Chapter 3
Assessment of Transmission Rates and Routes, and the Implications for Management

Peter Caley, Glenn Marion, and Michael R. Hutchings

3.1 Introduction

3.1.1 Preamble

Being able to estimate disease transmission rates and determine the underlying mechanisms of transmission is fundamental to the effective management of wildlife disease – transmission rates drive disease dynamics and persistence, and thus determine the level of control or vaccination necessary to achieve disease eradication, or predict the likely impact of a biocontrol agent. The mechanisms of transmission determine where management efforts can be targeted. Not knowing and not being able to estimate transmission rates when trying to manage disease in wildlife is analogous to managing overpopulated wildlife without knowing the intrinsic rate of population increase. Being able to estimate transmission rates allows us to determine whether management actions are achieving their aims. This chapter looks at the measures of disease transmission and how they can be calculated. We recommend that the non-mathematical readers skim through Section 3.2 without trying to follow the mathematics, and refer back to it when needed.

3.1.2 Measures of Transmission

The term disease transmission means many things and can be quantified in different ways. Exactly what measure is required will depend on the aims of the investigator/manager. The following terms are all measures that result from disease transmission:

P. Caley
National Centre for Epidemiology & Population Heath, Australian National University, ACT 0200, Australia

G. Marion
Biomathematics Statistics Scotland, The Kings Buildings, Edinburgh, EH9 3JG, UK

M.R. Hutchings
SAC, West Mains Road, Edinburgh, EH9 3JG, UK

R.J. Delahay et al. (eds.), *Management of Disease in Wild Mammals,*
DOI:10.1007/978-4-431-77134-0_3, © Springer 2009

Force of Infection (λ) – the instantaneous *per capita* rate at which susceptible individuals acquire infection. Also called the instantaneous incidence.
Basic Reproduction Number (R_0) – the expected number of secondary infections produced by a typical infected individual over the course of their infectious period when among a population where every individual is equally susceptible. Also called the basic reproduction ratio and basic reproduction rate.
Effective Reproduction Number (R) – the actual number of secondary infections produced by an infectious individual.
Disease Prevalence (p) – the proportion of the population that is infected at a given time.
Attack Rate (α) – the proportion of the population infected over the course of an epizootic.
Transmission Coefficient (β) – the model-dependent constant that as part of the transmission function, determines the rate at which susceptible individuals become infected in the population.

Note that despite being related, knowing the value of one measure of transmission does not necessarily mean any other is also known. Also, in general, measures such as p and λ are dependent on the prevailing conditions (e.g. numbers of infectives and susceptibles) – they are not fixed parameters. Conversely, R_0, whilst essentially being a fixed parameter that underlies the number of secondary infections an infected individual produces (which is a random variable), is often specific to the population from which it is estimated, and usually changes with host density or numbers. Hence the usefulness of R_0, for all its laudings, becomes tempered when applied to wildlife populations inhabiting different environments and/or locations from those used in its estimation. An analogous problem occurs with the intrinsic (maximum) rate of increase (r_m) of a wildlife population, which is specific to the particular environment in which it is measured (Caughley and Birch 1971). To get any measure of transmission that can be generalised to changed conditions (e.g. post-intervention or to a different population) requires that we relate these measures to an underlying model of transmission (described by the transmission function) that can account for changed conditions. In its most basic form, the transmission function describes how transmission scales with population size and/or density and is where the transmission coefficient β is found. As such, β is typically the only intrinsic measure of transmission. It is also the most difficult to estimate; estimation necessarily being achieved via a model. Commonly considered forms of the transmission function are shown in Table 3.1. Clearly these functions do not accommodate variation in transmission relating to factors such as environmental conditions influencing pathogen survival, strain-specific differences in transmission, population immunity/susceptibility or local influences on the spatial arrangement (and hence mixing) of hosts. Most of these influences are all subsumed within β, which is typically assumed a constant (or if using Bayesian statistics, a distribution expressing belief in it's likely value). Violation of these assumptions may go partway to explaining why transmission rates often differ between sites for unknown reasons; underlining the simplistic nature of our models in many cases. It should also be stressed that where spatial information is available it is possible to infer contact rates within spatial

Table 3.1 Proposed forms for the transmission function. Adapted from McCallum et al. (2001) and references therein (reproduced with permission)

Number	Function[a]	Comments
1	βsi	Density-dependent transmission (also termed mass action)
2	$\beta si/n$	Frequency-dependent transmission
3	$\beta s^p i^q$	Power relationship; constants: $0 < p < 1$, $0 < q < 1$. Phenomenological. Sometimes considered to account for spatial effects such as local depletion of susceptibles
4	$\beta i(n - 1/q)$	Constant: $0 < q < 1$. Embodies a refuge effect (q = proportion of population potentially susceptible)
5	$ks \ln\left(1 + \frac{\beta i}{k}\right)$	Negative binomial. Small k corresponds to highly aggregated infection. As $k \to \infty$, expression reduces to Function 1
6	$\frac{n}{1 - \varepsilon + \varepsilon n} \frac{F(s,i)}{n}$	Asymptotic contact function separated from the mixing term $F(s,i)$ which may be any of Functions 1–5 above. If the constant $\varepsilon = 0$, the contact rate is proportional to n. If $\varepsilon = 1$, contacts are independent of n
7	$\frac{\beta si}{c + s + i}$	Asymptotic transmission where c is a constant

[a] i is the density of infected hosts, s is the density of susceptible hosts, and n is the total host density. β is the transmission coefficient. Other parameters are described where necessary under comments.

models. Additionally, if a management intervention aims to change the behaviour of animals (e.g. increasing mating frequency as reported by Caley and Ramsey (2001)), then clearly β will change and can no longer be considered fixed.

3.1.3 Practical Difficulties in Field Measurement

Disease transmission is typically an unobservable event – even if we observed a known infected "contacting" a known susceptible, we would be none the wiser as to whether transmission occurred. Thus we have to infer transmission from observable data such as evidence of prior or current infection (e.g. diagnostic testing) or surrogate markers for infection such as the onset of clinical signs or death. Such data usually requires that animals can be captured and samples collected, or easily observed. Obtaining such data for free-ranging mammals is often problematic, making large-scale replicated field experiments difficult and smaller pen studies the only feasible type of experimentation. Considerable difficulties, however, are often experienced when extrapolating transmission rates estimated from experimental conditions to field conditions (McCallum 2000).

