calculated as the probability of the parameters based on the prior distribution divided by the sum of the product of the previous weights of particles multiplied by a kernel that returns smaller values for parameter values that are further apart. Thus, the kernel penalizes particles that are too close to highly weighted particles in the previous population, ensuring that the variance of the posterior distributions are not underestimated.

The dataset simulated in step 4 is derived from 500 stochastic simulations of one of the competing models (see below), and is a frequency distribution of outbreak sizes that are adjusted to account for the detection probability by sampling a new detected outbreak size from a binomial distribution. Any detected outbreak sizes of zero were dropped from the sample as they represent undetected outbreaks. The distance statistic was calculated as $\rho = \sum_i (S_i - O_i)^2$, where S_i and O_i are the density of simulated (S) and observed (O) detected outbreak sizes (expressed as a proportion of the initial number of susceptible animals), for each bin, i, in a histogram with 0.05 width bins (see Figure 4.1).

A key assumption of the ABC method is that the summary statistic describes the data without loss of important information (i.e. that it is close to sufficient), and that the distance statistic is unbiased and inversely proportional to the likelihood. Summary statistics that are very specific are closer to being sufficient, but result in much higher rejection rates in the ABC algorithm because simulated data is unlikely to match the observed data exactly. Conversely, less specific summary statistics may fail to provide the resolution to adequately discriminate between simulated and observed data, which may result in biased parameter estimates or estimates with high variance. There is therefore a balance that must be found between specificity and rejection rates when designing a summary statistic. Inevitably, this is somewhat subjective, and for this reason it is essential to validate the parameter estimates to ensure that they can reproduce the observed data.

The final population of particles is an estimated sample from the posterior distribution of the parameters. Kernel smoothing is applied to identify the point estimate of the maximum likelihood value of the parameter. Validation of the estimates involved comparing the observed and expected outbreak size distributions based on 50000 stochastic simulations of the highest ranked model using the estimated maximum likelihood parameter values.

4.3.4 Human intervention model

This model is designed to explicitly incorporate two mechanisms by which human intervention might limit the severity of an outbreak. First, our model includes two

infectious states, representing dogs that suffer normal rates of disease mortality (I) and those that are detected by people (H) and therefore suffer increased mortality rates as a result of direct human intervention (people killing infectious dogs). Second, we introduce a parameter κ, that represents the proportional reduction in transmission rates among dogs following detection of rabies. κ is initially 1, but switches to a lower value in the range [0,1] following a trigger event. The model dynamics are determined by the following equations that govern the rates of change between the four epidemiological states:

$$dS/dt = -\kappa\beta SI/N \tag{4.8}$$

$$dE_1/dt = \kappa\beta SI/N - m\sigma E_1 \tag{4.9}$$

$$dE_j/dt = m\sigma E_{j-1} - m\sigma E_j, \quad (j = 2, ..., m) \tag{4.10}$$

$$dI_1/dt = (1-\phi)m\sigma E_m - n\alpha I_1 \tag{4.11}$$

$$dI_j/dt = n\alpha I_{j-1} - n\alpha I_j, \quad (j = 2, ..., n) \tag{4.12}$$

$$dH_1/dt = \phi m\sigma E_m - n\gamma H_1 \tag{4.13}$$

$$dH_j/dt = n\gamma I_{j-1} - n\gamma H_j, \quad (j = 2, ..., n) \tag{4.14}$$

where $E = \sum_{i=1}^{m} E_i$, $I = \sum_{i=1}^{n} I_i$, $N = S + E + I$, and $\beta SI/N$ represents frequency dependent transmission. E_j and I_j are the *j*th stage of the exposed and infectious periods respectively. Assuming human intervention with probability ϕ, the duration of the infectious state is reduced from 3.70 ($1/\alpha$) to 2.75 ($1/\gamma$) days (see above).

The model was initialized with a single, newly infectious dog that was assigned to state H_1 with probability ϕ, and state I_1 with probability $1 - \phi$. This reflects the fact that the initial source animal can also be targeted by people for disease control. Stochastic simulations of the model were performed using a Gillespie algorithm (Gillespie, 1976) and were run until fadeout of the disease. Survey data on the outcomes of infectious dogs was used to estimate the proportion of infectious dogs that are controlled or killed by people ($\phi = 0.555$).

We evaluate five competing models of the trigger that reduces κ, corresponding to different hypotheses about the timing of this event. The trigger occurs at the transition between states: $E \rightarrow H_1$ (Model 1), $H_1 \rightarrow H_2$ (Model 2), $H_2 \rightarrow H_3$

(Model 3), $H_3 \rightarrow$ death (Model 4), and the death of the second detected dog (Model 5), corresponding to models 1-5 respectively. The trigger therefore spans timings ranging from the instant a detected dog first becomes infectious, to the death of the second infectious dog in the system. Thus, we estimate both the timing of the change in contact rates, and the magnitude of the change.

A trivial solution of this problem is to allow β and κ to vary freely, resulting in a system where R_0 is less than 1. We controlled the value of R_0 by describing the combination of β and κ that ensured the empirical and simulated R_0 values were the same ($R_0 = 1.14$). Stochastic simulations of each model (1-5) were used to estimate R_0 at regular intervals in the 2-dimensional parameters space for β and κ. At each interval of κ (0 to 1 in 0.05 increments) a point estimate of the required value of β was made. A third-order polynomial line was fit to this sample using maximum likelihood. Thus, κ was a free parameter with prior $U(0, 1)$ that was estimated by the ABC-SMC algorithm, and β was calculated deterministically based on the polynomial equation. The algorithm was run with 5000 particles, and the distance thresholds for each sequential population of particles were defined as $\epsilon = (384, 192, 96, 48, 24, 12)$. Appropriate threshold values are determined by running trials to define suitable start and end values.

4.3.5 Structured population model

To evaluate the effect that heterogeneous mixing among dogs has on outbreak sizes, we introduce a structured SEIR model in which the dog population is divided into equally sized groups that are hypothesized to correspond to socially and/or spatially mediated groups in the host population. Transmission between dogs is permitted with a susceptible dog in the same group or an immediately adjacent neighbouring group, using a hexagonal grid to identify connections among neighbouring groups (Figure 4.4). The within-group transmission rate, β, is higher than between-group transmission rate, $\mu\beta$, where μ is a scaling parameter ranging from [0,1). This is similar to a metapopulation model (Park *et al.*, 2001), but applied at a smaller scale.

The compartment model for each group, i, is:

$$dS_i/dt = -\beta S_i \left[I_i + \mu \sum_{k,i \neq k}^{G_i} I_k \right] \tag{4.15}$$

$$dE_{i,1}/dt = \beta S_i \left[I_i + \mu \sum_{k,i \neq k}^{G_i} I_k \right] - m\sigma E_{i,1} \tag{4.16}$$

$$dE_{i,j}/dt = m\sigma E_{i,j-1} - m\sigma E_{i,j}, \quad (j = 2, ..., m) \tag{4.17}$$

$$dI_{i,1}/dt = m\sigma E_{i,m} - n\alpha I_{i,1} \tag{4.18}$$

$$dI_{i,j}/dt = n\alpha I_{i,j-1} - n\alpha I_{i,j}, \quad (j = 2, ..., n) \tag{4.19}$$

where $N_i = S_i + E_i + I_i$, $E_i = \sum_{j=1}^{m} E_{i,j}$, and $I_i = \sum_{j=1}^{n} I_j$. In this model we assume transmission is density dependent because group sizes are small and the assumption that transmission is proportional to group size is reasonable. G is the number of neighbouring groups for group i, and will vary according to the specific arrangement of groups (Figure 4.4). $E_{i,j}$ and $I_{i,j}$ are the *j*th stage of the exposed and infectious periods respectively for group *i*.

In this model human-influences on the duration of the infectious period are not explicitly modelled, thus $1/\alpha$ is the mean duration of the infectious period among all dogs (3.12 days), and the duration of the incubation period ($1/\sigma$) remains the same (22.5 days).

In the absence of empirical data on group sizes and contact rates among dogs, we evaluate four models of structuring of the population. The total population size is always 288 dogs, which is the average population size of dogs in SD. The four models we evaluate correspond to the following combinations of the number of groups and group sizes respectively: 6 groups of 48 dogs (Model 6), 12 groups of 24 dogs (Model 7), 24 groups of 12 dogs (Model 8), and 48 groups of 6 dogs (Model 9). This spans a wide range of possible population structures (Figure 4.4). Because we summarize outbreak sizes as the proportion of the total population, the simulation results are insensitive to the initial population size.

The value of R_0 was controlled by describing the combination of β and μ that ensured the empirical and simulated R_0 values were the same ($R_0 = 1.14$). Thus, μ was a free parameter with prior $U(0, 1)$ that was estimated by the ABC-SMC

algorithm, and β was calculated deterministically. The outbreak size distribution for each model was quantified using 500 stochastic simulations of a continuous time Gillespie algorithm (Gillespie, 1976), and were run until fade-out of the disease in the population. The ABC-SMC algorithm was run with 5000 particles, and the distance thresholds for each sequential population of particles were defined as $\epsilon = (768, 384, 192, 96, 48, 24, 12)$. β and μ are free parameters that are estimated by the ABC-SMC algorithm.

4.3.6 Efficacy of vaccination

We contrasted the implications of the simple SEIRV model, and the highest ranking human intervention model and structured population model, on the efficacy of vaccination using 10,000 stochastic simulations of each model at different levels of vaccination coverage. Each of these three models was initialized with a population of 288 susceptible individuals into which a single infectious individual is introduced. The mean outbreak size (expressed as a proportion of the population size) was calculated for each level of vaccination, and efficacy was calculated as the proportional reduction in mean outbreak size relative to outbreak size when no vaccination occurred.

4.4 Results

Unlike the stochastic simulations of the simple SEIRV model, which are characterised by a bimodal distribution of outbreak sizes, no major outbreaks were observed in the real outbreak data (Figure 4.1a, b). Although only 13.4% of outbreaks were major in the simulations, they accounted for 83.7% of the total incidence of rabies.

In the human intervention model, the timing of the trigger determined the strength of the trade-off between κ and β (Figure 4.5a). If the value of κ was reduced by an early trigger (e.g. Models 1 and 2), β was strongly positively correlated with the size of the reduction. If the trigger was later (e.g. Models 4 and 5), κ had no influence on R_0.

The human intervention model with the greatest support included a reduction

in transmission rate triggered when the first detected infectious dog transitioned from state $H_1 \rightarrow H_2$ (Model 2, Figure 4.5b). There was little support for the models with an earlier or later trigger. In the highest ranked model the estimated maximum likelihood parameter values were $\kappa = 0.0169$ (Figure 4.5d) and $\beta = 0.518$. This implies that human intervention would need to reduce the transmission rate among dogs by 98.3% soon after first detection of an infectious dog to prevent major outbreaks from occurring. Stochastic simulations using the estimated maximum likelihood values of β and κ, and in a population with the mean observed population size of 288 susceptible dogs, resulted in a distribution of outbreak sizes that was similar to the observed data (Figure 4.5d).

In the structured model, to maintain a constant R_0 the strength of coupling between groups (μ) decreased as the transmission parameter (β) increased. This trade-off was non-linear, and was more pronounced for the models with smaller, more numerous groups (Figure 4.6a). The structured model with the greatest support was the model with 48 groups of 6 dogs (Figure 4.6b). There was little support for the models with fewer groups. The estimated maximum likelihood values for the highest ranked model were $\mu = 0.0670$ (Figure 4.6c) and $\beta = 0.0633$. Thus, this implies that within-group transmission rates were approximately 15 times higher than between-group transmission rates. Stochastic simulations of the highest rank model with the estimated maximum likelihood values of β and μ, and in a population with the mean observed population size of 288 susceptible dogs, resulted in a distribution of outbreak sizes that also closely matched the observed data (Figure 4.6d). Only 1.1% of simulated outbreaks were larger than the maximum observed outbreak. Larger values of κ than the maximum likelihood value, corresponding to a weaker reduction in β following the trigger event, resulted in outbreak size distributions with a greater proportion of simulated outbreak sizes that were larger than the maximum observed outbreak. For instance, κ values of 0.1,0.3, and 0.5 resulted in outbreak size distributions that predicted 1.9, 4.9 and 9.7% of outbreaks would be larger than the maximum observed outbreaks.

The simple SEIRV model, and the highest ranked human intervention and structured population models, suggest widely different estimates of the efficacy of

vaccination based on the expected change in mean outbreak size following different levels of vaccination (Figure 4.7a, b). Mean outbreak size fell by 81% following 17% vaccination coverage for the simple SEIRV model, and further vaccination resulted in only incremental changes in efficacy. In contrast, outbreak size for the human intervention model was insensitive to vaccination at coverage levels below 83%, while the change in efficacy for the structured model was approximately directly proportional to vaccination coverage levels.

These results were insensitive to changes in the detection probability (Electronic Supplementary Material). For both the human intervention and structured population models, the highest ranked models were the same at all detection levels. For the human intervention model the estimated maximum likelihood values of κ were 0.0283 and 0.0105 for the 40 and 60% detection rates respectively. For the structured population model the estimated maximum likelihood values of μ were 0.0780 and 0.0599 for the 40 and 60% detection rates respectively. Thus, different detection rates resulted in small quantitative, not qualitative, differences to the results.

4.5 Discussion

We have demonstrated that a simple SEIRV compartment model that included realistic incubation and infectious period distributions, and that was based on estimates of the numbers of susceptible and vaccinated dogs in SD villages, predicted that 13.9% of outbreaks would be major (Figure 4.1b). Of 185 observed outbreaks we would, therefore, expect approximately 26 major outbreaks. The absence of any major outbreaks in the observed populations (Figure 4.1a) is an indication that the simple model fails to capture important drivers of outbreak dynamics, and therefore may result in incorrect or biased insights into this system. Major outbreaks would have a disproportionately strong influence on evaluations of control efficacy because they account for over 80% of simulated incidence. Using a simple SEIRV model, similar to the one we have used, to design control programmes and identify target vaccination levels could, therefore, result in incorrect predictions and recommendations. In this case, such models would over-estimate the efficacy of vaccination

because vaccination eliminates major outbreaks in the simulations that do not occur in reality.

We have evaluated two possible explanations for the discrepancy between these observed and simulated outbreak size distributions. An SEI model that incorporates a hypothesized role of human intervention in reducing transmission rates can reproduce the observed distribution of outbreak sizes (Figure 4.5d). However, to do so requires that this intervention result in an approximately 98% reduction in transmission rates among dogs over and above the reduction in outbreak sizes resulting from the killing of infectious dogs, which was already reflected in the mean duration of the infectious period. Furthermore, the model assumes that the change occurs soon after the first detection of an infectious dog, and instantaneously throughout the population. A more gradual change would require an even greater reduction in transmission rates. We suggest that an effect of this magnitude is unrealistic and that there is no empirical or anecdotal evidence of such an effect size. While some owners may restrict the movement of their dogs for short periods of time, we suggest that it is unlikely that this effect size is of the order required to prevent major outbreaks from occurring. Smaller reductions in transmission rates result in outbreak size distributions that predict a higher frequency of simulated outbreaks larger than the largest observed outbreak, and are therefore less plausible than the estimated maximum likelihood value.

The second explanation we investigated was the role of heterogeneous contact rates and structure in the dog population in limiting outbreak sizes. There was greater support for the models with more numerous, smaller groups compared to the models with fewer, larger groups. Approximating what is probably a complex contact network among dogs as a structured population in which transmission rates within groups is higher than with immediately adjacent neighbouring groups was able to reproduce the observed distribution of outbreak sizes (Figure 4.6d). We favour the structured model as the more plausible and parsimonious hypothesis because of the unrealistic effect sizes required in the human intervention model. However, the formulation of our structured model is undoubtedly an oversimplification of reality. Variation in group sizes and connectivity among groups is likely

to influence transmission dynamics. Further fieldwork is required to quantify the structuring of dog populations in different communities.

What are the implications of this work for vaccination programmes designed to control or eliminate rabies? To some degree, both the human intervention model and the structured population model are an improvement over the simple SEIR compartment model as they provide a better phenomenological description of outbreaks. However, while the dynamics of the human intervention model are not sensitive to population size, the dynamics of the structured model change considerably with population size. This affects the estimate of efficacy of vaccination in reducing incidence. Importantly, although these two models were able to reproduce the observed distribution of outbreak sizes, they made considerably different predictions about the efficacy of vaccination (Figure 4.7b). This work demonstrates that different models can have profoundly different implications for the design and assessment of control measures. Developing a better mechanistic understanding of transmission in host populations, especially the role of structure in driving heterogeneity in transmission rates, is essential in order to gauge the risk of major outbreaks and to optimize disease control strategies for particular host populations.

Because we are interested in outbreak dynamics and not the behaviour of the model at the endemic equilibrium, and because our human intervention models include an event-triggered parameter change, deterministic solutions to these models would not have provided useful insight. The SMC-ABC method is a powerful approach for both parameter estimation and model comparison in models for which likelihood functions cannot be developed. It provides a mechanism for identifying correlations among parameters that provides insight into the dynamics of the system, it allows us to incorporate uncertainty in a similar way to state-space models, and it is an effective way of exploring parameter space and complex likelihood surfaces. The danger of ABC methods is that a poorly designed summary statistic could bias parameter estimation, resulting in an estimate that is not similar to the maximum likelihood estimate. Stringent statistics that are close to being sufficient are more likely to yield posterior distributions that are a good reflection of the true distribution, but more stringent summary statistics typically result in increased

processing times. Validation of the parameter estimates through simulation and comparison with observed data, as we have done here, is therefore essential when using ABC methods.

While the effect of host population structure on disease dynamics near the endemic equilibrium has been explored in large populations (Bolker & Grenfell, 1995; Keeling & Gilligan, 2000; Park *et al.*, 2002; Kao *et al.*, 2006), it is often assumed that homogeneous mixing is a reasonable assumption in small populations. This work suggests that heterogeneous mixing may play an important role in outbreak dynamics even in small populations. The domestic dog population we studied is probably structured at several social or spatial levels, all of which may be important factors influencing the spread of infection. While structure at some levels is obvious (e.g. dogs are distributed among villages that form a regional metapopulation), fine scale structure may be difficult to perceive and quantify. Yet it appears that fine scale structure in the dog population may be important in limiting outbreak size. Quantifying the contact structure among dogs and the movement behaviour of rabid dogs would allow us to better establish how structure and disease transmission combine to influence outbreak dynamics. The assumption of a well mixed population and the use of simple SEIRV compartment models is not a good representation of disease dynamics in this system, and is therefore not a good basis for establishing control targets such as the number of animals to vaccinate. By capturing some of the transmission heterogeneity in real populations, structured population models have the potential to be closer approximations of reality than simple unstructured models, and to therefore provide more relevant insight into intervention strategies.

Two important assumptions were required to quantify within-village outbreak sizes: that movements of infectious dogs among villages are balanced within the course of a single outbreak (no net loss or gain from this movement), and that this movement does not alter the size of the outbreak. The first assumption implies that there are concurrent infections in multiple villages, and that movement of dogs among villages occurs frequently enough that they may be balanced. Concurrent infections in villages were common (Figure 4.2), and high resolution epidemiological data based on contact-tracing (Hampson *et al.*, 2009), indicates that approximately

20% of infections cross village boundaries. It is difficult to assess the validity of the second assumption given the absence of relevant quantitative data. However, it is questionable whether these two assumptions are reasonable in this system, in particular because the results obtained are sensitive to under-reporting of outbreak sizes. A better approach would, therefore, be to explicitly account for the spatial distribution of dogs in villages, and the movement rates among villages. Competing spatially explicit SEIR models could be evaluated using the ABC approach to estimate maximum likelihood parameter values and for model comparison. The difficulty with this approach is the much longer processing times that would be required relative to the simpler within-village outbreak models.

4.6 Acknowledgements

Financial support was provided from the Leverhulme Trust, the National Science Foundation and National Institutes of Health (awards DEB0225453 and DEB0513994), the Wellcome Trust and Lincoln Park Zoo. Thanks are due to the Ministries of Health and Social Welfare, and of Livestock Development and Fisheries in Tanzania, TANAPA, TAWIRI, NCA Authority, the Tanzanian Commission for Science and Technology, and National Institute for Medical Research for permissions and collaboration; Intervet for providing vaccines; Frankfurt Zoological Society, Lincoln Park Zoo, Sokoine University of Agriculture, and the Mwanza Veterinary Investigation Centres for technical and logistical support; and the Serengeti Viral Transmission Dynamics team, medical officers and healthworkers, livestock field-officers, paravets, and village officers in Serengeti District. We thank Louise Matthews and Rowland Kao for useful discussions on estimating R_0.

Author contributions. KH, TL, MK, SC: carried out the fieldwork, compiled the empirical data, conducted demographic surveys and assisted with the delivery of vaccination campaigns. All authors discussed the work. DTH and HLB designed the models. HLB wrote the paper.

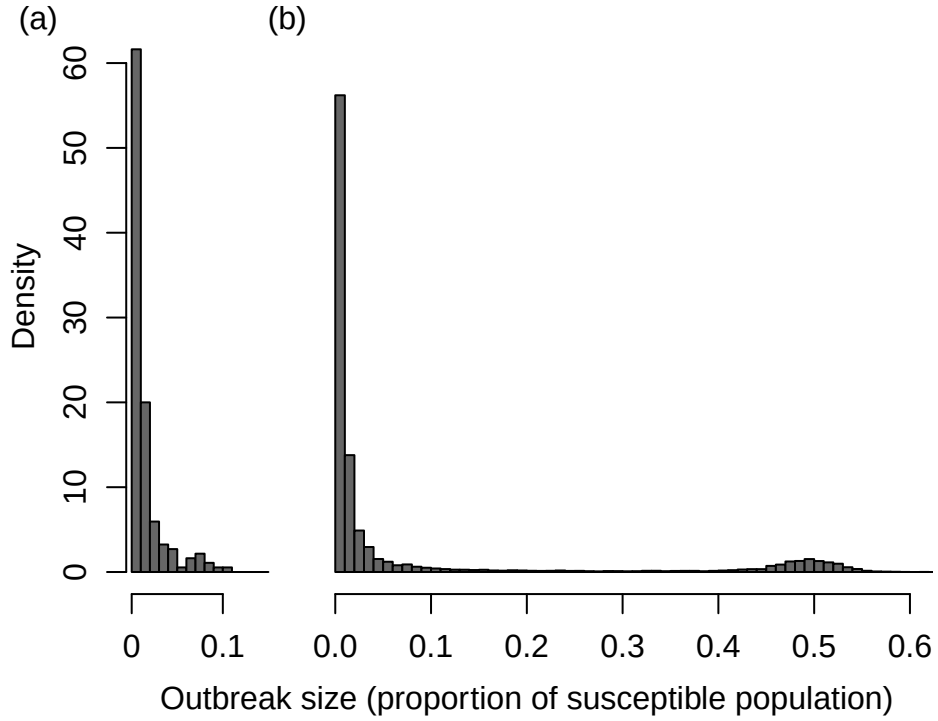

Figure 4.1: Distributions of sizes of observed (a) and simulated (b) outbreaks, expressed as the proportion of the susceptible population that becomes infected during the outbreak. The observed outbreaks are based on monitoring of 75 villages in Serengeti District, Tanzania, over a 5 year period (2002-2006). The simulated outbreaks are based on stochastic simulations of a SEIR compartment model that includes realistic incubation and infectious period distributions, and takes into account the number of susceptible and vaccinated dogs in the observed villages at the start of each outbreak. We adjusted the simulated outbreak distribution assuming a case detection rate of 50%. This was incorporated into the stochastic model to facilitate comparison of the observed and simulated samples (see Methods).

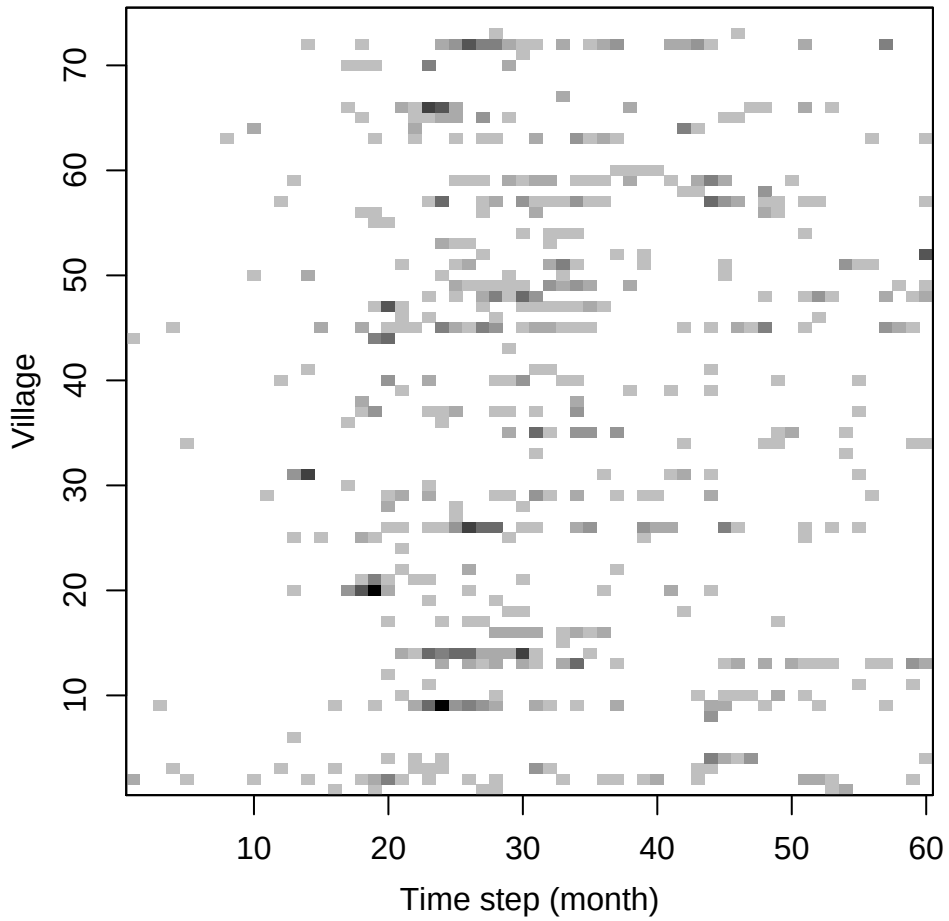

Figure 4.2: Graphical depiction of observed rabies incidence (squares) among the 75 villages in Serengeti District, Tanzania, over a 5 year period (2002-2006). Villages are ordered alphabetically (y axis), and incidence is quantified in one month intervals (x axis). The shading of squares (light grey to black) is proportional to incidence, ranging from 1-10 observed cases per month.

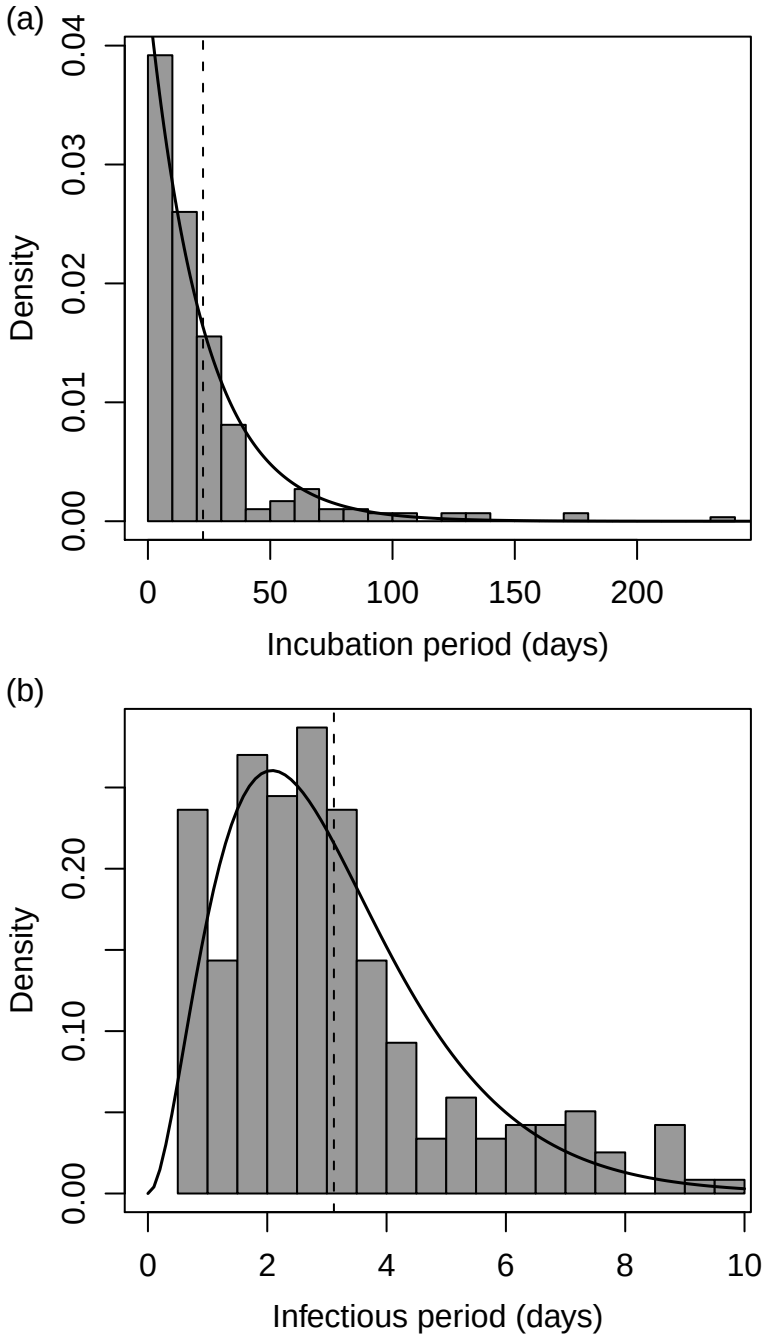

Figure 4.3: Distributions of the durations of incubation (a) and infectious (b) stages of rabid dogs in Serengeti District, Tanzania. Solid lines represent the maximum likelihood fits of gamma distributions with integer shape parameters, and the dashed line indicates the mean of the distributions.

