University of Salford MANCHESTER

Canine echinococcosis in Kyrgyzstan: detection, diagnosis, and dynamics.

"Multa novit canis, verum echinus unum magnum"

Alexander James Mastin

Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy, 5th May 2015

University of Salford,
School of Environment and Life Sciences

Supported by wellcome trust

Contents

Figures

Tables

Acknowledgements and declaration

"No man is an island, entire of itself. Every man is a piece of the continent, a part of the main."

John Donne (1572 – 1631)

Acknowledgements

[This thesis is nearly 100,000 words long already, so I'll try to keep this short!]

Professor Phil Craig – for being an excellent supervisor, field biologist and general all-round inspiration. You gave me the freedom to take this thesis in the direction I wanted to, which has made it what it is. It has truly been an honour to work with you (and to have been your last PhD student before your retirement).

Dr. Freya van Kesteren – what can I say? The best colleague I have ever had, or will ever have. We complemented each other perfectly, and without you there is literally no way I could have finished the field work (and that's not to mention the fact that you are responsible for all of the sample testing upon which this whole thesis is based). I can't thank you enough for your friendship and support. Give me a call once you've got over this behavioural ecology phase and are ready to return to echino!

My wife (and best friend), **Kim Casement-Mastin** – for offering infinite support when things were challenging, for giving me the freedom to work when I needed to, for pulling me away from work when I didn't need to, and for keeping things in perspective throughout (whilst also expertly managing two small children!).

My two sons, **Rowan and Oran Mastin** – one of whom arrived six weeks before I started working at Salford, and one of whom got me out of a fieldwork trip 19 months later – for bringing joy (amongst a range of other emotions!) into my life.

All my friends at Salford University, especially the original crew of **Patrick Killoran**, **Gemma Lace-Costigan**, **Manisha Patel**, and **Chris Harrison**. You made my time at Salford an absolute joy – I can't thank you all enough.

Our team of Professors: **Paul Torgerson**, **Patrick Giraudoux**, **Mark Danson**, and **Mike Rogan**. Thank you all for giving me inspiration and support. I'm so grateful I managed to avoid vomiting on you all that night in Sary-Mogol!

Professor Dirk Pfeiffer – for starting me off on this course, and for helping craft my philosophy of science. I never forgot I was an epidemiologist, even when I was living amongst the parasitologists!

Our other collaborators, inside and outside Kyrgyzstan.
Iskender Ziadinov for making sure fieldwork ran smoothly (and for accepting frantic phonecalls from me in the middle of the night from Osh airport!).
Akjol Tagaibekov and **Turdumamat Sultanov** (and your fantastic families), for incredible hospitality in the field. I hope to see you all again someday soon.
Bermet Mytynova, **Dastanbek Sydykov**, and **Almaz**, for helping out with fieldwork, translating, and for being great field companions.

The rest of the Cestode Zoonoses Research Group at Salford (past and present):
Dr. Belgees Boufana, **Helen Bradshaw**, **Dr. Tony Bodell**, **Judy Mwangi**, and **Rebecca Rushworth.** It's been an absolute pleasure working with you all.

The Wellcome Trust – for funding this project and for therefore giving me the opportunity to visit the beautiful country of Kyrgyzstan, as well as to attend a range of wonderful conferences and workshops.

And last but not least, my parents, **Keith and Brenda Mastin**, for always being there for me, for supporting me through vet school (and beyond), and for encouraging me to always ask questions.

Declaration

This thesis is the product of approximately three and a half years' worth of work, commencing in November 2011, and was generously funded by the Wellcome Trust, (grant number #094325/10/Z/10). Whilst the author (with a little help from Google) was responsible for the direction the thesis took and the methods applied, the project itself was a collaborative one.

Freya van Kesteren assisted with collection of most of the faecal samples used in the study, along with Professors Phil Craig and Mike Rogan and Iskender Ziadinov in the early stages of the project. Along with Iskender, Bermet Mytynova and Dastan Sydykov were responsible for administering questionnaires over the course of the project.

Freya van Kesteren was also responsible for collection of the faecal samples from Xinjiang, China, used in chapter 3. Professor Phil Craig conducted many of the dog necropsies from which the panel used in this chapter and in the subsequent chapters (especially chapter 4) were derived.

All faecal sample testing was conducted solely by Freya van Kesteren, with laboratory support where needed from Dr Belgees Boufana, who also provided many of the control samples used. Professor Paul Torgerson and Dr Ruby Chang offered some initial statistical advice for chapter 4, and Professor Patrick Giraudoux made some suggestions for the model selection strategy used in chapter 6.

Abstract

The point is not merely to understand the world,
but to change it.

Karl Marx (1818 - 1883)

Human echinococcosis is an increasing public health issue in Kyrgyzstan, where *Echinococcus granulosus* and *Echinococcus multilocularis* are coendemic and domestic dogs are considered the primary source of human infection. A control scheme based upon dosing dogs with praziquantel was commenced in Kyrgyzstan in 2012 and was evaluated using ELISA tests to measure levels of *Echinococcus*-specific 'coproantigens' in canine faeces. The current study describes methods of interpretation of coproELISA test results, both prior to and during a control scheme, using data collected from dogs in southern Kyrgyzstan over a period of three years.

Current methods of coproELISA test interpretation based upon selection of a single cut-off value are described and found to have considerable limitations. To address this, Bayesian mixture modelling was used to transform raw coproantigen data into a metric which approximates the possible worm burden in individual dogs and reduces test misclassification at the population level. This approach was validated using data from a panel of faecal samples of known status and was applied to data from samples of unknown status collected from Kyrgyz dogs. Multiple correspondence analysis was used to characterise the Kyrgyz study sites and identify possible associations with canine infection status (incorporating both coproELISA and coproPCR results), but did not identify any strong relationships. A mixed effects logistic regression modelling approach combined with model averaging was used to identify temporal and seasonal trends in coproantigen and coproPCR prevalence. A trend of decreasing test prevalence over time with pronounced seasonality was found for some test results. Finally, a mathematical model of transmission of both *Echinococcus granulosus* and *Echinococcus multilocularis* in Kyrgyzstan was developed and used to simulate the effects of a number of different dog dosing strategies. Canine echinococcosis surveillance and control could be improved by tailoring methods of diagnostic test interpretation (population-level/individual-level, categorical/continuous) to the situation at hand.

Chapter 1: Introduction and Literature Review

"Study the science of art. Study the art of science.
Develop your senses—especially learn how to see.
Realize that everything connects to everything else."

Leonardo da Vinci (1452 – 1519)

1.1 Introduction

Echinococcosis, resulting from infection with cestodes of the genus *Echinococcus*, is an important disease of people and animals worldwide. It has been classified by the World Health Organisation (WHO) as a 'Neglected Zoonotic Disease' (NZD) (WHO, 2009), within the wider group of 'Neglected Tropical Diseases' (NTDs) (Craig *et al.*, 2007a; WHO, 2010a; b, 2013a; Molyneux, 2012), which have the potential to 'impede development within a country'. Although species classification is currently in a state of flux, two main species groups are generally involved with human infection: *Echinococcus granulosus s*ensu lato (the cause of cystic echinococcosis) and *Echinococcus multilocularis* (the cause of alveolar echinococcosis). *Echinococcus* species have an indirect lifecycle (see figure 1.1), requiring both an intermediate and a definitive host.

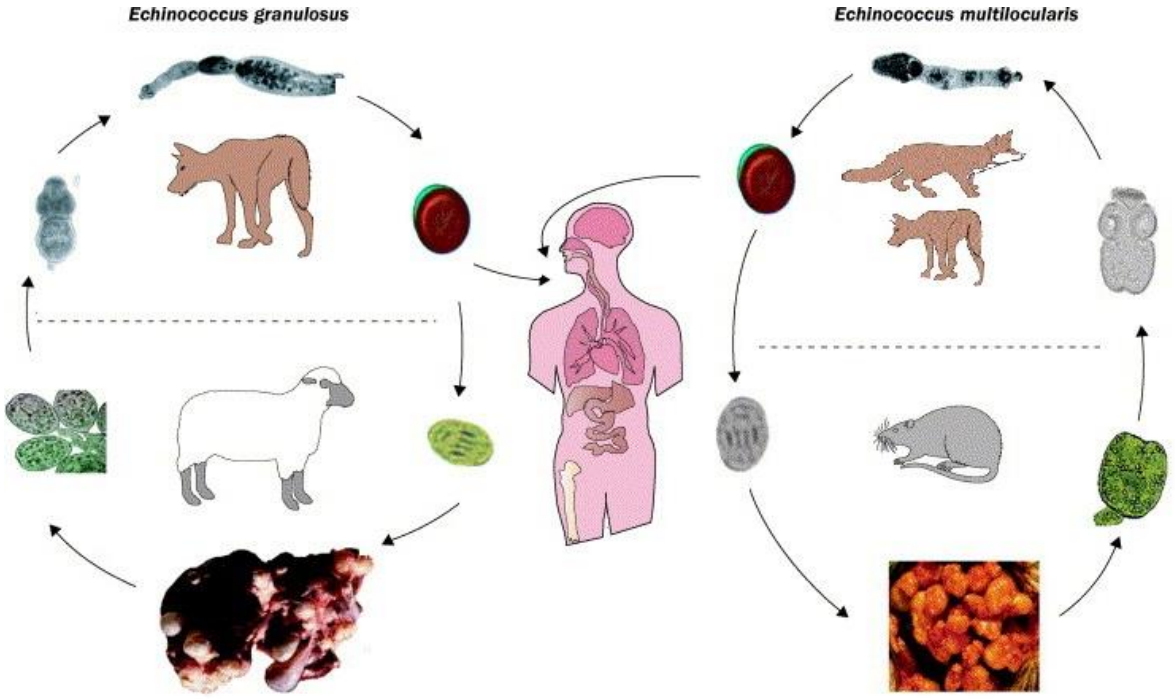


Figure 1.1. Lifecycles of *Echinococcus granulosus* and *Echinococcus multilocularis*. Adapted from McManus *et al.* (2003)

In the case of *E. granulosus* (s.l.) and *E. multilocularis*, the definitive hosts are canid species (primarily domestic dogs in the case of *E. granulosus*, and traditionally foxes in the case of *E. multilocularis*), in which infection is usually asymptomatic. Intermediate hosts for *E. granulosus* are usually ruminant livestock species, in which infection can lead to clinical effects and can lead to offal condemnation and reductions in productivity. Infection in humans also often leads to clinical effects, and the overall economic impact of both human and animal echinococcosis worldwide is considerable (Torgerson *et al.*, 2000, 2001, 2010; Torgerson and Dowling, 2001; Torgerson, 2003a; Budke *et al.*, 2004, 2005c, 2006). Additionally, within developing countries, echinococcosis tends to predominantly affect poorer pastoral communities, and as such may exacerbate poverty (Craig *et al.*, 2007b; Maudlin *et al.*, 2009). The geographical distribution of *E. granulosus* and *E. multilocularis* is shown in figure 1.2 (Eckert and Deplazes, 2004).

Despite its considerable public health and economic burden worldwide. echinococcosis is rarely given priority in disease control schemes nationally or internationally. One possible reason for this is the fact that human infection is commonly found to be heterogeneously clustered within rural communities (Craig *et al.*, 1992), which may be geographically isolated and physically difficult to reach as well as politically marginalised. Additionally, it has been argued that a high disease burden alone should not necessarily indicate a priority candidate for disease control – instead, the availability of cost-effective control strategies should be considered (Canning, 2006). Whilst echinococcosis is considered to be a 'preventable disease' by the World Health Organisation due to the availability of effective treatments for canine infection and practical measures which can reduce human infection risk (WHO, 2013b), instigating an echinococcosis control scheme is not a trivial issue. Due to the persistence of the parasite in intermediate hosts, and the difficulties in treating all definitive hosts in a community, control of echinococcosis can take long periods of time even in optimal situations (Craig and Larrieu, 2006), and in many cases may need to run indefinitely.

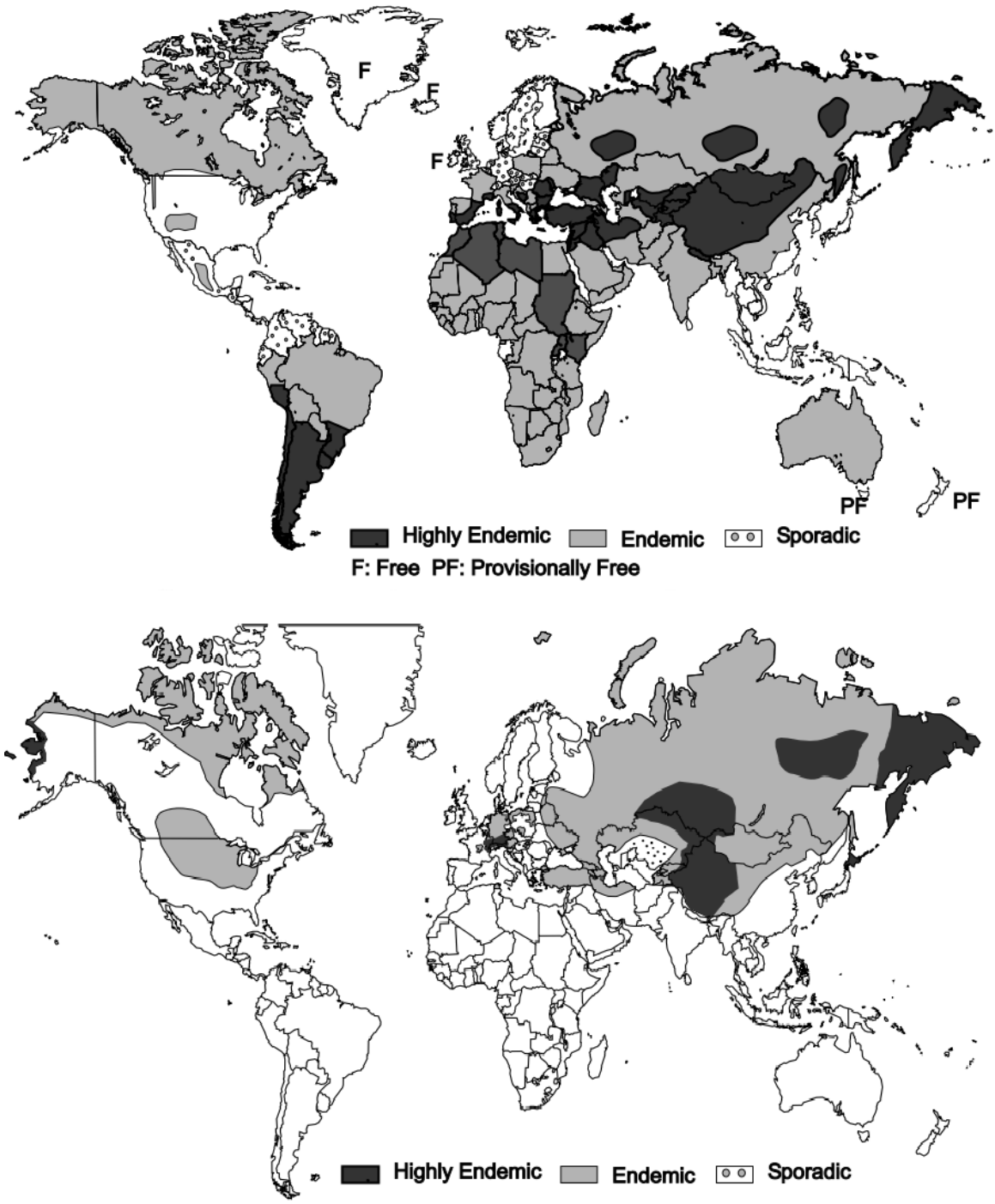


Figure 1.2. Geographical distribution of *E. granulosus* (top) and *E. multilocularis* (bottom). Taken from WHO/OIE (2001a); Torgerson and Budke (2003); and Eckert and Deplazes (2004)

An important component of disease control is the ability to detect disease. As described above, echinococcosis in definitive hosts is asymptomatic, and therefore

diagnostic tests are required to detect infection in order to assess the potential risk of human infection (through eggs released by adult worms in the definitive host) or the effect of a control scheme. The faecal coproantigen test (Deplazes *et al.*, 1992; Allan *et al.*, 1992) offers an effective method of surveillance of definitive host infection during control campaigns, but relatively little attention has been given in recent years to the optimal approach for surveillance. This is a central concept explored throughout the current thesis, and will be described in more detail below and in later chapters.

1.2 *Echinococcus* biology

1.2.1 Basic lifecycle

The lifecycle and epidemiology of *Echinococcus* spp has been well described elsewhere (Smyth, 1964; Thompson, 1995; WHO/OIE, 2001b), and as such will be only briefly covered here. As for all taenid tapeworms, *Echinococcus* spp have an indirect lifecycle, requiring mammalian intermediate and definitive hosts (although some species can act as both intermediate and definitive hosts), with transmission through predator-prey interactions. The intermediate hosts for the different species differ, but small ruminants and small rodents are the intermediate hosts most commonly responsible for definitive host infection with *E. granulosus* (s.l.) and *E. multilocularis*, respectively. Domestic dogs are the main definitive host for *E. granulosus* (s.l.) (which includes a number of recently classified different species of *Echinococcus*), and although foxes are the traditional definitive host of *E. multilocularis*, domestic dogs are known to be susceptible to infection (Kapel *et al.*, 2006; Matsumoto and Yagi, 2008), and are thought to play a major role in human infection with this parasite (Craig *et al.*, 2000; Li *et al.*, 2005; Wang *et al.*, 2006a; Craig and The Echinococcosis Working Group in China, 2006; Yang *et al.*, 2006).

Eggs released by definitive hosts contain an onchosphere within a protective layer known as the embryophore, which offers protection from environmental conditions (Gemmell and Lawson, 1986). Following ingestion (Dévé, 1949) or inhalation (Borrie *et al.*, 1965) by an intermediate host, the onchosphere is released from the embryophore

and is activated. In the case of enteral activation (the most common route), the onchosphere is able to penetrate the small intestine and enter the bloodstream or lymphatic system, through which it can be carried to an eventual site of encystment such as the liver or lungs (Heath, 1971; Harris *et al.*, 1989). Here, development proceeds to the metacestode stage, which comprises singular or multiple cystic structures formed of parasite tissue (the germinal and laminated layers). In the case of *Echinococcus granulosus*, the cysts are surrounded by a fibrotic response from the host, known as the adventitial layer (Cameron and Webster, 1969), although this is not present in the case of *Echinococcus multilocularis* (Sakamoto and Sugimura, 1970). Through proliferation of the germinal layer (endogenously in the case of *E. granulosus*, and both endogenously and exogenously in the case of *E. multilocularis*), 'brood capsules' are formed. Proglottids develop within these through further proliferation and budding. Although proliferation of *E. multilocularis* metacestodes in normal definitive hosts is curtailed, this does not occur in the case of human infection, where proliferation tends to continue indefinitely (Rausch and Wilson, 1973).

Following ingestion of a fertile cyst by a definitive host, evagination of the protoscolex suckers, rostellum and hooks takes place in the upper duodenum, and the protoscolices move into the crypts of Lieberkühn, where they attach to the mucosa through the use of suckers and hooks. Development to the adult worm follows over a period of weeks, with each protoscolex being potentially able to form an individual adult worm. *E. granulosus* worms tend to be located in the proximal small intestine (Gemmell *et al.*, 1986c; Lymbery *et al.*, 1989), whereas *E. multilocularis* are generally found in the distal small intestine (Thompson and Eckert, 1983; Morishima *et al.*, 1999a; Umhang *et al.*, 2011). The adult worms are hermaphroditic and are able to self-fertilise, although in some situations sexual reproduction may also take place. Eggs are fertilised and develop in the terminal proglottid, and comprise an embryo (known as the onchosphere) surrounded by several layers. Eggs are released in the faeces of the definitive host, either free or within the released terminal proglottids (likely in different stages of maturation and therefore varying infectivity), and the cycle continues as described above (Thompson, 1995).

1.2.2 Species and strains of *Echinococcus*

Wide variations in the phenotypic characteristics and behaviour of different members of the *Echinococcus* genus are recognised, suspected to be due largely to genetic variation, but which may also be induced during the lifecycle of the parasite (Thompson, 1991, 1995). As such, appropriate classification of different species and strains within the genus has been complex, and there is ongoing debate regarding the most appropriate method of taxonomic classification. It has been known for some time that the original classification system used for *E. granulosus* was inappropriate (Bowles *et al.*, 1992, 1995), and current evidence suggests that there are at least six separate species of *Echinococcus*: *E. granulosus*, *E. equinus*, *E. ortleppi*, *E. multilocularis*, *E. vogeli* and *E. oligarthus* / *E. oligartha* (Kumaratilake and Thompson, 1982; Thompson and McManus, 2002; Thompson, 2008; Nakao *et al.*, 2013). Other distinct species have been suggested, including one affecting camels and/or pigs as intermediate hosts (currently classified as two 'strains' of *E. granulosus*: G6 and G7) (Thompson *et al.*, 1995; Nakao *et al.*, 2007), and one affecting pika (*E. shiquicus*, which appears to be related to *E. multilocularis*) (Xiao *et al.*, 2006b). It has also been proposed that G6, G7, G8 and G10 be combined into a single species known as *E. canadensis* (Moks *et al.*, 2008), or that (based on nuclear DNA) G8 and G10 should together be termed *E. canadensis*, and G6, G7 and G9 termed *E. intermedius* (Saarma *et al.*, 2009). The former classification is more generally accepted, and will be used in the current study.

Genetic variation within *E. multilocularis* has until relatively recently been overlooked largely due to the relatively low variation compared to that within *E. granulosus* (s.l.) (which as described above, actually represents a number of different species) (Bowles *et al.*, 1995; Nakao *et al.*, 2009). Investigation of *E. multilocularis* has identified a number of different isolates found in different locations – suggesting development in geographic isolation from each other. The main three isolates identified have been named as the 'European', 'Asian' and 'North American' clades, with their names describing their geographical location (central-eastern Europe; central-eastern Asia and Alaska; and Alaska and central North America). The Asian clade is likely to have

given rise to the European and North American forms, with both Asian and North American clades being found in Alaska, although a separate, distinct clade has also been identified in Inner Mongolia (Nakao *et al.*, 2009).

Different species and strains of *Echinococcus* vary in their predilection for definitive and intermediate hosts (including humans), as well as in their distribution. A recent paper (Nakao *et al.*, 2013) summarised some these differences, as shown in table 1.1. Due to the complexities of classification and the limited relevance of some of these species/strains to the current investigation, only *E. granulosus* and *E. multilocularis* (and their accompanying strains/associated species) will be considered further here.

Table 1.1. Strains and species of *Echinococcus*, based upon Nakao *et al.* (2013)

Species/strain	Distribution	Main intermediate host(s)	Main definitive host(s)	Human infection
Echinococcus granulosus (G1,G2,G3)	Worldwide	Sheep, goat, cattle	Dog	Most common
Echinococcus equinus (G4)	Worldwide	Horse	Dog	Unknown
Echinococcus ortleppi (G5)	Worldwide	Cattle	Dog	Uncommon
Echinococcus canadensis (G6,G7)	Worldwide	Pig, camel, cattle, goat, sheep	Dog	Common
Echinococcus canadensis (G8)	Northern arctic/boreal	Moose, wapiti	Wolf	Uncommon
Echinococcus granulosus/intermedius (G9)	Poland	Pig	Dog	Uncommon
Echinococcus canadensis (G10)	Northern arctic/boreal	Moose, reindeer, wapiti	Wolf, dog	Uncommon
Echinococcus multilocularis	Holarctic	Arvicoline rodents	Red fox, arctic fox, dog	Common
Echinococcus shiquicus	Tibetan plateau	Pika	Tibetan fox	Unknown
Echinococcus oligarthus/oligarthra	Neotropical	Agouti	Wild felids	Uncommon
Echinococcus vogeli	Neotropical	Paca	Bush dog	Uncommon
Echinococcus ortleppi	Worldwide	Cattle	Dog	Uncommon
Echinococcus felidis	Africa	Unknown	Lion	Unknown

1.2.3 Biological parameters

The prepatent period of *E. granulosus* in definitive hosts has been reported as between six and twelve weeks (Gemmell *et al.*, 1986c; Gemmell and Roberts, 1995; WHO/OIE, 2001b), and that for *E. multilocularis* is around four weeks (WHO/OIE, 2001c; Kapel *et al.*, 2006). Mathematical modelling techniques have estimated the lifespan of *E. granulosus* and *E. multilocularis* in the definitive host to be around nine months and around 3-4 months, respectively (Ziadinov *et al.*, 2008). The duration of egg production in foxes and dogs experimentally infected with *E. multilocularis* has been estimated as around four to six weeks (Kapel *et al.*, 2006). In sheep, *E. granulosus* growth is slow, requiring around seven years for half of the infections to reach fertility, whereas fertile cysts of *E. multilocularis* in rodents may be present after only two months (WHO/OIE, 2001c).

1.3 Echinococcosis

1.3.1 Human infection

Infection of humans with *E. granulosus* or *E. multilocularis* can result in the production of tissue cysts – a condition named cystic echinococcosis (CE) in the case of *E. granulosus*, and alveolar echinococcosis (AE) in the case of *E. multilocularis*. Humans, whilst acting as an intermediate host in these cases, are commonly referred to as a 'dead end host', as they very rarely play any further role in parasite transmission following infection. CE is characterised by individual, well-encapsulated cysts, whereas AE presents with numerous proliferating small cysts which are not well encapsulated and are able to metastasise in the blood or lymphatic systems to other sites (Thompson, 1995). Although *E. multilocularis* metacestodes in humans rarely contain protoscolices (Rausch and Wilson, 1973), AE is generally a much more serious condition than CE, and is less responsive to the usual surgical techniques used to treat CE (Torgerson *et al.*, 2008, 2010). The primary cysts of both species most commonly develop in the liver, although *E. granulosus* cysts may also develop in the lung (or, in rare cases, in other organs). Secondary spread of infection from these primary sites can occur following cyst rupture, or through blood or lymph-borne metastasis in the case

of AE. Due to the slow-growing nature of the cysts, many years may pass before clinical disease is observed (generally resulting from the space-occupying effect of the cyst or due to organ pathology), and although AE is generally ultimately fatal if left untreated (Torgerson *et al.*, 2008), this is not the case for CE, which can in some cases remain asymptomatic indefinitely (Schaefer and Khan, 1991) – especially in the case of liver cysts (Larrieu and Frider, 2001).

As mentioned earlier, CE and AE are a considerable disease burden in terms of morbidity, surveillance, treatment costs and mortality in endemic communities (Torgerson *et al.*, 2000, 2001, 2010; Torgerson and Dowling, 2001; Torgerson, 2003a; Budke *et al.*, 2004, 2005c, 2006). Although treatment is available for both CE and AE, this can be costly and challenging, often requiring surgical intervention and/or long periods of chemotherapy with albendazole (Brunetti *et al.*, 2010). Infection of livestock with *E. granulosus* is also a cause of additional economic losses due to reduced productivity and condemnation of animal products (Torgerson, 2003a). These economic issues are exacerbated by the fact that those individuals and communities most affected tend to be poorer, rural communities which are often geographically and/or behaviourally isolated to some degree from healthcare systems (Craig *et al.*, 2008; Maudlin *et al.*, 2009; Molyneux *et al.*, 2011). In particular, nomadic or seminomadic groups with livestock and people in close contact with dogs are more commonly affected by echinococcosis. Dog ownership has frequently been found to be a risk factor for human infection with *E. granulosus* (s.l.) (Campos-Bueno *et al.*, 2000; Larrieu *et al.*, 2002; Yang *et al.*, 2006; Moro *et al.*, 2008), although this association is not invariably found (Carmona *et al.*, 1998; Dowling and Torgerson, 2000; Dowling *et al.*, 2000; Yamamoto *et al.*, 2001; Torgerson *et al.*, 2003a, 2009b). Domestic dog ownership or contact with dogs has also been identified as a risk factor for human infection with *E. multilocularis* in a number of studies (Craig *et al.*, 2000; Li *et al.*, 2005; Wang *et al.*, 2006a; Craig and The Echinococcosis Working Group in China, 2006; Yang *et al.*, 2006).

1.3.2 Detection of echinococcosis

Detection of echinococcosis in intermediate and definitive hosts can be challenging. Due to the longevity of cysts in intermediate hosts, the most accurate reflection of the current infection pressure is obtained by diagnosing infection in definitive hosts (i.e. canids in the case of the species and strains of interest here)(Walters, 1978; Palmer *et al.*, 1996; Craig and Larrieu, 2006), although livestock (the normal intermediate hosts of *E. granulosus*, and accidental hosts of *E. multilocularis*) may be useful as sentinels, for assessing levels of environmental contamination with eggs, and for surveillance in low endemic areas (such as in the later stages of a control scheme). Diagnostic testing for echinococcosis has been reviewed in a recent paper (Torgerson and Deplazes, 2009).

1.3.2.1 Definitive hosts

As mentioned above, for the species of *Echinococcus* of interest to this project, the definitive hosts are canid species. Detailed reviews of diagnostic testing for echinococcosis in these species are available (WHO/OIE, 2001d; Craig *et al.*, 2003; Torgerson and Deplazes, 2009), and so only a brief description will be given here. The 'gold standard' test for infection in these hosts is considered to be necropsy and examination of the small intestine for adult worms using the sedimentation and counting technique (SCT) (WHO/OIE, 2001d; Eckert, 2003). This method involves gross examination of the small intestinal mucosa for adult worms, followed by stripping of mucosa, sedimentation and microscopic examination of the sediment. However, this method is relatively time-consuming, is biohazardous if performed with fresh intestines, and requires culling of dogs (which is both logistically challenging and problematic in communities reliant on their dogs). A similar method which is less time consuming is the intestinal scraping technique (IST), in which 15 deep mucosal scrapings are taken from equal intervals along the intestine, and the squash preparations are examined microscopically. This method has been reported to have a sensitivity of 78% and a specificity of 100% from a sample of 170 foxes (87 of which were found to infected by the SCT) (Hofer *et al.*, 2000).

For much of the 20th century, arecoline purgation was the mainstay of investigation of canine infection in living dogs (with the additional benefit of being a treatment method), despite having a low diagnostic sensitivity (Schantz, 1997; Lahmar *et al.*, 2007b). Enteral administration of preparations of arecoline can both induce paralysis in any cestodes present and encourage the expulsion of intestinal contents. The material evacuated can then be inspected for worms using sieving techniques. However, as for necropsy, this method is biohazardous and time-consuming, requires skilled personnel, carries a potential risk of dog death, and is not always successful. One study in Tunisia estimated that arecoline purgation had a sensitivity of around 65% after one dose and around 78% after two doses for detection of *E. granulosus* (compared to necropsy), with a specificity of 100% (Schantz, 1997). This and another study found that less than 70% of dogs purged after one dose, and less than 90% purged after two doses (Schantz, 1997; Lahmar *et al.*, 2007b). Little is known about the performance of purgation in the detection of *E. multilocularis*, although analysis of data collected from Kyrgyzstan using a variety of diagnostic tests has estimated the sensitivity of purgation to be around 40% for *E. granulosus* and 20% for *E. multilocularis* (with overlap in the 95% credible intervals for these two estimates) (Ziadinov *et al.*, 2008). A method based on latent class analysis of data collected from the Tibetan plateau has given point estimates of purgation sensitivity of between 30% and 55% for *E. granulosus*, and between 55 and 75% for *E. multilocularis* (again, often with overlap in the 95% credible intervals) (Hartnack *et al.*, 2013).

In order to address some of the challenges associated with purgation or necropsy, methods based upon analysis of normally voided faeces have been developed, and offer a potentially useful method of diagnosing infection easily with minimal invasiveness. Traditional parasitological techniques such as examination of faeces for eggs or proglottids are problematic due to irregular excretion of eggs, their small size and the inability to distinguish *Echinococcus* spp eggs from those of *Taenia* spp (which are also likely to be commonly found in endemic communities, but which rarely pose a risk to humans) (Allan and Craig, 2006). As well as having a low sensitivity and specificity, these methods are also quite labour-intensive (Craig *et al.*, 1988).

Therefore, detection of infection from faecal samples is predominantly based upon molecular methods such as the detection of faecal 'coproantigens' and PCR-based methods (Deplazes *et al.*, 2003).

Coproantigens are large carbohydrate-based molecules, thought to be excretory/secretory or turnover material derived from the surface glycocalyx of the adult tapeworm (Elayoubi *et al.*, 2003; Elayoubi and Craig, 2004) (although it appears they are also present to some degree in prepatent infections (Deplazes *et al.*, 1992; Malgor *et al.*, 1997; Lahmar *et al.*, 2007b)). They are passed out in the faeces of infected dogs and remain relatively stable in both faeces and in a variety of climatic conditions, meaning that faeces do not need to be fresh (Deplazes *et al.*, 1990; Jenkins *et al.*, 2000; Raoul *et al.*, 2001). As with many ELISAs, antigen detection is based upon a reaction involving a colour change, which is quantified by recording the optical density (OD) at a specified wavelength in a plate reader or spectrophotometer. Evidence has been found of a broad linear correlation between coproantigen ELISA OD values and worm burdens when worm burdens are high (Deplazes *et al.*, 1992; Allan *et al.*, 1992; Craig *et al.*, 1995; Ahmad and Nizami, 1998; Morishima *et al.*, 1999a; Raoul *et al.*, 2001; Reiterová *et al.*, 2005; Buishi *et al.*, 2005b). Whilst this means that test sensitivity is expected to be lower in cases of low worm burden (often stated to be less than 50 or 100 worms) (Allan *et al.*, 1992; Deplazes *et al.*, 1994; Nonaka *et al.*, 1996; Reiterová *et al.*, 2005; Allan and Craig, 2006)), this also indicates that OD data could be interpreted in a semi-quantitative manner (Raoul *et al.*, 2001). A number of ELISA tests for the detection of *Echinococcus* spp coproantigens are available (Deplazes *et al.*, 1992; Allan *et al.*, 1992; Malgor *et al.*, 1997; Casaravilla *et al.*, 2005; Huang *et al.*, 2007; Morel *et al.*, 2013), and due to their ease of use (hundreds of samples can be tested per day), these now provide the mainstay of large-scale surveillance of echinococcosis (Deplazes *et al.*, 2003), as has been recommended by the WHO and the FAO (WHO/OIE, 2001d), as well as the PAHO (Morel *et al.*, 2013).

A crude meta-analysis (Allan and Craig, 2006) of a number of studies comparing coproantigen ELISA results to those of necropsy (Deplazes *et al.*, 1992, 1999; Allan *et al.*, 1992; Malgor *et al.*, 1997; Morishima *et al.*, 1999a; El-Shehabi *et al.*, 2000; Jenkins *et*

al., 2000; Machnicka *et al.*, 2003; Reiterová *et al.*, 2005; Buishi *et al.*, 2005b) has suggested that coproantigen ELISA methods have a sensitivity of around 80% and a specificity of over 95% for both *E. granulosus* and *E. multilocularis*. However, these characteristics would be expected to vary according to the worm burdens of individual canids (with higher sensitivity for detection of higher burden infections), and therefore according to the overall distribution of burdens in the community under study. Despite this, the coproantigen test is usually interpreted in a dichotomous manner, classifying individual samples as 'positive' or 'negative' for *Echinococcus* coproantigens. The difficulties associated with this dichotomous classification of a continuous variable (the OD value) in any diagnostic situation are well recognised, and are especially true in the case of coproantigen testing of faecal samples in the field situation, where the quality and quantity of sample material may be lower than those used in initial test evaluation (indeed, evaluation of these tests is commonly performed using experimental infections rather than field data, and the performance of the test in the field may be unknown (Torgerson and Deplazes, 2009)).

To date, coproantigen ELISA tests are unable to distinguish between different species or strains of *Echinococcus*, meaning that their use is limited in coendemic areas. PCR-based methods of detection of *Echinococcus* spp egg DNA in faecal samples ('coproPCR' testing) (Craig *et al.*, 1988; Bretagne *et al.*, 1993; Mathis and Deplazes, 2006) are currently the main method of *Echinococcus* species/strain determination based on analysis of faecal samples. These approaches are generally unsuitable for high-throughput situations such as routine surveillance due to the difficulties associated with both extraction of DNA from faeces and from the PCR protocol itself, meaning that the rate of sample analysis is an order of magnitude lower than for coproantigen methods (Deplazes *et al.*, 2003; Torgerson and Deplazes, 2009). The relatively low numbers of eggs passed in the faeces, the presence of inhibitory substances in faeces (Opel *et al.*, 2010), the protection afforded by the embryophore (Bretagne *et al.*, 1993), and the lack of any cell-free DNA in faeces can all have detrimental impacts on test performance; meaning that additional processing steps are required before PCR can be undertaken. These may consist of faecal DNA

extraction methods (for 'copro-DNA PCR' (Bretagne *et al.*, 1993; Monnier *et al.*, 1996; Dinkel *et al.*, 1998; van der Giessen *et al.*, 1999; Abbasi *et al.*, 2003; Boufana *et al.*, 2013)), or taeniid egg extraction using flotation/sieving procedures (for 'egg-DNA PCR' (Mathis *et al.*, 1996; Stefanić *et al.*, 2004; Trachsel *et al.*, 2007; Al-Sabi' *et al.*, 2007; Boubaker *et al.*, 2013)).

Despite the issues described above, coproPCR remains a useful tool as it generally has a very high specificity (although coprophagia has been suggested as a source of false positive results (Ziadinov *et al.*, 2008; Hartnack *et al.*, 2013)), and allows identification of individual species and strains of *Echinococcus*, which cannot be achieved currently using ELISA methods. A variety of PCR primers have been developed for faecal testing for *E. multilocularis* (Bretagne *et al.*, 1993; Monnier *et al.*, 1996; Dinkel *et al.*, 1998; van der Giessen *et al.*, 1999; Boufana *et al.*, 2013) and various strains/species of *E. granulosus* (s.l.) (Abbasi *et al.*, 2003; Stefanić *et al.*, 2004; Dinkel *et al.*, 2004; Boufana *et al.*, 2013) (some of which have been evaluated at the species/strain level (Boufana *et al.*, 2008)). A number of studies have suggested that, as for coproantigen detection, the probability of detection of DNA is positively correlated with worm burden (Dinkel *et al.*, 1998; Lahmar *et al.*, 2007b). More recent developments using quantitative real-time PCR techniques may allow quantification of DNA levels and therefore potentially estimate the worm burden (Knapp *et al.*, 2014). DNA detection has been reported prior to patency, although this appears to have little relationship to coproantigen results (Deplazes *et al.*, 2003; Lahmar *et al.*, 2007b; Boufana *et al.*, 2008). Due to these issues, any statement regarding the sensitivity of PCR tests will depend on a wide variety of factors – including the species/strain(s) of *Echinococcus* present, the stage of infection, and the worm burden in the animal itself; as well as the exact processing approaches used for DNA extraction prior to PCR. Estimates of sensitivity for most PCRs are generally higher than 50%, although estimates as low as 20%, or as high as 100%, have been reported (Bretagne *et al.*, 1993; Mathis *et al.*, 1996; Dinkel *et al.*, 1998; van der Giessen *et al.*, 1999; Abbasi *et al.*, 2003; Stefanić *et al.*, 2004; Mathis and Deplazes, 2006; Trachsel *et al.*, 2007; Ziadinov *et al.*, 2008). A concurrent evaluation of three *E. granulosus*-specific primers (Abbasi *et al.*, 2003; Stefanić *et al.*,

2004; Dinkel *et al.*, 2004) has estimated PCR sensitivity amongst naturally infected dogs of between 50% and 100%, with variable specificities for the different species and strains of *Echinococcus* (Boufana *et al.*, 2008). The latent class model mentioned earlier has also been used to estimate the sensitivity and specificity of *E. granulosus*- (Abbasi *et al.*, 2003) and *E. multilocularis*- (Dinkel *et al.*, 1998) –specific primers, which gave point estimates of sensitivity of between 80 and 90%, and specificity of over 80%, for each (Hartnack *et al.*, 2013).

Finally, it should be noted that some work has been conducted on serological diagnosis of infection in canid hosts (Gasser *et al.*, 1988, 1993, 1994; Jenkins *et al.*, 1990; Gottstein *et al.*, 1991). Although these tests often gave a high sensitivity, their specificity was low (considerably lower than the coproantigen ELISA (Craig *et al.*, 1995)). As such, further work is required in order to develop a useful serological test for canine infection – which will likely require the development of better recombinant antigens (reviewed in (Carmena *et al.*, 2006)).

1.3.2.2 Environment

The most commonly used form of environmental sampling for egg contamination is based upon the collection and testing of environmental canid faeces – usually fox faces, in order to test for *E. multilocularis* infection and gain an estimate of the overall levels of infection in an area (Morishima *et al.*, 1999b; Tsukada *et al.*, 2000; Raoul *et al.*, 2001; Knapp *et al.*, 2014). Although many coproantigen surveillance schemes based on the collection of faecal samples from domestic dogs can be considered to be a form of environmental sampling, these samples are usually matched to individual dogs, with interpretation accounting for this rather than being set at the 'ecological' level (although there are some exceptions to this (Vaniscotte *et al.*, 2011; van Kesteren *et al.*, 2013)). When environmental faeces are studied, the molecular methods described above such as coproantigen ELISA and coproPCR are usually used for diagnosis. Relatively few studies have directly investigated levels of 'true' environmental (e.g. soil) contamination with *Echinococcus* eggs, despite this being of potential importance to human exposure and ongoing transmission. The reason for this is that most

approaches require initial isolation of eggs, which can be very labour-intensive if egg densities are low (although methods are available (Matsuo and Kamiya, 2005)). An early study of environmental contamination used immunological methods to differentiate *Echinococcus* spp eggs from those of other taeniids (Craig *et al.*, 1988), but PCR-based methods such as those described above are most commonly used currently (Matsudo *et al.*, 2003; Shaikenov *et al.*, 2004).

One other method of detection of environmental contamination is through the use of 'sentinel animals': usually sheep, pigs or goats which are known to not be infected with *Echinococcus* spp prior to the study. These are left in the field site of interest for a short period of time (such as two weeks), and are then removed to an *Echinococcus*-free location and killed after a suitable period of time in order for detailed necropsy to be undertaken (Gemmell and Johnstone, 1977; Eckert *et al.*, 1982; Lloyd *et al.*, 1991, 1998). Sheep are a good sentinel for pasture egg contamination, as they do not appear to develop immunity to reinfection (Torgerson *et al.*, 2003b). Measuring infection of cattle with *E. granulosus* has also been suggested as a similar method of surveillance for egg contamination, despite cattle rarely playing a role in the transmission of *E. granulosus* (s.s.). From a diagnostic testing perspective, the presence of cysts in this animal is more likely to represent *E. granulosus* than other pathologies, and these animals are more likely to enter the abattoir system than small ruminants. Therefore, by combining data on cattle infection with GIS data relating to the farm of origin of infected cattle, surveillance and control efforts may be better targeted to those areas with a higher transmission intensity (Cringoli *et al.*, 2007; Rinaldi *et al.*, 2008; Temple *et al.*, 2013; Cassini *et al.*, 2014).

1.3.2.3 Intermediate hosts, excluding humans

As infection in domestic livestock is usually relatively asymptomatic, infection is most commonly identified at necropsy. However, ultrasound examination has also been used to detect liver cysts in some cases (Maxon Sage *et al.*, 1998; Lahmar *et al.*, 2007a; Dore *et al.*, 2014), with one study estimating a sensitivity of 54% and specificity of 98% compared to necropsy (Maxon Sage *et al.*, 1998), although differentiation between the

metacestodes of *E. granulosus* and *Taenia hydatigena* may be difficult (Maxson *et al.*, 1996). Two other problems with examination using either of these methods is the small size of young cysts (which may mean that new infections or infections in young animals are missed) and the length of time required for cysts to develop (which will result in a temporal lag between infection and detection). One solution to this problem in the case of necropsy is to thinly slice the organs of interest for close examination and histological examination as required (Lloyd *et al.*, 1991, 1998). As described above, PCR techniques ('tissue-DNA PCR') are also available to confirm whether detected cysts are those of *Echinococcus spp*, or to determine the strain or species of *Echinococcus* present (Dinkel *et al.*, 2004; Boufana *et al.*, 2008).

Attempts to develop serological screening tests for echinococcosis in livestock intermediate hosts have faced difficulties, with reported low sensitivities and cross-reactions with other cestodes (Yong *et al.*, 1984; Ibrahem *et al.*, 1996; Craig, 1997; Kittelberger *et al.*, 2002). One particular issue is that many studies have compared serology to visual inspection at slaughter. For the reasons described above, accurate identification of infected individuals (or distinction of *E. granulosus* lesions from those caused by other taeniid species) may be difficult, meaning that the 'gold standard' test used for comparison may be imperfect. One study which made a concerted effort to reduce this problem found that serology gave reasonable estimates of sensitivity (85%) and specificity (97%), compared to histological examination and Western Blotting (Gatti *et al.*, 2007).

Medical imaging approaches have also been used to detect *E. granulosus* infection in intermediate hosts. Early work used thoracic radiography to detect lung cysts (Wyn-Jones and Clarkson, 1984), with a view towards identification of infected sheep for pharmacological studies. However, most recent attention has focussed on abdominal ultrasound scanning (US) to detect liver cysts (Craig, 1993). Studies comparing US to inspection at slaughter suggested that this strategy has a reasonable sensitivity and specificity (Maxon Sage *et al.*, 1998; Dore *et al.*, 2014). Another potential advantage of US for diagnosis of infection is that cyst types may be determined, in order to

distinguish 'active' forms from 'inactive' or 'transition' forms (WHO/OIE, 2001e; Lahmar *et al.*, 2007a; Dore *et al.*, 2014).

In the case of *E. multilocularis* infection in short-lived small mammals, careful necropsy of captured animals is the most commonly used method of detection, often combined with tissue-DNA PCR techniques (Stieger *et al.*, 2002; Abdyjaparov and Kuttubaev, 2004; Afonso *et al.*, 2015). As the prevalence of infection is commonly very low, large sample sizes are often required in order to detect infected individuals. This means that comprehensive studies of small mammal infection are relatively rarely undertaken.

1.3.2.4 Humans

A large number of human cases of CE or AE are detected through passive surveillance (i.e. diagnosis in hospital after presentation to a physician), which will tend to result in the preferential detection of more advanced, clinical, cases of disease (Schantz, 1997). However, recent improvements in ultrasound scanning technology and development of portable ultrasound scanners has made active surveillance in the form of screening campaigns possible, even in the remote communities traditionally affected by the disease (Macpherson *et al.*, 1987). A system of grading and classifying CE cysts according to their ultrasonographic appearance has also been developed by the World Health Organisation (WHO/OIE, 2001e; Wang *et al.*, 2003; WHO Informal Working Group, 2003). Whilst ultrasonography is suitable for the detection of hepatic cysts, it is less useful for the detection of pulmonary cysts, which may require methods such as radiography or computed tomography.

A variety of serological tests are also available for the detection of both CE and AE, and are well reviewed elsewhere (Gottstein, 1992; Lightowlers and Gottstein, 1995; WHO/OIE, 2001e; Zhang *et al.*, 2003). As with most serological tests, options are available for the detection of antibodies against the parasite (Kagan, 1968; Gottstein, 1985), or against parasite antigens themselves (Gottstein, 1984; Craig, 1986). Antibody detection for *E. multilocularis* is generally viewed as more reliable than that for *E. granulosus* (Gottstein *et al.*, 1993; Ito *et al.*, 2003a), making it more appropriate for

use alone in screening and surveillance campaigns, whereas surveillance for *E. granulosus* will generally require the use of multiple tests and/or modalities (Eckert and Deplazes, 2004).

1.3.3 Risk factors for canine infection

A review of risk factor studies for canine echinococcosis has recently been published (Otero-Abad and Torgerson, 2013), and so only a brief description of some of the most commonly recognised risk factors will be given here. It should be noted that different studies measured different outcomes – with some measuring coproantigen positivity, some measuring 'true' positivity, and others measuring worm burdens. All of these outcomes will be grouped together here as 'canine infection', although they do represent slightly different measures. The most commonly identified risk factor for canine infection with *E. granulosus* was access to infected offal – whether this is due to purposeful feeding of offal/home slaughtering (Moro *et al.*, 1999; Buishi *et al.*, 2006; Acosta-Jamett *et al.*, 2010), lack of restraint/free roaming (Buishi *et al.*, 2005a, 2006; Guzel *et al.*, 2008; Huang *et al.*, 2008; Mastin *et al.*, 2011), proximity to possible infected offal (Bchir *et al.*, 1987; Wang *et al.*, 2001; Elshazly *et al.*, 2007; Acosta-Jamett *et al.*, 2010), or dog type (farm/working dogs and stray dogs frequently had a higher probability of positivity) (Moro *et al.*, 1999; Shaikenov *et al.*, 2003; Buishi *et al.*, 2005b; Inangolet *et al.*, 2010). A number of studies have also found that older dogs had a lower probability of positivity than younger dogs, which may suggest some degree of acquired immunity (Sharifi and Zia-Ali, 1996; Torgerson *et al.*, 2003c; Buishi *et al.*, 2005b, 2006; Inangolet *et al.*, 2010; Acosta-Jamett *et al.*, 2010). As expected, a lack of knowledge about echinococcosis and a lack of recent praziquantel dosing were also associated with increased probability of positivity (Buishi *et al.*, 2005a; b; Huang *et al.*, 2008; Acosta-Jamett *et al.*, 2010).

Studies of *E. multilocularis* infection in domestic dogs have found similar risk factors to those for *E. granulosus*: with free roaming (Budke *et al.*, 2005a; Ziadinov *et al.*, 2008), proximity/access to intermediate hosts (Wang *et al.*, 2007, 2010; Antolová *et al.*, 2009) and other spatial factors (Dyachenko *et al.*, 2008) commonly identified.

1.4 *Echinococcus* control

1.4.1 Methods of control

The main eim of echinococcosis control is generally to reduce the number of human. In the remote communities where human echinococcosis is most common, the logistics of the required surveillance and provision of suitable medical and surgical care can present a considerable barrier to effective control of human infection. As a result, control of human echinococcosis in these areas is often achieved by management of the definitive host (usually domestic dogs) in an attempt to perturb the parasite lifecycle (Craig *et al.*, 2000). Praziquantel is an isoquinolone drug which causes tetanic muscle contractions and damage to the tegument of adult *Echinococcus* worms, resulting in worm detachment and death (Conder *et al.*, 1981; Elsheikha *et al.*, 2011). Treatment is effective against all species and strains of adult *Echinococcus* spp (although it does not appear to be particularly effective against metacestode forms (King and Mahmoud, 1989)). To date, there have been few reports on the development of resistance amongst *Echinococcus* spp to praziquantel, as has been reported for many other anthelmintic drugs. Despite these promising attributes, instigating an effective praziquantel dosing scheme is very challenging. Praziquantel has no residual action on worms, meaning that repeated treatment is required to prevent reinfection, and it has been shown that when dosing is not supervised by trained operatives, control of *Echinococcus* is often not achieved (likely due to dogs not being dosed in these cases) (Craig and Larrieu, 2006). Additionally, this approach will generally not impact upon unowned dogs in the community, which can also be a source of infection (Inangolet *et al.*, 2010). Techniques involving slow-releasing praziquantel (such as subcutaneous implantation of praziquantel-loaded bars (Wei *et al.*, 2005; Cheng *et al.*, 2010)) have been attempted, but can be labour intensive and have not yet been fully evaluated. Despite these issues, praziquantel dosing of dogs has become the predominant control strategy for domestic dog-associated echinococcosis worldwide (Economides and Christofi, 2000; WHO/OIE, 2001f; Jenkins, 2005; Craig *et al.*, 2007b; Zhang *et al.*, 2009, 2015; Larrieu and Zanini, 2012; Barnes *et al.*, 2012).

Dog population management – in particular, culling campaigns – are also commonly used as a method of control of disease and other problems associated with dogs (Deplazes and Hegglin, 2004; Kachani and Heath, 2014). These campaigns are most commonly focussed on reducing the risks of dog bites and rabies, especially due to unowned dogs, but can also be applied to echinococcosis control. A number of methods of euthanasia are still in use worldwide, including strychinine and cyanide poisoning which are currently considered inhumane and so should be avoided (Tasker, 2008; OIE, 2010). More humane methods of culling include shooting with a free bullet (with an accurate shot to the head, which will require an experienced marksman) and use of injectable anaesthetics (Tasker, 2008). A well-organised culling campaign based tailored to the community in question and carried out with full knowledge of the local community can be an effective supplemental method of controlling echinococcosis, but culling as a control strategy remains controversial (Johansen and Penrith, 2009). Therefore, it is important to work with a community when controlling echinococcosis. Dog culling is a contentious subject, and implementing culling campaigns without the consent of a community can risk damaging relations and result in reduced ongoing cooperation (Atema and Hiby, 2015).

Another method of control is vaccination of intermediate livestock hosts against *E. granulosus* (Lightowlers *et al.*, 1996, 1999; Heath *et al.*, 2003, 2004; Zhang and McManus, 2008). Although this is relatively labour intensive and would not be expected to have an immediate effect on the risk of human infection, this can be a useful adjunct to dog dosing campaigns (Torgerson, 2003b, 2006a). It is also possible that it can be combined with ongoing livestock vaccination schemes (such as for brucellosis or peste des petits ruminants), in order to minimise logistical and financial hurdles.

Methods of infection control in the small mammal intermediate hosts of *E. multilocularis* are much less effective and more challenging than in other hosts. Culling campaigns are likely to have considerable ecological repercussions, and may have little effect on the levels of infection in definitive hosts (since some species of

intermediate host are predated preferentially despite their population density (Hegglin *et al.*, 2007; Raoul *et al.*, 2010)). One possible avenue for exploration is modifying land management practices in order to indirectly alter intermediate host population densities (Viel *et al.*, 1999; Wang *et al.*, 2006b; Giraudoux *et al.*, 2006; Hegglin and Deplazes, 2013).

Reducing human exposure directly would not be expected to impact upon the natural lifecycle of the parasite, but could be a useful adjunct to control schemes. Education of people about the risks from dogs, encouraging avoiding contact with dogs, and/or ensuring hands are washed after dog contact are also potential strategies for directly reducing transmission to humans. Additionally, increasing public knowledge of the risks would be expected to increase compliance with other control measures. As such, education campaigns and community involvement are important tools in the control of echinococcosis. However, the failure of most control schemes using this as a sole method of control suggests that it should be combined with other approaches in order to be effective at actually reducing *Echinococcus* burdens (Craig and Larrieu, 2006).

1.4.2 Economics of control

The economics of *Echinococcus* control have a considerable impact on the ultimate effect of a control scheme, as any control scheme will need to be run for a considerable period of time (in some cases, indefinitely) in order to be effective (Craig and Larrieu, 2006). Also, due to the time lag between infection and development of disease in humans or livestock, the economic benefits would not be expected to become apparent immediately. Although praziquantel is relatively cheap, the logistical costs required for effective (i.e. supervised) dosing, and the relatively high frequency and long durations required for an effective dosing scheme can make a dosing scheme expensive. Another challenge associated with the long periods required is that of public perception: if a control scheme is successful, then the perception of risk in the community will tend to decrease and so will demand for control. However, ending an effective control scheme prematurely can result in a complete failure of control, as was seen in Wales (Craig and Larrieu, 2006).

As *E. granulosus* affects both livestock and humans, its control may become economical from both a health (measured using Disability-Adjusted Life Years (DALYs)) and a financial perspective. A study modelling the control of brucellosis (which also has effects on both livestock and humans) in Mongolia suggested that a benefit-cost ratio of around 3.2 (with a net present value of around US$ 8million) can be achieved, as benefits result from both reduced human infection and from reduced animal losses (Roth *et al.*, 2003). A study on the eastern Tibetan plateau (where *E. granulosus* and *E. multilocularis* are coendemic) suggested that combined anthelmintic dosing of livestock and praziquantel dosing of dogs offered a financial benefit, as well as being cost-effective for reduction of DALYs, especially if cost sharing between the public health and agricultural sectors was attempted (Budke *et al.*, 2005c). The economics of *E. granulosus* control from a financial perspective have been reviewed elsewhere (Torgerson *et al.*, 2000, 2001; Torgerson and Dowling, 2001; Torgerson, 2003a).

1.5 Echinococcus ecology

1.5.1 Overdispersion

Overdispersion has been described as 'one of the most important features of the epidemiology of helminth parasites' (Anderson and May, 1991a), and therefore a considerable amount of attention has been (and continues to be) focussed on this aspect of parasite ecology (Anderson and May, 1978, 1985; May and Anderson, 1978; Pacala and Dobson, 1988; Quinnell *et al.*, 1995, 1990; Medley, 1992; Barbour and Kafetzaki, 1993; Grenfell *et al.*, 1995; Shaw *et al.*, 1998; Galvani, 2003; Churcher *et al.*, 2005). Overdispersion is a characteristic of the relationship between many metazoan parasites ('macroparasites') and their hosts, and is broadly described as the situation where parasite biomass appears to be 'clustered' within certain hosts. In the most extreme possible example of an overdispersed distribution, the entire parasite population would be found within only one host (Anderson and Gordon, 1982) (conversely, 'underdispersion' would be seen in each host in the population harboured the exact same number of parasites). The host-parasite relationship, as with any

ecological relationship, can be best understood by study at the population level (Crofton, 1971a) – in particular, through investigation of the frequency distribution of parasite burdens amongst different hosts (Shaw and Dobson, 1995). Given that counts of parasite burdens could be considered to follow a Poisson distribution if randomly distributed between hosts, overdispersion can be framed in a statistical context, where it represents the situation in which the variance of this frequency distribution is greater than the mean (in a Poisson distribution, the mean and the variance would be expected to be equal). This would appear visually as a distribution with a strong right skew – such as that shown in figure 1.3, which is a (theoretical) right-skewed distribution of burdens with a mean of 114 and a variance of 30,000.

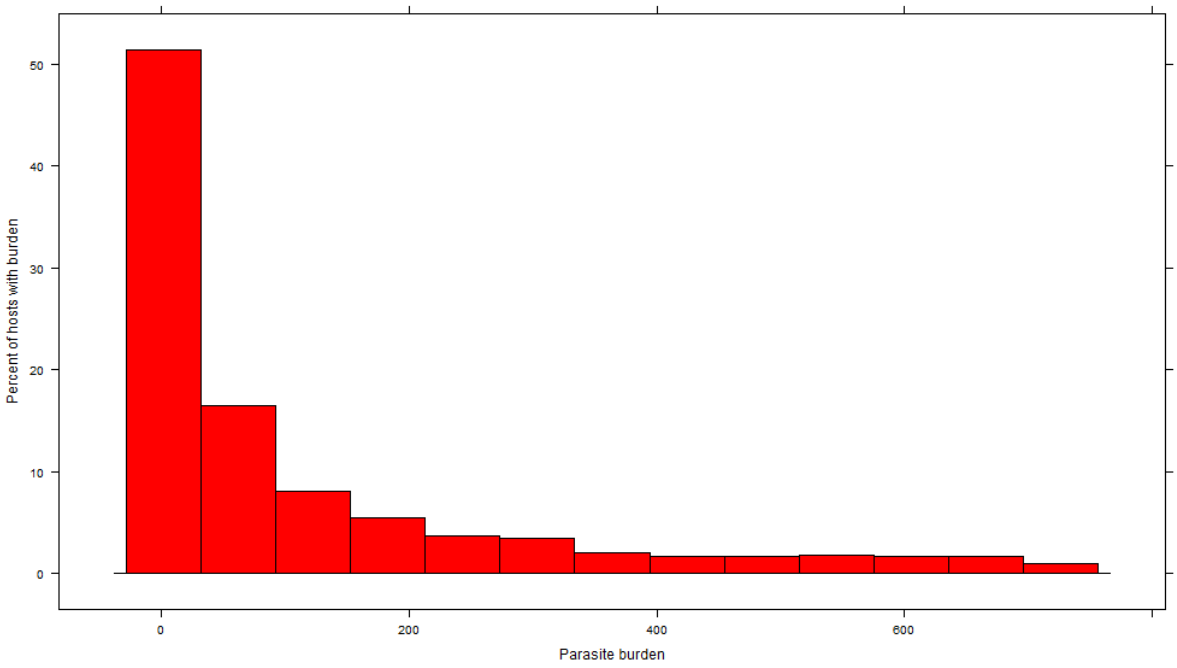


Figure 1.3. Representation of theoretic right-skewed distribution, as is commonly observed with parasite burdens

It has been noted that 'it is difficult, if not impossible…to try to reach conclusions about the biological mechanisms generating a particular distribution pattern by simply examining the resultant observed distribution of parasite numbers per host' (Anderson and May, 1978). Despite this, identification of overarching patterns and relationships is a key area of exploration in parasite ecology (Poulin, 2007). A power

law relationship (see later) has been identified between the mean and the variance of various animal and plant populations (Taylor, 1961) as well as in host-parasite relationships (Shaw and Dobson, 1995; Poulin, 2007, 2013). However, most work on the investigation of overdispersion in parasitism is based upon the use of mathematical models (Anderson and May, 1985; Medley, 1992). Early mathematical modelling work suggested that overdispersion conferred some stability on macroparasites (Anderson and May, 1978), with further work suggesting that the relationship between the mean and the variance (the 'dispersion'), rather than aggregation *per se*, was the key factor in generating this stability (Adler and Kretzschmar, 1992; Kretzschmar and Adler, 1993). A wide range of mathematical models have been developed over the last 40 years to further investigate overdispersion in parasite distribution - a full description of which is beyond the scope of the current report. However, some general points will be made below.

Particular consideration has been given to identification of the processes which give rise to overdispersion, with particular attention focussed on variations of the construct of stochasticity in host susceptibility to infection, and stochasticity in host exposure to infection (such as 'clumping' in the number of parasites acquired per infection) (Anderson *et al.*, 1978; Anderson and Gordon, 1982; Anderson and May, 1985; Quinnell *et al.*, 1990). These are examples of 'environmental stochasticity', and result from variation in the transmission processes above and beyond those solely expected due to the natural probabilistic nature of these events (described as 'demographic stochasticity') (Anderson and Gordon, 1982; Engen *et al.*, 1998). A recent study investigated the relative effects of these processes by modelling the infection process as the product of 'encounters' (i.e. the number of times a host is exposed to parasites) and 'successes' (which could be considered to represent the number of parasites acquired per exposure). When these two processes were allowed to vary randomly (representing demographic stochasticity only), an overdispersed distribution resulted – suggesting that demographic stochasticity alone is able to produce overdispersion (Gourbière *et al.*, 2015). The generating processes underlying overdispersion are of potential importance when considering the effect of control, since variation in host

susceptibility has been proposed to result in increased parasite stability compared to that resulting from clumping of infection (Rosà and Pugliese, 2002). However, the interplay of other related processes such as density dependence within hosts would also be expected to have an effect on stability and should also be considered when investigating parasite overdispersion (Keymer, 1982; Churcher *et al.*, 2005).

1.5.2 Spatial overdispersion

The Russian parasitologist and geographer Yevgeny Pavlovsky developed the original concept of spatial epidemiology in the 1930s, which he described as 'landscape epidemiology' (Pavlovsky, 1966). This concept can be summarised in three general points: firstly, zoonotic and vector-borne diseases (i.e. those not dependent solely on human-human contact) tend to be geographically clustered; secondly, this clustering results from variation in the physical or biological environment; and thirdly, if these factors can be mapped, then current and possible future risk of disease can be predicted (Ostfeld *et al.*, 2005). Despite the apparent ease with which these general concepts can be described, the consequence of this idea is that study of diseases should be conducted at the landscape level, accounting for both static and dynamic characteristics of the relevant factors (along with their interactions) at a variety of spatial scales. Application of these concepts to a number of case studies of zoonotic disease resulted in the identification of ten important principles requiring consideration when studying diseases from this 'ecological' perspective. These were summarised in a figure, reproduced here as Figure 1.4 (Lambin *et al.*, 2010), and were as follows:

1. Landscape attributes may influence transmission
2. Presence, area and spatial configuration of habitats affect the spatial distribution of transmission risk
3. Transmission risk also depends on the connectivity of vector and host habitats
4. In the case of multi-host pathogens, the landscape can be considered a proxy for specific associations of different hosts
5. Pathways of pathogen transmission between different hosts and the environment are of importance to spatial variations in transmission risk

6. Different factors acting at different scales affect the spatiotemporal emergence and distribution of transmission
7. Landscape and meteorological factors affect the emergence, the spatial concentration, and the spatial diffusion of transmission risk
8. Land use as well as land cover is of importance to transmission risk
9. The relationship between land use and the probability of contact between vectors and animal hosts and human hosts is influenced by land ownership
10. Human behaviour is a crucial controlling factor of vector-human contacts, and therefore transmission

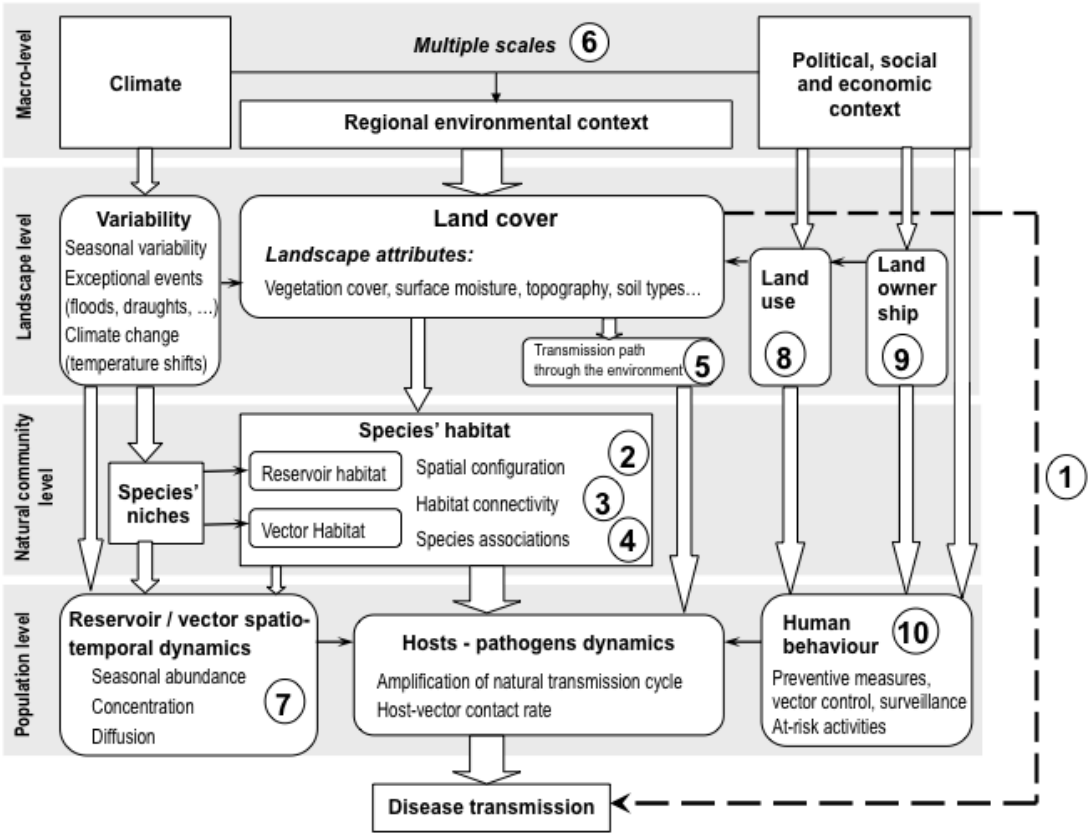


Figure 1.4. 'Ecological' determinants of pathogen transmission. From Lambin *et al.*, 2010. Numbers relate to the ten principles identified in the text.

These general concepts are equivalent to those identified in the ecological study of organisms. This is logical, as infection and disease can be considered to be a form of ecological interaction between pathogens and hosts, and is likely to be particularly true in the case of macroparasites – the study of which exists on the frontier of epidemiology and ecology. As expected, spatial aggregation is commonly seen with

macroparasites (May, 1978; Woolhouse and Chandiwana, 1989; Shaw and Dobson, 1995), and (similarly to aggregation within hosts) is thought to stabilise parasite dynamics to some degree, possibly in association with variability in acquisition of parasites by hosts (Pacala *et al.*, 1990; Hassell *et al.*, 1991a; Holt and Hassell, 1993; Brockhurst *et al.*, 2006). Indeed, it has been suggested that the majority of the spatial variation observed may be due more to variation in acquisition than it is to some 'true' underlying spatial variation in parasite location *per se* (Reeve *et al.*, 1994). In particular, it has been found that fragmented yet interconnected distribution of smaller populations (i.e. a 'metapopulation') can stabilise parasite presence (Hassell *et al.*, 1991b; Bonsall *et al.*, 2002).