Estimating epidemic threshold parameters (e.g. critical host population size N_T or critical host density K_T) from whether an introduced pathogen establishes and gives rise to a major epizootic as opposed to a minor epizootic is difficult to achieve experimentally (Lloyd-Smith et al. 2005b). The reasons being the epizootics are by nature dichotomous (either the epizootic is major with many cases or is minor with a trivial number of cases)

and stochastic (an epidemic may not take off despite $R_0 > 1$). Many experiments may therefore be required to estimate where the threshold may lie with any kind of precision. The result is that many researchers are forced to undertake observational experiments of host/pathogen systems as a means of quantifying disease transmission.

Exactly what measure of disease transmission is estimated will depend on the aims of the investigation and logistical constraints. If the aim is simply to determine whether a management intervention is reducing transmission or whether a particular pathway of transmission occurs (a hypothesis testing question), then bias may not be problem and precision more important. Fitting more parsimonious models is a way of achieving this (though increasing bias). For example, ignoring the effect of disease-induced mortality when modelling age-prevalence data biases estimates of the force of infection downwards, though it facilitates straightforward model fitting (Caley and Hone 2002). If the bias of an estimator is consistent across experimental treatments, then such an estimator may suffice for estimating relative changes in underlying transmission. If the purpose of the investigation is to identify risk factors contributing to disease transmission (as typically measured by either the prevalence or time to infection), then robust statistical frameworks such as logistic regression (e.g. Joly and Messier 2004) or Cox's proportional hazards model (e.g. Calvete et al. 2004b) will suffice. Such models typically do not explicitly include a transmission component and hence cannot be used to estimate rates of transmission. Conversely, if the aim is to investigate predicted changes to the host(s)/pathogen system of a mechanistic nature (e.g. introducing vaccination), then unbiased estimates of transmission coefficients will be required along with knowledge of the correct underlying transmission function, and models will need to be specified accordingly.

3.2 Estimating Transmission Rates for Directly Transmitted Pathogens

Quantifying disease transmission is simplest for directly transmitted pathogens, particularly if only one or two hosts are involved, and this is the focus of this section.

3.2.1 Estimating the Force of Infection (λ)

The force of infection experienced by a susceptible individual will depend on the infection status of other individuals that the susceptible mixes with (as quantified by prevalence of infection or density of infectives), and the form of the transmission function. For this reason, estimates of λ in isolation are of little use for quantifying underlying transmission rates. However, relating λ via a model to host density and the relative abundance of susceptibles and infecteds, in combination with other demographic parameters, is a practical approach for estimating parameters (e.g. β) that determine transmission rates (McCallum et al. 2001).

Using Age-Prevalence Data In general, methods to estimate disease transmission rates from age-prevalence data assume steady-state (c.f. epizootic) conditions.

This is a strong assumption that needs to be applied with care, as it is difficult to distinguish between age-dependent and time-dependent variation in disease incidence. Most models developed for analysing age-specific prevalence data were developed for diseases of humans, and assume that mortality due to infection can be ignored (e.g. Farrington et al. 2001). This is less often the case with wildlife diseases, and accounting for disease-induced mortality introduces additional complications. Disease-induced mortality tends to flatten age-prevalence curves (Heisey et al. 2006) as does loss of evidence of prior infection (or recovery from infection for chronic diseases), resulting in the force of infection being underestimated if ignored. This may not be a problem in a hypothesis testing application (e.g. answering "does the intervention significantly reduce transmission?"), but will be an issue if estimation is the main aim of the investigation (Caley and Hone 2002).

If disease-induced mortality can be ignored, and the system is in equilibrium, then the modelled probability of an individual being infected (or showing signs of past infection) at age a when subjected to age-dependent force of infection $\lambda(a)$ is

$$p(a) = 1 - \exp\left(- \int_{t=0}^{t=a} \lambda(t) dt \right) \tag{3.1}$$

The form of $\lambda(a)$ may be as simple or complex (data willing) as the scientific investigation requires, and may change as a function of age, time and other covariates. The underlying form chosen for $\lambda(a)$ may be flexible (e.g. Grenfell and Anderson 1985; Heisey et al. 2006) or consistent with how transmission is known or hypothesised to occur (e.g. Caley and Hone 2002 and see Box 3.1). For simple forms of $\lambda(a)$ it is often possible to express Eq. (3.1) as a generalised linear model and obtain estimates of λ and factors influencing it directly (see Box 3.2). For more complex forms of $\lambda(a)$ and if additional demographic parameters are included, analytical solutions for the prevalence usually do not exist and numerical methods are used, although the parameters may still be estimated via standard maximum likelihood techniques. For n samples of individuals of ages a_j ($j = 1, \ldots, n$) where each sample contains N_j individuals of which I_j are infected (or shows signs of previous infection), the likelihood assuming that the probability of infection for a given age is binomially distributed is

$$L = \prod_{j=1}^{n} p\left(a_j\right)^{I_j} \left(1 - p\left(a_j\right)\right)^{N_j - I_j} \tag{3.2}$$

Maximum likelihood estimates of the parameters are obtained by minimising the negative of the log-likelihood function with respect to the parameters that determine $p(a)$:

$$-\ln(L) = -\sum_{j=1}^{n} \left[I_j \ln\left(p\left(a_j\right)\right) + \left(N_j - I_j\right) \ln\left(1 - p\left(a_j\right)\right) \right]. \tag{3.3}$$

This is usually achieved numerically by a standard numerical algorithm. Likelihood theory also enables estimation of the precision of these estimates, and comparison of models via likelihood ratio tests or information-theoretic methods (e.g. Akaike's Information Criterion). Alternatively, the likelihood function may be used within a Bayesian estimation framework (e.g. Markov Chain Monte Carlo) to obtain posterior

Box 3.1 Estimating the rates of rabbit to rabbit transmission of *Mycobacterium avium* subspecies *paratuberculosis (Map)*

European rabbits (*Oryctolagus cunniculus*) have been increasingly linked to the persistence of *Mycobacterium avium* subspecies *paratuberculosis* (*Map)* (Johne's disease) in domestic ruminants in the UK. Quantifying the routes of rabbit to rabbit transmission of *Map* is a key step to establishing whether rabbits are a persistent source of infection (i.e. a reservoir). Judge et al. (2006) fitted an SI (Susceptible-Infected) epidemiological model to field data to estimate the probabilities of vertical (vertical + pseudo-vertical) and horizontal transmission. *Map* was isolated from various tissues and excreta from a study site in Scotland suggesting the potential for vertical, pseudo-vertical and horizontal rabbit-to-rabbit transmission routes. The overall prevalence of *Map* in rabbits was high at both sites studied, with an average of 39.7%.