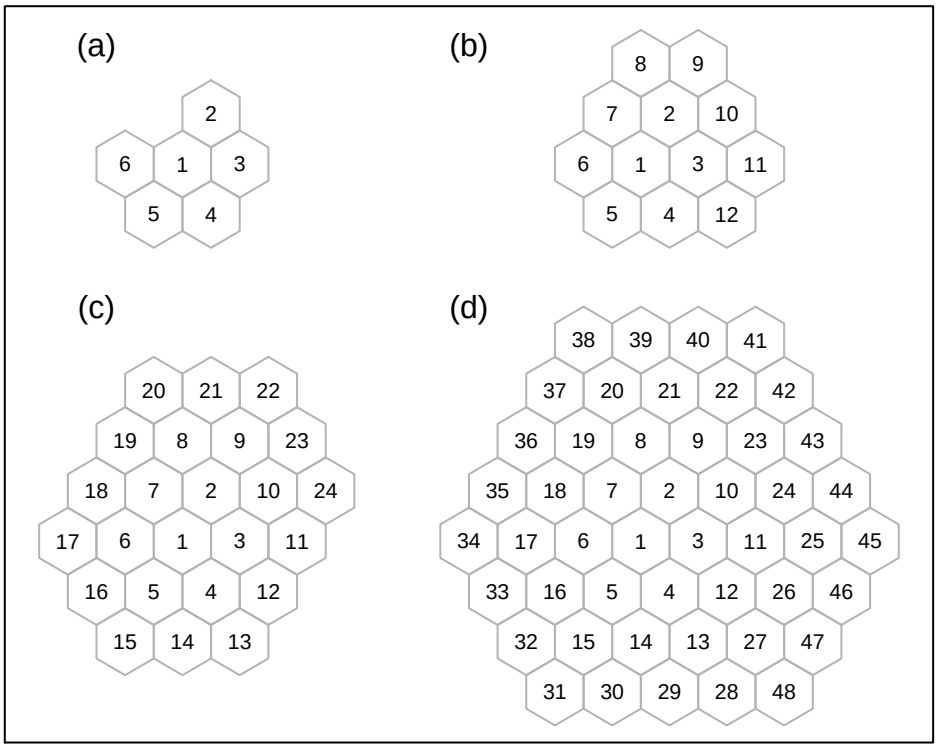

Figure 4.4: Structure in dog populations was imposed using a hexagonal grid to define connections between adjacent groups. The total population of dogs was 288 (the mean of the observed number of susceptible dogs in Serengeti District villages). Four scenarios regarding group size and the number of groups were evaluated: (a) 6 groups of 48 dogs, (b) 12 groups of 24 dogs, (c) 24 groups of 12 dogs, and (d) 48 groups of 6 dogs.

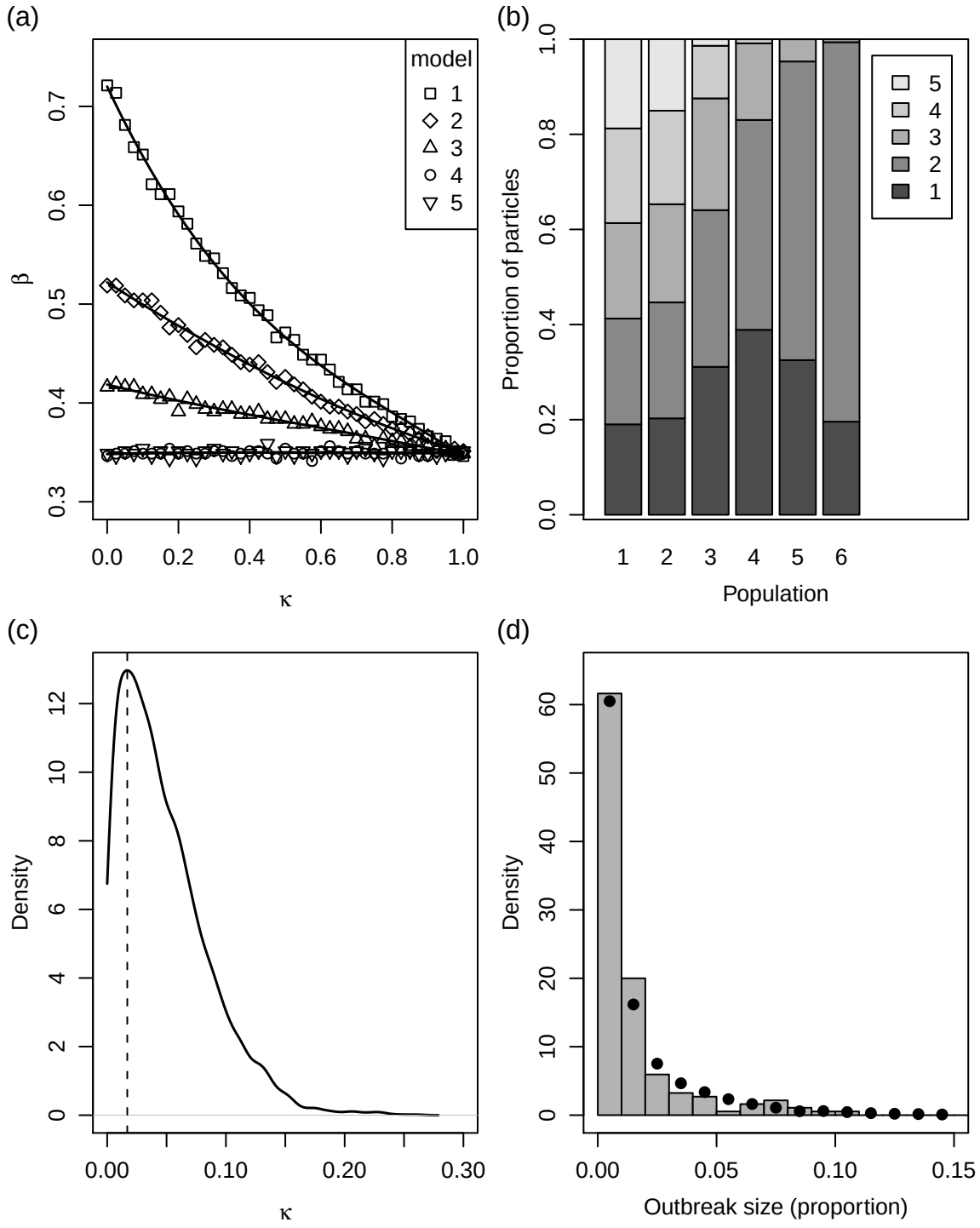

Figure 4.5: The output from the ABC-SMC algorithm for the human intervention models. (a) Lines that describe the values of β and κ that result in an R_0 of 1.14 for each of the 5 models. These lines were fit to point estimates (symbols) at regular intervals of κ, based on stochastic simulations of the models. (b) Proportion of particles associated with each model (1-5) in each of the six populations of particles (x axis). The distance measures between observed and simulated datasets become increasingly stringent in this progression of populations. The relative frequency of particles for each model in the final population is used as an indication of the relative likelihood of the models. The strongest support was found for Model 2. (c) The estimated posterior distribution of κ based on the density of particles in the final population of particles. The dashed line represents the estimated maximum likelihood estimate of κ. (d) Validation of the model was based on 50,000 simulations using these estimated maximum likelihood parameter values. The predictions from the simulated data (black dots) were similar to the observed data (histogram).

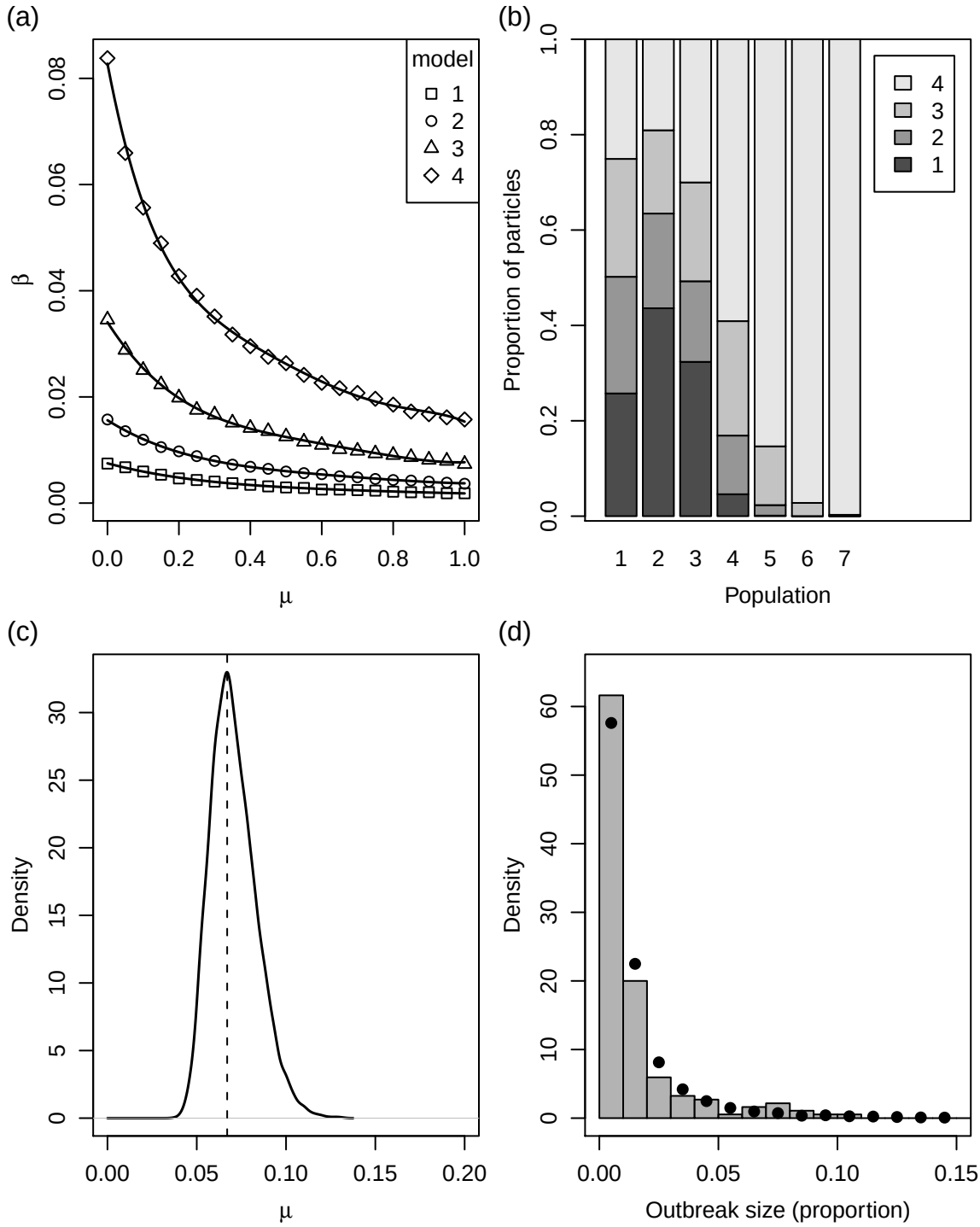

Figure 4.6: The output from the ABC-SMC algorithm for the structured population models. (a) Lines that describe the values of β and μ that result in an R_0 of 1.14 for each of the 4 models. These lines were fit to point estimates (symbols) at regular intervals of μ, based on stochastic simulations of the models. (b) Proportion of particles associated with each model (1-4) in each of the five populations of particles (x axis). The distance measures between observed and simulated datasets become increasingly stringent in this progression of populations. The relative frequency of particles for each model in the final population is used as an indication of the relative likelihood of the models. The strongest support was found for Model 4, followed by Model 3. (c) Plots of parameter values (β, μ) for each particle in the final population of particles, and from the two highest ranked models. The lines represent the 25, 50 and 75% isopleths of the kernel density estimate of these points. The solid black dot represents the estimated maximum likelihood estimate of these parameters. (d) Validation of the model was based on 10,000 simulations using these estimated maximum likelihood parameter values. The predictions from the simulated data (black dots) matched the observed data (histogram) closely.

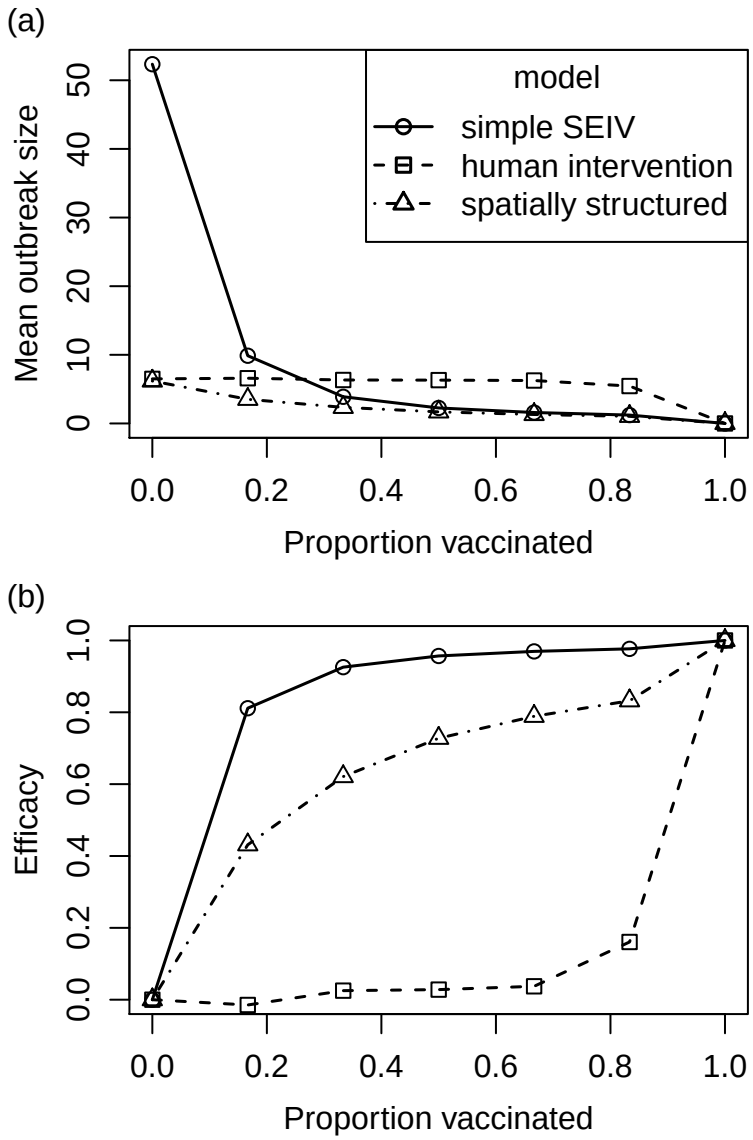

Figure 4.7: Quantifying the efficacy of vaccination using three alternative models: a simple SEIRV compartment model, a model that includes the effects of human intervention, and a model in which the host population is structured (see Methods for details). The change in mean outbreak size based on 10,000 stochastic simulations (a) was used to calculate the efficacy of vaccination in reducing outbreak size (b). The efficacy of vaccination differs substantially among models.

Chapter 5

Metapopulation dynamics of rabies and the efficacy of vaccination

5.1 Abstract

The common assumption in simple epidemiological models that a population is well mixed is often not valid. Spatial structure in a host population results in heterogeneity in transmission dynamics. We used a Bayesian framework to evaluate competing metapopulation models of rabies transmission among domestic dog populations in Serengeti District, northern Tanzania. A proximate indicator of disease, medical records of animal-bite injuries, is used to infer the occurrence (presence / absence) of suspected rabid dog cases in one month intervals. State-space models are used to explore the implications of different levels of reporting probability on model parameter estimates. We find evidence for a relatively high rate of infection of these populations from neighbouring inhabited districts or from other species distributed throughout the study area, rather than from adjacent wildlife protected areas. Stochastic simulation of our highest ranked models in vaccinated and hypothetical unvaccinated populations indicated that pulsed vaccination campaigns (2002-2007) reduced rabies occurrence by 57.3% in vaccinated villages in the one year following each pulse, and that a similar regional campaign would deliver an 80.9% reduction in occurrence. This work demonstrates how a relatively coarse, proximate sentinel of rabies infection is useful for making inferences about spatial disease dynamics and the efficacy of control measures.

Key words: rabies; vaccination efficacy; patch occupancy; state-space model; spatial transmission

5.2 Introduction

Rabies exerts a major public health and economic burden: it is responsible for 55,000 deaths worldwide (predominantly in Africa and Asia), and expenditure on treatment and control exceeds US$500 million per annum (Coleman *et al.*, 2004; Knobel *et al.*, 2005). Although effective post-exposure prophylaxis exists, it is expensive, often scarce, and must be administered shortly after exposure to be effective. Prevention of rabies infection in humans is therefore problematic in developing countries. Yet rabies is a pathogen that can be effectively controlled or eliminated by vaccinating hosts (Eisinger & Thulke, 2008). Rabies virus is a multi-host pathogen that infects a wide range of mammals (Hanlon *et al.*, 2007) and is therefore also an important threat to animal populations of conservation concern (Woodroffe, 2001; Randall *et al.*, 2006; Cleaveland *et al.*, 2007). A single rabies epidemic can eliminate a large proportion of a population (Randall *et al.*, 2004; Haydon *et al.*, 2006). Our interest in rabies control is therefore motivated by both human health and conservation concerns.

Epidemiological models are frequently used to estimate basic parameters (Anderson & May, 1991), evaluate alternative control strategies (Haydon *et al.*, 1997; Ferguson *et al.*, 2003; Keeling *et al.*, 2003; Haydon *et al.*, 2004; Feng *et al.*, 2009), and set levels for control measures such as vaccination (Coleman & Dye, 1996; Kitala *et al.*, 2002; Haydon *et al.*, 2006), culling (Matthews *et al.*, 2003), or quarantine/isolation (Feng *et al.*, 2007). However, many applications of epidemiological models to disease control apply to human or agricultural systems where detailed information about movement, transmission, and host populations is available (e.g. Ferguson *et al.*, 2001; Kao, 2002, 2003; Matthews *et al.*, 2003; Medlock & Galvani, 2009). This quantity and quality of epidemiological data is usually unavailable for diseases in developing regions where formal monitoring, reporting and diagnosis can be ineffectual or

absent (Knobel *et al.*, 2005).

There are two common approaches to resolving this issue. First, theoretical or general models can be used to explore the relative efficacy of different control strategies, and to devise approximate rules for setting control targets (e.g. Coleman & Dye, 1996; Roberts, 1996; Vial *et al.*, 2006; Feng *et al.*, 2009). For instance, one frequently used approximation of the proportion of a population that must be vaccinated to reduce the basic reproductive number, R_0, below 1 is $1 - 1/R_0$ (Anderson & May, 1991). The second approach is to collect epidemiological data, which is often difficult and costly. This approach is typically applied to the development of non-spatial models (e.g. Kitala *et al.*, 2002; Cleaveland & Dye, 1995; Zinsstag *et al.*, 2009) because spatial models require the estimation of more parameters and therefore require more extensive data collection. Also, although valuable, detailed individual-level epidemiological data (e.g. diagnostic tissue testing, sequence data, and case histories) often cannot be collected retrospectively or over large areas.

However, disease dynamics and the efficacy of control measures are influenced by the spatial distribution of the host populations and interventions. Spatial structuring of the host population resulting from social organization (e.g. family groups) or a patchy physical environment (e.g. islands) violates the assumption of many simple models that the population is well mixed. Metapopulation models explicitly model this spatial structure as a system of loosely coupled discrete populations or patches with different rates for within and between patch transmission (Bolker & Grenfell, 1995; Lloyd & May, 1996; Grenfell *et al.*, 2001; Fulford *et al.*, 2002; Cross *et al.*, 2007; Colizza & Vespignani, 2008). Disease persistence in the metapopulation is profoundly influenced by these spatial dynamics (Swinton, 1998; Park *et al.*, 2001). Thus, the promise of spatially explicit epidemiological models is that, because they are locale-specific and capture some of the spatial dynamics of transmission, they allow us to maximize the efficacy of control designs and therefore the deployment of limited control resources.

Our focus is the control of rabies in a multi-host African ecosystem (Serengeti District, Tanzania) in which domestic dogs are thought to be the primary disease reservoir (Lembo *et al.*, 2008) and are therefore the target of control measures (vacc-

ination). This study presents methods for using an existing, indirect measure of disease occurrence (medical records of animal-bite injuries) to parameterize and evaluate competing spatially explicit models of disease occurrence and transmission among dogs at a regional scale, and to quantify the efficacy of a control programme. Although insights into the transmission dynamics of rabies in domestic dogs have been presented previously (Cleaveland & Dye, 1995; Hampson *et al.*, 2009), the spatial dynamics at a more regional scale are still poorly understood. This is important because control measures are often targeted at these larger scales. No predictive models of host-pathogen metapopulation dynamics have yet been developed for this system.

We are also interested in sources of infection of the domestic dog population because of their importance to maintaining a disease-free state. We evaluate evidence for three sources: infected domestic dogs from neighbouring (unmodelled) districts, interactions with wildlife originating from neighbouring wildlife protected areas, and inter-species transmission with other potential hosts (domesticated and wild) that occur throughout the district. Each of these sources results in different testable predictions about the spatial distribution of infections from outside the system.

5.3 Methods

5.3.1 Assessing the occurrence of rabies

This study took place in Serengeti District (SD), northern Tanzania, which borders wildlife protected areas to the south and east (Serengeti National Park and the Ikorongo and Grumeti Game Reserves), and other inhabited districts to the north and west (Bunda, Musoma and Tarime Districts). SD consists of 75 villages (Figure 5.1, and Table D.1 in Appendix D) and is inhabited by approximately 174,400 people (Population and Housing Census of Tanzania 2002) in primarily agro-pastoralist communities that use domestic dogs for guarding households and livestock.

Medical records of patients reporting with animal-bite injuries were collected from local hospitals and medical dispensaries and were used to identify bites from suspected rabid dogs (Cleaveland *et al.*, 2002, 2003; Hampson *et al.*, 2008, 2009). Most

records indicate the date of the bite, the biting animal, and the village from which a patient reported, although this may not always be recorded accurately or represent the location where the person was bitten. There are also several other ways in which medical records may misrepresent actual cases of rabies in domestic dogs: not all rabid dogs bite humans, bite-victims do not always report to hospital, there may be misidentification of whether an animal was really rabid, and it was often not possible to determine if bites occurring close in time were from the same dog. We therefore interpret these bite records conservatively, and explicitly modelled the effects of uncertainty in detection using a state-space modelling approach.

We summarized the data as the occurrence (presence or absence) of exposures by suspected rabid dogs in one-month intervals in each village over a six year period (2002-2007). Occurrence is synonymous with occupancy in patch occupancy models, and is not a measure of the number of infectious dogs (incidence). An occupancy may correspond to the presence of more than one infectious dog in that month. We identified 243 monthly occurrences of rabies among all 75 villages and across all 72 months (Figure 5.2).

5.3.2 Dog demography and vaccination history

The initial number of susceptible dogs in each village in January 2002 was estimated based on the human population size and the average number of dogs per household in this region (Knobel *et al.*, 2008; Lembo *et al.*, 2008), and the numbers of dogs that were vaccinated. The number of susceptible dogs in subsequent months (see Figure 5.3) was modelled as a function of the birth and death rate, the number of dogs vaccinated during vaccination campaigns, and the rate at which vaccination coverage wanes (refer to Hampson *et al.*, 2009, for details).

Following an initial vaccination campaign in 2000 that resulted in low (35-40%) and patchy coverage, subsequent campaigns targeted villages within 10km of the wildlife protected areas (Figure 5.1) and increased coverage levels to between 40-80% (Hampson *et al.*, 2009). Specifically, there were four vaccination campaigns: August 2003 (4179 dogs, 33 villages), June/July 2004 (12975 dogs, 67 villages), Aug/Sep 2005 (7998 dogs, 39 villages), and Aug/Sep 2006 (8030 dogs, 36 villages).

5.3.3 Modelling disease dynamics

Rabies infections have two stages: an incubation period when the animal is infected but not infectious and exhibits no clinical signs, and an invariably fatal infectious period where the animal displays the clinical signs of rabies and can transmit the virus in its saliva to uninfected animals. In domestic dogs in this region the mean duration of the incubation and infectious periods are 22.3 days (95% CI: 20.0-25.0 days) and 3.1 days (95% CI: 2.9-3.4) days respectively (Hampson *et al.*, 2009). We therefore assume that transmission from a village with infectious animals results in incubating animals in the same time period (t), which become infectious animals in the next time period ($t+1$). A one-month time step is also convenient because some medical records can only be used to assign a suspected rabies case to a calendar month. There will be cases where an animal is bitten and becomes infectious in the same month, or long incubation periods that delay the infectious period for more than one month, and these exceptions will add some error to parameter estimates. We believe, however, that the one month time-step between acquiring infection and becoming infectious is applicable to the majority of cases and allows us to capture the important dynamics of the system.

We define $\mathbf{H}$ as the matrix of observed occupancy (1/0) of each village in each of the 72 time periods (one-month time steps), which is a function of the unobserved, true occupancy matrix, $\mathbf{I}$, and the probability of detection of an occurrence (ρ):

$$\mathbf{H}_{i,t} \sim Bern[\rho \mathbf{I}_{i,t}] \tag{5.1}$$

This reflects the fact that if disease was present, it is observed with probability ρ, but if disease was absent it could not have been observed. If reporting is perfect ($\rho = 1$) then $\mathbf{H} = \mathbf{I}$. The probability of detecting infectious dogs based on medical records is influenced by a complex interaction among human social and educational factors, dog behaviour, and the quality of the medical records. Our detection parameter, ρ, encapsulates all of this uncertainty in the simplest possible (one parameter) data model as we have no quantitative basis for constructing and parameterizing a more complex model.

We hypothesize transmission of infection from another village may be mediated by the distance between villages (d), and the population size (S) of susceptible dogs in the villages that could receive and transmit infection. We therefore test the hypotheses that transmission is negatively associated with distance between villages, that larger populations of susceptible dogs are more likely to encounter infectious animals and acquire infection, and that larger populations of susceptible dogs are associated with larger outbreaks and are therefore more likely to transmit infection.

The infectious state, $\mathbf{I}$, of the ith village at time t is modelled using an exponential failure distribution:

$$\mathbf{I}_{i,t} \sim Bern[1 - e^{-c_{i,t}}] \tag{5.2}$$

where c is the hazard rate. The full transmission model is:

$$c_{i,t} = \beta \left(1 - e^{-\delta S_{i,t-1}}\right) \sum_{j}^{V} \left[\mathbf{I}_{j,t-1} e^{-\kappa d_{i,j}} \left(1 - e^{-\psi S_{j,t-1}}\right)\right] + \tau e^{-\mu g_i} \tag{5.3}$$

where V is the total number of villages, $d_{i,j}$ is the Euclidean distance between the centres of ith and jth villages (km), $S_{i,t}$ is the number of susceptible dogs in the ith village at time t. The parameters κ, δ and ψ determine the relative contribution of the distance between villages and the size of the receiving and transmitting village to the probability of acquiring infection, and τ and μ correspond to an external source of infection into this system (see below).