1.5.3 *Echinococcus* in definitive hosts

Early experimental work on *E. granulosus* in domestic dogs was conducted by Gemmell, Lawson and Roberts (Gemmell *et al.*, 1986c), which found that worm burdens were overdispersed. Linear relationships were found between the log of the number of protoscolices administered and the resultant log mean worm burden, and between the log mean worm burden and the log of the dispersion (based on the ratio of the variance to the mean of the burden distribution), for each dose of protoscolices administered.

Due to the challenges associated with obtaining a sufficient sample size to effectively evaluate levels of overdispersion (Kapel *et al.*, 2006), many studies are based upon field data rather than experimental data. These studies have invariably found evidence of overdispersion, with the majority of the population harbouring no *Echinococcus* worms, and the majority of the total *Echinococcus* biomass being found in a small proportion of the infected population (Jenkins and Morris, 1991; Hofer *et al.*, 2000; Stieger *et al.*, 2002). Where *E. granulosus* data were fitted to the negative binomial distribution, estimates of the negative binomial constant, k, were generally less than 0.1, suggesting considerable overdispersion (Jones and Walters, 1992; Ming *et al.*, 1992; Gasser *et al.*, 1994; Parada *et al.*, 1995; Eslami and Hosseini, 1998; El-Shehabi *et al.*, 2000; Lahmar *et al.*, 2001; Torgerson *et al.*, 2003c; Torgerson and Heath, 2003; Budke *et*

al., 2005b; Azlaf *et al.*, 2007; Ziadinov *et al.*, 2008). A similar result was seen in *E. multilocularis* infections of domestic dogs (Budke *et al.*, 2005b; Ziadinov *et al.*, 2008) and foxes (Hofer *et al.*, 2000).

1.5.4 *Echinococcus* in intermediate hosts

Many studies of echinococcosis in intermediate hosts have investigated the numbers of cysts. However, it should be remembered that the infectious agent in the metacestode stage of *Echinococcus* spp is not the cyst, but the protoscolices within the cyst. Another issue with the investigation of cyst burdens is the fact that not all cysts will be fertile; meaning that there are two possible measurements to consider: the total number of cysts and the number of viable cysts. Experimental work by Gemmell, Lawson and Roberts found that both the total number of cysts and the number of viable cysts were overdispersed. As was the case for canine infection, broadly linear relationships between the log of both total cyst and viable cyst numbers and the log number of eggs administered were found, and a positive correlation between the number of cysts (either total or viable) and the dispersion were also observed. It should be noted however that only three different doses of eggs were administered to these animals, so the numbers of data points are few (Gemmell *et al.*, 1986c). Interestingly, a study of the number of cysts in sheep of different ages in Tunisia found evidence of overdispersion, but that the degree of overdispersion reduced as age increased – meaning that aggregation of cysts was lower in older animals (Lahmar *et al.*, 1999). This 'density dependence' was hypothesised to result from either space constraints limiting further cyst development or immunological effects. A study of the protoscolex burden of Kyrgyz sheep also found evidence of considerable overdispersion, with older sheep constituting only 28% of the sampled population, but containing around 80% of all protoscolices (Torgerson *et al.*, 2009a).

Despite the difficulties in sampling rodent hosts of *E. multilocularis* (as described above), overdispersion of this parasite has also been observed in these hosts (Roberts and Aubert, 1995; Burlet *et al.*, 2011). This has been suggested to result from spatial factors, and mathematical modelling of *E. multilocularis* infection of intermediate

hosts and foxes has suggested that landscape characteristics resulting in heterogeneous inactivation of eggs (rather than variation in host location or susceptibility) were most able to reproduce empirical data (Hansen *et al.*, 2004), which would also be expected to ultimately result in a heterogeneous distribution of infected foxes (Hansen *et al.*, 2003).

1.5.5 Spatial factors associated with echinococcosis

It is well known that spatial heterogeneity is an important characteristic of *E. multilocularis* infection in human accidental hosts (Craig *et al.*, 2000; Danson *et al.*, 2003, 2006; Graham *et al.*, 2005; Giraudoux *et al.*, 2006, 2013a; b), and in fox definitive hosts (Staubach *et al.*, 2001; Stieger *et al.*, 2002; Pleydell *et al.*, 2004). Evidence of spatial aggregation in domestic dog infection with *E. multilocularis* has also been found (Wang *et al.*, 2012). Little is known of the spatial risk factors for infection in wild intermediate hosts due to the difficulties inherent in diagnosis of infection in these animals, but as mentioned above, differential egg survival (for example, due to soil moisture content) has been postulated to be of importance (Hansen *et al.*, 2003, 2004).

Attempts to identify risk factors for the spatial distribution of *E. multilocularis* have identified suggested that different processes may operate at different spatial scales (Tackmann *et al.*, 1998; Danson *et al.*, 2003; Giraudoux *et al.*, 2003; Graham *et al.*, 2005). The majority of the work investigating this has been conducted in China, where three spatial scales have been suggested: the continental scale; the more local 'regional' scale; and the very narrow 'patch' scale. At the continental scale, the presence of *E. multilocularis* is generally associated with climatic and landscape factors – in particular, the presence of grassland and meadows (Zhou *et al.*, 2000; Danson *et al.*, 2003; Giraudoux *et al.*, 2013a). However, within these areas, at the regional scale, parasite presence appears to be associated with the presence and the dynamics of suitable intermediate hosts. The ratio of optimal to marginal habitat has been suggested to be associated with the population dynamics of small mammals (Lidicker, 2000), and may therefore be of relevance to the risk of *E. multilocularis* infection (Giraudoux *et al.*, 2003, 2013b). In particular, *E. multilocularis* risk to humans has been

found to be associated with areas of low intermediate host biodiversity and therefore a predisposition to large population increases (Giraudoux *et al.*, 2013a). One consequence of this relationship is that human activities which affect small mammal dynamics and distribution, such as deforestation, may have a considerable impact upon the risk of *E. multilocularis* infection (Giraudoux *et al.*, 2003, 2006; Yang *et al.*, 2012). It has been suggested that echinococcosis in intermediate hosts is stabilised through the "metapopulation" of hosts, which links processes occurring at different spatial and temporal scales, as shown in figure 1.5.

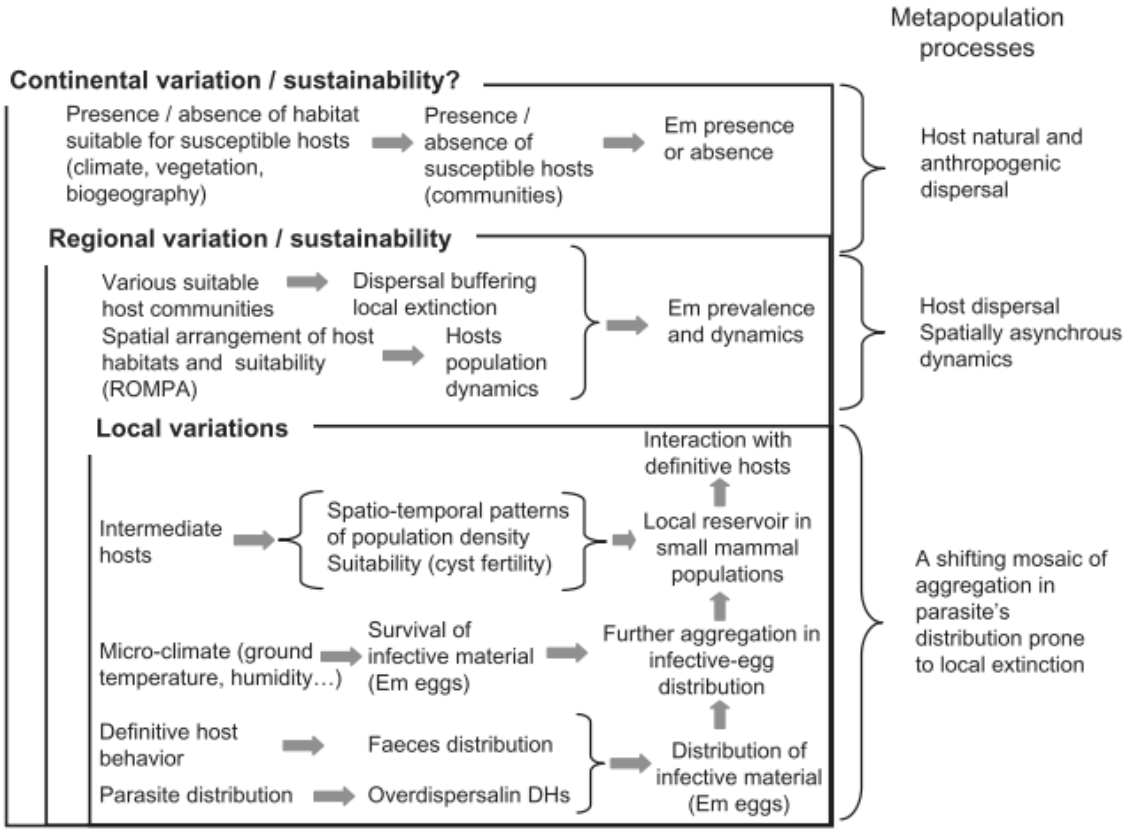


Figure 1.5. Processes affecting *Echinococcus* transmission stability. Taken from Giraudoux *et al.*, 2006.

Despite the associations identified above, overdispersion and clustering of *E. multilocularis* presence at the narrow spatial scale remains, making the exact distribution of the parasite difficult to predict. This has been postulated to result from a combination of variation in the presence of suitable intermediate hosts (with

clumped infections); movement of infected foxes (also with clumped infections); and environmental-associated variation in egg persistence due to microclimatic factors (Hansen *et al.*, 2000, 2003, 2004; Danson *et al.*, 2003, 2006; Giraudoux *et al.*, 2006; Shaikenov, 2006).

1.6 Mathematical modelling of *Echinococcus*

1.6.1 Modelling macroparasites

The study of the epidemiology of helminths has undergone periods of growth and relative decline over the last 100 years (Anderson and May, 1985). Early developments in the field were likely facilitated by the availability of relatively crude diagnostic techniques such as faecal egg detection using flotation and microscopy. However, in order to fully understand helminth epidemiology, data relating to worm burden is required. This results from the fact that most helminths do not reproduce directly within their host, meaning that the burden will often reflect the balance of parasite immigration (infection) and death. This can result in a variety of different patterns of infection within individual hosts – with effects upon the processes such as transmission and morbidity.

A wide variety of modelling approaches are available which can lend an insight into the biology and epidemiology of helminths, but only mathematical models of helminth population dynamics will be described here. These can broadly be classified into two types: differential equation models and agent-based models, of which most focus here will be placed upon the former. Differential equations can be used to model the instantaneous rate of change of the variables of interest (rather than focussing on modelling these directly) over time. The advantage of this strategy is that time is modelled in a continuous nature, rather than as a series of discrete time steps (which could result in compounding errors over time and can become overly complex in the case of large systems).

The central aim of mathematical modelling (as with any form of modelling) is the representation of a system or process in a simplified form. There has been some

debate about the validity of mathematical modelling when applied to more complex systems such helminth infections. As described above, in order to effectively model these systems, the burden of infection should be accounted for. Other challenges relate to aggregation of parasites (both within hosts and spatially), the presence of more complex lifecycles (such as those requiring intermediate hosts), and dioecy in some cases. It has been argued, however, that in order to definitively state that mathematical modelling approaches are unsuitable, they should be attempted and evaluated for the system in hand. The presence of apparently complex dynamical processes *per se* is not proof that these were themselves generated by complex processes (indeed, it has been shown that complex patterns can be generated from relatively simple models (May, 1976; Bolker and Grenfell, 1993)). Additionally, some relevant aspects of epidemiological processes may be identified from simple models even if the complete system under study cannot be accurately modelled. As such, it is important to remember that mathematical models are specialised tools with a defined purpose – whether this is to shed light on data gaps; predict the future (such as the evaluation of possible control schemes); identify characteristics of a host:parasite relationship; or to quantify what would otherwise be unmeasurable (such as the force of infection).

Although the concept of mathematical modelling is therefore appropriate, parameterising a model can be challenging – and this is a particular issue for helminth infections. Despite the importance of incorporating burden data into a model, as described above, these data are rarely available. Therefore, a large amount of work to date (especially in the case of helminth infections of humans) has relied upon the use of indirect measures of abundance such as faecal egg counts. As these do not offer an exact estimate of the true burden, this should be accounted for in the analytic process or in the final interpretation of the results.

It is important to consider the limitations of models when interpreting their results: in the words of George Box, “Essentially, all models are wrong, but some are useful” (Box and Draper, 1987), and the quality of any model is largely dependent on the quality of the data used for parameterisation and validation (Hollingsworth, 2009).

1.6.2 Modelling overdispersion

As described by Crofton, 'one of the few methods of expressing a quantitative relationship between hosts and their parasites is by the use of frequency distributions' (Crofton, 1971a), and a variety of frequency distributions are available to enable this (although robust methods of fitting data to these distributions have only relatively recently become available). Earlier attempts to represent overdispersion in parasite burdens have either log-transformed burden estimates (Wilson *et al.*, 1996), or concentrated on the use of the negative binomial distribution (Fisher, 1941; Bliss and Fisher, 1953; Crofton, 1971a; Roberts *et al.*, 1986; Shaw and Dobson, 1995; Budke *et al.*, 2005b) or its limiting form, the Log series distribution (Williams, 1964; Crofton, 1971a).

Because the numbers of *Echinococcus* worms present, unlike with many macroparasitic infections, can commonly reach into the thousands (Gemmell *et al.*, 1986c), early fitting procedures usually required grouping of the data in order to fit to a negative binomial distribution (O Carroll, 1962). One way to view the negative binomial distribution is as a *compound* probability distribution, created from a Poisson distribution where the rate parameter, lambda, itself follows a gamma distribution (Boswell and Patil, 1970). Although the use of the negative binomial is a largely phenomenological construct, it can be considered in biological terms, resulting from environmental stochasticity in the infection pressure (through whatever mechanism – whether variation in host susceptibility or in clumping of infection). If the "force of infection" (and therefore the 'expected' infection burden) could be considered to vary between the individual hosts in a population according to a gamma distribution, then the resultant parasite burdens would be expected to follow a negative binomial distribution. This is because the actual parasite burden (for any given force of infection) would be expected to vary according to a Poisson distribution (i.e. demographic stochasticity, conditional on the force of infection), and the Poisson 'sampling' of a gamma process results in the negative binomial. Despite this broadly biologically plausible background (although there is little evidence to suggest that the distribution of infection pressure varies according to a gamma distribution), fitting a negative binomial distribution to the data will sometimes result in the

underestimation of the number of negative individuals (possibly associated with the inability of a Poisson distribution to have a rate parameter of zero). One approach to remedy the issue of these 'excess zero' counts is through the use of zero-inflated distributions, such as the zero-inflated negative binomial distribution (Heilbron, 1994; Nødtvedt *et al.*, 2002; Denwood *et al.*, 2008; Ziadinov *et al.*, 2010). The sources of zero burdens (whether 'true' zeros, due to animals which have never been infected, or 'false' zeros, due to animals which have been exposed but not infected or which are infected but not detected) can also be modelled in some cases, which may help improve the accuracy and interpretability of the results obtained (Tyre *et al.*, 2003; Martin *et al.*, 2005; Zuur *et al.*, 2009).

An alternative, more biologically sound, approach to modelling parasite burdens has been proposed by Heinzmann and others (Heinzmann *et al.*, 2009, 2011a), and uses 'compound processes': a compound mixed Poisson process for intermediate host infection, and a shot-noise process for definitive host infection. The compound mixed Poisson process for intermediate hosts estimates the number of cysts acquired by time point t as the sum of N_t independent and identically distributed random variables (S_j). N_t is a mixed Poisson process with a randomly distributed rate parameter which represents the number of 'clumps' of eggs ingested on pasture up to time t. The S_js follow a zero-truncated negative binomial distribution, and describes the number of successfully established cysts per ingested clump. (Heinzmann *et al.*, 2009). The shot noise process for definitive hosts similarly models infection with 'clumps' of parasites (protoscolices in this case) over time, but allows a decrease in parasite burden over time. The model structure remains very similar to the compound mixed Poisson process described above, but the rate parameter of the Poisson process (representing the number of clumps of protoscolices acquired by time t) was fixed, and the independent and identically distributed random variables (representing the number of successfully established worms per ingested clump of protoscolices) were first multiplied with a decay function (representing the loss of parasites over the time period in question). When fitted to data from Kazakhstan, China and Libya, the number of parasites per clump was best described using a lognormal distribution, and

the decay function of established parasites was best described using either a Poisson or a Uniform process. (Heinzmann *et al.*, 2011a).

1.6.3 The Reproduction Ratio

The basic reproduction ratio, R_0, is very commonly used in epidemiological analysis in order to quantify the transmissibility of a pathogen, and in particular whether it is likely to spread within a population. In the case of microparasite epidemiology, it can be broadly defined as the average number of secondary cases resulting from each infection in a totally susceptible population (Anderson and May, 1982). However, in cases where heterogeneities in transmission are present in the population, the estimation of R_0 can become more challenging. As a result of this, the R_0 has been defined mathematically as the 'dominant eigenvalue of a positive linear operator' relating the number of infected hosts in one generation to that in the next generation (Diekmann *et al.*, 1990). Estimation of R_0 for agents with a complex lifecycle (where transmission may be mediated through vectors or intermediate hosts), and/or macroparasites (where the burden of infection is of relevance to transmission) is also challenging (Roberts and Grenfell, 1991, 1992; Mollison *et al.*, 1994; Heesterbeek and Roberts, 1995). Anderson and May defined R_0 for macroparasites as 'the average number of offspring … produced throughout the reproductive life span of a mature parasite that themselves survive to reproductive maturity in the absence of density-dependent constraints on population growth' (Anderson and May, 1991b), which is the same interpretation given to the quantity Q_0 developed by Roberts, Grenfell and Heekesterbeek (Roberts and Grenfell, 1991; Heesterbeek and Roberts, 1995). In a mathematical construct, the 'Q_0' is the dominant eigenvalue of a matrix (K) of transmission functions, raised to the power k, where k is the number of stages in the model (Heesterbeek and Roberts, 1995; Roberts and Heesterbeek, 1995). As well as retaining the threshold properties of R_0 (which is arguably the main output of importance), this approach allows the incorporation of different stages of parasite, different types of hosts and/or different parasites (e.g. competition between similar species), as required, in the model.

Further discussion of R_0 estimates, and methods of estimation for *Echinococcus* spp, is included in the appendix (A1).

1.6.4 Models of *Echinococcus granulosus*

1.6.4.1 Modelling parasite burden

The first model of definitive host infection with *Echinococcus* focussed on modelling the numbers of worms in individuals (Roberts *et al.*, 1986), in a similar fashion to the first mathematical model of helminth infection (described in "Symbiose, Parasitisme et Évolution" by Kostitzin (1934) (Anderson and May, 1985)). Details of this original 'Roberts, Lawson and Gemmell' model will be given here (Roberts *et al.*, 1986), as this model (or slight variations of it) is still commonly used. This model considered four general host statuses for definitive hosts: infected, noninfected, immune (y) and nonimmune (x). Within infected individuals, the number of worms (n) was explicitly modelled:

$$\frac{\partial x_n}{\partial t} = -(\beta + \mu)x_n + \beta(1 - \alpha_{n-1})x_{n-1} + \gamma y_n$$

$$\frac{\partial y_n}{\partial t} = -(\gamma + \mu)y_n + \beta\alpha_{n-1}x_{n-1}$$

The differential equations for the numbers of uninfected dogs over time are as follows:

$$\frac{\partial x_0}{\partial t} = -\beta x_0 + \gamma y_0 + \mu \sum_{n=1}^{\infty} x_n + \mu\delta_n \sum_{n=1}^{\infty} y_n$$

$$\frac{\partial y_0}{\partial t} = \gamma y_0 + \mu(1 - \delta) \sum_{n=1}^{\infty} y_n$$

Where x denotes susceptible dogs, and y denotes immune dogs. n is the worm burden; β is the infection pressure; μ is the rate of complete parasite loss (including host death); α is the probability of development of immunity upon exposure; and δ is the probability of loss of immunity upon loss of parasites (or replacement of dead

hosts with naive individuals). γ denotes the rate of loss of immunity amongst immune animals, and is related to β (since reinfection is likely to boost immunity). t may either represent time (if creating an overall model of transmission) or age of dog (in a closed population, where the relationship between age and level of infection is being investigated).

In the absence of immunity, the following equations can be used:

$$\frac{\partial x_n}{\partial t} = -(\beta + \mu)x_n + \beta x_{n-1}$$

$$\frac{\partial x_0}{\partial t} = -\beta x_0 + \mu \sum_{n=1}^{\infty} x_n$$

For the investigation of infection in intermediate hosts, it can be assumed that infections are permanent (i.e. μ=0), meaning that the rate of change in the mean number of cysts (m) over time (t) can be modelled as the product of the infection pressure (in terms of the rate of acquisition of parasites in the absence of density-dependent constraints) and the proportion of susceptible individuals. If it is assumed that the infection pressure is constant, the differential equation can be formulated as:

$$\frac{\partial m}{\partial t} = \left(\frac{\gamma h}{\gamma + ah}\right) + \frac{ah^2}{\gamma + ah}(exp^{-(\gamma + ah)t})$$

This model differentiates between exposures and infections, due to the clustered nature of infection. N is the number of parasites which become established per exposure; β is the infection pressure in terms of rate of exposure (which varies with time); h is the infection pressure in terms of rate of acquisition of parasites ($h = N\beta$); and a is the rate of development of immunity per parasite ($a = {}^{\alpha}/_{N}$).

1.6.4.2 Estimating model parameters from field data

Estimation of model parameters can be achieved by fitting age-stratified field data collected at an endemic 'steady state' to a model using maximum likelihood or

Bayesian techniques. This approach is known as 'catalytic modelling', as movement between groups (such as the 'uninfected' group to the 'infected' group) takes at a particular rate (in this example, the 'force of infection'), which is not dependent upon the numbers of individuals in the groups themselves (Muench, 1959). This will therefore differ from the 'force of infection' used in transmission models, which will depend upon the numbers of infectious and susceptible individuals in the population.

For both intermediate and definitive hosts, the instantaneous rate of change in the proportion of susceptible animals over time is estimated as the balance of the rate of loss of immunity amongst immune individuals $(1-S)$ and that of acquisition of immunity amongst susceptible individuals (S) (this equation can also be obtained by summing the equations for x_n and x_0 above):

$$\frac{\partial S}{\partial t} = \gamma(1-S) - ahS$$

This differential equation can be solved to obtain the following formula for estimating the proportion of susceptible animals by age t (Torgerson *et al.*, 2003c):

$$S(t) = \frac{1}{\gamma + ah}(\gamma + exp^{-(\gamma+ah)t})$$

As the rate of change of parasite abundance (M) over time will be related to the balance of the basic infection pressure (h) and the parasite death rate (μ), this can be represented by the following differential equation:

$$\frac{\partial M}{\partial t} = hS - \mu M$$

These two equations can be combined and solved to give an equation for the expected variation in the number of parasites over time (Torgerson *et al.*, 2003c):

$$M(t) = \frac{ah^2}{(\gamma+ah)(\mu-\gamma-ah)}\left(exp^{-(\gamma+ah)t} - exp^{-\mu t}\right) + \frac{\gamma h}{\mu(\gamma+ah)}(1-exp^{-\mu t})$$

In the absence of immunity, a=0 and the equation becomes:

$$M(t) = \frac{h}{\mu}(1 - exp^{-\mu t})$$

This approach was used in a Bayesian framework to parameterise a transmission model of *E. granulosus* in Kazakhstan, which suggested that farm dogs were developing immunity, whereas village dogs were not (Torgerson *et al.*, 2003c). A similar strategy was used in Kyrgyzstan, incorporating the results of a number of imperfect tests in order to evaluate test characteristics and estimate the force of infection for *E. granulosus* (and *E. multilocularis)* in dogs (Ziadinov *et al.*, 2008).

1.6.4.3 Modelling prevalence of infection

Due to the highly overdispersed nature of infection in both intermediate and definitive hosts, attempting to model the prevalence of infection rather than the parasite burden is not ideal. However, in some cases it is not possible to obtain suitable estimates of worm burden and so only prevalence data may be available. A model based upon the burden model described above has been developed, and has been applied to field data in China and Kyrgyzstan (Budke *et al.*, 2005b; Ziadinov *et al.*, 2008). Infected dogs are classified in one of two groups:

$$\frac{\partial Y}{\partial t} = -(\gamma + \mu)Y + \alpha\beta S$$

$$\frac{\partial X}{\partial t} = -(\beta + \mu)X + \beta(1 - \alpha)S + \gamma Y$$

Where Y represents the proportion of dogs which are infected but immune, and X represents those which are infected but susceptible to further infection. The total proportion of dogs which are susceptible to further infection (whether already infected or not) is represented as S and is modelled in the same fashion as described earlier:

$$\frac{\partial S}{\partial t} = \gamma(1 - S) - \alpha\beta S$$

The prevalence of infection, $P(t)$, can be either estimated as $X(t) + Y(t)$, or estimated directly in the absence of immunity from the following equation:

$$P(t) = \frac{\beta}{\beta + \mu}(1 - exp^{-(\mu+\beta)t})$$

The equation used to estimate susceptibility above can also be adjusted in order to be used for prevalence data as follows:

$$S(t) = \frac{1}{\gamma + \alpha\beta}(\gamma + \alpha\beta exp^{-(\gamma+\alpha\beta)t})$$

1.6.4.4 Simulation modelling

Individual-based simulation models which operate at the level of the individual animals involved in the transmission cycle have also been developed in an attempt to explicitly account for stochasticity in individual infection with *E. granulosus* (Heinzmann *et al.*, 2011b; Huang *et al.*, 2011). These model individual animals as autonomous units and aim to reproduce the complex behaviour of these units using simple rules, and so differ from the population-based differential equation methods described above. The first of these models is based on the compound process models described earlier (Heinzmann *et al.*, 2009, 2011a; b), which were fitted to data from Kazakhstan (Torgerson *et al.*, 2003b; c). Independent models of the infection of sheep by eggs and of the infection of dogs by protoscolices were linked by simulating a contact pattern between the two hosts in order to create a complete simulation model. Dogs become exposed through ingesting sheep offal, with the infection risk (and therefore the proportion of dogs becoming infected/reinfected in each time step) being related to the number of fertile cysts, modelled as one minus the probability that none of the cysts are fertile. Sheep become exposed through contact with dog faeces, with the probability of contact with faeces selected for each individual sheep at birth/start of simulation from a gamma distribution. Infectivity of canine faeces is not assumed to be related to the worm burden. Deaths of sheep and dogs result in removal from the population and replacement with a new uninfected animal of age zero.

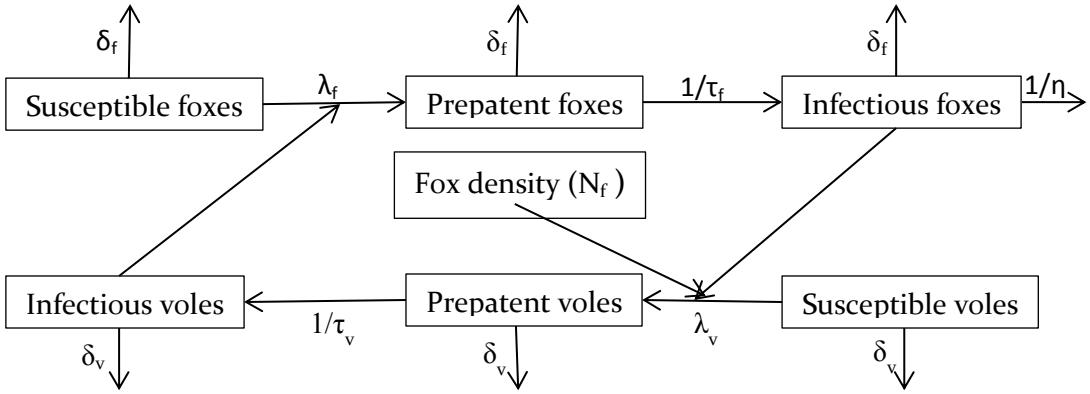
Model fit and parameter estimation was based on the component models, and the model was used to evaluate the possible effects of praziquantel dosing of dogs and of seasonality on transmission of infection (Heinzmann *et al.*, 2011b).

The model by Huang (Huang *et al.*, 2011) was a largely theoretical model using parameters (often point estimates) from other studies and surveys. Dogs, intermediate hosts, parasites and egg contaminations were specified as the agents of interest, each of which functions according to defined characteristics (including survival, ageing and acquisition of infection). The model also includes 'objects' (in this case, the number of egg contaminations and the number of deaths of infectious intermediate hosts), which are not autonomous but are dependent upon the status of the agents, and the environment (a community in western Sichuan province). The model was used to investigate the possible effects of a wide range of control strategies.

1.6.5 Models of *Echinococcus multilocularis*

The approaches used in the modelling of *E. multilocularis* depend to some degree on whether the intention is to capture the sylvatic cycle which exists between foxes and small mammals, or the semi-domestic cycle between domestic dogs and small mammals (no models to date have attempted to capture both simultaneously, which is an issue addressed in chapter 7 of the current thesis).

One of the first models of the *E. multilocularis* sylvatic cycle attempted to combine the dynamics of infection in both foxes and voles with transmission between these species in one formulation (Roberts and Aubert, 1995). This model is based upon the following structure:



Where λ_f is the infectious contact rate for foxes, λ_v is the infectious contact rate for voles, τ_f is the prepatent period (time to worm maturity) in foxes $(=1/_{rate\ of\ maturation\ of\ worms})$, τ_v is the prepatent period (time to cyst maturity) in voles $(=1/_{rate\ of\ maturation\ of\ cysts})$, η is the duration of egg production $(=1/_{life\ expectancy\ of\ adult\ worm})$, δ_v is the mortality rate of voles, δ_f is the mortality rate of foxes, and N_f is the density of foxes.

More recent models have also attempted to incorporate the spatial and/or temporal factors which can affect the distribution and location of intermediate and definitive hosts. Temporally-explicit models have accounted for the variability in parasite and host survival at different times of the year (Ishikawa *et al.*, 2003; Ishikawa, 2006; Nishina and Ishikawa, 2008), and spatially-explicit models have accounted for either the locations of different hosts types in relation to each other (Hansen *et al.*, 2003, 2004), or habitat suitability from an ecological perspective (Milner-Gulland *et al.*, 2004). One challenge faced when constructing a spatial model is deciding on which spatial scale the model should be developed: intermediate hosts often have a smaller home range than definitive hosts, and so some models have been developed based on the definitive host range. However, working at the level of the range of a single definitive host will result in the modelling of few (possibly only one) definitive hosts. As variation in susceptibility between hosts is suspected to be an important factor affecting the distribution of parasites within hosts, ideally a reasonable number of

hosts should be modelled (Morgan *et al.*, 2004). However, increasing the scale of the model will also increase the computational load. A major challenge for spatially explicit modelling of *E. multilocularis* therefore lies in finding a suitable balance between these two issues.

The most commonly adopted approach to the modelling of the semi-domestic cycle of *E. multilocularis* is to treat it in a similar way to *E. granulosus*, as described above. Although no models to date have attempted to model transmission dynamics over time, the force of domestic dog infection with *E. multilocularis* in Tibetan and Kyrgyz communities has been estimated in the same way as that described for *E. granulosus* above (Budke *et al.*, 2005b; Ziadinov *et al.*, 2008).

1.6.5.1 Compartmental models

The models developed by Roberts and Ishikawa are forms of compartmental model, which rather than specifically modelling the numbers of worms, classify fox infection into compartments (uninfected, prepatent infection, infectious for the Roberts model (Roberts and Aubert, 1995); uninfected, prepatent infection, peak egg production, and declining egg production for the Ishikawa model (Ishikawa *et al.*, 2003); non-infectious and infectious for the Vervaeke model (Vervaeke *et al.*, 2006)). This approach is similar to the simple 'SEIS' (susceptible, exposed, infectious, recovered) compartmental model used for microparasites such as bacteria and viruses. In the case of the Ishikawa model, both juvenile foxes and adult foxes are also modelled, with different mortality rates. Similarly, voles were classified as belonging to one of two or three compartments: uninfected, infected but not yet infectious (left out in the final form of the Vervaeke model), and infectious (Roberts and Aubert, 1995; Vervaeke *et al.*, 2006), with five different age classes (0-1 months; 1-2 months; 2-3 months; 3-4 months; > 4 months) considered in the Ishikawa model. The Vervaeke model included a compartment for abundance of eggs in the environment (Vervaeke *et al.*, 2006), and the density of foxes and voles were varied according to season in the Ishikawa model (Ishikawa *et al.*, 2003). More recently, stochastic models of transmission have been developed in order to account for individual variation between foxes (Nishina and

Ishikawa, 2008). A compartmental model classifying foxes as either infected or uninfected was recently incorporated with an economic model for the evaluation of fox anthelmintic dosing schemes (Kato *et al.*, 2010).

1.6.5.2 Measuring the force of infection

A recent model used a Bayesian framework to identify the force of vulpine infection with *E. multilocularis* from prevalence data. Conceptually, this was an extension of the force of infection models described earlier (Budke *et al.*, 2005b; Ziadinov *et al.*, 2008), but estimated transmission parameters (probability of immunity upon exposure, number of insults per unit time, rate of immunity loss, parasite mortality rate) directly from the application of the system of differential equations to age-stratified prevalence data, rather than using the algebraic solutions of the differential equations. Using this technique, a number of different models were compared (with different immunity and force of infection structures). This model suggested seasonal and geographical differences in the force of infection, with periodic increases in the force of infection during the winter months and higher forces of infection for foxes in non-urban areas (Lewis *et al.*, 2014).

1.6.5.3 Mean worm burden models

Although these can be considered a form of compartmental model, they will be described separately here as the formulation is quite different. A common approach to modelling transmission of other helminths is based upon the mean worm burden (MWB), as first described by Macdonald (1965) (despite a number of challenges incorporating overdispersion into this framework (Gurarie *et al.*, 2010)). However, this approach has not been commonly adopted in the modelling of echinococcosis. The only examples of models for echinococcosis based upon the MWB at the population level are the models of Takumi and others (Takumi and van der Giessen, 2005; Takumi *et al.*, 2008). These models describe the dynamics of *E. multilocularis* infection by modelling the parasite biomass within a 1km^2 area, divided according to the stage of the parasite: total eggs, total protoscolices, and total adult worms. The latter two

estimates are adjusted according to the number of available hosts in order to estimate the "mean worm" or "mean protoscolex" burden within these hosts. As this model therefore functions at the level of the parasite rather than the host, it is able to directly incorporate a lag period prior to patency for both intermediate and definitive host infection as well as the persistence of eggs in the environment – making it useful for evaluation of potential control schemes (Takumi and van der Giessen, 2005). However, it does not account for seasonality in infection, the age structure of the population, or density dependence in transmission. This model was developed further in chapter 7 of the current thesis in order to incorporate the *E. granulosus* transmission cycle, along with some seasonal effects. Another modified version of the model included a spatial spreading component of the model, allowing the spread of infection from an initial focus to be modelled (Takumi *et al.*, 2008).

1.6.5.4 Simulation models

The 'Echi' model developed by Hansen is a spatially explicit simulation model combining individual- and grid- based modelling approaches, allowing it to incorporate both spatial factors and individual fox movements (Hansen *et al.*, 2003, 2004). Despite incorporating a lot of information, the general rules governing the model remain relatively simple. This model has been useful in both the evaluation of different control measures (Hansen *et al.*, 2003) and in the investigation of various characteristics relevant to the transmission of *E. multilocularis* (Hansen *et al.*, 2004). The importance of incorporating spatial factors in modelling of the sylvatic (fox-based) cycle *E. multilocularis* is well-recognised (Milner-Gulland *et al.*, 2004; Morgan *et al.*, 2004; Pleydell *et al.*, 2004), but the importance of this in the case of the semi-domestic cycle is less clear.

1.6.5.5 Metapopulation models

Another spatially-explicit model framework, developed by Milner-Gulland *et al*, used an ecological approach derived from the concept of metapopulation dynamics in order to investigate the transmission of *E. multilocularis* (Milner-Gulland *et al.*, 2004).

Metapopulations are 'populations of populations', which although spatially separated by unsuitable habitat types, interact on some level (Hanski, 1991, 1998). Although this concept was originally introduced from an ecological perspective, it can also be applied to infectious pathogens and parasites (Hess, 1996; Grenfell and Harwood, 1997). The Milner-Gulland model adapted this approach to *E. multilocularis* by modelling populations of parasites (whether in the adult, metacestode or egg form) with hosts (and the environment, in the case of eggs) treated as habitats. Therefore, like the Takumi models (Takumi and van der Giessen, 2005; Takumi *et al.*, 2008), this model operates at the parasite level rather than the host level.

1.6.6 Modelling coinfection

Coinfection between two parasite species may need to be explicitly modelled if there is evidence to suggest that infection with one affects infection with the other. In particular, the presence of immunity in sheep towards *T. hydatigena* has been suggested to act to reduce infection with *T. ovis*, despite the R_0 of this species being greater than unity (Roberts *et al.*, 1987). In this particular case, the effect of a control scheme including education campaigns aimed at reducing the feeding of sheep offal to dogs was to reduce exposure to *T. hydatigena*, which resulted in a loss of natural immunity against this parasite. The effects of this were, firstly, more infections with *T. hydatigena* in older animals, and secondly, an increase in infection with *T. ovis*. Alongside this, infection with *E. granulosus*, which does not appear to be considerably regulated by immunity, was pushed towards the extinction steady state (Roberts *et al.*, 1987). Although coinfections between *E. granulosus* and *E. multilocularis* are relatively rarely reported and tend to affect different areas in the small intestine (Thompson and Eckert, 1983; Gemmell *et al.*, 1986c; Lymbery *et al.*, 1989; Morishima *et al.*, 1999a; Umhang *et al.*, 2011), this may also need to be considered when developing a mathematic model for areas where both species are coendemic.

1.6.7 Modelling control measures

As mentioned earlier, the intended output of a large number of the mathematical models of *Echinococcus* spp transmission is some idea of the effect of various control measures. Integration of these control measures within a model require that various model parameters are adjusted. Precise adjustments will depend on the form of the model in question (Torgerson, 2003b, 2006a; Takumi and van der Giessen, 2005; Heinzmann *et al.*, 2011b; Huang *et al.*, 2011), and so will not be described in detail here. However, it is worthy of note that whilst most models have evaluated the effect of control strategies on parasite dynamics, one model has explicitly combined a transmission model and an economic model in order to identify economically optimal strategies for control in different settings and at different time points during the control process (Kato *et al.*, 2010).

1.6.8 Complexity, self-organised criticality and fractal analysis

Echinococcus spp transmission (as with many systems in epidemiology and ecology (Anderson, 1994; Horwitz and Wilcox, 2005)) has a number of characteristics of a 'complex system', with numerous interconnected objects and processes operating at different scales in a nonlinear manner (for example, through nesting and feedback loops) (Anderson, 1994; Goldenfeld and Kadanoff, 1999; Horwitz and Wilcox, 2005; Pearce and Merletti, 2006). Parasites exist within their hosts, which themselves exist within a local ecosystem, which exists within the regional ecosystem, and so on. Each of these levels do not exist in isolation, and changes at one level can have repercussions in the others. As such, study of the constituent components in isolation is unlikely to be able to fully capture the full dynamics of transmission, even if relatively simple rules and patterns are apparent in the system when viewed at an appropriate scale. This "nested" relationship has been demonstrated through spatial analysis of risk factors for *E. multilocularis* infection in people: at the continental scale, climatic conditions and availability of grassland is of importance; at the local scale (kilometres), proximity of human populations to suitable landscapes is important; at the patch scale (villages/households), human behaviour is important; and at the

individual person level, genetic and immunological factors are important (Danson *et al.*, 2003). Considering this, attributing any form of 'causality' to infection with *E. multilocularis* is challenging, as it would be expected to differ at different spatial levels. Indeed, the concept of 'causality' in epidemiological studies is currently undergoing a transition, with a move away from the identification of individual-level exposure-outcome relationships towards a more 'ecological' interpretation, set in a wider scale (such as at the societal level) (Susser and Susser, 1996a; b; Rothman and Greenland, 2005). This paradigm shift is of particular relevance in the presence of complexity, where traditional techniques are likely to give variable and potentially misleading results (Glattre and Nygård, 2004; Glattre *et al.*, 2012).

The concept of 'criticality' is increasingly being applied to epidemiological and ecological scenarios, since it provides a possible explantion for two commonly observed characteristics: threshold behaviour and spatial 'patchiness' in distribution (Pascual and Guichard, 2005). Criticality describes the situation in which large systemic changes can occur in a system in response to small changes in the underlying system conditions, and commonly results in a scale-invariant distribution of outcomes (that is, a power law relationship between the size of event and the frequency of event). Different forms of criticality have been identified, and the form may have considerable repercussions for potential control measures (Zinck *et al.*, 2011). For example, 'self-organising criticality' was introduced by Per Bak and others (Bak *et al.*, 1987; Bak and Chen, 1991) as a possible method whereby complexity arises in nature. Under this theory, dynamic systems naturally evolve into a 'critical' state, which is only barely stable, and can therefore destabilise given particular conditions. The example given in the original paper was that of a pile of sand, with grains being continually added to it. The pile will tend to exist at a particular height and slope (any less than which, and more sand can be added; any greater than which, and 'avalanches' of sand of varying magnitudes will occur). It can be shown that a power law relationship exists between the size of 'avalanches' and the frequency of these events (Bak *et al.*, 1987). However, other forms of criticality have been proposed which are not self-organising, and which may develop in response to changes or variation in

underlying parameters and conditions (Pascual and Guichard, 2005; Zinck *et al.*, 2011). This difference is of great importance to the predicted efficacy of control measures, since a system governed by self-organising criticality would tend towards the critical state regardless of intervention, whereas other systems may be more conducive to particular interventions, depending upon their position in relation to the criticality 'threshold'. This issue is discussed largely in relation to wildfire dynamics in Zinck *et al.* (2011), but equally could apply to epidemiological issues. There are possible parallels between the concept of criticality and that of the endemic persistence of a pathogen in a population (including 'endemic stability', whereby high levels of infection conversely can result in lower incidence of clinical disease (Gemmell, 1978; Coleman *et al.*, 2001)). In the endemic/hyperendemic situation, there is little change in levels of infection over time, yet a disturbance to the system (for example, with praziquantel dosing, in the case of hyperendemic cysticercosis in New Zealand described above (Gemmell, 1978)) can lead to large changes in infection levels.

One method of investigating possible complexity is through fractal analysis, which is based upon investigation of the 'scale invariance' often seen in complex systems. This entails identifying a pattern within the 'fractal dimension' which may be a more appropriate method of description of the situation than those available using more traditional Euclidean or Gaussian techniques. The fractal dimension can be viewed as a form of scaling parameter which measures the 'complexity' in a system – whether the scaling is on a spatial, temporal or some other level. The first example of this concept was described by Benoit Mandelbrot, in an investigation of the 'self-similar' nature of the British coastline. The central concept identified here was that the measured length of the British coastline will vary according to the length of the item used to measure it, due to differences in fine 'resolution' at different spatial scales. The fractal dimension describes the relationship between the measured length of the coastline and the length of the measuring tool used, with higher values (closer to 2) indicating a greater effect of the length of the tool on the measured length (and therefore greater 'complexity'), and lower values (closer to 1) suggesting lower complexity (Mandelbrot, 1967). This same concept can also be applied in a non-geographical context, for

example to epidemiology (Skjerve and Glattre, 2006; Glattre *et al.*, 2008, 2012) and physiology and medicine (Goldberger, 1996).

One characteristic of fractal processes is self-similarity, which may present as a power law relationship: a scale-invariant relationship between two quantities which presents as a linear function on the logarithmic scale. The possible power law relationship governing parasite distributions in hosts has been known for some time, with Anderson and May reporting that 'it is not uncommon to find 80 per cent or more of the macroparasites contained within 20 per cent or fewer of their human hosts' (Anderson and May, 1991c), followed by the assertion being substantiated through empirical analysis (Woolhouse *et al.*, 1997; Perkins *et al.*, 2003). This '80:20' pattern is characteristic of a Pareto distribution: a type of power law relationship. Identification of the form of these relationships may be of great relevance to statistical testing and modelling of *Echinococcus* spp, and may also assist in establishing whether there is evidence of self-organised criticality in *Echinococcus* transmission. Attempts were made during the current thesis to investigate some of these concepts further, especially in the context of coproELISA OD distributions. However, time constraints prevented comprehensive investigation, and therefore this area of exploration is briefly mentioned as an area worthy of potential further investigation in chapter 8 only.

1.7 Kyrgyzstan

1.7.1 Background

Kyrgyzstan (Кыргызстан) is a poor, mountainous country in Central Asia. It is totally landlocked and is bordered to the north by Kazakhstan, to the southeast by China, to the southwest by Tajikistan, and to the west by Uzbekistan. As over 90% of the country is mountainous (Schmidt, 2001), seminomadic pastoralism was traditionally practiced in order to make full use of the pastureland available, with people and their livestock moving between higher and lower ground with the seasons (Schillborn-van Veen, 1995). The country became part of the Russian empire in the late eighteenth

century, and subsequently became a constituent republic of the Soviet Union. Seminomadic practices remained common in rural communities despite attempts at settlement under soviet rule (Farrington, 2005). Independence from the Soviet Union was declared in 1991, and was followed by economic and political instability as attempts were made to develop a democratic republic. Since independence, the proportion of ethnically Kyrgyz inhabitants in the country has gradually increased, leading to nationalism and intermittent outbreaks of sectarian violence. Kyrgyzstan was the first of the Commonwealth of Independent States to join the World Trade Organisation, since which time, exports of gold, cotton, electricity and tobacco have supported economic recovery (World Bank, 2005). Although many of the inhabitants of Kyrgyzstan live in urban centres, agriculture remains an important economic sector, and market reforms since independence have been more progressive than in other ex-Soviet states. Nomadic or semi-nomadic pastoralism remains a way of life in rural communities, and is the optimal method of livestock husbandry in many areas due to the poor quality of the pasture and the mountainous environment. However, due to the costs associated with movement of livestock to summer pastureland ('Jailoo'), there has been a trend towards sedentarisation in recent years. As livestock ownership has remained common, this led to overgrazing of the land around many settlements (World Bank, 2006; Liechti, 2012), and reduced use of Jailoos (Dörre and Borchardt, 2012). Despite most people (especially in the south of the country) being nominally Muslim, dogs are often tolerated due to their guarding (and, to a lesser degree, herding) abilities. However, they are often minimally cared for.

The Alay valley is a high mountain valley (around 3,000m above sea level) situated in the south of Kyrgyzstan (see figure 2.1); bordered to the south by the Pamir Mountains, to the north by the Alay Mountains, and connecting the Xinjiang Uyghur Autonomous Region of the People's Republic of China in the east with Tajikistan in the west. Most villages in the Alay Valley were founded in the 1940s as supply bases for the predominantly Kyrgyz settlement of Murghab in modern day Tajikistan (Paarmann, 2009). Although the valley lies within the oblast (province) of Osh (Ош области), the valley is administratively divided into two raions (districts), with the

westernmost portion of the valley forming Chon-Alay raion (Чоң-алай району; capital: Daroot-Korgon/Дароот-коргон), and the eastern portion (along with the area within the Alay Mountains to the north of the valley) forming Alay raion (Алай району; capital: Gulcha/Гүлчө). Roads to the Alay valley have until recently been of relatively poor quality, although recent restoration work has been conducted in order to improve these as the area provides a useful overland link between Osh and both China and Tajikistan. One mountain pass must be crossed before reaching the Alay valley from Osh or Gulcha, which in winter and poor weather may become non-traversable.

A discussion of healthcare and livestock management in Kyrgyzstan is provided in the appendix (A2).

1.7.2 Echinococcosis in Kyrgyzstan

Passive surveillance of hospital records and active surveillance through ultrasound scanning campaigns have suggested that cystic and alveolar echinococcosis (CE and AE, respectively) is highly endemic in Kyrgyzstan (Torgerson *et al.*, 2003a; Kuttubaev *et al.*, 2004; Usubalieva *et al.*, 2013), and has been increasing in prevalence in recent years. This is thought to be associated with the loss of controls over animal ownership and slaughter, reduced coordinated surveillance and management of animal disease, and increased poverty since independence from the soviet regime (Torgerson *et al.*, 2003a; Torgerson, 2013). The long latent period between infection and clinical CE or AE has resulted in a 'lag period' before the increases in infection seen around the time of independence have become apparent. In response to this (and increases in the prevalence of a number of other zoonotic diseases such as brucellosis (Pappas *et al.*, 2006)), a World Bank–funded project aiming to improve surveillance and control of a number of zoonotic pathogens (including *Echinococcus* spp) has recently been implemented in the country (World Bank, 2005, 2010). Whilst the majority of the interventions (foot and mouth disease, anthrax, brucellosis, sheep pox, peste des petits ruminants, tuberculosis) were targeted at livestock and were based on vaccination, the echinococcosis control scheme targeted domestic dogs and was based largely on

regular praziquantel dosing of domestic dogs every three months by local government veterinarians and paraveterinarians. Other components of the echinococcosis control scheme were based upon education, dog registration, stray dog control, use of slaughterhouses and vaccination of lambs with the EG95 vaccine (WHO, 2011).

Studies of canine infection have recently been conducted in Naryn oblast in the centre of the country, which have demonstrated high levels of infection (Ziadinov *et al.*, 2008), and recent surveillance has also suggested that there are high levels of human infection with AE in the Alay Valley (Professor Bakhadyr Bebezov, Kyrgyz-Russian Slavic University, personal communication) (Torgerson *et al.*, 2015). This led to a Kyrgyz-led expedition to the Alay valley in order to conduct ultrasound surveillance of the inhabitants of two villages in Alay raion: Sary-Mogol (Сары-Могол; on the border with Chon-Alay raion), and Taldu-Suu (Талды-Суу; around 7km to the east of Sary-Mogol). These villages were selected as some of the first reported cases of human infection in the area came from Sary-Mogol, and this scanning campaign found high prevalences of AE in both communities. It is noteworthy that these villages are some of the most isolated from central administration in Kyrgyzstan, with Sary-Mogol situated around 140km from the raion capital, and around 200km from the oblast capital. Until recently, Sary-Mogol was nominally owned by Tajikistan, and was leased to Kyrgyzstan (as the vast majority of its occupants were Kyrgyz). However, although still Tajik in appearance, the village is now in full Kyrgyz ownership, and its inhabitants are almost invariably ethic Kyrgyz.

1.8 Aims of current study

1.8.1 Study setting

As described above, an epidemic of human echinococcosis appears to be developing in Kyrgyzstan, due to factors associated with independence from the Soviet Union over 20 years ago (Torgerson *et al.*, 2003a), with indications that Kyrgyzstan may be a substantial focus of *E. multilocularis* transmission in particular (Usubalieva *et al.*, 2013; Giraudoux *et al.*, 2013b; Torgerson *et al.*, 2015). Despite this, relatively little work has

been conducted in Kyrgyzstan, especially in the southern districts of Osh province, which appear to have a high prevalence of human alveolar echinococcosis (Professor Bakhadyr Bebezov, Kyrgyz-Russian Slavic University, personal communication). Following on from pilot studies in the area, a decision was made to conduct a multidisciplinary study of echinococcosis in the area: focussing on canine infection, human infection and rodent infection/distribution. Along with the original two villages of Sary-Mogol and Taldu-Suu, two additional villages were visited: Kashka'Suu (Кашка'Суу) and Kara-Kabak (Кара-Кабак); both of which are located to the west of Sary-Mogol and Taldu-Suu, in the adjacent district of Chon-Alay. Although these villages are further from Osh (around 220km), they are considerably closer to their raion capital, Daroot-Korgon (around 40km away).

In May and October 2012, dogs in these four villages were registered and faecal samples collected. Dogs in Sary-Mogol and Taldu-Suu were given praziquantel, and those in the other two did were not given praziquantel. Coincidentally, a World Bank–funded canine praziquantel dosing campaign (World Bank, 2010) was commenced in the area in late 2012, and offered an opportunity to investigate the epidemiology of *Echinococcus spp* in this area. Praziquantel dosing continued in all four villages under this campaign, which continued for the remainder of the study (until September 2014). Faecal samples were collected from dogs in all four villages twice annually over this time.

1.8.2 Study aims

The central aims of the current study are to investigate the epidemiology of *Echinococcus spp* infection amongst dogs in rural communities in the Alay valley, evaluate an ongoing praziquantel-based control scheme, and identify possible methods of improving surveillance and control in remote areas such as those in the Alay valley. This latter focus is the major thesis aim, since whilst surveillance is well accepted to be a central component of any control scheme, little work to date has investigated how to achieve this in the remote, rural communities most impacted by echinococcosis, and where obtaining high quality data will be challenging.

1.8.3 Chapter description

Chapter 2 describes the collection, testing, and initial interpretation of faecal samples over the 28 months of the study. All samples underwent coproantigen ELISA testing (Deplazes *et al.*, 1992; Allan *et al.*, 1992), which gives an optical density (OD) value for each sample. Evaluation and identification of methods of interpretation of this output was the main focus of chapters 3 and 4. A selection of samples also underwent coproPCR testing, using three primers: two of which have been previously described (Boufana *et al.*, 2013), and one of which was developed as part of the current work (van Kesteren, 2015). A method of combining PCR results with coproELISA results was investigated in chapter 5, and both PCR and ELISA data were interpreted independently in chapter 6.

Chapter 3 describes an evaluation of different strategies for determination of a cut-off value for coproELISA OD data. This was evaluated using faecal samples collected from dogs of known *Echinococcus* status (as identified by necropsy) in Xinjiang province (van Kesteren *et al.*, 2015). Different approaches were then applied to a number of samples of unknown status in order to evaluate the effects of differences in cut-off on the estimated coproprevalence and the adjusted true prevalence of infection.

Chapter 4 develops a novel strategy for interpretation of coproantigen ELISA data which does not depend upon dichotomisation, using a Bayesian mixture model. This approach uses a panel of concurrently tested samples of suspected negative status to estimate the probability of any individual sample being a true 'positive' sample, and also gives an estimate of the true prevalence of infection amongst the samples tested. This output is then combined with the output of a Bayesian logistic regression in order to allocate a score relating to the expected log worm burden to each sample. This approach therefore can allow extra information to be obtained from the coproantigen test.

Chapter 5 describes a method of combining the results of ELISA and PCR testing whilst also identifying types of dogs in the study villages. A multiple correspondence

analysis was conducted on the individual dog data collected prior to the start of the control scheme in order to broadly identify dog types, and associations between these and the results of diagnostic testing were identified. It is hoped that strategies such as this will help to better characterise communities and improve ongoing surveillance and control.

Chapter 6 describes a logistic regression analysis of trends in coproantigen and coproPCR prevalence over time, during the control scheme. As individual dog identity was not known for individual samples, data were aggregated by household, which may offer a useful strategy for evaluation of control schemes in areas with multi-dog households where it is not possible to sample individual dogs directly.

Chapter 7 details a framework for a novel mathematical model of *Echinococcus* transmission in areas coendemic for both *E. granulosus* (sensu lato) and *E. multilocularis*, incorporating seasonal effects and prepatent periods prior to fertility in the intermediate and definitive host. Whilst most of the chapter discusses the model framework and parameterisation, simulation outputs for trends over time in the presence and absence of dog dosing strategies are shown and described.

Chapter 8 gives an overview of the thesis, drawing general conclusions from the previous chapters and details possible strategies for ongoing surveillance and control in Kyrgyzstan. Areas worthy of further work and investigation are also described.

Chapter 2: Research Methodology

"What we observe is not nature itself,
but nature exposed to our method of questioning."

Werner Heisenberg *(1901 – 1976)*

2.1 Kyrgyz study sites

As described in the previous chapter, the Alay Valley was selected for the current study due to a high reported prevalence of human alveolar echinococcosis (AE) in the area (Professor Bakhadyr Bebezov, Kyrgyz-Russian Slavic University, personal communication; Usubalieva *et al.*, 2013). The Alay valley is bordered to the north by the Alay Mountains, and to the south by the Pamir Mountains. The Pamir Mountains form the border with Tajikistan, which also borders the valley to the west, and to the east is the border with Xinjiang province, China. Two study communities were initially selected, based upon reports of high human AE prevalences: Sary-Mogol (Сары-Могол [39.68°, 72.89°]) and Taldu-Suu (Талды-Суу [39.70°, 72.98°]). Two additional communities were selected based upon proximity to the primary study villages: Kashka'Suu (Кашка'Суу [39.64°, 72.67°]), and Kara-Kabak (Кара-Кабак [39.66°, 72.72°]). Figure 2.1 shows the location of Kyrgyzstan in the Eurasian landmass, and the locations of the four study communities (and associated settlements). These four communities all lie along a major road, the A372, which runs the length of the Alay valley, from the Chinese border to the Tajik border.

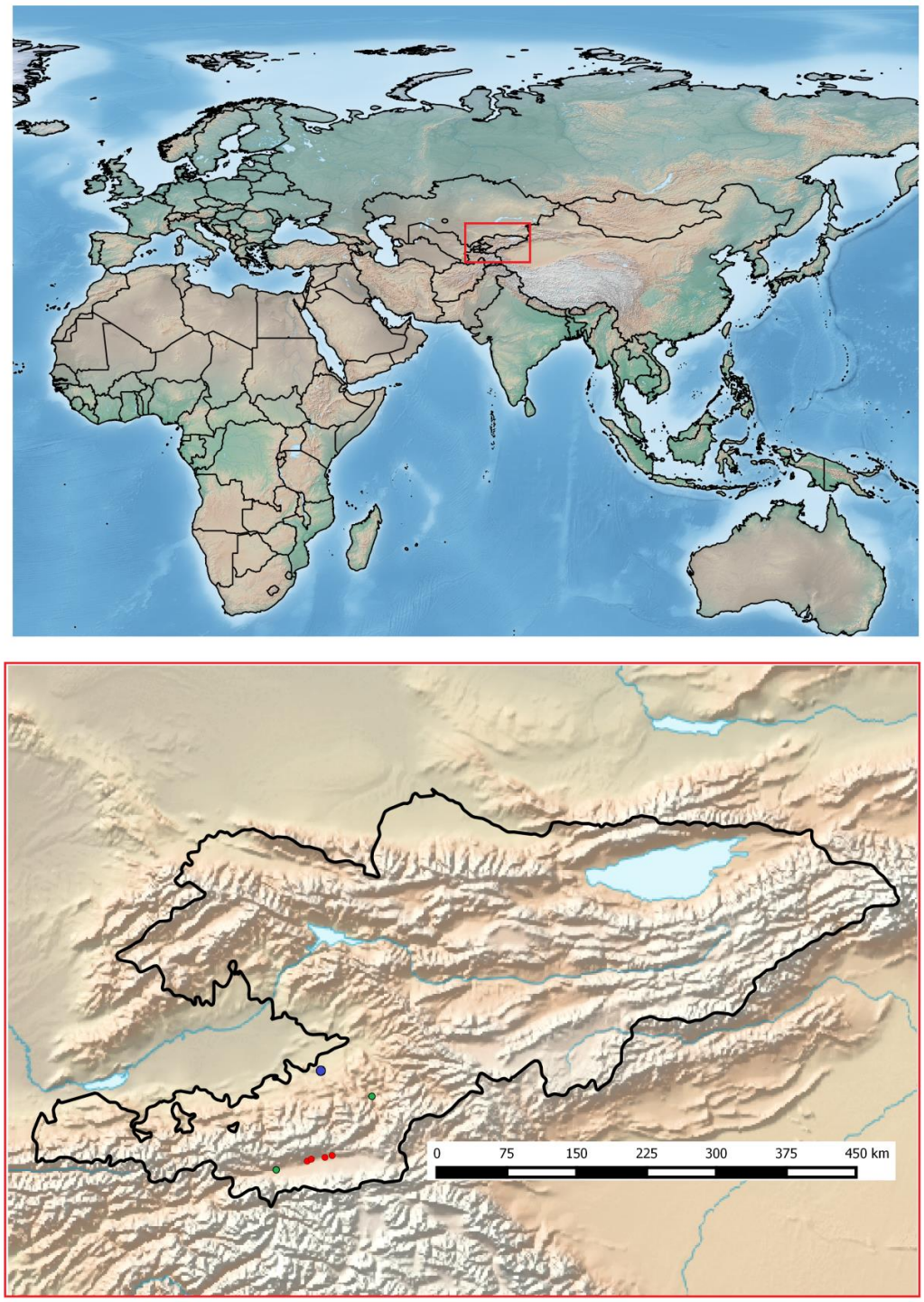


Figure 2.1. Location of Kyrgyzstan (top) and location of areas of interest within the country (bottom). In the lower map, the four study villages are shown in red, and are (from west to east) Kashka'Suu, Kara-Kabak, Sary-Mogol and Taldu-Suu. Raion (district) capitals are shown in green: Daroot-Korgon (west) and Gulcha (east). The oblast (province) capital, Osh, is shown in blue. Map imagery provided by Natural Earth (naturalearthdata.com)

2.2 Kyrgyz sample collection

Villages were first visited in May 2012 (van Kesteren *et al.*, 2013), by a team of researchers from the University of Salford, the University of Zurich, the University of Franche-Comté and Bishkek Veterinary Institute. In Sary-Mogol (SM) and Taldu-Suu (TS), researchers travelled from house to house, georeferenced each occupied house using a Garmin® GPS60 unit, and administered an oral questionnaire in Kyrgyz or Russian with the person who answered the door (or the head of the household, if this was preferred). Details of the questionnaires administered are given in the appendix (A3). The "household questionnaire" contained information regarding household demographics, behaviour and dog and livestock. If dogs were owned and not based permanently in mountain pasture ("Jailoo"), a copy of the "dog questionnaire" was administered for each dog. This contained questions about dog demographics, dog management and recent praziquantel treatment. An attempt was also made to collect a faecal sample from each dog registered. If dogs were present, an attempt was made to collect faeces per rectum (by experienced veterinarians). However, if dogs were not present, could not be restrained, or had recently defaecated, samples were collected from the floor (with attempts made to ascertain faeces 'ownership' with the owner wherever possible). Fresh samples were obtained wherever possible due to the risk of DNA degradation in older samples. Samples were divided upon collection, with some stored in 35ml universal tubes containing 0.3% PBS Tween (Fisher Scientific, Loughborough, UK) buffer with 10% formalin (sourced locally) for coproantigen ELISA; and some stored in bijoux tubes or 15ml polypropylene tubes containing 70% ethanol (sourced locally) for PCR testing. Finally, a random sample of 40 households from Sary-Mogol and Kashka'Suu and 20 households from Taldu-Suu were selected to receive an additional questionnaire relating to healthcare and economic issues.

If dogs were present at the time of visit, they were dosed with praziquantel at a dosage of 5mg/kg. If dogs were not present (either in an unknown location or temporarily at summer pasture), an estimate of their weight was given by the owner and a suitable number of tablets were left with the owner for dosing upon return.

The same process was repeated in Kashka'Suu (KS) and Kara-Kabak (KK), but without praziquantel dosing (with the exception of a number of dogs in KK selected to undergo arecholine purgation). In KS, households were selected in groups of six due to proximity to a number of randomly generated points within the village boundaries (as estimated visually from recent 'SPOT' satellite imagery taken from Google Earth). A broad estimate of the required sample size for each of these two villages were estimated using the equation: $n = \frac{P(1-P)}{d^2}$. Using an expected coproantigen prevalence (P) of 20%, and 95% confidence intervals ($1.96 \times d$) of ±10%, this was estimated at around 60-70 dogs. As population estimates for the villages were not available prior to the visit, no adjustment for finite population size was made, but approximately 25% of the households in KS were ultimately visited, and due to its small size all households in KK were visited. Maps of all sampled dogs in each of the four villages are shown in figures 2.2 and 2.3.

In TS and KK, a total of 33 dogs (convenience sampled with the assistance of the local veterinarian) underwent arecoline purgation. Owners restrained the dogs whilst they were administered a 0.4% solution of arecoline hydrobromide in water orally (7mg arecoline/kg body weight), which was repeated if there was no purge within 30 minutes. The initial faecal void was collected and stored as described above, and the purge was filtered and closely inspected for adult worms by an experienced veterinarian (Iskender Ziadinov). Due to logistical difficulties, it was not possible to match the purges collected from dogs in TS to individual households, although this was possible in KK. As a public health precaution, all purged dogs (including those in Kara-Kabak) were given praziquantel.

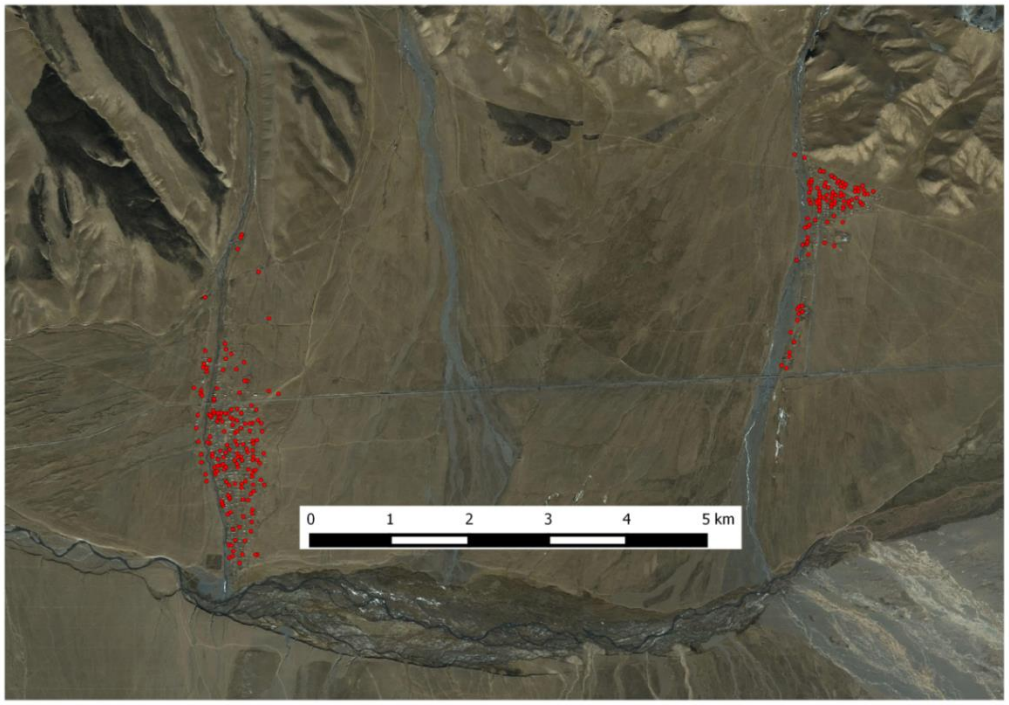


Figure 2.2. Locations of dogs sampled from Sary-Mogol (west) and Taldu-Suu (east) in May 2012. Imagery from Google Earth (satellite image taken 20th Jan, 2012 from SPOT 5 satellite)

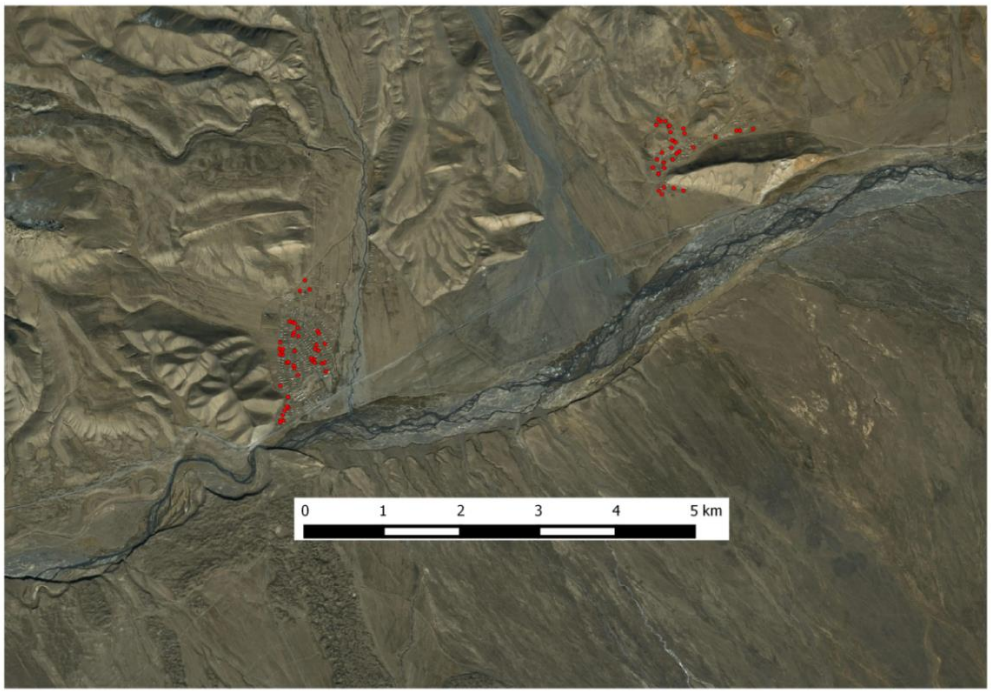


Figure 2.3. Locations of dogs sampled from Kashka'Suu (west) and Kara-Kabak (east) in May 2012. Imagery from Google Earth (satellite image taken 20th Jan, 2012 from SPOT 5 satellite)

A second visit to the field site was made in October 2012 by two researchers from the University of Salford and one veterinarian from the Bishkek Veterinary Institute. Each household in SM, TS and KK which reported owning dogs during the previous visit was revisited again (including any household which had not been visited previously), and all dogs present were identified. Those in SM or TS were dosed with praziquantel as before (either supervised dosing, or tablets left with the owner). In KK, any dogs which had previously undergone arecholine purgation were also given praziquantel. A shortened version of the previous questionnaire was also administered (see appendix), and faecal samples were collected from all available dogs. Due to the difficulties previously experienced with rectal sampling, a decision was made to collect most faecal samples from the floor during this visit. Due to a large culling campaign in the village, fewer dogs were present at this visit, and therefore all available dogs were able to be registered and sampled in these three villages.