Estimating rates of transmission: A maximum likelihood fitting procedure was used to fit the SI model to the data on the proportion of infected rabbits per age group (2 month blocks) from the random sample to derive probabilities of vertical/pseudo-vertical and horizontal transmission (Fig. 3.1).

In order to model the variation of the mean infection prevalence with age, Judge et al. (2006) assumed that both the number of individuals at any given age and the number of infected individuals at any given age remain constant at least on the time scale of an individual's lifetime. This was consistent with the finding that the overall prevalence of infection in rabbits did not increase across the years of sampling (Judge et al. 2005a). Given this assumption it was then possible to pool the prevalence data taken on each visit and treat the inferred prevalence

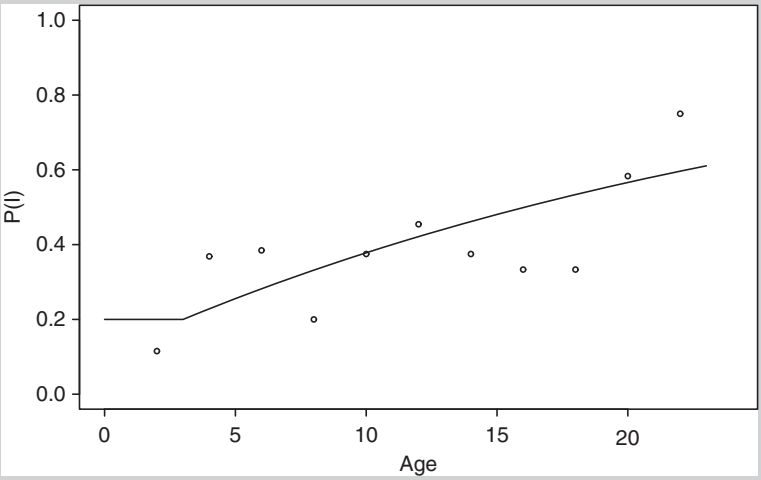

Fig. 3.1 Fitted prevalence of *Map* infection in rabbits as a function of age. Data are categorised in age ranges of two months (from Judge et al. 2006)

for each age as being equal to the prevalence that would be measured if it were possible to track a cohort of individuals from birth to death, measuring the prevalence in that cohort. A model for the spread of disease over time in a group of individuals exposed to a constant level of infection could therefore be used.

The model was constructed by assuming a two-stage infection process; individuals are exposed only to vertical and pseudo-vertical infection up until time t_0, when all vertical/pseudo-vertical infection ceases and they become exposed to horizontal infection by infected rabbits they are in contact with. The absence of sufficient data from pre-weaned individuals prevented using a detailed model of the vertical processes, so the combined effect of vertical and pseudo-vertical transmission was represented by a single probability P_v that individuals are infected at age t_0. The horizontal infection process was modelled as a homogeneous Poisson process (representing the simplest mathematical form for horizontal infection within a homogeneously mixing social group of rabbits, see below for group size) with a constant infection rate in which I is the (constant) number of infected individuals in the population as a whole and λ_0 is the per capita rate of infection.

$$\lambda_0 = \beta I \tag{3.17}$$

In a homogeneous Poisson process with rate parameter λ_0, the probability that an event occurs in the time interval $(0, \tau)$ is

$$1 - e^{-\lambda_0 \tau}. \tag{3.18}$$

Including the effect of vertical transmission there are two ways that an individual could be infected at time τ – by being infected vertically/pseudo-vertically from its mother, or horizontally, with a combined probability

$$P_v + (1 - P_v)\left(1 - e^{-\lambda_0 \tau}\right) \tag{3.19}$$

whereas, in order to escape infection up to time τ an individual must avoid infection through both routes, leading to a probability of being uninfected

$$(1 - P_v)\left(e^{-\lambda_0 \tau}\right). \tag{3.20}$$

Combining these probabilities with the data, the likelihood

$$\begin{aligned} L(\lambda_0, P_v, t_0) = \prod_{i=1}^{N} \Big\{ & I\left[y_i = 1\right]\left(\mathrm{P_n} + \mathrm{I}\left[\mathrm{t} > \mathrm{t}_0\right](1 - \mathrm{P_n})\left(1 - \mathrm{e}^{-\mathrm{I}_0(\mathrm{t_i} - \mathrm{t}_0)}\right)\right) \\ & + I\left[y_i = 0\right](1 - P_v)\left(I\left[t \le t_0\right] + I\left[t > t_0\right]e^{-\lambda_0 (t_i - t_0)}\right)\Big\}, \end{aligned} \tag{3.21}$$

is formed which was maximised numerically in order to obtain maximum likelihood estimates of the parameters λ_0 and P_v. Note that in practice it is the negative of the logarithm of the likelihood that is minimized. The data consist

(continued)

Box 3.1 (continued)

of the infection status y_i ($y_i = 1$ corresponds to infection and $y_i = 0$ to susceptible) of $I = 1$, N individuals and their estimated ages t_i. $I[..]$ denotes the indicator function which is unity if the expression in square brackets is true and zero otherwise. The first line of Eq. (3.21) corresponds to the probability that susceptibles become infected, whilst the second line represents the probability that susceptibles escape infection.

Maximum likelihood estimates were $\lambda_0 = 0.037$ and $P_v = 0.14$ when using a weaning age of $t_0 = 1$ month. These values can be expressed in terms of the underlying transmission probabilities. This per capita rate of horizontal infection per month (λ_0) is specific to the study site and will vary depending on the number of infectious (I) and susceptible animals in regular contact. The generic horizontal transmission coefficient per month (β) can be estimated as

$$\begin{aligned} \lambda_0 &= \beta I = \beta N p \\ \beta &= \lambda_0 / Np, \end{aligned} \tag{3.22}$$

where p is the overall prevalence and N is the total population size.