We identify three competing models. Model 1 includes only the distance component:

$$c_{i,t} = \beta \sum_{j}^{V} \left[\mathbf{I}_{j,t-1} e^{-\kappa d_{i,j}}\right] + \tau e^{-\mu g_i}. \tag{5.4}$$

Model 2 includes the distance and size of the receiving village components:

$$c_{i,t} = \beta \left(1 - e^{-\delta S_{i,t-1}}\right) \sum_{j}^{V} \left[\mathbf{I}_{j,t-1} e^{-\kappa d_{i,j}}\right] + \tau e^{-\mu g_i}. \tag{5.5}$$

Model 3 is the full model (Eq. 5.3). We also fit a reference model where β is the

only parameter ($c_{i,t} = \beta$) to gauge the performance of the three competing models. The implication of the reference model is that all infection arises randomly, with no inter-village transmission and no population size effects.

Infections from outside the system are defined by the expression $\tau e^{-\mu g_i}$, where τ is proportional to the rate at which these infections arise, and μ scales τ as a function of distance to the source of infection. There are three hypotheses regarding the source of infectious animals that can trigger outbreaks in the domestic dogs in SD. The source may be spill-over infections from species that occur throughout this landscape, indicated by randomly distributed infections (source "R"). Alternatively, the wildlife protected areas to the east and south (source "P"), or the inhabited adjacent districts to the north and west (source "D"), may be the source of infected animals in which case transmission should be related to proximity to these areas (Figure 5.1). For source R, $\mu = 0$, and τ represents the rate at which a village acquires infection from an external source per time step. For sources P and D, g_i represents the distance to the protected area boundary or nearest adjacent district boundary respectively (km). In these models μ was allowed to vary as a free parameter, and τ was calculated deterministically so that the overall rate of external infection among all villages remained constant. For all sources we evaluated rates of external infections into the system of 2, 6 and 10 infections per year, which we consider to span the range of low to high estimates of the true rate of external infection. This parameter is difficult to estimate empirically, therefore we selected three rates that allow us to make a qualitative assessment of the impact of different rates of external infection on model dynamics.

The exponential terms are a convenient yet versatile form for scaling the influence of distance among villages and the population size of villages on transmission dynamics. Each exponential function requires only a single parameter which facilitates model fitting compared to more complex multi-parameter functions. Negative exponential distributions are appropriate for representing effects that decay, such as the probability of transmission between villages as a function of the distance between them, and have the flexibility to represent very rapid decay, or almost no decay at all. This form can be inverted by taking one minus the negative exponential distribution

which is appropriate for capturing effects that increase up to a saturation point. Again, this form can accommodate patterns that saturate very quickly, or that have almost no effect. Although more complex functional forms could more accurately characterize the influence of distance or population size on transmission dynamics, we have little quantitative basis for selecting or fitting more complex forms. As a first approximation, therefore, the exponential form is appropriate.

Models were fit using WinBUGS (Lunn *et al.*, 2000) using uninformative prior distributions ($\beta \sim U(-4, 0)$, $\kappa \sim U(0, 0.5)$, $\mu \sim U(0, 0.7)$, $\delta \sim U(-8, -2)$, $\psi \sim U(-8, -2)$). For β, δ and ψ the prior distributions were log-transformed so that the posterior distribution was approximately normally distributed, which facilitates the estimation of the effective number of parameters (pD) for model comparison. We generated 37500 samples from the posterior distributions of all parameters using 3 chains, a burn-in period of 200 samples, and a conservative thinning rate of 1 in 50 to ensure the resulting 750 samples were not autocorrelated. Chain convergence was quantified using the $\hat{R}$ statistic (values close to 1.0 indicate convergence). Models were ranked using DIC. However, because the calculation of pD can be unstable (Spiegelhalter *et al.*, 2002), model comparison was based on both DIC and an interpretation of model fit based on changes in deviance in relation to the number of parameters in the model.

Uniform priors were selected because we had no previous data upon which to base informative priors. Although informative priors can help to facilitate chain convergence, that was not an issue encountered in fitting these models. Furthermore, we did not wish subjective assumptions about prior distributions to bias parameter estimates (the posterior distributions). Importantly, we ensured that the bounds of the uniform prior distributions used did not limit how the chains explored parameter space.

We use state-space models to evaluate how reporting error influences parameter estimation in the highest ranked models (processing time constraints prevented us from running all models as state-space models). The transmission component of the models remained the same, but we evaluate reporting probabilities of $\rho = 0.6$ and $\rho = 0.8$.

We use 1,000 stochastic simulations of each of the four highest ranked models to quantify the efficacy of pulsed vaccination campaigns that took place in SD from 2002-2007 (Hampson *et al.*, 2009). Using the mean parameter values for each model, occurrence of rabies was quantified in three scenarios: hypothetical unvaccinated populations, vaccinated populations with a continuing external infection source that corresponds to the vaccination campaign that took place, and hypothetical vaccinated populations with an external infection source that ends 6 months after the first set of vaccinations. The last scenario corresponds to hypothesized vaccination at a regional scale that eliminates cross-district transmission, or that reduces the incidence of rabies in other species that could then infect dogs. For the unvaccinated population scenario, the number of susceptible dogs was estimated based on the vaccination history and demographic parameters (Figure 5.3). The difference in overall disease occurrence (the total number of months in which disease is observed) between the vaccinated and unvaccinated population simulations is a measure of the expected efficacy of vaccination. We measure efficacy at two scales: among the villages targeted for vaccination in the 12 month period following a vaccination campaign, and over the entire district from the first month in which vaccination occurred until the end of the study period.

Simulations were initialized by randomly assigning infections to three villages in the first time step, then running the simulation over a 72 month burn-in period with constant population sizes (this data was discarded) before recording simulated occurrence over the following 72 month period in which population sizes varied as described above. These simulations were performed in R (R Development Core Team, 2009).

5.4 Results

We obtained good chain convergence for all models ($\hat{R} < 1.1$ for all variables in all models). Sampling the prior distributions for β, δ, and ψ on a log-transformed scale was essential for obtaining reasonable estimates of the effective number of parameters (pD).

Of the 28 models tested (Table 5.1), the highest ranked model was the model in which probability of transmission was a function of both inter-village distance and the number of susceptible dogs in the village receiving infection, and where the probability of acquiring an external infection declined as a function of distance to neighbouring districts. However, three other models performed similarly well (ΔDIC < 2 relative to highest ranked model) and therefore also warrant consideration. We infer from these four models that there is strong support for the role of village distances and the size of the village receiving infection in driving transmission dynamics (components of all four top ranked models), but weaker support for the role of the size of the village transmitting infection (a component only of the models ranked third and fourth). The model that included only the village-distance component consistently ranked the lowest, providing further evidence of the importance of population sizes in transmission.

Overall, there was most support for the district-source of external infection (Table 5.1), especially at the lowest rate of infection (2 yr^{-1}). At the higher rates of infection the district and random-source models of external infection performed similarly: although the district model had a lower DIC value in five out of six comparisons (models 1-3, for rates 6 and 10 yr^{-1}), the difference was generally less than 2. We found only weak support for the wildlife protected area source models, which consistently ranked lower than the other source models for each model and rate combination.

There was also strongest support for the highest rate of external infection (Table 5.1). On average, 10 external infections per year would account for 24.7% of all observed occurrences (60 of 243 over the six year study period). However, the inferences regarding the source of external infection and the important components of transmission dynamics were consistent among the three external infection rates (Table 5.1).

Although the highest ranked models have different structures and therefore do not all share the same set of parameters, there was high consistency in parameter values among these models (Table 5.2 and Table D.2 in Appendix D). The implications of the parameter values on the probability of transmission are shown

in Figure 5.4. Although this figure is based on the parameter values of the most complex model, which ranked third, it is representative for all four top ranked models. For all models, the probability of transmission is negatively associated with the distance between villages and positively associated with the population size of susceptible dogs in the village receiving transmission. The probability of acquiring external infection declines with distance from the district boundary in the district-source models. Finally, the population size of susceptible dogs in the village transmitting infection is important only for small populations whereby very small populations (< 150 dogs) have a lower probability of transmission.

Within the set of villages that were targeted for vaccination (all but 5 of the 75 villages in SD) and in the 12 month period following a vaccination campaign, vaccination reduced the occurrence of rabies by 57.3% (59.0, 51.9, 60.0, and 58.1% for the four highest ranked models respectively) relative to the occurrence predicted if no vaccination had occurred (Figure 5.5). Under the alternative assumption that regional-scale vaccination occurred (thereby eliminating the external infection source after 6 months), vaccination reduced the occurrence of rabies by 80.9% (81.7, 83.9, 79.0, and 78.9% for the four highest ranked models respectively) relative to the unvaccinated population. Vaccination also reduced the variance in the size of outbreaks (for instance, the standard deviation in the count of occurrences was reduced from 51.4, 38.6, 54.8, and 51.7 to 17.4, 16.0, 19.1, and 19.4 respectively for each of the four highest ranked models). Over the entire district, and including all months following the first vaccination campaign, vaccination reduced the occurrence of rabies by 50.0% (51.0, 44.9, 52.6, and 51.5% for the four highest ranked models respectively), and assuming regional scale vaccination the occurrence of rabies was reduced by 81.7% (82.2%, 84.1%, 80.3%, and 80.3% respectively).

Explicitly assuming that the reporting probability is only 60% or 80% relative to perfect reporting (100%) resulted in a marginal increase in the estimates of all parameters (Figure 5.6 and Table D.3 in Appendix D). This corresponds to a reduction in the spatial transmission kernel (a reduced probability of transmission over longer inter-village distances), and an increase in the probability of a village receiving infection as population size increases, for all four top ranked models. For

the models with the neighbouring district source of infection there was a decrease in the spatial transmission kernel from that source. Finally, there was also a reduced effect of population size on the probability of transmission from a source village in model 3.

5.5 Discussion

This work demonstrates that a relatively coarse, proximate sentinel (i.e. medical records of animal-bite injuries) of rabies infection can be used to make inferences about spatial transmission dynamics of rabies and the efficacy of control measures. This has important practical implications for identifying drivers of disease transmission, and the design and assessment of control protocols when only limited, indirect epidemiological and demographic data are available. Medical bite records are widely available in Tanzania (and many other countries), and therefore, if they are sufficient to make useful epidemiological inferences, then a great deal of progress could be made using information that is already available without necessarily prioritizing further investment in the acquisition of expensive surveillance data.

In this part of Tanzania, medical bite records appear to be a useful proxy for quantifying the occurrence of rabies at a village scale. Although medical records could also be used to estimate incidence, this requires distinguishing among infectious dogs when multiple, closely-timed bite injuries are reported, which can be difficult. Furthermore, because only a single medical record is needed to infer occurrence, it is less sensitive to low reporting rates than is incidence. Medical records are, therefore, a more accurate indicator of occurrence than incidence. Thus, rather than adopt a traditional individual-based, SEIR (susceptible, exposed, infectious, removed) formulation in which transmission is modelled as a function of the number (or density) of infectious individuals, we adapted a metapopulation dynamics approach (sensu Ovaskainen & Hanski, 2003) to an epidemiological process, whereby the size of patches (villages) in our model is measured by the number of susceptible dogs. A further benefit of this approach is that medical bite records provide limited data for parameterizing an SEIR model, or identifying a suitable transmission model (e.g.

density versus frequency dependent transmission). Simple SEIR models also fail to include the influence of structure in the dog population, or the possible role of human management in limiting outbreaks. By modelling dynamics at a slightly coarser level in this study we avoid the need to explicitly specify fine-scale dynamics that are currently not well understood.

Although there are several examples of the application of epidemiological models to rabies control problems in Africa (e.g. Kitala *et al.*, 2002; Cleaveland & Dye, 1995; Zinsstag *et al.*, 2009), none of them are spatially explicit. While non-spatial models provide approximate rules for control measure targets (e.g. the proportion of a population that must be vaccinated to reduce R_0 below 1 is $1 - 1/R_0$), the efficacy of controls in a specific context (such as Serengeti District) is influenced by the spatial distribution of the host population, of other host populations, and of control measures. The promise of spatially explicit models is the potential to maximize the efficacy of controls in a specific circumstance, thereby optimising the deployment of limited intervention resources. Moreover, these models provide novel insights into the importance of local population size and coupling, and proximity to wildlife host species on disease dynamics. Developing a detailed, more mechanistic understanding of disease dynamics also provides new opportunities for understanding how disease dynamics in different regional contexts may differ.

Spatially explicit models can also provide insight into drivers of regional-scale transmission dynamics. A subjective interpretation of the spatial distribution of rabies occurrences (Figure 5.1), which is highest near the wildlife protected areas, might conclude that infection of the domestic dog reservoir from wildlife in these protected areas was implicated in long-term disease persistence. However, our models demonstrate that, once inter-village transmission dynamics are accounted for, there is only weak evidence of a link between the protected areas and infection of dogs. The apparent proximity of rabies cases to the protected areas may be largely due to the distribution of dogs and villages in that area that results in a "hot-spot" of infection. This may be exacerbated if the protected area boundary encourages higher levels of inter-village movement of people and dogs. Because movement is restricted to the east and south this may concentrate movement among villages to

the north and west resulting in higher levels of inter-village transmission.

There is stronger evidence that the source of infection is the inhabited neighbouring districts, or that a source of infection is distributed randomly throughout the district. Both of these hypotheses are plausible and consistent with previous studies in this region that indicate domestic dogs are the reservoir for rabies (Cleaveland & Dye, 1995; Lembo *et al.*, 2008). Inter-district infection could result from movement of infected domestic dogs, either on foot or in vehicles. Randomly distributed within-district infection could result from inter-specific transmission between several species, e.g. Lembo et al. (2008) report that rabies is found in domestic cats and eight wild carnivore species in that region.

However, it is not clear to what extent other wild and domestic species contribute to disease persistence. Our models indicate that the rate of external infection into the SD dog population may be quite high, although this rate includes transmission from domestic dogs in adjacent districts and is therefore not specific to wildlife. An external infection rate of 10 occupancies per year would, on average, account for 60 of the 243 occupancies observed (24.7%), implying that inter-village transmissions are only four times more common than transmissions from external sources. Lembo *et al.* (2008) estimate that dog to dog transmissions are approximately eight times as common as transmissions between dogs and other carnivores, therefore occasional infection of the dog population from wildlife is plausible. However, given that dogs are the primary reservoir, rabies outbreaks in wildlife that could infect the dog population are likely to have originated from the dog population. Thus, if vaccination reduces the reservoir dog population below the critical threshold required for endemic rabies to persist, then this should also eliminate rabies outbreaks in other species and remove one source of infection of the dog population. Improving our understanding of inter-species transmission rates is a priority for future work.

Another reason the protected areas do not appear to be a significant source

Although the mean incubation period is typically 22.3 days (95% CI: 20.0-25.0 days) (Hampson *et al.*, 2009), incubation periods of months or years are possible, although rare, in mammalian hosts (Lakhanpal & Sharma, 1985). These long incubation times, which violate our assumption of a one-month delay between being

infected and becoming infectious, could account for some of the observed occupancies (Figure 5.2), implying the frequency of transmission events between villages or from the external source may be overestimated in our models. Some dogs may also become infectious in the same month they are infected, but this will have no influence on our estimate of dynamics if this occurs in a month in which occurrence has already been detected. For instance, if an infectious dog bites a susceptible dog that, in turn, becomes infectious in the same village and month as the first dog, then the recorded occurrence captures both animals. Stochastic simulations of individual-based based SEIR (susceptible, exposed / incubating, infectious, removed) models with realistic incubation and infectious period distributions could be used to further assess the sensitivity of these results to the duration of the model time-step.

Within our model specification new occurrences could arise as a result of either within-system transmission (transmission within a single village from one time-step to the next, or between-village transmission) or transmission from the external source of infection. If the parameters associated with these processes are free an identifiability problem can arise as a result of the direct trade-off between these processes. For instance, if the rate of infection arising from the external source was high enough it could account for all of the observed occurrences. The primary symptom of this problem is a lack of MCMC chain convergence. To prevent this problem it was necessary to constrain the parameters associated with one of these two processes. As we were primarily interested in metapopulation transmission dynamics we constrained the parameters associated with the external source of infection such that the rate of infection entering the system was constant for a given model. By evaluating three different rates corresponding to estimated low, medium and high rates of external infection we are able to evaluate the effect of variation in this parameter without it being a truly free parameter. The results obtained are conditional upon the assumptions we have made about the rate of external infection. If our representation of that process is grossly incorrect there is the possibility that it could bias the estimates of other parameters in the model. We have argued, however, that the three rates we evaluated are reasonable in this system, and in lieu of quantitative data that could be used to better parameterize this component of the

model the pragmatic approach we have taken is reasonable.

Assessing the efficacy of vaccination is not straightforward because disease transmission is a stochastic process that can result in highly variable spatial and temporal patterns of occurrence. Field observations of occurrence before and after vaccination provides an important measure of the realized efficacy, but this measure is based on only a single realization of a stochastic process and therefore may not be a good representation of the efficacy that would be expected in general. Our approach, using stochastic simulations of the vaccinated population and a hypothetical, unvaccinated population, provides an estimate of the expected efficacy resulting from the pulsed vaccination campaigns that took place between 2002-2007. This measure of expected efficacy may be more relevant when planning future interventions as it describes the expected mean reduction in occurrence resulting from the vaccination campaigns.

Although we found that the four vaccination campaigns between 2002-2007 resulted in a 57.3% (or 80.9% assuming a regional-scale vaccination programme) decrease in our measure of occurrence, it is important to recognize that the reduction in incidence will be greater than this. Mean outbreak size is positively associated with the number of susceptible dogs (Hampson *et al.*, 2009), therefore occurrence in the unvaccinated populations is likely to correspond to a larger number of infectious dogs than occurrence in the vaccinated populations. This non-linear relationship between our measure of occurrence and outbreak size implies that the estimate of the efficacy of the vaccination campaigns would be higher if we were able to monitor incidence at the individual animal level. For instance, Cleaveland *et al.* (2003) estimate that vaccination campaigns in SD in the decade prior to this study reduced the incidence of rabies by approximately 90% based on the incidence of bites of humans by suspected rabid dogs.

These models suggest a potentially complex relationship between vaccination coverage levels and the reduction in disease occurrence. As expected, we found strong evidence that the population size of susceptible dogs was an important predictor of the probability of transmission: smaller populations were less likely to acquire infection, and this effect was approximately linear. However, we also

found support for a strong reduction in the probability of transmission in small populations (fewer than 150 dogs), indicating a possible threshold beyond which vaccination may have increasing benefits. A possible explanation of this effect is that the density of susceptible dogs may become so low in these highly vaccinated populations that fade-out of the disease becomes increasingly likely. Alternatively, this effect could result from human social factors that might vary as a function of population size. Understanding this effect warrants further investigation as it has the potential to be exploited to improve disease management.

The most ambitious zoonotic disease control programmes aim to achieve disease eradication at regional scales. Although we have used our four highest ranked models to quantify the efficacy of the pulsed vaccination campaigns (2002-2007) there is clearly scope to apply them to optimise the design of vaccination programmes in metapopulations (Asano *et al.*, 2008), to design responses to subsequent disease outbreaks in disease-free populations, and to predict what the large scale implications of intervention actions might be. The application of metapopulation models to inform management decisions has the potential to increase both the efficacy and cost-effectiveness of control and eradication programmes.

State-space models provided a rigorous method for quantifying the effect of measurement error on parameter estimates and model inferences. Because processing times were considerable, we evaluated the influence of measurement error only on the highest ranked models and suggest this approach provided a reasonable trade-off between expediency and confidence in inferences. Metapopulation models provide a powerful framework for investigating disease dynamics in spatially structured populations, and for evaluating the efficacy of control strategies. This work demonstrates that these powerful models can be developed based on proximate measures of disease occurrence when more specific and detailed epidemiological data is unavailable.

5.6 Acknowledgements

Financial support was provided from the Leverhulme Trust, the National Science Foundation and National Institutes of Health (awards DEB0225453 and DEB0513994), the Wellcome Trust and Lincoln Park Zoo. Thanks are due to the Ministries of Health and Social Welfare, and of Livestock Development and Fisheries in Tanzania, TANAPA, TAWIRI, NCA Authority, the Tanzanian Commission for Science and Technology, and National Institute for Medical Research for permissions and collaboration; Intervet for providing vaccines; Frankfurt Zoological Society, Lincoln Park Zoo, Sokoine University of Agriculture, and the Mwanza Veterinary Investigation Centres for technical and logistical support; and the Serengeti Viral Transmission Dynamics team, medical officers and healthworkers, livestock field-officers, paravets, and village officers in Serengeti District. We thank Jonathan Dushoff, Louise Matthews, Hans Heesterbeek, and two anonymous reviewers for their constructive comments.

Author contributions. KH, TL, MK, SC: carried out the fieldwork, compiled the empirical data (including collection of hospital records), conducted demographic surveys and assisted with the delivery of vaccination campaigns. All authors discussed the work. DTH and HLB designed the model. HLB wrote the paper.

Table 5.1: Summary of competing patch occupancy models, the deviance, the number of parameters in the model, the effective number of parameters (pD), the deviance information criteria (DIC) value, and the difference in DIC value relative to the highest ranked model (Δ DIC). Probability of transmission is modelled as a function of the distance between villages (model 1), and the number of susceptible dogs in the receiving village (model 2), and the number of susceptible dogs in the village from which infection was transmitted (model 3). These models include an external source of infection which is random in space (R), or arises from wildlife protected areas (P) or inhabited districts (D) adjacent to Serengeti District. The rate of external infection from these sources is fixed at 2, 6, or 10 yr^{-1} among all villages. A simple reference model in which all infection arises randomly from an external source (model 0) is included to gauge the performance of the other models. Models are ranked by the Δ DIC values.

model	source	rate	deviance	parameters (free/fixed)	pD	DIC	Δ DIC
2	D	10	1836	5 (3, 2)	3.32	1839	0.00
3	D	10	1836	6 (4, 2)	3.67	1840	0.69
2	R	10	1837	4 (3, 1)	2.44	1840	0.71
3	R	10	1838	5 (4, 1)	2.67	1840	1.22
2	P	10	1840	5 (3, 2)	3.08	1843	3.95
3	P	10	1840	6 (4, 2)	3.33	1843	4.50
2	D	6	1847	5 (3, 2)	3.29	1850	11.4
3	D	6	1847	6 (4, 2)	3.53	1851	11.9
2	R	6	1850	4 (3, 1)	2.44	1852	13.5
3	R	6	1850	5 (4, 1)	2.66	1853	14.0
2	P	6	1852	5 (3, 2)	3.23	1855	16.4
3	P	6	1852	6 (4, 2)	3.56	1856	17.1
2	D	2	1881	5 (3, 2)	3.32	1884	45.1
3	D	2	1881	6 (4, 2)	3.58	1885	45.6
2	R	2	1885	4 (3, 1)	2.31	1887	48.2
3	R	2	1885	5 (4, 1)	2.65	1888	48.9
2	P	2	1886	5 (3, 2)	3.16	1889	50.1
3	P	2	1886	6 (4, 2)	3.60	1890	51.1
1	R	10	1896	3 (2, 1)	1.88	1897	61.8
1	D	10	1895	4 (2, 2)	2.96	1898	62.9
1	P	10	1900	4 (2, 2)	2.84	1902	66.9
1	D	6	1911	4 (2, 2)	2.90	1914	78.2
1	R	6	1912	3 (2, 1)	1.91	1914	78.4
1	P	6	1915	4 (2, 2)	2.71	1917	81.8
1	D	2	1947	4 (2, 2)	2.86	1950	114.0
1	R	2	1950	3 (2, 1)	1.95	1952	116.0
1	P	2	1951	4 (2, 2)	2.81	1954	118.2
0	R	0	1958	1 (1, 0)	0.97	1959	123.2

Table 5.2: Estimated mean parameter values and 95% credible intervals for the four highest ranked rabies disease transmission models. A dash represents a model in which that parameter was omitted. (The full table is provided in Appendix D.2.)

model	infection source	DIC	$\beta \times 10^{-1}$ (95% CI)	$\kappa \times 10^{-2}$ (95% CI)	$\delta \times 10^{-3}$ (95% CI)	$\psi \times 10^{-1}$ (95% CI)	$\mu \times 10^{-1}$ (95% CI)
2	R	1840	2.11 (0.923, 4.31)	5.28 (2.89, 7.75)	0.608 (0.345, 1.49)	-	-
2	D	1839	2.14 (0.936, 4.46)	5.23 (3.06, 7.76)	0.603 (0.344, 1.47)	-	0.365 (0.0858, 0.734)
3	R	1840	2.15 (0.889, 4.55)	5.25 (2.90, 7.72)	0.607 (0.343, 1.63)	0.319 (0.0604, 1.21)	-
3	D	1840	2.19 (0.884, 4.31)	5.16 (2.88, 7.55)	0.591 (0.346, 1.50)	0.334 (0.0607, 1.28)	0.369 (0.0826, 0.743)

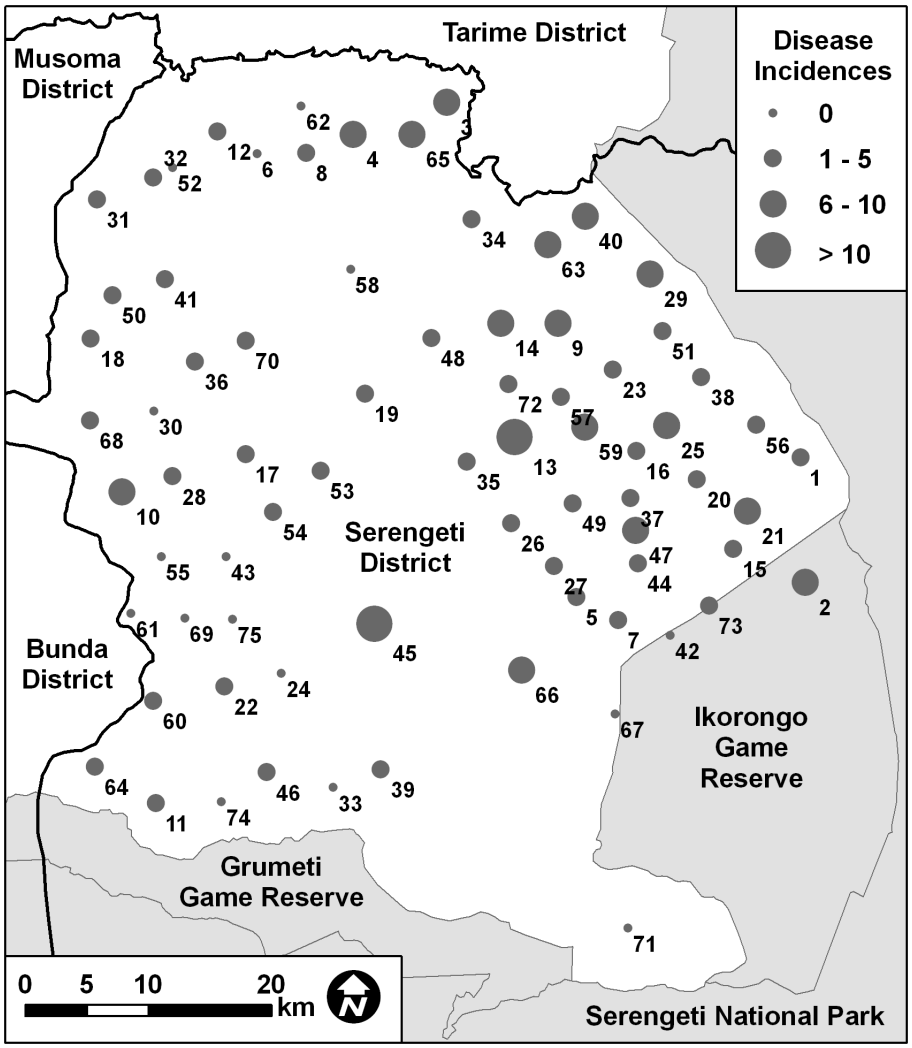

Figure 5.1: Distribution of the 75 villages in Serengeti District, Tanzania. The size of the symbol relates to the total number of occurrences of rabies observed (2002-2007), whereby an occurrence is defined as the presence of at least one suspected rabid dog in a village in a one month period. These villages are bordered by wildlife protected areas (grey) to the south and east, and other inhabited districts (white) to the north and west. Black lines depict District boundaries. Village names (indexed by the ID number beside each village in this map) are included in Appendix D.