A random sample of households in KS were visited in order to sample any dogs present and conduct a questionnaire for each dog. Households were selected randomly from a sampling frame of houses generated from imagery collected by the 'SPOT5' satellite in 2010 (images taken from Google Earth), and households were sequentially visited until around 60 dogs had been sampled. Households without dogs were recorded but did not undergo a questionnaire.

Following the October visit, a World Bank -funded praziquantel dosing scheme was commenced in the area. Raion capitals provided praziquantel and a consent form for canine dosing (detailing the effect of the tablet and required faeces management measures following dosing), although local veterinarians and paraveterinarians needed to collect the tablets and make photocopies of the consent forms themselves. All signed consent forms were to be returned to the raion capital for monitoring. Therefore, on the subsequent visits to the field site in April and September 2013 and April and September 2014, a decision was made to not interfere with the ongoing dosing campaign (which was also providing funding to local vets). As such, only faeces were collected (from the ground) and questionnaires completed, with no overall census conducted and no praziquantel administered.

2.3 Chinese samples

Sample collection for this study has been described in a recent paper (van Kesteren *et al.*, 2015). In April/May 2012, a total of 38 unwanted domestic dogs from northern Xinjiang province in the People's Republic of China were euthanased using ketamine followed by intravenous injection of air. Necropsy was performed on each dog and the intestines were closely studied by experienced fieldworkers in order to estimate the burden of infection with various helminths (including *Echinococcus* and *Taenia* spp). Faecal samples were collected rectally for coproantigen ELISA testing and stored in 10% formal saline, and the presence and burden of *Echinococcus* spp was recorded, together with other intestinal helminths (van Kesteren *et al.*, 2015). These samples will be referred to as the 'necropsy panel'.

Additionally, a sample of owned dogs in the area was made using a spatial technique based upon the World Health Organization Expanded Program on Immunization (EPI) cluster survey method for evaluation of vaccination coverage (Henderson *et al.*, 1973). Investigators moved from the approximate centre of the village in a randomly selected direction, and selected houses using a systematic sampling method. The inclusion criterion was dog ownership, and after each house was visited, the process was repeated again in a random direction from that house. A total six villages were visited, from each of which a sample of between 19 and 26 canine faeces were collected from the floor (originally in order to evaluate coproantigen prevalence using a Lot Quality Assurance Sampling (LQAS) approach (Valadez *et al.*, 2002; Hedt *et al.*, 2008; Pagano and Valadez, 2010)). These samples will be referred to as the 'field panel', and totalled 125 faecal samples.

2.4 Sample testing

All samples were transported back to the University of Salford, England, where they were decontaminated at -80°c for a minimum of four days prior to further analysis (WHO/OIE, 2001d). Full details of the coproantigen ELISA protocol are described elsewhere (van Kesteren, 2015). Samples were homogenized with a wooden spatula,

shaken, centrifuged, and the supernatant tested using a well-known sandwich ELISA protocol for coproantigen detection (after Allan *et al.* (1992), with modifications in that the capture and conjugate antibodies were raised from two different hyperimmune rabbit sera). All samples undergoing coproantigen ELISA were tested using the same reagents in the same batch period (no more than four days), with each sample tested in duplicate (in adjacent wells). A plate of samples of known status (from the Xinjiang necropsy panel described above) was tested with each batch of other samples. Samples from non-endemic and low-endemic areas were added to each plate as negative controls, and known infected samples or samples spiked with *Echinococcus* spp. whole worm extract were added to each plate as positive controls.

After following the ELISA protocol (van Kesteren, 2015), substrate (SureBlue® TMB (Insight Biotechnology, Wembley, UK)) was added to each well, the plate was incubated for a further 20 minutes in darkness, and was then read on a Thermo Scientific Multiscan FC platereader at 620nm. The optical density (OD) readings for the blank well was manually subtracted from OD estimates for each well, but the coproantigen readings were not adjusted in any other way prior to interpretation.

All samples collected in ethanol in May 2012 underwent PCR testing, along with a random sample of 30% of subsequent samples (regardless of the ELISA status, contrary to normal protocol, which usually suggests the use of the ELISA test as a screening test and the PCR test as a confirmatory test (Eckert and Deplazes, 2001; Eckert, 2003; Craig *et al.*, 2003)). This strategy is discussed further in chapter 6. Full details of the PCR testing procedure are available elsewhere (van Kesteren, 2015). DNA was extracted from the ethanol-fixed faecal sample using a commercial DNA extraction kit (QIAamp® DNA stool kit); using 1g of faeces instead of the suggested 180-220mg and increasing the volume of lysis buffer, but otherwise following manufacturer's guidelines (Qiagen, 2010). PCR testing was attempted using highly specific primers targeting the NADH dehydrogenase subunit 1 (ND1) mitochondrial gene for *E. granulosus* G1 and for *E. multilocularis* (Boufana *et al.*, 2013). Primers for the detection of *E. canadensis* G6 were developed specifically for the current project, and are described in more detail elsewhere (van Kesteren, 2015).

2.5 Data processing

Initial data processing for questionnaire data, GPS data, and sample data was conducted using Microsoft Access 2010. Further data processing and analysis was conducted using R version 3.1.1 (R Development Core Team, 2014). The difference in coproantigen ELISA OD between the two duplicates for each sample was calculated and the Studentized residuals of an intercept-only linear regression were inspected for outliers. A Bonferroni correction was applied to the t-test p-value threshold of 0.05 using the "outlierTest" function in the "car" package for R (Fox and Weisberg, 2011), and results with p-values lower than this were classified as failures of replication. These results were removed from the dataset and the samples retested if possible.

Most final data processing and analysis varied according to the particular analysis being conducted, and therefore will be described in the relevant chapters.

Chapter 3: Methods of classification of *Echinococcus* coproantigen ELISA data.

"Errors using inadequate data are much less than those using no data at all"

Charles Babbage (1791 – 1871)

3.1 Introduction

3.1.1 Echinococcosis

As mentioned in chapter 1, echinococcosis has been identified by the World Health Organization as one of five main 'Neglected Zoonotic Diseases' in need of further attention (WHO, 2009), with a view towards the eradication of cystic echinococcosis as a public health problem in selected countries (WHO, 2013b). Domestic dogs play an integral role in human echinococcosis in a number of situations, being the main definitive host for most species and strains of *E. granulosus* sensu lato, and a host of importance for human infection with *E. multilocularis* in a number of locations. Although canine infection with the adult worms is asymptomatic, investigation of the prevalence of infection in dogs can be a useful measure of the risk to humans in an area (Cohen *et al.*, 1998), and is also invaluable for surveillance during a control scheme (Schantz *et al.*, 1995; Gemmell and Schantz, 1997; Schantz, 1997). Although a number of methods are available for detection of canine echinococcosis, coproantigen ELISA tests (Deplazes *et al.*, 1992; Allan *et al.*, 1992; Malgor *et al.*, 1997; Casaravilla *et al.*, 2005; Huang *et al.*, 2007; Morel *et al.*, 2013), are currently recommended as the mainstay of surveillance by the WHO and the FAO (WHO/OIE, 2001d), as well as the PAHO (Morel *et al.*, 2013). A coproantigen ELISA has been developed at Salford University, which uses polyclonal antibodies against *Echinococcus spp* to detect these coproantigens (Craig *et al.*, 1988, 1995; Allan *et al.*, 1992; Craig, 1997). Individual samples undergoing this test are usually interpreted in a dichotomous fashion by identifying an optical density threshold for positivity, three standard deviations higher than the mean of a known negative panel (the 'Gaussian approach') (Deplazes *et al.*, 1992; Allan *et al.*, 1992). However little attention to date has focussed on whether this is the optimal strategy for test interpretation.

3.1.2 Diagnostic testing

Diagnosis of infection or exposure to infectious agents is a fundamental concept in human and animal epidemiology, and ranges from individual-level diagnosis with a

view towards instigating appropriate treatment, through to surveillance at the population or regional level. ELISA tests are commonly used to detect antibodies or antigens in the sera or other compartments, and generally give results on a continuous scale (such as the optical density [OD] of a colour change reaction), before samples are classified as either 'negative' or 'positive' according to their OD in relation to a cut-off value. As an overlap in OD values of positive and negative samples is commonly observed, one major issue resulting from this dichotomised interpretation is that of misclassification of samples. These limitations are well known for the *Echinococcus* coproantigen ELISA test (Allan and Craig, 2006), and therefore care must be taken when interpreting coproantigen data, as the test prevalence is unlikely to represent the true prevalence. This can be a particular problem during eradication campaigns when the true prevalence is low, meaning that false positive results (in the case of an imperfect test specificity) can result in a low positive predictive value at the individual level (Christofi *et al.*, 2002; Eckert, 2003; Torgerson and Deplazes, 2009).

At the population level, the degree of misclassification associated with a diagnostic test can be quantified using estimates of the sensitivity and specificity of the method of test interpretation, which are the conditional probabilities of a positive result in a positive sample and a negative result in a negative sample, respectively (Altman and Bland, 1994a). These can then be taken to account in the final interpretation of the results if desired. The best known approach to adjustment of test results in order to account for test sensitivity and specificity is the Rogan Gladen estimator (Rogan and Gladen, 1978), which allows an estimate of the true prevalence (p) to be made, based upon the test prevalence (p'), sensitivity (Se) and specificity (Sp):

$$p = \frac{p' + Sp - 1}{Se + Sp - 1}$$

In order to account for random error resulting from sampling, confidence intervals can be estimated for p (Clopper and Pearson, 1934; Sterne, 1954; Blaker, 2000; Reiczigel *et al.*, 2010), and methods of incorporating random error in the estimates of Se and Sp

themselves (as described by (Cameron and Baldock, 1998)) have also been developed (Reiczigel *et al.*, 2010; Lang and Reiczigel, 2014).

More sophisticated approaches of diagnostic test interpretation based upon approaches such as latent class analysis have been developed, which do not necessarily require knowledge of test characteristics (Hui and Walter, 1980; Enøe *et al.*, 2000; Johnson *et al.*, 2001; Black and Craig, 2002; Toft *et al.*, 2005) and applied to *Echinococcus* data (Ziadinov *et al.*, 2008; Torgerson and Deplazes, 2009; Hartnack *et al.*, 2013). These methods allow an estimate of the true prevalence of infection to be made, as well as extracting information on test performance in the field. However, although these methods are powerful, they generally require that at least two tests have been conducted on a relatively large number of samples (and in the case of only one population being studied, three tests are required for identifiability (Hui and Walter, 1980; Johnson *et al.*, 2001)). It may not always be possible, due to limited resources, to conduct this many tests – especially during ongoing surveillance in the face of a control strategy.

The current report investigates three different general approaches to classify samples as positive or negative based on coproantigen ELISA (coproELISA) test data. Two of these methods: the Gaussian distribution method and ROC curves (or approaches based upon this principle) are already commonly used, whereas the other (mixture modelling) is a novel method of coproantigen data analysis.

3.1.3 Gaussian distribution cut-off

The method most commonly used for selection of an ELISA test cut-off is known as the 'Gaussian distribution method'. It is based upon the assumption that the distribution of OD values amongst negative samples is approximately Gaussian, and uses the properties of a Gaussian distribution in order to select a cut-off with a low probability of false positives (i.e. a high specificity). In order to calculate the cut-off, the mean and standard deviation of OD values for a panel of known negative samples are calculated, and the cut-off point is identified as the OD value three standard

deviations above the mean (Deplazes *et al.*, 1992; Allan *et al.*, 1992). This cut-off is usually determined during test optimisation (see below), and is then used for all subsequent tests without re-evaluation. According to the characteristics of a Gaussian distribution, only 0.1% of true negative samples would be expected to have an OD greater than or equal to this value, meaning that the test specificity would be expected to be 99.9% (although reported coproELISA test specificities are often lower than this, possibly due to cross-reactions with other cestodes (Deplazes *et al.*, 1992; Allan *et al.*, 1992; Allan and Craig, 2006)). It is unclear why three standard deviations are used rather than two standard deviations (as has been reported in some other ELISA studies – for example, (Richardson *et al.*, 1983)). As this approach does not explicitly account at all for the distribution of positive samples (and therefore does not account for test sensitivity), particular attention is paid to maximising the signal:noise ratio (i.e. the ratio of the OD of known positive samples to that of negative samples) during antibody screening and optimisation. If this ratio is high (at least 5.0), the difference between the OD values of positive samples and those of negative samples should be high, which would hopefully result in an adequate sensitivity.

3.1.4 ROC curves

Receiver operating characteristic (ROC) curves are a graphical method for the investigation of the effect of varying the cut-off point on two 'operating characteristics' of the test: the true positive proportion (TPP; i.e. the sensitivity) and the false positive proportion (FPP; i.e. one minus the specificity) (Zweig and Campbell, 1993; Greiner *et al.*, 2000). By plotting the FPP against the TPP for a variety of cut-offs, the optimal cut-off for the study in question can be determined. As for the Gaussian distribution method described above, this approach requires a panel of known negative samples, but additionally requires a panel of known positive samples. However, as the distributions of both negative and positive samples are explicitly accounted for, estimates of both sensitivity and specificity are obtained, and the cut-off can be adjusted in relation to these as required (for example, if a perfect specificity is required, the cut-off point for this can be estimated from a ROC curve). The area under a ROC curve can also provide useful information on the overall ability of the

test to discriminate between positive and negative samples, regardless of the cut-off chosen (Swets, 1988). Adjustments to ROC curve analysis have also been suggested, for example by allowing the determination of an 'intermediate' range of test results as well as positives and negatives (Greiner *et al.*, 1995). However, most commonly, a single cut-off point maximising both sensitivity and specificity (assuming equal weighting for both) is desired. This approach is the basis for the Youden index (Youden, 1950; Guezala *et al.*, 2009), which is calculated as the sum of the sensitivity and specificity at the cut-off point which maximises both of these, minus 1. The Youden index can therefore be used for both estimation of a cut-off point, and for comparison of the differentiating ability/performance of different tests (with values close to 1 indicating good differentiating ability, and those close to zero indicating poor differentiating ability).

3.1.5 Mixture models

As described above, the concept of the identification of component distributions within a group of biological samples was first introduced in 1894 (Pearson, 1894), in one of the first examples of the application of statistical principles to the analysis of biological data (McLachlan and Peel, 2000c). Mixture models (or similar approaches) have subsequently been frequently applied to the problem of diagnostic test interpretation (Rushforth *et al.*, 1971; Grannis and Lott, 1978; Parker *et al.*, 1990; Gay, 1996; Neuenschwander *et al.*, 2000; Baughman *et al.*, 2006; Vyse *et al.*, 2006; Hardelid *et al.*, 2008), where they have potential use as a method of classification in the absence of a gold standard test. Due to the logistical and practical difficulties associated with the identification of known infected and uninfected dogs in the field, mixture models were investigated here as a potential approach to coproELISA classification.

Finite mixture models (FMMs) are a form of cluster analysis method whereby a finite number of subpopulations ('components') can be identified within a population based on the distribution of the data rather than through association with external variables (meaning that they can also be described as a type of 'person-centred' rather than 'variable-centred' analysis tool (Muthén and Muthén, 2000; Jung and Wickrama,

2008)). FMMs can also be described as form of 'model-based clustering', which groups individuals based upon explicit assumptions regarding the distributional qualities of the components – most commonly, that these follow a Gaussian distribution (in the case of Gaussian Finite Mixture Models, as are used in the current report). This gives the model a clear statistical foundation, as well as potentially having some biological basis. The output of an FMM includes a description of the parameters of the component distributions, along with the *a priori* probability of membership of each component (that is, the 'relative size' of each component). From this, estimates of the posterior probability of component membership for individual samples can be made, and if desired, these can be allocated to particular components according to modal probability. Despite being first proposed over 100 years ago (Newcomb, 1886; Pearson, 1894), it is only in recent years that computational advances such as the expectation-maximisation algorithm (Dempster *et al.*, 1977; McLachlan and Peel, 2000a) and Markov Chain Monte Carlo (MCMC) methods (Hastings, 1970; McLachlan and Peel, 2000b) have allowed reasonable model fitting (Aitkin and Rubin, 1985; McLachlan and Peel, 2000c).

The statistical background to mixture models has been reviewed elsewhere (McLachlan and Peel, 2000c), and will be only briefly introduced here. As they are a model-based approach, a statistical model can be explicitly defined. For a simple univariate Gaussian FMM, Y represents a vector of length n, relating to a random sample of n individuals from a population $\left(Y_j;\ j = 1{:}n\right)$. The probability density function of y_j, $f\left(y_j\right)$, can be presented as the sum of g components, each of which has its own mixing proportion (or weight), π_i: each of which lie between zero and one, and sum to one. Each component is distributed with its own normal distribution, $f_i\left(y_j\right) \sim N(\mu_i, \sigma_i^2)$; $(i = 1{:}\,g)$. This can be presented as follows:

$$f\left(y_j\right) = \sum_{i=1}^{g} \pi_i f_i(y_j)$$

As such, in order for the model to be created, the number of mixture components, g, must be specified. This is one of the main difficulties encountered when constructing

a mixture model, as in most cases, it is unknown - leading to a roundabout problem of model assessment prior to model creation. Possible options for achieving this have been recently reviewed (Oliveira-Brochado and Martins, 2005), and will not be fully described here. A common method of comparing different numbers of groups in FMMs is by creating models with different numbers of components and comparing these using complexity-penalised information criteria such as Akaike's Information Criterion (AIC) (Akaike, 1973) or the Bayesian Information Criterion (Schwarz, 1978). Traditional hypothesis tests of the effect of adding an extra component to the model, such as the likelihood ratio test, are complicated by the fact that models with different numbers of components are not nested within one another (Aitkin and Rubin, 1985). This problem can be circumvented using bootstrapping approaches (McLachlan, 1987; Efron and Tibshirani, 1993). A bootstrap sample is taken from the data under the "null hypothesis" of g components in the model, and the likelihood estimated. This is repeated for the "alternative hypothesis" of $(g+1)$ components, and the likelihood ratio of the null and alternative hypotheses (λ) is estimated. From this, $-2\ln(\lambda)$ can be estimated (as is usually used in the likelihood ratio test). This process is then repeated multiple times, allowing the full distribution of $-2\ln(\lambda)$ to be estimated. Evidence against the null hypothesis can therefore be obtained if the likelihood ratio statistic obtained from the data differs from that predicted from these replications, as is the case with any null hypothesis test (Hope, 1968; Aitkin *et al.*, 1981; McLachlan, 1987). Given that the estimated p-value is below the significance threshold, this process is then repeated with the null hypothesis of $(g+1)$ components, and an alternative hypothesis of $(g+2)$ components, and continues until there is no evidence against the null hypothesis.

Assuming that the number of components is known, the remaining issue, as alluded to earlier, is fitting of the model. The likelihood of the model with the distributional parameters $(\mu_i, \sigma_i^2) = \theta_i$ is as follows:

$$\mathrm{L}(\theta_1 : \theta_g \,;\, \pi_1 : \pi_g \,|Y) = \prod_{j=1}^{n} \sum_{i=1}^{g} \pi_i f_i(y_j|\theta_i)$$

Where y_j indicates an individual observation. Maximising this likelihood in order to estimate the parameters θ_i and π_i can be facilitated through the use of the expectation-maximisation (EM) algorithm, which was developed by Dempster *et al* and has been described elsewhere (Dempster *et al.*, 1977; Jeff Wu, 1983; McLachlan and Peel, 2000a; Fraley and Raftery, 2002). Basically, the EM algorithm in the context of FMMs assumes that along with the dataset Y, there are missing/unobserved variables relating to component membership, which need to be taken into account when maximising the likelihood. This can be presented as each observation in the 'complete data', x_j , being comprised of the individual observations (y_j) and the n unobserved variables associated with these (z_j) which relate to component membership. Each z_j is a vector of length g ($z_j = (z_{j1}: z_{jg})$), where $z_{ji} = 1$ if y_j is in component i, and $z_{ji} = 0$ otherwise. The algorithm itself is an iterative procedure consisting of an 'expectation step', where z_{ji} is estimated, based upon Y and current estimates of θ_i and π_i; followed by a 'maximisation step', whereby z_{ji} is assumed fixed and the log-likelihood of θ_i and π_i are maximised, conditional on Y.

3.1.6 Aims and objectives

The aim of the current study is to evaluate the current approaches used for the dichotomous classification of dog faecal coproantigen data (Gaussian cut-off and ROC curves/Youden index), and to investigate the use of alternative methods of achieving this using mixture models. Finally, the effect of the different classification systems on both the coproantigen prevalence and the estimated true prevalence is investigated.

3.2 Materials and Methods

3.2.1 Samples

Samples were collected from a control scheme evaluation in northern Xinjiang province in the People's Republic of China, as described in the previous chapter and in recent reports (van Kesteren *et al.*, 2015). Both necropsy (n=38) and field samples

(n=125) were used in the current analysis. One 'field panel' sample, from village 'N', was removed from further study due to failure of replication.

3.2.2 Cut-off determination

The necropsy panel was used to estimate cut-off values for positivity using the three broad methods described earlier. All cut-offs were interpreted as being the threshold for positivity – meaning that any samples with the same OD as the cut-off were interpreted as being positive. Firstly, the standard protocol based on selecting a point three standard deviations above the mean of a selection of validation samples from a nonendemic area was used ('predetermined Gaussian' method). This same approach was then repeated using samples collected from the field in Xinjiang, from dogs which were negative on necropsy ('Gaussian 1' method). As the distribution of the OD values of these samples did not conform to the Gaussian distribution expected for this approach, outliers (defined as those points which were greater than 1.5 times the interquartile range above or below the upper or lower quartile, respectively) were identified using the 'boxplot' command in R, and were removed, before repeating the process ('Gaussian 2' method). Secondly, ROC curve analysis was used in order to select the cut-off point which maximises both the sensitivity (*Se*) and specificity (*Sp*) of the test simultaneously, using the R package 'ROCR' (Sing *et al.*, 2005) (although the Youden index was not calculated here, this method would be expected to give the same cut-off value).

Finally, a Gaussian finite mixture model was created using the R package 'mixtools' (Benaglia *et al.*, 2009). Selection of the appropriate number of components was achieved using an iterative bootstrap analysis of the effect of adding an extra class to the model (described above), using a p-value of 0.05 or less to suggest an improved fit. Three methods were then used to allocate samples to a 'positive' or 'negative' status. The first of these was based on characterisation of the 'negative' component, followed by selection of a cut-off point three standard deviations above the mean of this distribution in a similar manner to the Gaussian distribution method described earlier ('MM1' method). In the case of a solution including more than two components,

different cut-offs were estimated according to the distribution of the data. The second mixture model method was based on allocation of individual samples to either a 'negative' or 'positive' group according to the modal posterior probability of component membership ('MM2' method). The final mixture model approach ('MM3') was based upon a ROC curve analysis approach to the mixture model output. 'Positive' and 'negative' components were identified (in the case of models with more than two components, intermediate groups were ignored), and the posterior probabilities of membership in each of these components for each individual sample were estimated. Samples were then ordered according to OD, and a range of different OD cut-off points were applied to the data. For each cut-off point, the sum of all the probabilities of 'negative' component membership for samples classified as negative, and the sum of all the probabilities of 'positive' component membership for samples classified as positive were estimated, and these were expressed as a proportion of the total sum of all probabilities within the group in question (in order to ensure equal weighting of negative and positive groups in unbalanced studies). These estimates were then summed, and the cut-off which maximised this total was selected.

In order to assess the potential use of mixture models in the absence of a gold standard test, Gaussian mixture models were also created using the field data, and cut-off points were estimated as described for the MM1 and MM3 methods (allocation according to modal probability was not performed as it does not result in a cut-off, and so could not be validated). The Se, Sp, and overall accuracy (proportion of samples correctly classified) of each method was estimated using the necropsy panel results. Finally, the effect of a selection of these different methods on the coproantigen prevalence estimate from the field data (stratified by village, due to the stratified sampling approach used) was also estimated. The Rogan-Gladen approach (Rogan and Gladen, 1978) described above was used to give a point estimate of the true prevalence of infection, and exact Blaker confidence intervals (accounting for test sensitivity and specificity) were calculated (Blaker, 2000; Reiczigel *et al.*, 2010), using the 'epi.prev' command in the 'epiR' package for R (Stevenson *et al.*, 2013).

3.3 Results

The 'Gaussian 1' method, based on a panel of faeces from 43 dogs from nonendemic areas, estimated a cut-off OD of 0.065. A total of 38 faecal samples were collected by necropsy in Xinjiang, of which 16 (42%) were found to contain *Echinococcus* spp. The number of *Echinococcus* worms present amongst infected dogs ranged from 2 to over 10,000, with a median of 100, as shown in Table 3.1. A total of 22 dogs were negative on necropsy, and were used in the 'Gaussian 2' approach to give a cut-off OD of 0.331. After the removal of three high OD outliers, the 'Gaussian 3' approach gave a cut-off of 0.180. ROC curve analysis including all 38 samples suggested that a cut-off of 0.117 maximised the overall accuracy of test classification.

Application of mixture models to the necropsy data identified two components, as detailed in Table 3.2. Based upon the distribution of the 'negative' component, a cut-off of 0.149 was estimated with the 'MM1' approach. When the adjusted ROC curve approach was applied to the posterior probabilities of sample component membership (the 'MM3' approach), the optimal cut-off was found to be 0.117.

Table 3.1. Distribution of worm burdens and coproantigen ELISA OD values amongst the 16 *Echinococcus* spp positive dogs identified by necropsy.

OD	Number of *Echinococcus* worms	Number of *Taenia* spp
0.117	2	1
0.088	3	7
0.155	10	1
0.176	20	2
0.252	50	2
0.171	50	1
0.087	50	Not recorded
0.461	100	0
0.240	100	0
0.373	100	7
0.396	300	4
0.462	500	5
0.571	>5,000	6
0.793	>10,000	0
0.680	>10,000	6
0.665	>10,000	2

Table 3.2 Properties of components identified by the mixture models. Components have been ordered according to their means (smallest to largest).

Dataset	Mixture Model component	Proportion	Mean OD in component	OD standard deviation
Necropsy	**'negative'**	0.45	0.075	0.025
	'positive'	0.55	0.335	0.198
Living dogs	**'negative'**	0.81	0.091	0.041
	'positive'	0.19	0.339	0.132
	'negative'	0.22	0.044	0.012
	'intermediate'	0.59	0.108	0.034
	'high'	0.19	0.336	0.132

When a mixture model was applied to the field data (collected from living dogs), the optimal number of components was found to be three, with an 'intermediate' component between the negative and positive ones identified in the necropsy data (see Table 3.2). As this was unexpected, a mixture model was first created with only two components, as had been used for the necropsy data ('MM1a'), which gave a cut-off of 0.215. As the status of dogs in the intermediate component of the three component model was not clear, cut-off methods were applied including it as both a negative ('MM1b') and as a positive ('MM1c') group. These gave cut-off points of 0.079 and 0.209, respectively. Finally, the adjusted ROC curve approach ('MM3') applied to these models, which gave an optimal cut-off of 0.200 for the two component model and 0.180 for the three component model.

Figures 3.1 and 3.2 show the overall distribution of OD values for all samples (necropsy dogs and field dogs), both in the form of a histogram and a kernel density plot (created using the 'density' command in R). Cut-off estimates are overlaid (some of the less reliable and more problematic cut-offs have been excluded, for ease of interpretation).

Table 3.3 shows the estimated sensitivities and specificities of the different classification methods. The average of the sensitivity and specificity is also presented (rather than the overall classification accuracy of the test), in order to ensure that negative and positive samples have equal weighting. Finally, Table 3.4 and figure 3.3 show the effect of some of the different cut-offs on the point estimates of the coproantigen prevalence and the Rogan-Gladen 'true' prevalence (and exact 95% confidence interval) for the six villages visited. The 'Gaussian 1' and three-component mixture model approaches were not evaluated here, due to suspected limitations in their applicability. The 'predetermined Gaussian' approach was primarily included for comparison purposes rather than due to its suspected validity, since this is the current approach used for estimation of a cut-off point.

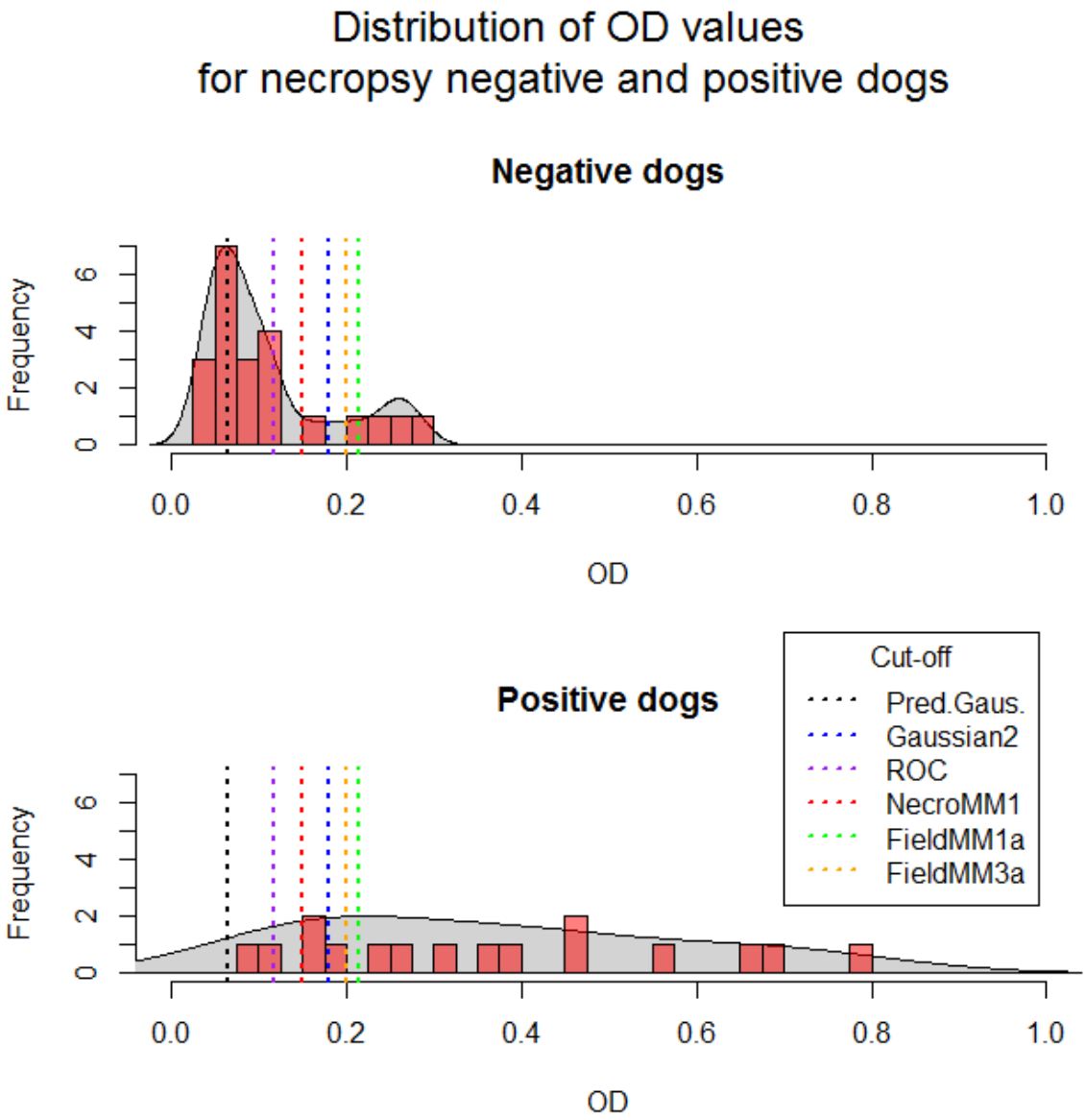


Figure 3.1. Distribution of OD values for all necropsied dogs. The top graph shows results for those with no *Echinococcus* spp. adult worms on intestinal inspection, and the graph below shows the results for all those with at least one adult worm.

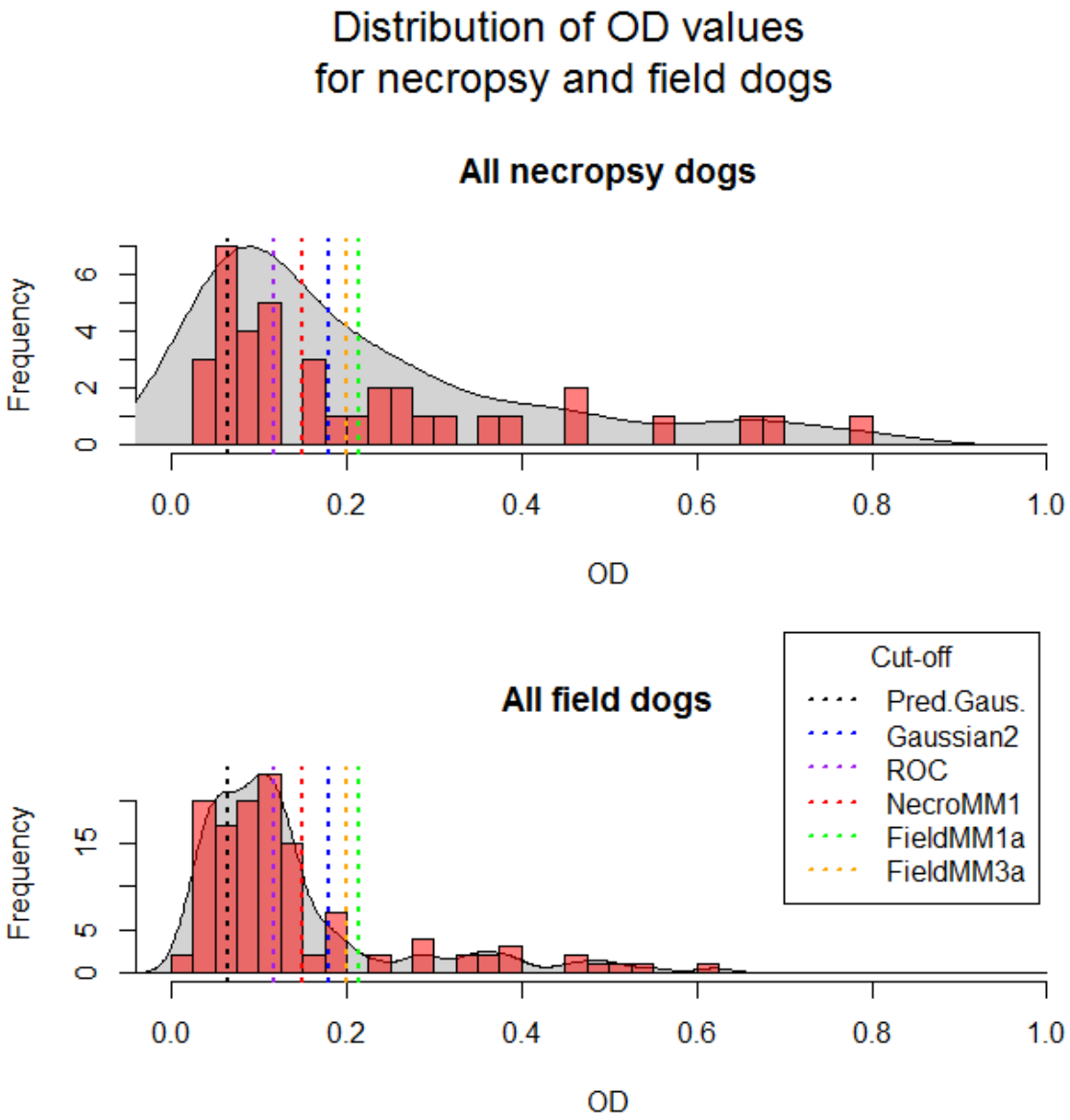


Figure 3.2. Distribution of OD values for all necropsied dogs (top) and all live dogs sampled in the field (bottom).

Table 3.3. Test characteristics using different methods of classification (sensitivity, specificity and overall accuracy estimated from necropsy data). The 'predetermined Gaussian' method is a Gaussian method using nonendemic controls; and the 'Gaussian 1' and 'Gaussian 2' methods use necropsy data (with the latter excluding outliers). The 'MM1' approach estimates a Gaussian cut-off from the 'negative' component of a mixture; the 'MM2' approach allocates individuals to mixture model components according to modal probability; and the 'MM3' approach uses a modified ROC curve-type approach to identify a cut-off which maximises the probability of membership in the 'negative' and 'positive' mixture model components.

Method		Cut-off	Sensitivity (n=16 positives)	Specificity (n=22 negatives)	Accuracy
	Predetermined Gaussian *(3sd above mean of nonendemic panel tested separately)*	0.065	16/16 = 100%	8/22 = 36%	68%
Estimated from necropsy panel	**Gaussian 1** *(3sd above mean of necropsy negative panel from field)*	0.331	8/16 = 50%	22/22 = 100%	75%
	Gaussian 2 *(as for Gaussian 1 but 'outliers' removed)*	0.180	11/16 = 69%	19/22 = 86%	78%
	ROC curve *(maximising Se + Sp; based on necropsy data from field)*	0.117	15/16 = 94%	17/22 = 77%	86%
	Necropsy MM1 *(3sd above mean of 'negative' component)*	0.149	14/16 = 88%	17/22 = 77%	82%
	Necropsy MM2 *(component allocation according to modal posterior probability)*	N/A	14/16 = 88%	17/22 = 77%	82%
	Necropsy MM3 *(ROC curve evaluation of posterior probabilities)*	0.117	15/16 = 94%	17/22 = 77%	86%
Estimated from living dogs	**Field data MM1a** *(two component model as for MM1)*	0.215	11/16 = 69%	19/22 = 86%	78%
	Field data MM1b *(three component model; as for MM1 with intermediate group classified as negative)*	0.079	16/16 = 100%	11/22 = 50%	75%
	Field data MM1c *(three component model; as for MM1 with intermediate group classified as positive)*	0.209	11/16 = 69%	19/22 = 86%	78%
	Field data MM3a *(ROC curve evaluation of two component model)*	0.200	11/16 = 69%	18/22 = 82%	75%
	Field data MM3b *(ROC curve evaluation of three component model)*	0.180	11/16 = 69%	18/22 = 82%	75%

Table 3.4. Effect of different cut-offs on point coproantigen prevalence (upper percentages) and estimated true prevalence (according to the Rogan-Gladen method) and exact 95% confidence intervals (using the Blaker method), whilst accounting for test sensitivity and specificity (using the Reiczel method) (lower percentages) for six villages in Xinjiang.

Method used (cut-off)	Village					
	A	**B**	**C**	**N**	**Q**	**T**
Predetermined Gaussian (0.065)	13/19 = **68%** **13%** (0 – 60%)	18/20 = **90%** **73%** (13 – 95%)	13/21 = **62%** **0%** (0 – 46%)	12/20 = **60%** **0%** (0 – 43%)	19/26 = **73%** **26%** (0 – 66%)	11/19 = **58%** **0%** (0 – 39%)
Gaussian 2 (0.180)	1/19 = **5%** **0%** (0 – 21%)	8/20 = **40%** **48%** (13 – 90%)	2/21 = **10%** **0%** (0 – 30%)	1/20 = **5%** **0%** (0 – 19%)	5/26 = **19%** **10%** (0 – 44%)	2/19 = **11%** **0%** (0 – 33%)
ROC /Necropsy MM3 (0.117)	6/19 = **32%** **12%** (0 – 46%)	14/20 = **70%** **67%** (35 – 89%)	8/21 = **38%** **22%** (0 – 52%)	3/19 = **15%** **0%** (0 – 20%)	12/26 = **46%** **33%** (8 – 61%)	8/19 = **42%** **27%** (0 – 60%)
Necropsy MM1 (0.149)	3/19 = **16%** **0%** (0 – 25%)	10/20 = **50%** **42%** (10 – 74%)	3/21 = **14%** **0%** (0 – 19%)	1/20 = **5%** **0%** (0 – 2%)	6/26 = **23%** **1%** (0 – 30%)	5/19 = **26%** **6%** (0 – 42%)
Field data MM 1a (0.215)	3/19 = **16%** **4%** (0 – 46%)	10/20 = **50%** **66%** (28 – 100%)	3/21 = **14%** **1%** (0 – 39%)	1/20 = **5%** **0%** (0 – 19%)	6/26 = **23%** **17%** (0 – 52%)	5/19 = **26%** **23%** (0 – 66%)
Field data MM3a (0.200)	2/19 = **11%** **0%** (0 – 26%)	8/20 = **63%** **43%** (5 – 89%)	2/21 = **10%** **0%** (0 – 24%)	1/20 = **5%** **0%** (0 – 11%)	5/26 = **19%** **2%** (0 – 39%)	2/19 = **11%** **0%** (0 – 26%)

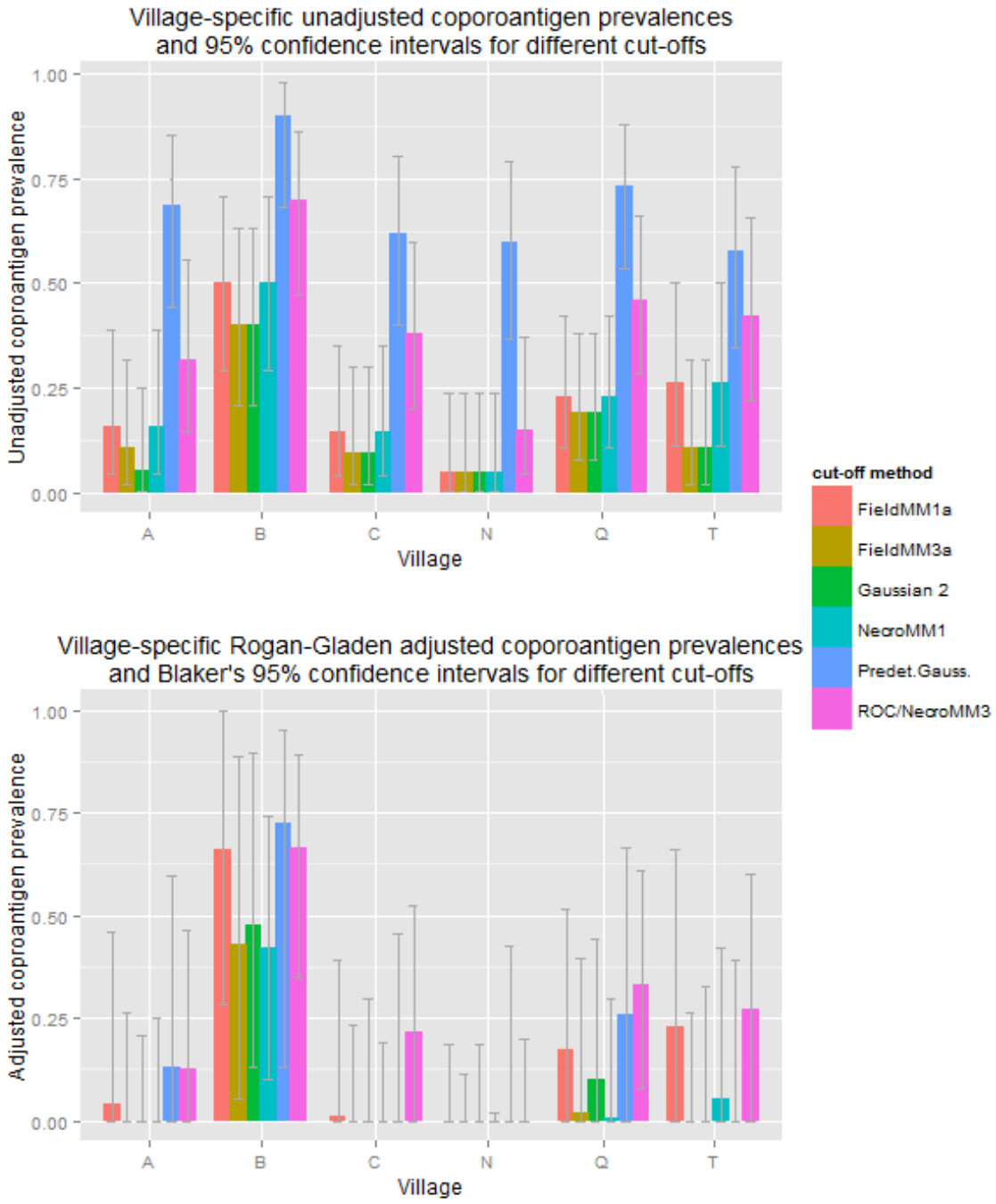


Figure 3.3. Unadjusted (top) and adjusted (bottom) estimates of the coproantigen prevalence for the different villages, using different cut-offs. Bars indicate 95% confidence intervals.

3.4 Discussion

The current paper details an attempt to improve the method of classification of canine faecal samples when using an *Echinococcus* coproantigen ELISA test. As well as reviewing two general approaches in common use currently, a number of novel methods based on mixture modelling are described. It is hoped that these methodologies will offer some prospect for improvements in canine echinococcosis surveillance, by both assisting in the selection of an appropriate approach to diagnostic test interpretation, and by illuminating some of the limitations associated with dichotomous interpretation of any data on a continuous scale, such as ELISA OD data.

3.4.1 Gaussian approaches

Based on the results shown here, the 'predetermined Gaussian' approach to cut-off determination, based upon selection of a cut-off three standard deviation above the mean OD of samples taken from a nonendemic area, does not perform well as a diagnostic test. The estimated cut-off using this strategy was low, and therefore the test specificity was also low. Although the sensitivity was 100% in this particular case, this strategy does not implicitly incorporate positive samples in its calculation, and therefore this could be an incidental finding. The low specificity resulted from the nonendemic negative panel generally having lower OD values than negative samples from the necropsied dogs. This may have resulted from variations in ELISA conditions (as the nonendemic samples were not tested at the same time as the others presented here, as per the usual protocol for the predetermined Gaussian approach), or may suggest that the variation and mean of these samples was lower than that observed in the field data (see Table 3.5). It is possible that dogs from nonendemic communities differ in various ways from those in endemic communities, and as such may not present an optimal panel for selection of a cut-off point to apply to field data. One possible way in which these dogs may differ is in terms of concurrent worm burdens – in particular, other cestodes such as *Taenia* spp. Although preliminary analysis of the necropsy data used here gave no evidence of an association between *Taenia* spp

burdens and OD values (following adjustment for the effect of *Echinococcus* spp burden – data not shown), previous studies have suggested that this may exist at a low level (Deplazes *et al.*, 1992; Allan *et al.*, 1992; Allan and Craig, 2006). The presence of this cross-reactivity could possibly result in greater variation in OD values amongst *Echinococcus* negative animals from endemic communities compared to those from nonendemic communities. One other possibility is that of misclassification of the necropsy panel (meaning that necropsy negative samples may not have been true negatives) – which, if true, would be expected to improve the estimated specificity of the predetermined Gaussian method. This issue is dealt with below.

The 'Gaussian 1' approach suffered the opposite problem to the predetermined Gaussian method – giving a high cut-off, and therefore a low sensitivity. The cause of this was the presence of a number of high-OD outliers amongst the otherwise relatively Gaussian-distributed negative samples from the necropsy panel, which increased both the standard deviation and estimated mean of these OD values. As the presence of these outliers violated the basic assumption of the Gaussian approach, the three most extreme values were removed (although a further two remained outside the expected Gaussian distribution). This demonstrates the dangers of not visually inspecting the distribution of data before applying a technique such as this. The reason for the outliers is unclear, but as they are rarely seen when using a panel of negative dogs from a nonendemic area (author's personal observation), they may represent dogs with low burdens which have been overlooked during necropsy. Alternatively, they may be taken from dogs which have been recently dosed with a cestocidal drug: as seen with other cestodes (Deplazes *et al.*, 1990; Allan *et al.*, 1990), it has been found that it can take 2-4 days for *E. granulosus* coproantigens to disappear following treatment (Jenkins *et al.*, 2000). This demonstrates a potential limitation with the use of field data as a negative control panel (see Table 3.5). Another possibility is that the distribution of negative samples truly does not follow a Gaussian distribution, as has been suggested from a study of coproantigen ELISA OD densities for fox faeces in France (where samples were diagnosed using the 'gold standard' of necropsy and sedimentation and counting technique (WHO/OIE, 2001d; Eckert,

2003)) (Raoul *et al.*, 2001). If this was the case, any approach based upon the Gaussian method (as well as the Gaussian mixture model) will be flawed.

As well as the specific issues described above, one clear limitation of any Gaussian cut-off method is that the distribution of 'positive' OD values is not accounted for at all. As such, selection of a cut-off aims to maximise the test specificity without any consideration of the impact this has on the sensitivity. An ideal coproantigen ELISA test would have complete separation between negative and positive OD values, and so this would not be an issue (indeed, this is often seen when evaluating the test using nonendemic negative samples and high burden or spiked positive samples). However, this is not likely to be the case in the field situation, in the presence of negative dogs with high concurrent worm burdens (and therefore possible cross reactions, as described above), and positive dogs with low worm burdens (as described below) or prepatent infections.

3.4.2 ROC curves

Traditional ROC curve analysis of the necropsy data appeared to give the best results of all of the methods assessed here, with a very high sensitivity and a good specificity (although definitive conclusions are difficult to make based on such a small sample size). ROC curve analysis is also the only method detailed here which specifically allows the determination of a cut-off point according to the requirements of the test (Greiner *et al.*, 2000), and which does not make any assumptions about the frequency distribution of OD values amongst negative or positive samples. Although the current approach has aimed to maximise both the sensitivity and specificity, in some cases (such as monitoring for introduced infection in a nonendemic community, where any possible positive dogs need to be identified quickly), it may be of greater use to select a cut-off giving a maximal sensitivity, even if this results in a reduced specificity. Additionally, as the sensitivity and specificity are explicitly estimated as part of the ROC curve estimation approach, a greater appreciation may be gained of the limitations in test interpretation.

Despite these positive aspects, it should be noted that the cut-off determined in the nonparametric ROC curve analysis used here can only take the value of one of the OD values in the panel investigated – which can lead to some loss of accuracy in the overlapping area between negative samples and positive samples (which is the area of interest), especially when sample sizes are relatively small (as was the case in the current study). For example, both ROC curve-based approaches towards classification of necropsy data identified the cut-off for positivity as 0.117 (to 3 significant figures), which relates to one particular sample (which had an OD of 0.11685). Therefore, all values with an OD of 0.11685 or more were classified as positive. The sample with the next highest OD value to this one had an OD of 0.11585 – meaning that it could equally be stated that the cut-off for *negativity* was 0.116 (i.e. all samples with an OD of 0.11585 or less were classified as negative). Whilst both of these methods give the same result when applied to the necropsy data here, the use of these slightly different cut-off interpretations to field data could give different results if intermediate OD values between 0.116 and 0.117 were present. An alternative to this would be to use parametric ROC curves, which assume that both negative and positive samples follow Gaussian distributions. This approach may not be appropriate with the relatively small sample sizes in this case, and may be problematic if the distribution of OD values amongst infected individuals (see below) or uninfected individuals (see above) did not follow a Gaussian distribution. Another approach is 'two-graph ROC' (TG-ROC) analysis, which can allow the estimation of an 'intermediate range' (Greiner *et al.*, 1995).

3.4.3 Selection of negative and positive panels

The two classification methods described above (Gaussian method and ROC curves) require the use of a panel of samples of known status (either a negative panel, in the case of the Gaussian approach, or both a negative and a positive panel, in the case of the ROC curve analysis). The panel is commonly either taken from necropsy samples from the area in question (purge samples are of limited use, since false negatives are relatively common (Schantz, 1997; Lahmar *et al.*, 2007b; Ziadinov *et al.*, 2008; Hartnack *et al.*, 2013)), or from either faeces collected from dogs in a known nonendemic area (negative panel) or faeces confirmed by necropsy/purgation to be

positive (positive panel). Selection of this panel can be problematic for a number of reasons, as detailed in Table 3.5. Bayesian approaches to ROC curve analysis which do not require known samples have been developed (Branscum *et al.*, 2008; Jafarzadeh *et al.*, 2010), and would be expected to resolve this problem to some degree.

Table 3.5. Advantages and disadvantages of different positive/negative panels for determination of cut-off points

Panel used	Advantages	Disadvantages
Samples from study area (confirmed by necropsy or purgation)	More likely to relate to epidemiological situation in the field. Positive results provide conclusive evidence of presence of infection (and good quality material for PCR analysis).	Logistically difficult to carry out. Necropsy panels likely to over represent unwanted/stray dogs (which may less represent the dogs of interest in the community). May include false negative dogs (especially those with low worm burdens). Less useful when prevalence of infection very low, as may not obtain many positive dogs. Samples collected will invariably be per rectum, which are likely to be of higher quality than ground samples (so there may be differences in OD values).
Negative panel from nonendemic area ***or*** **Positive panel from various necropsy/purge campaigns**	Logistically easier to use (do not need to be sampled for each field site). Can be used regardless of the echinococcosis situation in the field site.	May not represent the situation in the field, as dogs may differ in various ways from those in the areas under investigation, resulting in an inappropriate cut-off estimate (e.g. possible reduction in mean burden amongst positives in face of control scheme, or high OD values in negatives due to concurrent *Taenia* spp infection).
Samples from study area classified by ELISA (usually along with PCR confirmation)	Readily available in most cases.	No conclusive evidence of presence of infection (possible false positive PCR results). Methodologically questionable to use ELISA test as validation for itself.

3.4.4 Mixture modelling

The mixture model approach described here was an attempt to circumvent the problem of negative/positive panel selection by identifying the infection status from coproantigen field data alone. It appears, by relation back to the necropsy panel, that this approach gave reasonable estimates of sensitivity and specificity, although difficulties in objective interpretation arose when more than two components were identified in the mixture model. The 'ROC curve' approach to interpretation of mixture model results gave reasonable estimates of sensitivity and specificity compared to the other approaches, and did not require interpretation of the intermediate component in these cases. Additionally, this approach could also be adjusted in order to select cut-offs which maximised either the 'sensitivity' (sum of 'positive' component probabilities) or the 'specificity' (sum of 'negative' component probabilities), as required. However, this method may experience problems if the intermediate component is very large, as estimates will then be based on fewer samples. When a mixture model is applied to nonendemic negative samples only, only one component is identified (data not shown), suggesting that mixture modelling would also be appropriate for identification of truly uninfected communities.

Despite the potential for the use of mixture modelling in the classification of field data, there are considerable limitations. Firstly, the mixture model assumes that each of the components follow a Gaussian distribution. One feature of the Gaussian distribution is that it is unbounded, and therefore can take any value between $-\infty$ and $+\infty$. Since coproantigen OD data is expressed as the difference in OD from a "blank" well, it would be expected to invariably be positive. This could cause possible problems for the parameterisation of the "negative" component of a mixture model, as the mixture model could predict support for negative numbers. This could be a particular problem if low OD positive samples were included in the negative component, as this would widen the variance of the component and could therefore lead to support for negative OD values. Despite this, the Gaussian distribution was retained in the current example due to its ease of specification, and the fact that the

mean and variance estimates could be specified independently of each other. A truncated Gaussian distribution could alternatively be used.

Whilst it could be reasonably be expected that the OD values of negative samples would be distributed according to a Gaussian distribution (although even this may not be the case (Raoul *et al.*, 2001)), it is unlikely that the same is true for the positive samples. It is well-recognised that infections with *Echinococcus* spp, as with most parasitic infections and infestations, follow a highly aggregated distribution, whereby most infected hosts carry very few parasites (Crofton, 1971a; Anderson and May, 1978; Anderson and Gordon, 1982; Gemmell *et al.*, 1986c). It has been reported by numerous authors that there is a broad linear correlation between OD values and worm burdens when worm burdens are high (Deplazes *et al.*, 1992; Craig *et al.*, 1995; Raoul *et al.*, 2001; Reiterová *et al.*, 2005; Buishi *et al.*, 2005b), and inspection of the current data suggests that a linear relationship exists between the natural logarithm of the burden and the OD value, even at lower worm burdens (see chapter 4). Therefore, some degree of overdispersion in the distribution of OD values amongst positive individuals (which would be seen as a "right skew" in the distribution) would be expected. Therefore, it is not reasonable to suspect the distribution of OD values amongst infected dogs will follow a Gaussian distribution. Despite this, Gaussian mixture models are relatively flexible to these distributional issues, as many overdispersed distributions can be recreated using a mixture of Gaussians (Priebe, 1994) (although in order to effectively capture this, a large sample size may be required, which may not be available from a field survey). The converse potential problem with mixture modelling where the number of components is extracted from the data is the risk of overfitting, which could result in erroneous conclusions relating to the underlying components being drawn (Lin *et al.*, 2007). One solution to this issue would be to explicitly model the overdispersion in the positive samples (whilst attempting to ensure that the negative samples still follow a normal distribution, as expected). One potential method of achieving this would be to use the Skew Normal distribution (which includes the Gaussian distribution as a special case) in the mixture model (Azzalini, 1985; Lin *et al.*, 2007), or by modelling the positive component of the mixture model using

"nonparametric distributions" such as Polya trees (Ferguson, 1974; Hanson, 2006). This latter strategy is explored further in the next chapter.

3.4.5 Application to field data

There were considerable variations in the estimated canid coproantigen prevalence between the different villages sampled, and between different strategies of test interpretation. As mentioned earlier, the small sample sizes selected were chosen in an attempt to evaluate the potential use of LQAS methodology in the rapid appraisal of infection status in communities, and were not originally intended to be used for coproantigen prevalence estimation. As a result, the expected confidence intervals were very large, meaning that large apparent differences in point prevalence estimates are not necessarily statistically significant. However, one other possible reason for this variability in coproantigen prevalences with different cut-offs is the fact that the different cut-offs would be expected to result in different test sensitivities and specificities. Following adjustment for sensitivity and specificity, the results show less variation than the 'raw' coproantigen prevalence, although the wide confidence intervals make clear discrimination of prevalence estimates between villages rare. Most point estimates could be broadly subjectively classified as 'high' (village B), 'medium' (villages Q and T) or 'low' (villages C and N). Village A appeared to lie somewhere between the medium and low estimates. This general strategy may ultimately allow categorisation at the village level using broad bands of coproantigen prevalence, as is commonly used in the investigation of schistosomiasis (Montresor *et al.*, 1998; Mitchell and Pagano, 2012). An alternative approach would be to identify those villages which are likely to have some degree of infection (either following the protocol described in (Cameron and Baldock, 1998), or determined as a 95% confidence interval which does not include zero. In this case, there is only clear evidence for infection in village Q. Although adjustment of the coproantigen results according to point estimates of sensitivity and specificity which are based on a sample of only 38 samples is not ideal, this approach was considered to be the best approach for the available data, and at the very least go some way towards demonstrating the

potential dangers inherent in presenting point estimates of coproantigen prevalence without adjustment for test sensitivity and specificity.

3.4.6 Limits of detection

One of the primary challenges in the interpretation of this evaluation of coproantigen ELISA data is that although necropsy is considered a 'gold standard' test (WHO/OIE, 2001d; Torgerson and Deplazes, 2009), it is unlikely to have a perfect sensitivity, especially in the case of visual inspection (as was conducted here) rather than the use of the sedimentation and counting technique (Deplazes *et al.*, 1992; Hofer *et al.*, 2000; Allan and Craig, 2006). Indeed, one study comparing visual inspection and intestinal scraping with the sedimentation technique suggested a sensitivity of 78% - largely due to low burdens (less than ten worms) or prepatent infections (Hofer *et al.*, 2000). As dogs with low worm burdens would be expected to comprise a large proportion of the infected population in overdispersed infections such as echinococcosis, this could result in a considerable reduction in any estimate of infection prevalence based on visual inspection at necropsy. Although each of these individuals may have a relatively small impact on transmission, together they could be of importance, and so identification of them (at least in the case of prevalence estimation in the face of an intervention campaign) would be useful. From the perspective of coproantigen testing, this problem has been addressed by assuming a lower limit of detection of around 50 worms (Allan and Craig, 2006). Indeed, the five infected dogs which were not correctly identified by the normal Gaussian method (with outliers removed) or the three component mixture model with the intermediate group classified as negative, all had worm burdens of 50 or less. When an attempt is made to capture these dogs using a lower cut-off, the confluence of these low OD positive dogs and higher OD negative dogs may result in instability in the estimation of coproantigen prevalence – with small changes in the cut-off leading to large changes in the estimate. As described above, this effect may be responsible for some of the variation in coproantigen prevalence estimates from the field data. Further work is required to characterise those animals with OD values in this 'grey area', with a view towards an improved method of test interpretation which is stable to small changes in cut-off.

3.4.7 Coinfections

Finally, although *E. multilocularis* has been reported in Xinjiang, it is thought to exist at a low prevalence in the area investigated in the current study (adjacent to the Junggar basin), with highest endemicity found in the Altai, western Junggar, and Tianshan mountain ranges (Zhou *et al.*, 2000). As such, most of the infections described here are expected to be *E. granulosus* (sensu lato). However, in some of the areas in the world worst affected by echinococcosis, such as the Tibetan plateau and Kyrgyzstan, both *E. granulosus* and *E. multilocularis* coexist (occasionally in the same host (Xiao *et al.*, 2006a)). Although the coproELISA is known to detect all species of *Echinococcus*, the expected distribution of worm burdens amongst infected dogs is thought to differ between *E. granulosus* and *E. multilocularis* (Gemmell *et al.*, 1986c; Kapel *et al.*, 2006). However, the effect of this on coproantigen results is currently unknown. One possibility is that the distribution of OD values amongst 'positive' dogs will differ between those infected with *E. granulosus*, those infected with *E. multilocularis* (and those with mixed infections), which could cause difficulties in the fitting of mixture models. The development of species-specific coproantigen tests would be expected to resolve this problem (WHO/OIE, 2001d), but these are not yet available. In the meantime, methods of combining PCR results and ELISA results may be beneficial. This is discussed in chapters 4 and 5.

3.5 Conclusions

In conclusion, the current study details three different approaches to the dichotomous classification of coproantigen ELISA data, and addresses the major strengths and limitations of these. Although the Gaussian method has been used for some time with no apparent problems (mainly due to its easy application and requirement for only negative control samples for validation), there is little to recommend it over the other techniques assessed in the current study. ROC curve analysis offers a method of classification which can either maximise both sensitivity and specificity, or can allow selection of a cut-off point which is appropriate for the aims of the study in question, given that a suitable panel of positive and negative samples can be obtained (which is

a non-trivial issue). Alternatively, finite mixture models allow the classification of samples in the absence of any panel data without requiring the use of multiple tests (as are required for latent class analysis), and may be used to give an estimate of the prevalence of infection (based upon the mixture model component weights). However difficulties with application of mixture models may arise due to either overfitting or underfitting due to the skewed distribution of positive samples, and estimates of test sensitivity and specificity are not obtained. If multiple diagnostic tests (such as coproELISA, coproPCR and purge inspection) have been conducted, latent class analysis within a Bayesian framework likely offers a superior method of classification (Hartnack *et al.*, 2013), with the additional benefit of estimates of test sensitivity and specificity. Mixture modelling and ROC curve analysis have been conducted together, in a Bayesian framework which is able to incorporate covariates of interest (Branscum *et al.*, 2008). This strategy (which is introduced in chapter 4) has considerable potential for further development. It is hoped that further work on the application of mixture models to coproantigen ELISA data will address some of the issues identified, with a view towards the establishment of an accurate method of classification which is reliable and appropriate for use during surveillance in resource-poor communities.

Chapter 4: Development of a Bayesian mixture model to enhance interpretation of coproantigen ELISA data.

I think that progress is not possible without deviation. And I think that it's important that people be aware of some of the creative ways in which some of their fellow men are deviating from the norm, because in some instances they might find these deviations inspiring and might suggest further deviations which might cause progress, you never know.

Frank Zappa (1940 - 1993)

4.1 Introduction

4.1.1 Diagnosis of echinococcosis

As described in the previous chapter, coproantigen ELISA testing (Deplazes *et al.*, 1992; Allan *et al.*, 1992) is commonly used for the surveillance of canine echinococcosis in endemic areas and when monitoring the effect of hydatid disease control schemes (WHO/OIE, 2001d; Morel *et al.*, 2013). Interpretation of coproantigen data is commonly based upon classification of samples as 'positive' or 'negative', based upon a cut-off ELISA optical density (OD) value. Whilst there is evidence of a relationship between the coproantigen ELISA OD value and the worm burden (Deplazes *et al.*, 1992; Allan *et al.*, 1992; Craig *et al.*, 1995; Morishima *et al.*, 1999a; Raoul *et al.*, 2001; Reiterová *et al.*, 2005; Buishi *et al.*, 2005b), the corollary of this is that the test sensitivity is low when the worm burden of the sample is low (Allan and Craig, 2006; Huang *et al.*, 2013), and therefore the sensitivity of the test will vary depending upon the distribution of worm burdens in the population under study. This makes the estimation of an 'overall' test performance parameter problematic, and limits the ability to estimate the true prevalence from test results (for example, using the Rogan-Gladen estimator (Rogan and Gladen, 1978)). This problem would be expected to be particularly pronounced during the evaluation of a control scheme, where the prevalence of infection (and therefore the test sensitivity and positive/negative predictive values) would be changing.

This chapter investigates a possible alternative strategy for interpretation of coproELISA data, which avoids the need for dichotomisation (whilst retaining the possibility for dichotomisation if desired), and with potential benefits for both individual-level and population-level interpretation. This may have particular use for interpretation of longitudinal data collected during control campaigns.

4.1.2 Issues with dichotomisation

Given that the true sensitivity and specificity of a test can be estimated, the Rogan-Gladen approach (and associated methods) described in the previous chapter

can reduce some of the limitations associated with dichotomisation. However, several potential drawbacks remain – one of which is interpretation of an individual test result. Although the Rogan-Gladen adjustment may allow a reasonable estimate of the population prevalence to be made, it does not operate at the individual animal level, which could result in difficulties when reporting individual results to stakeholders or for risk factor studies which rely on individual-level interpretation. One method of accounting for this issue is through the use of positive and negative predictive values (PPV and NPV) (Altman and Bland, 1994b). These estimates are influenced by the test sensitivity (Se) and specificity (Sp) as well as the prior probability of infection in the individual (often estimated as the true prevalence of infection in the population, p):

$$PPV = \frac{(Se \times p)}{(Se \times p) + ((1 - Se) \times (1 - p))}$$

This approach clearly has use for dissemination of information back to stakeholders, and methods of incorporation of PPV and NPV into a regression model (within a Bayesian context) have also been described (Lewis *et al.*, 2012).