Adult rabbit social group sizes at the study site were conservatively estimated at between 2 and 7 individuals, equating to a conservative β value range of 0.013 to 0.046. The proportion of individuals entering the population after weaning (at 1 month old), which were infected via vertical and/or pseudo-vertical transmission (P_v), estimated from the maximum likelihood procedure, was 0.14. As only offspring from infected does can be infected vertically or pseudo-vertically, the probability of transmission via these routes can be calculated from the proportion of infected juveniles entering the population after weaning and the proportion of infected females of reproductive age. There was no significant difference in the prevalence of *Map* between sexes at either site therefore it was assumed that equal percentages of males and females were infected with *Map*. For adults of reproductive age (i.e. >6 months), 42.9% (85/198) were *Map* positive. Assuming that there is no effect of *Map* infection on either reproductive output or juvenile survival, this gives a probability of infection via vertical and/or pseudo-vertical transmission of up to 0.326 (14% of young infected when entering the population at 1 month /42.9% of infected females of reproductive age). These estimates of rabbit-to-rabbit routes of *Map* transmission were subsequently used in a modelling study to show that infection is highly persistent in rabbit populations (Judge et al. 2007) a critical step in understanding the role of rabbits in the epidemiology of paratuberculosis within the host community as a whole.

distributions for the parameters of interest, and incorporate prior belief regarding parameters (if available). Such models can be compared using the Bayesian Information Criterion (BIC) or deviance information criterion (DIC) as appropriate.

Using Longitudinal Data Estimating the force of infection from prospective studies of individuals (i.e. susceptible individuals are followed and their time to infection

Box 3.2 *Mycobacterium bovis* (bTB) in wild pigs – testing for treatment effects

The study data under consideration (Table 3.2 and Fig. 3.2) come from the Northern Territory, Australia, and estimate the proportion of wild pigs (*Sus scrofa*) with visible lesions typical of bovine tuberculosis (caused by *M. bovis*) during two territory-wide surveys. The first survey during the early 1970s (Corner et al. 1981), occurred at a time when bovine tuberculosis was highly prevalent in sympatric populations of wild cattle (*Bos spp*.) and water buffalo (*Bubalus bubalis*). The high prevalence observed in pigs was hypothesised to be a result of their association with these infected bovid populations. Subsequently, the populations of cattle and buffalo were dramatically reduced as part of the Brucellosis & Tuberculosis Eradication Campaign (BTEC). The second survey was undertaken in 1992, with the aim of determining whether

Table 3.2 Prevalence of wild pigs with lesions resembling bovine tuberculosis by age (in years). Adapted from McInerney et al. (1995)

Survey	Age	Sampled	Lesioned
1	0.5	128	21
1	1.5	132	59
1	2.5	117	55
1	3.5	83	47
1	4.5	105	66
1	5.5	82	56
1	6.5	45	35
2	0.5	251	8
2	1.5	227	9
2	2.5	131	10
2	3.5	113	13
2	4.5	38	2
2	5.5	16	4
2	6.5	14	3

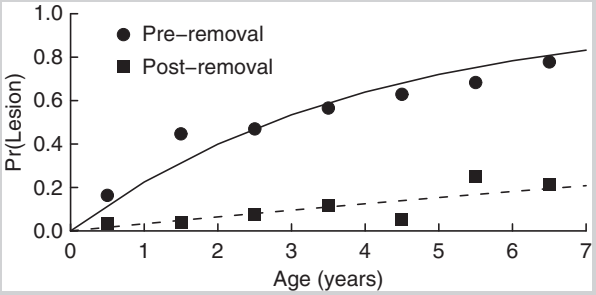

Fig. 3.2 Prevalence of lesions typical of bovine tuberculosis in wild pigs before (pre-removal) and after (post-removal) culling of sympatric cattle and buffalo populations known to be infected

(continued)

Box 3.2 (continued)

BTEC had indeed reduced the level of disease in wild pigs as predicted (McInerney et al. 1995).

There appears to be a difference in the age-specific prevalence of lesions between the two surveys (Table 3.2). How do we quantify this difference in terms of an underlying model that accounts for this data? This model is determined by how the age-specific prevalence relates to the force of infection. Assuming that animals are exposed to a constant force of infection from birth, then the prevalence (p) at a given age (a) is

$$p(a) = 1 - e^{-\lambda a} \tag{3.23}$$

This equation can be linearised with several simple algebraic operations:

$$\ln\left(-\ln\left(1 - p(a)\right)\right) = \ln(\lambda) + \ln(a) \tag{3.24}$$

This equation is straightforward to fit as a generalised linear model (see Crawley 2005 for details). The left hand side of this equation is what is known as a complementary log-log transformation of p. Hence the link function within the GLM is specified as complementary log-log. The age-dependent probability of infection ($p(a)$) is binomially distributed – so the error structure is specified as binomial. By specifying $\ln(a)$ as an offset (equivalent to fixing its slope to 1), we can directly assess the effect of factors and other explanatory variables on $\ln(\lambda)$. The steps to fitting this model are

1. Specify the proportion infected as the **response variable**.
2. Specify the **error structure** as binomial.
3. Specify the **link function** as complementary log-log.
4. Specify $\ln(a)$ as an **offset**.
5. Fit the model.

Two models are fitted, the first without any treatment effect, and the second including the factor "Survey" (which is a proxy variable for the removal of Tb positive buffalo & cattle). The inclusion of "Survey" is highly significant based on a likelihood ratio test ($\chi^2 = 266.9$, d.f. =1, $P < 0.001$). It is, however, the parameter estimates that are of most interest (Table 3.3).

Table 3.3 Parameter estimates and their standard errors

Parameter	Estimate	Standard error	Z value	Pr(>|z|)
Intercept	−1.37	0.06	−23.75	<2e − 16
Survey	−2.03	0.15	−13.19	<2e − 16

Note that the parameter estimates in Table 3.3 are on a logarithmic scale. That is

$$\lambda = e^{-1.37} = 0.26/Year \qquad \text{(Survey 1 - Pre-removal)}$$

$$\lambda = e^{-1.37-203} = 0.033/Year \qquad \text{(Survey 1 - Pre-removal)}$$

That is, force of infection post bovid removal was about 13% of that in pre bovid removal times – a substantial and statistically significant reduction (note though the lack of an experimental control in this simple before-after experimental design).