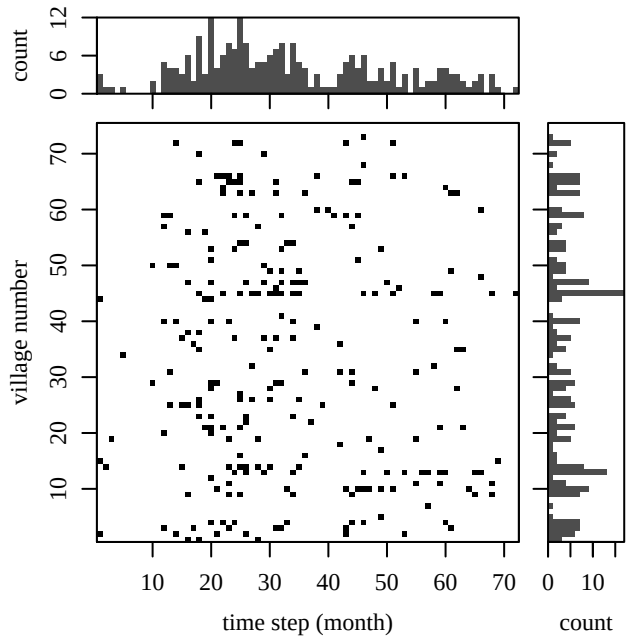

Figure 5.2: Graphical depiction of rabies occurrence (black squares) among the 75 villages in Serengeti District, Tanzania, over a 6 year period (2002-2007). Villages are ordered alphabetically (y axis), and occurrence is quantified in one month intervals (x axis). The histograms (top and right sub-plots) summarize the pattern of occurrence among villages and time periods respectively.

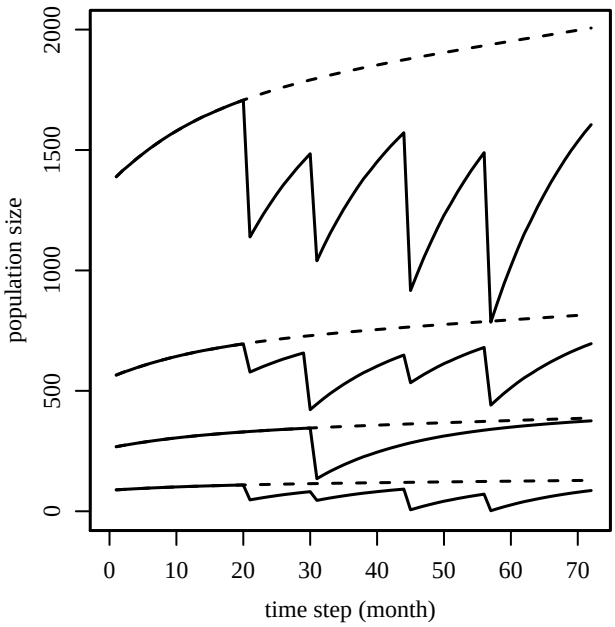

Figure 5.3: The estimated population size of susceptible dogs (solid lines) in four representative villages from 2002-2007. Declines in population numbers result from vaccination of dogs, and the increase in numbers results from both population growth and the waning of vaccination coverage. Included are villages with the smallest, largest and two intermediate dog populations. Note that not all villages were vaccinated in each vaccination campaign. The dashed line represents the estimated population of susceptible dogs in the absence of vaccination, which we use to quantify the efficacy of vaccination.

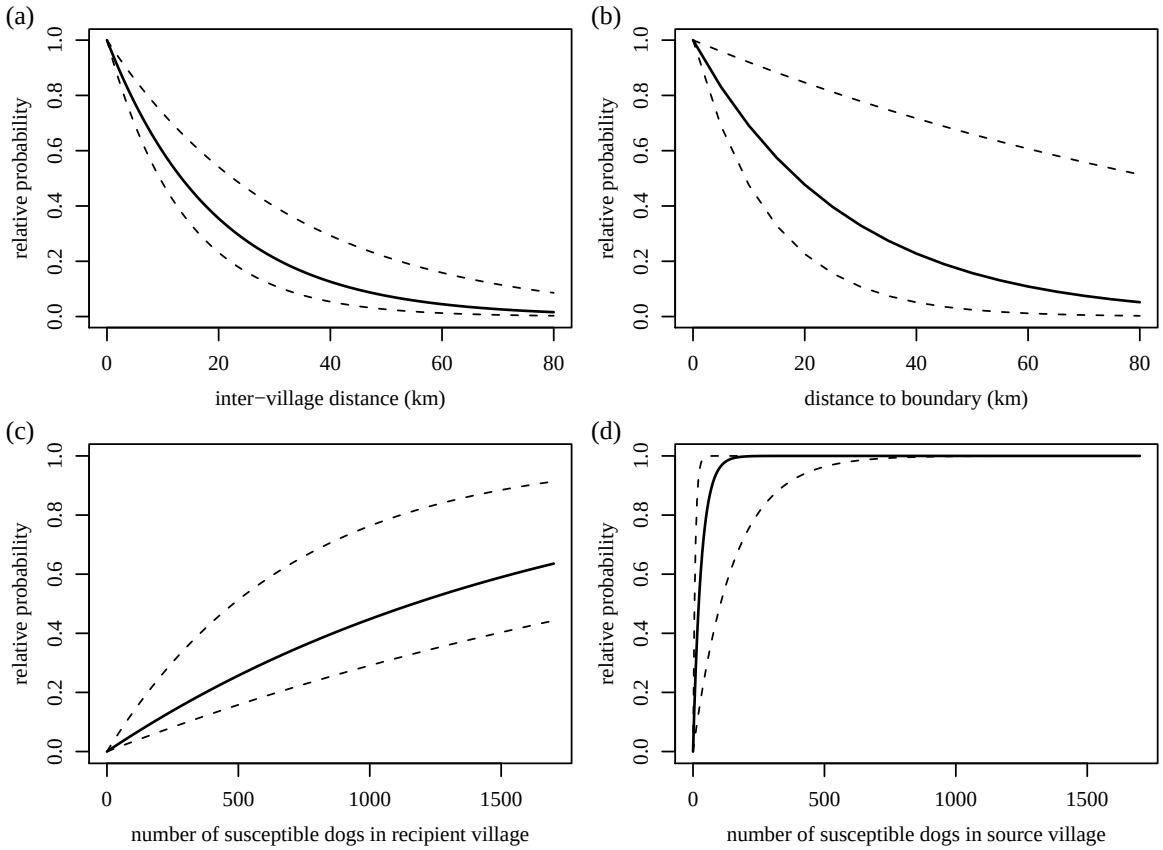

Figure 5.4: The relative probability of transmission between villages is negatively associated with the distance between villages (a) and the distance to neighbouring districts (b), and positively associated with the population size of susceptible dogs in the village receiving (c) and transmitting (d) infection. The mean (solid lines) and 95% credible intervals (dashed lines) are derived from the two highest ranked models of disease occurrence and transmission (see Methods). Graphs (a-d) are based on parameter estimates for κ, μ, δ and ψ respectively (Table 5.2), in the four highest ranking models. Note that the combination of these model components varies among these four models (Table 5.1), but parameter estimates were highly consistent among models therefore these four graphs are representative of the parameters in all four models.

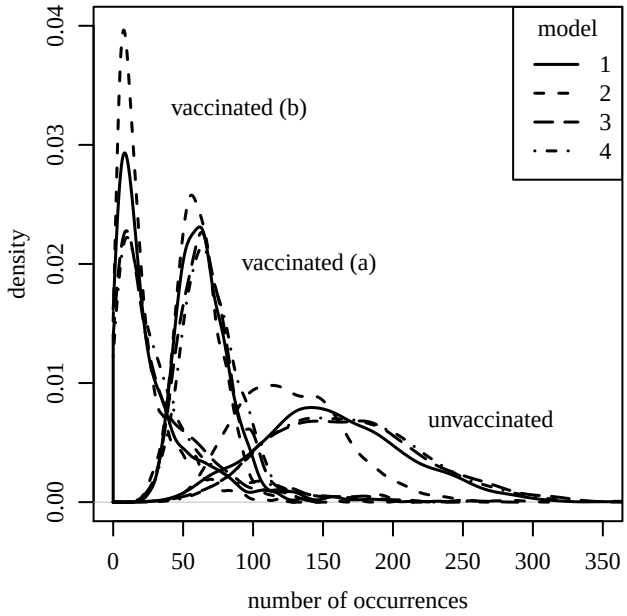

Figure 5.5: Frequency of disease occurrence based on 1000 stochastic simulations of the four highest ranking models (represented by different line styles) among the 75 villages in Serengeti District. To quantify the efficacy of vaccination occurrence of rabies was quantified in three scenarios: unvaccinated populations, vaccinated populations with a continuing external infection source (a), and vaccinated populations with an external infection source that ends 6 months after the first set of vaccinations (b). Occurrence was summed within the set of villages that were targeted for vaccination (all but 5 of the 75 villages in SD) for the 12 month period following a vaccination campaign in each village. Vaccination reduces both the mean and variance of disease occurrence.

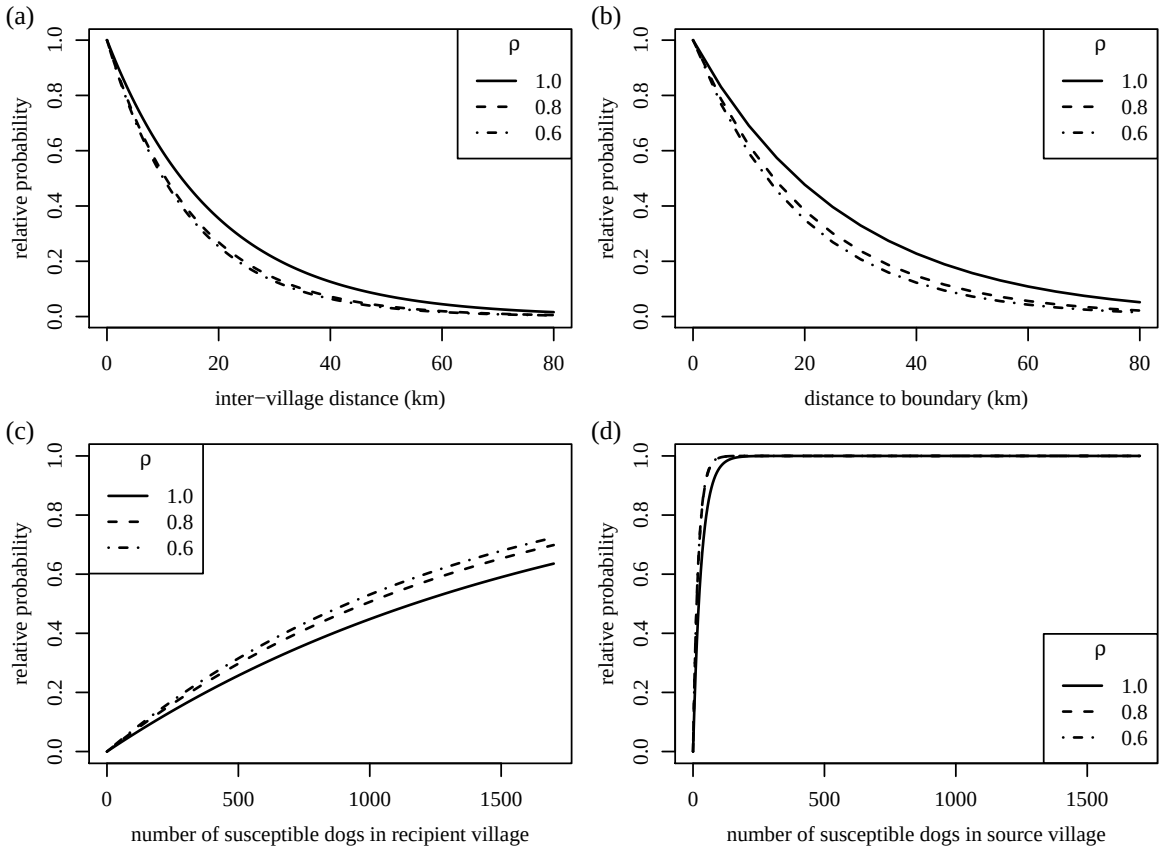

Figure 5.6: The effect of uncertainty in the proximate indicator of disease occurrence on model parameter estimates. Relative to a perfect indicator that captures 100% of occurrences (solid line), assuming the indicator captures only 80% (dashed line) or 60% (dashed dotted line) of occurrences results in a reduction to the spatial transmission kernel (a), and the external infection source kernel (b), but has little influence on the effects of the population size of susceptible dogs in the villages receiving (b) or transmitting (c) infection.

Chapter 6

The implications of metapopulation dynamics on the design of rabies vaccination campaigns

6.1 Abstract

The problem of how to most effectively deploy vaccine in metapopulations has not been resolved. We evaluate alternative strategies of pulse vaccination in order to maximize the reduction in the occurrence (presence / absence) of rabies in a metapopulation. We use metapopulation patch-occupancy models to quantify the contribution of each sub-population to disease occurrence ("risk"). The competing allocation strategies prioritise sub-populations based on population size, the risk metric, or the reduction in global (metapopulation) risk that would result from the vaccination. We also evaluate these three allocation strategies under the constraint that if villages are visited then 70% of susceptible individuals must be vaccinated. The allocation strategy that resulted in the greatest reduction in disease occurrence maximized the reduction in global risk, and was 30-50% more effective compared to all other allocation algorithms. Higher frequencies of smaller vaccination pulses were more effective at reducing occurrence than less frequent, larger pulses. Forcing a 70% vaccination level reduced the effectiveness of vaccination. This work demonstrates the strong potential for the role of metapopulation models in optimizing disease control strategies.

Key words: rabies; vaccination effectiveness; patch occupancy; allocation

6.2 Introduction

Epidemiological models are routinely applied to the design and evaluation of control measures like vaccination campaigns (e.g. Coleman & Dye, 1996; Roberts, 1996; Kao, 2002; Haydon *et al.*, 2006; Feng *et al.*, 2009). For instance, such models can be used to estimate how many vaccine doses need to be deployed to confer protection, and to understand how the frequency and intensity of vaccination pulses influence the long-term effectiveness of control programmes (Nokes & Swinton, 1997).

The simplest epidemiological models, which assume all individuals form a single, homogeneously mixed population, provide straightforward solutions to the problem of setting vaccination targets. One common estimate of the proportion of a population that must be vaccinated to reduce the basic reproductive number, R_0, below 1 is $1 - 1/R_0$ (Anderson & May, 1991). The problem with these models is that they are typically poor approximations of real-world epidemics, and it is, therefore, not clear whether inferences based on these models are valid.

The assumption of homogeneous mixing is often not valid because structuring of host populations results in heterogeneity in transmission rates that have an important influence on disease dynamics (Bolker & Grenfell, 1995; Lloyd & May, 1996; Swinton, 1998; Keeling, 2000; Fulford *et al.*, 2002). Spatially structured metapopulations can be modelled as a collection of loosely-coupled sub-populations (or patches), whereby homogeneous mixing is assumed to be a reasonable approximation within sub-populations and between-population transmission can be a function of distance and sub-population size. The dynamics of disease transmission in the metapopulation is dependent on the strength of coupling among sub-populations, the within-population value of R_0, and the size and spatial arrangement of sub-populations (Bolker & Grenfell, 1995; Ball *et al.*, 1997; Park *et al.*, 2001, 2002; Cross *et al.*, 2005). These models are rarely analytically tractable, thus it is difficult to draw general conclusions about disease dynamics and how to optimize the implementation of control measures (Watts *et al.*, 2005; Cross *et al.*, 2007).

Here, we address the problem of how to deploy rabies vaccine doses in order to reduce disease occurrence in a host metapopulation, where occurrence is defined as the presence of one or more infectious cases in a one-month period. Rabies

exerts a major public health and economic burden (Coleman *et al.*, 2004; Knobel *et al.*, 2005), and poses an important threat to animal populations of conservation concern (Woodroffe, 2001; Randall *et al.*, 2004, 2006; Haydon *et al.*, 2006; Cleaveland *et al.*, 2007). Yet rabies is a pathogen that can be effectively controlled or eliminated by vaccinating hosts (Eisinger & Thulke, 2008). We evaluate competing strategies for allocating vaccine doses among sub-populations in a metapopulation, and quantify how a trade-off between vaccination pulse frequency and pulse intensity (number of vaccine doses deployed) affects longer term effectiveness of the control programme, which we define as the reduction in occurrence of rabies resulting from vaccination relative to occurrence in a hypothetical unvaccinated population. We adopt a metapopulation patch-occupancy framework in which the occurrence of rabies is used to quantify the colonization-extinction dynamics of rabies in the metapopulation (Beyer *et al.*, in review).

6.3 Methods

This study took place in Serengeti District (SD), northern Tanzania, which borders wildlife protected areas to the south and east (Serengeti National Park and the Ikorongo and Grumeti Game Reserves), and other inhabited districts to the north and west (Bunda, Musoma and Tarime Districts). SD consists of 75 villages (Figure 6.1) and is inhabited by approximately 174,400 people (Population and Housing Census of Tanzania 2002) in primarily agro-pastoralist communities that use domestic dogs for guarding households and livestock.

Our evaluation of control strategies was based on models of rabies metapopulation dynamics presented in Beyer *et al.* (in review). In these models the presence and absence (occupancy) of rabies in a village in one month intervals is modelled as a function of four factors that influence disease transmission and persistence. First, the probability of transmission between villages is negatively associated with the distance between them. Second, the larger the sub-population of susceptible dogs in the village receiving transmission, the greater the probability of transmission. Third, for a subset of models, probability of transmission was also positively associated

with the dog population size in the sub-population transmitting infection. Finally, all models included one of three sources of infection external to the metapopulation of villages: the inhabited districts or protected areas adjacent to SD (Figure 6.1) whereby probability of infection is associated with proximity to the district or protected area boundaries, or a source that is distributed randomly throughout SD. As there was little support for the protected area source of infection models (Beyer *et al.*, in review) they are not evaluated here.

Specifically, the infectious state, **I**, of the ith village at time t is modelled as:

$$\mathbf{I}_{i,t} \sim Bern[1 - e^{-c_{i,t}}] \tag{6.1}$$

where c, the hazard rate, is proportional to the probability of transmission. The full transmission model is:

$$c_{i,t} = \beta \left(1 - e^{-\delta S_{i,t-1}}\right) \sum_{j}^{V} \left[\mathbf{I}_{j,t-1} e^{-\kappa d_{i,j}} \left(1 - e^{-\psi S_{j,t-1}}\right)\right] + \tau e^{-\mu g_i} \tag{6.2}$$

where V is the total number of villages, $d_{i,j}$ is the Euclidean distance between the centres of ith and jth villages (km), and $S_{i,t}$ is the number of susceptible dogs in the ith village at time t. The parameters κ, δ and ψ determine the relative contribution of the distance between villages and the size of the receiving and transmitting village to the probability of acquiring infection. In the models including a randomly distributed source of infection $\mu = 0$, and τ represents a constant rate of external infection among all villages. In the models where this external source is adjacent districts, g_i is the distance to the nearest inhabited district boundary, μ determines the relative probability of transmission as a function of this distance, and τ is set deterministically so that the overall rate of external infection among all SD villages is 10 infections / yr (see Beyer *et al.*, in review, for details). We use the four highest ranked models with the maximum likelihood parameter values (Table 6.2) as the basis for evaluating vaccination strategies.

Based on these models, the contribution of each sub-population to the metapopulation dynamics of disease occurrence ("risk") was quantified as the sum of two metrics that represent the risk of becoming infected, and the risk of transmitting

infection. The risk of a sub-population receiving infection is $1 - e^{-a_i}$, where:

$$a_i = \beta \left(1 - e^{-\delta S_i}\right) \frac{1}{V-1} \sum_{j,i \neq j}^{V} \left[e^{-\kappa d_{i,j}} \left(1 - e^{-\psi S_j}\right) p_j\right]. \tag{6.3}$$

The probability that a village contained infectious individuals and could thus transmit infection (p) was estimated using stochastic simulations of the meta-population model over a 72 month burn-in period, which were discarded, and then a 10000 month period that was used to calculate the probability of occurrence in each village. The proportion of simulated occurrences is an empirical estimate of p for a given population distribution in the metapopulaton, and it takes into account the influence of the external source of infection in determining patterns of occurrence.

Second, the risk of a sub-population transmitting infection (including to self) is $1 - e^{-b_i}$, where:

$$b_i = \beta p_i \left(1 - e^{-\psi S_i}\right) \frac{1}{V} \sum_{j}^{V} \left[e^{-\kappa d_{i,j}} \left(1 - e^{-\delta S_j}\right)\right]. \tag{6.4}$$

Thus, the risk (R) of sub-population i is $R_i = a_i + b_i$. We also define the global metapopulation risk as $R_G = \sum_i^V R_i$, where V is the total number of sub-populations. Formal analytical approximations for the contribution of individual sub-populations to persistence of metapopulations have been previously developed (Hanski & Ovaskainen, 2000; Ovaskainen & Hanski, 2001, 2003). However, they assume that there are no external sources of migrants (Ovaskainen & Hanski, 2001). Rabies in SD cannot be considered a closed system as an external source of infection is an essential component of the metapopulation model (Beyer *et al.*, in review). The quantities R_i and R_G are, therefore, an ad-hoc first approximation to quantifying the contribution of sub-populations to metapopulation dynamics in lieu of more formal metrics.

We evaluate six different allocation designs (A1-A6; Table 6.1). In all cases we assume a maximum of 70% of susceptible animals in each sub-population can be vaccinated, which reflects the fact that some dogs are stray, that some owners will not choose to vaccinate their dogs, and that some dogs will not be accessible for vaccination for other reasons (Lembo *et al.*, 2010). It is not possible to vaccinate

all dogs in a village, and the fraction of unvaccinated dogs could influence the effectiveness of control measures in the metapopulation. We therefore assume that a maximum of 70% will be vaccinated and explicitly account for the unvaccinated proportion of the population in our evaluation of the effectiveness of the allocation algorithms.

The first three algorithms allocate one dose at a time to the sub-population prioritized by the allocation metric. Thus, algorithms A1-A3 may not necessarily vaccinate sub-populations to the 70% level. Algorithms A4-A6 are based on the algorithms A1-A3 but with the additional constraint that if a sub-population is targeted that it must be vaccinated to the 70% level. We refer to these two strategies as "per-dose" and "per-village" allocation strategies respectively.

A1. The algorithm allocates one vaccine dose at a time to the sub-population with the most susceptible dogs, thereby reducing that sub-population size by 1. This process is repeated until all vaccine has been allocated.

A2. The algorithm allocates one vaccine dose at a time to the sub-population with the highest metapopulation risk value (R_i). After each allocation the number of susceptible dogs in that sub-population is reduced by 1 and the risk values are recalculated. This process is repeated until all vaccine has been allocated.

A3. The algorithm allocates one vaccine dose at a time to the sub-population that would result in the greatest decrease in the global risk score (R_G). After each allocation the number of susceptible dogs in that sub-population is reduced by 1 and the risk scores are recalculated. This process is repeated until all vaccine has been allocated.

A4. Vaccine doses are allocated to sub-populations in decreasing order of size, with 70% of susceptible dogs vaccinated in each sub-population in sequence until all of the doses have been allocated.

A5. Vaccine doses are allocated to sub-populations in decreasing order of risk (R_i), with 70% of susceptible dogs vaccinated in each sub-population in sequence until all of the doses have been allocated.

A6. Vaccine doses are allocated to sub-populations in decreasing order of the reduction in the global risk (R_G) that would result from vaccinating 70% of susceptible dogs in each sub-population, and repeated until all of the doses have been allocated.

The last sub-population to be vaccinated, however, receives the remainder of vaccine doses and therefore may not be vaccinated to the 70% level.

The initial number of susceptible dogs in each village in January 2002 was estimated based on the human population size and the average number of dogs per household in this region (Knobel *et al.*, 2008; Lembo *et al.*, 2008; Hampson *et al.*, 2009), resulting in 20,774 dogs distributed among the 75 villages. If the entire metapopulation was vaccinated at the 70% level this would require 14,542 doses of vaccine. Here we assume that 10,000 doses of vaccine are available. In many circumstances the availability of vaccine doses is likely to be only one of several possible limiting factors. For instance, the organisational and operational cost of vaccinating a village may be high relative to the per-dose cost of vaccine, thus could be a more important constraint on vaccination strategy. However, using vaccine doses as a constraint allows us to evaluate the relative effectiveness of the allocation algorithms and pulse designs, and to examine the importance of metapopulation dynamics on these control strategies. Furthermore, the additional constraint in algorithms A4-A6 that 70% of the village must be vaccinated if it is visited effectively limits the campaign to a small number of villages. We are therefore able to make a qualitative comparison between these different sets of constraints.

We evaluate the trade-off between vaccination intensity and frequency by contrasting the effectiveness of one-, two- and four-pulse vaccination designs over a 5 year period following the first vaccination event. The total number of vaccine doses is identical among the three scenarios: one pulse of 10,000 vaccinations in month 1, two pulses of 5,000 vaccinations in months 1 and 13, and four pulses of 2,500 vaccinations in months 1, 13, 25 and 37.

The number of susceptible and vaccinated dogs in each month of our five year study period was modelled as a function of the initial sub-population size, the

dog birth rate ($b = 0.538$ yr^{-1}) and death rates ($d = 0.450$ yr^{-1}), the number of dogs vaccinated during vaccination campaigns, and the rate at which vaccination coverage wanes ($w = 0.4$ yr^{-1}). The dynamics of the susceptible (S) and vaccinated (V) components of the sub-population were described by $dS/dt = b(S+V)+wV-dS$, and $dV/dt = -wV - dV$, which were solved using numerical simulation (the "lsoda" command; R Development Core Team, 2009).

Using the maximum likelihood parameter values for each metapopulation model (Table 6.2) we use stochastic simulations of the four highest ranked models to quantify the effectiveness of these simulated pulsed vaccination campaigns. Simulations were initialized by randomly assigning infections to three villages in the first time step, then running the simulation over a 72 month burn-in period with constant sub-population sizes (these data were discarded) before recording simulated occurrence among all sub-populations over the 60 month period following first vaccination. These simulations were performed in R (R Development Core Team, 2009). For comparison, we also evaluate the frequency of disease occurrence in an unvaccinated metapopulation. The difference in overall disease occurrence (the total number of months in which disease is observed) between the vaccinated and unvaccinated metapopulation simulations is a measure of the expected effectiveness of vaccination.

Villages are administrative boundaries that may not relate closely to heterogeneity in dog densities. Dogs typically can be found throughout this agro-pastoralist landscape so representing villages as point locations and quantifying transmission dynamics as a function of distances among villages may be an over-simplification, particularly for adjacent villages where there may be little difference in between- and within-village transmission. To test the sensitivity of our models to these assumptions village distances between adjacent villages (those in which polygon boundaries touch) were set to 0 and the analysis was repeated.

6.4 Results

The estimated contribution of sub-populations to metapopulation disease dynamics (R_i) was positively correlated with sub-population size (Figure 6.2a-d), although there was considerable variability in values resulting from the spatial proximity of sub-populations. For instance, of the villages with 400-600 dogs there was an approximately three-fold difference in values of R_i (Figure 6.2a-d). The relative importance of the roles of sub-population size and proximity are illustrated in Figure 6.3. The risk values based on the metapopulation models that included the size of the transmitting village (models 3 and 4, Figure 6.2c, d) were similar to the risk values based on the models omitting this term (models 1 and 2, Figure 6.2a, b). However, for algorithms A2, A3, A5, and A6, even small changes in the estimates of R_i can alter the way vaccine is deployed in the metapopulation. While algorithms A2 and A5 target the sub-populations with the largest values of R_i, algorithms A3 and A6 target the sub-populations where the vaccination of susceptible individuals results in the largest decrease in R_i.

The effectiveness of vaccination (the percent reduction in occurrence of rabies relative to the unvaccinated metapopulation) varied between 13.2% and 33.4% among the different allocation algorithms and pulse designs (Table 6.3). The highest effectiveness was predicted for the algorithm that allocated vaccine based on the largest per-dose reduction in R_G (A3) in the 4-pulse design (Figure 6.4).