Another issue resulting from dichotomisation is the loss of potentially useful information regarding the probability of infection (Choi *et al.*, 2006b), or even in some cases the level of infection. For example, for a continually measured diagnostic test result for which higher values indicate infection, an animal with a very high test result would be more likely to be infected than an animal with a test result just above the cut-off. However, these two animals would both just be classified as 'positive' under a dichotomous interpretation. Similarly, in situations where levels of infection are not homogenously distributed amongst infected individuals, as is seen with overdispersed macroparasitic infections (Crofton, 1971a; Anderson and May, 1978), the test result may offer some insight into the level of infection (for example, animals with higher parasite burdens may tend to have higher test results, as is seen with the coproantigen ELISA (Deplazes *et al.*, 1992; Allan *et al.*, 1992)). In these cases, dichotomisation could result in the loss of information with potential implications for the risk of pathogen

transmission (an animal with a higher parasite burden may pose a greater risk of transmission than one with lower burdens).

In the case of the former of these two issues, a method of interpretation of test results at the individual level through the use of likelihood ratios has been suggested (Deeks and Altman, 2004). The 'likelihood ratio' of a positive test result (LR^+) can be calculated as the ratio of the conditional probability of a positive test result (T^+) given the individual is infected (D^+) (the sensitivity, in the case of a dichotomous result) to the conditional probability of a positive test result given that the individual is not infected (D^-) (which is $(1 - Sp)$, in the case of a dichotomous interpretation):

$$LR^+ = \frac{p(T^+|\, D^+)}{p(T^+|\, D^-)} = \frac{Se}{(1 - Sp)}$$

Through the use of Bayes' theorem, an estimate of the post-test odds of disease can then can be estimated by multiplying the LR^+ by the pre-test odds of disease, which can then be converted to a probability if desired. The 'raw' test result (without dichotomisation) can also be used to estimate a likelihood ratio, using the same principles as described above. Another method, termed probability diagnostic assignment (PDA), has been developed which incorporates test results from known infected and uninfected individuals in order to estimate the individual-level probability of infection and the population-level prevalence of infection using a frequentist application of Bayes' theorem (Thurmond *et al.*, 2002). This approach has been developed into a fully Bayesian framework, which is computationally easier and which may be less dependent on the availability of data of known status (given there is reasonable separation in the distribution of test results between infected and uninfected individuals) (Choi *et al.*, 2006b).

As can be seen from the equations above, these approaches require clear estimates of the sensitivity and specificity of the test (and in the case of predictive values, also an estimate of the prevalence of infection in the community). As described above, there is evidence that the sensitivity of the *Echinococcus* coproantigen ELISA test is correlated with the worm burden, with lower sensitivities in the case of low burdens (Allan and

Craig, 2006). As the burden of *Echinococcus* in the definitive host is commonly highly overdispersed (Jenkins and Morris, 1991; Torgerson and Heath, 2003; Budke *et al.*, 2005a), it would be expected that the majority of infected individuals would have low burdens, and as such would be less likely to be identified using a dichotomous interpretation than those with higher burdens. The overall test performance would therefore be expected to depend upon the distribution of worm burdens in the population of interest.

Another challenge when attempting to estimate the sensitivity and specificity of any test for canine echinococcosis is the lack of a readily available gold standard test (i.e. a test with perfect sensitivity and specificity). As described in previous chapters, the gold standard test for canine echinococcosis is necropsy of dogs and examination of intestines using the sedimentation and counting technique (WHO/OIE, 2001d; Torgerson and Deplazes, 2009). This is rarely possible in the field situation as it is logistically challenging, potentially biohazardous, and requires culling of dogs. Alternative strategies of estimating test performance based upon application of Bayesian modelling strategies and latent class analysis to multiple imperfect test results have been described (Ziadinov *et al.*, 2008; Hartnack *et al.*, 2013).

4.1.3 Finite mixture models

Finite mixture models (FMMs) are a statistical tool for the identification and quantification of subpopulations within a larger population. As FMMs are described in the previous chapter and elsewhere (McLachlan and Peel, 2000c), only extensions of relevance to the current study will be detailed here. Due to difficulties interpreting the results of FMMs when the number of components was greater than two (see chapter 3), the current study fixed the number of components to two – indicating 'negative' and 'positive' individuals. For a univariate FMM with two components, Y represents a vector of length n, relating to a random sample of n individuals from a population $(Y_j;\ j = 1{:}n)$. The probability density function of y_j, $f(y_j)$, can be presented as the sum of two components, each of which has its own 'mixing proportion', π_0 and π_1,

which lie between zero and one and sum to one. Each component is distributed according to its own specified distribution, G_i:

$$f_i(y_j) \sim G_i; (i = 0,1)$$

$$f(y_j) = (1 - \pi)G_0 + \pi G_1$$

This basic model structure detailed above is true for both Bayesian and frequentist mixture models, with the Bayesian approach modelling the parameters of G_0 and G_1 (i.e. mean and variance, in the case of a Gaussian mixture model), and π as random variables, and allowing the incorporation of prior information for these. Due to both the more 'philosophical' advantages of Bayesian methodology over frequentist approaches, and its increased computational ease when dealing with complex models, it was decided to develop a Bayesian model in the current report.

Finite mixture models have been used in the classification of diagnostic test results in a number of reports (Choi *et al.*, 2006a; b; Erkanli *et al.*, 2006; Branscum *et al.*, 2008; Hanson *et al.*, 2008). Although in some cases, a Gaussian model may be appropriate, this is commonly not the case – for example, infected individuals may represent a relatively heterogenous group of individuals compared to noninfected individuals (possibly due to variable times since infection, or due to different burdens of infection) (Branscum *et al.*, 2008; Hanson *et al.*, 2008). As such, alternative, non-Gaussian, approaches to modelling should be considered. In particular, semi-parametric/nonparametric approaches such as those based on Dirichlet processes (Erkanli *et al.*, 2006; Hanson *et al.*, 2008) or Polya Trees (Branscum *et al.*, 2008; Hanson *et al.*, 2008) have been developed.

4.1.4 Polya trees

A Polya tree is a method of modelling 'nonparametric' distributions in a parametric fashion, and is conceptually based upon repeated subdivision of the sample space and allocation of Beta-distributed conditional probabilities of subset membership. Polya trees have been covered in detail elsewhere (Ferguson, 1974; Lavine, 1992, 1994;

Mauldin *et al.*, 1992; Hanson, 2006; Christensen *et al.*, 2008), and so will be only covered briefly here. As described above, a Polya tree can be created by dividing the sample space repeatedly (each division representing one 'stage'). Therefore, at stage one, the whole sample space is divided into two subsets; whereas at stage two, each of these two subsets is further divided into two; and so on. At each stage, the conditional probability of membership in the particular set given membership in the 'parent' set in the stage above is estimated (assuming that these follow a Beta distribution); ensuring that each of these paired conditional probabilities (belonging to each 'parent' set in the stage above) sum to unity. Polya trees can also be created by the generalisation of a parametric distribution, which is the approach used in the current report, basing the Polya tree upon a Gaussian distribution. In this case, at stage one, the sample space is split at the median; at stage two it is split at the quartiles; and so on. Although the parameters (π and σ^2) of the original distribution remain unchanged, the probabilities of subset membership are permitted to vary – allowing deviation from the Gaussian distribution if warranted (of course, if the probability of subset membership was fixed at 0.5 for each subset at each level, the original Gaussian distribution will be retained). A weighting parameter, c, indicates the level of deviation from the original distribution, and takes a high value (e.g. >5) if there is minimal deviation from this; and a low value (<1) if the resultant distribution is to be largely nonparametric (Branscum *et al.*, 2008). In order to 'smooth out' the resultant distribution at the boundaries of the subsets, a 'mixture of Polya Trees' (MPT) can be created by setting priors on π and σ^2 and allowing some variation in the centering measures used (Paddock *et al.*, 2003; Hanson, 2006).

4.2 Materials and Methods

Samples were collected from a total of 38 unwanted domestic dogs from northern Xinjiang province in the People's Republic of China, as described in the previous chapters and in other recent papers (van Kesteren *et al.*, 2015). No failures of replication were identified in the current study.

The mixture model was developed using JAGS version 3.4.0 (Plummer, 2003), from within the statistical package R (R Development Core Team, 2014) by using the rjags package (version 3-11) (Plummer, 2013). Model code is provided in the appendix (A4). The aim of the model was to incorporate known panel data from necropsy (and/or purgation) -confirmed positive and negative dogs. This part of the model is derived largely from a mixture model which incorporates known panel data (Choi *et al.*, 2006a; b), with additional code to account for uncertainty amongst negative sample data due to the imperfect sensitivity of these approaches (meaning that some apparent necropsy/purge negative dogs could in fact be infected). This was achieved by using a method based on the examination of standardised residuals within the negative panel (Birkes and Dodge, 1993), and selection of a cut-off for exclusion based upon this Z-level.

The aggregated distribution of worm burdens (and therefore the right skew in OD values amongst positive samples) was accounted for using code developed for the creation of Polya trees and MPTs (Christensen *et al.*, 2011). As described in a previous report (Choi *et al.*, 2006b), the resultant model is able to both estimate the prevalence of infection within the population as a whole, and is able to estimate the predicted probability of infection for any individual based upon their OD value, based upon Bayes' theorem:

$$p(D^+ | OD) = \frac{p(OD \mid D^+) \times p(D^+)}{p(OD \mid D^+) + p(OD \mid D^-)}$$

Where $p(OD \mid D^+)$ is the likelihood of membership in the positive component for that particular OD value (and $p(OD \mid D^-)$ is the likelihood of membership in the negative component for that particular OD value), and $p(D^+)$ is the estimated prevalence.

A simple linear regression model of the relationship between OD value and the natural logarithm of the worm burden of positive samples was also incorporated into the model, based on code developed in a recent Bayesian textbook (Kruschke, 2011), where j represents the vector of confirmed positive samples:

$$\ln(burden_j) \sim N(\mu_j, \sigma^2)$$

$$N(\mu_j, \sigma^2) = \beta_0 + (\beta_1 \times OD_j)$$

A schematic network diagram for the modelling approach is shown in figure 4.1. The Z-scores of the OD values for samples which were negative on necropsy were calculated, and those with z-scores higher than would be expected (i.e. which reduced the fit of the data to a Gaussian distribution) were removed from further analysis. The remaining samples were then used to parameterise the negative component of a mixture model. Field samples were then applied to the mixture model, with the positive component parameterised using Polya trees. This gave parameter estimates for the two mixture model components (mean and variance for the negative and positive components, along with c in the case of the positive component), along with the overall estimated prevalence (i.e. the mixture model weight of the positive component). For each individual sample, an estimate of the probability of membership in each component was obtained.

The samples found to be positive on necropsy were used to parameterise a linear regression model of the relationship between the OD and the log of the estimated worm burden. The output of this model was used to predict the log burden for each of the field samples. This estimate was then multiplied with the probability of membership in the positive component in order to create a "sample score".

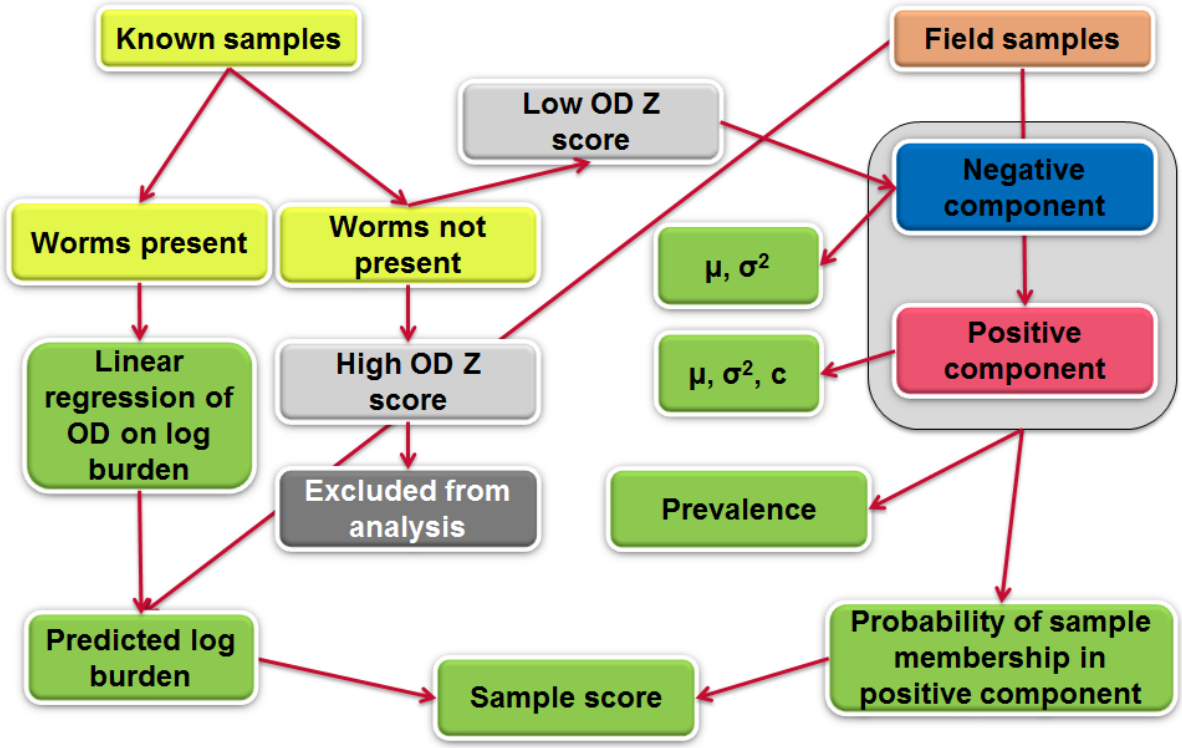


Figure 4.1. Conceptual modelling approach used in the current study. Although known samples and field samples are shown separately, in the current case, the same samples were used in both cases due to the limited number of samples available.

An adaptive Markov Chain Monte Carlo procedure with Gibbs sampling was used to obtain estimates of the posterior distributions of all model parameters of interest. Noninformative priors were used and a total of 100,000 iterations were run, with the Markov Chain thinned to one in every ten iterations (total 10,000 iterations retained), following a burn-in of 50,000 iterations. As model output was obtained for each individual sample tested and for 100 OD values from 0.01 to 1.00, MCMC diagnostics (visual inspection of the Markov chain) were only conducted on the population-level parameters (mean and standard deviation of negative and positive distributions; the Polya Tree weighting parameter, *c*; Z-level for exclusion of outliers in negative panel; estimated proportion of positive samples; intersect and slope of linear regression model).

In order to evaluate the model, the same dataset (taken from dogs in Xinjiang, as described above) was used twice: for both model 'training' (i.e. parameterisation of both the negative distribution of the mixture model and the linear regression model),

and validation. Therefore, the "known samples" in figure 4.1 are identical to the "field samples" in this case. This is a highly problematic approach, but was considered necessary due to the limited size of the dataset.

4.3 Results

The collection of the sample panel used here has been described elsewhere (van Kesteren *et al.*, 2015), and so will not be described in detail here. Of the 38 dogs investigated, *Echinococcus* spp were found in 16 (42%), with a range of worms from 2 to over 10,000, and a median of 100. As described in chapter 3, the distribution of OD values amongst the necropsy negative animals was not Gaussian-distributed, and three samples in particular were clear outliers (with OD values more than 1.5 times the interquartile range higher than the upper quartile of the 'negative' sample distribution). Exclusion of these three samples from analysis altogether would give a prevalence estimate of 16/35 = 46%; and inclusion of them as likely true positive samples would give a prevalence estimate of 19/38 = 50%.

The distribution of OD values and the mixture model component estimates (based upon the median estimates of the posterior distribution) is shown in Figure 4.2. All posterior distributions of the mixture model parameters (including the value of *c* for the positive Polya Tree distribution and the Z-level for the negative samples) from MCMC sampling are shown in Figure 4.3, with the mode and the 95% high density interval highlighted (using code for graphical output provided in Kruschke (2011)). Similar posterior estimates of the overall prevalence and the parameters (intercept and slope) of the linear regression model are shown in Figure 4.4. All suspected negative samples with OD values of less than 0.1 were included in the fitting of the Bayesian model, whereas the posterior probability of acceptance of negative panel samples with OD values greater than this decreased as the OD value was increased, as shown in table 4.1. The overall median estimated prevalence was 49.6%, with a 95% HDI of 28.7% to 69.1%.

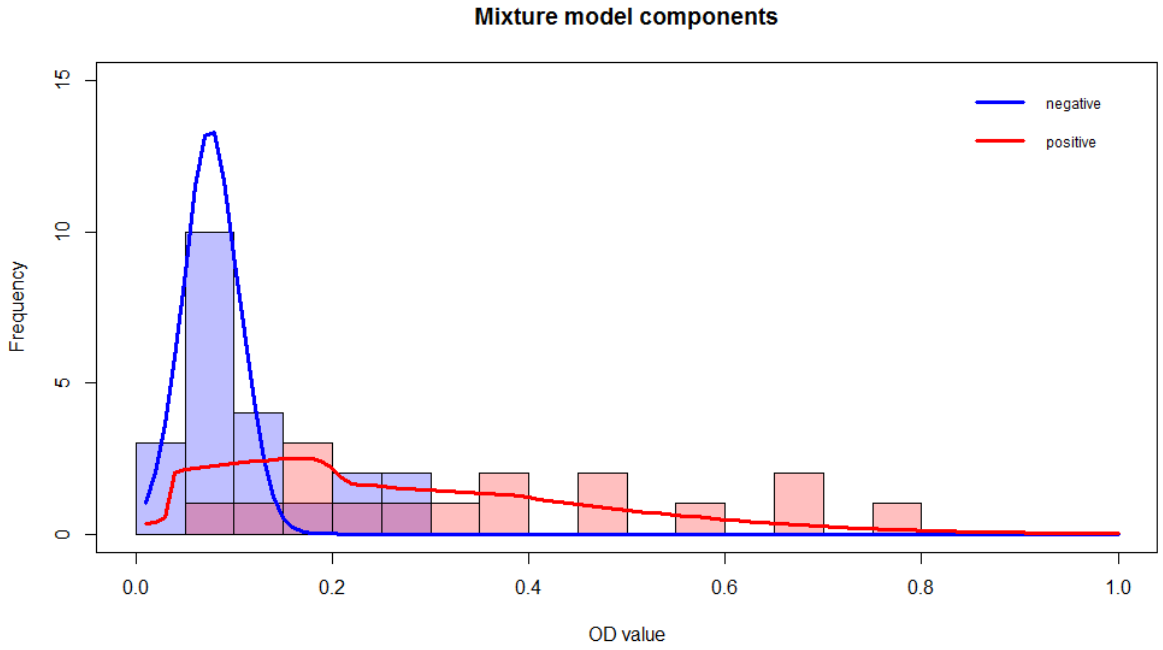


Figure 4.2. Distribution of test results and median estimates of the mixture model components

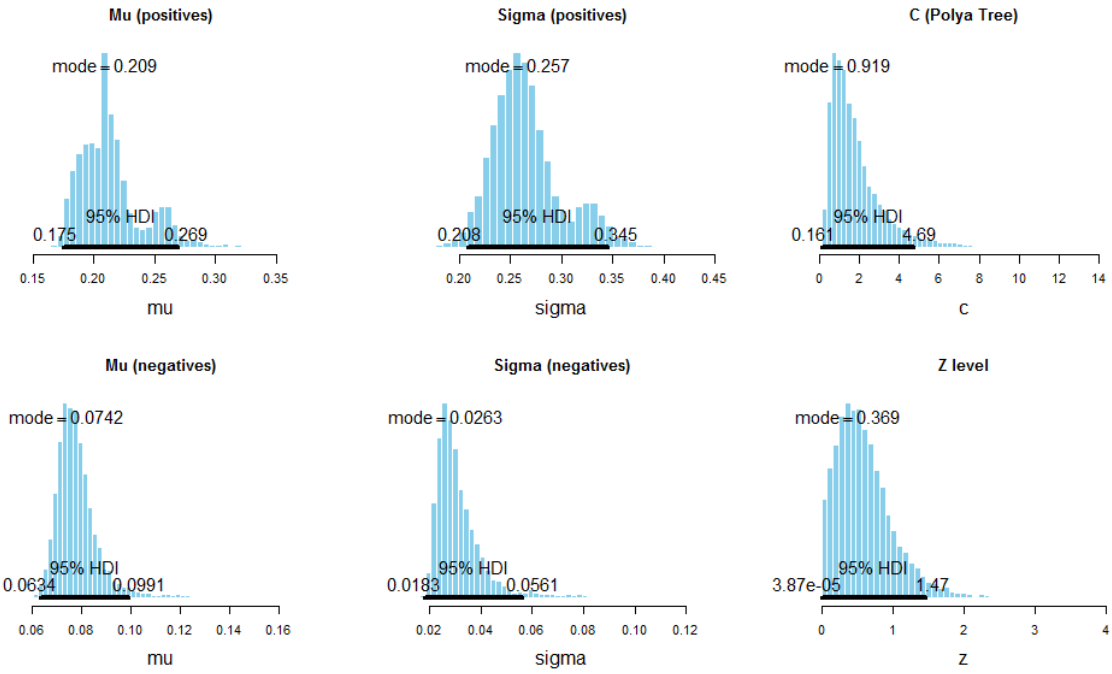


Figure 4.3. Posterior estimates of mixture model parameters

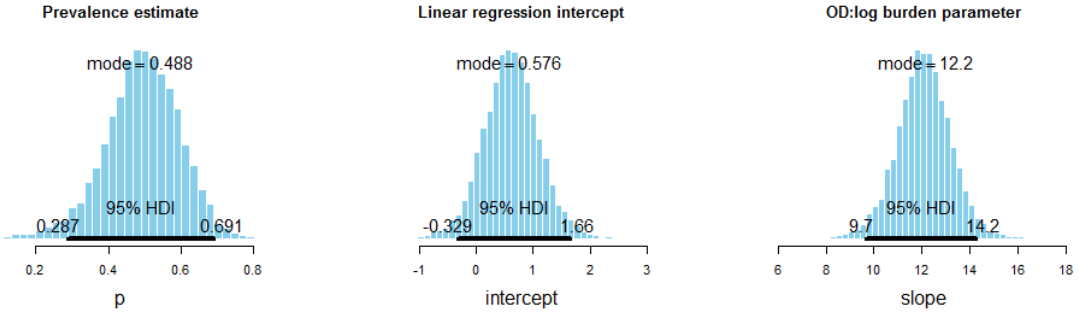


Figure 4.4. Posterior estimates of the overall prevalence of infection from the mixture model, and estimates of the parameters of the linear regression model

Table 4.1. Mean proportion of high-OD negative panel samples (OD>0.1) included in the final mixture model

OD value (4 d.p.)	Posterior probability of inclusion in the mixture model
0.1039	0.9999
0.1063	0.9996
0.1116	0.9975
0.1159	0.9874
0.1656	0.2187
0.2055	0.0522
0.2478	0.0184
0.2595	0.0157
0.2782	0.0127

The distribution of posterior probabilities of infection according to the mixture model for the individual samples and for all OD values between 0 and 1 is shown in Figure 4.5. Figure 4.6 demonstrates the median posterior probabilities of positivity for all samples.

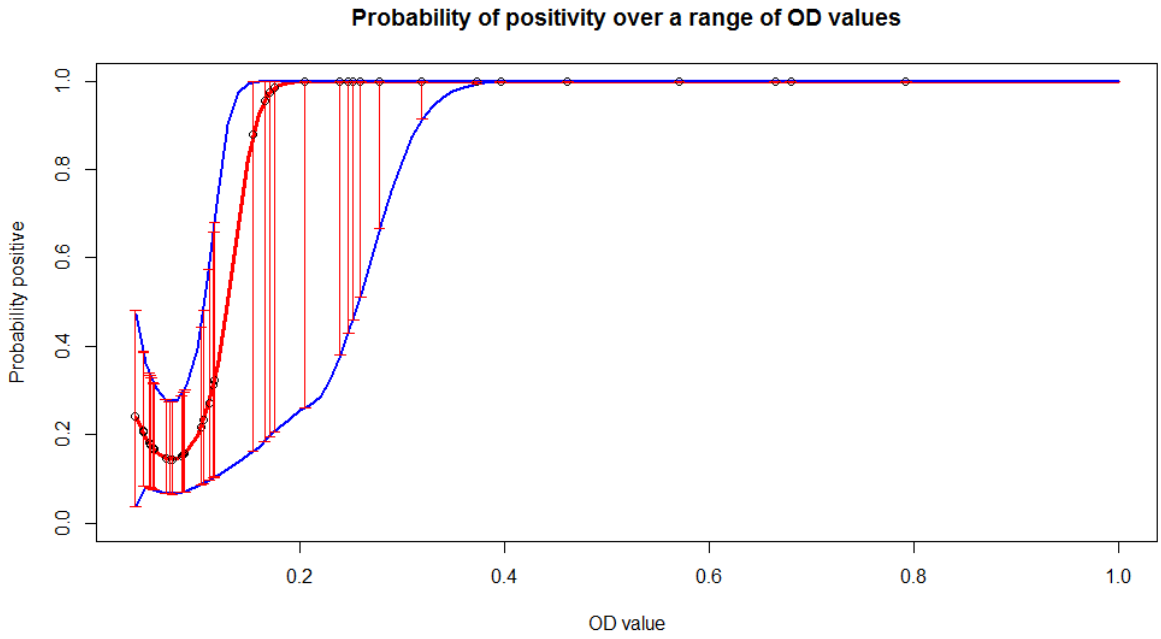


Figure 4.5. Median posterior probabilities of positivity for samples (black circles) and for all OD values between 0.0 and 1.0 (red line). Blue lines and red 'whiskers' indicate the interquartile range.

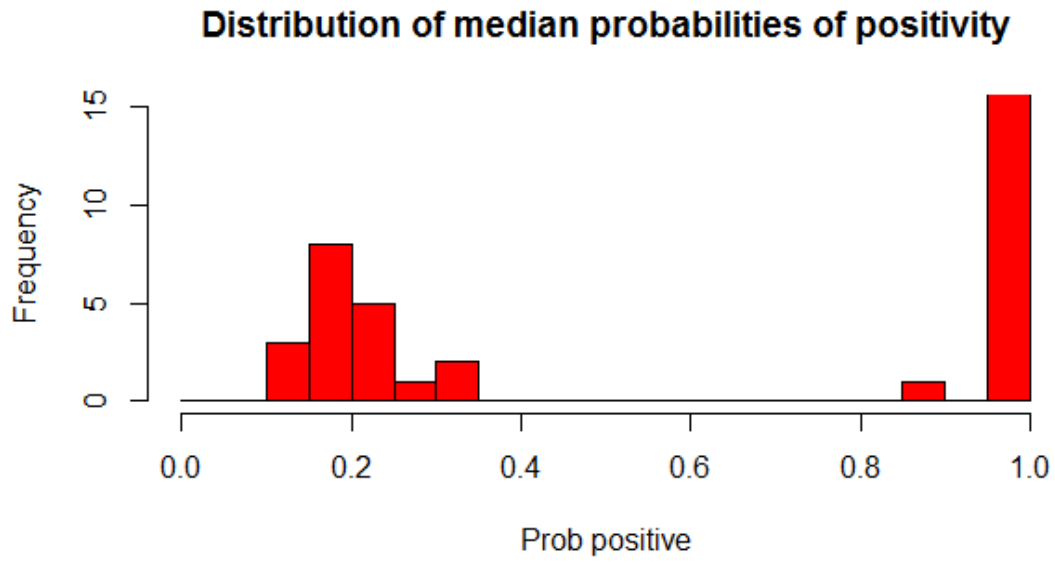


Figure 4.6. Distribution of median posterior probabilities of infection for all samples, according to the mixture model.

The distribution of posterior scores based on the product of the posterior probability of positivity and the predicted log worm burden for the individual samples and for all OD values between 0 and 1 is shown in Figure 4.7. Figure 4.8 demonstrates the distribution of median posterior score estimates for all samples.

Finally, a comparison was made between the median score estimates and the worm burden for all samples. Figure 4.9 shows this relationship both for the unadjusted burden and for the log burden.

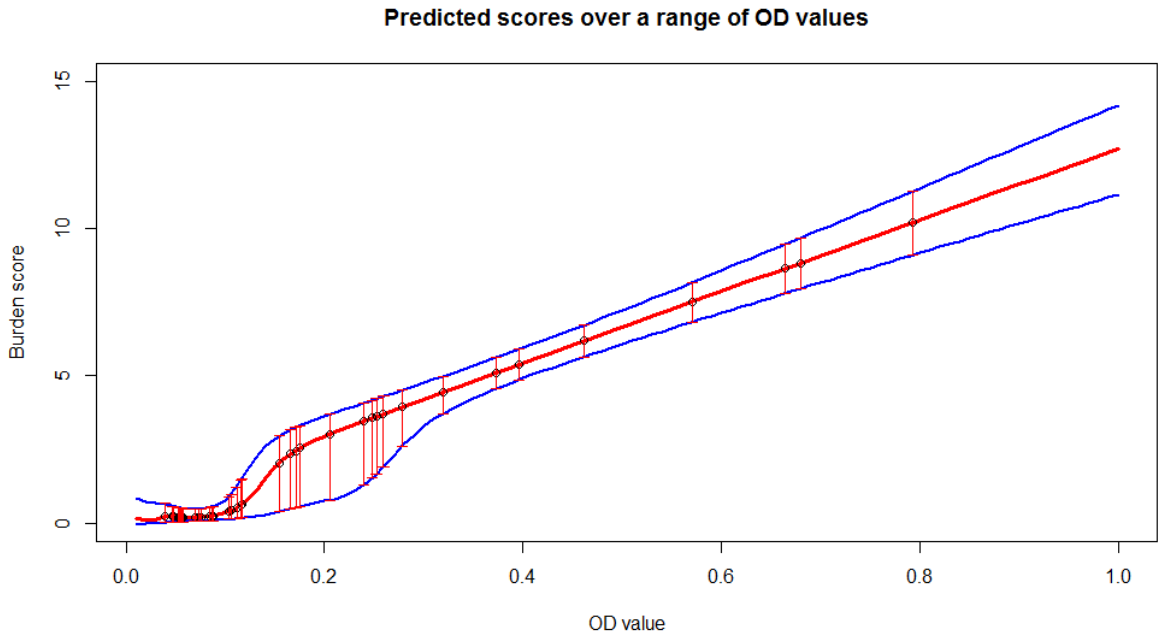


Figure 4.7. Median posterior scores for samples (black circles) and for all OD values between 0.0 and 1.0 (red line). Blue lines and red 'whiskers' indicate the interquartile range. The 'burden score' is the ouput of the Bayesian mixture model, and can be considered to broadly relate to the log of the expected burden.

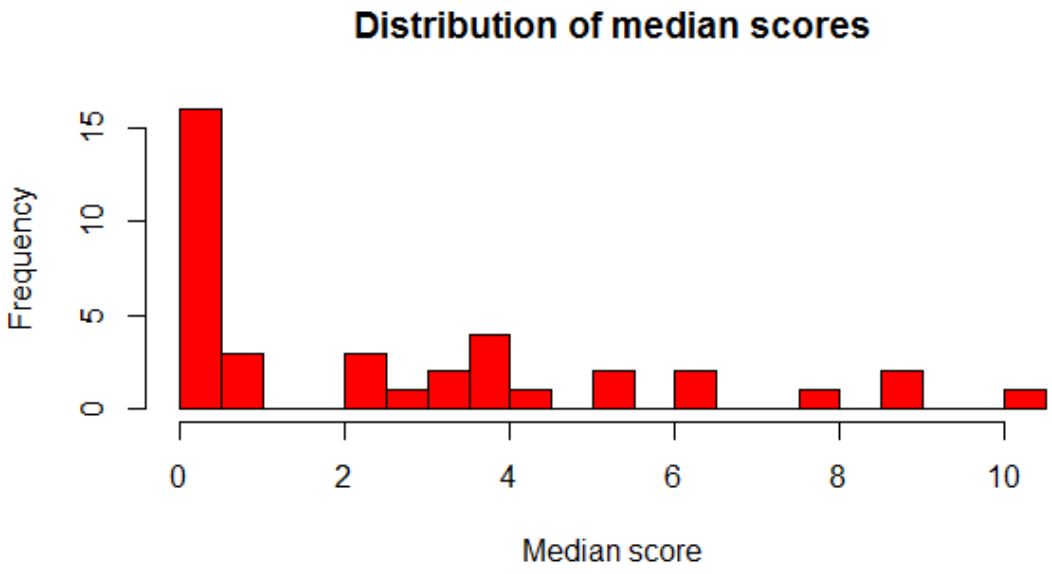


Figure 4.8. Distribution of median posterior scores for all samples

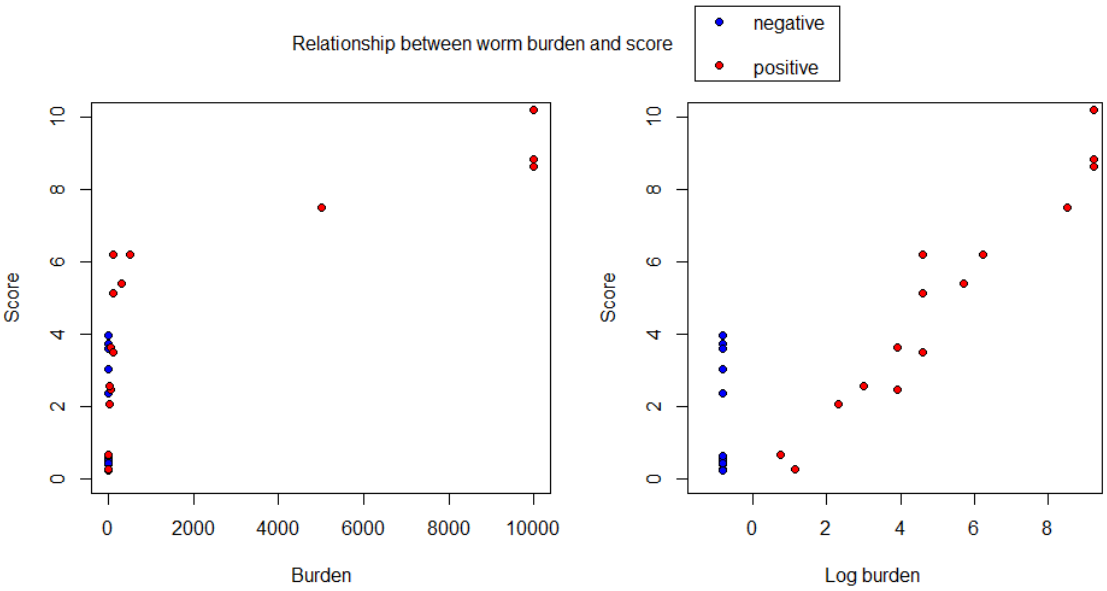


Figure 4.9. Relationship between worm burden and median score for both unadjusted and log-adjusted burden estimates. For right-censored burden estimates, the lower end of the interval is used. Samples taken from dogs with observable worms are shown in red, and those taken from dogs with no worms are shown in blue.

4.4 Discussion

Optimal interpretation of canine *Echinococcus* coproELISA test data is important for epidemiological investigation, as well as for surveillance and control activities (Hartnack *et al.*, 2013). The most commonly used method of interpretation of coproantigen ELISA data currently is based upon dichotomisation using a cut-off point three standard deviations above the mean OD of a known negative panel (Deplazes *et al.*, 1992; Allan *et al.*, 1992). The advantage of this approach is that the results are easy to understand – giving a clear estimate of which individual animals are coproantigen 'positive' or 'negative', and an estimate of the coproantigen prevalence at the population level. However, as well as the limitations associated with this particular strategy of dichotomisation (as described in the previous chapter), any dichotomisation approach is likely to result in imperfect sensitivity and specificity (Allan and Craig, 2006; Huang *et al.*, 2013; Hartnack *et al.*, 2013). Additionally, by dichotomising results, valuable information on the probability of infection is lost (Choi *et al.*, 2006b), along with possible information on the burden of infection (in the case of testing for echinococcosis). Information on potential burdens is known to be of great value considering infection status (Hofer *et al.*, 2000; Raoul *et al.*, 2001). Finally, dichotomisation is known to reduce the statistical power (i.e. the ability to detect an effect if it is truly present) of any further analysis conducted on the data (Altman and Royston, 2006).

The current paper describes a novel method of modelling *Echinococcus* spp coproELISA OD data, and uses Bayesian finite mixture modelling in order to identify the constituent infected and uninfected groups. This produces output which can be interpreted in a continuous or a dichotomous fashion, as desired, and doesn't require the use of multiple diagnostic tests (Hartnack *et al.*, 2013) and/or samples taken from populations with different levels of infection (Ziadinov *et al.*, 2008). As such, it may be useful for ongoing surveillance activities (where the cost of multiple testing may be prohibitive), although there are limits to its use in coendemic areas (see below). Whilst not yet fully evaluated (due to a relative lack of high quality parasitological data), it is hoped that the modelling approach described in the current study will offer

a framework for improving the interpretation of test results, whilst still offering a transparent approach to dichotomous interpretation if desired.

4.4.1 Population level interpretation

One output from the Bayesian finite mixture model is an estimate of the weights of the two component distributions – which in biological terms will relate to the prevalence of infection (and the prevalence of non-infection) in the population. These should not be affected by 'overlap' between the negative and positive distributions to the same degree as approaches based on cut-offs, and as such should offer a better approximation of the true prevalence of infection in a community. Despite the wide HDI of the prevalence parameter (which will result from the relatively small number of samples evaluated in the model) (Figure 4.4), the modal prevalence estimate of 49% obtained from the current model is close to the suspected 'true' prevalence of around 46%.

Another possible method of interpretation of the output at the population level is to account for the distribution of burdens. Population-level interpretation of OD values directly have been studied recently (Raoul *et al.*, 2001), but is difficult to interpret from a biological perspective, and would require some form of standardisation of OD data if to be used to compare testing conducted in different locations or different times. The output of the current model could be used to approximate the distribution of burdens (or log burdens) in the population under study (see figure 4.8). Investigation of the distribution of model 'scores' obtained from the model could shed light on aspects of the host-parasite relationship (Crofton, 1971a), and is discussed further in relation to overdispersion below. From a practical perspective, the output of the current model could be a useful step towards the development of a meaningful classification system for interpretation of the levels of canine echinococcosis in a community, as is recommended by the WHO for monitoring and control of schistosomiasis (WHO Expert Committee, 2002; Olives *et al.*, 2012).

4.4.2 Individual level interpretation

The current model also facilitates interpretation at the individual animal level, as an indication of the conditional probability distribution of coproantigen positivity for any OD value (or for any sample included in the model) is obtained. This may be interpreted in a dichotomous fashion by taking the modal estimate of the median probability for each individual (i.e. classify as positive if the median probability is greater than 0.5; as negative if it is less than 0.5; and decide upon a rule for classification of those with a probability of exactly 0.5). Using this approach. the posterior probabilities of infection in the individual samples tested here showed a reasonably clear distinction between 'low' and 'high' probabilities (for example, as shown in figure 4.6 for the median estimates). This suggests that, in this particular case at least, individual samples could reasonably be classified in a dichotomous fashion, as is commonly required for risk factor studies. If these distinctions were not apparent, then alternative approaches for categorical interpretation (such as inclusion of an 'unknown' category) could be considered. As the true necropsy status of the samples included in the current model is known, the performance of the dichotomous assignment of sample status can be estimated. The estimated sensitivity is 14/16 = 88%, and the estimated specificity is 17/22 = 77% (or 17/18 = 94% if the high OD 'outliers' are removed from the negative group – see below). These results are identical to those derived from a frequentist Gaussian mixture model, and very similar to those resulting from ROC curve analysis, of the same data, as described in the previous chapter.

Test results interpreted in a dichotomous fashion are easy to understand because they considerably simplify the true situation. Whilst this may simplify the communication of results, this approach is less useful for the investigation of patterns of macroparasite infection, and will generally increase the sample size needed to detect a significant effect when conducting analytic studies (Altman and Royston, 2006). As described earlier, macroparasitic infections, including echinococcosis, usually show a highly aggregated distribution within a population (Torgerson and Heath, 2003; Budke *et al.*, 2005b). As such, 'positive' individuals do not represent a homogenous group – with most individuals carrying low worm burdens, and a minority of individuals usually

carrying the majority of the parasite biomass (Jenkins and Morris, 1991; Hofer *et al.*, 2000). This has implications for interpretation at the individual level (since the infection risk associated with high burden hosts would be expected to be greater than that in low burden hosts), and at the population level (since the majority of the parasite biomass may be found in a minority of the host population), and has repercussions for control (Woolhouse *et al.*, 1997). This also means that the total biomass of *Echinococcus* spp could be reduced (through a moderately effective control scheme, for example) without necessarily observing a similar change in the estimated prevalence, as has been seen with other overdispersed helminth and macroparasitic infections (Guyatt *et al.*, 1990; Shaw and Dobson, 1995).

One method of interpretation of infection status at the individual animal level which can be relatively easily conveyed to stakeholder is the probability of positivity (Choi *et al.*, 2006b), which is similar in concept to the positive predictive value. This expands upon the idea of a dichotomous interpretation by accounting for uncertainty when interpreting OD values which lie in the 'overlap' region between negative and positive samples. The level of overlap may be greater in more aggregated distributions, where more animals have low burdens and therefore where a traditional dichotomous interpretation based upon a cut-off would be expected to have a lower sensitivity and/or specificity. However, this approach was unable to distinguish individuals with very high burdens from those with moderate burdens – animals with high OD values invariably had a 100% probability of positivity (figure 4.5)). It was for this reason that estimates of the probability of positivity and the expected worm burden were combined in order to create a 'score' for each individual (see figure 4.7). The score estimate may be a useful tool for the communication of the predicted burdens within individual dogs, as well as within a community as a whole as described above. However, further work would be required in order to investigate how best to manage, utilise and disseminate this information.

4.4.3 Overdispersion

Whilst care should be taken when directly attempting to interpret the score estimates obtained from the model in relation to the predicted worm burden (due to the small panel of samples upon which this estimation is based), these score estimates could be useful for investigation of overdispersion in parasite burdens in a population, as alluded to earlier. As described in chapter 1, overdispersion is a major characteristic of macroparasites (Woolhouse *et al.*, 1997; Perkins *et al.*, 2003; Poulin, 2007), and is a known characteristic of echinococcosis (Gemmell *et al.*, 1986c; Hofer *et al.*, 2000; Budke *et al.*, 2005b). Based on the conservative estimates of the right-censored worm burdens shown in figure 4.9, the three dogs with the highest scores carried an estimated 83% of the total parasite biomass; and the four dogs with the highest scores carried an estimated 97% of this. This finding agrees with the following statement made by Anderson and May: 'it is not uncommon to find 80 per cent or more of the macroparasites contained within 20 per cent or fewer of their … hosts' (Anderson and May, 1991c).

This overdispersion is likely to have a particular effect on the transmission ecology of *Echinococcus* spp, and vice versa. As described in chapter 1, mathematical and statistical models are indispensable for the investigation of the processes which may give rise to overdispersion, the stability of host-parasite interactions, and the potential effect of control schemes. However, the availability of high quality data often limits the ability to parameterise these models. Although the current model is not a substitute for high-quality parasitological data, such as that obtained through necropsy or purgation, it does increase the amount of information which may be obtained from a simple ELISA test. As such, it may improve the interpretation of surveillance data collected routinely during control schemes. On a basic level, the score estimates obtained from the current model can be considered to be a form of standardised OD data, offering the potential for interpretation of longitudinal data using simple regression models. This could be useful for the identification of dogs with higher burdens (whether at the individual dog or household level, or in the context of associated covariates). Control targeted at these individuals could have a

disproportionate effect on the total parasite biomass in a community and therefore improve the effectiveness of a control scheme (Woolhouse *et al.*, 1997). If no risk factors are identified, then this in itself could be accounted for in a control scheme, and a focus could be placed on maximising praziquantel coverage throughout a community. These concepts are introduced and expanded upon in chapters 5 and 6. The effect of control schemes on parasite dynamics is an important area of current research (Basáñez *et al.*, 2012a; b), and generally necessitates the use of mathematical models. The output of the current model could potentially be used to parameterise a mathematical model of transmission, in order to investigate the effect of a control scheme and better target ongoing surveillance and control. This idea is explored further in chapter 7.

4.4.4 Model limitations

As mentioned earlier, the largest constraint to full interpretation of the model output described here is the use of the same data for both model fitting and validation. This was considered unavoidable, as little other parasitologically-confirmed data were available. As such, whilst the modelling framework described here may be reasonable, extra care should be taken when interpreting the results of this preliminary evaluation. The difficulty in obtaining good quality parasitological data is a known issue for echinococcosis. As the positive predictive value of purgation for identification of echinococcosis would be expected to be 100%, this approach could therefore be used to identify positive samples. However, the accurate identification of negative samples is very challenging, and itself would be worthy of further investigation. Whilst some account for this is incorporated into the model structure (using the Z-score to exclude potential false negatives), the presence of false negatives (which would be expected to overrepresent those individuals with low burdens which comprise a large proportion of the population, if not a large proportion of the total parasite biomass) is a considerable problem for full evaluation. Although the use of faecal samples from known nonendemic / echinococcosis-negative areas is a simple solution to this problem, these dogs are unlikely to be representative of negative dogs in a highly

endemic area. Further work investigating optimal methods of parameterising the current model in the face of this challenge would be beneficial.

One central concept of the model developed here is that the distribution of coproELISA OD values from noninfected dogs will be broadly Gaussian distributed. The issues associated with the use of the (unbounded) Gaussian distribution to model an outcome which can only take positive values were introduced in the previous chapter, and remain a potential issue here. However, these issues were considered relatively trivial. Analysis of coproantigen data from nonendemic sites has repeatedly suggested that the distribution of OD values amongst true *Echinococcus* spp negative dogs follows a Gaussian distribution (author's own observation) – hence the original practice of calculating a cut-off for positivity which is three standard deviations above the mean of a known negative panel (Deplazes *et al.*, 1992; Allan *et al.*, 1992). The necropsy negative samples used here did not follow a Gaussian distribution – with at least three dogs having higher than expected OD values. As described in the previous chapter, the data in the current study are of high quality (having been based upon necropsy and visual inspection of intestines by experienced individuals), but were not derived using the 'gold standard' test (the sedimentation and counting technique). As such, it is plausible that some infected animals (especially those with low burdens) may not have been detected, and as such will be classified as false negatives (Allan and Craig, 2006). However, a similar non-Gaussian distribution of OD values amongst negative samples has also been reported from a study of foxes in France (Raoul *et al.*, 2001). This data was based upon the sedimentation and counting technique, and therefore the sensitivity would be expected to be high (Raoul *et al.*, 2001).

One other possible explanation for the observed lack of Normality amongst negative samples relates to the disparity between infection and the presence of coproantigens. The coproantigen ELISA test detects *Echinococcus* coproantigens rather than the presence of worms *per se*. Coproantigens may be present for some days after removal of the worms themselves with a cestocidal drug (Deplazes *et al.*, 1990; Allan *et al.*, 1990; Jenkins *et al.*, 2000). As a praziquantel dosing campaign was in place in the study area at the time of the study (van Kesteren *et al.*, 2015), those dogs with high OD

readings may have recently been treated with praziquantel (leaving them free of worms but still with residual coproantigens). The possibility of cross reaction with other cestodes such as *Taenia* spp was considered unlikely in the current case as necropsy was conducted and *Taenia* are large worms which would be difficult to overlook.

Further work is required to evaluate the usefulness of the current Bayesian finite mixture model as a method of interpretation of ELISA data, in particular with regards to the incorporation of worm burden data. One other issue of relevance is coendemicity of different strains or species of *Echinococcus* in an area. The data used here was taken from an area principally endemic for *E. granulosus* sensu lato (which includes a number of different species and strains, but which all have a similar lifecycle). However, there are major foci of *Echinococcus* spp infection where both *E. granulosus* and *E. multilocularis* coexist (including Kyrgyzstan). Due to differences in the lifecycles of these two species (in particular in terms of intermediate host preference, but also in terms of patterns of infection in domestic dogs (Kapel *et al.*, 2006) and potential immunity (Budke *et al.*, 2005b), the distributions of parasites in infected dogs and the effect of an intervention campaign may differ between species. As the coproELISA does not allow species identification, PCR techniques are required to distinguish these species. Methods of incorporation of PCR data into the current model, possibly using latent class models (Hartnack *et al.*, 2013) and/or Bayesian strategies (Praet *et al.*, 2013) will be investigated in future work.

4.5 Conclusions

The current paper describes a novel approach for interpretation of canine coproantigen ELISA data based upon Bayesian finite mixture modelling. The model can be made identifiable through the incorporation of samples of known status taken from endemic areas. The limited sensitivity of the methods of diagnosis available can be incorporated into the Gaussian-distributed negative component of the mixture model, and the skewed distribution of positive samples can be explicitly accounted for by using Polya trees. The output of the model can be used for traditional dichotomous

interpretation of sample data (along with estimation of the sensitivity and specificity of the test), but it is suggested that attention is given to the possibility of interpretation of results on a continuous scale. Methods of interpreting this data at the population level and at the individual level are discussed, along with potential areas of further application – in particular, by incorporating the model output into statistical and mathematical models. Despite these promising signs, further work is required to evaluate this approach, using data collected from other areas, and also considering incorporation of PCR data in order to allow identification of species of *Echinococcus* present.

Chapter 5: Use of multiple correspondence analysis to classify dog ownership and potential risk factors for canine echinococcosis

"Science may be described as the art of systematic over-simplification
— the art of discerning what we may with advantage omit"

Karl Popper (1902 – 1994)

5.1 Introduction

5.1.1 Echinococcosis in Kyrgyzstan

As described earlier, independence from the Soviet Union in 1991 led to a loss of regulation of animal management (including disease control) and an increase in poverty in Kyrgyzstan. This has resulted in a variety of ecological and epidemiological problems. In particular, a reduction in transhumant movement has led to overgrazing of pastures around human settlements, and a lack of veterinary disease control has resulted in an increase in the prevalence of a number of zoonotic diseases in humans (in particular, brucellosis and echinococcosis) (World Bank, 2005, 2010). The increase in the prevalence of human alveolar echinococcosis (AE) is a considerable concern due to the high case fatality rate if this disease is left untreated. These increases in prevalence have appeared to be particularly pronounced in the south of the country (Torgerson, 2013; Usubalieva *et al.*, 2013), where canine infection with *E. granulosus* G1, *E. canadensis* G6, and *E. multilocularis* has recently been described (van Kesteren *et al.*, 2013). In response to this, a World Bank–funded project aiming to improve pasture management and strengthen the agricultural services was instigated in 2010 (World Bank, 2010). One component of this intervention was focussed on the control of a number of veterinary and zoonotic pathogens, including *Echinococcus* spp. (World Bank, 2010). The cornerstone of this campaign was regular praziquantel dosing of dogs, although a variety of other strategies were also planned (WHO, 2011).

5.1.2 Surveillance prior to control

Surveillance is essential for the monitoring and evaluation of any control scheme, and in the case of echinococcosis is commonly based largely upon testing of canine faecal samples in order to estimate the coproantigen prevalence in a community (Deplazes *et al.*, 1992; Allan *et al.*, 1992), although monitoring of infection in other hosts can be very useful (Gemmell and Schantz, 1997; Gemmell *et al.*, 2001). A major consideration when planning an intervention campaign is the estimation of the baseline prevalence of infection (or, more commonly, the prevalence of coproantigen positivity), and the

identification of risk factors associated with this (which can improve understanding of the transmission dynamics in the communities of interest). This can assist with the 'risk profiling' of a community, and as such is of vital importance to the implementation and evaluation of a control scheme. Whilst a component of the World Bank strategy was the collection and testing of faecal samples prior to control (WHO, 2011), the lack of identification of potential risk factors for infection could lead to difficulties for full evaluation of the efficacy of the control scheme. As echinococcosis is a disease of communities as much as it is a disease of individuals, characterisation of the communities of interest could useful in understanding local transmission ecosystems.

Risk factor studies can be useful in the identification of relevant features of the transmission cycle of *Echinococcus* spp. in a particular study area, and are commonly based upon regression modelling techniques. However, these strategies, although useful for gaining an overall idea of risk factors for infection, can overlook some of the complex interactions and interdependencies between potential risk factors. The wide variety of different findings from these studies (many of which have been reviewed in a recent article (Otero-Abad and Torgerson, 2013)) further support these limitations in conventional risk factor studies. Therefore, rather than focussing solely on individual risk factors of interest, it may be beneficial to identify and characterise particular features of the community which may have relevance for the risk of canine infection, or which may be of concern to the implementation of a praziquantel dosing campaign. This is commonly implicitly conducted as part of the natural fieldwork process, by speaking with locals and gaining a general understanding of the local environment and livelihoods. However, describing these findings clearly and succinctly in relation to a large number of variables of potential interest is generally not possible, and as such these important findings are often either not presented in the analysis or mentioned only in relation to specific identified risk factors.

The current study uses a novel technique, multiple correspondence analysis, to characterise patterns of dog ownership in four study villages in the Alay valley of southern Kyrgyzstan, prior to the implementation of a praziquantel-based dog dosing

scheme. An attempt was also made to identify possible risk factors for canine echinococcosis by investigating associations between patterns of dog ownership and canine coproantigen and PCR positivity.

5.1.3 Multiple correspondence analysis

Multiple correspondence analysis (MCA) can be considered a method of data exploration which aims to identify relationships between a number of categorical variables in a similar fashion to the way factor analysis (FA) or principal components analysis (PCA) deal with continuous variables. It can be viewed as either a generalisation of correspondence analysis (CA), or a generalisation of PCA. The latter approach will be used for description of MCA here, but a full review of all of these techniques can be found in (Husson *et al.*, 2011).

At a conceptual level, CA can be understood as the deconstruction of a chi-square analysis, followed by the use of orthogonal rotation or transformation in order to better represent the variance in the data. If two variables are considered, with n and m categories each, an $n \times m$ contingency table of relationships between these categories can be created (as would be performed when manually conducting a chi-square test). From this, estimates of the 'row masses' can be made for each of the n categories of the row variable by dividing the marginal row frequencies by the total number of observations. The same can be done for each of the m categories of the column variable (in order to give the 'column mass'). Under the assumption of independence between the two variables, the product of any row mass and any column mass will give the expected proportion for the particular cell at the intersection. In chi-square testing, this estimate is then multiplied with the total number of observations to give the expected cell count, and the chi-square statistic is calculated as the sum of the squared differences between the observed cell counts and the expected cell counts, weighted according to the expected counts. If this same procedure is instead performed on the observed and expected *proportions* rather than the counts, the 'Pearson's mean square contingency', or ϕ^2, is estimated (which is equal to the chi-square statistic divided by the total number of observations). This can be considered

to be a measure of the overall intensity of the relationship between the two variables, or the 'total inertia' in the data. CA is based upon the identification of the contribution of each of the cells (i.e. particular combinations of row and column levels) to this total inertia.

Correspondence analysis is based upon the singular value decomposition of the matrix of standardised residuals (which are the differences between the observed and expected values for each cell in the table, and which give an indication of the magnitude and direction of each cell's deviation from independence). One way to approach this is to consider row and column profiles, which are a method of normalising the data and can be useful for identifying the contribution of the variables under investigation to the total inertia. The 'row profile' for each row can be considered as a vector of the (conditional) frequencies of column membership for that row. If each of these n vectors could be plotted together as coordinates in m–dimensional space, a geometric interpretation of the relationship between the different rows could be developed (a 'cloud' of n points). The vector of column masses represents the 'average' row profile, and therefore the point of origin of the cloud. The points (which each represent individual rows) are each weighted according to the row mass, meaning that rows containing a higher proportion of the total number of observations contribute more. The same approach can also be conducted for the columns, in order to create a cloud of m points in n-dimensional space.

The measures of departure from the independence model used in the creation of the row and column profile clouds are related to the chi-square statistic. This relationship becomes more apparent when estimates of the distance between each point (i.e. each row or column) and the cloud origin are made. In the case of the row profile cloud, this distance can be calculated as the sum of the squared differences between each row profile vector entry and the corresponding entry in the vector of column masses, weighted by the row profile. Since the vector of column masses represents the 'expected' row profile vector under an assumption of independence, this distance measure is known as the 'χ^2 distance' (d^2). Multiplication of the χ^2 distance with the weight allocated to each point (the row mass) gives a measure of the inertia of the

point. When these individual row point inertias are summed up for all rows, the total inertia (ϕ^2) is returned. As before, the same principle applies for the column profile cloud, which will give the same estimate of total inertia. The aim of MCA, as for related techniques such as PCA, is then to find the way to best represent the n-dimensional cloud of points in fewer than n dimensions whilst maintaining these distances between points. This is achieved by specifying the origin (the coordinates of the average row or column profile) as the centre of gravity (the 'barycentre') of the cloud, and creating a set of orthogonal axes around this which maximise the inertia captured, in each successive dimension.

MCA can be approached using a similar approach to CA, by creating the 'Burt matrix' which is a symmetric matrix representing all possible cross tabulations (i.e. contingency tables) for the variables under investigation, and analysing these separately. However, another way of conducting MCA is to apply the methodology described above to an indicator matrix (also known as the 'complete disjunctive matrix') of all individuals, which comprises the indicator matrices for all variables under investigation. Here, rows represent individuals, columns represent variable levels for all variables under investigation, and each cell will contain either a zero or a one – representing either presence or absence of the factor level for the individual in question. The cloud of individuals can be developed and analysed as required, and also a cloud of variable categories can be created. This presents the locations of the barycentres of individuals positive for each variable category. The barycentre of all categories within a particular variable will be equal to the point of origin of the axis.

5.2 Materials and methods

5.2.1 Samples

In May 2012, four communities in the Alay valley of southern Kyrgyzstan were visited. All occupied households in Sary-Mogol, Taldu-Suu and Kara-Kabak, and a random selection of households (approximately 25%) in Kashka'Suu were visited. For each household visited, a questionnaire was administered relating to details such as general

demographics (age, sex, occupation of interviewee), dog ownership (number of dogs currently owned, management of these dogs), dog demographics (dog age, dog sex, dog weight), and perception of echinococcosis (recent administration of praziquantel to dogs, understanding of source of human echinococcosis). Not all questions were answered by all interviewees. Of 692 households registered, a total of 329 individuals reported owning dogs, and a total of 388 dogs in total were registered. A total of 318 dog faecal samples included a subsample stored in saline buffer, and these were used for the remainder of the analysis.

5.2.2 Data processing

Sample processing was as described in chapter 2. Of the 318 samples, 23 could not be matched to an individual questionnaire (due to illegible or damaged sample labels), but were retained in the model as the village was known. Receiver-operating characteristic (ROC) curve analysis (Zweig and Campbell, 1993; Greiner *et al.*, 2000) was used on a panel of parasitologically defined dog faecal samples taken from Xinjiang province in China during an evaluation of a control scheme (van Kesteren *et al.*, 2015), and the Youden index approach (i.e. maximisation of both test sensitivity and specificity) (Youden, 1950) was used to determine the optimal cut-off point. The resultant cut-off point (OD 0.07635) gave an estimated test sensitivity of 96% and specificity of 83%, based upon the panel evaluated.

A Bayesian mixture model (described in the previous chapter) was also used to obtain risk scores for each sample, calculated from the OD of the sample and using OD data from a number of parasitologically-confirmed positive and negative samples from Xinjiang, China (van Kesteren *et al.*, 2015) in order to ensure identifiability and parameterise the linear regression component of the model. In order to account for possible differences in the distribution of OD values for positive samples between the four villages under investigation, each village was fit to a separate mixture model, with a distinct Polya tree created for each village but the same distribution of negative samples assumed for all four villages. Full code for the mixture model is provided in the appendix (A5). In order to simplify the interpretation of the model output, the

median estimates of the risk scores were extracted for each sample and used in the current study.

The number of available samples taken from each community is shown in table 5.1. Prior to analysis, the number of variables with missing data was assessed. Two variables were commonly left unanswered: "heard of hydatid disease" (only answered by 38 people – all of whom answered "no") and "dog fed by neighbours" (only answered by four people). These two variables were therefore removed from further study. All remaining variables were answered by at least 267 people. Variables were then inspected for the distribution of outcomes, and all variables with fewer than 10 responses in any category were removed, as these can contribute disproportionately to total inertia and lead to inappropriate conclusions in MCA (Husson *et al.*, 2011). This resulted in the removal of variables relating to a perceived source of human hydatid disease in other humans, food, cats, and other sources; a history of hydatid disease in the household; the burning of organs; the feeding of commercial food and scraps; and dog handling by strangers. As MCA requires categorical input, all continuous variables were categorised using biologically and demographically reasonable cutpoints (keeping the number of categories to a minimum wherever possible, as variables with more categories will tend to result in greater estimates of inertia). This process resulted in a total of 52 variables of interest, as shown in table 5.2.

Table 5.1. Numbers of samples analysed from the four study villages

Village	Number of samples	Proportion
Sary-Mogol	155	0.49
Taldu-Suu	86	0.27
Kara'Kabak	42	0.13
Kashka-Suu	35	0.11

Table 5.2. Variables considered in the risk factor modelling process. "Supp." indicates the the variable was included as a supplementary variable rather than an active variable in the MCA analysis

Variable type	Variables	Supp.
Location	Village	*
Sample	Purge sample	*
Animal ownership	Number of dogs owned in last 10 years (1, 2, 3, ≥4)	
	Number of dogs currently owned (1, 2, ≥3)	
	Sheep owned	
	Goats owned	
	Cattle owned	
	Horses owned	
	Yaks owned	
	Donkeys owned	
Dog demographics	Dog age (≤1y, 1-3y,3-4y, ≥5y)	
	Dog size (small, medium, large)	
	Dog weight (≤10kg, 10-20kg, >20kg)	
	Dog sex	
	Hunting dog	
	Guard dog	
	Pet dog	
	Sheep dog	
Dog management	Dog wormed in last six months	
	Dog known to eat rodents	
	Dog fed meat	
	Dog fed offal	
	Dog chained (always, day only, never)	
	Dogs handled by adults in the household	
	Dogs handled by children in the household	
	Dogs handled by friends of the family	
	Dogs not handled	
	Dog visited Jailoo (summer pasture) previous year	
	Dog will visit Jailoo this year	
Animal slaughter	Slaughter own animals	
	Slaughter other people's animals	
	Organs from slaughtered animals thrown away	
	Organs from slaughtered animals given to dogs	
	Organs from slaughtered animals buried	
Knowledge about human echinococcosis	Dogs perceived source of hydatid disease	
	Livestock perceived source of hydatid disease	
	Source of hydatid disease not known	
Diagnostic test results	ELISA status (pos/neg)	*
	ELISA OD value	*
	Median Bayesian mixture model score	*
	E. granulosus G1 PCR status(pos/neg)	*
	E. canadensis G6 PCR status(pos/neg)	*
	E. multilocularis PCR status(pos/neg)	*

5.2.3 MCA model

MCA was conducted using the package "FactoMineR" version 1.29 (Lê *et al.*, 2008; Husson *et al.*, 2015). Village, purge status and diagnostic test results were entered as supplementary variables into the model, and therefore did not contribute to the output themselves. All remaining variables were entered into the model as 'active' variables, and therefore were used in the construction of the MCA dimensions.

MCA was initially used to visualise the pattern of missing data (which was included as a separate level for each variable with missing data). In order to remove these missing data levels from the analysis, the "estim_ncpMCA "procedure in the R package "missMDA" (Husson and Josse, 2014) was used to identify the optimal number of dimensions from which to compute the missing values. As this procedure indicated that zero dimensions should be used to impute missing data, the "imputeMCA" procedure in the same package was used to construct a new disjuntive matrix by replacing missing data with the proportion of positive responses for the category in question (Josse *et al.*, 2012).

MCA was run using the adapted disjunctive matrix, and the barchart of eigenvalues for sequential dimensions was visually inspected in order to identify a reasonable number of dimensions to retain for interpretation. Initial interpretation was conducted on a visual basis, using scatter plots of two consecutive dimensions in pairs. Firstly, a 'cloud of variables' (Husson *et al.*, 2011) was created based upon the correlation ratioestimate for individual categories within that variable in relation to the dimension scores. This allowed identification of potential variables of importance to the dimensions. A 'cloud of categories' was then created, which represents the output of the process described earlier, and represents the barycentres of those individuals positive for that category for each of the dimensions. To aid interpretation of these graphs, only those variable levels which were most strongly associated with the dimension were represented by selecting points according to their squared cosine coefficient (cos^2). The cos^2 can be considered to be a measure of the correlation between individual points and the dimension in question: the name is derived from the geometric properties of MCA

output. All levels of active variables with a cos^2 estimate of 0.1 or less were not labelled in the graph or interpreted in the model output. Once an idea was obtained of what the different dimensions represented, the relationship between dimensions and the supplementary variables was investigated graphically using the same approach, including all variable levels regardless of cos^2.

Quantitative interpretation of the MCA output was achieved using the "dimdesc" procedure in FactoMineR, and results were only presented for active variables for which there was reasonable evidence of an association with the dimension in question ($p<0.05$). Output was presented for all supplementary variables due to the exploratory nature of this part of the study and the relatively low number of variables investigated. A number of estimates were obtained. For each variable, the correlation ratio was estimated and a one-way analysis of variance was used to identify significant associations. For particular variable levels, the mean coordinates on the dimension in question for individuals positive for the level in question were estimated and compared to the mean coordinates overall using a t-test. This procedure was repeated for both active and supplementary variables. For quantitative supplementary variables, the correlation coefficient between individual scores and the variable was estimated.

5.3 Results

As expected, the distribution of OD values from these samples showed a clear right skew, as shown in figure 5.1. Of the 318 samples included, 78 (25%) were classified as being coproantigen positive, with the distribution of positivity between villages detailed in table 5.3. The predicted distributions of samples from the four villages from the Bayesian mixture model are shown in figure 5.2, and estimates of the prevalence of infection for the four villages taken from the model are shown in table 5.3. The Bayesian mixture model scores ranged from 0.04 to 7.61, and were highly overdispersed, as is shown in figure 5.3.

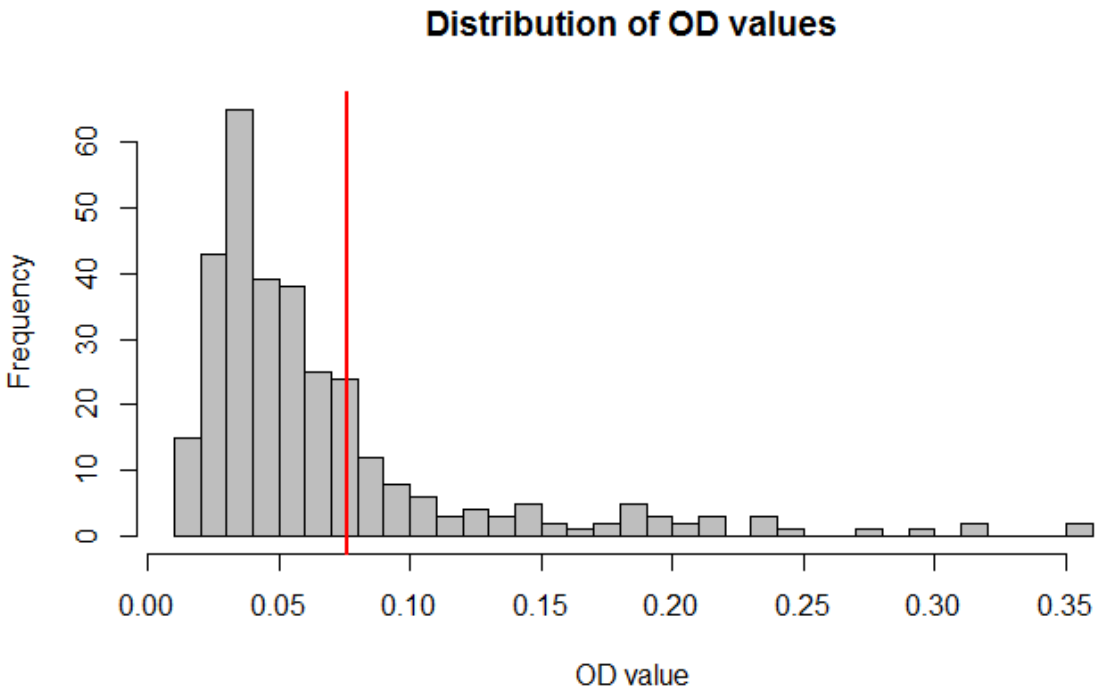


Figure 5.1. Distribution of OD values for all samples tested (n=318). The red line indicates the cut-off for positivity.