This basic exponential model fitted in this way can be expanded to include further factors and covariates. However, if for example, the mortality rate of animals increases as a result of being diseased, then the new expression for the prevalence of infection is intrinsically non-linear and unable to be fitted as a GLM. It can still, however, be fitted by standard maximum likelihood techniques (see Caley and Hone 2002 and Heisey et al. 2006 for details).

measured, or their infection status after a known length of time is determined) is similar to modelling age-prevalence data, only with the exposure time substituted for age (e.g. Ramsey 2007). A rich family of models exists for analysing this type of data centred on Cox's proportional hazards model (Cox 1972) and variants thereof. Note, however, that Cox's model is primarily concerned with estimating the proportional effects of different factors on the force of infection, rather than the baseline force of infection, which is sometimes the variable of intrinsic interest.

3.2.2 Estimating β

Since β is dependent on the underlying transmission function, for it to be estimated requires that the relevant variables (e.g. densities of the different infective classes) and/or parameters are also known or estimable.

Estimating β Directly from Trajectory of Prevalence or Cases There are several approaches to estimating transmission coefficients from such data, which typically includes additional data on temporal changes in the population size. Often enough simplifying assumptions can be made to enable the model to be fitted as a generalised linear model and coefficients estimated directly, with the response variables being either the prevalence of infection (e.g. Caley and Ramsey 2001) or the density of infectious classes (e.g. Begon et al. 1999). If the model cannot be solved analytically, then typically the series of differential equations that describe the host/pathogen dynamics will be solved numerically to yield the fitted number and/or density of animals in the relevant disease classes at the times of observation. If prevalence is chosen as the

response variable then the model may be fitted by minimising the binomial negative log likelihood where estimable parameters enter into the likelihood through the fitted prevalence (e.g. Arthur et al. 2005). Likewise, Miller et al. (2006) modelled temporal changes in the number (or cumulative number) of cases of chronic wasting disease in elk (*Cervus elaphus*) as a means of estimating disease transmission rates.

Estimating β from the Force of Infection If the underlying transmission function is known (or assumed), then estimates of λ in conjunction with other variables enables estimation of β. For example, under density-dependent transmission for a single-species model, and assuming the area occupied by the study populations is constant over time (Begon et al. 2002), the rate of conversion from susceptibles to infecteds (di/dt) must equate with the term βsi, where i is the density of infectious animals. That is, $\lambda s = \beta si$, hence $\lambda = \beta i$ where β has the units "potentially infectious contacts per infectious individual per unit area per unit time". Under frequency dependent transmission, infecteds are created at rate $\beta si/n$ where n is the density of all individuals. The rationale is that there are βi potentially infectious contacts per unit area of which a proportion s/n will be with a susceptible individual and hence lead to transmission. Caley and Ramsey (2001) apply both transmission models to leptospirosis infection of brushtail possum (*Trichosurus vulpecula*) populations.

Where a host species may be infected from several sources, the observed force of infection is the summation of the contribution of the different sources of infection. In the case where both intra- and inter-species transmission is density-dependent, the force of infection experienced by the jth species is the sum of the products of the inter-specific transmission coefficients and their densities

$$\lambda_j = \sum_{k=1}^{n} \beta_{jk} i_k \tag{3.4}$$

In Eq. (3.4), n is the number of species, i_k is the density of infectious individuals of species k, and β_{jk} is the transmission coefficient from species k to species j (this follows the notation order of Dobson and Foufopoulos (2001)). Where there are independent estimates of λ_j and i_k, then estimates of β_{jk} can be obtained by regression. An application of this model to a two host (possum, ferret) one pathogen (*M. bovis*) system is given by Caley and Hone (2005). Clearly one could have a mix of frequency-dependent & density-dependent transmission processes occurring in a multi-host system.

3.2.3 Estimating R_0

Estimating R_0 from λ or β Anderson and May (1991) provide a number of steady-state solutions for the basic disease reproductive number. Under Type I mortality (death rate consistently low until the older age classes) and assuming a constant force of infection, they derive the following expression:

$$R_0 = \frac{\lambda L}{1 - e^{-\lambda L}}, \tag{3.5}$$

where L is the life expectancy (clearly disease-induced mortality is assumed to be negligible). However, under Type II mortality, where life expectancy declines exponentially with increasing age, they obtain (again under steady-state assumption and with negligible disease-induced mortality):

$$R_0 = 1 + \lambda L. \tag{3.6}$$

As under these conditions λ is simply the reciprocal of the mean age of first infection (A), Eq. (3.6) can be rewritten in terms of L and A:

$$R_0 = 1 + \frac{L}{A}. \tag{3.7}$$

Anderson and May (1991) also provide a general argument relating R_0 for a microparasite in a homogeneously mixed host population to the overall fraction who are susceptible at equilibrium (x^*) (Eq. (3.8)). The parameter p is the proportion of hosts that are infectious. Note that $x^* = S^*/N^*$, where S^* and N^* are equilibrium densities of the susceptible and total population respectively.

$$\begin{aligned} R_0 &= \frac{1}{x^*} \\ &= \frac{1}{1-p} \end{aligned} \tag{3.8}$$

Applications of these estimators for estimating R_0 in wildlife are hard to find, and note that they are inappropriate for making inference on host status as they are greater than or equal to one for all non-negative values of λ, L, A, or p. This is because these estimators for R_0 assume the system is in a steady-state with a non-zero prevalence – clearly the disease must be persisting, and hence R_0 must be unity or greater.

Assuming that the rate of conversion from the susceptible to the infected class is described by density-dependent transmission, βsi, with horizontal transmission only, a more general estimate of the basic reproduction number of the disease is given by Anderson et al. (1981):

$$R_0 = \frac{\beta S}{\delta + b + \gamma}, \tag{3.9}$$

where β is the transmission coefficient, b is the natural mortality rate, S equals the number of susceptible animals (that can be replaced by the density s), γ is the rate of disease recovery, and δ is the rate of disease-induced mortality. The latent period is assumed equal to zero. Host population dynamics assume exponential population growth, with the exponential rate of increase $r = a–b$, where a and b are the instantaneous *per capita* birth and death rates respectively. Many studies have estimated R_0 using Eq. (3.9) or variants of it. If host population growth follows the simple logistic model, the solution for R_0 is essentially the same, although S may be

replaced by K (population carrying capacity), and a replaces b, and a disease latency period ($1/\sigma$) incorporated if required (e.g. Anderson et al. 1981; Pech and Hone 1988). Anderson and Trewhella (1985) used Eq. (3.9) to estimate the R_0 of *Mycobacterium bovis* infection in badgers (*Meles meles*) assuming generalised logistic growth. Equation (3.9) can be interpreted as one infected animal, equivalent to population density $I = 1/H$ (where H is the home-range area), making $\beta s/H$ infectious contacts per unit area per unit time for its infectious life expectancy $1/(\delta + b + \gamma)$, over an area H. This term for life expectancy (whilst diseased) assumes δ, b and γ are additive.