Of the pulse designs, the effectiveness of the 4-pulse design was up to 15.9% (mean 6.5%) higher than the effectiveness of the 2-pulse design, and up to 52.7% (mean 32.5%) higher than the effectiveness of the 1-pulse design. The effectiveness of the 2-pulse design was up to 46.3% (mean 27.9%) higher than the effectiveness of the 1-pulse design. A higher frequency of vaccination pulses was therefore more effective at reducing disease occupancy over the 5 year study period, even though the total number of vaccines deployed was the same.

For the vaccine allocation algorithms based on the size of the sub-population (A1, A4), or on R_i (A2, A5), applying vaccination on a per-village basis resulted in similar vaccination effectiveness to the per-dose algorithms (comparing A1 with A4, and A2 with A5 for all three pulse designs). However, the algorithms that allocated

vaccine based on the largest reduction in R_G had greater effectiveness with the per-dose allocation strategy (A3) than the per-village strategy (A6). For instance, in the 4-pulse design the per-dose allocation strategy was 8.4% more effective than the per-village allocation strategy (Table 6.3).

There was wide variation in the number of villages visited and the proportion of susceptible dogs in each village vaccinated (Table 6.5). Allocating vaccination doses using the per-village strategy resulted in 21-59% fewer villages visited compared to the per-dose strategies. The greatest variation among allocation algorithms in the number of villages vaccinated and the number of doses allocated to villages was seen in the 1-pulse design. For instance, A1 and A2 allocated doses among numerous villages and resulted in a variable proportion of susceptible animals vaccinated in each village in contrast to A3, which allocated doses to fewer villages and consistently vaccinated the maximum of 70% of susceptible individuals, even though it was not constrained to do so (Table 6.5). These differences diminished, however, as the number of pulses increased.

The spatial and temporal pattern of vaccine allocation differed widely among the allocation algorithms. For instance, the pattern of allocation for the algorithm that allocated vaccine based on the largest reduction in metapopulation risk (A3) in the 4-pulse design was highly clumped (Figure 6.4). In contrast, the algorithm that prioritized by population size (A1) resulted in a much more dispersed pattern of allocation (Figure 6.5). Of particular note, the largest sub-population with over 1300 dogs (see arrows in Figures 6.4 and 6.5) was allocated no vaccine by algorithm A3, but was the most heavily vaccinated sub-population using algorithm A1.

Metapopulation dynamics are highly variable and there was considerable overlap in the frequency distributions of disease occurrence among all models and scenarios (Figure 6.6). Although vaccination reduces this variability, high levels of simulated occurrence were possible even in the vaccinated metapopulation.

These results were not sensitive to assumptions about how how distances between neighbouring villages were specified. Under the alternative village distance model where adjacent villages were assigned as distance of 0 and non-adjacent villages retained the original Euclidean distance, effectiveness among the pulse and

allocation strategies was qualitatively similar but approximately 3-4% higher than that of the original model (Table 6.4). This reflects the fact that changing the distance matrix effectively alters metapopulation structure.

6.5 Discussion

The algorithm (A3) that allocated vaccine doses based on metpopulation risk (R_G) was up to 52% more effective at reducing disease occurrence than the other allocation algorithms. The greater effectiveness of this algorithm (A3) compared to the algorithm that prioritized sub-populations based on size (A1) demonstrates the importance of host population structure and metapopulation dynamics on the design of, and evaluation of the effectiveness of, control measures. The implication is that, using metapopulation models, it may be possible to improve the effectiveness of control programmes by optimizing vaccination effort in spatially structured host populations. This improvement in effectiveness, however, was reduced to approximately 30% under the constraint that sub-populations were prioritized using the per-village (A6) versus the per-dose (A3) allocation strategies. Thus, the potential for improving the effectiveness of control programmes may be limited by operational constraints if the number of villages that can be visited is a strong limited factor in the design of a campaign.

In the real world, deploying a limited number of vaccines in a village, or focusing on vaccinating more, smaller villages, may not be tenable strategies if they create resentment among dog owners, competition for vaccines, or if they are cost ineffective. If the set of constraints in a specific vaccination problem can be specified then formal optimization of allocation could be achieved using techniques such as fuzzy logic (Massad *et al.*, 1999) or simulated annealing (Kirkpatrick *et al.*, 1983). For instance, Massad *et al.* (1999) characterize constraints to measles vaccination programmes in terms of compliance (the expected maximum number of vaccines that can be delivered), human resources, transportation, communication (advertising and education), and cost, and use a fuzzy logic framework to contrast competing strategies.

The single village that was considerably larger than the other villages had the largest risk value (R_i). However, for the smaller villages there was considerable variation in risk as a function of dog population size. Of particular importance are the larger villages with low risk and the smaller villages with high risk. The former represent villages that conventional approaches to vaccine allocation might suggest should be targeted, but where control efforts yield relatively little protection to the metapopulation per unit investment. The latter represent villages that might typically not be prioritized by conventional approaches, but that in fact are important from a metapopulation disease dynamics perspective.

No vaccine was allocated to the largest village (see arrows in Figure 6.4) using the most effective algorithm (A3), despite this village having the largest risk value (R_i). This village was not prioritized by algorithm A3 because a single vaccine dose resulted in a small reduction in the global risk value (R_G) relative to other villages. Thus, this work challenges the intuitive expectation that it is important for the largest population to be vaccinated. Several factors determine transmission dynamics in these models, and R_G is a function of the interaction between population size, the probability of occurrence of rabies in each village, and the spatial arrangement of villages in relation to each other and to any other possible sources of rabies infection.

Our objective was to reduce the occurrence of rabies (the presence of one or more infectious cases in a one-month period), which is different to reducing prevalence. Rabies outbreaks in SD tend to be minor (sensu Anderson & Watson, 1980), with a prevalence that is uncorrelated with the population size of susceptible dogs (data not shown), which is one reason why it is justifiable to model metapopulation dynamics of rabies as occurrence rather than prevalence (Beyer *et al.*, in review). For diseases where prevalence is strongly correlated with population size, however, population size may be a stronger driver of transmission dynamics and the largest populations may be more likely to be prioritized by the risk algorithms. A strategy that was specifically designed to reduce prevalence might, therefore, result in a different allocation of vaccine doses among villages.

The effectiveness of all vaccination allocation algorithms improved as the number of pulses increased, with the 4-pulse design being approximately twice as effective

as the 1-pulse design for the best performing algorithm (A3). The single pulse design that allocates all vaccine at once provides the greatest protection, but the susceptible population recovers from this pulse within about 2 years (Figure 6.7). In contrast, the 2- and 4-pulse designs provide less protection at any one time, but spread that protection more evenly across the 5 year study period, thereby providing better long-term protection (Figure 6.7). In this study recruitment of susceptible dogs results from both birth and from the waning of vaccination in vaccinated dogs. The best pulse strategy is, therefore, determined by a complex interaction between the number of doses administered, the number of pulses, the interval between pulses, and the rates of birth and waning of vaccination.

The metapopulation models we evaluated included an external source of infection of the dog population, representing transmission from neighbouring inhabited districts, or transmission from alternative host populations distributed throughout SD. In order to use these models to design strategies to eradicate rabies it is necessary to first better establish how vaccination of the dog population is likely to influence this source of infection. If neighbouring districts are the primary source, then vaccinating dogs on a larger regional scale would result in a rapid elimination of the source of reinfection. If, however, other host populations present throughout SD are a source of reinfection, there may be a considerable lag time between elimination of rabies in dogs, and fade-out of rabies from these other host populations. Current evidence suggests other domesticated species within SD and wildlife species in the adjacent protected areas do not constitute a separate host population (Lembo *et al.*, 2007, 2008), but they could play a role in helping to facilitate the persistence of rabies in the domestic dog population in the short-term. Understanding the likely source of external infection is key to devising eradication designs and, equally importantly, to devising responses to outbreaks arising in the disease-free population following initial eradication. This is an essential area for further work.

Metapopulation models are particularly suitable for the design of control or eradication programmes at large scales in spatially structured populations. Simple approximate rules for determining the number of individuals to vaccinate in a sub-population may be effective at preventing major outbreaks in that sub-population.

For instance, one frequently used approximation of the proportion of a population that must be vaccinated to reduce the basic reproductive number, R_0, below 1 is $1-1/R_0$ (Anderson & May, 1991). But this work suggests that, given a limited supply of vaccine, allocation strategies that take into account metapopulation structure can be substantially more effective at reducing occurrence than simple population size-based strategies.

While models can be useful in informing the design of control strategies, they are inevitably simplifications of reality and must be interpreted in the context of the broader social, political and organisational factors that influence the design and implementation of control programmes. Many of these factors are important but cannot be explicitly included in models so must be accounted for in a an ad-hoc way. Models are also sensitive to the assumptions upon which they are based. Here, we examined the sensitivity of our analysis to assumptions about how the distances among villages were quantified and found our results to be robust to those assumptions. There are many more assumptions, however, that may influence the allocation of vaccine therefore it is also important to consider the risk of model recommendations, particularly for strategies that involve high investment of resources in just a few populations. Furthermore, disease outbreaks are a highly variable stochastic process (Figure 6.6), and, under any but the most comprehensive control programmes, significant outbreaks remain possible. Yet despite these cautionary provisos, this work indicates there are significant potential benefits to using metapopulation models to identify effective control strategies.

Metapopulation models are key to understanding disease dynamics in spatially structured populations at regional scales, and to maximising the efficiency of control efforts. Although it can be difficult to obtain the data required to parameterize such models, we suggest that initial survey efforts could be used as a basis for creating simple metapopulation models, while more detailed data could be collected through time to build more detailed and informative models. Thus, we recommend that collecting metapopulation data be included as a specific objective in control programmes designs. This would include sub-population surveys, estimates of key demographic parameters as a function of sub-population size, and the relative

frequency of within- and between-population transmission rates.

6.6 Acknowledgements

Financial support was provided from the Leverhulme Trust, the National Science Foundation and National Institutes of Health (awards DEB0225453 and DEB0513994), the Wellcome Trust and Lincoln Park Zoo. Thanks are due to the Ministries of Health and Social Welfare, and of Livestock Development and Fisheries in Tanzania, TANAPA, TAWIRI, NCA Authority, the Tanzanian Commission for Science and Technology, and National Institute for Medical Research for permissions and collaboration; Intervet for providing vaccines; Frankfurt Zoological Society, Lincoln Park Zoo, Sokoine University of Agriculture, and the Mwanza Veterinary Investigation Centres for technical and logistical support; and the Serengeti Viral Transmission Dynamics team, medical officers and healthworkers, livestock field-officers, paravets, and village officers in Serengeti District.

prioritization metric:	population size (S)	metapopulation risk (R_i)	global risk (R_G)
allocation size:			
per-dose	A1	A2	A3
per-village	A4	A5	A6

Table 6.1: Vaccine was allocated to villages according to 6 algorithms (A1-A6) that differ according to whether doses are allocated on a per-vaccine or per-village basis, and according to the metric used to prioritize villages. For per-dose allocation one dose is allocated at a time, whereas for per-village allocation 70% of the population of susceptible dogs are allocated at each step. Villages were prioritized based on the population size of susceptible dogs, the magnitude of village-level metapopulation risk (R_i), or in order of the greatest reduction in global risk (R_G). See Methods for further details.

model	infection source	DIC	$\beta \times 10^{-1}$ (95% CI)	$\kappa \times 10^{-2}$ (95% CI)	$\delta \times 10^{-3}$ (95% CI)	$\psi \times 10^{-1}$ (95% CI)	$\mu \times 10^{-1}$ (95% CI)
2	R	1840	2.11 (0.923, 4.31)	5.28 (2.89, 7.75)	0.608 (0.345, 1.49)	-	-
2	D	1839	2.14 (0.936, 4.46)	5.23 (3.06, 7.76)	0.603 (0.344, 1.47)	-	0.365 (0.0858, 0.734)
3	R	1840	2.15 (0.889, 4.55)	5.25 (2.90, 7.72)	0.607 (0.343, 1.63)	0.319 (0.0604, 1.21)	-
3	D	1840	2.19 (0.884, 4.31)	5.16 (2.88, 7.55)	0.591 (0.346, 1.50)	0.334 (0.0607, 1.28)	0.369 (0.0826, 0.743)

Table 6.2: Estimated mean parameter values and 95% credible intervals for the four highest ranked metapopulation rabies disease transmission models. A dash represents a model in which that parameter was omitted.

allocation method	model	occurrence reduction (%) 1×10k	2×5k	4×2.5k
A1	1	14.8	20.6	21.5
A1	2	13.4	16.9	16.9
A1	3	14.8	19.2	20.7
A1	4	14.1	19.6	20.6
A2	1	17.2	21.5	24.2
A2	2	14.2	20.9	20.4
A2	3	16.0	22.7	24.1
A2	4	17.3	21.1	23.2
A3	1	17.3	27.7	32.1
A3	2	15.5	24.2	27.4
A3	3	17.0	28.0	33.3
A3	4	15.8	29.4	33.4
A4	1	14.9	18.4	19.6
A4	2	14.2	17.9	17.6
A4	3	14.3	19.3	20.1
A4	4	14.2	18.4	20.7
A5	1	16.6	23.0	22.9
A5	2	13.2	19.0	20.9
A5	3	16.8	23.3	25.0
A5	4	15.6	23.3	25.6
A6	1	15.6	23.0	23.7
A6	2	15.2	20.8	19.5
A6	3	15.6	21.5	25.5
A6	4	18.2	22.2	24.1

Table 6.3: The reduction in the simulated occurrence of rabies as a result of vaccination relative to simulated occurrence in an unvaccinated population. Occurrence is defined as at least one case in a sub-population in a one-month period, and was assessed over 5 years following first vaccination. Vaccine doses were allocated according to 6 algorithms (A1-A6) over three pulse schedules (one pulse of 10,000 doses, two pulses of 5,000 doses, or four pulses of 2,500 doses). Models 1-4 refer to four competing metapopulation patch-occupancy models. See Methods for details.

allocation method	model	occurrence reduction (%) 1×10k	2×5k	4×2.5k
A1	1	17.3	22.3	23.6
A1	2	14.8	18.2	19.6
A1	3	15.0	21.8	22.5
A1	4	18.1	24.0	23.5
A2	1	17.3	27.0	27.5
A2	2	16.0	21.6	23.2
A2	3	16.7	25.1	27.3
A2	4	18.8	27.7	28.1
A3	1	19.1	31.9	37.4
A3	2	17.0	27.4	31.7
A3	3	17.0	32.4	38.3
A3	4	19.0	32.4	38.7
A4	1	16.9	23.0	23.5
A4	2	14.1	18.7	20.2
A4	3	12.9	19.3	22.5
A4	4	18.5	22.5	23.1
A5	1	18.5	26.8	27.2
A5	2	14.8	22.0	22.9
A5	3	16.6	23.5	26.1
A5	4	19.4	27.2	28.6
A6	1	19.7	25.2	29.4
A6	2	14.9	21.8	23.7
A6	3	16.2	25.3	26.7
A6	4	18.2	27.4	29.6

Table 6.4: The reduction in the simulated occurrence of rabies as a result of vaccination relative to simulated occurrence in an unvaccinated population for the alternative village-distance model (see Methods). These results are qualitatively the same as those for the original distance model but approximately 3-4% higher. Occurrence is defined as at least one case in a sub-population in a one-month period, and was assessed over 5 years following first vaccination. Vaccine doses were allocated according to 6 algorithms (A1-A6) over three pulse schedules (one pulse of 10,000 doses, two pulses of 5,000 doses, or four pulses of 2,500 doses). Models 1-4 refer to four competing metapopulation patch-occupancy models. See Methods for details.

allocation		1×10k pulse			2×5k pulses			3×2.5k pulses		
method	model	N_{vill}	P_{vacc}	s.d.	N_{vill}	P_{vacc}	s.d.	N_{vill}	P_{vacc}	s.d.
A1	1-4	68	43.7	17.0	41	31.6	24.4	24	22.8	22.4
A2	1	61	48.8	32.2	33	29.8	31.3	22	17.8	21.1
A2	2	59	51.6	21.3	31	39.8	29.3	21	21.8	21.8
A2	3	60	48.4	32.3	30	29.6	31.3	22	18.2	21.7
A2	4	59	51.1	19.9	30	38.5	28.7	24	20.2	21.4
A3	1	56	68.8	9.4	31	40.4	29.9	20	25.9	24.6
A3	2	56	68.8	9.4	30	40.4	29.9	19	26.4	24.9
A3	3	57	69.4	4.9	32	40.4	29.9	21	26.6	25.0
A3	4	56	70.0	0.6	32	40.4	29.9	20	26.5	24.9
A4	1-4	40	69.0	5.9	17	38.5	30.0	10	23.3	23.3
A5	1	43	48.9	32.5	20	32.0	31.0	10	20.1	22.7
A5	2	43	66.6	14.9	20	37.1	29.9	12	21.7	22.6
A5	3	43	50.5	31.8	20	30.1	30.7	11	19.7	22.6
A5	4	43	68.4	10.7	20	37.1	29.9	10	23.2	23.6
A6	1	43	68.5	10.1	19	35.7	30.2	11	22.1	23.3
A6	2	43	69.8	1.8	19	37.4	29.7	12	21.6	23.1
A6	3	43	68.6	9.7	19	35.7	30.2	12	20.7	22.9
A6	4	44	67.6	11.9	21	36.1	29.5	12	21.8	22.5

Table 6.5: The number of villages visited during vaccination campaigns (N_{vill}), and the mean (P_{vacc}) and standard deviation (s.d.) of the percent of the susceptible population vaccinated on each visit. Each visited village was counted only once if it was visited in more than one pulse. Vaccine doses were allocated according to 6 algorithms (A1-A6) over three pulse schedules (one pulse of 10,000 doses, two pulses of 5,000 doses, or four pulses of 2,500 doses). Models 1-4 refer to four competing metapopulation patch-occupancy models. See Methods for details.

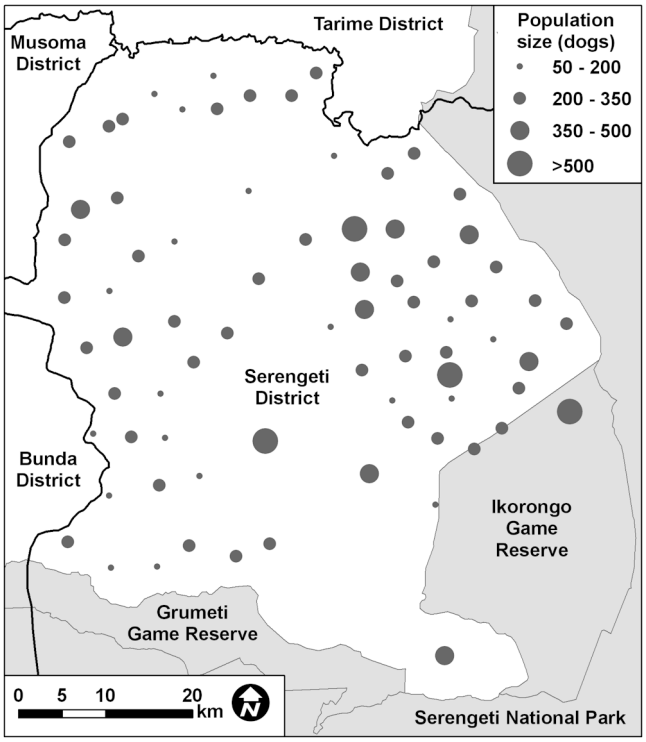

Figure 6.1: Distribution of the 75 villages in Serengeti District, Tanzania. The size of the symbol relates to the estimated domestic dog population sizes (2002). These villages are bordered by wildlife protected areas (grey) to the south and east, and other inhabited districts (white) to the north and west. Black lines depict District boundaries.

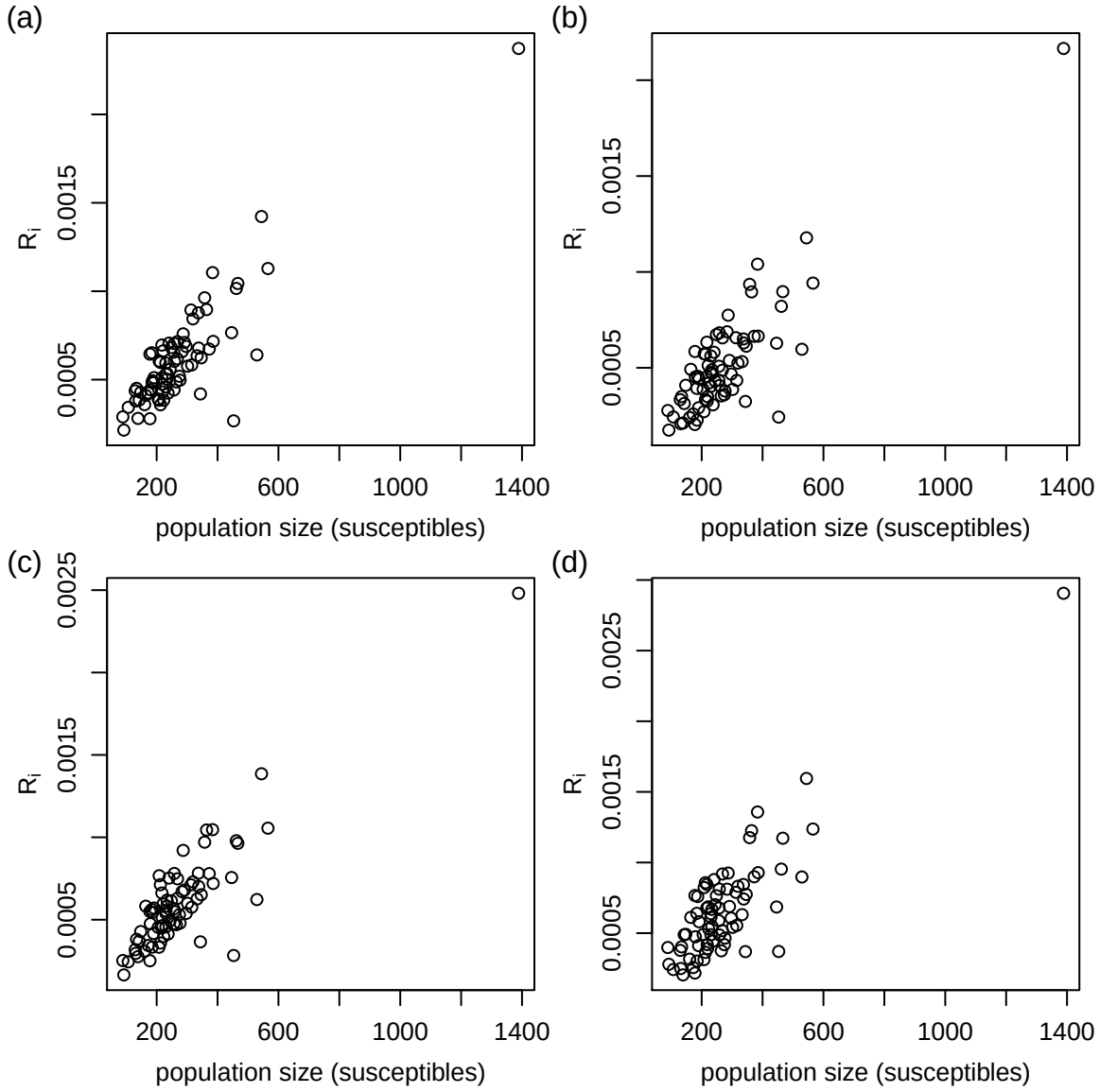

Figure 6.2: The estimated contribution of each SD village to metapopulation disease dynamics (R_i) based on four metapopulation models (a-d; see Methods) that included the number of susceptible dogs in the receiving population and the distance to neighbouring populations (a, b), and that also include the size of transmitting populations (c, d). Risk was positively correlated with population size, although there was considerable variability in values resulting from the spatial proximity of populations.

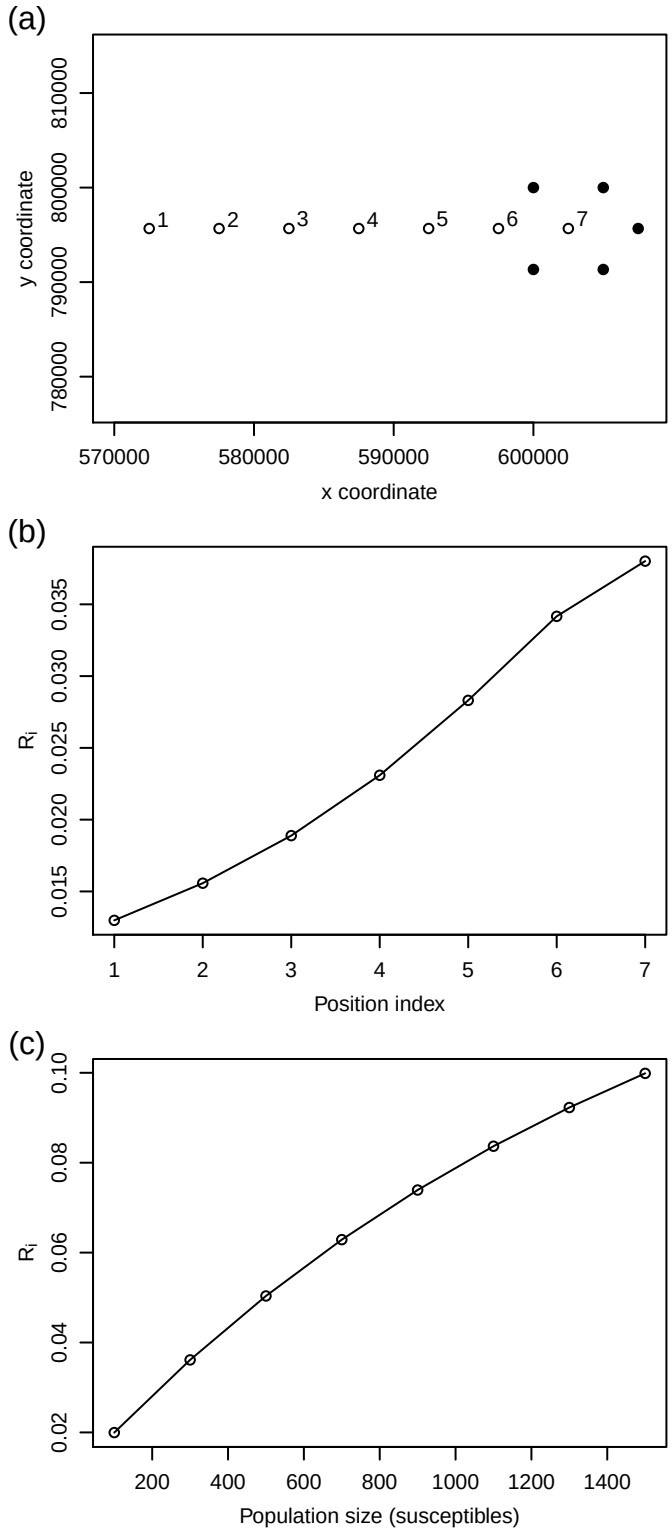

Figure 6.3: Demonstration of the relative importance of spatial proximity and population size to the contribution of a sub-population to metapopulation disease dynamics (R_i) calculated using the highest ranked model (see Methods for details). In this simple hypothetical example (a) five populations of 200 dogs are fixed (solid circles) and the location of one further village changes (open circles). R_i for this additional village increased as proximity to the other populations increased (b). The labels in (a) correspond to the x axis of (b). The influence of population size was quantified using village position 5 and calculating R_i for a range of population sizes (c). The population size has a stronger influence on the magnitude of R_i than the proximity of the populations.