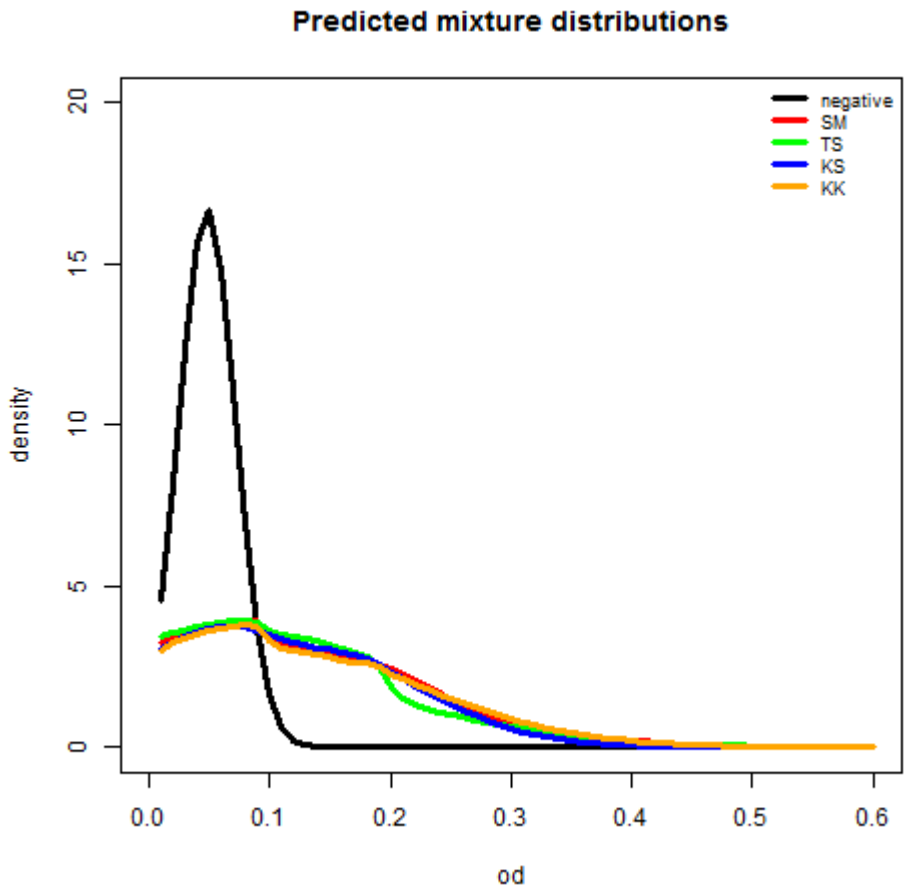


Figure 5.2 Predicted distribution of mixture model components for the four villages

Table 5.3. Coproantigen prevalence/prevalence estimates for the four study villages

Village	Coproantigen cutoff	Mixture model (mode and HDI)
SM	42/155 = 27%	14% (9-22%)
TS	16/86 = 19%	9% (4-18%)
KS	10/42 =24%	13% (5-28%)
KK	10/35 = 29%	15% (6-31%)

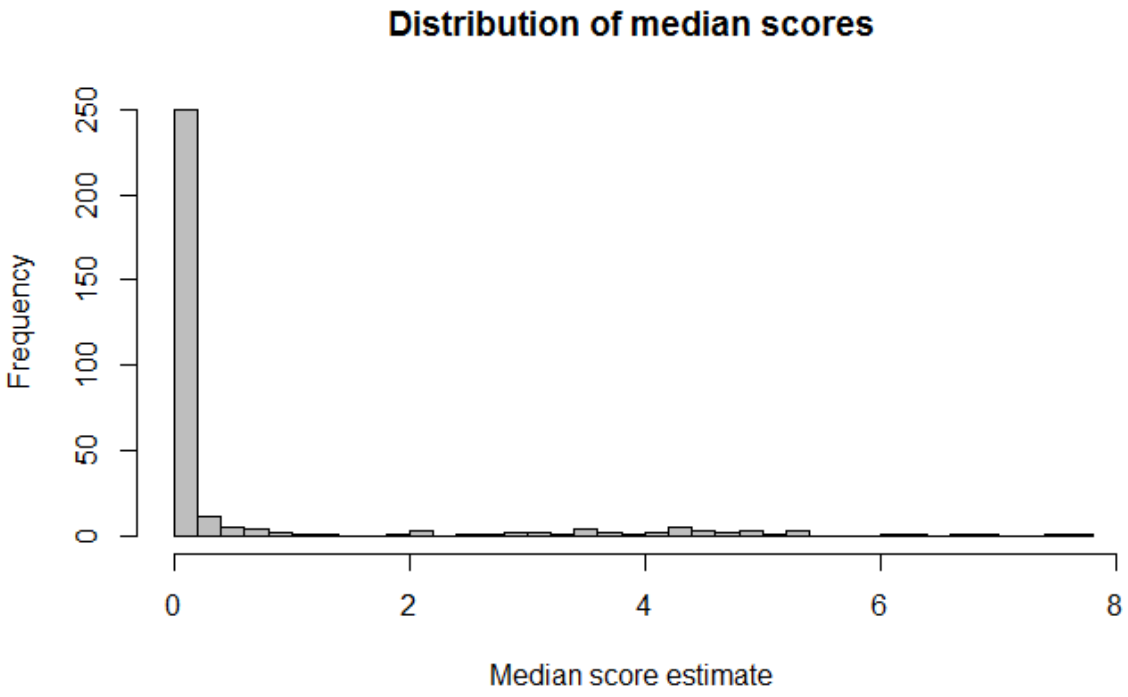


Figure 5.3 Distribution of median score estimates from the mixture model

As is expected for MCA analysis, the eigenvalues of the dimensions were all relatively low, as shown in figure 5.4. A decision was made to select the first four dimensions for further interpretation, as after this there was a sudden drop in eigenvalue estimates. The estimate of total variance explained by these four dimensions was 27% (although this is likely an underestimate, as is usually seen when MCA is conducted on a disjunctive matrix (Husson *et al.*, 2011)).

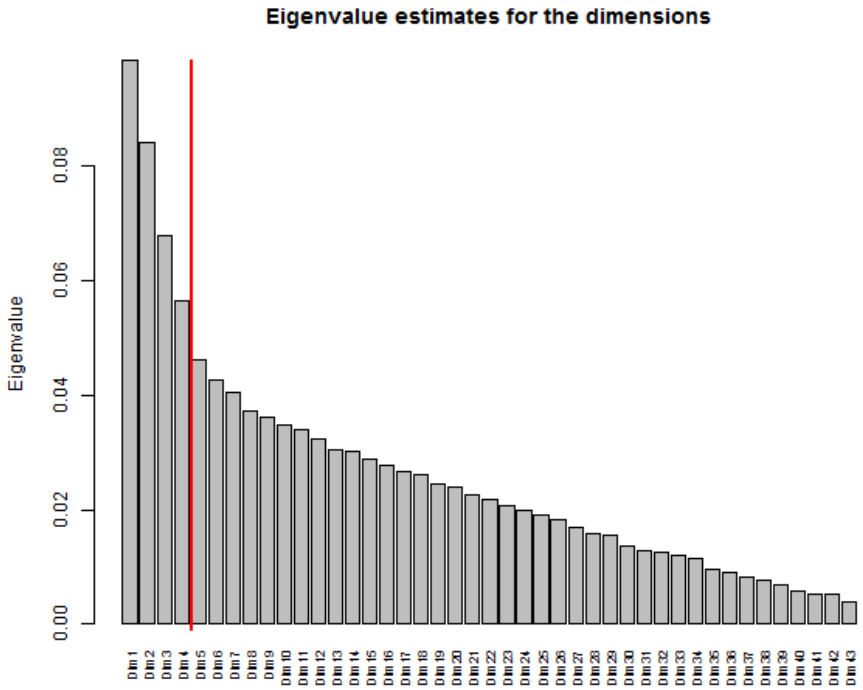


Figure 5.4 Eigenvalue estimates for the dimensions created in MCA. The red line indicates the cut-off for dimensions to interpret

Table 5.4. Eigenvalues and inertia ('variance') explained by the first five dimensions

Dimension	Eigenvalue	Percentage of variance	Cumulative percentage of variance
1	0.10	8.77	8.77
2	0.08	7.48	16.25
3	0.07	6.04	22.29
4	0.06	5.02	27.31

Scatterplots of the individual, variable and category clouds of the first four dimensions are shown in figures 5.5-5.9. Estimates of dimension scores and cos^2 estimates (for variables with a cos^2 of greater than 0.1) are shown in tables 5.5-5.8. All p-values for the association between these variables and the dimension were less than 0.001, and so are not shown. Tables 5.9 and 5.10 show either cos^2 (for categorical variables) or correlation coefficients (for continuous variables), dimension scores and p-values for the supplemental variables.

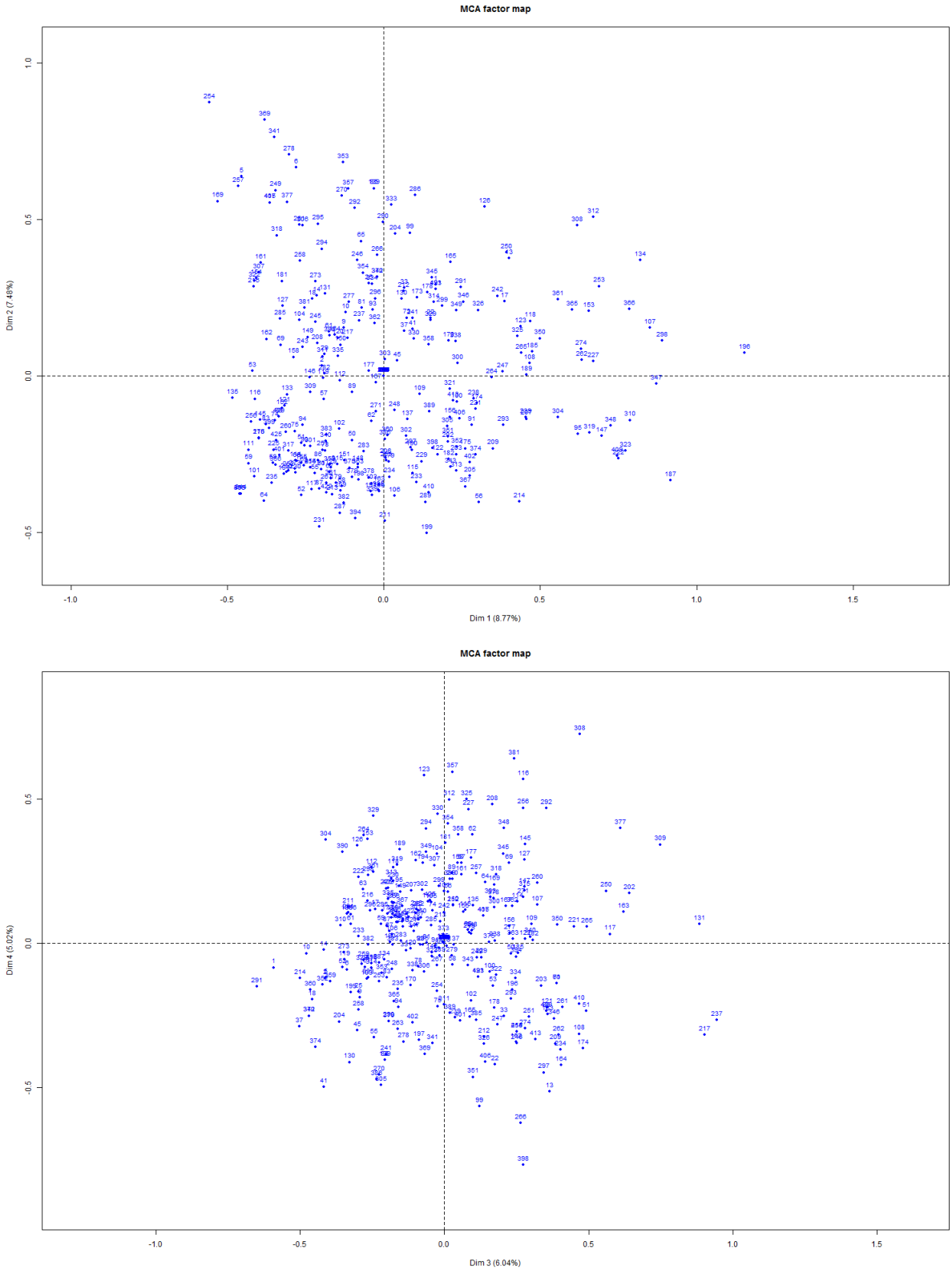


Figure 5.5. Individual clouds for the first four dimensions

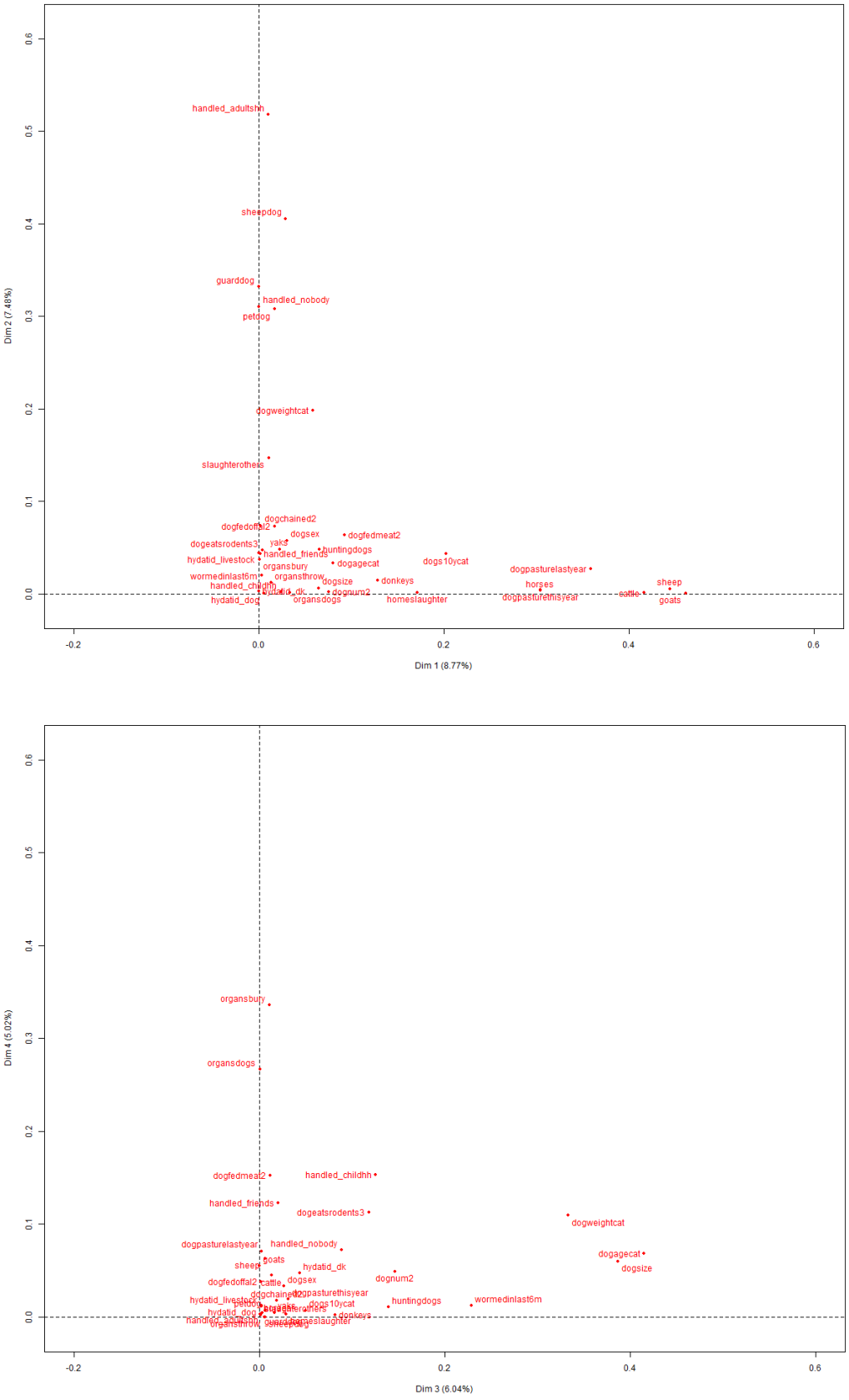
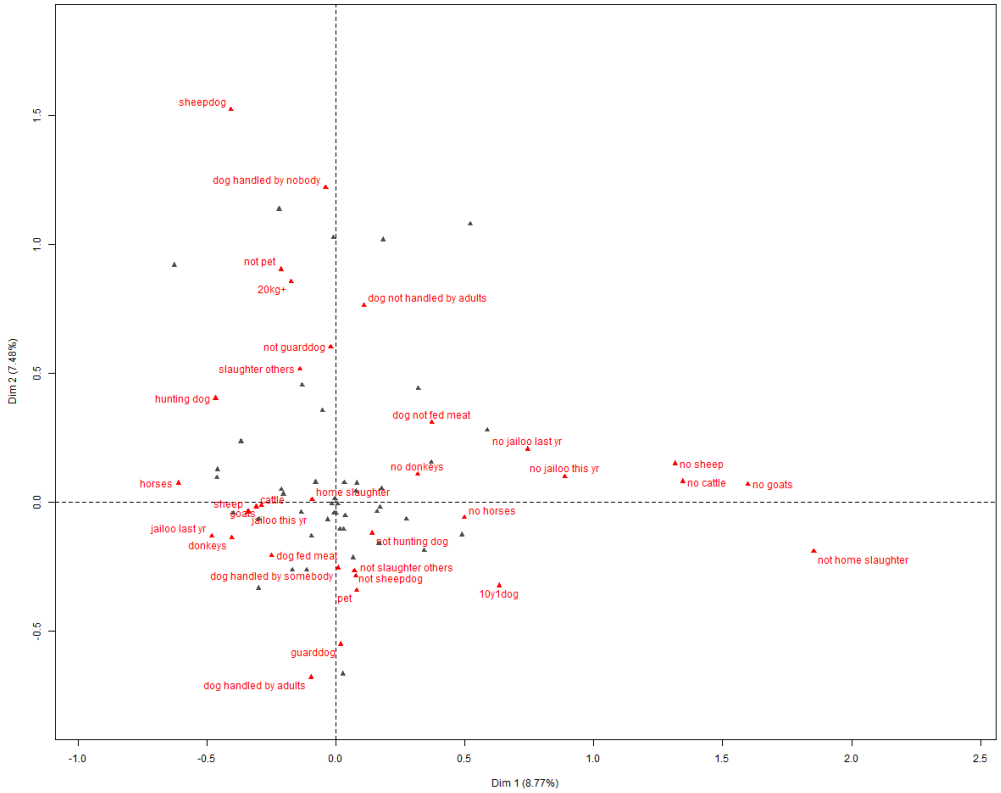


Figure 5.6. Variable clouds for the first four dimensions



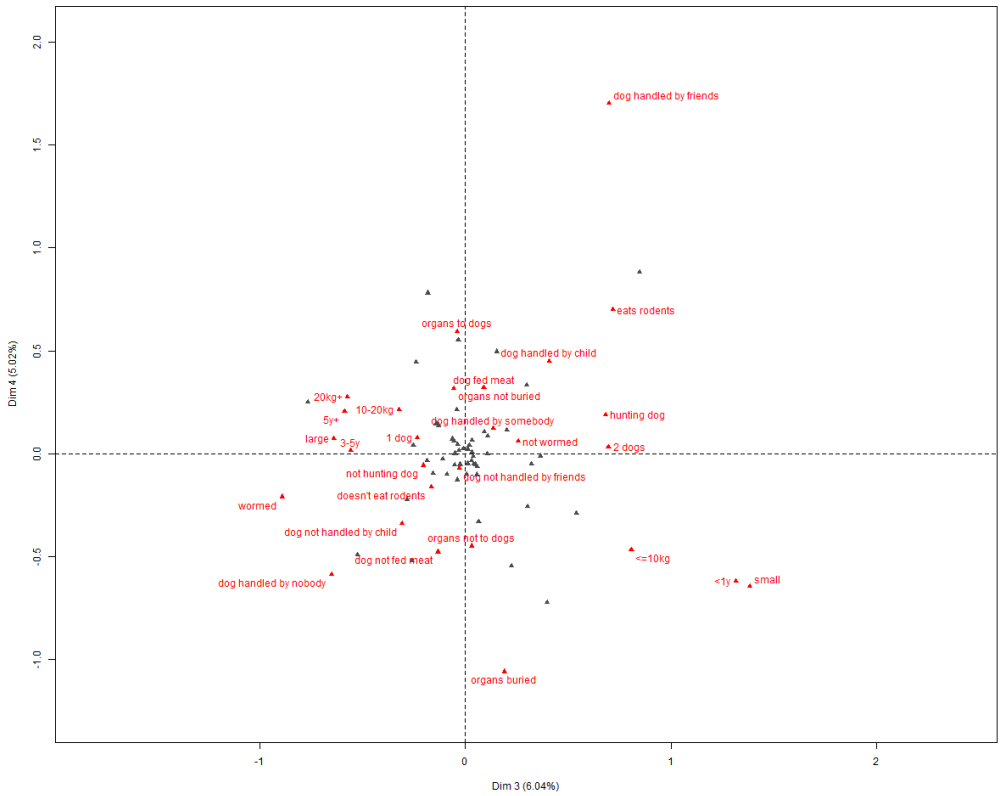


Figure 5.7. Category clouds for the first four dimensions. Labels are not included for variables with a *cos*^{2} estimate of 0.1 or less (grey points)

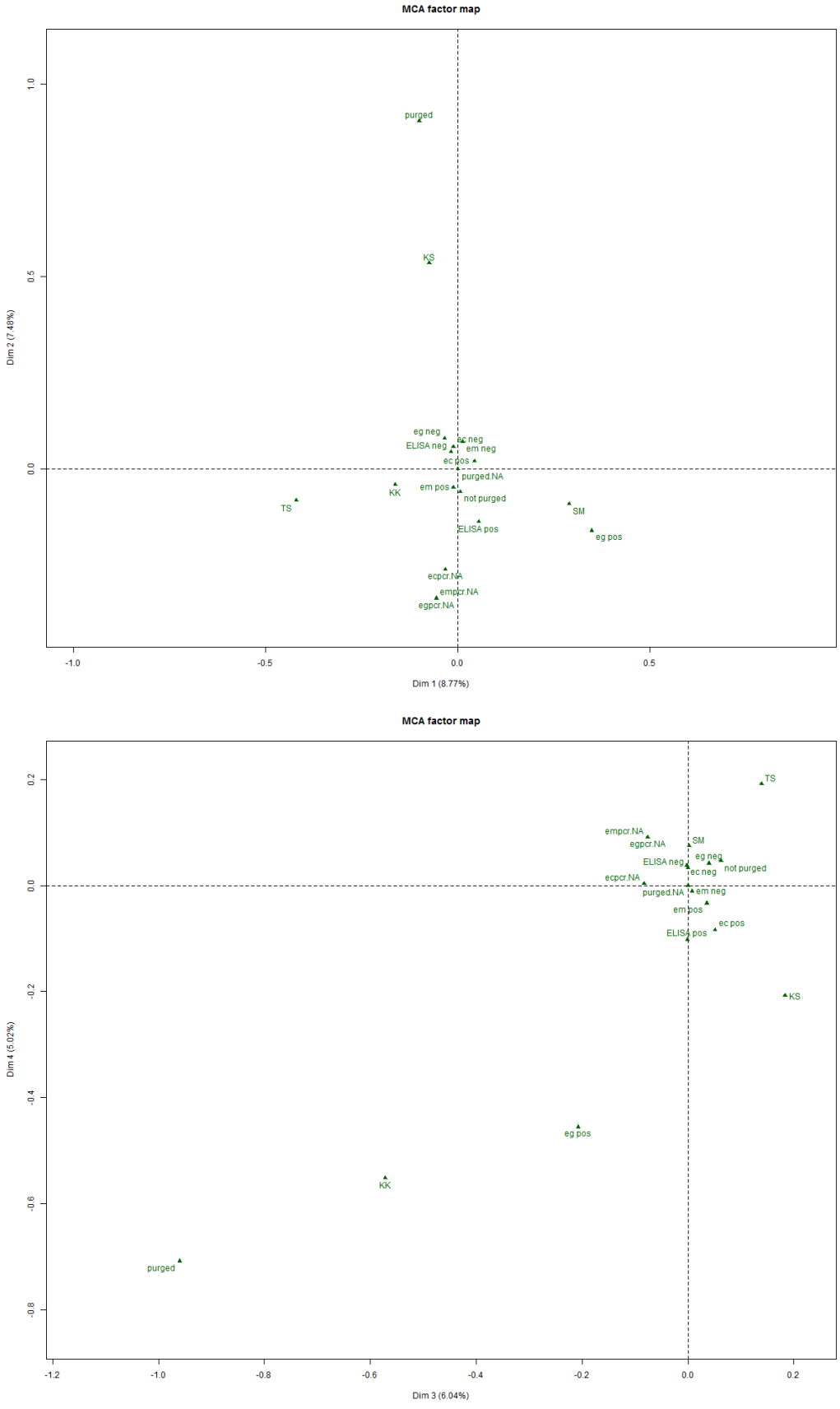


Figure 5.8. Supplementary variable category clouds for the first four dimensions

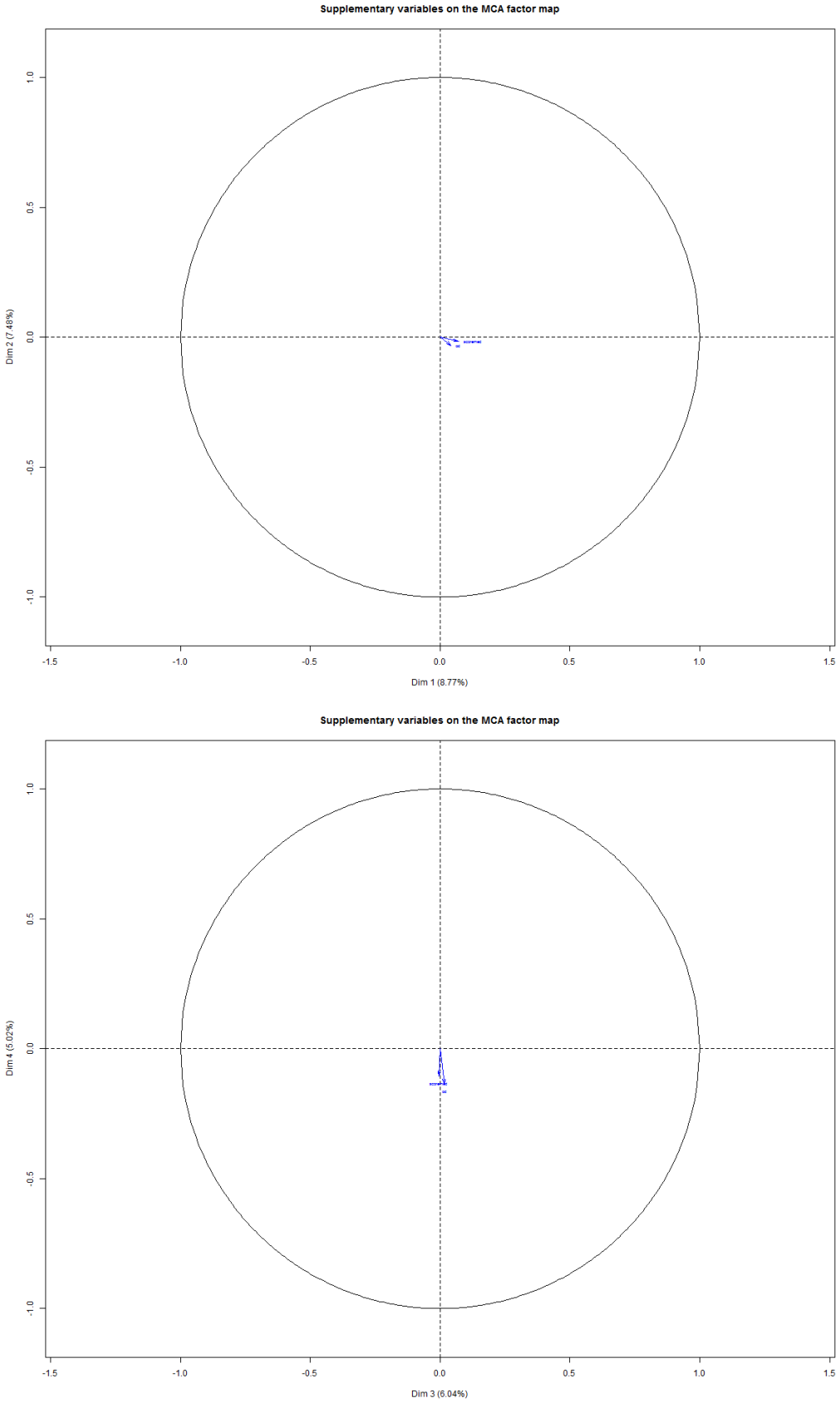


Figure 5.9. Correlation circles for supplementary quantitative variables for the first four dimensions

Table 5.5. Variable categories associated with dimension 1

Category	Dimension score	cos^2
goats	-0.25	0.50
home slaughter	-0.23	0.18
sheep	-0.23	0.48
cattle	-0.23	0.45
horses	-0.20	0.33
jailoo last year	-0.19	0.39
jailoo this year	-0.18	0.33
donkeys	-0.13	0.14
dog fed meat	-0.11	0.10
dog not fed meat	0.11	0.10
no donkeys	0.12	0.14
no horses	0.18	0.33
10y1dog	0.21	0.13
no jailoo last year	0.22	0.39
no jailoo this year	0.24	0.33
no sheep	0.33	0.48
no cattle	0.34	0.45
no goats	0.39	0.50
not home slaughter	0.43	0.18

Table 5.6. Variable categories associated with dimension 2

Category	Dimension score	cos^2
dog handled by adults	-0.22	0.56
not sheepdog	-0.21	0.44
dog handled by somebody	-0.18	0.33
guard dog	-0.18	0.36
pet	-0.17	0.33
not slaughter others	-0.11	0.16
slaughter others	0.14	0.16
not guarddog	0.18	0.36
not pet	0.22	0.33
dog not handled by adults	0.23	0.56
20kg+	0.24	0.23
dog handled by nobody	0.28	0.33
sheepdog	0.34	0.44

Table 5.7. Variable categories associated with dimension 3

Category	Dimension score	cos^2
large	-0.24	0.25
wormed	-0.19	0.25
3-5y	-0.19	0.13
20kg+	-0.17	0.10
1 dog	-0.13	0.15
not hunting dog	-0.10	0.15
doesn't eat rodents	-0.10	0.13
dog not handled by child	-0.10	0.14
dog handled by child	0.11	0.14
not wormed	0.13	0.25
2 dogs	0.14	0.13
hunting dog	0.15	0.15
eats rodents	0.15	0.13
<=10kg	0.24	0.37
small	0.36	0.31
<1y	0.37	0.33

Table 5.8. Variable categories associated with dimension 4

Category	Dimension score	cos^2
organs buried	-0.21	0.36
dog not handled by friends	-0.16	0.13
organs not fed to dogs	-0.13	0.29
<=10kg	-0.11	0.12
dog not fed meat	-0.10	0.17
dog not handled by child	-0.10	0.16
doesn't eat rodents	-0.08	0.12
dog fed meat	0.10	0.17
dog handled by child	0.11	0.16
organs fed to dogs	0.14	0.29
eats rodents	0.14	0.12
organs not buried	0.14	0.36
dog handled by friends	0.30	0.13

Table 5.9. Associations between the first four MCA dimensions and the categorical supplementary variables. Variables with a t-test p-value of less than or equal to 0.05 are shown in blue. E.g = *E. granulosus* G1; E.c = *E. canadensis* G6; E.m = *E. multilocularis*

Category	Dimension 1			Dimension 2			Dimension 3			Dimension 4		
	cos^2	score	p	cos^2	score	p	cos^2	score	p	cos^2	score	p
SM	0.08	0.12	<0.01	0.01	-0.05	0.12	0.00	0.02	0.97	0.01	0.05	0.20
TS	0.07	-0.10	<0.01	0.00	-0.05	0.38	0.01	0.05	0.13	0.01	0.07	0.04
KS	0.00	0.01	0.61	0.04	0.13	<0.01	0.01	0.06	0.20	0.01	-0.02	0.15
KK	0.00	-0.02	0.31	0.00	-0.04	0.80	0.04	-0.13	<0.01	0.04	-0.10	<0.01
ELISA(-)	0.00	-0.01	0.58	0.01	0.03	0.16	0.00	0.00	0.99	0.00	0.02	0.30
ELISA(+)	0.00	0.01	0.58	0.01	-0.03	0.16	0.00	0.00	0.99	0.00	-0.02	0.30
E.g PCR(-)	0.00	-0.04	0.27	0.02	0.06	0.01	0.01	0.03	0.20	0.01	0.04	0.18
E.g PCR(+)	0.01	0.08	0.04	0.00	-0.01	0.35	0.00	-0.03	0.23	0.02	-0.08	0.01
E.c PCR(-)	0.00	0.00	0.81	0.00	0.03	0.22	0.00	0.00	0.97	0.00	0.01	0.43
E.c PCR(+)	0.00	0.01	0.64	0.00	0.02	0.83	0.00	0.02	0.58	0.00	-0.02	0.37
E.m PCR(-)	0.00	0.01	0.71	0.01	0.05	0.04	0.00	0.00	0.82	0.00	-0.01	0.75
E.m PCR(+)	0.00	0.00	0.94	0.00	0.02	0.75	0.00	0.01	0.81	0.00	-0.01	0.82
Purge(-)	0.00	-0.02	0.66	0.02	-0.10	0.01	0.03	0.09	<0.01	0.01	0.06	0.03
Purge(+)	0.00	0.01	1.00	0.05	0.18	<0.01	0.06	-0.17	<0.01	0.03	-0.12	<0.01

Table 5.10. Associations between the first five MCA dimensions and the continuous supplementary variables

Variable	Dimension 1		Dimension 2		Dimension 3		Dimension 5	
	correlation	*p*	*correlation*	*p*	*correlation*	*p*	*correlation*	*p*
OD	0.04	0.45	-0.03	0.56	0.02	0.76	-0.14	0.02
Score	0.07	0.21	-0.02	0.78	0.00	0.94	-0.11	0.06

5.4 Discussion

This chapter describes a strategy for the investigation of types of households and dog owners in the Alay valley of southern Kyrgyzstan. It is intended to demonstrate the use of multivariate techniques in the integration of datasets containing relatively large numbers of variables of potential interest – such as those obtained at the start of a study, when little is known of the study area. As echinococcosis is a community-level problem, methods which can identify community-level risk factors of potential importance can be useful for planning and assessing control and intervention schemes. Whilst this is no substitute for local knowledge gained by speaking with and involving local people, it offers an additional strategy for the exploration and analysis of complex situations, and may also be useful for combination with risk factor investigations. Multiple correspondence analysis identified four different dimensions of dog types, as summarised in table 5.11.

Table 5.11. Basic description of first five dimensions extracted from MCA

Dimension	Low values	High values
1	Livestock ownership; visits Jailoo; fed meat; owners slaughters animals	No livestock; only owned one dog in last ten years; doesn't visit Jailoo; not fed meat; owner doesn't slaughter animals
2	Guard dogs and pets; handled; owners don't slaughter other people's animals	Sheepdogs; not handled; greater than 20kg; owners slaughter other's animals
3	Older, larger dogs from single dog households; received praziquantel; not handled by children; not used for hunting; not seen eating rodents	Younger, smaller dogs from two dog households; hasn't received praziquantel; handled by children; used for hunting; seen eating rodents
4	Owners bury organs rather than give to dog; small dogs; not handled; not known to eat rodents	Owners feed dog organs; known to eat rodents; handled

5.4.1 Comparison of prevalences

As shown in table 5.3, there were considerable differences in the estimated prevalences between the ROC curve analysis (using a single cut-off) and the mixture model (which estimated the prevalence as the modal weight of the positive component). The prevalence estimates using the cut-off were generally double those estimated from the mixture model, which likely represents limitations in the test specificity using this particular, low, cut-off (i.e. false positives). However, as the mixture model has not yet been comprehensively validated, it is also possible that the mixture model predictions are inaccurate. Interestingly, when the ROC curve approach was repeated classifying dogs with low burdens (<50 worms) as negative (see chapter 6), the estimated prevalences were similar to those obtained from the mixture model. This would suggest that the mixture model strategy here is underrepresenting the true prevalence. There was no clear evidence of any particular differences in prevalence between villages, and the distribution of the positive components of the mixture model for the four villages were also all very similar (figure 5.2).

5.4.2 Interpretation of dimensions

The first dimension represents differences in animal ownership, and distinguishes dogs in livestock-oriented households from those in households which do not keep livestock. As expected, only those dogs from households with livestock visit Jailoo (as the purpose of visiting Jailoo is to graze livestock) and slaughter their own animals.

The second dimension represents differences in dog use, and distinguishes sheepdogs (which are generally heavier in weight) from pets and guard dogs. There does not appear to be a clear distinction between pets and guard dogs (van Kesteren *et al.*, 2013), which may indicate overlapping responsibilities (or lack thereof) for these animals. These dogs tend to be more commonly handled by their owners, and come from households which tend to not be involved in slaughtering other people's livestock.

The third dimension represents differences in dog demographics, and distinguishes older dogs from single dog households from younger dogs (possibly new acquisitions) in multi-dog households. The older dogs are not handled by children but tend to have recently received praziquantel, whereas younger dogs are more likely to have been observed eating rodents and are commonly handled by children. As these dogs were also less likely to have been wormed, this could indicate an increased risk of transmission of *Echinococcus* spp.

The fourth dimension appears to be related to aspects of owner's knowledge of echinococcosis, and distinguishes dogs whose owners appear to know about echinococcosis and act accordingly (avoids feeding offal to dogs, avoids dog contact), from those whose owner does not. Alternatively, this may represent differences in level of dog management –with low values indicating people who take little interest in their dog. These two possible explanations would be worthy of further investigation, since they would be expected to result in different risks of dog infection. If this dimension does represent owner's knowledge of echinococcosis, the fact that this is acted upon is very promising, and suggests that education campaigns (whether formal or informal) are having an effect on these households. Dogs from these households tend to be smaller but not necessarily younger (contrasted with the smaller dogs identified in dimension 3), which may indicate underfeeding or breed differences. Dogs from households with less apparent echinococcosis knowledge were also more commonly reported to eat rodents than those with more knowledge. Given that there is probably little owner control over this behaviour, this may indicate response bias in the case of households with more knowledge (i.e. they may pretend to not observe these risky behaviours in their own dogs), or may represent differences in owner interaction with dogs.

5.4.3 Investigation of associations with supplementary variables

The next stage of analysis was the investigation of associations between the four identified dimensions and the supplementary variables. Although the supplementary variables were found to be significantly associated with dimensions in some cases,

their correlation with the dimensions (as measured using cos^2) were invariably low (less than 0.1). Despite this, it may still be possible to extract some potentially useful information from these associations, which may be worthy of further investigation.

Sary-Mogol was clearly distinguished from Taldu-Suu on dimension 1, which suggested that dogs from Sary-Mogol tended to be from less livestock-oriented households than those in Taldu-Suu. This feature was evident from personal observation during household visitation. Dogs which tested positive for *E. granulosus* G1 by PCR had significantly higher scores on dimension 1, which may suggest a direct relationship with households which did not own their own livestock and did not visit Jailoo. One possibility cause of this association is dog type, which will be discussed below in relation to dimension 2. Another possible reason for this association is visitation of Jailoo. Investigation of the potential risk associated with travel to "summer pastures" such as this has produced varied outcomes. A study of *E. granulosus* in Narenhebuke in Xinjiang, China, suggested that dogs at summer pasture had a lower coproantigen prevalence than those in winter pasture (Wang *et al.*, 2001). However, since these samples were collected at different times of the year, there is a possibility for confounding due to inherent seasonality in infection. Conversely, it has been suggested that summer pasture presents a focus of *E. multilocularis* transmission in Kazakhstan (Rysmukhambetova *et al.*, 2004). If the risk of canine infection with *E. granulosus* G1 is reduced when visiting the Jailoo, then those dogs which remain in the villages over the summer period may have a relatively higher prevalence of infection than those which visit Jailoo. Since the life expectancy of *E. granulosus* can be in the order of 6-20 months (Harris *et al.*, 1980), then it is plausible that worms could still remain at the time of visitation (May). A final possibility is that of socioeconomic status. It has been suggested that travel to Jailoo is not economically feasible for poorer livestock-owning families (Farrington, 2005; Kerven *et al.*, 2012), and a complete lack of ownership of livestock may predominantly identify particularly poor families. Therefore, the identified association may also be representative of socioeconomic factors, which may have an impact upon canine infection.

Dimension 2 distinguished Kashka'Suu from the other villages, suggesting that dogs from this village were more likely to be described as sheepdogs (rather than guard/pet dogs). As would be expected, sheepdogs were also represented by low scores on dimension 1, but the location of Kashka'Suu on dimension 1 was not noticeably different from the barycentre of this dimension (which therefore suggests that this village had an 'average' level of livestock ownership/Jailoo visitation). Another possible explanation for this association with dog types is that houses in Kashka'Suu were generally of a higher build quality than those in other villages, which could mean that guard /pet dogs were not needed for protection of possessions. The variable indicating whether dogs were purged was also associated with this dimension, and suggested that purged dogs were more likely to be identified as sheepdogs. Purging was only conducted in two villages – Taldu-Suu and Kara-Kabak – but selection of dogs was mediated through the local government and private veterinarian, respectively. This association, especially given that these villages did not score highly on dimension 2 *per se*, possibly suggests a possible selection bias in favour of sheepdogs. This may be due to closer relationships between the veterinarians and sheepdog owners than owners of other dogs, or may indicate relative availability of people when dogs were needed.

PCR negativity for both *E. granulosus* G1 and *E. multilocularis* were associated with higher scores on dimension 2, suggesting that sheepdogs were less likely to be infected with these species. This is unexpected, and differs from previous studies which have commonly identified sheepdogs or farm dogs as having a higher probability of coproantigen positivity or infection than non-sheepdogs (Moro *et al.*, 1999; Shaikenov *et al.*, 2003; Torgerson *et al.*, 2003c; Buishi *et al.*, 2005b). However, this result does agree with the association between dimension 1 and *E. granulosus* G1 PCR positivity, since dogs in households which do not own sheep are unlikely to be described as sheepdogs. One possible reason for this association is that sheepdogs are more highly valued than guard/pet dogs, and therefore are better fed. This would also correlate with the high scores on dimension 2 for larger dogs. Another possibility is that sheepdogs may be more commonly in work, herding sheep, and therefore may be less likely to roam through the village in search of food. One other possibility is the

complex interplay between age and dog type. Dimension 2 identified sheepdogs as well as larger dogs (>20kg). As mentioned above, a previous study in Kazakhstan found that farm dogs had much higher infection pressures for *E. granulosus* than village dogs (Torgerson *et al.*, 2003c). However, it was suggested that this provoked immunity in these dogs, and resulted in a reduction in worm burdens amongst older individuals. In the absence of immunity (as was suggested for village dogs), the worm burden increased to a plateau as age increased. Although the burdens themselves still appeared to remain higher in farm dogs than village dogs in this case, the possibility of an age-related reduction in burden in the face of high infection pressure cannot be excluded. Further work to characterise age-related trends in prevalence and/or coproantigen levels amongst dogs which visit Jailoo would be worthy of further investigation.

Dimension 3 differentiated dogs from Kara-Kabak from those from the other villages, and suggested that dogs from Kara-Kabak tended to be older dogs in single dog households. This suggests that the replacement rate for dogs in this community was lower than in the others, which may relate to previous culling campaigns. Although no direct questions were asked about previous culling campaigns during this visit, these campaigns were commonly reported in the villages of Sary-Mogol and Taldu-Suu over the period 2012-2013, and in Kashka'Suu in 2014. This result therefore suggests that widespread culling campaigns are not being implemented in Kara-Kabak (and could also indirectly suggest that these campaigns in the other villages ultimately only result in the replacement of culled dogs with new dogs). As low scores on this dimension were also associated with recent praziquantel dosing, this may suggest that praziquantel dosing is predominantly being used in Kara-Kabak rather than culling. The reason for this is unclear. However, one main difference between Kara-Kabak and the other three villages was that there was no resident government veterinarian in this community, although a private veterinarian remained. Culling campaigns appeared to be generally implemented based on governmental advice (Akjol Tagaibekov, personal communication), and therefore the lack of a government veterinarian in the community may be a possible reason for reduced culling in Kara-Kabak. The

association between this dimension and purged dogs is likely to result from the fact that many of the purge samples were collected from dogs in Kara-Kabak.

Dimension 4 differentiated Taldu-Suu from Kara-Kabak and suggested a possible difference in knowledge of echinococcosis (with associated preventive action), or differences in attitude towards dogs, between these villages. As mentioned above, this may be suggestive of more effective education campaigns in Kara-Kabak than in Taldu-Suu. The finding that Sary-Mogol had a similar (although nonsignificant) positive score to Taldu-Suu and that Kashka'Suu had a similar negative score to Kara-Kabak may be suggestive of differences in educational campaigns between those villages in Alay district and those in Chon-Alay. Kashka'Suu and Kara-Kabak, being located in Chon-Alay, are geographically close to their district centre, Daroot-Korgon (which lies just 50km to the east in the Alay valley), whereas the district centre of Alay district is Gulcha – located in the Alay mountains around 130km to the northwest of Taldu-Suu (and which requires traversing a mountain pass to reach). Education campaigns regarding zoonoses in Kyrgyzstan are largely based on communication between state veterinarians and livestock owners (Stammbach, 2009), and therefore require reciprocal trust and respect (which has been a problem since independence and a loss of control of livestock diseases). The relative isolation of Taldu-Suu (and Sary-Mogol) from their district centres could have repercussions on the implementation and delivery of educational campaigns (which are an important component of any control scheme (Craig and Larrieu, 2006)).

However, an alternative explanation for these differences in dimension 4 may be that people in Kara-Kabak are less involved with looking after their dogs than those in Taldu-Suu. This possibility would be supported by the finding of an association between low values on this dimension and *E. granulosus* G1 PCR positivity. The correlation coefficient for OD value was -0.14, suggesting that lower scores on this dimension were associated with higher OD values. This finding, combined with the PCR results, may also be suggestive that the causative agent responsible for the coproantigen ELISA results is *E. granulosus* G1. One possible explanation for this association is that reduced owner involvement with dogs may result in underfeeding

and therefore increased scavenging behaviour (even though a feature of this dimension was a reduced feeding of offal to dogs). An effective educational campaign leading to these behavioural patterns would be expected to be associated with a reduced, rather than an increased, risk of infection as has been identified in previous studies (Buishi *et al.*, 2005b; Huang *et al.*, 2008). One other possibility worthy of mention is reverse causality. Although canine infection with *E. granulosus* is asymptomatic and there are no reports of large numbers of human cystic echinococcosis cases in the area, cystic echinococcosis in intermediate hosts would be expected to be identifiable during slaughter (especially amongst the older animals more commonly slaughtered in these areas). Therefore, it is possible higher levels of infection with *E. granulosus* in the community are in fact the driver for the apparent increased knowledge. However, this is likely to require some degree of education: a study in Morocco found that whilst ruminant cysts were very commonly identified, their association with infection in dogs or humans was invariably unknown (Kachani *et al.*, 2003).

5.4.4 Diagnostic test interpretation

A major aim of the current study was to attempt to identify relationships between the different test results. Latent class methods are commonly used to simultaneously interpret the results of different diagnostic test results, and can reduce biases in overall prevalence estimation as well as assessing test performance (Hui and Walter, 1980; Johnson *et al.*, 2001; Toft *et al.*, 2005; Ziadinov *et al.*, 2008; Hartnack *et al.*, 2013). This is particularly true for cases where the tests measure different outcomes (in this case, coproantigens and DNA), which in theory makes coproantigen and coproPCR data well-suited for latent class analysis. Although a total of four tests have been applied here, the three PCR tests are measuring different outcomes, and therefore cannot be compared with each other using normal latent class approaches. Therefore, effectively, three groups of two tests (the coproantigen test and each one of the PCR tests) were available. In this situation, the latent class model is not identifiable when conducted on a single population (Johnson *et al.*, 2001), and it is preferable to have at least three distinct populations (Toft *et al.*, 2005).

Although previous studies have stratified datasets into populations with different prevalences according to identified risk factors for positivity (Ziadinov *et al.*, 2008), this may violate the assumption of fixed test sensitivity and specificity in all situations. As it is well known that a relationship between the optical density value of the coproantigen ELISA and the worm burden exists (Deplazes *et al.*, 1992; Raoul *et al.*, 2001; Reiterová *et al.*, 2005; Buishi *et al.*, 2005b), this means that the test sensitivity would be expected to be lower in a population where the mean burden is low than that in a population with a high mean burden (Allan *et al.*, 1992; Reiterová *et al.*, 2005). An alternative approach is to include covariates with known relationship with the outcome in the model (Hartnack *et al.*, 2013). However, no clear risk factors for positivity were identified in the current study (data not shown), and this approach was therefore not attempted. Instead, multivariable methods were in the hope that in cases where ELISA OD values were high due to a particular species of *Echinococcus*, dimension scores for ELISA positivity (and/or ELISA OD/score) would be similar to those for PCR positivity for that species. This association was only observed in the current study for dimension 4, where a clear association with both coproantigen ELISA OD and *E. granulosus* G1 PCR status was found, and which may suggest that the increased OD values were due to increased *E. granulosus* G1 prevalence. The lack of association with the dichotomised ELISA results is understandable, as dichotomisation is known to reduce study power (Altman and Royston, 2006), and demonstrates the potential benefits for interpretation of coproELISA data in a continuous rather than dichotomised fashion (see chapter 4).

One alternative approach potentially worthy of possible further investigation is the application of MCA to the test results directly (i.e. the inclusion of test results only as active variables in the MCA). In the current case, this would rearrange the four test 'dimensions' (one ELISA and three PCR statuses) into four dimensions ranked in order of inertia explained. These results could be interpreted in relation to variables of interest, included as supplementary variables. If a smaller number of these dimensions were found to be useful in representing the original data (measured according to

amount of inertia explained), these dimension scores could also be allocated to each individual (see below) and further analysis conducted.

5.4.5 Individual cloud interpretation

The analysis conducted here has focussed only on the interpretation of the 'category cloud', which describes relationships between the dimensions and categories of variables of interest. This is related to the 'individual cloud' of all individual dogs (figure 5.5), and this individual cloud itself may be worthy of further study. Although data exploration and identification of possible community-level associations was the main aim of the current study, one other use of MCA is the investigation of patterns in individual-level dimension scores. One approach is to allocate these scores to each individual dog and investigate relationships between these scores and outcomes of interest (in particular, the results of ELISA and PCR testing). Another approach is to use a clustering algorithm to identify groups of individuals within the population. Two commonly used clustering approaches are hierarchical clustering and partitional clustering, and are reviewed in (Husson *et al.*, 2010, 2011). Although clustering strategies could be applied to the raw dataset, principal components methods such as MCA can be a useful 'pre-processing' tool for clustering methods, since they can be useful for removalof the 'noise' in a dataset (the initial dimensions would be expected to identify the signal, and the later ones will tend to identify the noise). They can also be useful for visualising the data following clustering and therefore assist interpretation. Once allocated to clusters, cluster membership for individual dogs could be used as a predictor variable for outcomes of interest. Agglomerative hierarchical clustering using Ward's method followed by a K-means clustering for consolidation on the current MCA output identified a total of four clusters in the first four dimensions (data not shown), which is a considerable reduction from the original 49 active variables.

5.4.6 Caveats

As can be seen from the above discussion, this approach does not give exact answers, but rather is a method of data exploration for a relatively large dataset. It is important to note that the dimensions identified would be expected to differ between different communities and areas, and therefore this strategy has limited use if extrapolation to other areas is intended. However, it may be useful for hypothesis generation, or for categorising communities in a quantitative framework. One important consideration when interpreting these results is that although many of these relationships are specified at the household level, the analysis was conducted using dog-level data. This was conducted because of the intention to associate the results with those of faecal testing, which was conducted at the individual level. Also, only dog owners were included in the analysis, meaning that these results cannot be extrapolated to those who do not own dogs. A fuller analysis would possibly also conduct an initial MCA on household-level data collected from all people interviewed, in order to improve understanding of the study villages as a whole. This could also be useful in the identification of types of people who own dogs, and (if data was available from an ultrasound scanning campaign), of associations with human echinococcosis in a field setting.

Another potential issue relates to the inertia captured by each dimension. The percentages of inertia explained by MCA are often much lower than in other principal components methods, since many more dimensions are usually required in order to explain all of the variance (Husson *et al.*, 2011). Also, since analysis in this case has been based upon application of CA methods to an indicator matrix, a single categorical variable will be expressed in multiple columns of this matrix and the variance explained by each dimension will therefore be underestimated. A total of 78 categories were included in the indicator matrix in the current example, and a total of 43 dimensions were extracted. It has been shown that all dimensions with eigenvalues less than or equal to the reciprocal of the number of categories are simply coding these additional columns (Abdi and Valentin, 2007). In the current case, this relates to the last 12 dimensions, which contribute 8% of the total variance. Despite this, the

percentages of inertia captured are relatively low for each dimension in the current study (see table 5.4), which constrains how much information can be extracted regarding the complexities of dog ownership in a small number of dimensions.

As described above, the associations between the qualitative supplementary variables and the dimension scores, measured by cos^2, are low. This suggests that, even if significant by t-test, these variables are not strongly related to the dimensions (and vice versa), which therefore limits their interpretability. This is one reason why MCA should be considered to be an exploratory approach, and should be combined with other methods. The association with quantitative variables was generally higher. Whilst this estimate of correlation is based upon an assumption of a linear relationship with the dimension score (which is unlikely to be the case), this may suggest that interpretation of coproantigen ELISA results in a continuous fashion offers benefits over dichotomised results, as mentioned above (and in chapter 4).

5.5 Conclusions

Multiple correspondence analysis (MCA) is a multivariate technique which can identify associations between categorical variables (based upon the concept of deviation from independence). A dataset of dog owners collected prior to a praziquantel-based dosing campaign in the Alay valley, Kyrgyzstan, was analysed using MCA, in order to better understand patterns in dog ownership in four study villages. A number of variables of interest were evaluated, relating to both household-level and dog-level factors, and a total of four dimensions were extracted. These related to differences in livestock ownership, dog types and use, dog demographics and management, and knowledge of echinococcosis/dog management. Associations between *Echinococcus* faecal coproantigen and coproPCR test results and dimensions relating to livestock ownership, dog type and education were identified, which suggested possible risk factors for dog infection. Further work using clustering algorithms to identify and classify types of dogs would be a useful next stage of analysis, and could be combined with conventional regression modelling to obtain more quantitative estimates of risk factors of importance. However, it is advised that

MCA is used in combination with other approaches in order to maximise the information which can be obtained during surveillance.

Chapter 6: Temporal dynamics of canine echinococcosis in southern Kyrgyzstan during a praziquantel dosing scheme.

"Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces for they are not at all alike, but differ much from themselves in regard to their changes. Then the winds, the hot and the cold, especially such as are common to all countries, and then such as are peculiar to each locality."

Hippocrates (460 – 377 BC)

6.1 Introduction

6.1.1 Echinococcosis

As described in the previous chapters, *Echinococcus granulosus* (sensu lato) and *Echinococcus multilocularis* are endemic in Kyrgyzstan, with high prevalences of human alveolar echinococcosis (AE) reported in Osh province in the south of the country (Usubalieva *et al.*, 2013). Canine infection with *E. granulosus* G1, *E. canadensis* G6, and *E. multilocularis* has also recently been detected in the Alay valley of southern Osh province (van Kesteren *et al.*, 2013). In 2010, a World Bank–funded project aiming to improve surveillance and control of a number of zoonotic pathogens was implemented (World Bank, 2005). One component of this campaign focussed on the control of echinococcosis through educational campaigns, dog population management, livestock slaughter controls and dosing of dogs with praziquantel four times annually by local government veterinarians and paraveterinarians (World Bank, 2005; WHO, 2011). The World-Bank funded praziquantel dosing scheme was commenced in the Alay valley towards the end of 2012, and was continued for at least two years from this time. The current study details an investigation of temporal and seasonal trends in the *Echinococcus* coproantigen and coproPCR test prevalence amongst dogs over the course of this dosing scheme.

6.1.2 Control and surveillance

Due to the long periods required for echinococcosis control (which may be an indefinite process in some cases), it is important from an economic and a disease control perspective to conduct ongoing surveillance. Information gained from this process can be used to identify areas of control scheme failure and target control activities as required (Gemmell *et al.*, 1986a). A number of potential 'data streams' are available for the surveillance and monitoring of *Echinococcus* infection, due to its complex lifecycle (involving two hosts) and zoonotic nature (Craig *et al.*, 2015). These include passive surveillance through hospital records and abattoir inspection, and active surveillance through planned surveys of human and animal infection. Whilst

surveillance should include as many of these data streams as possible, data collection can be challenging in the remote communities most affected by echinococcosis, and therefore surveillance of infection in definitive hosts is commonly considered the best measure of the level of infection in a community (as well as being a measure of the potential risk to humans).

Due to logistical and practical difficulties in obtaining good quality parasitological data (such as purge or necropsy samples) from dogs, coproantigen testing is often used to approximate canine infection status during surveillance activities (WHO/OIE, 2001d; Morel *et al.*, 2013). This has the advantage of being relatively quick and easy to conduct, but is unable to differentiate different species of *Echinococcus*. As the Alay valley is known to be coendemic for at least three species of *Echinococcus* (van Kesteren *et al.*, 2013), methods of diagnosing canine echinococcosis to the species level would be useful. This can be achieved through coproPCR analysis, but as conducting PCR is a labour-intensive process, it is less suited for surveillance activities (Deplazes *et al.*, 2003; Torgerson and Deplazes, 2009). The outcome of the standard PCR testing used in the current study is dichotomous, based upon the visual identification of a band of appropriate molecular weight on agarose gel following electrophoresis. As described in previous chapters, interpretation of coproantigen data is also generally conducted in a dichotomous fashion – classifying samples as coproantigen 'negative' or 'positive' according to where the OD value lies in relation to a defined cut-off value. Despite the potential limitations with a dichotomised interpretation of OD data (as discussed in chapter 4), a decision was made to interpret the data in this way for the current study, since this form of interpretation is most conducive to easy dissemination of the study findings. However, it is hoped that further work will build upon the results of this study and investigate non-dichotomous interpretation of the coproantigen ELISA results, as discussed in section 6.4.7.

6.1.3 Regression modelling

Logistic regression is a type of generalised linear model (GLM), which aims to model a dichotomous outcome, y. The usual interpretation of a logistic regression model is that it models the probability of a 'success' for each individual ($y_i = 1$):

$$P(y_i = 1) = logit^{-1}(X_i\beta)$$

Where X_i is the matrix of predictor variables and β is the matrix of regression coefficients. The modelling strategy used in the current paper is based upon an adjusted formulation of this model, which can be described as "logistic binomial modelling" (Gelman and Hill, 2006a). Rather than modelling a dichotomous outcome for individuals (i.e. a Bernoulli process), y_i is assumed to represent the count of positive outcomes amongst n_i individuals. This is assumed to follow a binomial distribution with n_i trials and a probability of 'success' of p_i:

$$y_i \sim Binomial(n_i, p_i)$$

$$p_i = logit^{-1}(X_i\beta)$$

This model assumes that each of the n_i observations are independent, and each have a fixed probability of occurrence, p_i. There is no error term (ε_{ij}) specified in this formulation because the error distribution (i.e. the variance) for a binomial process is determined solely by p_i:

$$var(y_i) = n_i p_i (1 - p_i)$$

Because the variance is fixed according to n_i and p_i, when $n_i > 1$, there is a possibility of overdispersion in a logistic binomial model. A common cause of overdispersion is a lack of independence between the n_i observations within each group of study (for example, in the case of a highly transmissible infectious disease where either all animals within a group are infected or uninfected), although small group sizes have also been identified as a possible cause of apparent overdispersion (Wright, 1997). Overdispersion can be dealt with either at the model formulation stage (for example,

by using a quasibinomial distribution, or by including individual-level random effects (Elston *et al.*, 2001; Browne *et al.*, 2005)); or at the interpretation stage (by adjusting the standard error estimates) (Gelman and Hill, 2006a). More information on strategies of correcting for overdispersion in count data is available elsewhere (Lindsey, 1999).

Generalised linear mixed models (GLMMs) extend GLMs by incorporating both fixed effects (as found in GLMs) and 'random effects' (the coefficients of which vary between different groups or individuals within the population). Mixed models are useful for the investigation of clustered data and longitudinal data, where the assumption of independence of observations is not considered appropriate.

During the current study, repeated observations were made for most households. The observation number ('level 1', denoted by the subscript j) can be considered to be 'nested' within individual households ('level 2', denoted by the subscript i). Using this terminology, a simple logistic binomial regression mixed model with a random intercept for each individual household can be described as follows:

$$y_{ij} \sim Binomial(n_{ij}, p_{ij})$$

$$p_{ij} = logit^{-1}\left(X_{ij}\beta + v_i\right)$$

$$v_i \sim N(0, \sigma^2)$$

Most of the terminology is as before: y_{ij} represents the number of positive animals; n_{ij} is the number of animals tested per household; p_{ij} is the probability of positivity; X_{ij} is the matrix of predictor variables [which would be expected to include a variable relating to time of sampling in this case]; and β is the matrix of regression coefficients. The addition of the subscript j to these variables is due to the identification of each variable according to both individual household (i) and repeated observation (j). In this simple case, the individual household-level random effects, v_i are presumed to be distributed according to a Gaussian distribution – meaning that the linear predictor is adjusted for each individual household (regardless of which repeated observation) by a

particular amount (v_i). This model can be further developed to allow individual subject-level variation according to other variables, and can also incorporate different random effect structures over time (such as autocorrelation). These will not be described in any more detail here.

6.1.4 Model development and selection

Identification of risk factors for an outcome of interest from a dataset is usually based upon the creation of one or more statistical models in an attempt to summarise the data using as few variables as is considered appropriate for the particular question being asked. No model will be able to perfectly represent the complex realities of the true situation, but some may be useful for identification of associations of relevance or for making predictions. When fitting a statistical model to a dataset, there is invariably a conflict between maximising the fit of the model to the data (which will almost always increase as more explanatory variables are included), and the precision with which parameters can be estimated (since the inclusion of 'unnecessary' variables in the model will tend to reduce this precision) (Forster, 2000; Burnham and Anderson, 2004). This can also be viewed as the balance between underfitting a model (i.e. including too few variables to explain the outcome of interest, resulting in poor model fit and possibly biased coefficient estimates), and overfitting a model (whereby artefactual associations may be identified due to random associations between the variables and the 'noise' around the outcome of interest).

A number of methods of model development are available, but one common approach is to add or remove variables of interest sequentially to a model framework, assessing their effect on the model likelihood and on the coefficients of other variables in order to determine whether or not they should be retained in the final model. These 'stepwise' approaches are relatively commonly used in epidemiological and ecological studies, but are known to have considerable limitations from both a practical and theoretical perspective (Madigan and Raftery, 1994; Whittingham *et al.*, 2006; Flom and Cassell, 2007; Gelman, 2014). One issue with stepwise approaches is that they are based largely on the repeated application of null hypothesis tests, which were

originally developed for single hypothesis testing rather than the repeated use seen in stepwise model selection (Flom and Cassell, 2007), and which some have argued are conceptually flawed (Anderson *et al.*, 2000). Another issue with stepwise regression is that the vast majority of model selection is conducted on a purely empirical basis. Models are selected solely from the data available and it is possible (indeed, sometimes encouraged) to therefore create models with little *a priori* consideration of plausible mechanisms behind these associations (whilst this approach may be considered acceptable in a pure exploratory investigation where no prior information is available, this attitude is problematic from a model development perspective). Finally, the usual outcome from stepwise regression is a single 'best' model, from which all parameter coefficients are then estimated with no account being made of alternative models.

Considering the issues associated with stepwise strategies, alternative approaches to model selection should be considered when deciding upon a framework for model development, some of which are summarised in a recent paper (Burnham and Anderson, 2004). One increasingly commonly used strategy for model selection is based upon information-theoretic approaches based upon Kullback-Leibler (K-L) information (Kullbank and Leibler, 1951), which describes the information lost when a model, M_2 is used to approximate another model, M_1. For discrete probability distributions (such as the binomial distribution), this is estimated as:

$$I(M_1, M_2) = \sum_{r \in R} M_1(r) \cdot ln\left(\frac{M_1(r)}{M_2(r)}\right)$$

Where 'model' M_1 is considered the (unknown) true situation, and $M_1(r)$ represents the data vector (with length R). M_2 represents the approximating model and $M_2(r)$ the vector of model output. The K-L statistic describes the information lost when model M_2 (or whichever other models are evaluated) are used to approximate reality. This cannot be calculated exactly as firstly the 'true' situation is unknown, and secondly, the model parameters are only estimated from the particular dataset used.

Akaike's 'An Information Criterion' (AIC) (Akaike, 1973) estimates the relative expected K-L information using the maximised log-likelihood of the model(s) in question (and therefore links information theoretic methods and statistical modelling). It is calculated as:

$$AIC = -2\ln(\mathcal{L}(\hat{\theta}|data) + 2K$$

If K is large relative to the total sample size, n, the following correction should be used to calculate AIC_c (Hurvich and Tsai, 1989):

$$AIC_c = -2\ln(\mathcal{L}(\hat{\theta}|data) + 2K + \frac{2K(K+1)}{n-K-1}$$

The absolute value of AIC (or its associated forms) has no meaningful interpretation, and therefore model comparisons using AIC are based upon the differences in AIC estimates (Δ) for different models. Rather than adopting a null hypothesis framework, whereby each model is considered according to a null hypothesis and is therefore either selected or discarded, Thomas Chamberlin's concept of 'multiple working hypotheses' (Chamberlin, 1965) can be used to develop a number of *a priori* models before data is added. These can then be compared using the AIC in order to identify which ones are most supported by the data. Models which do not have empirical support by the data can be removed, but a number of models may be retained and examined (Burnham and Anderson, 2004). In these cases, comparison of the Δ values for the 'best' model and other models can be useful for model assessment. $\Delta \leq 2$ suggests that the models are similarly supported by the data; whilst those ≥ 10 have little support compared to the best model (Burnham and Anderson, 2004).

The Δ estimates can be used further in order to estimate the relative likelihood of the model, out of all models evaluated (Akaike, 1981). As the AIC is calculated from the expected log likelihood of the model multiplied by -2, the relative likelihood of any individual model given the data can be calculated by reversing this calculation:

$$\mathcal{L}(M_i|data) = \exp(-\frac{1}{2}\Delta_i)$$

Whereas the AIC was calculated from the likelihood of the parameters (θ) over the whole parameter space given the data and the model ($\mathcal{L}(\widehat{\theta_i}|data, M_i)$), this estimated likelihood is a function over the whole model set. For ease of interpretation, the likelihood estimates for all models considered are often standardised so that they sum to 1:

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_i \exp(-\Delta_i/2)}$$

The estimate of this standardised likelihood for each model is known as the 'Akaike weight', w_i, and can be viewed in a partially Bayesian context as a measure of the probability that the model is the best fit to the data (based upon the K-L information), given the data and the set of models evaluated. True Bayesian posterior probabilities ($P(M_i|data)$) can be calculated using a similar strategy, and will give identical results to the Akaike weights if particular prior distributions are used (Raftery, 1995; Burnham and Anderson, 2004). The Akaike weights were used in the current study as part of the model averaging process, which is described later.