For the frequency-dependent approximation of the transmission process, the initial maintenance of disease is independent of the population size because the density of susceptibles is assumed to be equivalent to the population density, and occurs (May and Anderson 1979) when $\beta' > (\delta + b + \gamma)$. It follows that the basic reproduction number may be calculated (Roberts and Heesterbeek 1993; Heesterbeek and Roberts 1995) as:

$$R_0 = \frac{\beta'}{\delta + b + \gamma}. \tag{3.10}$$

A heuristic explanation of Eq. (3.10) is an infective individual meeting β' susceptible individuals per unit of time, and it does this for a period of $1/(\delta + b + \gamma)$ (Heesterbeek and Roberts 1995). Assuming local population density does not vary (and hence affect the contact rate), this expression for R_0 is considered to be independent of population size (De Jong et al. 1995). This is also the case if local population density does vary; however, individuals have a fixed number of infectious contacts per unit time (as may be the case for sexually transmitted diseases).

Estimating R_0 from Case Notifications If T_G, the mean serial interval between infections or the generation length is known and the rate of increase (r) of cases in the epizootic can be estimated, then the effective reproduction number during the course of the epizootic may be estimated as

$$R(t) \approx e^{r(t)T_G}, \tag{3.11}$$

providing there are not substantial heterogeneities in transmission. An estimate of R_0 can be obtained during the early phase of epidemic growth when depletion of susceptibles is insignificant. It is commonly assumed that T_G is simply the reciprocal of the recovery rate added to the latent period (defined as infected but not infectious). This assumes that infectivity is constant throughout the infectious period whose length is distributed exponentially – unlikely in practice but difficult to measure. A more realistic pattern, particularly of directly transmitted infectious diseases of animals, is for infectivity to peak early during the infectious period. Unfortunately, estimates of transmission rates and hence R_0 are highly sensitive to the assumed shape of the infectivity function and the associated serial interval – overestimating the serial interval overestimates R_0 and *vice versa*. If the form of the infectiousness function $\beta(x)$ is known then R_0 may be obtained by solving Lotka's equation (here modified to include the proportion of the population that is susceptible (s)

$$s\int_0^{\infty} e^{-ru}\beta(u)du = 1/R_0 \tag{3.12}$$

If the form of the infectiousness function is known or can be reasonably assumed, there have been recent advances in using case notifications to estimate the effective reproduction number of the course of a completely observed (Wallinga and Teunis 2004) or truncated (Cauchemez et al. 2006) epizootic. The method of Wallinga and Teunis (2004) is reasonably robust to substantial under-reporting of cases (see Caley et al. 2008), which will inevitably be the case except in captive populations of wildlife (e.g. see Miller et al. 2006). Where there is a prolonged though variable delay between infection and case diagnosis, methods of back-calculation may be used to reconstruct the infection curve and thus estimate disease transmission rates (e.g. Isham 1989).

Estimating R_0 from Epizootic Attack Rates If an epizootic occurs over a period of time short enough for births and deaths to be considered negligible and the population is reasonably well mixed, the final size equation (Diekmann and Heesterbeek 2000) describes the relationship between the attack rate (α – the overall proportion of the population infected), the initial proportion of the population that is susceptible (s_0) (not to be confused with the initial density of susceptibles $s(0)$), and R_0:

$$\alpha = s_0(1 - e^{-\alpha R_0}) \tag{3.13}$$

For known values of R_0 and s_0, an estimate of α is obtained numerically – predicting α may be of interest where a pathogen is being deliberately introduced into a population (e.g. bio-control). Alternatively, when estimation of transmission rates are of interest, rearranging Eq. (3.13) gives an expression for R_0:

$$R_0 = -\frac{\ln\left(1 - \alpha/s_0\right)}{\alpha} \tag{3.14}$$

where α and s_0 are estimated with error (as will often be the case), the variance of the estimate can be approximated using the delta method. The final size equation for α has been shown to be robust to quite a range of spatial contact structures including spatially isolated patches of susceptibles (Ma and Earn 2006), although it breaks down if inter-patch coupling (movement of infectious individuals) is insufficient. A stochastic equivalent of the final size equation (Becker 1989; Yip 1989) has been applied to wildlife disease modelling (Hone et al. 1992), and has the added attraction of enabling straightforward estimation of the variance:

$$\hat{R}_0 = X\left[\frac{1}{X} + \frac{1}{X-1} + \ldots + \frac{1}{X-Z+1}\right] \Big/ Z \tag{3.15}$$

$$\operatorname{var}\left(\hat{R}_0\right) = \left[X^2\left(\frac{1}{X^2} + \frac{1}{(X-1)^2} + \ldots + \frac{1}{(X-Z+1)^2}\right) + \hat{R}_0{}^2 Z\right] \Big/ Z^2 \tag{3.16}$$

where X is the initial number of susceptibles and Z is the final number of individuals infected. The attraction of both approaches is their simplicity (see Box 3.3). The attack rate may be measured by the observed proportion of animals that die (in diseases with high case fatality rates or if the case fatality rate is known) or the proportion with serological or clinical (e.g. scars) evidence of infection at the completion of the epizootic. Estimating the proportion of animals that die is difficult in many situations as animals are cryptic at the best of times and carcasses are often difficult to locate. Where the attack rate is very high (near one), as was the case for some populations of European harbour seals (*Phoca vitulina*) during the phocine distemper virus epizootic in 1988 (Swinton et al. 1998), the precision of the estimated R_0 is poor.

Box 3.3 Classical swine fever (CSF) in wild boar – comparing estimators

The data set used (Inayatullah 1973) documents the number of wild boar (*Sus scrofa*) found dead on each day following the reported release of a single wild boar inoculated with CSF into a population inhabiting a forest plantation in Pakistan (Table 3.4). Prior to release, the number of wild boar in the population was estimated by a drive count at 465. In the days following the release, a total of 77 wild boar were found dead (Table 3.4). However, approximately 6 months later the population was estimated at 87. There is uncertainty as to whether as many as 379 (= 465 + 1 − 87) boar died from CSF during the epizootic, or whether the wild boar unaccounted for had simply moved out of the area (quite possible considering the forest plantation was only 44.6 km^2). Suitable methods for estimating R_0 using the data include the final size equation (Eq. (3.14)), the method of Wallinga and Teunis (2004) (assuming the time from infection to death has little variance) and trajectory matching.