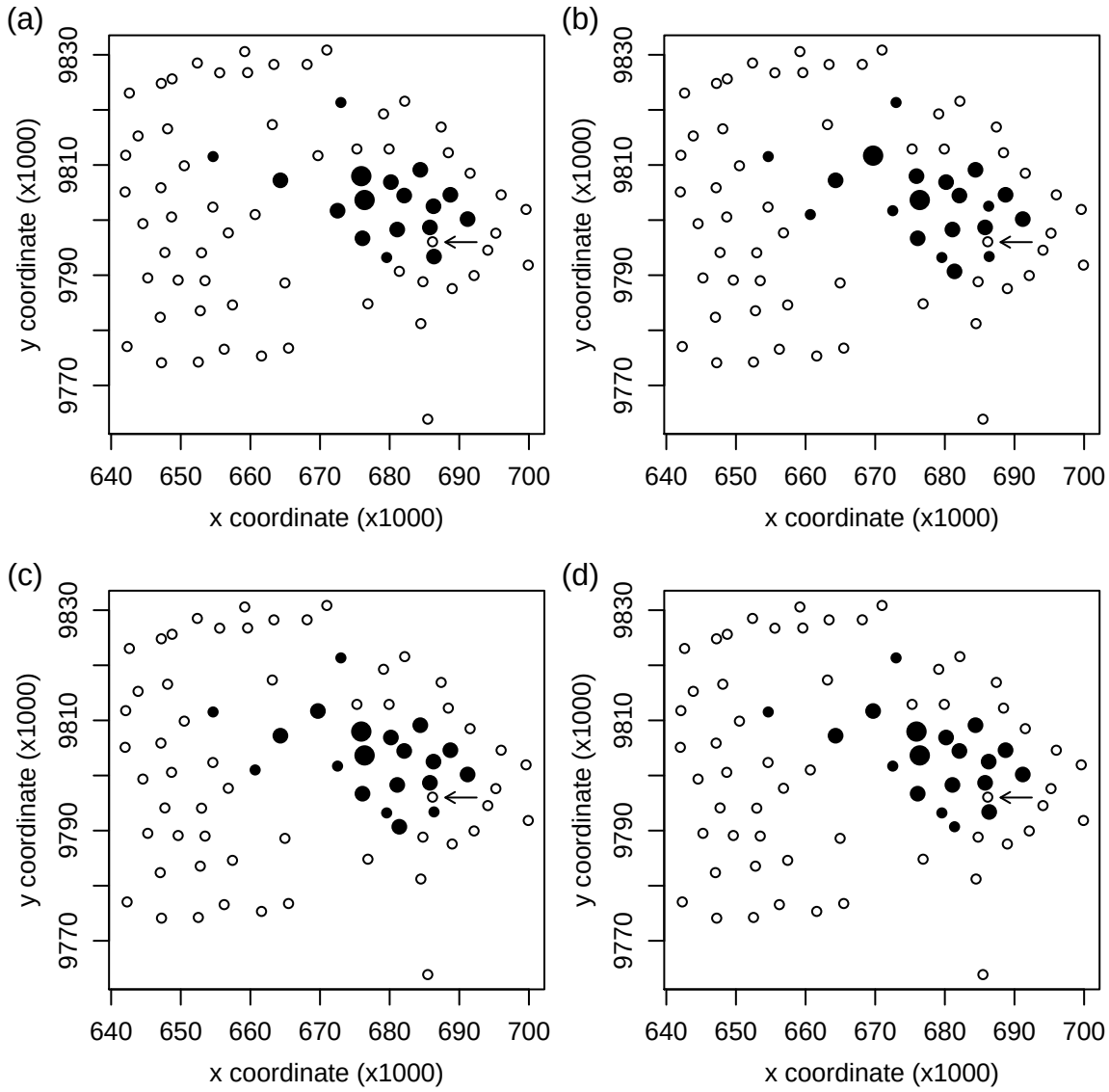

Figure 6.4: Allocation of 10,000 vaccines among villages, in four pulses of 2,500 vaccines in months 1 (a), 13 (b), 25 (c) and 37 (d), by the allocation algorithm (A3) that resulted in the greatest decrease (30.4%) in the occurrence of disease in simulations. Many populations received no vaccine doses (open circles). Small, medium and large sizes of solid circles correspond to villages that received 0-100, 100-200, or 200-300 vaccine doses respectively. The arrow indicates the largest village with an initial population of 1389 dogs.

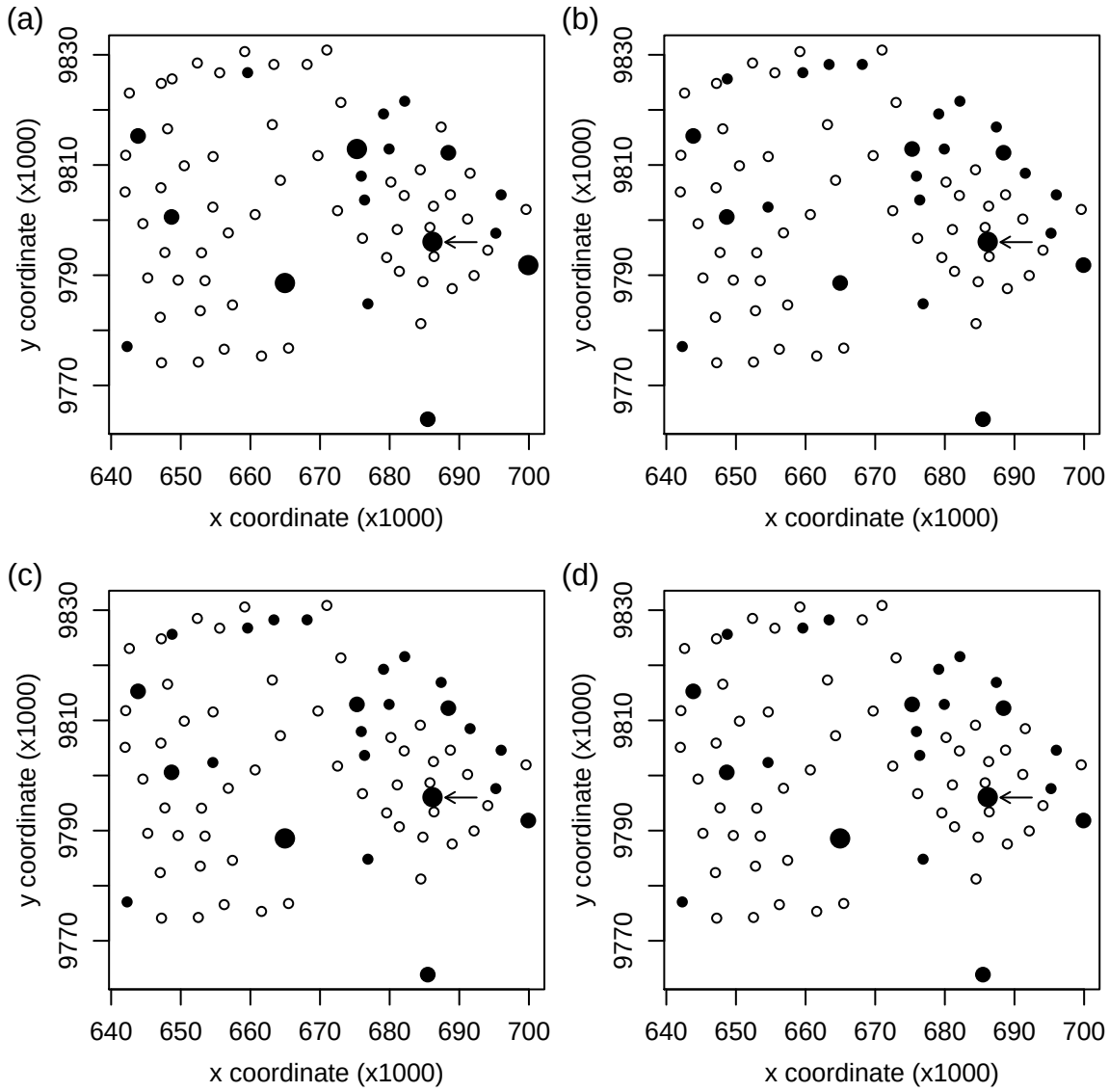

Figure 6.5: Allocation of 10,000 vaccines among villages, in four pulses of 2,500 vaccines in months 1 (a), 13 (b), 25 (c) and 37 (d), by the allocation algorithm (A1) that prioritises the largest populations, resulting in a 19.3% decrease in occurrence of disease in simulations. Many populations received no vaccine doses (open circles). Small, medium and large sizes of solid circles correspond to villages that received 0-100, 100-200, or 200-1000 vaccine doses respectively. The arrow indicates the largest village with an initial population of 1389 dogs.

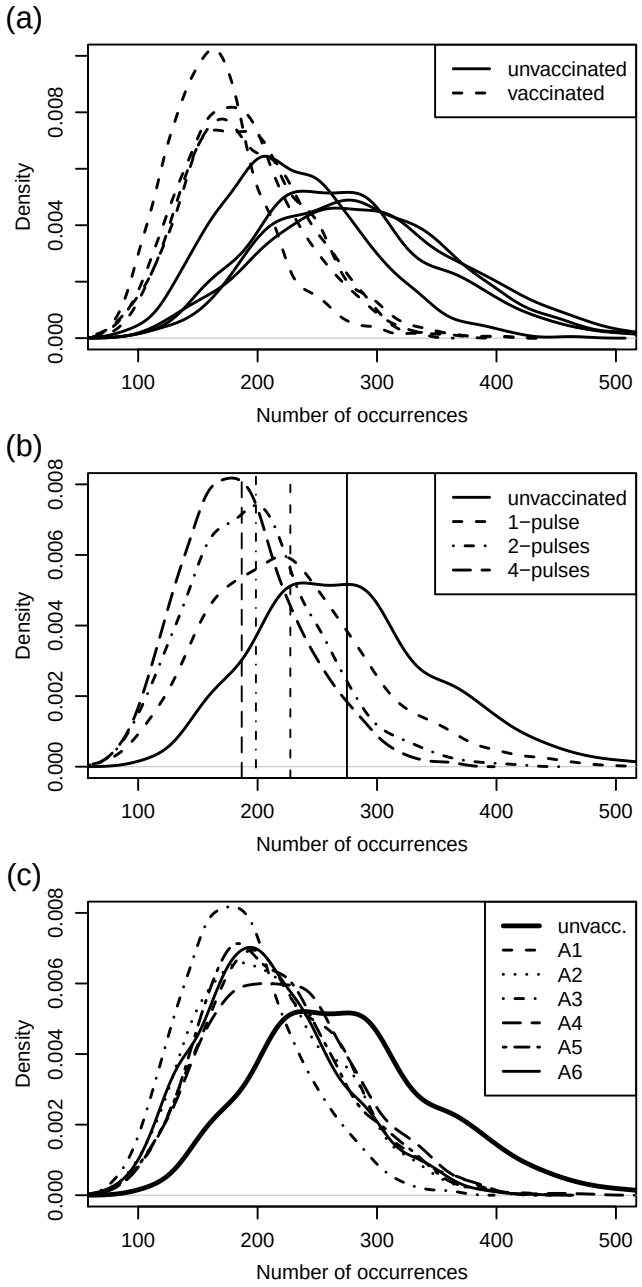

Figure 6.6: The reduction in the occurrence of rabies as a result of vaccination. Occurrence is defined as at least one case in a sub-population in a one-month period, and was assessed over 5 years following first vaccination. (a) Frequency distribution of disease occurrence based on simulations of four metapopulation models (see Methods) in unvaccinated (solid lines) and vaccinated populations (dashed lines). For simplicity, only the results of a single vaccine allocation algorithm (A3) applied using the 4-pulse design are shown. (b) Frequency distribution of disease occurrences based on simulations of the highest ranked metapopulation model (model 1) in unvaccinated populations (solid line) and populations vaccinated using four alternative vaccine pulse designs (1-, 2- and 4-pulses). For simplicity, only the results of a single vaccine allocation algorithm (A3) are shown. Vertical lines represent mean values of each distribution. (c) Frequency distribution of disease occurrences based on simulations of the highest ranked metapopulation model (model 1) in unvaccinated populations (thick line) and populations vaccinated using six alternative vaccine allocation algorithms (A1-A6). For simplicity, only the results of the 4-pulse vaccination design are shown.

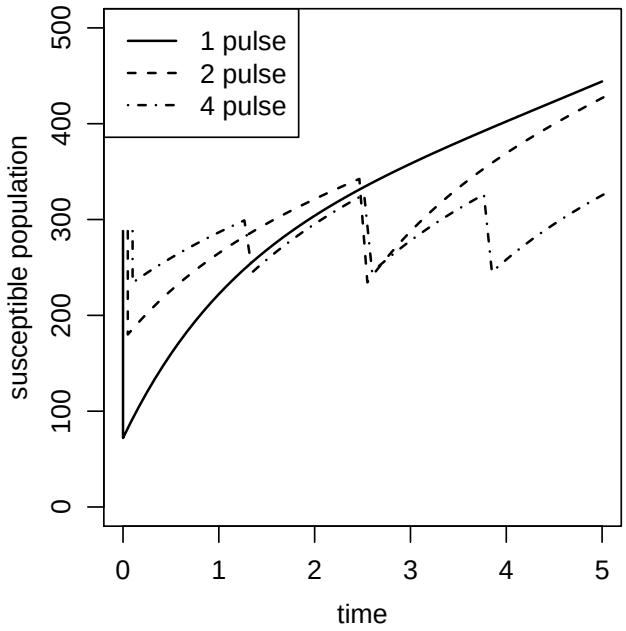

Figure 6.7: The number of susceptible individuals in a village is reduced by vaccination, but increases as new individuals are recruited to the population (birth or immigration) and due to the waning of vaccination. To illustrate the role of vaccination pulse frequency on the population of susceptible dogs, 216 vaccines were administered to a hypothetical population of average size (288 susceptible dogs) under four pulse strategies over a 5 year period. In the 1-pulse design (solid line) all 216 doses are administered at time 0, representing a 70% coverage. In the 2-pulse design (dashed line) 108 vaccines are administered at time 0 and 2.5, and in the 4-pulse design (dashed-dotted line) 54 vaccines are administered at times 0, 1.25, 2.5, and 3.75. (A small offset was added to the x-axis for the 2 and 4-pulse lines to prevent overlap.) Thus, the total number of vaccines administered is identical under all three scenarios. The mean number of susceptible dogs over the 5-year period is 311, 312 and 285 dogs respectively for the 1, 2, and 4-pulse designs. Thus, the 4-pulse design is the most effective of the three in minimizing the susceptible population.

Chapter 7

Discussion

One of the themes to emerge from this thesis is the sensitivity of SEIR models to the simplifying assumptions upon which they are based and the methods used to solve them. The assumption of exponentially distributed transition periods is one problematic assumption. I suspect that rabies is unusual in having a highly variable incubation period such that the exponential model turned out to be a better fit to the distribution than a gamma distribution (with a shape parameter greater than 1, which is the exponential distribution). It is perhaps likely that the majority of pathogens would have incubation and infectious period distributions that are considerable less-dispersed than the exponential distribution. As the method of stages (Cox & Miller, 1965; Lloyd, 2001b) provides an obvious and straightforward solution to incorporating realistic distributions into models there is little excuse for not using it. Even if data is not available for fitting a gamma distribution, the method of stages can be used to evaluate the sensitivity of the results to the assumption of exponentially distributed transition periods.

Representing a complex and continuously changing infection process as discrete states is another problematic assumption. Again, I suspect rabies is unusual in apparently having considerable variability in the infection process. There is convincing experimental evidence of possible mechanisms facilitating long incubation times and the ability of some infected animals to recover and develop immunity (at least, prior to CNS infection). Yet, there is also considerable uncertainty about the pathology of the rabies, that is fuelled in part by the complex interaction of different strains in different hosts. In particular, lab strains of RV appear to have much lower

pathogenicity than some wild ("street-virus") strains, and it is questionable how directly applicable the lab strain work is to disease in the wild.

One issue that has not been explored in this thesis is the effect that variable infectivity might have on outbreak dynamics. Transmission of rabies among dogs appears to be assisted by the neurological and therefore behavioural changes resulting from infection such as increased aggression, and tolerance to pain. As these changes appear to develop gradually during the infectious period, the probability of transmission may be relatively low in the early stages of the infectious period, and considerably higher up until the time the animal dies. This could be important in SD because the human population is quite effective at killing infectious dogs, thereby truncating the infectious period. Although we took this reduction in the duration of the infectious period into account in our models, we may have underestimated the importance of this effect if the days truncated from the end of the infectious period represent the majority of transmission events. Based on contact-tracing data (Hampson *et al.*, 2009) I suspect that is not the case, but it remains an issue that would be interesting to explore.

Another theme to emerge from this thesis is the importance of transmission heterogeneity at multiple scales. It is often assumed that the assumption of homogeneous mixing in small populations is reasonable. While this is obviously not true in some circumstances, particularly in the context of agricultural systems (e.g. cows on farms), it is very difficult to gauge the variability of transmission rates in SD dog populations. Dogs are often free to roam and mix in rural villages, although this may not be true in the more urban settings. Yet we found support for the hypothesis of population structuring in small populations (< 300 dogs). Even though this did not change the estimate of R_0 compared to the simple (unstructured population) SEIR models, it reduced outbreak severity substantially. Modelling the dog population as groups of dogs where within-group transmission was approximately 15 times higher than between-group transmission resulted in much more frequent stochastic fade-out of disease.

There is little doubt about the importance of population structure at the district level. The patch-occupancy approach was highly effective in modelling meta-

population dynamics. Although many metapopulation disease models are based on compartment models (Park *et al.*, 2001), the inability of these models to generate realistic outbreak size distributions (Chapter 4) precluded this approach here. Furthermore, there is little relationship between populations size and outbreak size (data not shown). If an outbreak occurs, it tends to result in a small number of cases, regardless of population size. Thus, modelling occurrence of rabies rather than incidence of rabies is a valid approach, and does not suffer from subjectivity in defining a suitable SEIR model.

Although epidemiological data can be difficult to collect, especially in developing countries, we have shown that metapopulation models can be fit based on proximate indicators of disease occurrence (medical records of human bite injuries). This data is much more readily available than direct information about rabies incidence in dogs, and has the added advantage of being a record of incidence over many years. Bayesian state-space models are a rigorous framework for relating this noisy indicator data to the true variable of interest (occurrence or incidence). The data model component of our state-space formulation is very simple, including only a single parameter representing the probability of detection. While this is a reasonable starting point, long-term projects could collect supplementary data that could be used to develop a better models of the link between unobserved, true incidence, and the proxy variable we observe.

Thus, the absence of direct epidemiological data does not always prevent us from developing more realistic models that might offer greater insight into disease dynamics and control, and there is clearly potential to apply these methods to many disease problems in developing countries. Given the potential trade-off between obtaining a detailed, one-time "snapshot" of the infectious status of a population (e.g. through field surveys), versus obtaining a much coarser, but long-term estimate of disease incidence, it is not obvious to me that the detailed data necessarily offers greater insight. Detailed but short-term incidence data offers limited scope for developing an understanding of transmission dynamics, which is so important from an intervention perspective. The possible exception to this is the collection of sequencing data over wide regions, which can address fundamentally important

questions of the origin and spread of pathogens over large scales (Biek *et al.*, 2007).

From the perspective of using the patch-occupancy metapopulation models for designing disease eradication programmes at a larger regional scale there are three issues that it would be useful to investigate further. Two of these issues relate to the problem of applying the model to population sizes that fall outside of the range of population sizes that were used to fit the model. First, it is not clear how the probability of transmission declines at very small population sizes (< 50 dogs). There is some evidence of quite strong non-linearity in transmission dynamics at small population sizes (Chapter 5). A better definition of the changes in transmission rates in these very small populations would be useful. At the other end of the spectrum, there is also uncertainty about how transmission dynamics change in larger populations. It would be interesting to fit these models to data from adjacent, unvaccinated villages, which have larger populations of susceptible dogs. There is also the question of whether population mixing is fundamentally different in the large populations, and in particular whether it would be more appropriate to develop different models for rural and urban populations. Dogs may be much more restricted in the latter, which would have a profound influence on transmission dynamics. Thus, very fine-scale data on within-village social and spatial organisation of the dog population would also be very useful for improving our understanding of this system.

Appendix A

Deterministic solutions of the method of stages

Systems represented by a series of ordinary differential equations, such as the SEIR model presented here with gamma distributed incubation and infectious periods implemented using the method of stages, can be solved deterministically using numerical simulation algorithms. Here we present an example of how to generate the deterministic solution to an SEIR model with 5 incubation and infectious stages using R (R Development Core Team, 2009).

The equations are built into a function that can be referenced using the "odesolve" library:

```
modelSEIR_5_5 <- function(t, y, p){
dS <- p[1]*(y[1] + y[2] + y[3] + y[4] + y[5] + y[6] + y[7] + y[8]
+ y[9] + y[10] + y[11] + y[12]) - p[3]*y[1]*(y[7] + y[8] + y[9]
+ y[10] + y[11]) - p[2]*y[1]
dE1 <- p[3]*y[1]*(y[7] + y[8] + y[9] + y[10] + y[11])
- (p[6]*p[4] + p[2])*y[2]
dE2 <- p[6]*p[4]*y[2] - (p[6]*p[4] + p[2])*y[3]
dE3 <- p[6]*p[4]*y[3] - (p[6]*p[4] + p[2])*y[4]
dE4 <- p[6]*p[4]*y[4] - (p[6]*p[4] + p[2])*y[5]
dE5 <- p[6]*p[4]*y[5] - (p[6]*p[4] + p[2])*y[6]
dI1 <- p[6]*p[4]*y[6] - (p[7]*p[5] + p[2])*y[7]
dI2 <- p[7]*p[5]*y[7] - (p[7]*p[5] + p[2])*y[8]
dI3 <- p[7]*p[5]*y[8] - (p[7]*p[5] + p[2])*y[9]
dI4 <- p[7]*p[5]*y[9] - (p[7]*p[5] + p[2])*y[10]
dI5 <- p[7]*p[5]*y[10] - (p[7]*p[5] + p[2])*y[11]
dR <- p[7]*p[5]*y[11] - p[2]*y[12]
list(c(dS, dE1, dE2, dE3, dE4, dE5, dI1, dI2, dI3, dI4, dI5, dR))
}
```

The vector t represents the times at which the values of each of the components of the model are reported, y is a vector that contains the initial conditions of each component in the system, and p is a vector of parameters which in this case corres-

ponds to: the birth rate (b), death rate (d), the transmission parameter (β), the rate parameter for the incubation period (σ) and infectious period (α), and the number of incubation stages (m) and infectious stages (n). The solution to the model is calculated as follows:

```
library(odesolve)
times <- seq(0,600,0.1)
parms <- c(0,0,0.00015,0.1,0.1,5,5)
init <- c(1000, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0)
sol <- lsoda(init,times,modelSEIR_5_5, parms, rtol=1e-4, atol=1e-6)
```

The resulting object "sol" is a dataset where each row represents the state of the system at a time specified in the vector "times", which are recorded in the first column of this dataset, and the subsequent columns contain the value of each of the stages at that time.

Appendix B

Sensitivity analysis (Chapter 4)

The sensitivity of our analysis to the assumption of a 50% detection rate was evaluated by repeating the analysis with detection rates of 40 and 60%. The following two figures describe the model selection and parameter estimation results for the human intervention model (first figure) and the structured population model (second figure) at these two detection probability levels. Differences in detection probabilities resulted in small quantitative, not qualitative, differences to the analysis.

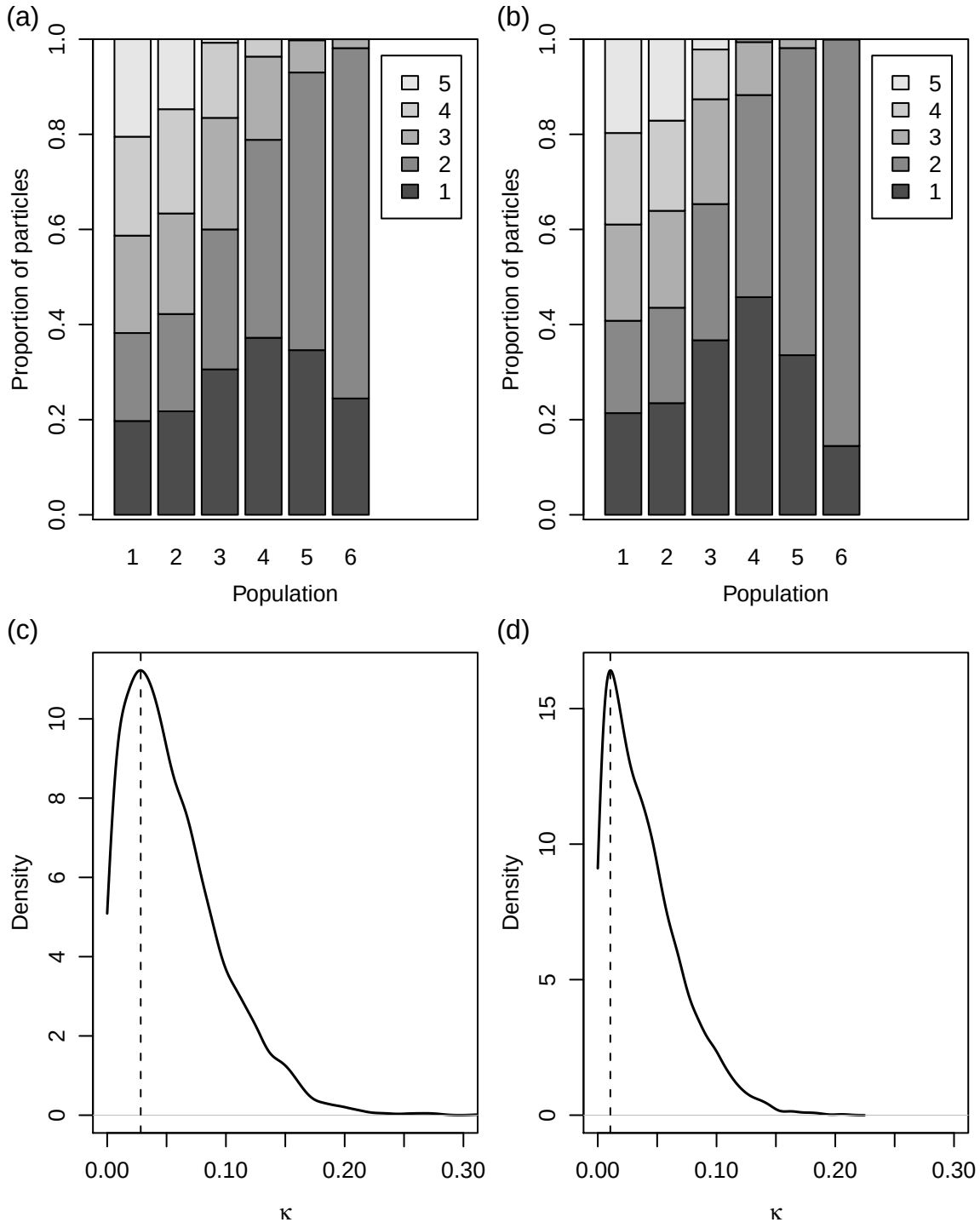

Figure B.1: The output from the ABC-SMC algorithm for the human intervention models assuming detection rates of 40% (a, c) or 60% (b, d). (a, b) Proportion of particles associated with each model (1-5) in each of the six populations of particles (x axis). The distance measures between observed and simulated datasets become increasingly stringent in this progression of populations. The relative frequency of particles for each model in the final population is used as an indication of the relative likelihood of the models. The strongest support was found for Model 2 at both detection probability levels. (c, d) The estimated posterior distribution of κ based on the density of particles in the final population of particles. The dashed line represents the estimated maximum likelihood estimate of κ.