6.1.5 Assessing model fit

Assessment of model fit can be challenging for generalised mixed effects models. A common method of assessing model fit for simple linear regression models based upon ordinary least squares is the R^2 estimate, which quantifies the amount of total variance which is explained by the model, and is estimated as 1 minus the amount of 'unexplained' variance in the model (which is itself estimated as the ratio of the residual variance for the model in question to that of a 'null', intercept-only, model). This same approach cannot be used in the case of generalised linear models with non-Gaussian outcomes (such as logistic regression), since it is not possible to estimate the 'residual variance', and therefore models are not fit with the aim of minimising this. One strategy to estimating a form of pseudo-R^2 for mixed effects models is to attempt to directly quantify the variance explained by the model, and was developed by

(Nakagawa and Schielzeth, 2013). Due to the difficulties in estimating residual variance at the model scale, variance estimates can be made on the latent/link scale and partitioned according to their derivation. For a logistic regression model with random intercepts, the partitioning is as follows:

$$R^2_{GLMM(m)} = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_v^2 + \sigma_d^2}$$

$$R^2_{GLMM(c)} = \frac{\sigma_f^2 + \sigma_v^2}{\sigma_f^2 + \sigma_v^2 + \sigma_d^2}$$

Where σ_f^2 is the variance of the fixed effect component (which can be calculated as the product of the design matrix of the fixed effects with the vector of fixed effects estimates ($X_{ij}\beta$)); σ_v^2 is the variance of the random intercepts (calculated from the model); and σ_d^2 is the distribution-specific variance (which, in the case of a logistic regression model, is $\pi^2/3$). $R^2_{GLMM(m)}$ is the marginal R^2, which measures the proportion of total variance which is due to the fixed effects; and $R^2_{GLMM(c)}$ is the conditional R^2, the proportion of total variance which is due to fixed and random effects. An extension of this strategy to allow random slopes as well as random intercepts has also been described (Johnson, 2014).

Another approach to investigating model fit is to use Receiver-operator characteristic (ROC) curve analysis (Swets, 1988; Zweig and Campbell, 1993; Greiner *et al.*, 2000) to assess the predictive ability of the model (Agresti, 2007). Logistic regression can be considered as a form of nonlinear modelling, whereby the binary outcome for each individual (y_i) depends upon the value of a 'latent variable' related to the probability of positivity (π_i), which takes a value of 0 or 1 depending upon the value of π_i in relation to an unknown threshold. This is therefore based upon the same concept as the selection of a suitable cut-off for a diagnostic test, for which ROC curve analysis can be useful (as has been described previously). As well as allowing a suitable cut-off to be determined, ROC curve analysis provides an estimate of the overall discriminatory ability of the test in the form of the "area under the curve" (actually the

area between the curve and the line of equivalence ($y = x$). This relates to the 'concordance index': the probability that the model output (the predicted probability of positivity) for a randomly selected positive individual is greater than that for a randomly selected negative individual (Agresti, 2007), and therefore is a useful measure of model fit.

6.1.6 Study aims

The current study is a descriptive and analytic investigation of temporal trends in coproantigen and PCR prevalence of canine echinococcosis in a selection of villages in the Alay valley over a period of 28 months during a control scheme, incorporating the effect of reported recent praziquantel dosing. Seasonal variations in the test prevalence over this time were also investigated. It is hoped that the output of this analysis will help to improve ongoing surveillance and control measures in the area.

6.2 Materials and methods

6.2.1 Data analysis

As mentioned above, all test data were interpreted in a dichotomous fashion for the current study. Despite the issues associated with this (see chapter 4), it remains the most common method of test interpretation in these situations, and can be useful for the identification and communication of trends in coproantigen positivity over time. ROC curve analysis (as described in chapter 3) was used to identify an appropriate cut-off. A panel of samples taken from Xinjiang province in China during an evaluation of a control scheme (van Kesteren *et al.*, 2015) was tested at the same time as each batch of field samples. Samples containing fewer than 50 worms were classified as negative due to the known analytic sensitivity of the coproantigen test (Allan and Craig, 2006), and a cut-off for each batch of samples was selected based on visual inspection of the ROC curves (figure 6.1), in order to maximise comparability (i.e. similar test sensitivities and specificities). As such, the cut-off used in the current study differs from that used in the 'baseline' investigation described in the previous

chapter. The resultant OD cut-offs and estimated test sensitivities and specificities (based on the panel evaluated) are shown in table 6.1.

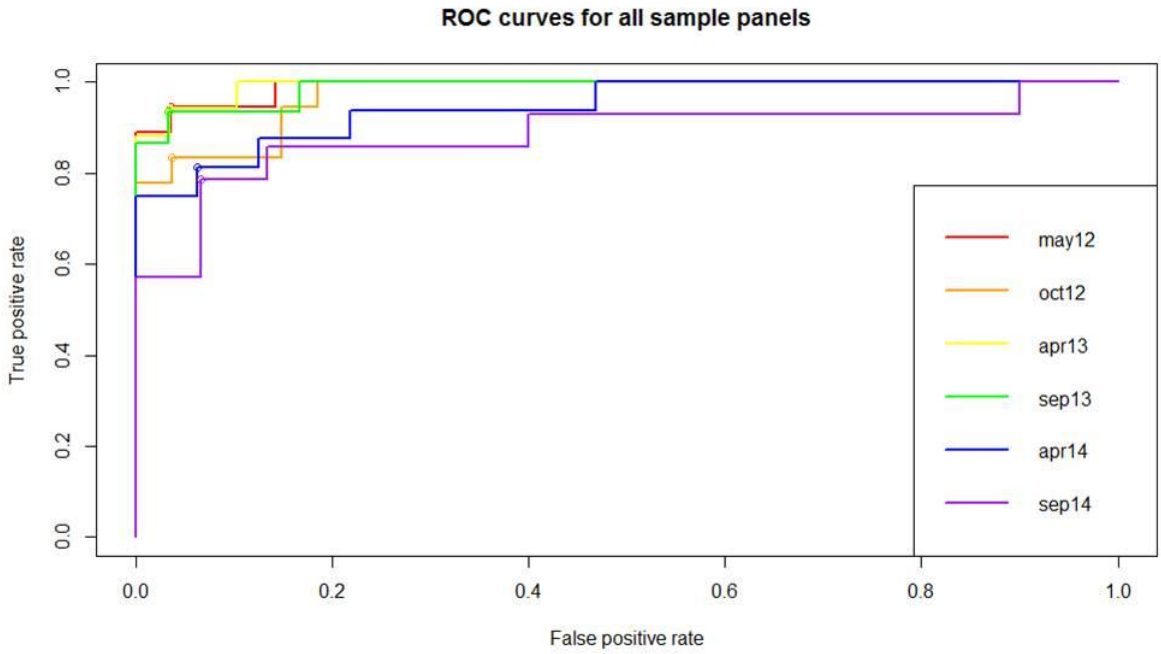


Figure 6.1. ROC curves for all sample batches, based on a panel of 'known' samples from Xinjiang, China. Circles indicate the cut-off selected.

Table 6.1. Cut-off points and estimated sensitivities and specificities for each batch of samples.

Date	Cutoff	Sensitivity	Specificity
May 2012	0.101	94%	96%
Oct 2012	0.183	83%	96%
Apr 2013	0.145	94%	97%
Sep 2013	0.172	93%	97%
Apr 2014	0.138	81%	94%
Sep 2014	0.165	79%	93%

All questionnaire and sample results were entered into Microsoft Access 2010 and matched according to the household and sample code, and questionnaire data which

could not be matched to faecal samples were removed from further investigation (n=245 individuals from 108 households).

6.2.1 Data exploration and processing

Temporal trends in praziquantel dosing behaviour; dog age, sex and weight; and *Echinococcus* positivity (coproantigen and PCR status), stratified by village, were initially investigated graphically. Further investigation of temporal patterns in coproantigen and PCR status was then conducted using a logistic binomial regression model (described below). Individual samples could generally not be matched to individual dogs as faeces were generally collected from the ground around the house rather than sampled per rectum from the dog. Therefore, all risk factor analysis was conducted at the household level rather than the individual dog level. Potential explanatory variables (see table 6.2) were identified as either household-level or dog-level variables, and all dog-level data were categorised as described in table 6.2. The dataset was collapsed according to household (with indicator variables used to indicate the presence or absence of dog-level variables from the household). The variable relating to praziquantel use since the last visit was completed for the first visit (May 2012) by asking whether praziquantel had been administered in the previous six months. The total number of dogs in the household was included as a numeric variable, with households containing more than three dogs aggregated due to the rarity of this outcome. The outcome of interest was the two-column matrix describing the number of positive and negative dogs in each household (according to either the coproantigen ELISA, or each of the three PCRs used).

Table 6.2. Variables considered for inclusion in the current study

Variable type	Variable
Echinococcus infection presence	Coproantigen status *E. granulosus* (G1) PCR status *E. canadensis* (G6) PCR status *E. multilocularis* PCR status
Temporal	Season (Spring/Autumn) Month after start of study
Spatial	Household Village (SM, TS, KS, KK)
Praziquantel dosing	Received praziquantel since last visit (Y/N)
Dog demographics	Dog age (≤1y, >3y) Male dog (Y/N) Dog weight(≤10kg, >20kg) Number of dogs in household (1, 2, 3, >3) Dog visits Jailoo (summer pasture) (Y/N)

6.2.2 Model development

The outcome of interest in the current study was the household-level probability of coproantigen or PCR positivity, and in particular how this changed over the course of the study, during different seasonal sampling points, and in the face of reported praziquantel dosing. As household-level variation in the probability of infection due to factors not captured in the questionnaire was considered a reasonable assumption, and because the sampling strategy was a longitudinal study (with many households visited multiple times), a mixed effects model was created, with household included as a random effect.

A decision was made in the current study to use information theoretic (IT) and model averaging approaches for model selection and parameter estimation. Prior to model development, temporal trends in overall test prevalence were inspected. The data structure lends itself to modelling using either a logistic regression model (which models the probability of positivity given y outcomes from n trials) or a Poisson model (which models the 'rate' of development of positivity amongst n given y outcomes over time period t) (Drolette, 1974). As the logistic regression construct appeared

more intuitive for the current study, this framework was chosen. Model development proceeded in three main stages: specification of the random effects structure; identification of interactions; and model averaging.

Assessment of the random effect structure for all four models used a 'framework model', containing the major variables of interest (month, season, village, history of praziquantel dosing since the last visit, and an interaction between the month and praziquantel dosing variables as described below). Household-level random intercepts were incorporated into this model, and additional models were created incorporating random slopes with respect to month of sampling (with and without correlation between intercepts and slopes). These models were then recreated whilst nesting the household effects within villages, resulting in a total of six possible models for comparison. Only G-side random effects were considered (rather than R-side random effects such as temporal autocorrelation), since it was considered more likely that individual households differed due to unmeasured variables than directly due to previous infection status (given the relatively short expected lifespan of the adult worm, and the ongoing praziquantel dosing campaign in place), and due to the relative small datasets available (especially for PCR data). The model with the lowest AIC_c was found to be the model with random intercepts only. In the case of the *E. multilocularis* model, a model with random intercepts and slopes was found to be the best fit to the data, as measured by AICc, but led to problems with model convergence, and therefore random intercepts only were modelled for all PCR models, due to the much reduced sample size.

The second stage of model selection involved characterisation of the full model and identification of interactions to be retained in this model (as model averaging outputs would not be interpretable if models with different interactions were included). For the ELISA data, the fixed effects incorporated in the full model (along with the month of sampling) were season (spring/autumn); village; the total number of dogs in the household (including unsampled dogs); the presence of dogs in the household which had received praziquantel since the previous visit; and whether any dogs in the household visit Jailoo. Dog demographic variables included related to the presence of

male dogs; dogs of less than one year of age; dogs older than three years of age; dogs less than 10kg in weight; and dogs greater than 20kg in weight; in the household. As mentioned above, an interaction between month of sampling and having received praziquantel was also fixed in this full model as it was considered to be of particular interest to the study. Consideration of other potential interactions led to the identification of three plausible relationships:

- The effect of season may differ dependent upon whether the dogs visited Jailoo (which, for dogs which moved between village and Jailoo, was only ever visited in the summer months)
- The effect of recent PZQ administration may differ by dog weight, since heavier dogs (>20kg) may be more likely to be underdosed with PZQ than those of 'average' weight (15-20kg)
- The apparent effect of recent PZQ administration may appear to differ according to the presence of younger dogs in the household due to the aggregated nature of the dataset. Young animals are less likely to have been originally registered (especially as the time of sampling increases), and may be less likely to have actually received praziquantel. As history of praziquantel administration was classified at the household level, this could result in a disparity between the individual dog exposure and the apparent household exposure.

The full model was expanded to create models containing all possible combinations of these interactions, and these were compared using the AICc. A model with an interaction between recent praziquantel dosing and age (≤1y) had the lowest AICc, and was therefore selected for the final stage of analysis.

For the PCR data, the variable indicating the number of dogs in the household caused problems for model convergence, and so was removed from the model as suggested by Grueber *et al.*, 2011. Only the *a priori* interaction between month of sampling and recent praziquantel dosing was assessed for these models for the same reasons of model convergence. Visualisation of preliminary model predictions for the

E. granulosus G1 model suggested a poor fit, in particular due to an apparent increase in PCR prevalence in September 2013. Therefore, an additional variable relating to sampling in this month was included for this outcome only. The justification for this was that there may have been some unrecorded event prior to this which resulted in an unusual increase in *E. granulosus* G1 infection. This issue is discussed further in section 6.4.6.

All full models were assessed prior to further analysis in order to ensure they were reasonable fits to the data. A broad idea of any possible overdispersion was gained by comparing the squared Pearson residuals to the residual degrees of freedom (using the "overdisp_fun" function described in (http://glmm.wikidot.com/faq, 2014)). This approach gave no indication of overdispersion in the data (whilst this is a flawed approach given the sparse structure of the data at the household level, this finding combined with the large number of households with only single dogs meant that overdispersion was considered unlikely). Due to the large number of low expected values, methods of assessing model fit based upon the chi-square distribution were considered inappropriate, and ROC curve approaches (Agresti, 2007) using the "pROC" package (Robin *et al.*, 2011) were instead used. Due to the aggregated nature of the data, and the fact that the ROC procedure requires a binary outcome, some adjustments to the dataset needed to be made before ROC curve analysis could be undertaken. Households were first classified according to whether or not they contained a positive dog, meaning that the probability estimates obtained from the regression model also needed to be similarly adjusted to estimate the probability of at least one tested dog in the household being positive, $\widetilde{\pi_i}$. These were estimated as $\widetilde{\pi_i} = (1 - (1 - \pi_i)^{n_i})$, where π_i is the original model output (the predicted probability of positivity for a dog in household i), and n_i is the number of dogs tested within the household. When $n_i = 1$, $\widetilde{\pi_i} = \pi_i$. Nakagawa and Schielzeth's R^2 measure for mixed effects models (Nakagawa and Schielzeth, 2013) was also estimated for this model, using code developed by (Lefcheck and Casallas, 2014).

A model averaging approach was used to estimate the final model parameters, in order to reduce some of the problems associated with selection of a single model to

represent a complex situation (Lukacs *et al.*, 2010). Variables relating to the month and season of sampling, whether praziquantel dosing had been conducted since the previous visit and the interaction between this and month of sampling were all selected as variables of particular interest, and were therefore retained in all models. In the case of the ELISA model, variables associated with the additional identified interaction (i.e. presence of young animals in household and the interaction between this and praziquantel use) were also fixed in all models. Due to the possible confounding of the effect of the presence of young dogs by that of small dogs, this latter variable was also fixed in all ELISA models. In the case of the PCR models, models containing only one of the 'young' and 'small' variables were not included in the model evaluation. This resulted in the evaluation of a total of $2^7 = 128$ *E. granulosus* G1 models, and $2^6 = 64$ models for each of the other three outcomes. Models were standardised prior to analysis by dividing input variable values by twice their standard deviation (Gelman, 2008) using the 'standardize' function in the 'arm' package (Gay and Su, 2014), and different models were generated and evaluated using the "dredge" command in the MuMIn package (Barton, 2014).

A brief inspection of the identified models was made, but most interpretation of model output was based upon a model averaging approach, using the 'model.avg' command in MuMIn. All models contributing to 95% of the total Akaike weights (Burnham and Anderson, 2002; Grueber *et al.*, 2011) were selected for model averaging, which proceeded by taking the sum of all coefficient estimates for all selected models, weighted according to the Akaike weights (Buckland *et al.*, 1997):

$$\hat{\bar{\theta}} = \sum_i w_i \hat{\theta}_i$$

Variance estimates for the coefficients were based upon the modified approach described in (Burnham and Anderson, 2004). In the case of models where the parameter did not appear, the coefficient and variance were estimated to be zero (which ensures that each variable was effectively present in each model) (Lukacs *et al.*, 2010):

$$var(\hat{\bar{\theta}}) = \sum_i w_i [var(\hat{\theta}_i | g_i) + \left(\theta_i - \hat{\bar{\theta}}\right)^2]$$

Standard errors were estimated by weighting the variance estimate for each component model using the squared ratio of critical values for a t-distribution (with the appropriate degrees of freedom for the model in question) and the critical value for a z-distribution (i.e. 1.96 for a 95% confidence interval) (Burnham and Anderson, 2002):

$$ase\left(\hat{\bar{\theta}}_i\right) = \sqrt{\sum_i w_i \left(\frac{t_i}{1.96}\right)^2 [var(\hat{\theta}_i | g_i) + \left(\theta_i - \hat{\bar{\theta}}\right)^2]}$$

Where t_i is the critical value of a t-distribution the appropriate number of degrees of freedom for model i. The confidence interval was then estimated assuming a Gaussian distribution, as is commonly performed (confidence limits: $\hat{\bar{\theta}} \pm 1.96(ase\left(\hat{\bar{\theta}}\right))$).

Due to the use of model averaging techniques rather than those based upon removal of variables from the model, model coefficient estimates for all variables were presented in the output. This can lead to some difficulties in interpretation. Following the advice of (Gelman and Hill, 2006b), all coefficients for variables of particular interest with an expected sign were interpreted, rather than excluding those which were not deemed 'statistically significant'. In cases where the sign did not match with the expectation, possible reasons for this were postulated. In order to aid interpretation of the output, predictions were made of the probability of positivity over time, including seasonal fluctuations, the effects of praziquantel use, and any other variables considered significant/informative. Prediction intervals were estimated from the estimated variance-covariance matrix of predictions, as advised in (http://glmm.wikidot.com/faq, 2014).

6.3 Results

6.3.1 Data description

Following removal of unwanted observations, coproantigen data were available for a total of 1,404 dogs from 1,188 households. The number of dogs per household was distributed as shown in figure 6.2, and the number of samples tested (at the individual dog and at the household level) visited in the four villages over the six sampling times is shown in tables 6.3 and 6.4.

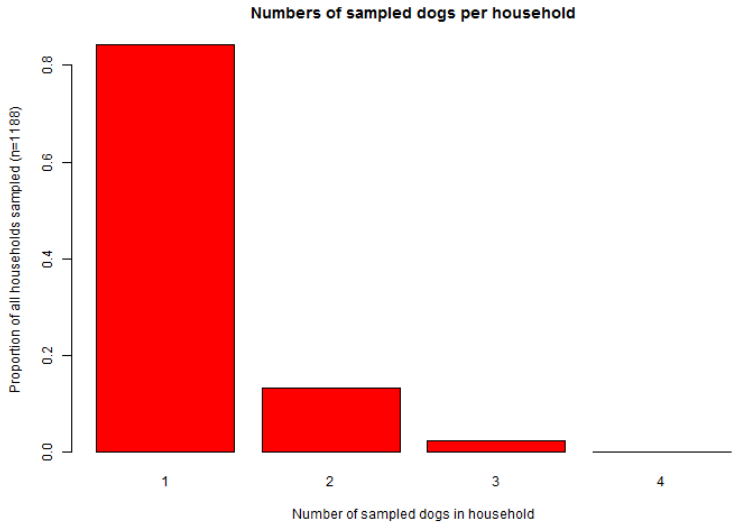


Figure 6.2.Numbers of dogs undergoing coproantigen testing per household

Table 6.3. Numbers of samples included in the ELISA analysis. Numbers of households from which samples were collected are shown in parentheses – for example, from Sary-Mogol in May 2012, a total of 155 samples from 142 households were included in the ELISA analysis.

Sampling point	Village				*Total samples*
	Sary-Mogol	Taldu-Suu	Kashka'Suu	Kara-Kabak	
May 2012	155 (142)	86 (76)	42 (40)	35 (31)	318 (289)
October 2012	63 (55)	70 (61)	56 (49)	33 (30)	222 (195)
April 2013	69 (61)	84 (72)	59 (49)	31 (24)	243 (206)
September 2013	64 (51)	80 (67)	59 (46)	27 (22)	230 (186)
April 2014	83 (66)	102 (74)	45 (38)	31 (25)	261 (203)
September 2014	42 (36)	46 (38)	24 (19)	18 (16)	130 (109)

Table 6.4. Numbers of samples which underwent PCR analysis. Numbers of households from which samples were collected are shown in parentheses

Sampling point	Village				*Total samples*
	Sary-Mogol	Taldu-Suu	Kashka'Suu	Kara-Kabak	
May 2012	126 (117)	74 (66)	41 (39)	33 (29)	274 (251)
October 2012	18 (18)	16 (15)	21 (19)	11 (10)	66 (62)
April 2013	17 (17)	34 (32)	13 (13)	8 (10)	75 (72)
September 2013	23 (21)	21 (21)	19 (17)	8 (8)	71 (67)
April 2014	28 (25)	34 (31)	16 (15)	9 (8)	87 (79)
September 2014	13 (13)	11 (10)	10 (10)	7 (6)	41 (39)

A total of 419 different households were sampled over the study period, with the visit frequency for each household distributed shown in figure 6.3. Over 30% of all households were only visited once (which will largely reflect the sampling strategy on the first visit), and fewer than 10% of households were visited all six times. Approximately 15% of all households were visited each of 2, 3, 4 or 5 times. The temporal trends in dog demographic characteristics (sex, age, and weight) are shown in figures 6.4-6.6, and those relating to praziquantel dosing are shown in figures 6.7 and 6.8. The proportion of dogs which tested coproantigen positive using the cut-off described earlier over the study period is shown in figure 6.8. Details of trends in coproPCR positivity are shown in figure 6.9.

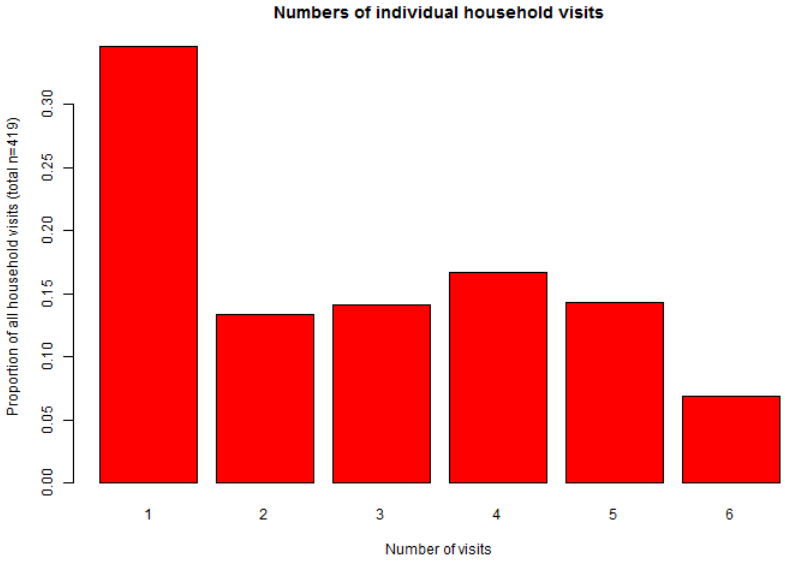


Figure 6.3. Relative frequencies of individual household visits over the study period

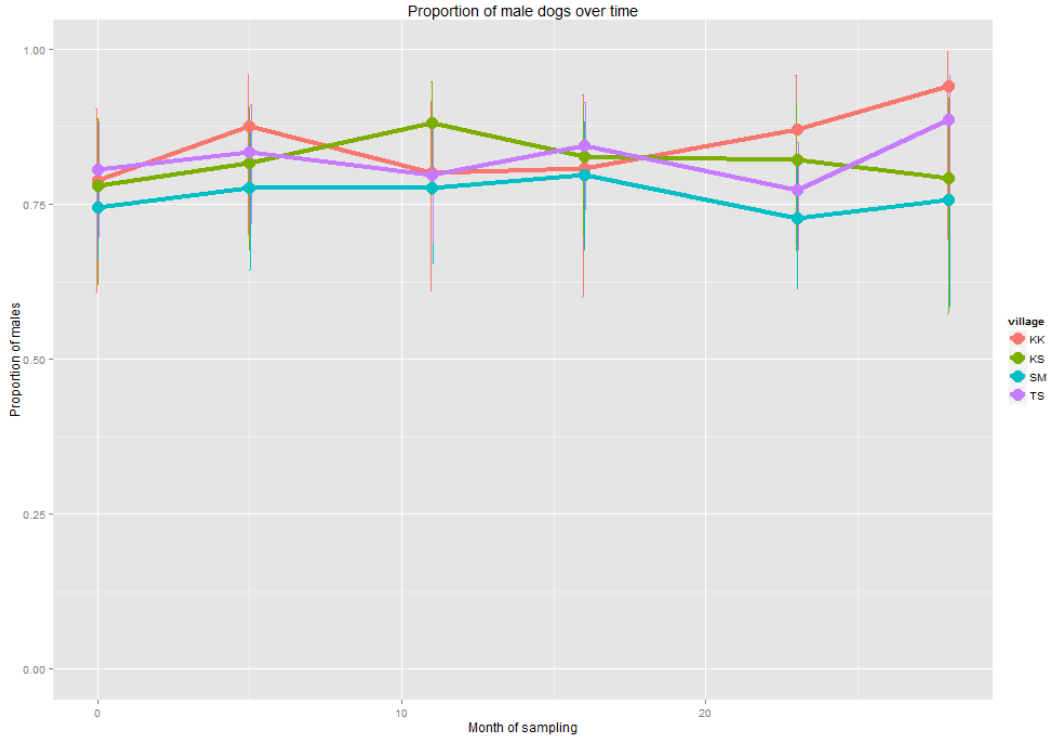
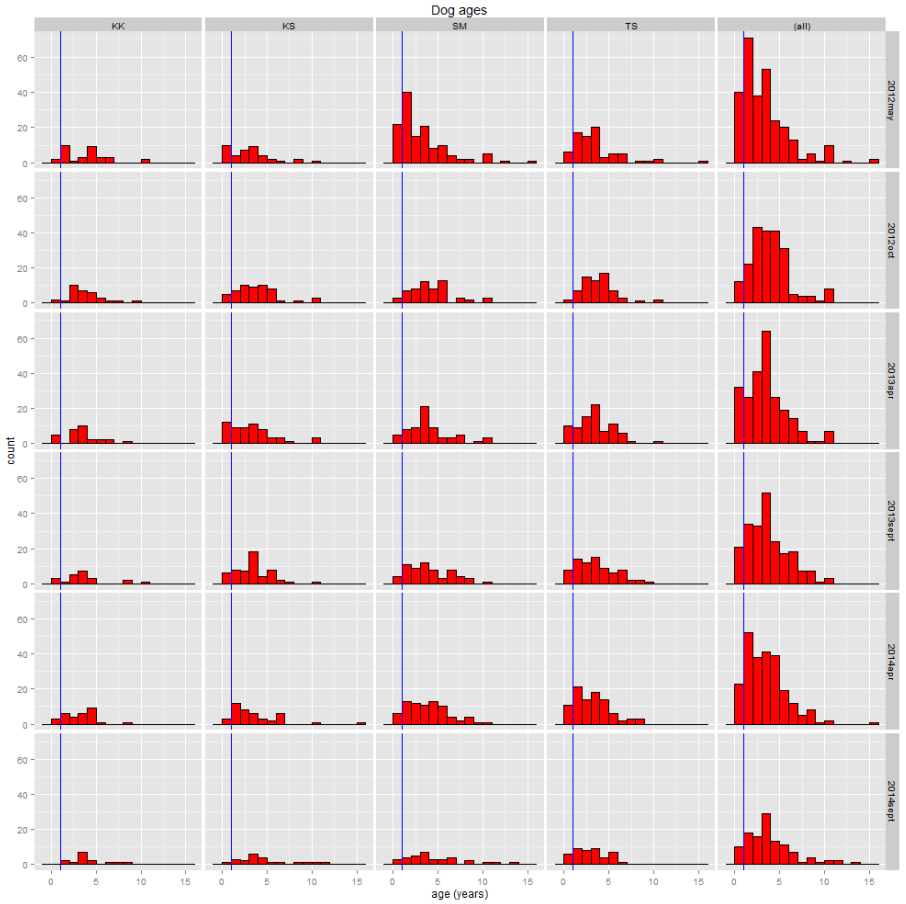


Figure 6.4. Temporal trends in the proportion of male dogs over the sampling period for the four study villages



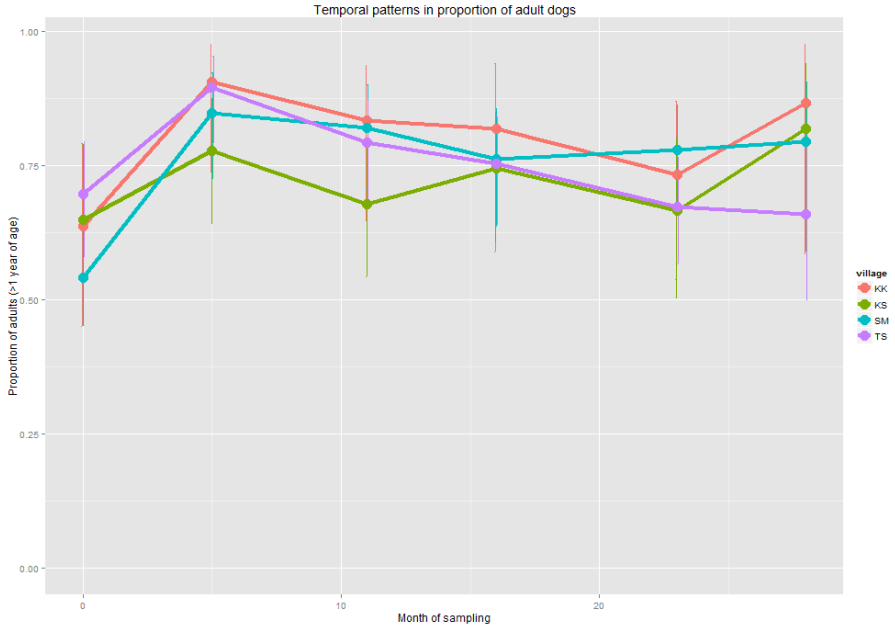


Figure 6.5.Temporal trends in the age distribution and proportion of adult dogs (≥1y) over the sampling period for the four study villages

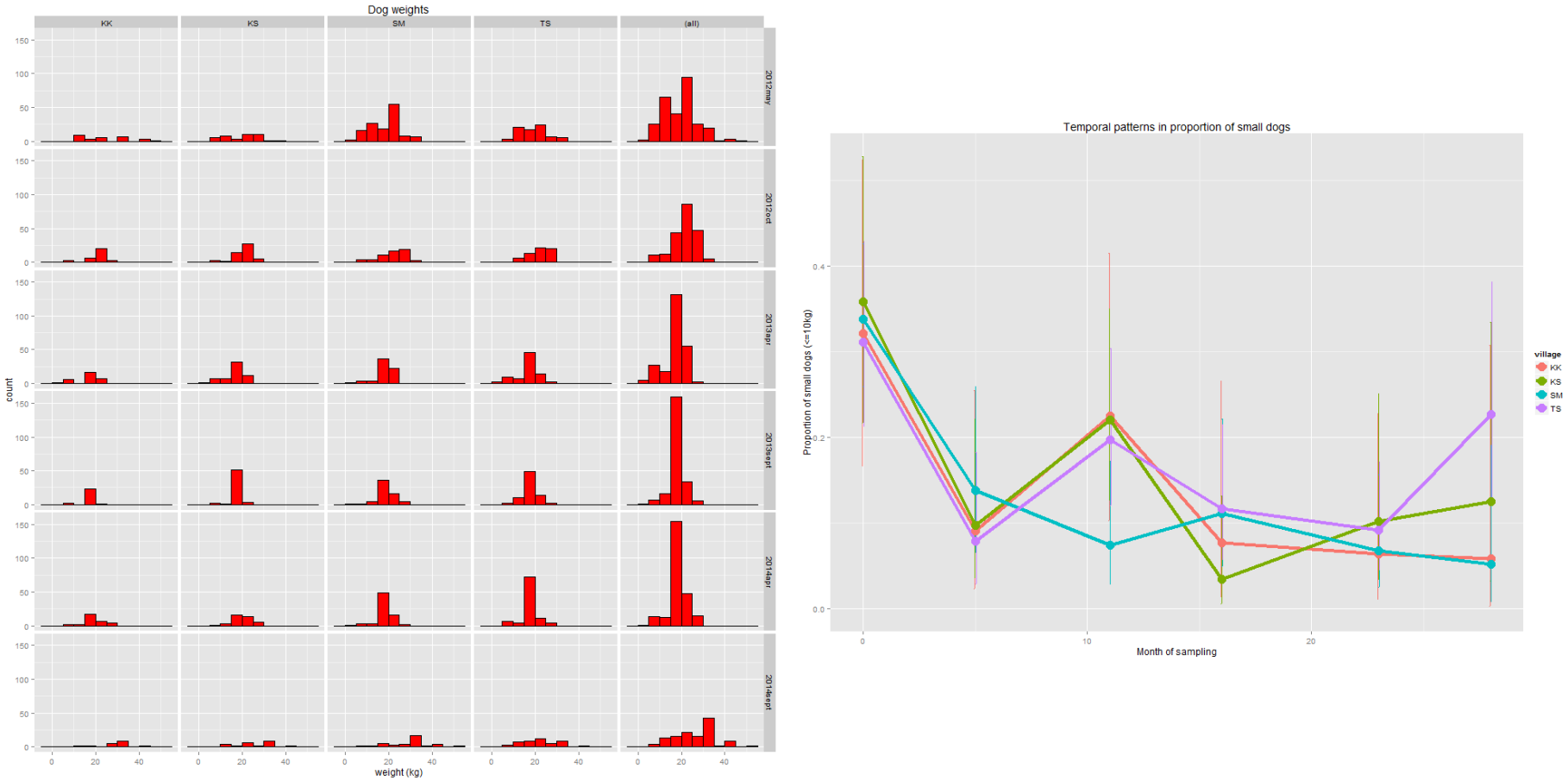
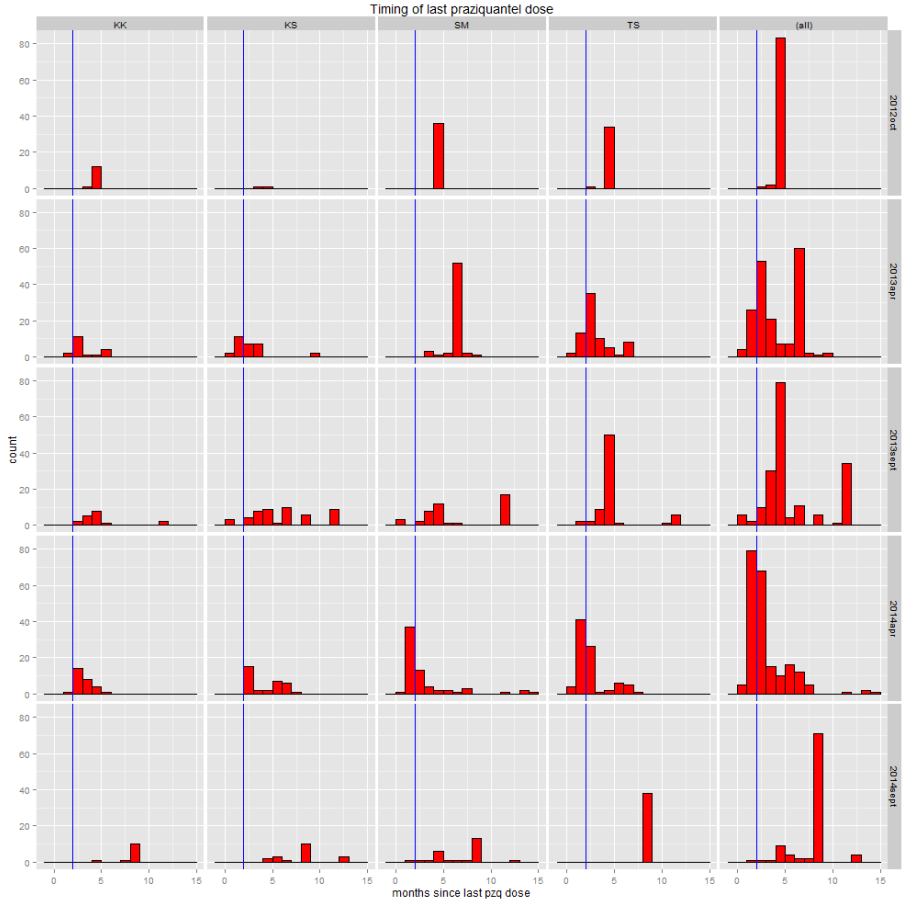


Figure 6.6.Temporal trends in the weight distribution and proportion of small dogs (≤10kg) over the sampling period for the four study villages



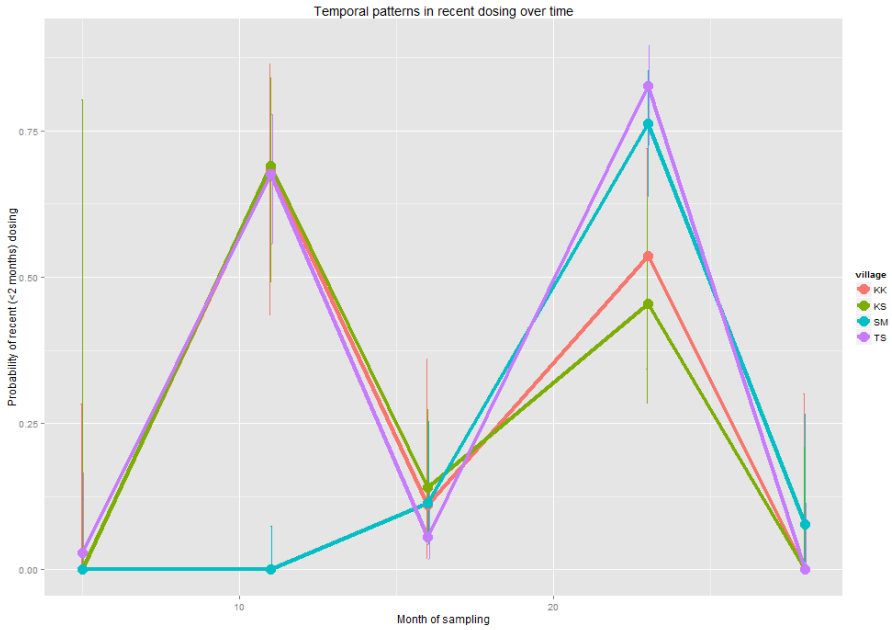


Figure 6.7. Temporal trends in the distribution of most recent praziquantel dosing, and in the proportion of recently dosed dogs (<2 months) over the sampling period for the four study villages

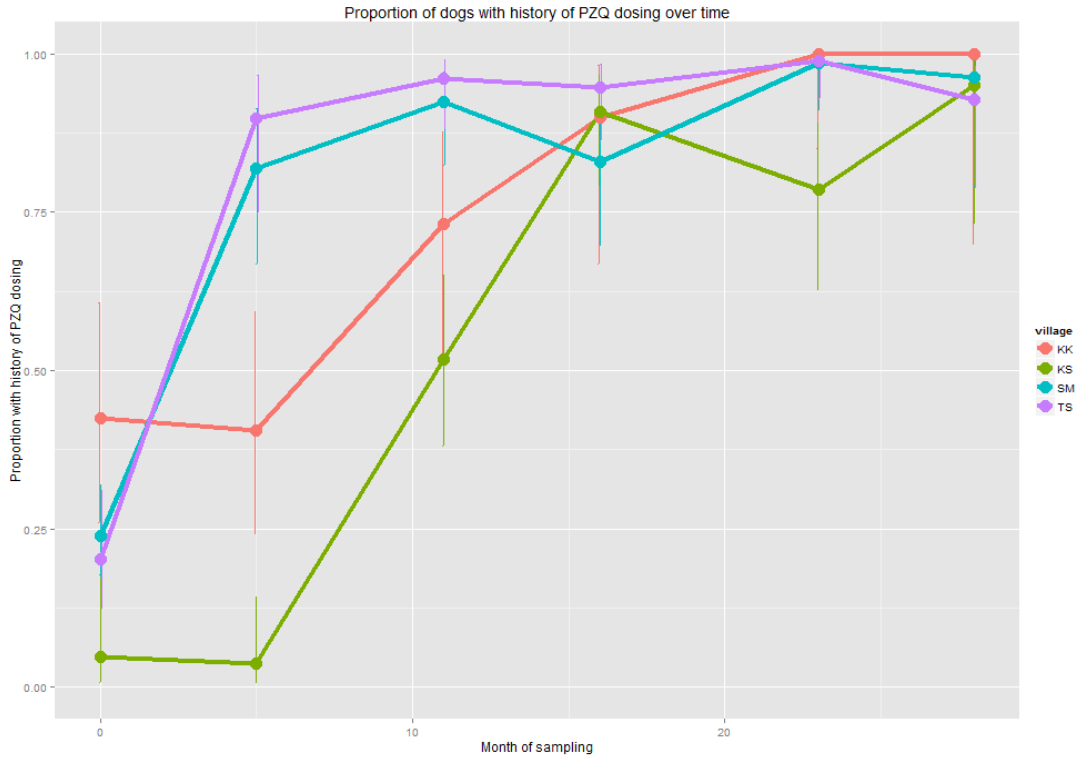


Figure 6.8. Temporal trends in the proportion of dogs which had received praziquantel at some point in the past, over the sampling period for the four study villages

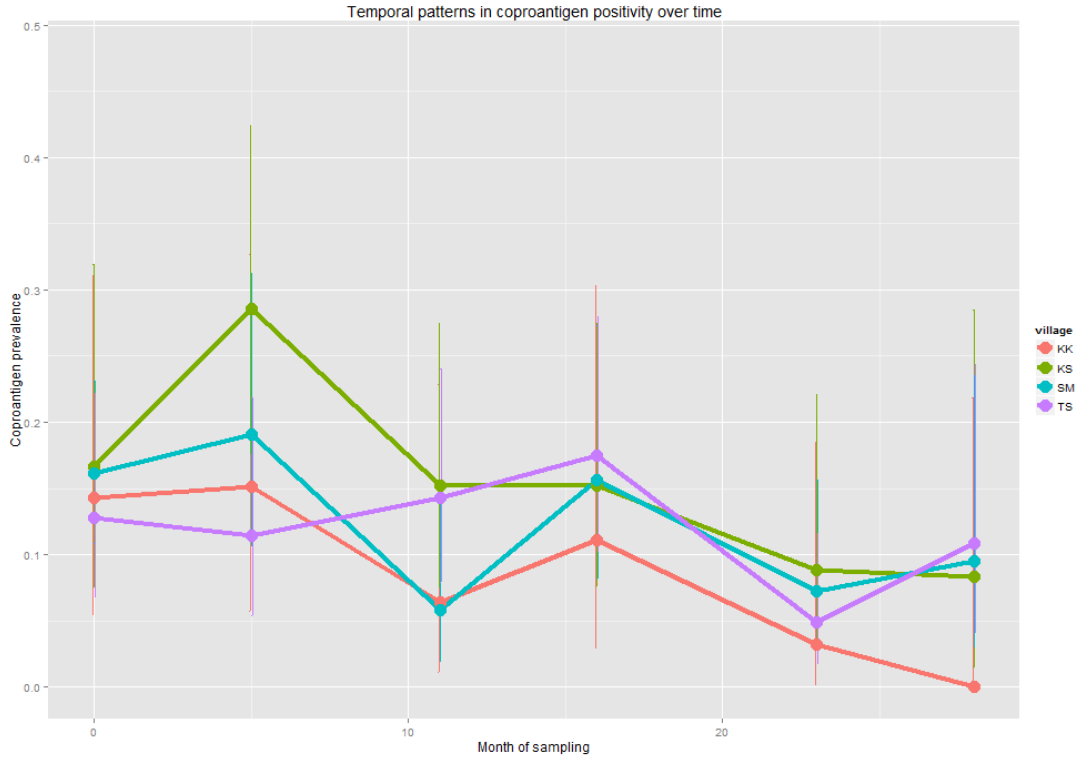


Figure 6.9. Temporal trends in coproantigen prevalence over the sampling period for the four study villages

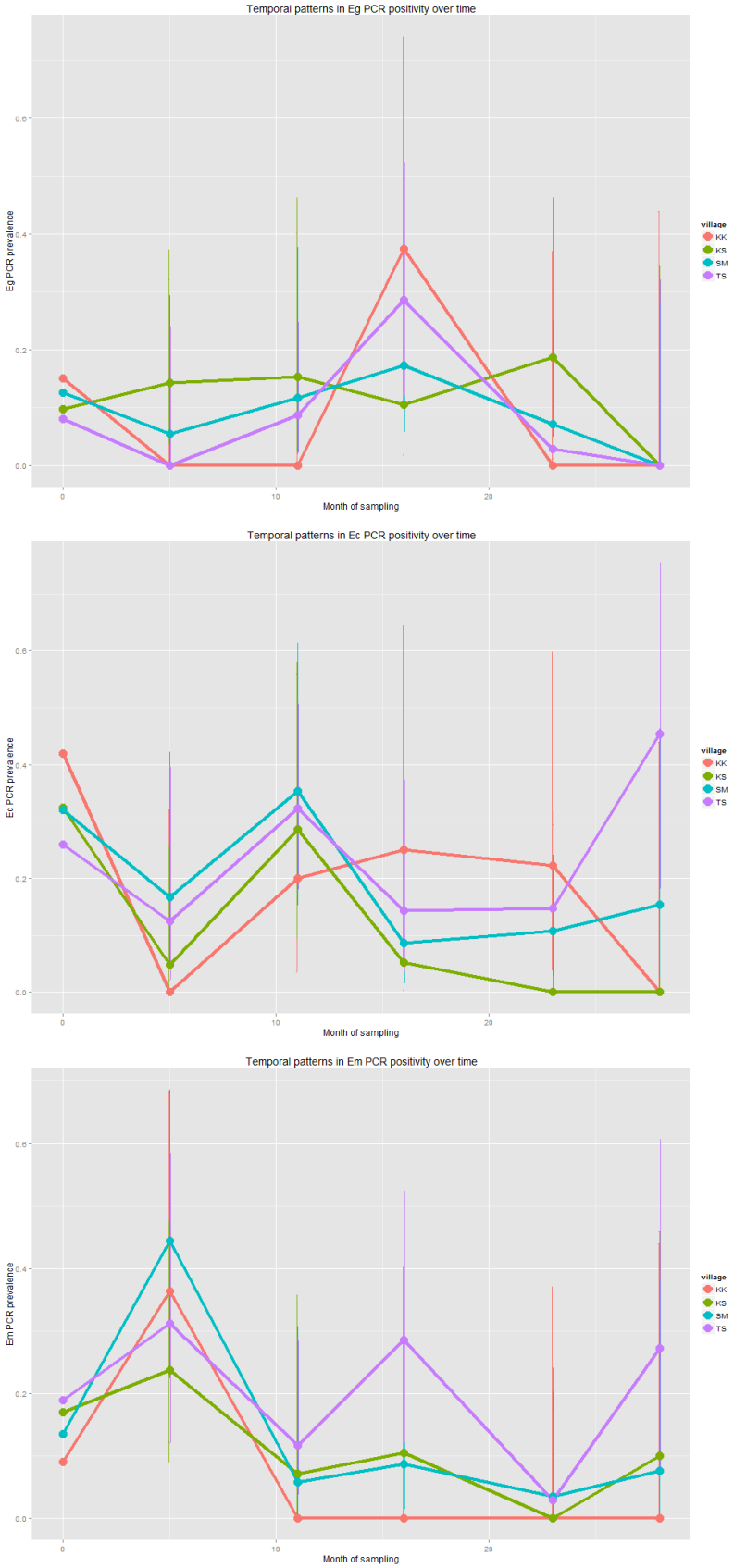


Figure 6.10. Temporal trends in *E. granulosus* G1 (top), *E. canadensis* G6 (middle) and *E. multilocularis* (bottom) PCR prevalence over the sampling period for the four study villages

6.3.2 Model checking

Model selection proceeded as described in the materials and methods. ROC curves for the four full models are shown in figure 6.11, and table 6.5 shows estimates of the concordance indices and Nakagawa and Schielzeth's R^2 estimates for the models. No evidence was found of any overdispersion in any of the models (the ratio of squared Pearson residuals to the residual degrees of freedom was less than 1.1 in all cases).

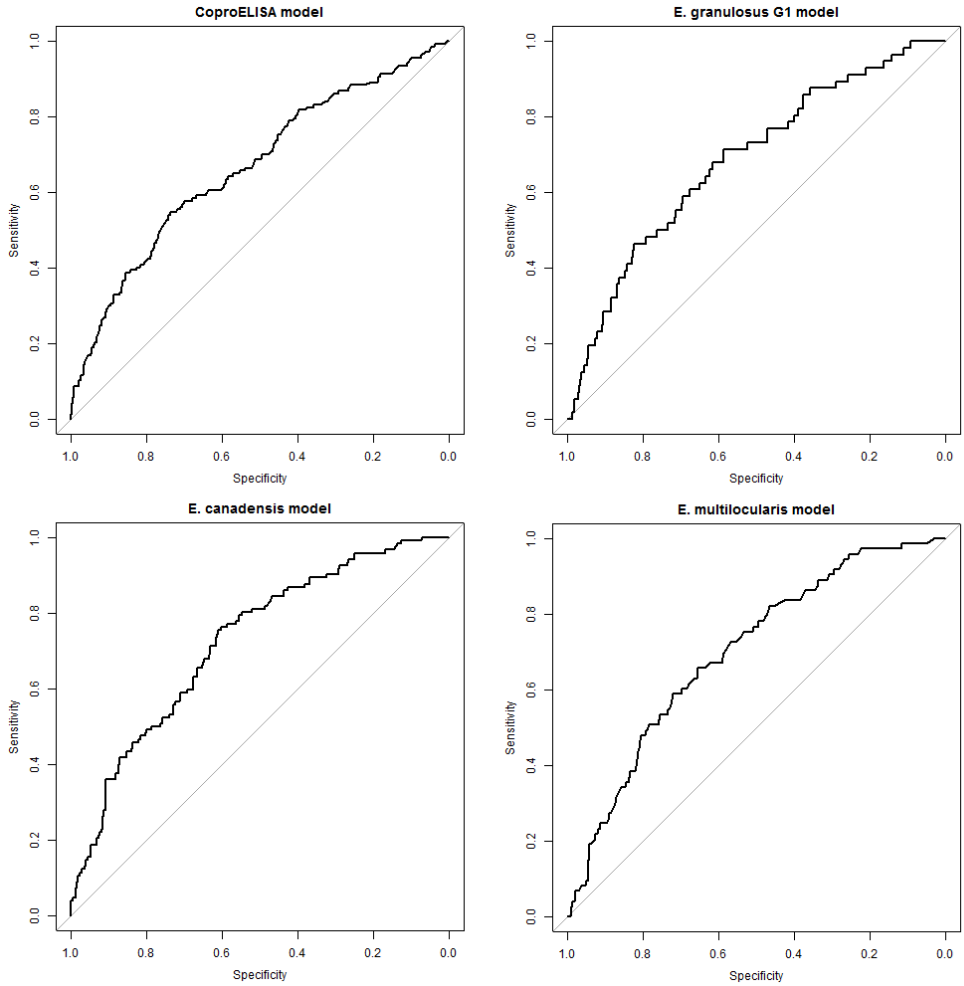


Figure 6.11. ROC curves of comparison between model predictions and data.

Table 6.5. Concordance indices (area under the ROC curve) and Nakagawa and Schielzeth's R^2 estimates for the four full models

Model	Concordance index	Conditional pseudo-R^2
Coproantigen ELISA	0.66	0.08
E. granulosus G1 PCR	0.68	0.11
E. canadensis G6 PCR	0.72	0.15
E. multilocularis PCR	0.70	0.17

6.3.3 Models generated

A total of 45 models for coproELISA, 62 for *E. granulosus* G1 PCR, 23 for *E. canadensis* G6 PCR, and 49 for *E. multilocularis* PCR were included in the final model averaging process. Tables 6.6 – 6.9 show the variables included in those models which were the best fit to the data (with a ΔAICc of 2 or less) for each outcome.

Table 6.6. Best models of coproantigen positivity, as determined by those with AICc values within 2 of the model with the lowest value. All models contained month, season, praziquantel use, presence of young dogs, and interactions between praziquantel use and both month and presence of young dogs.

	Additional variables included	df	Log-Likelihood ratio	AICc	ΔAICc	AIC weight
1	*Large dogs (≥20kg) in household*	10	-385.01	790.25	0	0.12
2	*none*	9	-386.49	791.17	0.92	0.08
3	*Older dogs (≥3y) in household* *Large dogs (≥20kg) in household*	11	-384.57	791.42	1.17	0.07
4	*Visits Jailoo* *Large dogs (≥20kg) in household*	11	-384.96	792.21	1.96	0.05

Table 6.7. Best models of *E. granulosus* G1 PCR positivity, as determined by those with AICc values within 2 of the model with the lowest value. All models contained month, season, praziquantel use, and interactions between praziquantel use and month.

	Additional variables included	**df**	**Log-Likelihood ratio**	**AICc**	**ΔAICc**	**AIC weight**
1	*Visited in September 2013* *Young dogs (≤1y) in household*	7	-169.64	353.51	0	0.12
2	*Small dogs (≤10kg) in household* *Visited in September 2013*	9	-167.99	354.36	0.85	0.08
3	*Visits Jailoo* *Visited in September 2013*	8	-169.36	355.03	1.52	0.06
4	*Older dogs (≥3y) in household* *Visited in September 2013*	8	-169.42	355.15	1.64	0.05
5	*Male dogs in household* *Visited in September 2013*	8	-169.43	355.16	1.65	0.05
6	*Young dogs (≤1y) in household* *Older dogs (≥3y) in household* *Small dogs (≤10kg) in household* *Visited in September 2013*	10	-167.49	355.45	1.94	0.05
7	*Large dogs (≥20kg) in household* *Visited in September 2013*	8	-169.6	355.51	1.99	0.04

Table 6.8. Best models of *E. canadensis* G6 PCR positivity, as determined by those with AICc values within 2 of the model with the lowest value. All models contained month, season, praziquantel use, and interactions between praziquantel use and month.

	Additional variables included	**df**	**Log-Likelihood ratio**	**AICc**	**ΔAICc**	**AIC weight**
1	*Young dogs (≤1y) in household* *Small dogs (≤10kg) in household*	8	-253.56	523.42	0	0.16
2	*Young dogs (≤1y) in household* *Small dogs (≤10kg) in household* *Large dogs (≥20kg) in household* *Visits Jailoo*	9	-252.82	524.03	0.6	0.12
3	*Young dogs (≤1y) in household* *Small dogs (≤10kg) in household* *Visits Jailoo*	9	-253.12	524.62	1.19	0.09
4	*Young dogs (≤1y) in household* *Small dogs (≤10kg) in household* *Large dogs (≥20kg) in household*	10	-252.38	525.23	1.81	0.07
5	*Young dogs (≤1y) in household* *Small dogs (≤10kg) in household* *Male dogs in household*	9	-253.43	525.25	1.82	0.07

Table 6.9. Best models of *E. multilocularis* PCR positivity, as determined by those with AICc values within 2 of the model with the lowest value. All models contained month, season, praziquantel use, and interactions between praziquantel use and month.

	Additional variables included	df	Log-Likelihood ratio	AICc	ΔAICc	AIC weight
1	*Older dogs (≥3y) in household*	7	-197.72	409.69	0	0.06
2	*Older dogs (≥3y) in household* *Village*	10	-194.69	409.86	0.17	0.05
3	*Village*	9	-195.78	409.94	0.25	0.05
4	*none*	6	-198.92	410.02	0.34	0.05
5	*Large dogs (≥20kg) in household*	7	-198.1	410.43	0.75	0.04
6	*Older dogs (≥3y) in household* *Large dogs (≥20kg) in household*	8	-197.11	410.52	0.83	0.04
7	*Male dogs in household* *Village*	10	-195.11	410.68	1	0.04
8	*Visits Jailoo* *Older dogs (≥3y) in household*	8	-197.3	410.9	1.21	0.03
9	*Older dogs (≥3y) in household* *Male dogs in household/ village*	11	-194.17	410.9	1.22	0.03
10	*Older dogs (≥3y) in household* *Male dogs in household*	8	-197.3	410.92	1.23	0.03
11	*Male dogs in household*	7	-198.36	410.95	1.26	0.03
12	*Large dogs (≥20kg) in household* *Village*	10	-195.32	411.11	1.42	0.03
13	*Older dogs (≥3y) in household* *Large dogs (≥20kg) in household* *Village*	11	-194.36	411.28	1.6	0.03
14	*Male dogs in household* *Large dogs (≥20kg) in household*	8	-197.59	411.49	1.8	0.02
15	*Visits Jailoo*	7	-198.65	411.54	1.86	0.02
16	*Visits Jailoo* *Older dogs (≥3y) in household* *Village*	11	-194.53	411.63	1.95	0.02

6.3.4 Model output

Coefficient estimates and confidence intervals for the model-averaged output are shown in tables 6.10-6.13. In order to assist visualisation of the relationship between the average household prevalence and that predicted by the model, these are shown for the four outcomes in figures 6.12-6.15.

Table 6.10. Coefficient estimates from model averaged results of models of *Echinococcus* coproantigen positivity. The intercept is shaded in black, and other variables with a 95% confidence interval which excludes zero are shaded in grey

Variable	Coefficient	95% confidence interval
Intercept	-2.05	-2.38 – -1.71
Increase of 1 unit in standardised month (baseline PZQ)	-0.61	-1.05 – -0.17
Large dogs (≥20kg) in household	-0.29	-0.91 – 0.33
Received praziquantel since last visit (baseline young dog / month)	-0.24	-0.65 – 0.16
PZQ:month (baseline young dog)	-0.16	-1.03 – 0.71
Sampled from KK (cf SM)	-0.06	-0.45 – 0.33
Sampled from TS (cf SM)	-0.05	-0.34 – 0.24
Young dogs (≤1y) in household (baseline PZQ)	-0.04	-0.50 – 0.42
Visits Jailoo	-0.02	-0.29 – 0.26
Increase in standardised dog number	-0.01	-0.17 – 0.16
Male dogs in household	0.00	-0.27 – 0.27
Older dogs (≥3y) in household	0.06	-0.21 – 0.33
Sampled from KS (cf SM)	0.07	-0.28 – 0.43
Small dogs (≤10kg) in household	0.51	0.01 – 1.01
Sampled in autumn (cf spring)	0.55	0.13 – 0.96
PZQ:young dog (baseline month)	0.69	-0.07 – 1.44

Table 6.11. Coefficient estimates from model averaged results of models of *E. granulosus* G1 PCR positivity. The intercept is shaded in black, and other variables with a 95% confidence interval which excludes zero are shaded in grey

Variable	Coefficient	95% confidence interval
Intercept	-2.15	-2.53 – -1.76
Increase of 1 unit in standardised month (baseline PZQ)	-0.72	-1.58 – 0.15
Sampled in Autumn (cf Spring)	-0.54	-1.78 – 0.69
PZQ:Month	-0.39	-1.91 – 1.14
Visits Jailoo	-0.08	-0.53 – 0.38
Sampled from TS (cf SM)	-0.04	-0.35 – 0.28
Small dogs (≤10kg) in household	-0.03	-0.52 – 0.46
Sampled from KK (cf SM)	-0.02	-0.35 – 0.30
Large dogs (≥20kg) in household	-0.02	-0.43 – 0.39
Sampled from KS (cf SM)	0.02	-0.24 – 0.27
Older dogs (≥3y) in household	0.08	-0.33 – 0.48
Male dogs in household	0.09	-0.48 – 0.66
Young dogs (≤1y) in household	0.24	-0.49 – 0.97
Received praziquantel since last visit (baseline month)	0.35	-0.32 – 1.02
Visited in September 2013	1.45	0.01 – 2.89

Table 6.12. Coefficient estimates from model averaged results of models of *E. canadensis* G6 PCR positivity. The intercept is shaded in black, and other variables with a 95% confidence interval which excludes zero are shaded in grey

Variable	Coefficient	95% confidence interval
Intercept	-1.21	-1.48 – -0.93
Sampled in Autumn (cf Spring)	-0.86	-1.57 – -0.15
Young dogs (≤1y) in household	-0.67	-1.26 – -0.07
PZQ:Month	-0.62	-1.71 – 0.48
Increase of 1 unit in standardised month (baseline PZQ)	-0.35	-0.92 – 0.22
Sampled from KS (cf SM)	-0.09	-0.58 – 0.39
Received praziquantel since last visit (baseline month)	-0.09	-0.60 – 0.43
Older dogs (≥3y) in household	-0.01	-0.24 – 0.22
Sampled from TS (cf SM)	0.00	-0.22 – 0.23
Sampled from KK (cf SM)	0.02	-0.29 – 0.33
Male dogs in household	0.04	-0.31 – 0.38
Visits Jailoo	0.08	-0.31 – 0.48
Large dogs (≥20kg) in household	0.14	-0.35 – 0.64
Small dogs (≤10kg) in household	1.27	0.64 – 1.89

Table 6.13. Coefficient estimates from model averaged results of models of *E. multilocularis* PCR positivity. The intercept is shaded in black, and other variables with a 95% confidence interval which excludes zero are shaded in grey

Variable	Coefficient	95% confidence interval
Intercept	-1.97	-2.46 – -1.48
Increase of 1 unit in standardised month (baseline PZQ)	-1.15	-1.89 – -0.41
PZQ:Month	-0.52	-1.81 – 0.78
Older dogs (≥3y) in household	-0.22	-0.81 – 0.36
Large dogs (≥20kg) in household	-0.15	-0.72 – 0.43
Male dogs in household	-0.14	-0.71 – 0.42
Sampled from KK (cf SM)	-0.11	-0.77 – 0.56
Received praziquantel since last visit (baseline month)	-0.09	-0.65 – 0.48
Young dogs (≤1y) in household	-0.08	-0.51 – 0.36
Sampled from KS (cf SM)	-0.04	-0.54 – 0.47
Small dogs (≤10kg) in household	0.04	-0.31 – 0.39
Visits Jailoo	0.08	-0.38 – 0.54
Sampled from TS (cf SM)	0.26	-0.45 – 0.97
Sampled in Autumn (cf Spring)	1.33	0.62 – 2.04

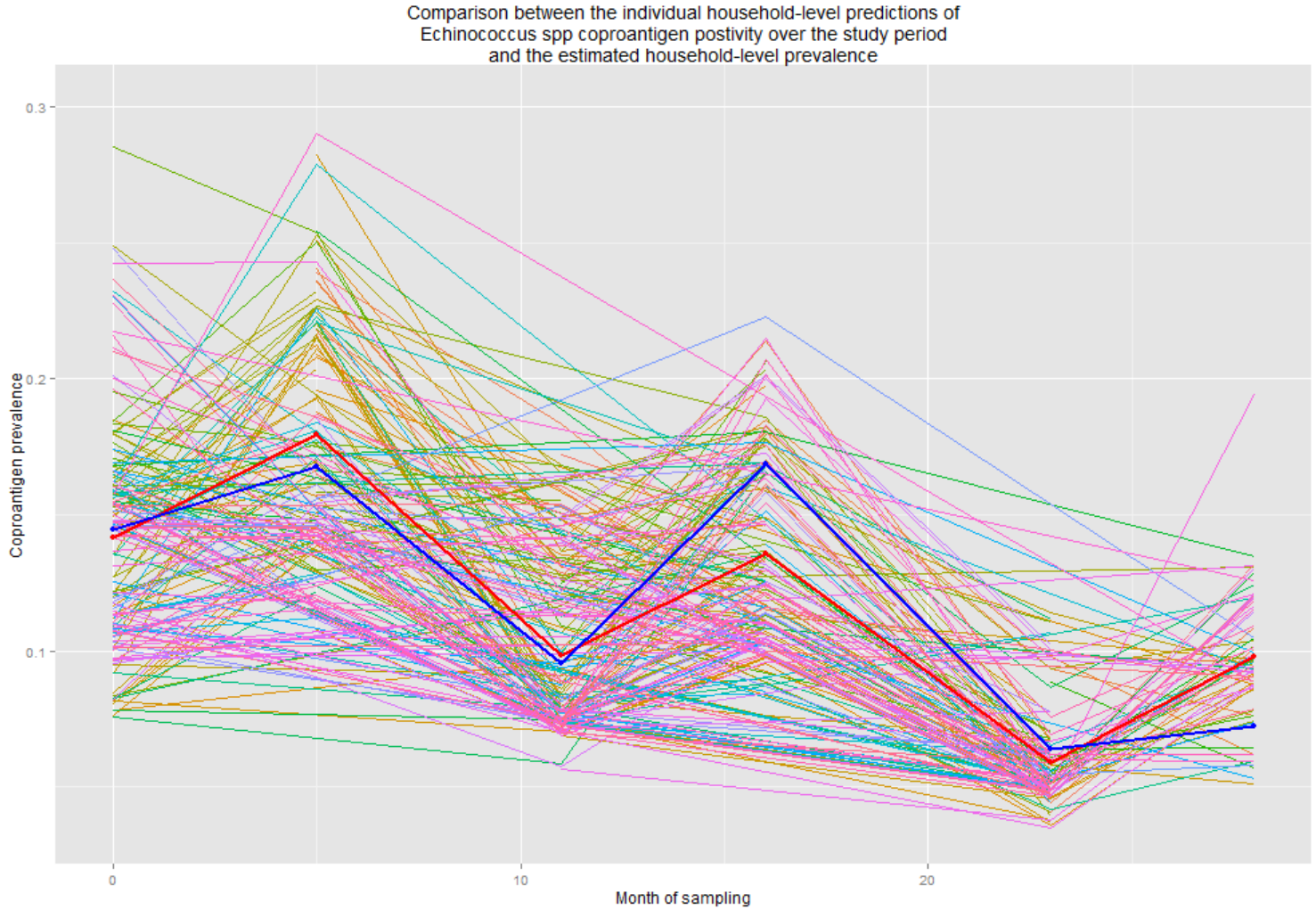


Figure 6.12. Comparison of averaged model predictions from *Echinococcus* coproantigen ELISA model (shown in blue) with average household prevalences (shown in red) over the study period. Thinner lines indicate individual household predictions.