Using the stochastic version of the final size equation, and assuming a case fatality rate of 90% (i.e. 9 in 10 wild boar that became infected died), then R_0 is estimated to be 1.1 ± 0.2 assuming 77 wild boar died, and 2.7 ± 0.2 assuming 379 boar died. In contrast, if we assume that the inoculated boar died 20 days following inoculation and the CSF infectiousness function is uniformly distributed between 5 and 20 days following infection (after Hone et al. 1992), then applying the method of Wallinga and Teunis (2004) estimates the effective reproduction number (R) of the early cases (what appears to be the1st generation) of the epizootic to be about 4 (Fig. 3.3). We would expect that at this stage the depletion of the susceptible population of wild boars would be minimal, and hence this estimate of the effective reproduction number would be close to R_0. The serial interval is uncertain, and if shortened would lead to a lower estimate of R_0, however, this would be inconsistent with the observed temporal distribution of cases.

All methods have strengths and weaknesses. The method of Wallinga and Teunis (2004) is independent of the epizootic attack rate and robust to consistent

Table 3.4 Observed deaths of wild boar in the days following the introduction of a single boar inoculated with classical swine fever. Adapted from Hone et al. (1992) (permission granted)

Days	Deaths
31	6
32	3
33	1
43	5
44	6
45	2
46	2
47	7
48	7
49	1
51	13
52	2
53	4
54	2
58	5
61	3
62	2
63	2
69	4

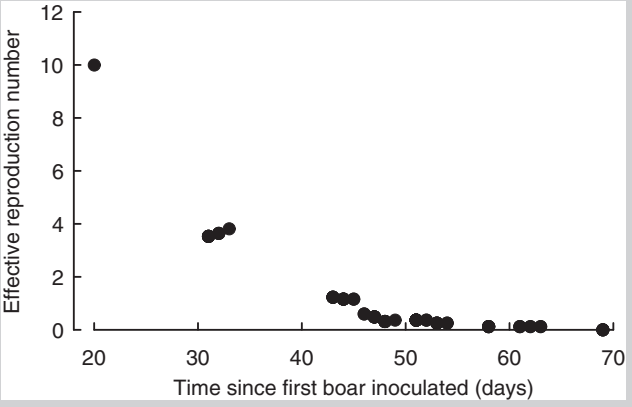

Fig. 3.3 Relationship between the estimated effective disease reproduction number and the day of carcass discovery for wild boars inoculated with classical swine fever

under-reporting of cases, but is heavily influenced by the assumed serial interval. Methods based on the attack rate are independent of nearly all parameters, but are strongly influenced by the assumption of the population being well mixed and the estimated attack rate being accurate. In this case the attack rate was based on the expected high case-fatality rate; doubts exist however,

(continued)

Box 3.3 (continued)

whether CSF causes uniformly high mortality in wild boar populations (Kramer-Schadt et al. 2007). Underestimating the attack rate will underestimate R and *vice versa*.

In summary, there is considerable uncertainty in our estimates of the R_0 of classical swine fever in wild boar – the method of Wallinga and Teunis (2004) strongly suggests a value of about 4, whereas methods based on the epizootic attack rate suggest an upper limit somewhere around 2. Can we reconcile the estimates? Yes, if we recognise that the wild boar population is structured into family groups with limited mixing between them, then it is quite possible for R_0 to be about 4, yet at a broader population level observe an attack rate consistent with a much lower value. It may well be that our assumption of homogeneous mixing is playing havoc with our estimation, reflecting uncertainty in how the system under study operates.

3.3 Dealing with Uncertainty

In a perfect world, different methods of estimating transmission rates should produce the same results. In practice this is rarely the case, and worse still, estimates using one method often lie outside reasonably wide confidence intervals estimated using another. Why? Because nearly all estimators are conditional on an assumed underlying model of how the host/pathogen system operates – and this is often subject to considerable uncertainty (we take a "leap of faith" across this lack of knowledge (McCallum 1995) by making assumptions). The estimation of the dynamics and rate of transmission of classical swine fever (CSF) in wild pigs (*Sus scrofa*) is a good example (see Box 3.3). As we try and fit more realistic disease transmission models containing a greater number of parameters, it will become imperative to incorporate as much prior information as possible to ease the burden on the likelihood functions. Hence Bayesian style model fitting that incorporate stochasticity will become the more common. Indeed, the use of stochastic models opens up alternative statistically rigorous options for parameter estimation and inference of unobserved features of the epidemic. If every event type represented in a stochastic model (e.g. infection, recovery etc.) were to be observed in a real epidemic, then it would be possible to construct a complete likelihood (based on this complete set of observations) from the definition of the model, and from which parameter values could be estimated as described in the examples above. However, in reality we typically have access to rather limited data; for example describing the prevalence or reported cases over time, and therefore we must infer not only the parameter values but also the missing infection (and other) events. Fortunately, it is possible to frame this problem in a Bayesian framework in such a way that the so-called posterior distribution of model parameters and missing events is known up to a normalising constant. Modern

stochastic integration techniques such as Markov Chain Monte Carlo (MCMC) can then be employed to generate true samples from the posterior for increasingly complex models. The Bayesian framework also allows (and requires to some extent) prior information about the value, or possible range, of parameter values obtained from the literature or particular empirical studies, to be taken into account. The samples generated from the posterior allow the calculation of essentially any statistic of interest based on the parameters, and/or the missing events. For example, Streftaris and Gibson (2004) employed such methods to fit non-Markovian stochastic models for the transmission dynamics of a particular strain of foot-and-mouth disease (FMD) virus to data from a controlled experiment. In addition to transmission rates they inferred the hidden transmission history of the epidemic. Cook et al. (2007) used such techniques to estimate multiple transmissions rates within and between crop species in a spatial context, using information theoretic measures of deviance to show that the best–fitting model requires a fully parameterized transmission rate matrix; that is different transmission rates from species A to B and *vice versa*.