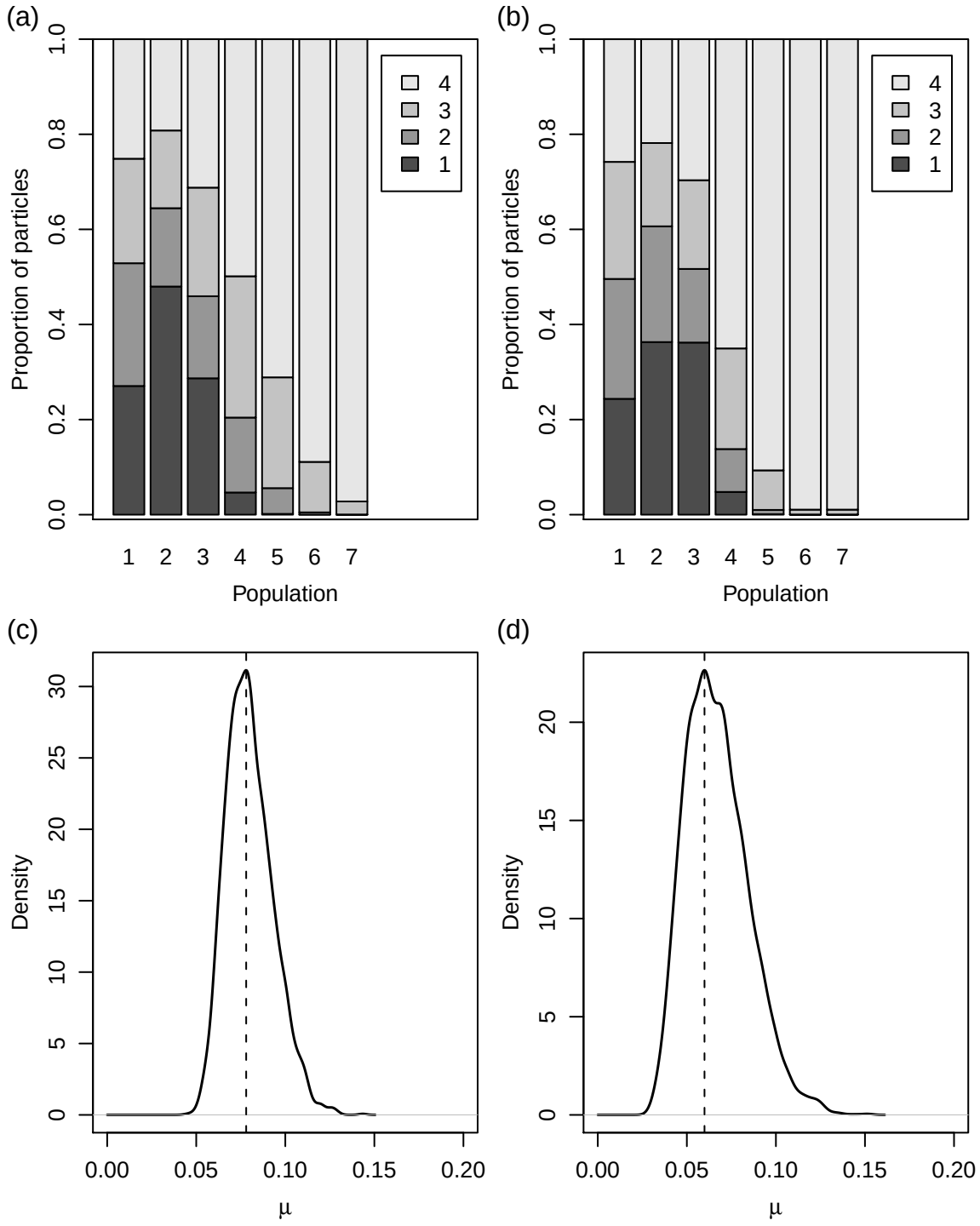

Figure B.2: The output from the ABC-SMC algorithm for the structured population models assuming detection rates of 40% (a, c) or 60% (b, d). (a, b) Proportion of particles associated with each model (1-4) in each of the six populations of particles (x axis). The distance measures between observed and simulated datasets become increasingly stringent in this progression of populations. The relative frequency of particles for each model in the final population is used as an indication of the relative likelihood of the models. The strongest support was found for Model 4 at both detection probability levels. (c, d) The estimated posterior distribution of μ based on the density of particles in the final population of particles. The dashed line represents the estimated maximum likelihood estimate of μ.

Appendix C

Supplementary material (Chapter 5)

Table C.1: Names of the 75 villages in Serengeti District, Tanzania. Numbers correspond to the numbers on the map in Figure 1. Alternative spellings are shown in parentheses.

1	Bisarara	39	Mbiso (Mbisso)
2	Bonchugu	40	Merenga
3	Borenga	41	Mesaga
4	Buchanchari	42	Miseke (Misseke)
5	Burunga	43	Monuna
6	Busawe	44	Morotonga
7	Bwitengi	45	Mosongo
8	Gentamome (Gantamome)	46	Motukeri
9	Gesarya	47	Mugumu
10	Gusuhi	48	Musati
11	Iharara	49	Ngarawani
12	Iseresere	50	Nyagasense
13	Itununu	51	Nyamakendo
14	Kebanchabache (Kebanchebanche)	52	Nyamakobiti
15	Kebosongo	53	Nyamatare
16	Kegonga	54	Nyamatoke
17	Kemgesi	55	Nyambureti
18	Kenyamonta	56	Nyamburi
19	Kenyana	57	Nyamerama
20	Kibeyo	58	Nyamitita
21	Kisangura	59	Nyamoko
22	Kitembere	60	Nyamsingisi (Nyamasingisi)
23	Kitunguruma	61	Nyankomogo
24	Kono	62	Nyansurumuti (Nyansurumunti)
25	Koreri	63	Nyansurura
26	Kwitete	64	Nyiberekera
27	Kyambahi	65	Nyiboko
28	Maburi	66	Nyichoka
29	Machochwe	67	Park Nyigoti
30	Magange	68	Remung'orori
31	Magatini	69	Rigicha
32	Majimoto	70	Ring'wani
33	Makundusi	71	Robanda
34	Marasomoche	72	Rung'abure
35	Masangura	73	Rwamchanga
36	Masinki	74	Singisi
37	Matare	75	Wegete (Wagete)
38	Mbalibali		

Table C.2: Estimated mean parameter values and 95% credible intervals for rabies disease transmission models. A dash represents a model in which that parameter was omitted. The four highest ranked models are displayed in bold.

model	infection source	rate (yr^{-1})	deviance	DIC	β (95% CI) $\times 10^{-1}$	κ (95% CI) $\times 10^{-2}$	δ (95% CI) $\times 10^{-3}$	ψ (95% CI) $\times 10^{-1}$	μ (95% CI) $\times 10^{-1}$
0	R	0	1958	1959	0.461 (0.405, 0.525)	-	-	-	-
1	R	2	1950	1952	0.321 (0.224, 0.449)	4.04 (2.49, 5.52)	-	-	-
1	R	6	1912	1914	0.357 (0.234, 0.525)	5.17 (3.19, 7.28)	-	-	-
1	R	10	1896	1897	0.41 (0.255, 0.649)	6.64 (4.19, 9.50)	-	-	-
1	P	2	1951	1954	0.323 (0.222, 0.457)	4.04 (2.45, 5.73)	-	-	0.311 (0.0424, 0.714)
1	P	6	1915	1917	0.348 (0.234, 0.52)	4.96 (3.10, 7.05)	-	-	0.251 (0.0355, 0.569)
1	P	10	1900	1902	0.38 (0.240, 0.594)	6.05 (3.68, 8.84)	-	-	0.201 (0.0193, 0.446)
1	D	2	1947	1950	0.329 (0.231, 0.467)	4.14 (2.63, 5.83)	-	-	0.627 (0.177, 1.170)
1	D	6	1911	1914	0.366 (0.244, 0.552)	5.24 (3.41, 7.40)	-	-	0.445 (0.123, 0.905)
1	D	10	1895	1898	0.419 (0.268, 0.653)	6.69 (4.27, 9.60)	-	-	0.322 (0.065, 0.638)
2	R	2	1885	1887	1.72 (0.747, 3.38)	3.40 (1.88, 4.92)	0.642 (0.346, 1.60)	-	-
2	R	6	1850	1852	1.85 (0.779, 3.67)	4.25 (2.37, 6.24)	0.640 (0.345, 1.61)	-	-
2	**R**	**10**	**1837**	**1840**	**2.11 (0.923, 4.31)**	**5.28 (2.89, 7.75)**	**0.608 (0.345, 1.49)**	**-**	**-**
2	P	2	1886	1889	1.70 (0.737, 3.31)	3.40 (1.84, 4.98)	0.654 (0.344, 1.71)	-	0.316 (0.0345, 0.714)
2	P	6	1852	1855	1.86 (0.810, 3.80)	4.13 (2.27, 6.11)	0.624 (0.345, 1.52)	-	0.271 (0.0239, 0.592)
2	P	10	1840	1843	2.01 (0.863, 4.13)	4.87 (2.83, 7.06)	0.606 (0.342, 1.50)	-	0.205 (0.0201, 0.470)
2	D	2	1881	1884	1.72 (0.741, 3.47)	3.49 (2.04, 5.13)	0.658 (0.346, 1.66)	-	0.678 (0.230, 1.22)
2	D	6	1847	1850	1.94 (0.856, 3.92)	4.31 (2.55, 6.19)	0.619 (0.345, 1.58)	-	0.535 (0.142, 1.04)
2	**D**	**10**	**1836**	**1839**	**2.14 (0.936, 4.46)**	**5.23 (3.06, 7.76)**	**0.603 (0.344, 1.47)**	**-**	**0.365 (0.0858, 0.734)**
3	R	2	1885	1888	1.72 (0.706, 3.51)	3.40 (1.88, 5.06)	0.656 (0.344, 1.71)	0.378 (0.0780, 1.26)	-
3	R	6	1850	1853	1.93 (0.829, 3.94)	4.26 (2.34, 6.46)	0.623 (0.347, 1.54)	0.341 (0.0644, 1.24)	-
3	**R**	**10**	**1838**	**1840**	**2.15 (0.889, 4.55)**	**5.25 (2.90, 7.72)**	**0.607 (0.343, 1.63)**	**0.319 (0.0604, 1.21)**	**-**
3	P	2	1886	1890	1.66 (0.705, 3.42)	3.35 (1.82, 4.97)	0.674 (0.350, 1.80)	0.356 (0.0784, 1.28)	0.329 (0.0406, 0.739)
3	P	6	1852	1856	1.84 (0.776, 3.89)	4.09 (2.32, 6.14)	0.641 (0.346, 1.75)	0.342 (0.0682, 1.25)	0.267 (0.0328, 0.589)
3	P	10	1840	1843	2.03 (0.845, 4.22)	4.85 (2.86, 7.17)	0.614 (0.345, 1.53)	0.333 (0.0668, 1.28)	0.203 (0.0216, 0.451)
3	D	2	1881	1885	1.76 (0.775, 3.57)	3.53 (2.04, 5.06)	0.655 (0.343, 1.74)	0.373 (0.0761, 1.28)	0.675 (0.226, 1.24)
3	D	6	1847	1851	2.02 (0.856, 3.90)	4.33 (2.58, 6.19)	0.605 (0.343, 1.54)	0.365 (0.0725, 1.26)	0.521 (0.144, 1.01)
3	**D**	**10**	**1836**	**1840**	**2.19 (0.884, 4.31)**	**5.16 (2.88, 7.55)**	**0.591 (0.346, 1.50)**	**0.334 (0.0607, 1.28)**	**0.369 (0.0826, 0.743)**

State-space models

Here we provide further information regarding the state-space model output. Considerable processing time is required to generate samples from the posterior distributions of these models. Furthermore, we adopted a conservative thinning rate of 1 in 50 to ensure there was no autocorrelation between samples. It was therefore difficult to generate a large number of samples from the posterior distributions. However, because chain convergence was good, generating additional samples from the posterior distribution is unlikely to alter the parameter estimates.

We present the thinned MCMC chain histories for the most complex model, which is representative of the behaviour of the other models. We also present the posterior density plots, and means and 95% credible intervals, for each of the estimated parameters in the models.

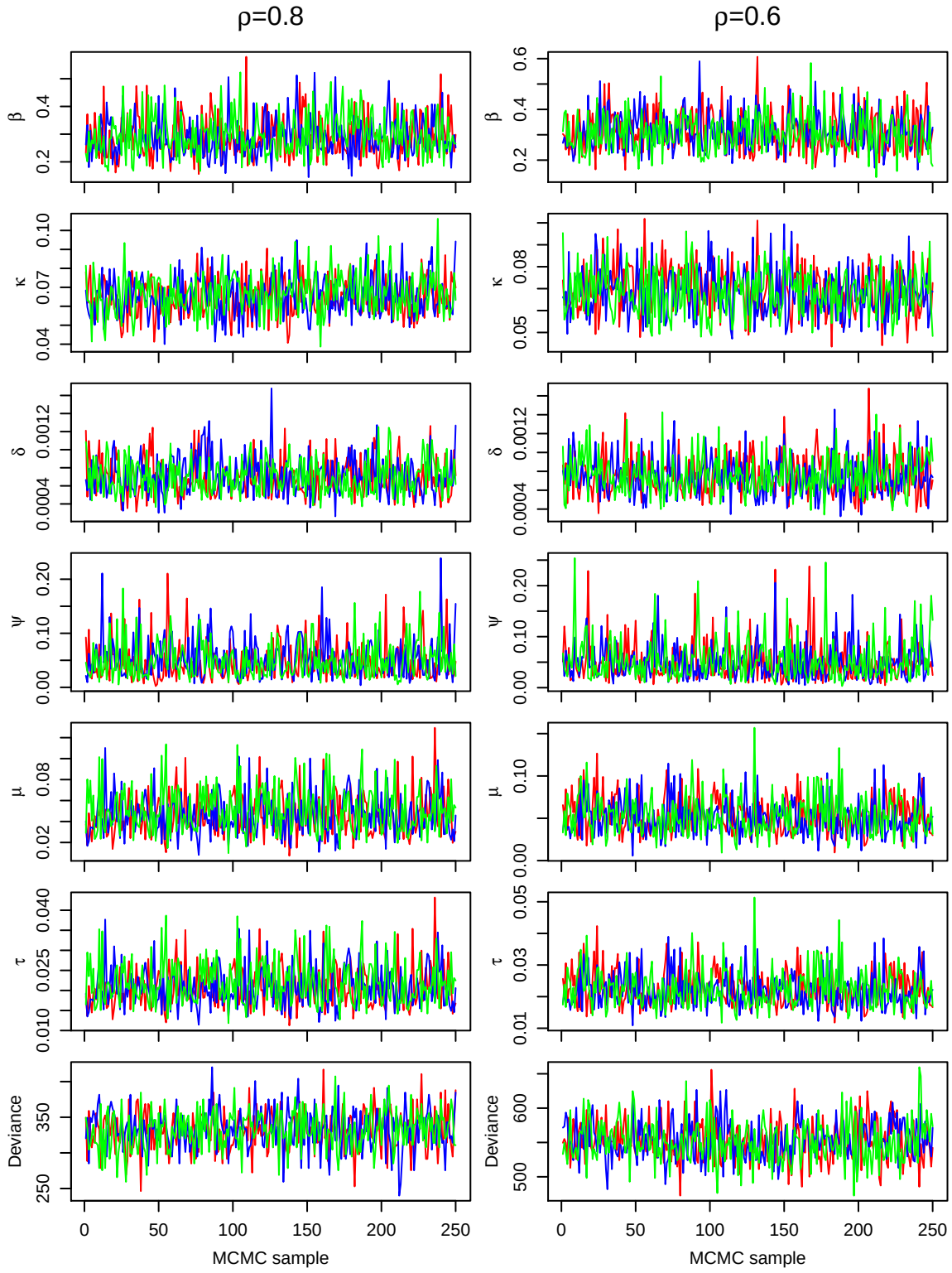

Figure C.1: Sample of MCMC chains for the state-space metapopulation models demonstrating chain convergence. Only the chains for the third model (in which the source of external infections are the adjacent districts and the rate of infection is 10 infections per year) are shown as it is the most complex model and is representative of the behaviour of the other models. The left and right columns of plots correspond to the 80% ($\rho = 0.8$) and 60% ($\rho = 0.6$) reporting probabilities.

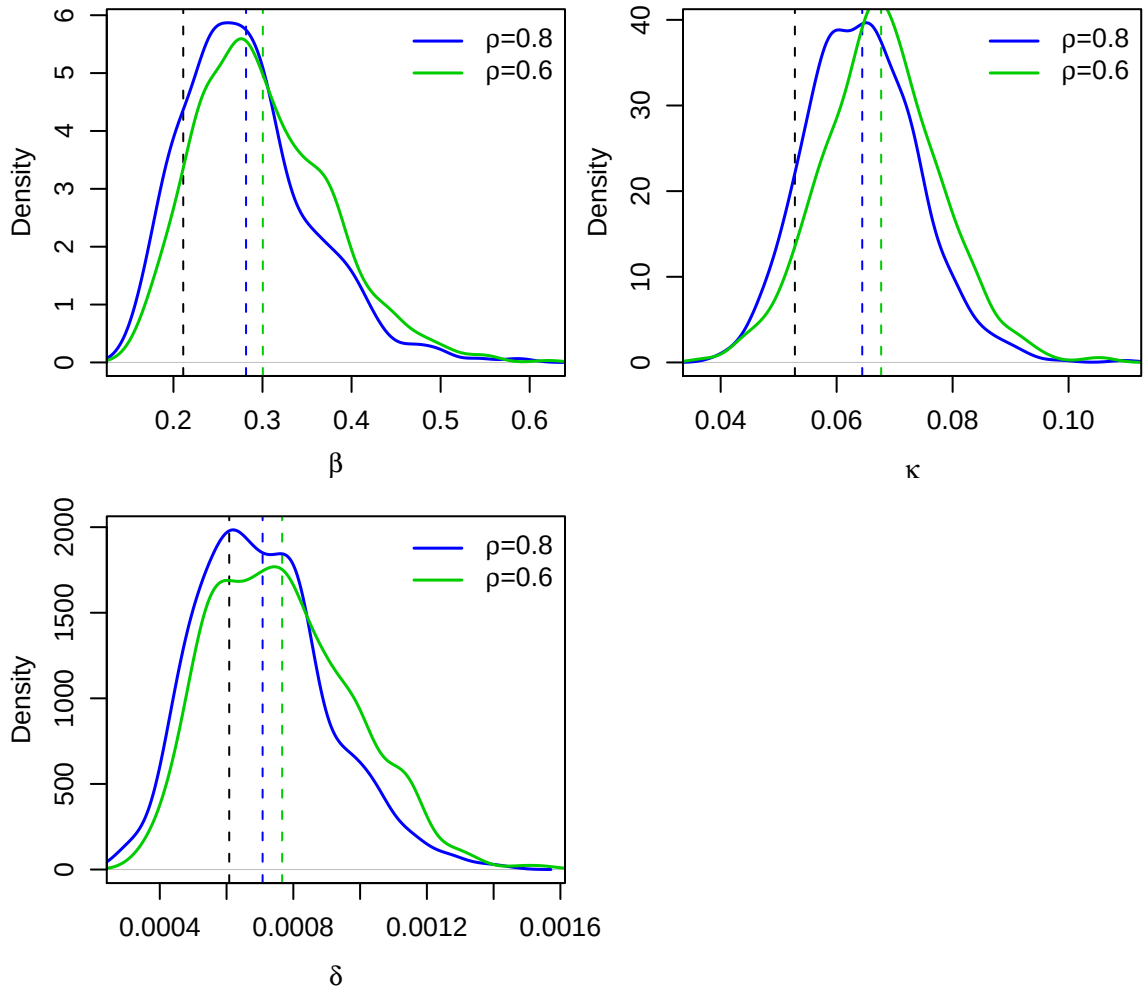

Figure C.2: Posterior densities for the estimated parameters (β, κ, δ) for model 2 in which the source of the external infection is randomly distributed at a rate of 10 infections yr^{-1}. The blue and green lines represent the models with 80% ($\rho = 0.8$) and 60% ($\rho = 0.6$) reporting probabilities. The vertical dashed lines are the mean of the distributions, with the black line representing the mean based on the model that assumes perfect reporting.

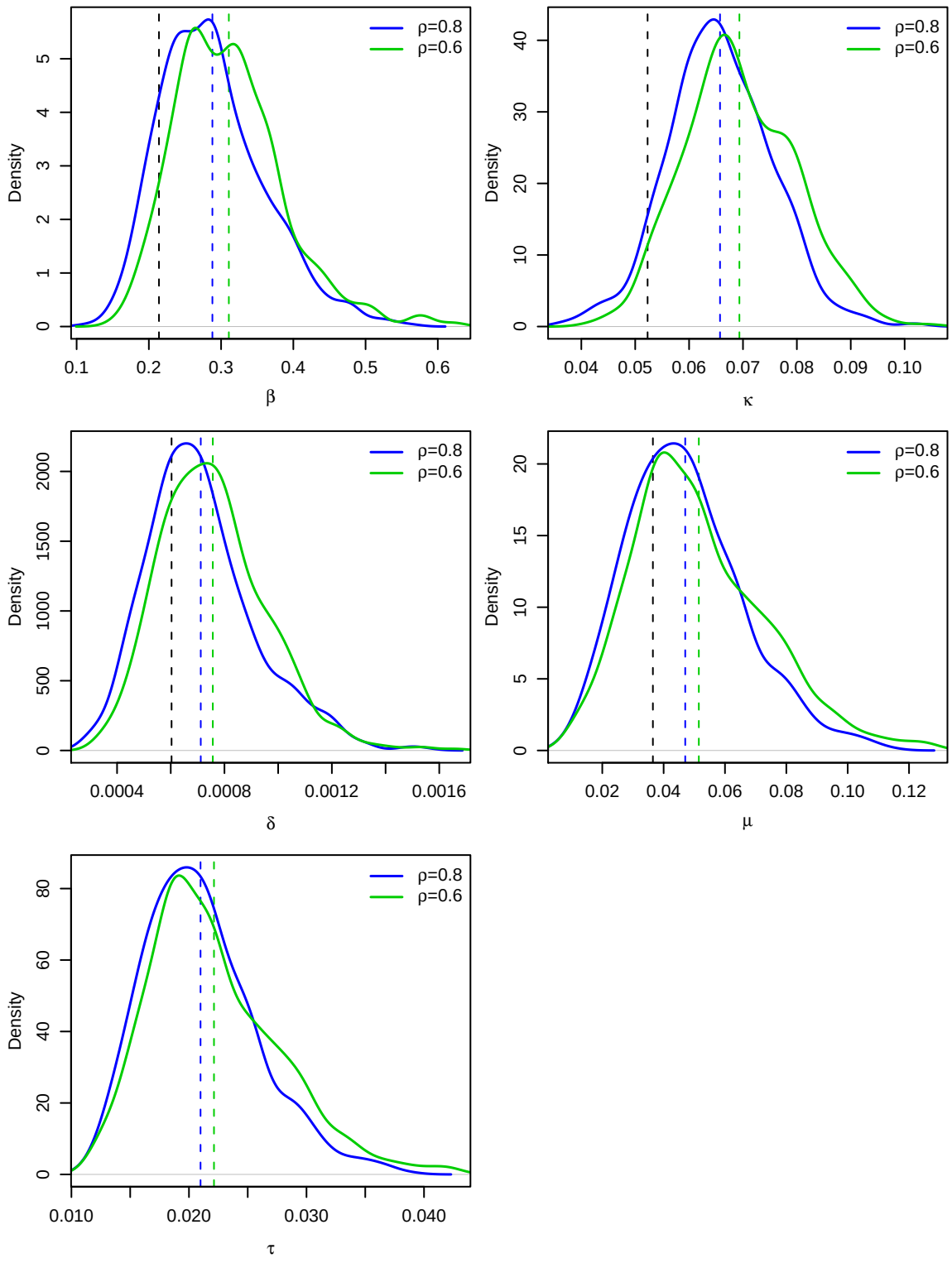

Figure C.3: Posterior densities for the estimated parameters ($\beta, \kappa, \delta, \mu$) and one derived parameter (τ, derived deterministically from μ) for model 2 in which the source of the external infection is the adjacent districts at a rate of 10 infections yr^{-1}. The blue and green lines represent the models with 80% ($\rho = 0.8$) and 60% ($\rho = 0.6$) reporting probabilities. The vertical dashed lines are the mean of the distributions, with the black line representing the mean based on the model that assumes perfect reporting.

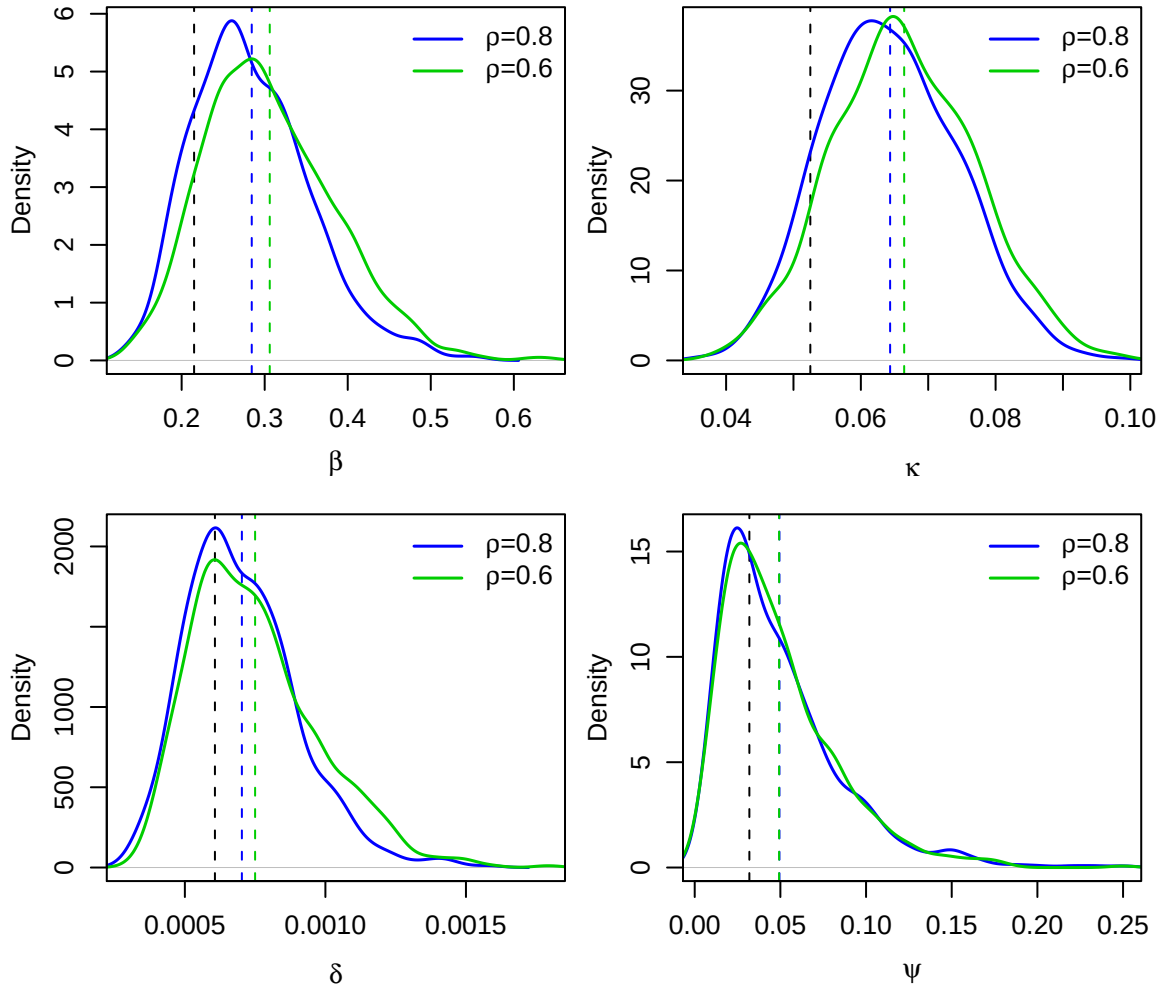

Figure C.4: Posterior densities for the estimated parameters $(\beta, \kappa, \delta, \psi)$ for model 3 in which the source of the external infection is randomly distributed at a rate of 10 infections yr^{-1}. The blue and green lines represent the models with 80% ($\rho = 0.8$) and 60% ($\rho = 0.6$) reporting probabilities. The vertical dashed lines are the mean of the distributions, with the black line representing the mean based on the model that assumes perfect reporting.