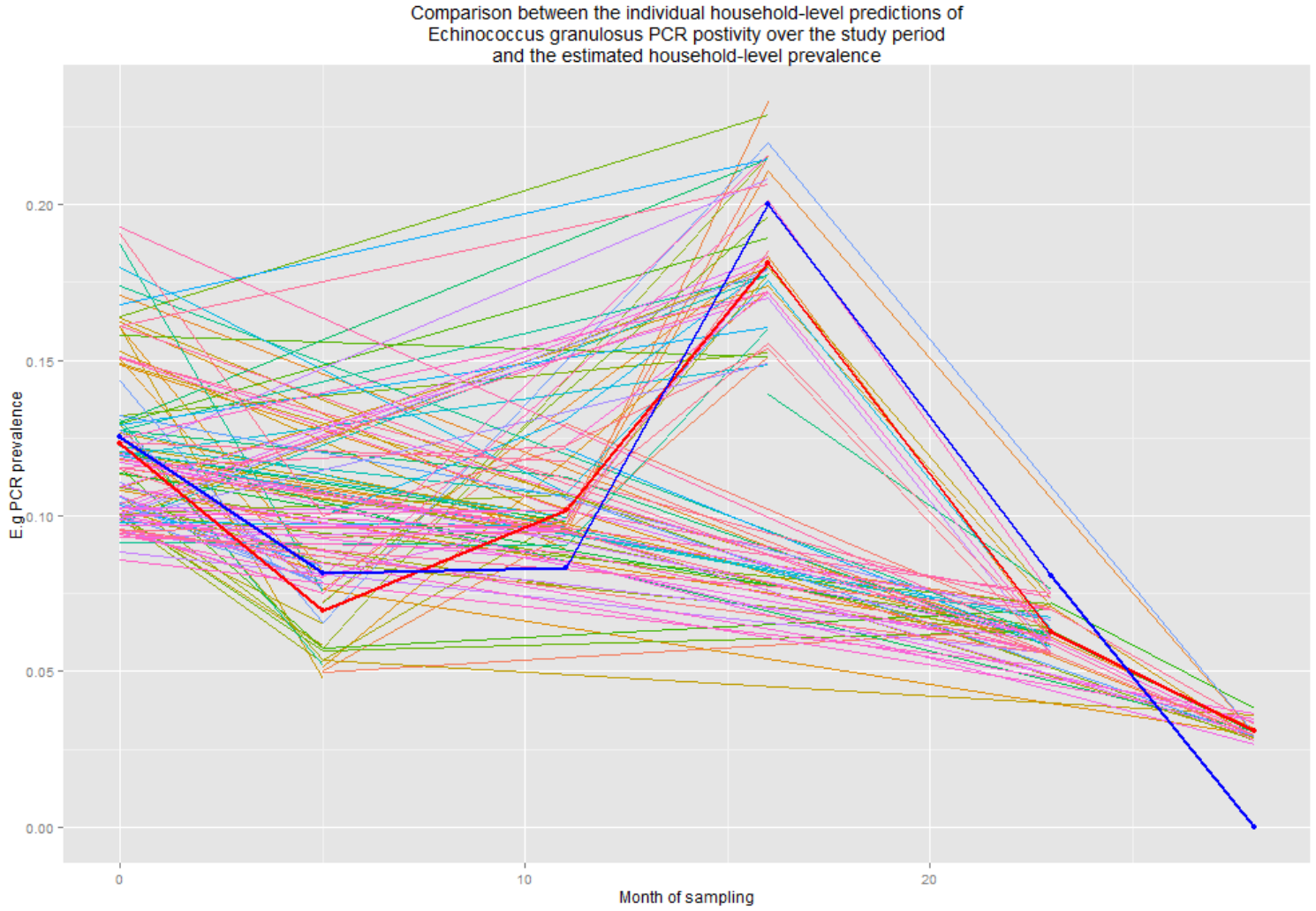


Figure 6.13. Comparison of averaged model predictions from *E. granulosus* G1 model (shown in blue) with average household prevalences (shown in red) over the study period. Thinner lines indicate individual household predictions.

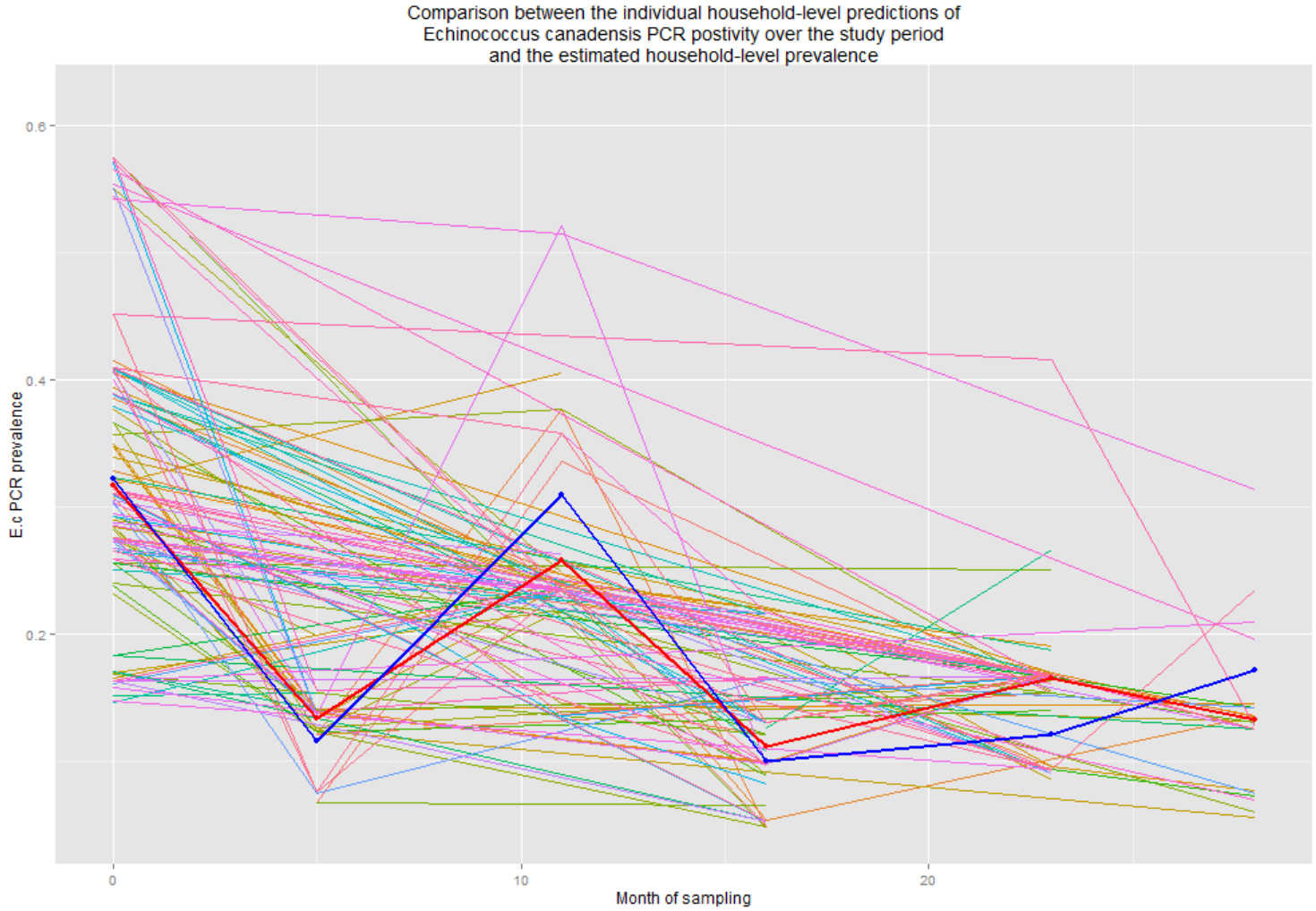


Figure 6.14. Comparison of averaged model predictions from *E. canadensis* G6 model (shown in blue) with average household prevalences (shown in red) over the study period. Thinner lines indicate individual household predictions.

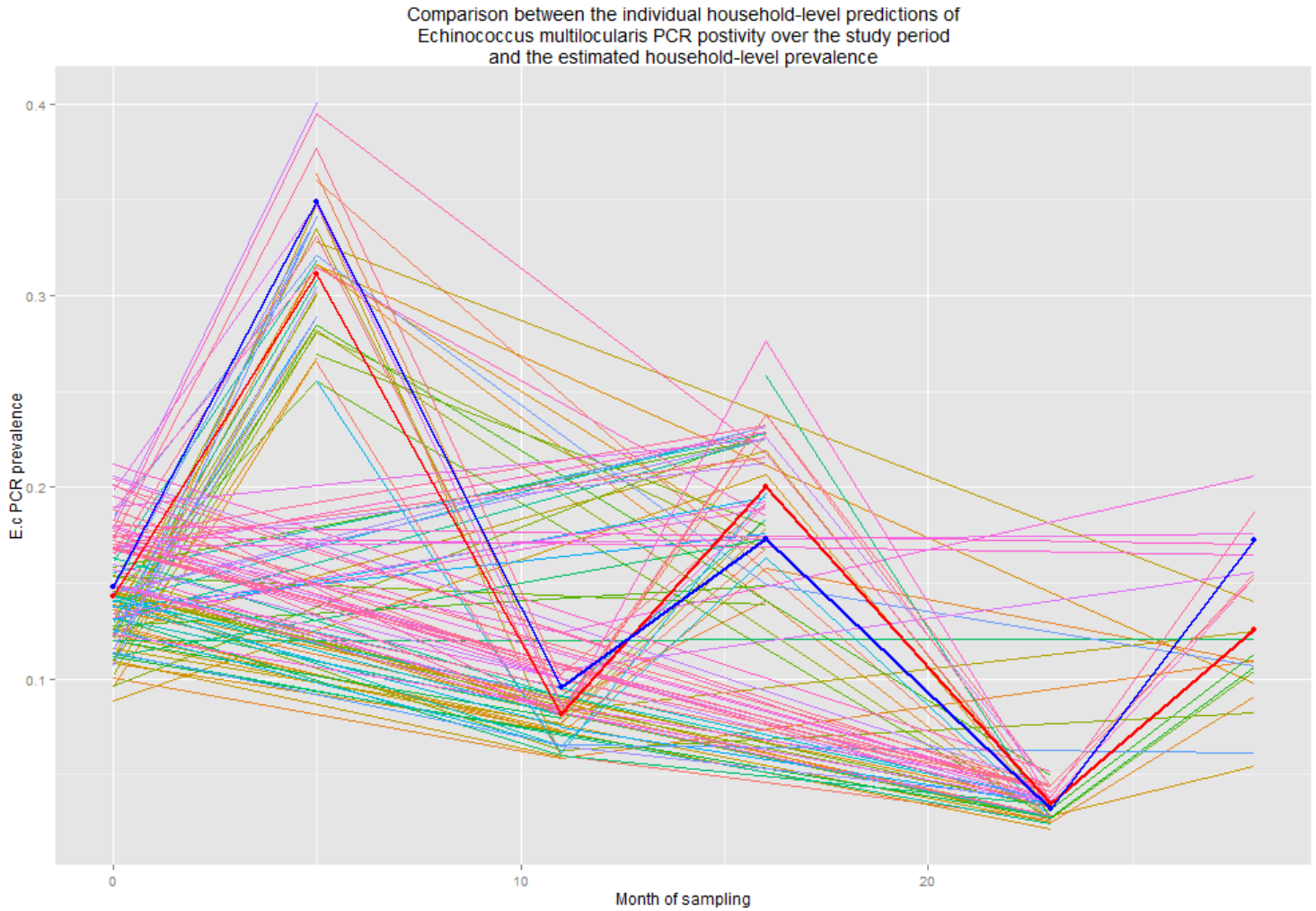


Figure 6.15.Comparison of averaged model predictions from *E. multilocularis* model (shown in blue) with average household prevalences (shown in red) over the study period. Thinner lines indicate individual household predictions.

6.3.5 Model predictions

Graphs of predicted temporal and seasonal trends in test positivity and the effect of praziquantel dosing, including confidence intervals, are shown in figures 6.16-6.19. Temporal and seasonal trends in coproantigen test positivity and the effect of presence of young dogs and small dogs (amongst households receiving praziquantel) are shown in figure 6.20, and those for the effect of the presence of young dogs and small dogs on *E. canadensis* G6 test positivity are shown in figure 6.21. For clarity of visualisation, predictions were based upon the estimates for Sary-Mogol only (selected as it was the largest village). However, the general predicted trends would be expected to be the same for all villages.

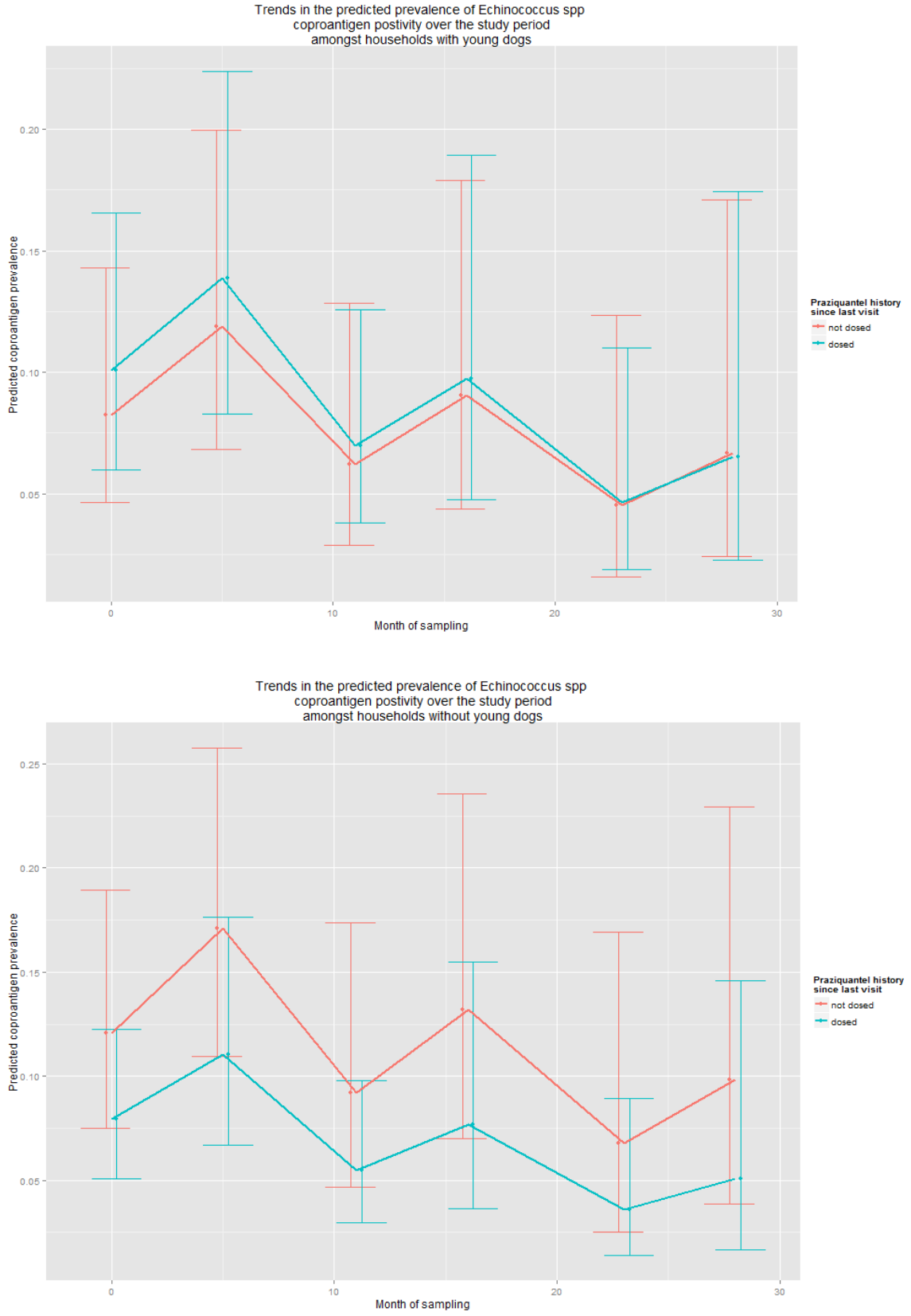


Figure 6.16. Temporal predictions of praziquantel effect from coproantigen model for households in Sary-Mogol with (top) and without (bottom) young dogs (≤1y). Bars show 95% prediction intervals.

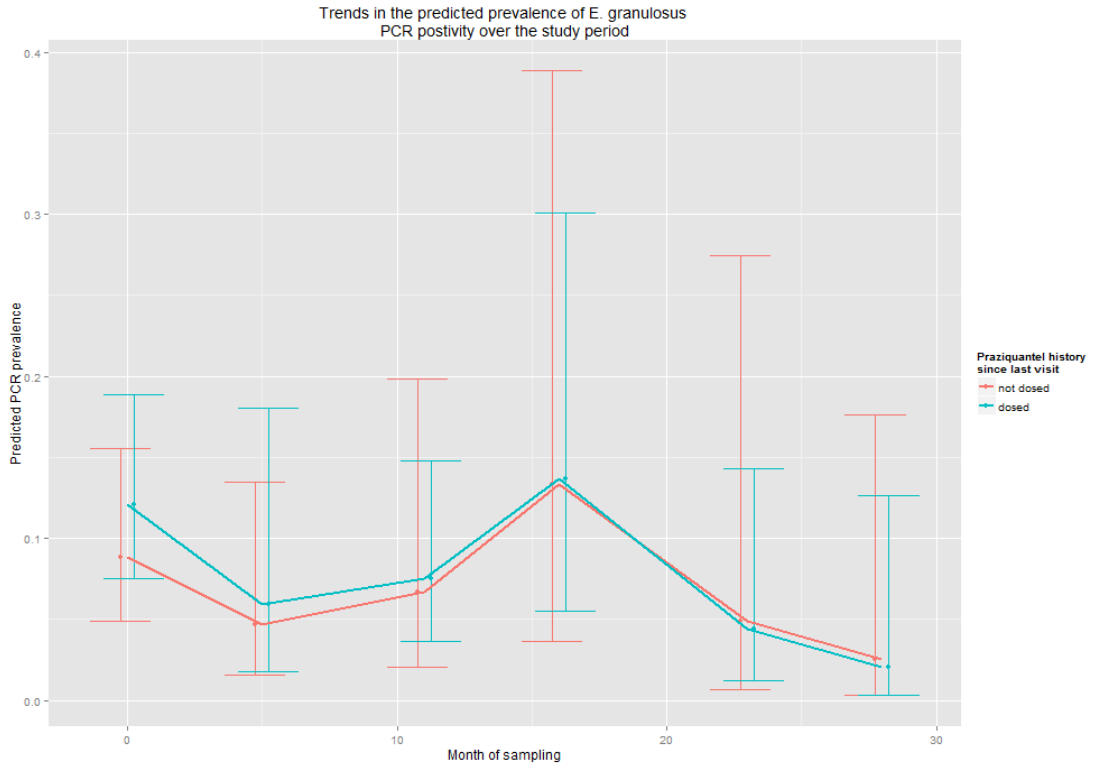


Figure 6.17. Temporal predictions of praziquantel effect from *E. granulosus* G1 model in Sary-Mogol. Bars show 95% prediction intervals.

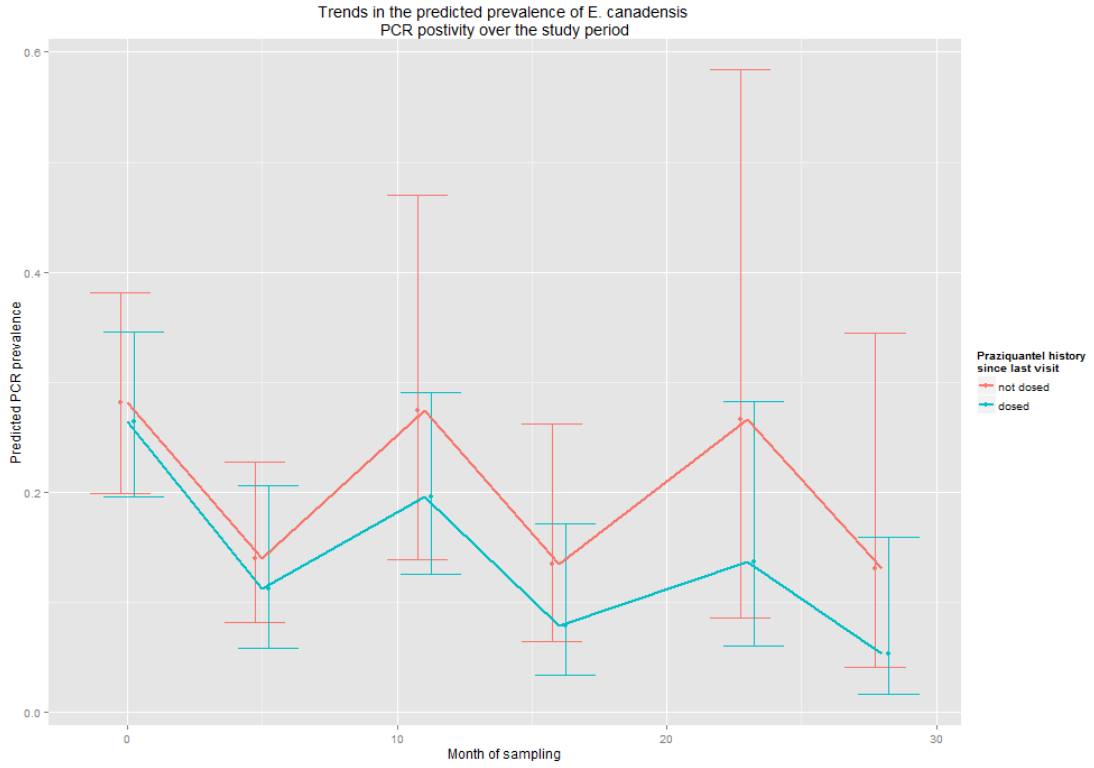


Figure 6.18. Temporal predictions of praziquantel effect from *E. canadensis* G6 model in Sary-Mogol. Bars show 95% prediction intervals.

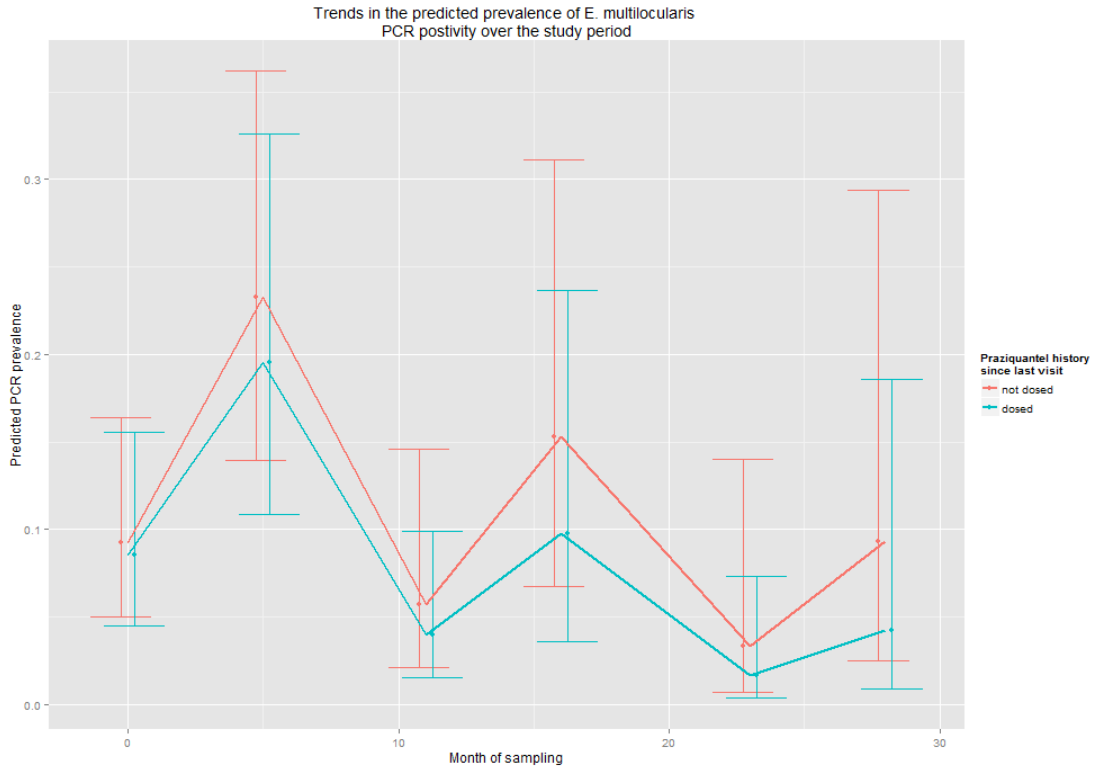


Figure 6.19. Temporal predictions of praziquantel effect from *E. multilocularis* model in Sary-Mogol. Bars show 95% prediction intervals.

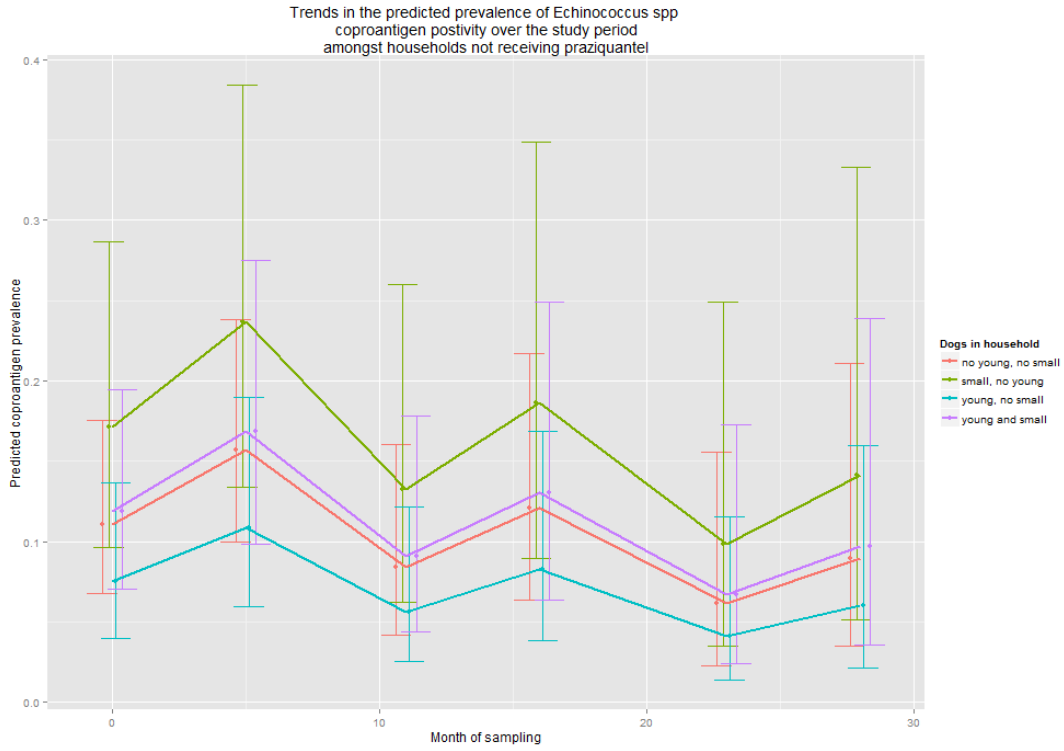


Figure 6.20. Temporal predictions of effect of presence of young dogs and small dogs from coproantigen model. Praziquantel dosing status is set to the average for the communities in Sary-Mogol. Bars show 95% prediction intervals.

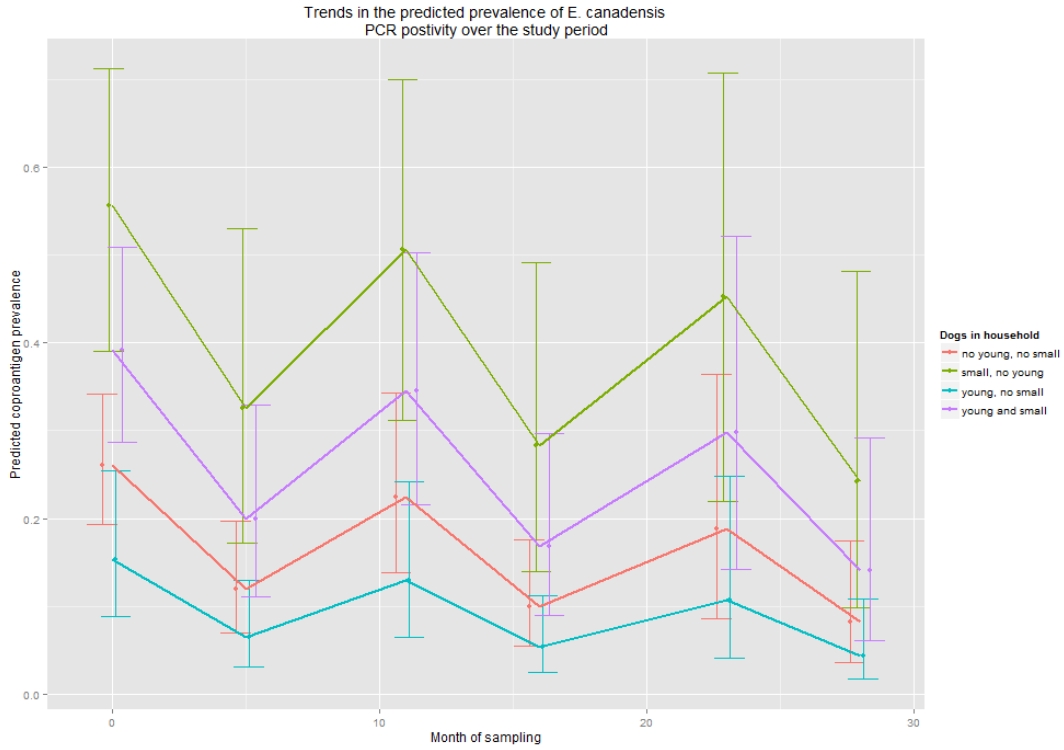


Figure 6.21. Temporal predictions of effect of presence of young dogs and small dogs from *E. canadensis* G6 model. Praziquantel dosing status is set to the average for the communities in Sary-Mogol. Bars show 95% prediction intervals.

6.4 Discussion

The current study used mixed effects logistic regression modelling to identify temporal and seasonal trends, and the effect of reported recent praziquantel dosing, on the prevalence of echinococcosis test positivity, using three different tests. These models predicted a decreasing trend in test prevalence over time in all cases, although this effect was only statistically significant for the coproantigen and *E. multilocularis* PCR models. Seasonality was also apparent in most cases, although this was not statistically significant in the *E. granulosus* G1 model. Seasonal trends differed between tests, with lower prevalences predicted during the spring months from the coproantigen and *E. multilocularis* PCR models, and the reverse for the *E. granulosus* (s.l.) (*E. granulosus* G1 and *E. canadensis* G6) models. Although not statistically significant, households containing dogs which had recently received praziquantel had lower predicted test prevalences (with the exception of coproantigen positivity in the case of households with young dogs – where predictions for 2012-2013 suggested a slightly higher prevalence amongst dosed households, but with very similar prevalence estimates to those for undosed households). Interestingly, there was evidence of an increase in *E. granulosus* G1 PCR positivity in September 2013, although the reason for this is unclear. Finally, in the case of coproantigen and *E. canadensis* G6 test results, households containing young dogs had lower predicted test prevalences, and households containing small dogs had higher test prevalences. General characteristics of the study communities have already been described elsewhere (van Kesteren *et al.*, 2013), and so will not be described in any further detail here unless associated with possible dynamic trends.

The current study describes one of the first full investigations of temporal trends in canine coproantigen and coproPCR prevalence during a praziquantel dosing scheme in an area coendemic for *Echinococcus granulosus* (sensu lato) and *Echinococcus multilocularis*. Due to the wealth of potential information which can be extracted from this analysis (especially since four different outcomes are evaluated), attention will first be focussed on those variables of primary interest – that is,

temporal and seasonal trends, and the effect of (reported) praziquantel dosing during the control scheme. Other identified associations will be discussed in less detail.

6.4.1 Quantifying infection status

It is first important to note the source of the outcomes of interest – ROC curve analysis was conducted on the coproantigen ELISA results in order to select suitable cut-offs, which were as comparable as possible between different visits. Since it is well known that an association exists between the worm burden and the OD value (Deplazes *et al.*, 1992; Raoul *et al.*, 2001; Reiterová *et al.*, 2005; Buishi *et al.*, 2005b), and that an 'overlap' in OD values is expected between high-OD negative samples and low-OD positive samples (see chapter 4), the sensitivity of the coproantigen test would be expected to be affected by the worm burden distribution in the community of interest. This is a potential problem when using the coproantigen ELISA to evaluate a control scheme, since in the face of a decreasing prevalence, the test sensitivity would also be expected to decrease. In order to address this problem, a decision was made to alter the outcome of interest and rather than model coproantigen positivity, to model "high" coproantigen positivity. A burden of 50 worms has been suggested to be suitable threshold for coproantigen test sensitivity (Allan *et al.*, 1992; Reiterová *et al.*, 2005), and therefore the ROC curve analysis used here was based upon the assumption that dogs with low burdens of less than 50 worms were (for the purposes of the study) negative. Although this will alter the distribution of 'negative' OD values in the panel used, the ROC curve approach is nonparametric and therefore will not be affected by this. By redefining positivity, the amount of overlap between 'negative' and 'positive' individuals will be reduced and the cut-off will generally be increased, which will hopefully make coproantigen prevalence estimates more robust in the face of a decreasing prevalence.

As a result of this adjustment, the model output should not be interpreted as a true estimate of the prevalence of infection in the community, but rather as a broad estimate of the prevalence of higher burden infections. Although the unadjusted prevalence may give some indication of the total parasite biomass in a community, the

overdispersed nature of *Echinococcus* infection means that a nonlinear relationship would be expected to be found between the mean worm burden and the prevalence of infection (Anderson and May, 1985; Guyatt *et al.*, 1990; Shaw and Dobson, 1995). Based upon this, it has been argued that the prevalence is not a useful measurement of the *Echinococcus* status of definitive hosts (Hofer *et al.*, 2000). Therefore, although the current approach will tend to underestimate the true prevalence of infection, it will also identify risk factors for higher burdens, rather than infection *per se*.

PCR testing is an expensive and laborious process, especially when being conducted on faecal samples, which contain PCR inhibitory substances and therefore require additional processing steps (Mathis *et al.*, 1996; Abbasi *et al.*, 2003). As such, it has been suggested that in the case of community-level surveillance, as was being conducted here, the ELISA test is used as a screening test, and the PCR as a confirmatory test applied to all ELISA positive samples (and a random selection of ELISA negative samples) (Eckert and Deplazes, 2001; Eckert, 2003; Craig *et al.*, 2003). This was not conducted in the current study due to lack of any apparent relationship between coproELISA and coproPCR results (data not shown), and the resultant difficulties in interpreting the status of samples which tested positive for one test and negative for the other. Whilst the 'screening and confirmation' strategy could be beneficial in the identification and genotyping of species and strains of *Echinococcus* in the field, it was therefore not considered a useful strategy for the estimation of the prevalence of infection in a surveillance context. Instead, with the exception of the first sample collection (when all available samples underwent PCR), a random selection of 30% of all samples (regardless of ELISA status) underwent PCR analysis. These results were then interpreted separately from the ELISA results, and used to build a picture of the transmission dynamics in the communities. It should be emphasised here that the sensitivity and specificity of these tests in the current conditions are unknown – especially in the case of the *E. canadensis* G6 primers, which were developed solely for this project and have as such not been evaluated elsewhere (van Kesteren, 2015). Whilst estimates of sensitivity and specificity for two of the PCR tests used here are available (Boufana *et al.*, 2013), the samples collected

from the field were often older than the samples used in this validation, and therefore would be more likely to contain degraded DNA which may not be detectable using the PCR.

As a result of all of these issues, the current study focussed primarily on trends over time (and in the face of other risk factors), rather than attempting to estimate the true prevalence of infection at any particular point. Although prevalence estimates are provided in figures 6.12-6.21, these are only provided to demonstrate trends and should be interpreted with care.

6.4.2 Model development

It was decided prior to model development that a random effects structure accounting for individual variation between households would be appropriate for the data. Due to difficulties in comparing the fit of models with and without random effects, and the desire to develop a general framework for analysis which could be applied to other situations, no attempt was made to compare this model to one without random effects. The random effect structure was identified by comparing the AICc of models with different structures (random intercepts only and random intercepts and slopes, with and without nesting within villages). This identified a "random intercept only" model in all cases except for the *E. multilocularis* model (for which a random intercept model was ultimately selected anyway due to convergence problems, as described above). It has been argued that when specifying random effects for GLMMs, the maximal random effects structure for the model in question should be used, regardless of measures of model fit such as AICc (Grueber *et al.*, 2011; Barr *et al.*, 2013). This was not conducted for the current study, but may be an area of exploration for further model development for the coproantigen ELISA model (convergence issues are likely for the PCR models, due to the small sample size). If this approach was used, R-side error structures (such as temporal autocorrelation) could also be considered, rather than the G-side error structure in the current model. The lifespan of *Echinococcus* spp in domestic dogs is not definitively known, but estimates of *E. granulosus* (s.l.) life expectancy of 9 months, and *E. multilocularis* expectancy of 3-6 months have been

made based on data from Kyrgyzstan (Ziadinov *et al.*, 2008). This relatively short life expectancy would be expected to reduce the degree of temporal autocorrelation in infection status between visits (even in the absence of praziquantel dosing), but may be worthy of further investigation in itself. Difficulties in incorporating R-side error structures in the current version of the 'glmer' package would mean that other packages or approaches may need to be considered if this was to be investigated further.

The decision to aggregate the test results by household was based initially upon an inability to match individual samples to individual dogs (in the case of a household containing two dogs, two faecal samples from the immediate vicinity would generally be collected, with an attempt to match for the reported sizes of the dogs). This approach will have implications for the estimation of overall prevalence (as presented in figures 6.12-6.15), which will not represent an estimate of the prevalence of infection amongst individual dogs, but the average household-level estimate. In the case of households containing only one dog, the current binomial logistic regression model reduces down to the conventional Bernoulli logistic regression form. This flexibility makes this modelling structure appropriate for a variety of situations, including those where more dogs per household are found, such as on the Tibetan plateau (Wang *et al.*, 2006a). Many communities affected by echinococcosis contain free roaming dogs (indeed, this is probably the most commonly identified risk factor for infection (Parada *et al.*, 1995; Buishi *et al.*, 2005a, 2006; Budke *et al.*, 2005a; Guzel *et al.*, 2008; Huang *et al.*, 2008; Ziadinov *et al.*, 2008; Antolová *et al.*, 2009; Mastin *et al.*, 2011)), meaning that identification of individual samples is likely to be a common problem when investigating canine echinococcosis. Additionally, since the current sampling strategy and intervention scheme were implemented at the household level, household-level risk factors were identified. This has a potential use in the context of a control scheme which is largely implemented at the household, rather than the individual dog, level. However, care should be taken when interpreting results of individual dog-level risk factors (such as presence of different dog types in the

household), due to the potential for ecological fallacy – where patterns and associations identified in aggregate do not relate to those at the individual level.

An information theoretic approach was used for model selection, largely due to the limitations of stepwise regression approaches (Burnham and Anderson, 2002; Whittingham *et al.*, 2006). The primary aim of the model selection process was to develop a predictive model which quantified the effect of the variables of primary interest (month of sampling, season of sampling, praziquantel use), whilst accounting for other factors of potential importance. In accordance with the information theoretic approach, this first relied upon the development of a number of different models ("multiple working hypotheses" (Chamberlin, 1965)). In the current study, "model dredging" was used to create these models. This is contentious, and is commonly not advised as it can discourage from careful consideration of potential models of interest (Burnham and Anderson, 2002). However, it was considered justified in the current case, as particular care was taken when selecting variables of potential importance to the model, and therefore relatively few variables were considered in the global model. As any combination of these variables was considered a reasonable model, a decision was therefore made to use model dredging to generate models (accounting for possible confounding between the presence of young dogs and small dogs by only including these two variables together). All main variables of interest were fixed in all models evaluated because models excluding these variables were not of particular interest to the current study.

Due to the conflicting problems of selection of only a single model to represent complex biological processes (Lukacs *et al.*, 2010), and the benefits of a clear model structure for dissemination of the output of analysis to a wider audience, model averaging was used to estimate coefficients for all variables whilst accounting for model uncertainty. Most of the analysis and interpretation is based upon the output of the model averaging process, but a secondary output were the results of model comparison based upon AICc estimates, as shown in tables 6.8-6.11 (these models had a ΔAICc of less than or equal to 2, which has been suggested to be a suitable threshold for model support by the data (Burnham and Anderson, 2002, 2004)). These models

may be useful for hypothesis generation but are only minimally discussed in the current report due to space limitations.

6.4.3 Temporal trends

All models were suggestive of a general decreasing trend in test prevalence over time, although this was only significant at the p≤0.05 level in the case of the coproantigen and *E. multilocularis* models. A conscious effort was made during the current analysis to not focus too much on measures of significance, since the aim was to identify possible trends. Since variables were standardised prior to model averaging, there is no clear interpretation of the coefficient for temporal trend in tables 6.10 – 6.13. This was not considered a problem, since interpretation of an annual or monthly decrease in the log odds of test positivity is similarly a difficult concept to convey. Instead, graphs of test prevalence predictions over time were provided (figures 6.16-6.21), which were considered easier to comprehend than tables of coefficient estimates.

The trend of decreasing prevalence over time is likely to represent the effect of the praziquantel dosing scheme to some degree, and has been found in previous studies of *E. multilocularis* infection of foxes during control schemes (Schelling *et al.*, 1997; Tackmann *et al.*, 2001; Hegglin *et al.*, 2003). However, it is not possible to identify the exact contribution of praziquantel administration *per se* to this trend due to a lack of any control communities which did not undergo dosing. As there is no residual effect of praziquantel, the reduction in prevalence in the face of a dosing scheme would be expected to operate through an initial removal of infection, followed by gradual reinfection over time. Previous work in Naryn province in Kyrgyzstan has suggested that the *E. multilocularis* infection pressure for free-roaming dogs is in the region of 1.1 – 1.3 infections per year, and that for *E. granulosus* (s.l.) is around 0.3 infections per year (Ziadinov *et al.*, 2008). Therefore, following dosing, the average time to infection with *E. multilocularis* would be expected to be in the region of one year, and that for *E. granulosus* (s.l.) would be around three years (i.e. the time course of the whole current study). As time progressed, the probability of any household having received praziquantel at some point increased (see figure 6.8), and therefore the overall

probability of infection would be expected to reduce. Therefore, it is plausible (and arguably likely) that the temporal trend is representative of an effective praziquantel dosing scheme. The lack of a strong association between recent dosing and probability of test positivity is also explainable due to the sampling strategy, and is discussed below.

However, although the dosing scheme is likely to be largely responsible for the decreasing temporal trend, other possible causes cannot be excluded. Dog culling campaigns have reportedly been ongoing in the villages for some time (Akjol Tagaibekov, personal communication), which would be expected to preferentially remove older dogs from households (since younger dogs often stayed closer to the households). Following culling, dogs were commonly replaced with younger dogs, which would be included in future samplings, but the change in dog would not be explicitly accounted for in the model (no attempt was made to match for individual dogs over time – just households). If younger dogs were less likely to be infected (as may be the case, as discussed below), this could result in an apparent reduction over time, as the probability of individual dogs being removed from the population increased.

Another possible cause of the decreasing trend over time is behavioural changes amongst dog owners. It is likely that the commencement of the study alerted local people to the problem of echinococcosis – and particularly, the role of dogs in the cycle. In June 2012, just after the initial round of sample collection, an ultrasound screening campaign was commenced in the area (based in Sary-Mogol and Taldu-Suu, but also including people from other villages if requested). It has been reported that these campaigns can have very positive educational benefits with regards to echinococcosis (Kachani *et al.*, 2003), and this appeared to be the case in the two main study villages (Sary-Mogol and Taldu-Suu) when semi-structured questionnaires were administered to local people in October 2012 (data not shown). As well as reducing the risk of human infection (which was not measured in the current study), these factors could lead to a reduction in the feeding of offal to dogs, which would be expected to reduce the probability of infection with *E. granulosus* G1 and/or *E. canadensis* G6.

Questions relating to knowledge of canine echinococcosis were asked at the start of the study (May 2012) and again towards the end of the study (April 2014), and further work is intended to investigate the effect of owner knowledge on dosing behaviour and test prevalence.

Whilst it is unlikely that the decreasing trend in prevalence over time seen here represents a decrease in the prevalence/intensity of intermediate host infection, a study based upon seasonally targeted dosing of dogs in Lithuania identified a trend of decreasing prevalence in both dogs and intermediate hosts (pigs) amongst intervention communities over a period of four years (Šarkūnas and Deplazes, 2014). Therefore, a reduction in infection pressure from intermediate hosts is possible, even over the short periods studied here. In the long term, a prolonged dosing campaign would be expected to reduce the levels of infection of ruminants with *E. granulosus* G1 / *E. canadensis* G6 (which are likely to exist solely in a domestic cycle). The effect on the prevalence of *E. multilocularis* in rodents is unknown – since the exact role of dogs in maintaining this cycle is unclear. If dogs were acting only as an overspill host from the sylvatic fox-rodent cycle, then dosing dogs would not be expected to affect the levels of infection of intermediate hosts. However, if they were contributing in some way then by controlling infection in dogs, the levels of infection in intermediate hosts would be expected to decrease. These longer-term effects of a dosing scheme would be worthy of further study themselves, and it would be useful to return to the study site after the control scheme has been running for a longer time in order to evaluate this further. The effects of dosing campaigns on parasite distributions is an area of active study (Basáñez *et al.*, 2012b), and the investigation of optimal strategies for measurement of infection pressure in the face of a control scheme would be useful.

6.4.4 Seasonal trends

A significant seasonal effect was observed for all models with the exception of the *E. granulosus* G1 model. However, the direction of this seasonality differed between the models. In the case of the coproELISA and *E. multilocularis* models, a higher test prevalence was found in the autumn months than in the spring months. However, in

the case of the *E. granulosus* G1 and *E. canadensis* G6 models, an increase in the test prevalence in the spring months was observed in comparison to that in the autumn months. This is an interesting finding, and may represent differences in the lifecycles of the two types of *Echinococcus* in the study area: *E. granulosus* (s.l.) and *E. multilocularis*.

Seasonality in *E. multilocularis* infection of foxes is well reported, and a trend of increased prevalence in autumn than spring has been identified in arctic foxes in Alaska (Fay and Rausch, 1964), juvenile red foxes in Switzerland (Brossard *et al.*, 2007; Hegglin *et al.*, 2007), and similar to that predicted from mathematical models of transmission in Japan (Ishikawa *et al.*, 2003; Nishina and Ishikawa, 2008). These variations were suggested to result from changes in availability of intermediate hosts (and resultant changes in dietary preference for intermediate hosts), changes in fox population density, and age effects. Amongst adult red foxes in Switzerland, the prevalence was lower in autumn than in spring (Brossard *et al.*, 2007), and other studies in Switzerland have suggested that the prevalence was higher in winter than in summer (Hofer *et al.*, 2000), or spring (Stieger *et al.*, 2002), but the months of sampling mean that these results are not directly comparable with the results of the current study. Data collected from Japan and France did not detect any seasonal changes in coproantigen prevalence or infection (despite the latter study finding evidence of increased ingestion of rodents in autumn than in spring) (Morishima *et al.*, 1999b; Robardet *et al.*, 2008).

The only published report of seasonality in *E. multilocularis* infection in domestic dogs to date was based on studies of hunting dogs in Kazakhstan (Bondareva *et al.*, 1975), as reported by Shaikenov (2004), which indicated a higher prevalence of infection in the spring than in the autumn – and therefore the converse trend to those observed here. Despite this, the finding of a higher prevalence amongst domestic dogs in the autumn (and likely, winter) months, as identified here, is logical due to expected seasonal trends in the availability of intermediate hosts to dogs. It has been suggested that the plasticity of host preference for suitable intermediate hosts of *E. multilocularis* is a key driving force in parasite dynamics (Hegglin *et al.*, 2007). The plasticity of domestic dog

preference for these intermediate hosts would be expected to be high. This means that domestic dogs would be expected to switch to other sources of nutrition in periods where access is limited (such as during the winter months, when population densities of these hosts would be expected to more readily decrease and snow cover would be expected to further reduce access). The effect of the cold weather on the age structure of rodent populations could also be of relevance for the infection pressure to dogs. On one hand, it is possible that during the harsh winter months, older intermediate hosts, which are more likely to carry mature cysts, will preferentially die off. However, the lack of reproduction during the winter months could equally increase the proportion of older animals, and therefore the overall prevalence amongst intermediate hosts (Burlet *et al.*, 2011). Further work on the population dynamics of rodent intermediate hosts (and on their interactions with dogs) in these communities would be useful in order to clarify these issues further.

Seasonality in *E. granulosus* (s.l.) infection of definitive hosts has received less attention than that for *E. multilocularis*, but a study in Kazakhstan suggested that the prevalence of farm dog infection in Zhambul oblast was higher in the autumn than the spring (although there was no significant difference in prevalence estimates for spring or summer samplings amongst farm dogs in two other areas) (Rysmukhambetova *et al.*, 2004). A study in Bangladesh identified a similar trend of increased prevalence in autumn (Islam, 1980). However, a reinfection study in Wales identified possible seasonal peaks in coproantigen prevalence in the spring and autumn months a year after the implementation of a supervised dosing scheme, with a higher estimated prevalence during the spring peak (Lett, 2013).

Reinfection studies in the eastern Tibetan plateau have identified possible seasonality in transmission of *Echinococcus* spp, but as this area is coendemic for both *E. granulosus* (s.l.) and *E. multilocularis* and coproantigen testing was used for diagnosis, it was not possible to determine the relative contribution of each species to these trends. One study following a single praziquantel dose in spring found an initial high reinfection prevalence (in late spring/early summer), followed by a lower prevalence, and then a slight increase the following spring (Moss *et al.*, 2013).

However, the exact role of seasonality in this pattern was difficult to determine since only one dose of praziquantel was administered. Another study in the same area where praziquantel was administered after each sample collection, found a similar trend – with highest apparent reinfection in late spring/early summer in one county, and in winter in the other (Wang, 2011). It was proposed that these patterns result from changes in the population density of intermediate hosts of *E. multilocularis* and increased mortality of livestock in the early spring.

6.4.5 Effect of praziquantel dosing

As described above, praziquantel dosing would be expected to be associated with a reduction in the prevalence of infection, as most studies have shown 100% efficacy against both *E. granulosus* (s.l.) and *E. multilocularis* when administered at a dose rate of 5.0 mg/kg (WHO/OIE, 2001d). Lack of praziquantel dosing has also been shown to be an important risk factor for infection in a number of previous studies (Buishi *et al.*, 2005a; b; Huang *et al.*, 2008; Acosta-Jamett *et al.*, 2010). Whilst the coefficient estimates of the effect of praziquantel dosing was negative in all models (suggesting lower prevalences amongst households which received praziquantel), this association was not found to be significant in any models. The most likely causes for this relationship (or lack thereof) are information bias from the questionnaire, ecological bias resulting from the aggregation of dogs within households, and an inability to definitively match faecal samples to households. The first of these may have resulted from people reporting dosing their dog when in fact they hadn't (because they felt they would be reprimanded for not dosing, or because the dog did not swallow the tablet despite it being offered), and vice versa (sometimes the person who answered the door was not the person responsible for dealing with the dog, and therefore may have reported no dosing when in fact the dog had been dosed by someone else). Ecological bias may result from the interpretation of praziquantel dosing history at the household level: a reported history of dosing, even if true, does not indicate that all dogs in a household had been dosed (see below for discussion of this issue in relation to the presence of young dogs in the coproantigen model). Another final possibility relates to difficulties experienced in ensuring that the samples were from the actual

registered dogs. As dogs were free to roam and defaecate throughout the village, the presence of dog faeces in the vicinity of a house does not necessarily indicate that those faeces are from that particular dog. Whilst this would not be a problem for variables identified at the community level (month of sampling, season of sampling), it would tend to reduce the magnitude of estimated model coefficients for household-level variables towards zero. Given these issues, and given the general trend towards a negative coefficient for all outcomes, it could be considered likely that recent dosing was associated with a reduction in test prevalence at the individual dog level.

6.4.6 Other identified risk factors

In the case of the coproELISA model, an identified interaction between the effect of having young dogs (less than 1 year of age) in a household and the effect of praziquantel dosing on coproELISA prevalence (see figure 6.16) suggested that the effect of praziquantel dosing amongst households with young dogs is less than that amongst households which do not have young dogs. This may indicate that younger dogs are not being dosed with praziquantel, even if praziquantel is being offered to other dogs in the household, and may therefore suggest a failure of the control scheme to reach all owned dogs in the community.

The presence of young dogs in a household appeared to be associated with a decrease in the probability of both coproantigen positivity (amongst undosed households) and *E. canadensis* G6 positivity. Interestingly, the presence of small dogs (≤10kg) in a household was found to be associated with an increased probability of test positivity for both of these tests. Due to expected collinearity between these exposures (i.e. households with young dogs were also more likely to report having small dogs), predictions for these effects were shown together in figures 6.20 and 6.21. These figures demonstrate that these predictions are complex: with households containing neither young nor small dogs, or both young and small dogs, having an 'average' test prevalence; households with young dogs but not small dogs having lower test prevalences; and households with small dogs but not young dogs having higher test prevalences. This finding was unexpected, and is difficult to interpret due to the risk of

ecological fallacy when making suggestions about the effect of dog-level effects at the household level. As such, possible reasons for this will only be briefly discussed. Age has been identified as a possible risk factor for infection in previous studies, with younger dogs having a *higher* probability of positivity (Sharifi and Zia-Ali, 1996; Buishi *et al.*, 2005b, 2006; Inangolet *et al.*, 2010; Acosta-Jamett *et al.*, 2010). However, none of these studies have categorised age with a threshold as young as one year (this age threshold was selected in an attempt to represent 'adult' dogs, but is itself an artificial construct). It has also been found that the relationship between age and prevalence/burden of infection with *Echinococcus* spp is not a linear pattern – with a low intensity observed in very young animals and a peak intensity in animals of 1-2 years of age, before reducing down again amongst older animals (if immunity is present) (Torgerson *et al.*, 2003c). It is possible that this 'peak' in young animals is being captured here. Few dog breeds in the area had an adult weight as low as 10kg, and therefore most of the households with 'non-young but small' dogs would likely have younger dogs (probably in the 1-2 year age range). Another associated possible explanation is that of feeding – with young dogs which are well fed (and which therefore have a higher bodyweight) having lower levels of scavenging, and dogs just above the age threshold which are underfed possibly having higher levels. However, further work would be required to investigate these patterns further. If individual faecal samples could be conclusively attributed to individual dogs (for example, if rectal or purge samples were taken from a selection of dogs of different ages), then further investigation could be conducted. This would be of particular importance if the force of infection was to be estimated, since modelling this (in the endemic steady state) often relies on the availability of age-stratified prevalence data (Muench, 1959; Hairston, 1965; Torgerson *et al.*, 2003c; Ziadinov *et al.*, 2008; Lewis *et al.*, 2014).

The apparent increase in *E. granulosus* G1 PCR prevalence in September 2013 was an interesting finding, although the reasons for this are unclear. In recent years, the festival of Курман айт (Eid al-Adha), during which large numbers of animals are slaughtered, has taken place during the autumn months. This would be expected to be followed by an increased in canine infection with *E. granulosus* G1, due to increased

access to offal. However, this festival took place in October in 2013, and so could not explain this trend. Whilst few dogs at this visit had received praziquantel within the previous two months (figure 6.7), this same trend was present in the other autumn visits and was not exceptional to this visit. One possibility which may be worthy of further investigation is the effect of culling campaigns in the area since a very large culling campaign was implemented in Sary-Mogol and (to a lesser extent) in Taldu-Suu in the summer and autumn of the previous year. However, this would not explain why only *E. granulosus* G1 prevalence appeared to increase at this time point. It is likely that the cause for this sudden increase in predicted *E. granulosus* G1 PCR prevalence will never be known, but this finding does demonstrate the importance of data checking and visualisation during model development.

Another method of identifying possible variables of importance is to look at the different models according to their AICc (tables 6.8-6.11), and identify a range of different models which were reasonably supported by the data. There is insufficient time to describe all of these possible associations here, but this could be worthy of further exploration. In particular, it is notable that differences between villages were only apparent for *E. multilocularis* (table 6.9), which could be due to spatial variation in the presence of suitable intermediate hosts between villages (Giraudoux *et al.*, 2002, 2003, 2006, 2013b). Further investigation of spatial patterns of *E. multilocularis* would be of interest and relevance to the planning and implementation of control campaigns.

6.4.7 Further development

One difficulty faced in the current strategy was reconciling the conflicting issues of model selection in the absence of hypothesis testing and coefficient interpretation when coefficients were 'not significant' (itself an interpretation derived from a hypothesis testing framework). The construction of a Bayesian model would reduce this problem, as coefficient estimates (and their associated credibility intervals or highest density intervals) could be interpreted directly as the best estimate of the true (uncertain) parameter values. Adopting a Bayesian approach may also reduce some of the issues regarding model convergence in the case of the PCR data, allowing

additional interaction effects to be investigated, and could also potentially allow diagnostic test limitations to be explicitly accounted for in the model. The deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) is a Bayesian measure of model fit, and behaves similarly to the AIC (Burnham and Anderson, 2004), and Bayesian methods of model averaging are also available (Hoeting *et al.*, 1999).

Another useful model development would be to interpret the coproantigen ELISA data on a continuous scale rather than dichotomising these results. As described in chapter 4, dichotomisation will invariably result in a loss of test sensitivity and/or specificity, and will also reduce the study power (Altman and Royston, 2006). Limitations in test sensitivity and specificity could lead to particular challenges for test interpretation in the face of a changing prevalence (as would be expected in the face of an effective control scheme). For example, an imperfect test specificity will result in a low positive predictive value if the prevalence is low (such as in the late stages of an effective control scheme) (Torgerson and Deplazes, 2009). As described earlier, another particular issue of relevance to echinococcosis is the fact that the test sensitivity is expected to be dependent upon the worm burdens amongst those animals tested. This means that if the burden of infection is low (again, as would be expected in the late stages of an effective control scheme), the test sensitivity will also be low. This could result in an overly optimistic interpretation of the effect of the control scheme, and therefore result in the premature cessation of control activities. A continuous outcome could be incorporated into the current model by modelling the outcome at the individual dog level, but maintaining the aggregated interpretation of risk factors (which would therefore avoid the need to match individual samples to individual dogs). Gamma and Inverse Gaussian regression models could be used to model OD data directly. Instead, if the output of the Bayesian mixture model described in chapter 4 was to be used rather than the OD, the exponent of the model scores could be used as an estimate of the worm burden in a sample, and negative binomial regression could be used. Finally, keeping with the theme of regression modelling, a variety of other outcomes could be investigated (such as reported praziquantel

administration, or dog culling) over the course of the study. These outputs could have direct benefits for the evaluation and improvement of the control scheme..

Finally, an alternative, novel, approach would be the investigation of "person-centred" latent variable methods such as Latent Class Growth Analysis (Nagin and Land, 1993), Growth Mixture Modelling (Muthén and Muthén, 2000; Muthén, 2002; Jung and Wickrama, 2008), or classification approaches for longitudinal data (Subtil *et al.*, 2014) in order to identify different household-level test prevalence trajectories over time, which can then be characterised in relation to overall household-level risk factors over the course of the study. These could also be combined with mixture modelling approaches in a Bayesian framework in order to avoid the need to dichotomise the test results, as mentioned above (Menten *et al.*, 2012). However, the use of these approaches may be constrained by the sample size – especially in the case of PCR data.

6.5 Conclusions

A mixed effects logistic regression was developed in order to investigate and quantify temporal and seasonal trends in test positivity using the coproantigen ELISA and three coproPCR tests. Interpretation was conducted at the household level, and identified a general decrease in test prevalence over time (although this was not significant in the case of *E. granulosus* G1 PCR results). Seasonal effects differed between the tests, with the coproantigen and *E. multilocularis* PCR tests suggesting lower prevalences in spring and higher prevalences in autumn. The reverse was found for *E. granulosus* (s.l.) PCR results (although again this was not significant in the case of *E. granulosus* G1). These differences may relate to seasonal variation in dog access to intermediate hosts. Whilst praziquantel use was not found to be significantly associated with a decrease in test prevalence, this likely results from a combination of information bias and ecological bias. There was some evidence that households containing young dogs had lower test prevalences, but that those containing smaller dogs had higher test prevalences. This may represent a combination of age- and feeding- related effects, and may be worthy of further study – for example, using non-categorised age and weight data at the individual dog level. A number of other possible variables of

interest are included in the model output, but are not discussed here due to space constraints. These would also be worthy of further study in order to better understand the effectiveness of the control scheme, and to possibly make some inference on parasite transmission dynamics.

These results suggest that the praziquantel dosing scheme is currently effective, and would suggest that it should be continued, with appropriate surveillance, over the coming years. As deficiencies in praziquantel administration (especially during the summer months) were identified, further work may be beneficial in order to better identify an optimal dosing strategy, accounting for *Echinococcus* transmission dynamics and logistics. As described above, a recent study in Lithuania suggested that annual targeted dosing during peak transmission times was effective in reducing the prevalence in both definitive and intermediate hosts (Šarkūnas and Deplazes, 2014), and an aspect of this is explored in chapter 7. It remains important to adopt an integrated approach to control of echinococcosis in affected areas (Torgerson, 2003b; Craig and Larrieu, 2006; Giraudoux *et al.*, 2007; Brisson *et al.*, 2011; WHO, 2011), and therefore ongoing control and surveillance in domestic dogs should involve local communities and be balanced with dog management and control strategies, ultrasound surveillance and education campaigns amongst people, and consideration should be given to methods of diagnosis of intermediate host infection if possible. This will require international collaboration between groups and institutions in different countries, whilst ensuring that all results are disseminated to relevant stakeholders (Ito *et al.*, 2003b).

Chapter 7: A mathematical modelling framework for the investigation of *Echinococcus granulosus* and *Echinococcus multilocularis* in a coendemic area.

"Those who have knowledge, don't predict. Those who predict, don't have knowledge."

Lao-Tzu (老子)

7.1 Introduction

7.1.1 Differential equation modelling

Ordinary differential equations (ODEs) are a method of describing the relationship between the rate of change of a 'dependent' variable and an independent variable (commonly, time). The solution of a differential equation will give an estimate of the value of the dependent variable for any value of the independent variable (that is, the solution is a function of the dependent variable in terms of the independent variable). Since the transmission of infectious agents can be considered in the form of a rate (the 'force of infection'), and as the dynamics of infection over a time period is often of particular interest, differential equations offer a useful framework for the modelling of infection processes (Kermack and McKendrick, 1927). A common strategy for the application of differential equations to infectious agents is to create a compartmental model, which represents the numbers of individuals in a population in different epidemiological states. For example, individuals may be classified as "susceptible" to infection, or "infected". The rate at which individuals move between these two compartments is determined by the force of infection and (in the case of infections which are not lifelong) the rate of recovery. This framework can be expanded in order to model multiple different compartments (for example, other epidemiological statuses or demographic factors), and therefore multiple different differential equations, in a single system.

Other developments from the basic model framework are stochastic models (which incorporate uncertainty and/or variability in the rates of transition between compartments), and spatial and metapopulation models (which explicitly model the spatial location/movement of hosts). The optimal choice of model framework will depend upon the system under study and the particular aims of the modelling process. It is important to note that no model will exactly represent the system under study, and it is unlikely that any individual model will be a 'perfect' model for all purposes. Instead, models can be developed to answer particular questions, and can be very useful (indeed, indispensable) for gaining a better understanding of otherwise

impenetrable transmission ecosystems; for making predictions of transmission dynamics over time and in the face of control schemes; and for relaying information to stakeholders. Mathematical models can also be useful in devising appropriate approaches to surveillance (Willeberg *et al.*, 2011), and have been suggested to be a useful tool for both the simulation of possible control schemes, and for the epidemiological and economic evaluation of these as they progress (Basáñez *et al.*, 2012a; Boatin *et al.*, 2012), including for *Echinococcus* spp (Roberts and Aubert, 1995; Gemmell *et al.*, 2001; Torgerson, 2006a; Kato *et al.*, 2010).

7.1.2 Modelling macroparasites

The history of the application of mathematical models to macroparasites (i.e. those parasites which generally do not multiply within the host) has been described in a number of publications (Anderson and May, 1985, 1991d; Basáñez *et al.*, 2012a), but will be briefly covered here. Early work was conducted by Kostitzin in 1934 who developed a model with each different worm burden represented as a compartment. In the 1960s, Hairston developed models to estimate transmission parameters for helminth infections using catalytic modelling approaches (developed by Muench (1959)), and developed the concept of R_0 for macroparasites (Hairston, 1962, 1965). Further mathematical modelling work was conducted by Macdonald (Macdonald, 1965), who developed a model of the mean worm burden (MWB) of schistosome worms in human hosts. Further investigation of overdispersion and host-parasite relationships was conducted by Crofton (Crofton, 1971b), and subsequently by May (May, 1977). These latter models accounted for both dynamic changes in the host population and in the parasite population within these hosts, and provided foundation for further work on the modelling of macroparasites (Anderson and May, 1978; May and Anderson, 1978). More recently, spatially explicit mathematical models have been developed, which may better represent the considerable impact of spatial heterogeneities on parasite transmission (Morgan *et al.*, 2004).

Mathematical modelling of macroparasites, including *Echinococcus* spp, is complicated by a number of factors:

- The complex lifecycle, involving different host species at different stages of development. Whereas domestic dog management is of particular important in both cases of infection, *E. granulosus* is also affected by livestock management practices, and *E. multilocularis* by the presence of small mammal communities (which in turn are affected by the particular environment in question: with both spatial and temporal/seasonal issues to consider). Wild canids such as foxes and wolves may also play a role in transmission dynamics and stability, despite not necessarily having a large effect on the risk of human infection.
- Difficulties in diagnosing infection in intermediate hosts (including in 'dead-end' intermediate hosts such as humans), as the cysts are slow growing, and serological tests may not detect early infections. Although the availability of portable ultrasound machines has greatly improved the ability to detect human infection (Macpherson *et al.*, 2003), this is predominantly only useful for the detection of liver cysts rather than those in other locations such as the lung.
- The need to model worm burdens within individuals explicitly in most cases. The numbers of parasites in each infected definitive or intermediate host is relevant to transmission (with possible effects on egg output, risk of host mortality, immune response, and parasite mortality, amongst others), and increases in burden generally occur only through reinfection (Heesterbeek and Roberts, 1995), rather than during multiplication within the host (as is the case with viruses and bacteria). This means that the simple compartmental models used for many microparasitic infections (such as viruses) are no longer appropriate (Anderson and May, 1991c).
- Clustering/aggregation of infection in both definitive and intermediate hosts: with most individuals having relatively low numbers of parasites, but some having very large numbers (Crofton, 1971a; Roberts *et al.*, 1986; Anderson and May, 1991b; Hansen *et al.*, 2004). This overdispersion has been described as 'one of the most important features of the epidemiology of helminth parasites' (Anderson and May, 1991a; Poulin, 2007), and is likely to have important

repercussions for parasite stability (Anderson and May, 1978; May and Anderson, 1978; Adler and Kretzschmar, 1992; Kretzschmar and Adler, 1993).

- Relatively short-lived immunity, which is generally lost following loss of the parasites.

Some of the more general difficulties associated with the attempted mathematical modelling of macroparasites have been addressed in a recent paper (Morgan *et al.*, 2004). Those issues which are suggested to be addressed include the biology, abundance and movement of host species; climatic issues; spatial and environmental issues; and parasite aggregation.

7.1.3 Mathematical modelling of *Echinococcus* spp

A number of approaches to modelling the dynamics of *Echinococcus* transmission have been described, many of which have been detailed in a recent review (Atkinson *et al.*, 2013) and in chapter 1 of the current thesis, and therefore will not be described in detail again. Although most models are compartmental models, one model of particular interest (Takumi and van der Giessen, 2005) explicitly modelled parasite biomass within these compartments, rather than focussing primarily on the host status. This model was selected for further development in the current study primarily because it avoided classifying definitive (and intermediate) host infection status in a dichotomous fashion (which, it is argued, is not the best approach for assessing the level of infection with overdispersed parasites such as *Echinococcus* – see chapter 4 and Hofer *et al.*, 2000). As environmental contamination with eggs was also explicitly modelled, the framework was also considered to be amenable to development in order to investigate the potential risk of human infection, which is often overlooked when modelling echinococcosis (Takumi *et al.*, 2012; Atkinson *et al.*, 2013). Finally, this framework explicitly incorporates the lag period between infection and infectiousness in intermediate and definitive hosts. This is of particular importance in the case of intermediate hosts, where this lag period can be months or years, and will have an impact upon control strategies (since control focussed on treatment of infection in dogs will generally need to be maintained for long periods before infection in

intermediate hosts is reduced). As a particular aim of the current modelling strategy was to evaluate the effect of praziquantel dosing on the transmission dynamics of *Echinococcus* spp, this was considered a useful feature of the model.

7.1.4 Model context

The current study aimed to develop the Takumi model (which focussed on the sylvatic lifecycle of *E. multilocularis* alone) in order to incorporate domestic dogs and therefore *E. granulosus* (sensu lato) and *E. multilocularis* infection in this host. A particular focus was placed on the investigation of trends in domestic dog infection and egg contamination, and on seasonality of transmission in the absence of control and temporal trends in the presence of praziquantel dosing. Although minimal data were available to parameterise the current model, it is hoped that further data collection will improve the model parameterisation (and that the model itself may be useful in guiding this process). The possibility of model parameterisation using canine *Echinococcus* coproantigen ELISA data is briefly explored, but this is largely an area for further exploration. As the coproELISA test is logistically easy and cheap to conduct and can be used on old faecal samples collected from the ground, it is commonly used for the evaluation of the progress of echinococcosis control schemes (such as that described in the previous chapter). The incorporation of data such as this into a mathematical model could offer considerable benefits for the evaluation of the control scheme, the planning of surveillance efforts, and in improving our understanding of the dynamics of *Echinococcus* transmission.