3.4 Assessing Host Status

Once the known host range of a disease has been established or extended there is a need to assess the role of these new hosts in the wider epidemiology of the disease. Assessing the host status in the epidemiology of a disease is crucial to its control (Caley and Hone 2005). Top of the agenda is determining whether the disease persists within a host population since all self-sustaining/persistent sources of infection (e.g. reservoirs) should be considered as part of a disease control strategy. Identification and quantification of transmission routes is often central to characterising the persistence of infection in wildlife populations. For example, the known host range of *M. avium* subspecies *paratuberculosis* has recently been extended to include a number of non-ruminant wildlife species (Daniels et al. 2003b). Of these new host species the European rabbit (*Oryctolagus cuniculus*) was identified as posing the greatest risk to sympatric livestock as rabbits are often abundant on livestock farms, they excrete high numbers of bacteria in their faeces and grazing livestock show no avoidance of rabbit faeces resulting in high exposure rates (Judge et al. 2005a). Given that paratuberculosis is a widespread and difficult disease to control in livestock populations and also has possible links to Crohn's disease in humans, the identification of a persistent wildlife source of infection would greatly impact on our understanding of current livestock control strategies. Judge et al. (2007) used a combination of field studies to quantify the rates of rabbit-to-rabbit transmission of paratuberculosis and mathematical modelling to show that infection can persist in rabbit populations for extended periods (see Box 3.4). This finding may go some way to explaining the persistent nature of the disease in livestock populations, and rabbits are now included in farmer led disease control strategies in the UK (e.g. The Premium Cattle Health Scheme).

Box 3.4 Persistence of *Mycobacterium avium* subspecies *paratuberculosis (Map)* in rabbit populations

European rabbits (*Oryctolagus cuniculus*) have recently been identified as a key wildlife species in terms of paratuberculosis transmission to the wider host community. Judge et al. (2007) tested the hypothesis that *Map* can persist in rabbit populations for extended periods of time. A spatially-explicit stochastic simulation model of a generic host-disease interaction was developed to quantify the inter-play between vertical and horizontal routes of transmission, needed for the persistence of *Map* in rabbit populations and to test the hypothesis. The model was parameterised based on empirical studies on rabbit population dynamics and on rabbit-to-rabbit routes of *Map* transmission. Predictions from the model suggest that *Map* persists in rabbit populations at all values of the horizontal and vertical transmission parameters in the range estimated from the field data (taken from Judge et al. 2006; see Box 3.1), and in many cases at all values within 95% confidence intervals around this range (Fig. 3.4). The persistence of *Map* infection in rabbit populations in the absence of external sources of infection suggests that they may act as a reservoir of infection for sympatric livestock. These findings, in combination with the ubiquitous distribution of rabbits in the United Kingdom and elsewhere, suggests that if *Map* becomes established in rabbit populations, they are likely

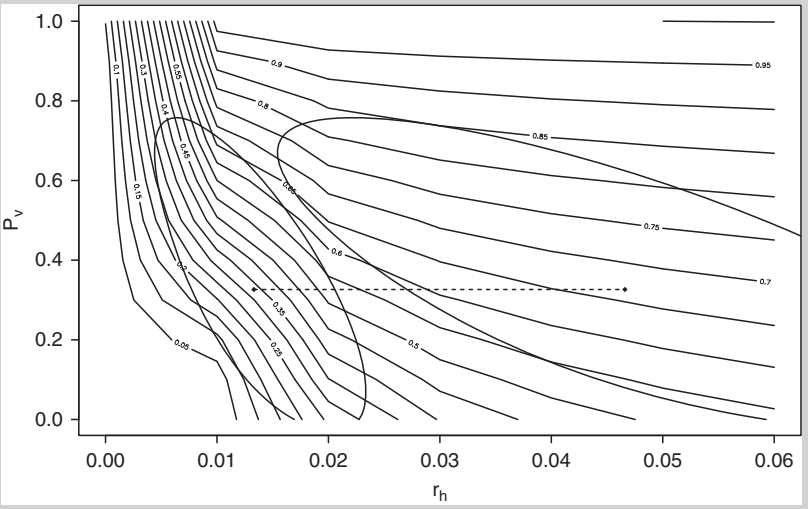

Fig. 3.4 Isopleths of equilibrium prevalence at differing vertical (P_v) and horizontal (r_h) transmission rates for a rabbit population approaching carrying capacity (dotted horizontal line is the estimated range of r_h values along the vertical transmission rate from field data). Ellipses represent 95% confidence intervals around the lower, and upper β estimates (from Judge et al. 2007)

to provide widespread and persistent environmental sources of infection, and thus a disease risk to livestock and potentially humans. Judge et al. (2007) conclude that where local rabbit populations are infected they should be included in any future paratuberculosis control strategies.

3.5 Implications for Management

Being able to quantify disease transmission and identify contributing factors is critical to not only evaluating management, but also designing management actions in the first place. Estimates of disease transmission coefficients are critically dependent on the assumed underlying model of transmission, and it is here that the greatest uncertainty is introduced. Where the mechanisms of transmission cannot be observed, or reasonably inferred by alternative means (e.g. disease pathology), data-based inference on the underlying mechanisms of transmission will need to be employed. This could take the form of critical experiments to identify routes of transmission. For example, Ramsey (2007) clearly demonstrated the importance of sexual transmission of leptospirosis (caused by *Leptospira interrogans*) in brushtail possums in a longitudinal experiment entailing castrating male possums to stop their sexual contacts. Likewise, Palmer et al. (2004) demonstrated the ability of *M. bovis* to be transmitted between white-tailed deer (*Odocoileus virginianus*) via contaminated feed. In doing so they overthrew the respiratory only paradigm of tuberculosis transmission in true reservoir hosts (Caley 2006).

Critical experiments needed to quantify the routes of transmission of wildlife diseases are typically difficult to undertake once let alone adequately replicate. Where critical experiments have not been undertaken, or are difficult to do, model selection techniques as applied to observational "experiments" may be the only way of (1) making inference on the underlying mechanism of transmission, and (2) estimating transmission parameters given a chosen model of transmission. Caley and Hone (2002) demonstrated how information-theoretic model selection techniques may be used to make inference on transmission routes by identifying how age-specific prevalence will vary as a function of age under different hypotheses. Miller et al. (2006) similarly used model selection techniques to demonstrate that a model that included indirect transmission of chronic wasting disease (CWD) amongst mule deer (*Odocoileus hemionus*) was the most supported model of transmission. These and other similar studies have greatly increased our understanding of the transmission of wildlife disease.