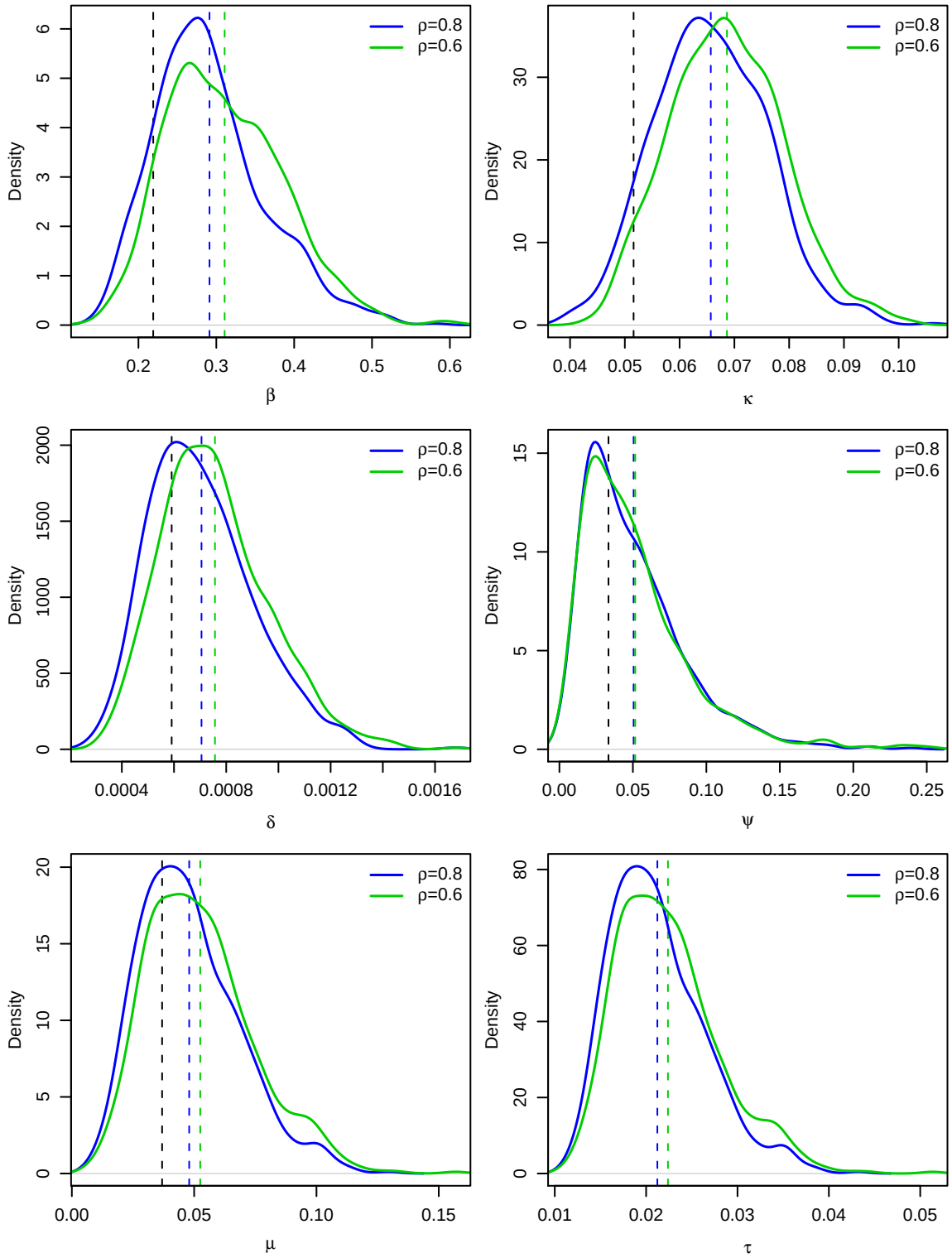

Figure C.5: Posterior densities for the estimated parameters ($\beta, \kappa, \delta, \psi, \mu$) and one derived parameter (τ, derived deterministically from μ) for model 3 in which the source of the external infection is the adjacent districts at a rate of 10 infections yr^{-1}. The blue and green lines represent the models with 80% ($\rho = 0.8$) and 60% ($\rho = 0.6$) reporting probabilities. The vertical dashed lines are the mean of the distributions, with the black line representing the mean based on the model that assumes perfect reporting.

Table C.3: Estimated mean parameter values and 95% credible intervals for the top four ranked rabies disease transmission models, assuming three levels of reporting probability ($\rho = 1.0$, 0.8, and 0.6). A dash represents a model in which that parameter was omitted. Note that $\rho = 1$ corresponds to the models presented in Tables 1 and 2.

Model	Infection source	Rate (yr^{-1})	ρ	β (95% CI) $\times 10^{-1}$	κ (95% CI) $\times 10^{-2}$	δ (95% CI) $\times 10^{-3}$	ψ (95% CI) $\times 10^{-1}$	μ (95% CI) $\times 10^{-1}$
2	R	10	1.0	2.11 (0.923, 4.31)	5.28 (2.89, 7.75)	0.608 (0.345, 1.49)	-	-
2	R	10	0.8	2.82 (1.75, 4.42)	6.44 (4.73, 8.39)	0.708 (0.400, 1.14)	-	-
2	R	10	0.6	3.00 (1.83, 4.65)	6.77 (4.80, 8.84)	0.766 (0.435, 1.18)	-	-
2	D	10	1.0	2.14 (0.936, 4.46)	5.23 (3.06, 7.76)	0.603 (0.344, 1.47)	-	0.365 (0.0858, 0.734)
2	D	10	0.8	2.88 (1.75, 4.53)	6.57 (4.71, 8.46)	0.712 (0.403, 1.18)	-	0.470 (0.166, 0.886)
2	D	10	0.6	3.11 (1.95, 4.85)	6.93 (5.15, 9.02)	0.757 (0.439, 1.17)	-	0.515 (0.176, 1.000)
3	R	10	1.0	2.15 (0.889, 4.55)	5.25 (2.90, 7.72)	0.607 (0.343, 1.63)	0.319 (0.0604, 1.21)	-
3	R	10	0.8	2.84 (1.78, 4.45)	6.43 (4.64, 8.45)	0.703 (0.389, 1.14)	0.495 (0.0923, 1.33)	-
3	R	10	0.6	3.06 (1.74, 4.72)	6.64 (4.63, 8.69)	0.750 (0.415, 1.27)	0.492 (0.0852, 1.30)	-
3	D	10	1.0	2.19 (0.884, 4.31)	5.16 (2.88, 7.55)	0.591 (0.346, 1.50)	0.334 (0.0607, 1.28)	0.369 (0.0826, 0.743)
3	D	10	0.8	2.91 (1.79, 4.51)	6.57 (4.74, 8.73)	0.705 (0.390, 1.15)	0.505 (0.0936, 1.35)	0.480 (0.176, 0.95)
3	D	10	0.6	3.11 (1.88, 4.68)	6.86 (5.02, 8.94)	0.757 (0.423, 1.22)	0.516 (0.0968, 1.47)	0.525 (0.186, 1.00)

References

Aghomo, H.O. & Rupprecht, C.E. 1990. Further-studies on rabies virus isolated from healthy dogs in Nigeria. *Veterinary Microbiology*, **22**, 17–22.

Anderson, D. & Watson, R. 1980. On the spread of a disease with gamma-distributed latent and infectious periods. *Biometrika*, **67**, 191–198.

Anderson, R.M. 1986. Rabies control - vaccination of wildlife reservoirs. *Nature*, **322**, 304–305.

Anderson, R.M. & May, R.M. 1991. *Infectious Diseases of Humans*. Oxford University Press.

Andersson, H. & Britton, T. 2000. Stochastic epidemics in dynamic populations: quasi-stationarity and extinction. *Journal of Mathematical Biology*, **41**, 559–580.

Asano, E., Gross, L.J., Lenhart, S. & Real, L.A. 2008. Optimal control of vaccine distribution in a rabies metapopulation model. *Mathematical Biosciences and Engineering*, **5**, 219–238.

Baer, G.M. & Cleary, W.F. 1972. Model in mice for pathogenesis and treatment of rabies. *Journal of Infectious Diseases*, **125**, 520–527.

Ball, F., Mollison, D. & Scalia-Tomba, G. 1997. Epidemics with two levels of mixing. *Annals of Applied Probability*, **7**, 46–89.

Baloul, L. & Lafon, M. 2003. Apoptosis and rabies virus neuroinvasion. *Biochimie*, **85**, 777–788.

Beaumont, M.A., Zhang, W.Y. & Balding, D.J. 2002. Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Beyer, H.L., Hampson, K., Lembo, T., Cleaveland, S. & Haydon, D.T. in review. Metapopulation dynamics of rabies and the efficacy of vaccination. *Proceedings Of The Royal Society of London B*.

Biek, R., Henderson, J.C., Waller, L.A., Rupprecht, C.E. & Real, L.A. 2007. A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **104**, 7993–7998.

Bolker, B. & Grenfell, B. 1995. Space, persistence and dynamics of measles epidemics. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences*, **348**, 309–320. (doi:10.1098/rstb.1995.0070).

Charlton, K.M., Casey, G.A. & Campbell, J.B. 1983. Experimental rabies in skunks - mechanisms of infection of the salivary-glands. *Canadian Journal of Comparative Medicine-Revue Canadienne De Medecine Comparee*, **47**, 363–369.

Charlton, K.M., NadinDavis, S., Casey, G.A. & Wandeler, A.I. 1997. The long incubation period in rabies: Delayed progression of infection in muscle at the site of exposure. *Acta Neuropathologica*, **94**, 73–77.

Cleaveland, S. & Dye, C. 1995. Maintenance of a microparasite infecting several host species: Rabies in the Serengeti. *Parasitology*, **111**, S33–S47. (doi:10.1017/S0031182000075806).

Cleaveland, S., Fevre, E.M., Kaare, M. & Coleman, P.G. 2002. Estimating human rabies mortality in the United Republic of Tanzania from dog bite injuries. *Bulletin of The World Health Organization*, **80**, 304–310.

Cleaveland, S., Kaare, M., Tiringa, P., Mlengeya, T. & Barrat, J. 2003. A dog rabies vaccination campaign in rural Africa: impact on the incidence of dog rabies and human dog-bite injuries. *Vaccine*, **21**, 1965–1973. (doi:10.1016/S0264-410X(02)00778-8).

Cleaveland, S., Mlengeya, T., Kaare, M., Haydon, D., Lembo, T., Laurenson, M.K. & Packer, C. 2007. The conservation relevance of epidemiological research into carnivore viral diseases in the Serengeti. *Conservation Biology*, **21**, 612–622. (doi:10.1111/j.1523-1739.2007.00701.x).

Coleman, P.G. & Dye, C. 1996. Immunization coverage required to prevent outbreaks of dog rabies. *Vaccine*, **14**, 185–186. (doi:10.1016/0264-410X(95)00197-9).

Coleman, P.G., Fevre, E.M. & Cleaveland, S. 2004. Estimating the public health impact of rabies. *Emerging Infectious Diseases*, **10**, 140–142.

Colizza, V. & Vespignani, A. 2008. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal Of Theoretical Biology*, **251**, 450–467. (doi:10.1016/j.jtbi.2007.11.028).

Cox, D.R. & Miller, H.D. 1965. *The Theory of Stochastic Processes*. John Wiley and Sons, Inc.

Cross, P.C., Johnson, P.L.F., Lloyd-Smith, J.O. & Getz, W.M. 2007. Utility of R-0 as a predictor of disease invasion in structured populations. *Journal of The Royal Society Interface*, **4**, 315–324. (doi:10.1098/rsif.2006.0185).

Cross, P.C., Lloyd-Smith, J.O., Johnson, P.L.F. & Getz, W.M. 2005. Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecology Letters*, **8**, 587–595.

Diekmann, O., Heesterbeek, J.A.P. & Metz, J.A.J. 1990. On the definition and the computation of the basic reproduction ratio r0 in models for infectious-diseases in heterogeneous populations. *Journal Of Mathematical Biology*, **28**, 365–382.

East, M.L., Hofer, H., Cox, J.H., Wulle, U., Wiik, H. & Pitra, C. 2001. Regular exposure to rabies virus and lack of symptomatic disease in serengeti spotted hyenas. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **98**, 15026–15031.

Eisinger, D. & Thulke, H.H. 2008. Spatial pattern formation facilitates eradication of infectious diseases. *Journal of Applied Ecology*, **45**, 415–423. (doi:10.1111/j.1365-2664.2007.01439.x).

Fekadu, M. 1975. Asymptomatic non-fatal canine rabies. *Lancet*, **1**, 569–569.

Fekadu, M. 1991. Latency and aborted rabies. In: *The Natural History of Rabies, 2nd edition* (ed. Baer, G.M.). CRC Press, chap. 9, pp. 191–198.

Fekadu, M., Shaddock, J.H. & Baer, G.M. 1981. Intermittent excretion of rabies virus in the saliva of a dog 2 and 6 months after it had recovered from experimental rabies. *American Journal Of Tropical Medicine And Hygiene*, **30**, 1113–1115.

Fekadu, M., Shaddock, J.H. & Baer, G.M. 1982. Excretion of rabies virus in the saliva of dogs. *Journal of Infectious Diseases*, **145**, 715–719.

Feng, Z.L., Xu, D.S. & Zhao, H.Y. 2007. Epidemiological models with non-exponentially distributed disease stages and applications to disease control. *Bulletin of Mathematical Biology*, **69**, 1511–1536. (doi:10.1007/s11538-006-9174-9).

Feng, Z.L., Yang, Y.D., Xu, D.S., Zhang, P., McCauley, M.M. & Glasser, J.W. 2009. Timely identification of optimal control strategies for emerging infectious diseases. *Journal Of Theoretical Biology*, **259**, 165–171. (doi:10.1016/j.jtbi.2009.03.006).

Ferguson, N.M., Donnelly, C.A. & Anderson, R.M. 2001. The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science*, **292**, 1155–1160. (doi:10.1126/science.1061020).

Ferguson, N.M., Keeling, M.J., Edmunds, W.J., Gant, R., Grenfell, B.T., Amderson, R.M. & Leach, S. 2003. Planning for smallpox outbreaks. *Nature*, **425**, 681–685. (doi:10.1038/nature02007).

Fowler, A.C. 2000. The effect of incubation time distribution on the extinction characteristics of a rabies epizootic. *Bulletin of Mathematical Biology*, **62**, 633–655.

Fulford, G.R., Roberts, M.G. & Heesterbeek, J.A.P. 2002. The metapopulation dynamics of an infectious disease: Tuberculosis in possums. *Theoretical Population Biology*, **61**, 15–29. (doi:10.1006/tpbi.2001.1553).

Gillespie, D.T. 1976. General method for numerically simulating stochastic time evolution of coupled chemical-reactions. *Journal of Computational Physics*, **22**, 403–434.

Grenfell, B.T., Bjornstad, O.N. & Kappey, J. 2001. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, **414**, 716–723. (doi:10.1038/414716a).

Hampson, K., Dobson, A., Kaare, M., Dushoff, J., Magoto, M., Sindoya, E. & Cleaveland, S. 2008. Rabies exposures, post-exposure prophylaxis and deaths in a region of endemic canine rabies. *Plos Neglected Tropical Diseases*, **2**, 1–9. (doi:10.1371/journal.pntd.0000339).

Hampson, K., Dushoff, J., Cleaveland, S., Haydon, D.T., Kaare, M., Packer, C. & Dobson, A. 2009. Transmission dynamics and prospects for the elimination of canine rabies. *PLoS Biology*, **7**, 462–471. (doi:10.1371/journal.pbio.1000053).

Hanlon, C.A., Niezgoda, M. & Rupprecht, C.E. 2007. Rabies in terrestrial animals. In: *Rabies (Second edition)* (eds. Jackson, A.C. & Wunner, H.W.). Elsevier, London, chap. 5, pp. 201–258.

Hanski, I. & Ovaskainen, O. 2000. The metapopulation capacity of a fragmented landscape. *Nature*, **404**, 755–758.

Haydon, D.T., Chase-Topping, M., Shaw, D.J., Matthews, L., Friar, J.K., Wilesmith, J. & Woolhouse, M.E.J. 2003. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proceedings of The Royal Society of London Series B-Biological Sciences*, **270**, 121–127.

Haydon, D.T., Kao, R.R. & Kitching, R.P. 2004. The UK foot-and-mouth disease outbreak - the aftermath. *Nature Reviews Microbiology*, **2**, 675–681. (doi:10.1038/nrmicro960).

Haydon, D.T., Randall, D.A., Matthews, L., Knobel, D.L., Tallents, L.A., Gravenor, M.B., Williams, S.D., Pollinger, J.P., Cleaveland, S., Woolhouse, M.E.J., Sillero-Zubiri, C., Marino, J., Macdonald, D.W. & Laurenson, M.K. 2006. Low-coverage vaccination strategies for the conservation of endangered species. *Nature*, **443**, 692–695. (doi:10.1038/nature05177).

Haydon, D.T., Woolhouse, M.E.J. & Kitching, R.P. 1997. An analysis of foot-and-mouth-disease epidemics in the UK. *Ima Journal of Mathematics Applied In Medicine and Biology*, **14**, 1–9. (doi:10.1093/imammb/14.1.1).

Heesterbeek, J.A.P. & Dietz, K. 1996. The concept of r0 in epidemic theory. *Statistica Neerlandica*, **50**, 89–110.

Heffernan, J.M., Smith, R.J. & Wahl, L.M. 2005. Perspectives on the basic reproductive ratio. *Journal Of The Royal Society Interface*, **2**, 281–293.

Hooper, D.C. 2005. The role of immune responses in the pathogenesis of rabies. *Journal of Neurovirology*, **11**, 88–92.

Jackson, A.C., Ye, H.T., Phelan, C.C., Ridaura-Sanz, C., Zheng, Q., Li, Z.S., Wan, X.Q. & Lopez-Corella, E. 1999. Extraneural organ involvement in human rabies. *Laboratory Investigation*, **79**, 945–951.

Johnson, N., Fooks, A. & McColl, K. 2008. Reexamination of human rabies case with long incubation, Australia. *Emerging Infectious Diseases*, **14**, 1950–1951.

Kao, R.R. 2002. The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends In Microbiology*, **10**, 279–286. (doi:10.1016/S0966-842X(02)02371-5).

Kao, R.R. 2003. The impact of local heterogeneity on alternative control strategies for foot-and-mouth disease. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, **270**, 2557–2564. (doi:10.1098/rspb.2003.2546).

Kao, R.R., Danon, L., Green, D.M. & Kiss, I.Z. 2006. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proceedings Of The Royal Society B-Biological Sciences*, **273**, 1999–2007.

Keeling, M.J. 2000. Metapopulation moments: coupling, stochasticity and persistence. *Journal of Animal Ecology*, **69**, 725–736.

Keeling, M.J. & Gilligan, C.A. 2000. Bubonic plague: a metapopulation model of a zoonosis. *Proceedings of The Royal Society of London Series B-Biological Sciences*, **267**, 2219–2230.

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

Keeling, M.J. & Grenfell, B.T. 1998. Effect of variability in infection period on the persistence and spatial spread of infectious diseases. *Mathematical Biosciences*, **147**, 207–226.

Keeling, M.J., Woolhouse, M.E.J., May, R.M., Davies, G. & Grenfell, B.T. 2003. Modelling vaccination strategies against foot-and-mouth disease. *Nature*, **421**, 136–142. (doi:10.1038/nature01343).

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J. & Grenfell, B.T. 2001. Dynamics of the 2001 uk foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science*, **294**, 813–817.

King, D.A., Peckham, C., Waage, J.K., Brownlie, J. & Woolhouse, M.E.J. 2006. Infectious diseases: Preparing for the future. *Science*, **313**, 1392–1393.

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. 1983. Optimization by simulated annealing. *Science*, **220**, 671–680.

Kitala, P.M., McDermott, J.J., Coleman, P.G. & Dye, C. 2002. Comparison of vaccination strategies for the control of dog rabies in Machakos District, Kenya. *Epidemiology and Infection*, **129**, 215–222. (doi:10.1017/S0950268802006957).

Knobel, D.L., Cleaveland, S., Coleman, P.G., Fevre, E.M., Meltzer, M.I., Miranda, M.E.G., Shaw, A., Zinsstag, J. & Meslin, F.X. 2005. Re-evaluating the burden of rabies in Africa and Asia. *Bulletin of The World Health Organization*, **83**, 360–368.

Knobel, D.L., Laurenson, M.K., Kazwala, R.R., Boden, L.A. & Cleaveland, S. 2008. A cross-sectional study of factors associated with dog ownership in Tanzania. *Bmc Veterinary Research*, **4**.

Lakhanpal, U. & Sharma, R.C. 1985. An epidemiologic-study of 177 cases of human rabies. *International Journal of Epidemiology*, **14**, 614–617. (doi:10.1093/ije/14.4.614).

Lembo, T., Hampson, K., Haydon, D.T., Craft, M., Dobson, A., Dushoff, J., Ernest, E., Hoare, R., Kaare, M., Mlengeya, T., Mentzel, C. & Cleaveland, S. 2008. Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem. *Journal of Applied Ecology*, **45**, 1246–1257. (doi:10.1111/j.1365-2664.2008.01468.x).

Lembo, T., Hampson, K., Kaare, M.T., Ernest, E., Knobel, D., Kazwala, R.R., Haydon, D.T. & Cleaveland, S. 2010. The feasibility of canine rabies elimination in Africa: dispelling doubts with data. *PLoS Neglected Tropical Diseases*, **4**, e626.

Lembo, T., Haydon, D.T., Velasco-Villa, A., Rupprecht, C.E., Packer, C., Brandao, P.E., Kuzmin, I.V., Fooks, A.R., Barrat, J. & Cleaveland, S. 2007. Molecular epidemiology identifies only a single rabies virus variant circulating in complex carnivore communities of the Serengeti. *Proceedings of The Royal Society B-Biological Sciences*, **274**, 2123–2130.

Lloyd, A.L. 1996. *Mathematical models for spatial heterogeneity in population dynamics and epidemiology.* Ph.D. thesis, DPhil. thesis, University of Oxford, UK.

Lloyd, A.L. 2001a. The dependence of viral parameter estimates on the assumed viral life cycle: limitations of studies of viral load data. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, **268**, 847–854.

Lloyd, A.L. 2001b. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, **268**, 985–993.

Lloyd, A.L. 2001c. Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics. *Theoretical Population Biology*, **60**, 59–71.

Lloyd, A.L. & May, R.M. 1996. Spatial heterogeneity in epidemic models. *Journal of Theoretical Biology*, **179**, 1–11. (doi:10.1006/jtbi.1996.0042).

Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. 2000. WinBUGS - a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337. (doi:10.1023/A:1008929526011).

MacQueen, J.B. 1967. Some methods for classification and analysis of multivariate observations. In: *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, vol. 1, pp. 281–297.

Marjoram, P., Molitor, J., Plagnol, V. & Tavare, S. 2003. Markov chain Monte Carlo without likelihoods. *Proceedings of The National Academy of Sciences of The United States of America*, **100**, 15324–15328.

Massad, E., Burattini, M.N. & Ortega, N.R.S. 1999. Fuzzy logic and measles vaccination: designing a control strategy. *International Journal of Epidemiology*, **28**, 550–557.

Matthews, L., Haydon, D.T., Shaw, D.J., Chase-Topping, M.E., Keeling, M.J. & Woolhouse, M.E.J. 2003. Neighbourhood control policies and the spread of infectious diseases. *Proceedings of The Royal Society of London Series B-Biological Sciences*, **270**, 1659–1666. (doi:10.1098/rspb.2003.2429).

McCallum, H., Barlow, N. & Hone, J. 2001. How should pathogen transmission be modelled? *Trends In Ecology & Evolution*, **16**, 295–300.

Medlock, J. & Galvani, A.P. 2009. Optimizing influenza vaccine distribution. *Science*, **325**, 1705–1708. (doi:10.1126/science.1175570).

Mollison, D. 1991. Dependence of epidemic and population velocities on basic parameters. *Mathematical Biosciences*, **107**, 255–287.

Murphy, F.A. 1977. Rabies pathogenesis. *Archives of Virology*, **54**, 279–297.

Nel, L.H. & Markotter, W. 2007. Lyssaviruses. *Critical Reviews In Microbiology*, **33**, 301–324.

Nokes, D.J. & Swinton, J. 1997. Vaccination in pulses: A strategy for global eradication of measles and polio? *Trends In Microbiology*, **5**, 14–19.

Ovaskainen, O. & Hanski, I. 2001. Spatially structured metapopulation models: Global and local assessment of metapopulation capacity. *Theoretical Population Biology*, **60**, 281–302.

Ovaskainen, O. & Hanski, I. 2003. How much does an individual habitat fragment contribute to metapopulation dynamics and persistence? *Theoretical Population Biology*, **64**, 481–495.

Park, A.W., Gubbins, S. & Gilligan, C.A. 2001. Invasion and persistence of plant parasites in a spatially structured host population. *Oikos*, **94**, 162–174. (doi:10.1034/j.1600-0706.2001.10489.x).

Park, A.W., Gubbins, S. & Gilligan, C.A. 2002. Extinction times for closed epidemics: the effects of host spatial structure. *Ecology Letters*, **5**, 747–755.

Pritchard, J., Seielstad, M.T., Perez-Lezaun, A. & Feldman, M.W. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, **16**, 17911798.

R Development Core Team 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Randall, D.A., Marino, J., Haydon, D.T., Sillero-Zubiri, C., Knobel, D.L., Tallents, L.A., Macdonald, D.W. & Laurenson, M.K. 2006. An integrated disease management strategy for the control of rabies in Ethiopian wolves. *Biological Conservation*, **131**, 151–162. (doi:10.1016/j.biocon.2006.04.004).

Randall, D.A., Williams, S.D., Kuzmin, I.V., Rupprecht, C.E., Tallents, L.A., Tefera, Z., Argaw, K., Shiferaw, F., Knobel, D.L., Sillero-Zubiri, C. & Laurenson, M.K. 2004. Rabies in endangered Ethiopian wolves. *Emerging Infectious Diseases*, **10**, 2214–2217.

Roberts, M.G. 1996. The dynamics of bovine tuberculosis in possum populations, and its eradication or control by culling or vaccination. *Journal Of Animal Ecology*, **65**, 451–464. (doi:10.2307/5780).

Roy, A. & Hooper, D.C. 2008. Immune evasion by rabies viruses through the maintenance of blood-brain barrier integrity. *Journal of Neurovirology*, **14**, 401–411.

Rupprecht, C.E., Hanlon, C.A. & Hemachudha, T. 2002. Rabies re-examined. *Lancet Infectious Diseases*, **2**, 327–343.

Shankar, V., Dietzschold, B. & Koprowski, H. 1991. Direct entry of rabies virus into the central-nervous-system without prior local replication. *Journal Of Virology*, **65**, 2736–2738.

Sisson, S.A. 2007. Genetics and stochastic simulation do mix! *American Statistician*, **61**, 112–119.

Spiegelhalter, D.J., Best, N.G., Carlin, B.R. & van der Linde, A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **64**, 583–616. (doi:10.1111/1467-9868.00353).

Swinton, J. 1998. Extinction times and phase transitions for spatially structured closed epidemics. *Bulletin of Mathematical Biology*, **60**, 215–230. (doi:10.1006/bulm.1997.0014).

Tepsumethanon, V., Lumlertdacha, B., Mitmoonpitak, C., Sitprija, V., Meslin, F.X. & Wilde, H. 2004. Survival of naturally infected rabid dogs and cats. *Clinical Infectious Diseases*, **39**, 278–280.

Tildesley, M.J., Savill, N.J., Shaw, D.J., Deardon, R., Brooks, S.P., Woolhouse, M.E.J., Grenfell, B.T. & Keeling, M.J. 2006. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the uk. *Nature*, **440**, 83–86.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M.P.H. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal Of The Royal Society Interface*, **6**, 187–202.

Tsiang, H., Ceccaldi, P.E. & Lycke, E. 1991. Rabies virus-infection and transport in human sensory dorsal-root ganglia neurons. *Journal Of General Virology*, **72**, 1191–1194.

Vial, F., Cleaveland, S., Rasmussen, G. & Haydon, D.T. 2006. Development of vaccination strategies for the management of rabies in African wild dogs. *Biological Conservation*, **131**, 180–192. (doi:10.1016/j.biocon.2006.04.005).

Wang, Z.W., Sarmento, L., Wang, Y.H., Li, X.Q., Dhingra, V., Tseggai, T., Jiang, B.M. & Fu, Z.F. 2005. Attenuated rabies virus activates, while pathogenic rabies virus evades, the host innate immune responses in the central nervous system. *Journal Of Virology*, **79**, 12554–12565.

Watts, D.J., Muhamad, R., Medina, D.C. & Dodds, P.S. 2005. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **102**, 11157–11162.

Wearing, H.J., Rohani, P. & Keeling, M.J. 2005. Appropriate models for the management of infectious diseases. *Plos Medicine*, **2**, 621–627.

Whittle, P. 1955. The outcome of a stochastic epidemic. *Biometrika*, **42**, 116–122.

Woodroffe, R. 2001. Assessing the risks of intervention: immobilization, radiocollaring and vaccination of African wild dogs. *Oryx*, **35**, 234–244.

Zhang, Y.Z., Fu, Z.F., Wang, D.M., Zhou, J.Z., Wang, Z.X., Lv, T.F., Xiong, C.L., Zou, Y., Yao, W.R., Li, M.H., Dong, G.M., Xu, G.L., Niezgoda, M., Kuzmin, I.V. & Rupprecht, C.E. 2008. Investigation of the role of healthy dogs as potential carriers of rabies virus. *Vector-Borne and Zoonotic Diseases*, **8**, 313–319.

Zinsstag, J., Durr, S., Penny, M.A., Mindekem, R., Roth, F., Gonzalez, S.M., Naissengar, S. & Hattendorf, J. 2009. Transmission dynamics and economics of rabies control in dogs and humans in an African city. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **106**, 14996–15001. (doi:10.1073/pnas.0904740106).