7.2 Materials and Methods

7.2.1 Setting of model

A decision was made to base the mathematical model on a 1km² area centred on a rural village in the Alay valley of Kyrgyzstan. The selected village was Taldu-Suu, which was chosen as it was one of the two primary study villages (see chapters 1 and 2), and was considered more representative of villages in the study area than Sary-Mogol. A map of the village, with this 1km² spatial frame of reference, is shown in

figure 7.1. The residential part of the village itself represents approximately 500m² of the 1km² area, with the remainder represented by degraded pasture (used for grazing of livestock and horses) to the south/east and the foothills of the Alay mountains to the north/east.

Figure 7.1. Geographical context for the mathematical model. The village of Taldu-Suu is shown within a 1km² box. The map on the left shows the full surrounding area, and that on the right just the area of interest. Image taken from Google Earth (satellite image taken 20th Jan, 2012 from SPOT 5 satellite)

7.2.2 Model structure

Full details of the original model structure are available in Takumi and van der Giessen (2005), and therefore most attention will be focussed here on additions and changes made to this structure. The first issue was adjusting the model in order to model infection in domestic dogs, and to incorporate *E. granulosus* (s.l.) as well as *E. multilocularis*. The original *E. multilocularis* model structure was duplicated in order to incorporate domestic dogs as an alternative definitive host, and a cycle including ruminants as intermediate hosts and domestic dogs as definitive hosts was added for *E. granulosus* (s.l.). The domestic dog cycle of *E. multilocularis* was identical in structure to the original fox cycle, but the *E. granulosus* cycle was adjusted slightly

to remove direct predation by dogs on ruminants (which was not commonly observed in the communities of interest). Instead, infection of dogs with *E. granulosus* was presumed to take place through scavenging of dead animals, or through feeding of ruminant offal to dogs following livestock slaughter (fig 7.2). In order to account for the differential effects of praziquantel dosing on the transmission dynamics of *E. granulosus* and *E. multilocularis* in dogs (potentially due to the reservoir of the latter in foxes), each of these species were modelled separately within this host. The resultant model compartments were:

- Dogs:
    - mean *E. granulosus* worm burden amongst dogs in 1km² area
    - mean *E. multilocularis* worm burden amongst dogs in 1km² area
- Foxes:
    - mean *E. multilocularis* worm burden amongst foxes in 1km² area
- Rodents:
    - mean *E. multilocularis* protoscolex burden amongst rodents in 1km² area
- Ruminants:
    - mean *E. granulosus* protoscolex burden amongst ruminants in 1km² area
- Environment:
    - mean number of *E. granulosus* eggs per km²
    - mean number of *E. multilocularis* eggs per km²

Any possible sylvatic cycle of *E. granulosus* (for example, involving wolves and wild ruminants) was ignored due to the lack of available data on this existence of this cycle in the study area. However, if evidence of this became available, the model structure could be easily modified.

A crude attempt to incorporate density dependence in the mean worm and mean protoscolex burdens was made. Although both worm numbers and protoscolex numbers within the host can become very high (thousands to tens of thousands), it would be expected that the mean of these estimates throughout the population as a whole would gradually reach a plateau. In order to not overcomplicate the current model, a logistic growth process was attached to the variables describing the mean

worm and mean protoscolex burden. This was repeated for both *E. granulosus* and *E. multilocularis*, assuming no effect of either of these on the other.

The model also differed from the original model in that removal of intermediate hosts was not assumed to be a completely random process throughout the whole population. Rather, the mean worm/protoscolex burden of animals being removed was adjusted using a scaling factor, κ. In the current model, this parameter was used to account for the fact that intermediate hosts removed from the population are more likely to be older (and therefore more likely to be infected), although it could also be used to represent differential mortality due to infection (Poulin, 1995; Vervaeke *et al.*, 2006). Parameterisation was based upon crude estimates of age-related differences in the probability of intermediate host infection (Torgerson *et al.*, 2009a; Burlet *et al.*, 2011).

For ease of interpretation, the parameters relating to proglottid production (per worm per day) and mean egg burden (per proglottid) included in the original model were consolidated, in order to estimate the rate of egg production per worm per day directly. It has been suggested that most eggs are liberated from proglottids prior to excretion (Wachira *et al.*, 1991), which makes this simplification reasonable. It was also assumed that removal of eggs from pasture due to ingestion by intermediate hosts, despite being included in the original model, had a minimal effect on the total egg contamination, and so this process was removed from the model.

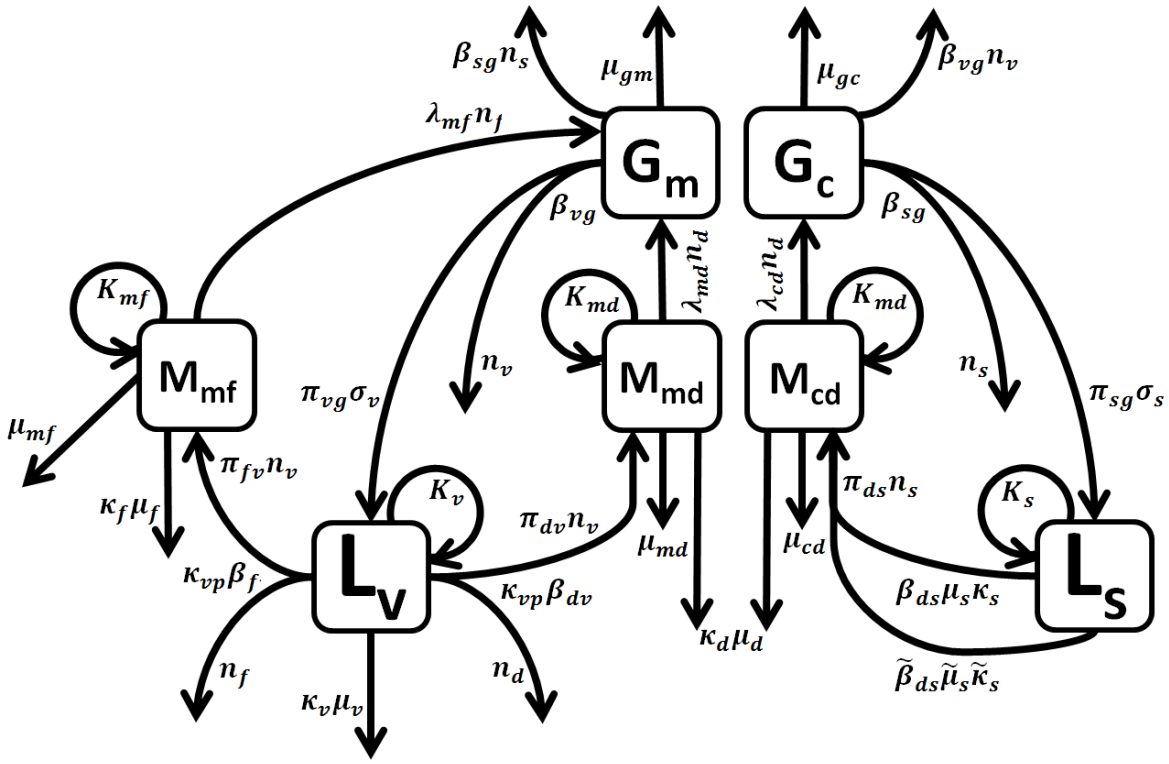


Figure 7.2. Conceptual structure of model. Parameters are described in table 7.4. G_m: *E. multilocularis* eggs on pasture; G_c: *E. granulosus* eggs on pasture; L_v: *E. multilocularis* protoscolices in small rodents; L_s: *E. granulosus* protoscolices in ruminants; M_{mf}: Adult *E. multilocularis* worms in foxes; M_{md}: Adult *E. multilocularis* worms in dogs; M_{cd}: Adult *E. granulosus* worms in dogs.

7.2.3 Differential equations

A schematic diagram of the basic model structure is shown in figure 7.2. The differential equations for the model are as follows:

Eggs:

$$\frac{dGm}{dt} = \lambda_{md} n_d M_{md} + \lambda_{mf} n_f M_{mf} - \mu_g G_m$$

$$\frac{dGc}{dt} = \lambda_{cd} n_d M_{cd} - \mu_g G_c$$

Larvae (protoscolices):

$$\frac{dLv}{dt} = \beta_{vg}\pi_{vg}\sigma_v G_{m(t-\tau_v)}\left(1 - {L_v}/{K_v}\right) - \mu_v\kappa_v L_v - \beta_{dv} n_d \kappa_{vp} L_v - \beta_{fv} n_f \kappa_{vp} L_v$$

$$\frac{dLs}{dt} = \beta_{sg}\pi_{sg}\sigma_s G_{m(t-\tau_s)}\left(1 - {L_s}/{K_s}\right) - (\mu_s\kappa_s + \tilde{\mu}_s\tilde{\kappa}_s)L_s$$

Adult worms:

$$\frac{dMmd}{dt} = \beta_{dv}\pi_{dv} n_v \kappa_{vp} L_{v(t-\tau_{md})}\left(1 - {M_{md}}/{K_{md}}\right) - \mu_d\kappa_d M_{md} - \mu_{md} M_{md}$$

$$\frac{dMcd}{dt} = \left(\beta_{ds}\mu_s\kappa_s + \tilde{\beta}_{ds}\tilde{\mu}_s\tilde{\kappa}_s\right)\pi_{ds} n_s L_{s(t-\tau_{cd})}\left(1 - {M_{cd}}/{K_{cd}}\right) - \mu_d\kappa_d M_{cd} - \mu_{cd} M_{cd}$$

$$\frac{dMmf}{dt} = \beta_{fv}\pi_{fv} n_v \kappa_{vp} L_{v(t-\tau_{mf})}\left(1 - {M_{mf}}/{K_{mf}}\right) - \mu_f\kappa_f M_{mf} - \mu_{mf} M_{mf}$$

7.2.4 Parameterisation

Model parameters are listed in table 7.4. The model was largely parameterised directly using published data, data collected from the Alay valley, and personal observation. The transmission parameters (β) were parameterised using data from force of infection models in Kyrgyzstan and Switzerland, and their calculation are described below.

The rate of ingestion of intermediate hosts by dogs and foxes was not known. However, the rate of infection of free-roaming domestic dogs in Naryn province, Kyrgyzstan, has been estimated at around 1.1 for *E. multilocularis* and 0.3 for *E. granulosus* (Ziadinov *et al.*, 2008). This can be used to broadly estimate the probability of ingestion of a suitable intermediate host (β_{dv}). If it is assumed that multiple infections do not occur (due to the low infection pressure on a daily basis per individual animal), and that all animals in the population at the steady state are infected (which, although epidemiologically implausible, is the basis of the model structure used here), the daily rate of infection can be estimated as the product of the rate of ingestion, the probability of infection given ingestion, and the number of intermediate hosts. For *E. multilocularis*, this can be approximated as $\beta_{dv}\pi_{dv}n_v$, and for *E. granulosus*, this will be $(\beta_{ds}\mu_s + \tilde{\beta}_{ds}\tilde{\mu}_s)\pi_{ds}n_s$. These can therefore be used to estimate the transmission parameters, based on the estimates mentioned before (Ziadinov *et al.*, 2008):

$$\beta_{dv}\pi_{dv}n_v = {}^{1.1}/_{365}$$

$$\beta_{dv} = \frac{{}^{1.1}/_{365}}{\pi_{dv}n_v} = 6.03 \times 10^{-6}$$

If it is assumed that 5% of slaughtered ruminants are fed to dogs (see section 7.4.3), the proportion of ruminants which die of natural causes and are scavenged by dogs can be estimated:

$$(\beta_{ds}\mu_s + \tilde{\beta}_{ds}\tilde{\mu}_s)\pi_{ds}n_s = {}^{0.3}/_{365}$$

$$\beta_{ds}\mu_s + \tilde{\beta}_{ds}\tilde{\mu}_s = \frac{{}^{0.3}/_{365}}{\pi_{ds}n_s}$$

$$\beta_{ds} = \frac{\left(\frac{{}^{0.3}/_{365}}{\pi_{ds}n_s} - \tilde{\beta}_{ds}\tilde{\mu}_s\right)}{\mu_s} = 0.05$$

A study in Naryn province of Kyrgyzstan found high levels of infection of red foxes with *E. multilocularis*, with an estimated prevalence of 64% and a mean burden of 8,669 worms (Ziadinov *et al.*, 2010), although the force of infection was not estimated in this study. Another study using data from Switzerland, where the estimated prevalence was 66% (in periurban areas) indicated a median force of infection of 3.8 infections/year (Lewis *et al.*, 2014). Assuming that due to the similar prevalences, this is broadly representative of the Kyrgyz situation, an estimate can be made of the rate of intermediate host ingestion using the same strategy described above:

$$\beta_{fv}\pi_{fv}n_v = {}^{3.8}/_{365}$$

$$\beta_{fv} = \frac{{}^{3.8}/_{365}}{\pi_{fv}n_v} = 5.21 \times 10^{-6}$$

Estimation of the rate of egg ingestion was based upon the approach described in (Takumi and van der Giessen, 2005), but incorporating the average number of protoscolices per egg ingested. If it can be assumed that the current prevalence of echinococcosis in intermediate hosts is at a steady state, then it would be expected that the number of new protoscolices is equal to the number of lost protoscolices (in the case of *Echinococcus*, this only takes place through death). Therefore, for *E. multilocularis*:

$$\beta_{vg}\pi_{vg}\sigma_v n_v \approx (\beta_{dv}n_d + \beta_{fv}n_f + \mu_v)$$

$$\beta_{vg} \approx \frac{(\beta_{dv}n_d + \beta_{fv}n_f + \mu_v)}{\pi_{vg}\sigma_v n_v} = 1.45 \times 10^{-8}$$

And for *E. granulosus*:

$$\beta_{sg}\pi_{sg}\sigma_s n_s \approx (\mu_s + \tilde{\mu}_s)$$

$$\beta_{sg} \approx \frac{(\mu_s + \tilde{\mu}_s)}{\pi_{sg}\sigma_s n_s} = 1.04 \times 10^{-6}$$

Seasonality in egg survival (table 7.1) was modelled based upon data obtained from Karakenja in Tajikistan (figure 7.3) (weatherbase.com, 2015), and using the formula in Ishikawa *et al.* (2003):

$$duration\ of\ infectivity = \exp(-0.135 \times |temp| - 43.49)$$

Table 7.1. Average monthly temperatures in Karakenja, Tajikistan and associated predicted durations of egg survival.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temp (°c)	-11	-9	-2.2	5.5	9.5	13.8	17.0	16.4	11.6	5.2	-1.4	-7.1
Egg survival (days)	80	105	264	169	98	55	36	39	74	176	294	136

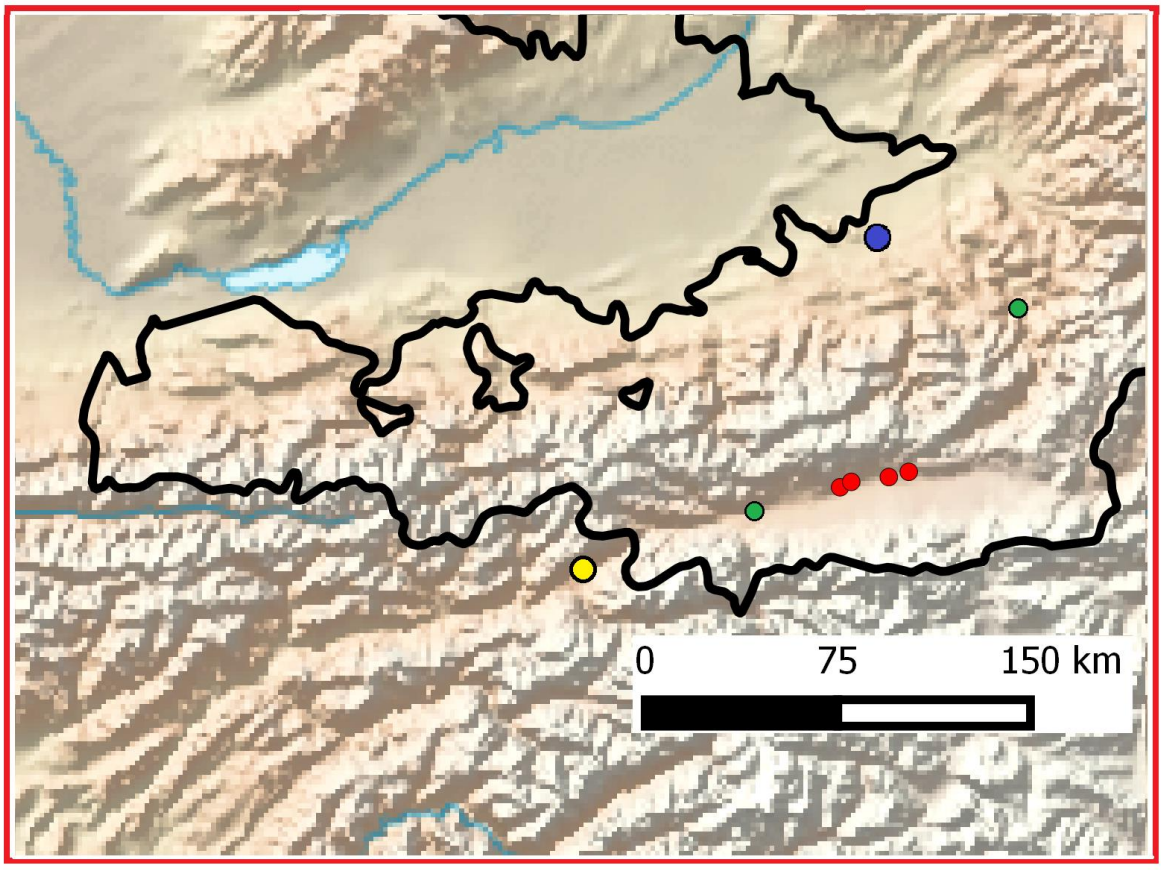


Figure 7.3. Location of Karakenja in Tajikistan (yellow circle), in relation to study villages (red circles). Taldu-Suu is the easternmost study village.

Fox population densities were modelled assuming a litter of five cubs per breeding pair of foxes emerge in June and disperse in October. For domestic small ruminants, whilst seasonal breeding would be expected to be associated with an increase in population size in the spring months, this also coincided with a general movement to Jailoo. Therefore, the population density within the village was assumed to remain broadly stable throughout the year. Rodent population densities were assumed to halve in the early spring (as young disperse), and then increase in the summer and early autumn, according to a study of the northern mole vole, *Ellobius talpinus* (Evdokimov, 2013). Although domestic dogs are not seasonal breeders, many dogs left the village in the summer months to travel to Jailoo. Based upon personal observation and estimates of dog faecal densities in the spring and autumn months (van Kesteren *et al.*, 2013), the population was presumed to halve during this time (table 7.2).

Table 7.2. Seasonal population density parameters used in the model. Variables are detailed in table 7.4.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rodent pop	n_v			$n_v/2$			$2n_v$			n_v		
Ruminant pop	n_s											
Fox pop	n_f					$n_f + (5n_f/2)$				n_f		
Dog pop	n_d				$n_d/2$					n_d		

Relative changes in the availability of rodents and (dead) ruminants for ingestion by definitive hosts were modelled by halving the beta parameters indicating reliance on these intermediate hosts for food. Therefore, the rate of ingestion of rodents reduced during the winter months (when rodents would be expected to be difficult to find due to snow cover (Heptner and Naumov, 1992)). Mortality amongst intermediate hosts was also modelled seasonally, assuming that mortality and livestock slaughter increased during the harsh winter months. Ruminant access by dogs would therefore be expected to increase in these months (table 7.3).

Table 7.3. Seasonal mortality and prey preference parameters in model

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rodent mortality	$2\mu_v$			$\mu_v/2$								$2\mu_v$
Reliance on rodents	$\beta_{fv}/2$			β_{fv}							$\beta_{fv}/2$	
	$\beta_{dv}/2$			β_{dv}							$\beta_{dv}/2$	
Ruminant mortality	$2\mu_s$			$\mu_s/2$								$2\mu_s$
Ruminant slaughter	$2\tilde{\mu}_s$			$\tilde{\mu}_s/2$						$2\tilde{\mu}_s$		

Due to the exploratory nature of the current study, and the very large numbers of parameters (many of which are not based on firm field data), the only form of

sensitivity analysis conducted was on the effect of removal of seasonality from the seasonal parameters. This was achieved by setting the values to μ_v , β_{fv}, β_{dv}, μ_s, and $\tilde{\mu}_s$ (see table 7.4 for values).

Table 7.4. Parameters included in the model

Parameter	Description	Value	Reference
μ_f	Rate of fox death	0.0022 foxes day^{-1}	(Takumi and van der Giessen, 2005)
κ_f	Relative mortality of infected foxes	1.0	
μ_d	Rate of dog death	1/(5*365) dogs day^{-1}	[personal observation]
κ_d	Relative mortality of infected dogs	1.0	
μ_{mf}	Rate of adult *E.m* death in foxes	1/30 worms day^{-1}	(Kapel *et al.*, 2006)
μ_{md}	Rate of adult *E.m* death in dogs	1/((365/3.4)-30) worms day^{-1}	(Ziadinov *et al.*, 2008)
μ_{cd}	Rate of adult *E.g* death in dogs	1/((365/1.3)-40) worms day^{-1}	(Ziadinov *et al.*, 2008)
μ_v	Rate of rodent death	Seasonal variable. Estimates based on 0.0055 rodents day^{-1}	(Roberts and Aubert, 1995)
κ_v	Relative mortality of infected rodents	2.0	(Burlet *et al.*, 2011)
μ_s	Rate of natural ruminant death	Seasonal variable. Baseline 1/(5*365) ruminants day^{-1}	(Torgerson *et al.*, 2009a)
$\tilde{\mu}_s$	Rate of ruminant slaughter	Seasonal variable. Baseline 4/(100*365) ruminants day^{-1}	
κ_s	Relative natural mortality of infected ruminants	2.0	(Torgerson *et al.*, 2009a)

Parameter	Description	Value	Reference
$\tilde{\kappa}_s$	Relative rate of slaughter of infected ruminants	2.0	(Torgerson *et al.*, 2009a)
μ_{gm}	Rate of *E.m* egg death on pasture	Seasonal variable. Estimate based upon formula given in (Ishikawa *et al.*, 2003)	Substantiated by (Veit *et al.*, 1995)
μ_{gc}	Rate of *E.g* egg death on pasture	Seasonal variable. Estimate based upon formula given in (Ishikawa *et al.*, 2003)	Substantiated by (Wachira *et al.*, 1991)
σ_v	Mean number of protoscolices in rodents resulting from infection with one *E.m* egg	926,239/81 = 11,435 protoscolices	(Stieger *et al.*, 2002)
σ_s	Mean number of protoscolices in ruminants resulting from infection with one *E.g* egg	437 protoscolices	(Torgerson *et al.*, 2009a)
π_{fv}	Proportion of rodent protoscolices expected to develop into adult worms after ingestion by foxes	0.4	(Takumi and van der Giessen, 2005)
π_{dv}	Proportion of rodent protoscolices expected to develop into adult worms after ingestion by dogs	0.1	(Kapel *et al.*, 2006)

Parameter	Description	Value	Reference
π_{ds}	Proportion of ruminant protoscolices expected to develop into adult worms after ingestion by dogs	0.05	(Gemmell *et al.*, 1990)
π_{vg}	Proportion of *E.m* eggs expected to result in infection in rodents	0.007	(Takumi and van der Giessen, 2005)
π_{sg}	Proportion of *E.g* eggs expected to result in infection in ruminants	0.003	(Gemmell *et al.*, 1990)
λ_{mf}	Rate of egg release from adult *E.m* in foxes	0.14 proglottid day^{-1} x 300 eggs $proglottid^{-1}$ = 42 eggs $worm^{-1}$ day^{-1}	(Matsudo *et al.*, 2003; Hansen *et al.*, 2003; Takumi and van der Giessen, 2005)
λ_{md}	Rate of egg release from adult *E.m* in dogs	42 eggs $worm^{-1}$ day^{-1}	Assumed to be comparable to other estimates (see discussion)
λ_{cd}	Rate of egg release from adult *E.g* in dogs	42 eggs $worm^{-1}$ day^{-1}	(Gemmell *et al.*, 1986c; Torgerson and Heath, 2003)
β_{fv}	Proportion of total rodents in 1km^2 area ingested by foxes	Seasonal variable. Baseline estimated as $5.2x10^{-6}$ fox^{-1} day^{-1}	(Ziadinov *et al.*, 2010; Lewis *et al.*, 2014)
β_{dv}	Proportion of total rodents in 1km^2 area ingested by dogs	Seasonal variable. Baseline estimated $6.0x10^{-6}$ fox^{-1} day^{-1}	(Ziadinov *et al.*, 2008)

Parameter	Description	Value	Reference
κ_{vp}	Relative predation risk of infected rodents	2.0	
β_{ds}	Proportion of total dead ruminants in 1km² area scavenged by dogs	0.02	
$\tilde{\beta}_{ds}$	Proportion of total slaughtered ruminants in 1km² area fed to dogs	0.05	[personal observation]
β_{vg}	Proportion of total *E. m* eggs in 1km² area ingested by rodents	1.45 x 10^{-8} rodent^{-1} day^{-1}	Used approach based upon (Takumi and van der Giessen, 2005)
β_{sg}	Proportion of total *E. g* eggs in 1km² area ingested by ruminants	1.05 x 10^{-6} ruminant^{-1} day^{-1}	Used approach based upon (Takumi and van der Giessen, 2005)
n_f	Mean fox population density	2 fox km^{-2}	Based upon estimates in (Takumi and van der Giessen, 2005)
n_d	Mean dog population density	Seasonal variable. Estimates based upon baseline of 100 dogs km^{-2}	[personal observation / (van Kesteren *et al.*, 2013)]
n_v	Mean rodent population density	5000 rodents km^{-2}	[personal observation]
n_s	Mean ruminant population density	500 ruminants km^{-2}	[personal observation]
τ_{mf}	*E.m* prepatent period in foxes	30 days	(Kapel *et al.*, 2006)

Parameter	Description	Value	Reference
τ_{md}	*E.m* prepatent period in dogs	30 days	(Kapel *et al.*, 2006)
τ_{cd}	*E.g* prepatent period in dogs	42 days	(Gemmell *et al.*, 1986c)
τ_v	Maturation time of *E.m* protoscolices in rodents	112 days	(Matsumoto *et al.*, 1998)
τ_s	Maturation time of *E.g* protoscolices in ruminants	(365*2) days	(Gemmell *et al.*, 1986c)
K_{mf}	Mean *E.m* burden at 'carrying capacity' in foxes	16000 worms	(Kapel *et al.*, 2006)
K_{md}	Mean *E.m* burden at 'carrying capacity' in dogs	2534 worms	(Kapel *et al.*, 2006)
K_{cd}	Mean *E.g* burden at 'carrying capacity' in dogs	2500 worms	(Torgerson and Heath, 2003)
K_v	Mean protoscolex burden at 'carrying capacity' in rodents	244400 protoscolices	(Stieger *et al.*, 2002)
K_s	Mean protoscolex burden at 'carrying capacity' in ruminants	9774 protoscolices	(Torgerson *et al.*, 2009a)

7.2.5 Model building

The model was created using the "deSolve" package (Soetaert *et al.*, 2010) in R version 3.1.1 (R Development Core Team, 2014). Model code is shown in the appendix (A6). The lag period between infection and infectiousness in intermediate hosts was modelled using the "lagvalue" function, and seasonal variation was modelled by specifying trends by month (see table 7.2) and then using the "approxfun" command to approximate this trend in a continuous nature for incorporation into the model code. The model was initially seeded with a mean worm burden of 100 *E. granulosus* and 100 *E. multilocularis* adult worms amongst dogs, and was run with a timestep of one day for a total period of 100 years.

7.2.6 Simulation of control strategies

Dog dosing interventions were added to the system once a steady state was reached (10 years into the simulation), and were modelled by directly changing the value of the mean burden in domestic dogs at particular time points. It was assumed that praziquantel efficacy was 100%, and that the only factor therefore affecting the resultant mean burden was the praziquantel coverage (the proportion of domestic dogs treated), ρ. The resultant mean worm burden after dosing was estimated as the product of the original mean worm burden and the difference between unity and the proportion of praziquantel coverage:

$$MWB_{t+1} = MWB_t \cdot (1 - \rho)$$

For the purposes of the current study, complete praziquantel coverage ($\rho = 1$) was assumed. Although praziquantel has no residual effect, the number of adult worms following administration would be expected to remain at zero until at least the prepatent period of the worm had passed. This lag period was considered to be of particular relevance to egg contamination, and was incorporated in the current model by fixing the MWB in dogs to zero for each day after initial dosing until the prepatent period of the worm had passed. Following this time, reinfection took place according to the status of intermediate hosts at the time of praziquantel dosing (accounting for

the temporal lag in development of patency) – representing the maturation of any protoscolices ingested immediately after dosing.

A number of control strategies based upon praziquantel dosing of dogs were briefly investigated. Firstly, a variety of frequencies of dosing were simulated. These were daily (i.e. the complete removal of all canine infection); monthly; every two, three, and six months; and once yearly. Using the current parameters for prepatent period, there was no difference between daily and monthly dosing, and therefore the daily dosing effect was no longer specifically investigated here (although it would provide a useful baseline for investigation of stochastic models, where even monthly dosing may not completely prevent reinfection due to stochasticity in the prepatent period – discussed later). Most control schemes are based upon routine, regular dosing throughout the year. In the Alay valley, the planned frequency of dosing was every three months, although the most common reported frequency of dosing was twice annually (personal observation).

Due to the inclusion of a number of seasonal parameters in the model, seasonal-based targeted dosing strategies were investigated, in the hope that these may offer a more economically and logistically viable control strategy. The timing of this dosing was selected using a manual stepwise process. The effect of a single praziquantel treatment of all dogs in the community was first investigated by simulating this dosing strategy in each month. The effect of each of these timed doses was evaluated by estimating the average mean worm burden and average egg contamination for each species of *Echinococcus* over a period of one year. The difference between each of these estimates and that predicted in the absence of any control is shown figure 7.4. Since reduction in egg output would reduce the risk of human infection, and since alveolar echinococcosis (caused by *E. multilocularis*) appears to be the predominant form of human echinococcosis in the Alay valley (Paul Torgerson, personal communication), a decision was made to prioritise reduction in *E. multilocularis* egg output in the current strategy (top left graph in figure 7.4). This suggested that a single dose in October maximised the reduction in *E. multilocularis* egg contamination.

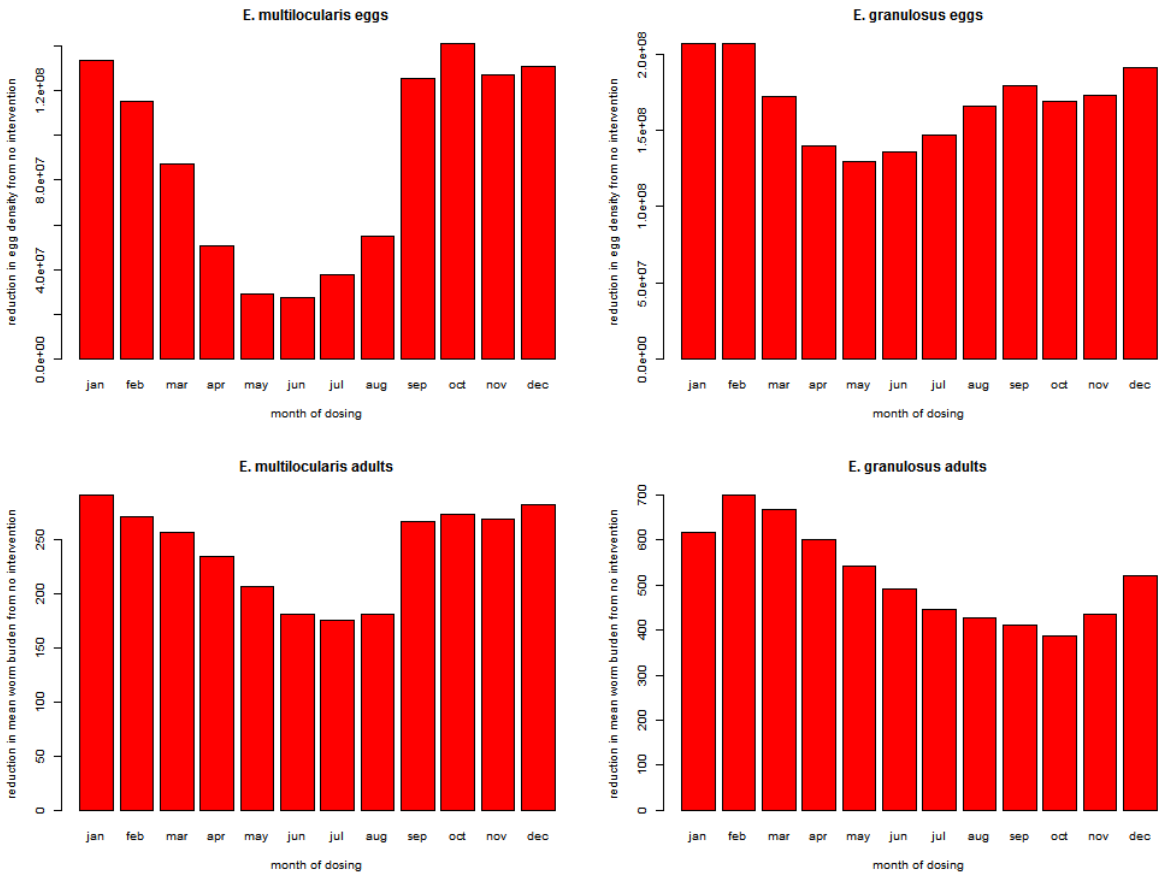


Figure 7.4. Reduction in average daily egg contamination (top) or mean worm burden (bottom) for *E. multilocularis* (left) and *E. granulosus* (right) over a year with different months of administration of a single dose of praziquantel to every member of the dog population

A planned single dosing event in October was then simulated and the process of selecting a second month was repeated as before. The October dosing strategy was used as the comparison minimising egg contamination. This approach identified January as the optimal second month, and the process was repeated again twice more – identifying August as the optimal third month, and September as the optimal fourth month. The dynamics of infection for these four targeted interventions (one, two, three, and four temporally targeted doses annually) were then directly compared with the dynamics in the absence of control and in the presence of regular dosing one, two, three and four times annually.

7.2.7 Interpretation of results

The model output was investigated in three ways. Firstly, general temporal trends and relationships between the different life stages of *E. granulosus* and *E. multilocularis* were investigated by scaling the burden estimates in each host (including the environment, in the case of eggs) as a proportion of the maximal burden (as was conducted in Takumi and van der Giessen, 2005). This allowed graphs of the different life stages to be overlaid, and facilitated visual identification of the sequential stages of infection in the lifecycle. Secondly, the predicted mean burden estimates in each host over time were estimated directly for each species of *Echinococcus*, in the presence and absence of control. For each of these approaches, the dynamics were investigated for two time periods for each of the control and non-control scenarios: the period immediately after the initial 'seeding' of a single infected dog into the population (or the initial period after commencement of the intervention); and once the 'steady state' had been reached (10 years after initial seeding, or 10 years after commencement of the control scheme). Finally, the relative effect of the control scheme on model predictions was estimated by dividing the model outputs under control by the maximal burden estimate in the absence of control).

7.3 Results

The relative dynamics of infection for the different species are shown in figure 7.5, which have been standardised to the maximal mean burden/density of each compartment for ease of visualisation. The initial trends in absolute mean burden/density following initial seeding are shown in figure 7.6, and seasonal trends in the steady state (10 years onwards) in the absence of any control are shown in figure 7.7. Figure 7.8 shows the effect of different frequencies of dosing on model components, over a period of three years, and figures 7.9-7.12 show these same effects following complete or selective removal of seasonal variation in the model seasonal parameters.

Figure 7.13 shows the predicted effect of a variety of control strategies on model compartments over the course of 100 years (from initial seeding), and figure 7.14 shows the predicted long-term levels of intermediate host infection (which indicates the level of persistence of transmission) under different seasonality assumptions. Figure 7.15 demonstrates the effect of administering a single dose of praziquantel annually during each month to all dogs in a community (and the simulated estimates in the absence of any control), at the new steady state for a period of three years. Figures 7.16 and 7.17 demonstrate the relative effect of targeted dosing (at the periods of greatest impact on *E. multilocularis* egg contamination) in comparison to both no dosing and untargeted (regular) dosing for three years at the new steady state, over a period of one year.

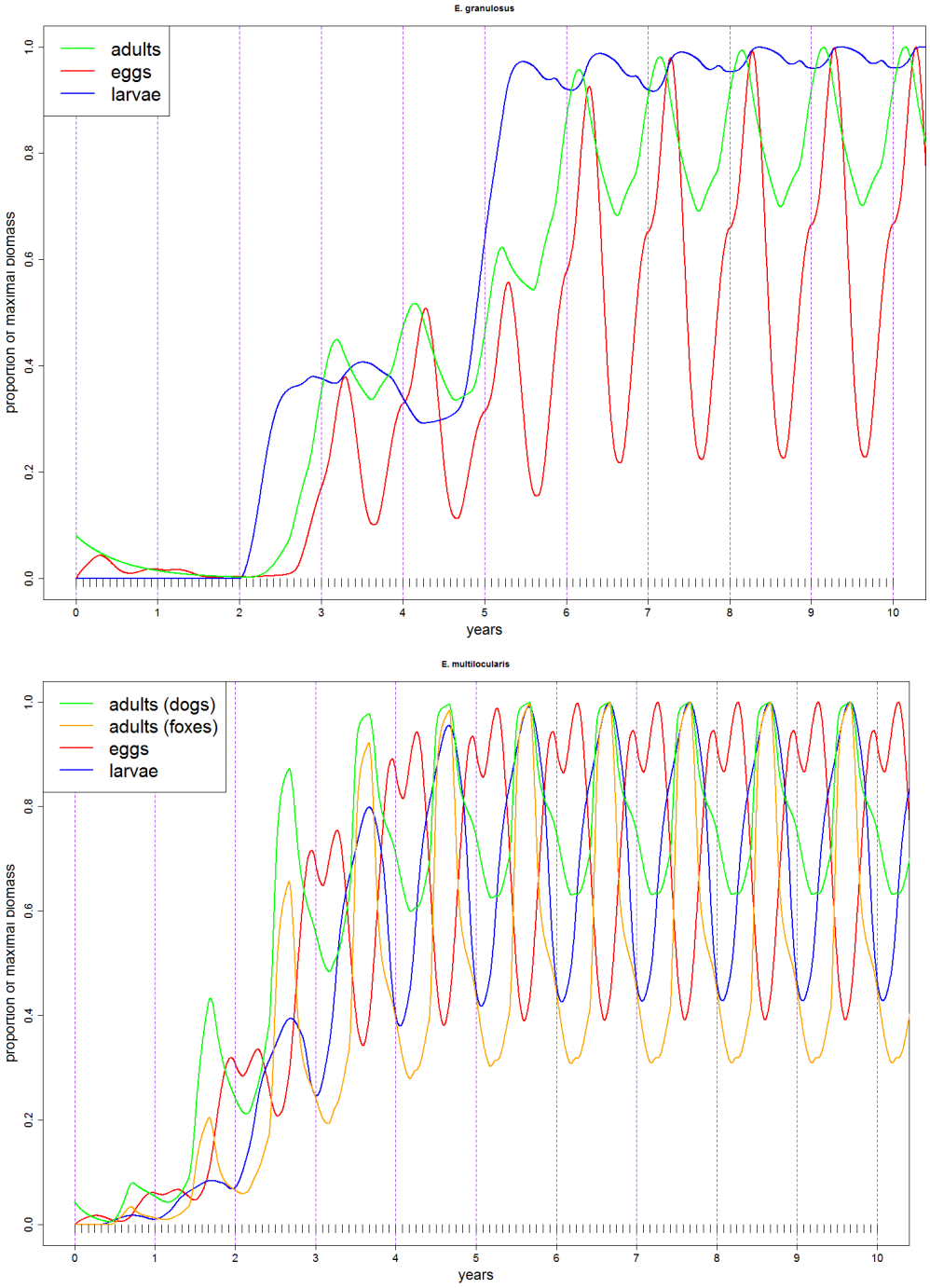


Figure 7.5. Relative dynamics of infection during the initial stages after seeding, over a course of 10 years. Values are expressed relative to the maximal estimate for the compartment in question.

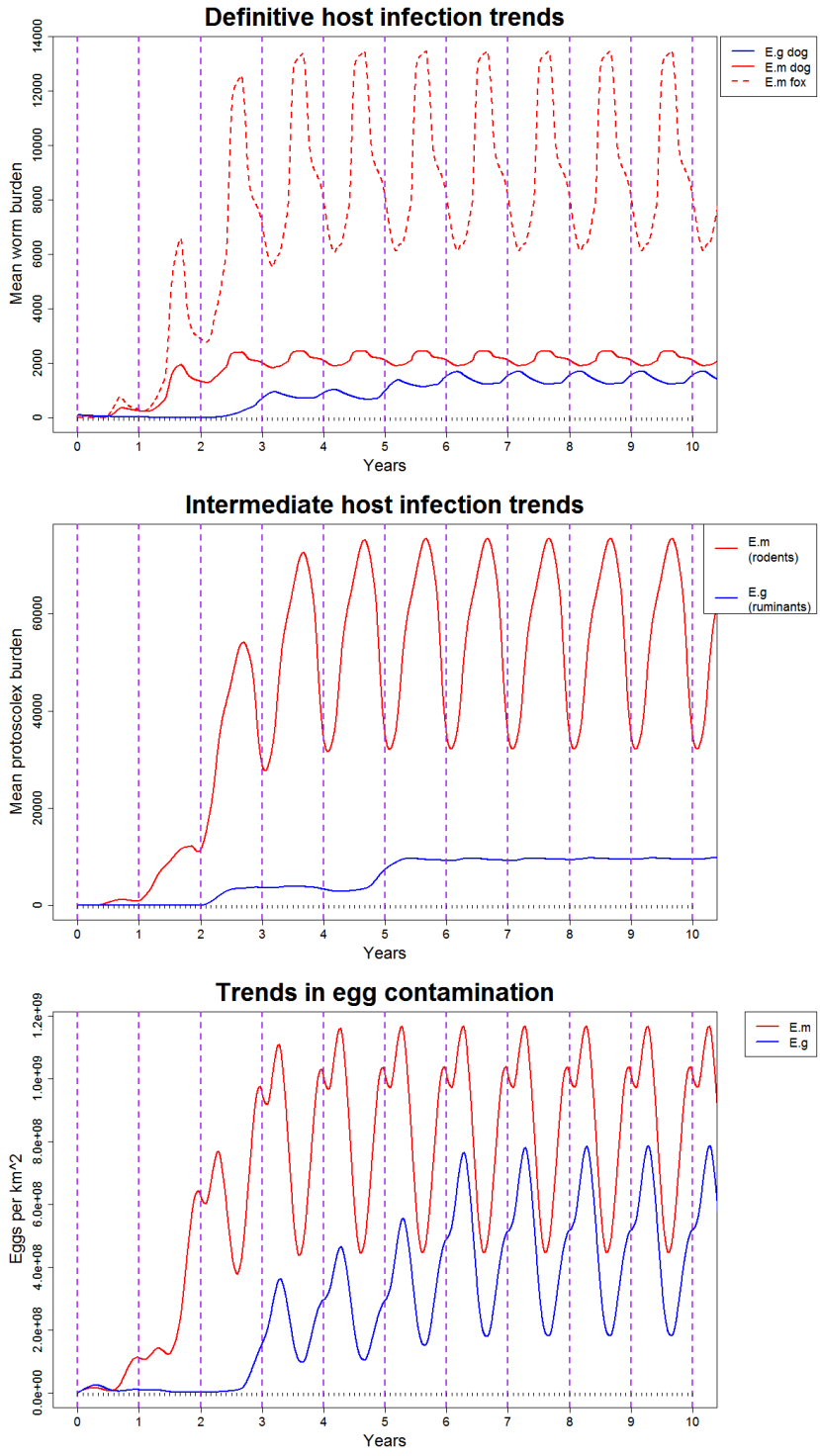


Figure 7.6. Initial trends (first 10 years) of *E. granulosus* and *E. multilocularis* worm burden (top), cyst burden (middle) and egg contamination (bottom) following seeding of the model with a mean worm burden for *E. granulosus* and *E. multilocularis* of 100.

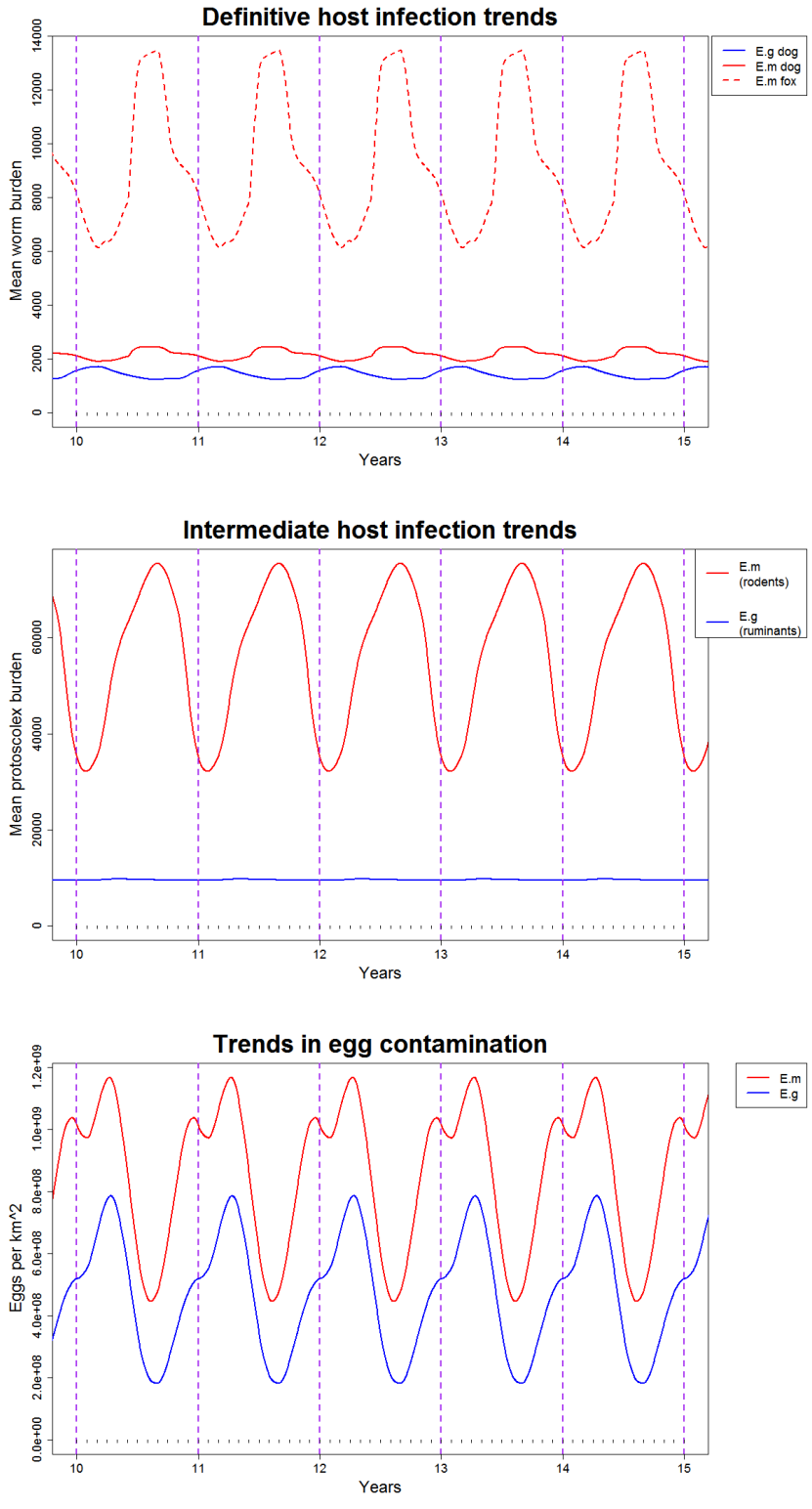


Figure 7.7. Trends in *E. granulosus* and *E. multilocularis* worm burden (top), cyst burden (middle) and egg contamination (bottom) 10 years after initial seeding.

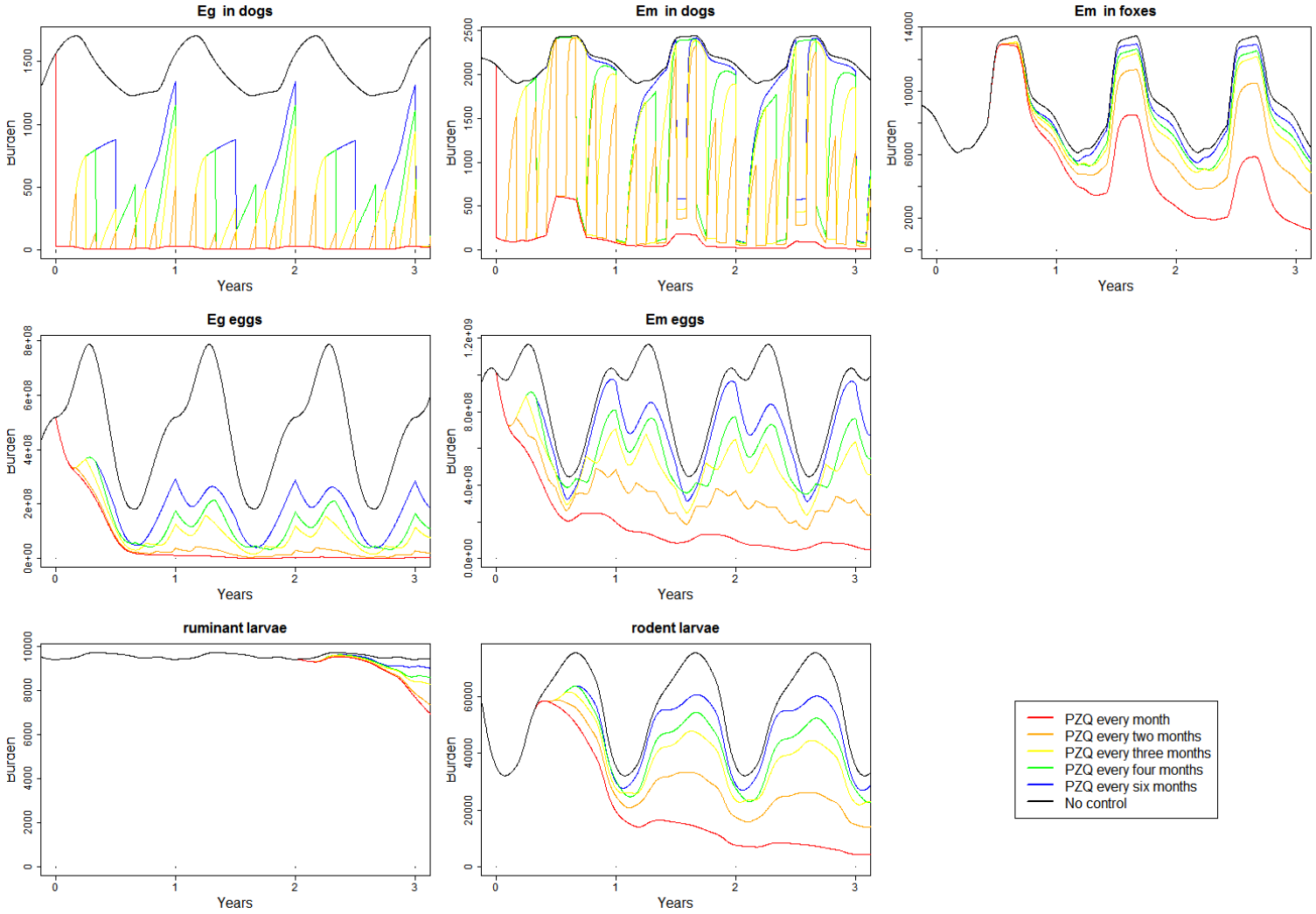


Figure 7.8.Effect of three years of regular praziquantel dosing on the simulated mean *E. granulosus* (left column) and *E. multilocularis* (centre and right columns) burden/contamination

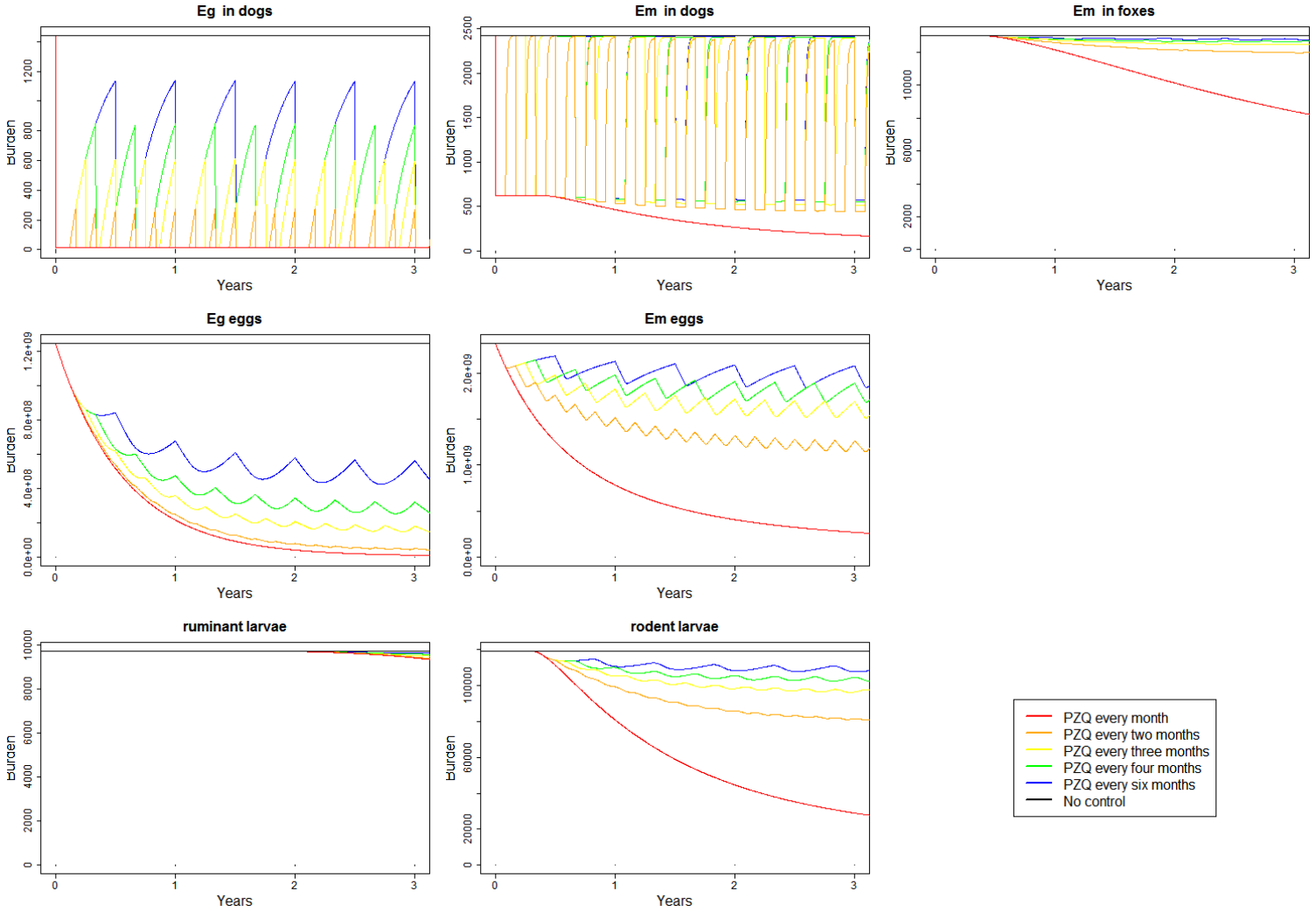


Figure 7.9. Effect of three years of regular praziquantel dosing on the simulated mean *E. granulosus* (left column) and *E. multilocularis* (centre and right columns) burden/contamination with no seasonality in any parameters

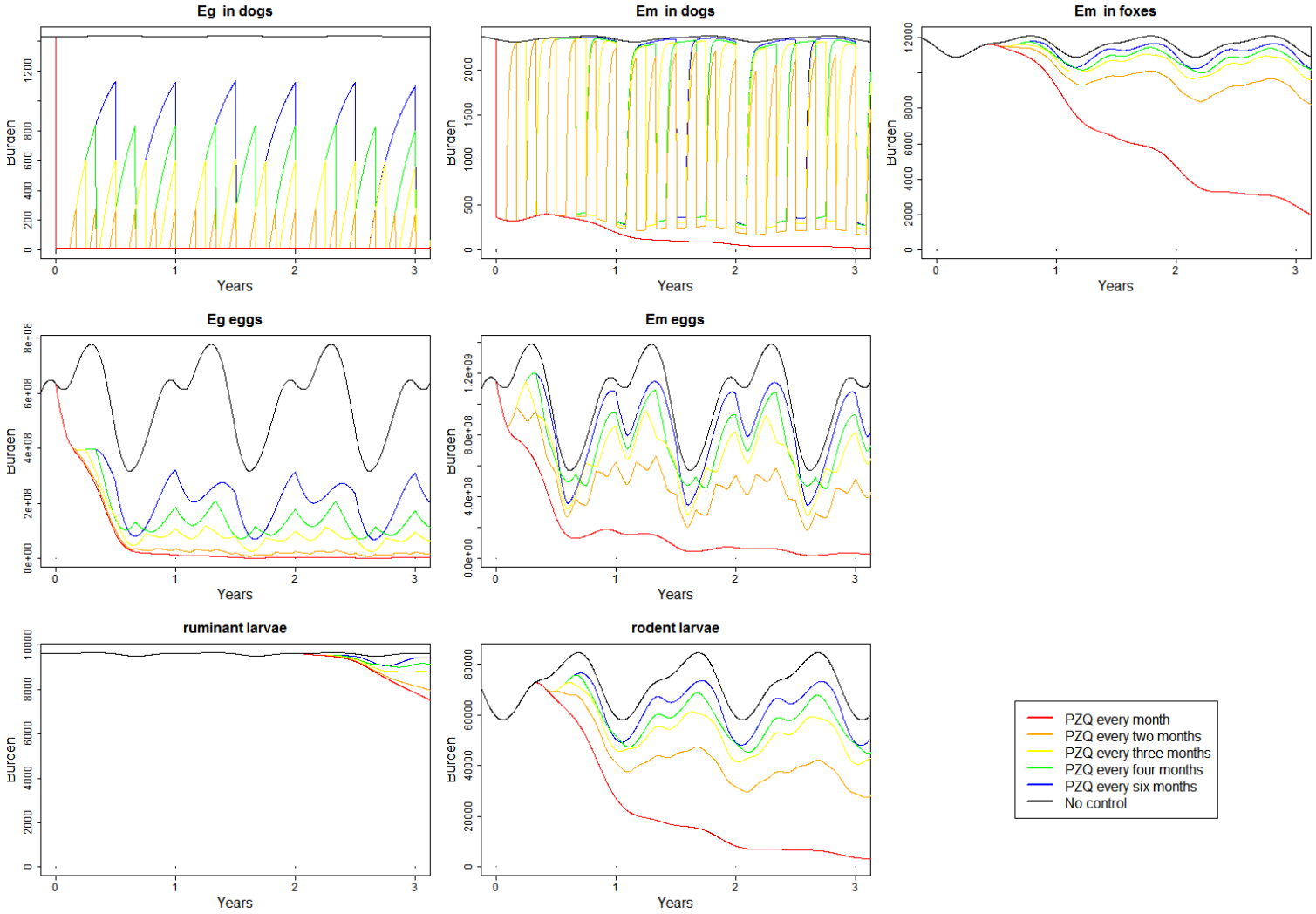


Figure 7.10. Effect of three years of regular praziquantel dosing on the simulated mean *E. granulosus* (left column) and *E. multilocularis* (centre and right columns) burden/contamination with seasonality in egg survival only

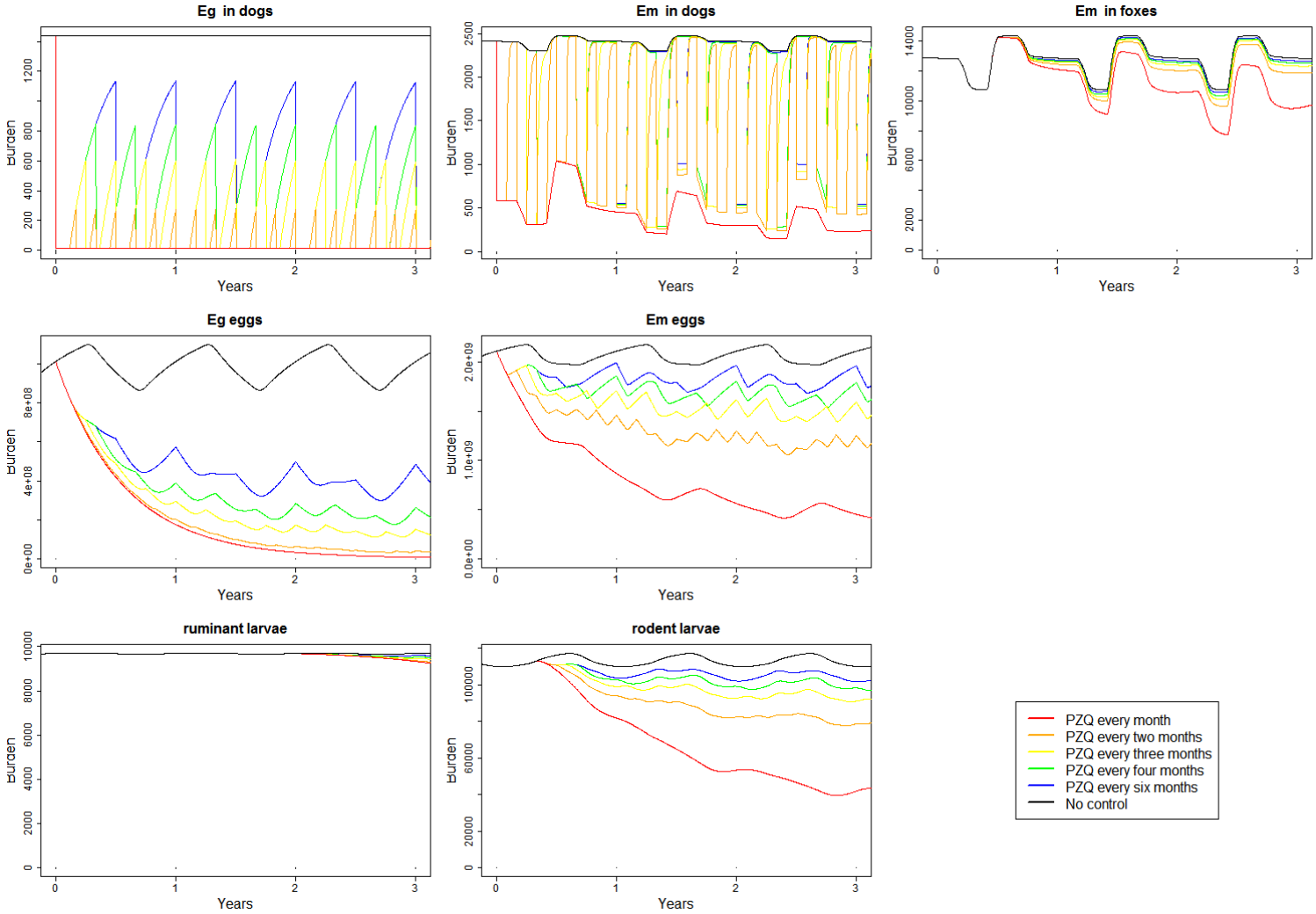


Figure 7.11.Effect of three years of regular praziquantel dosing on the simulated mean *E. granulosus* (left column) and *E. multilocularis* (centre and right columns) burden/contamination with seasonality in host population density only

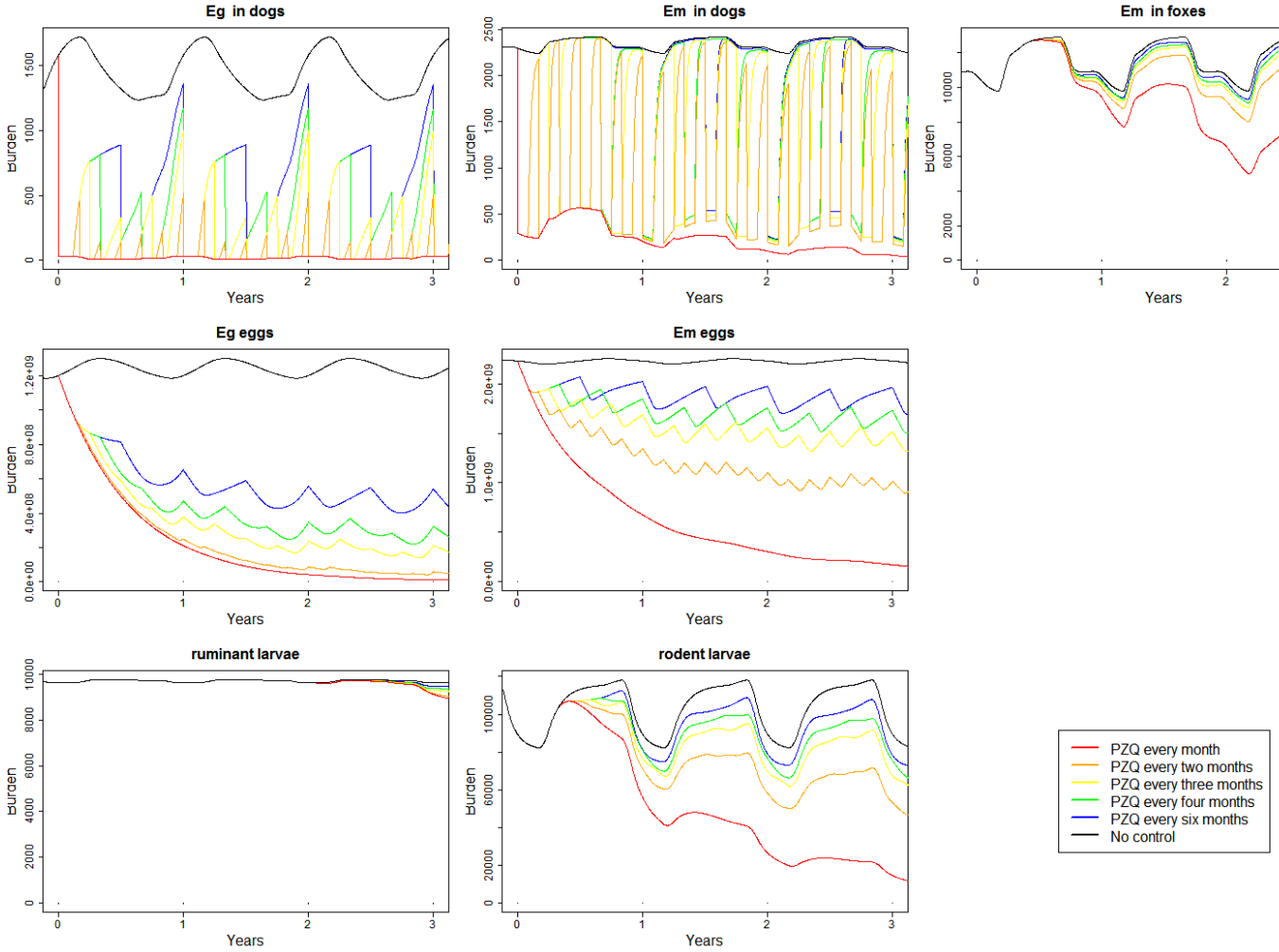


Figure 7.12.Effect of three years of regular praziquantel dosing on the simulated mean *E. granulosus* (left column) and *E. multilocularis* (centre and right columns) burden/contamination with seasonality in host mortality (table 7.3) only

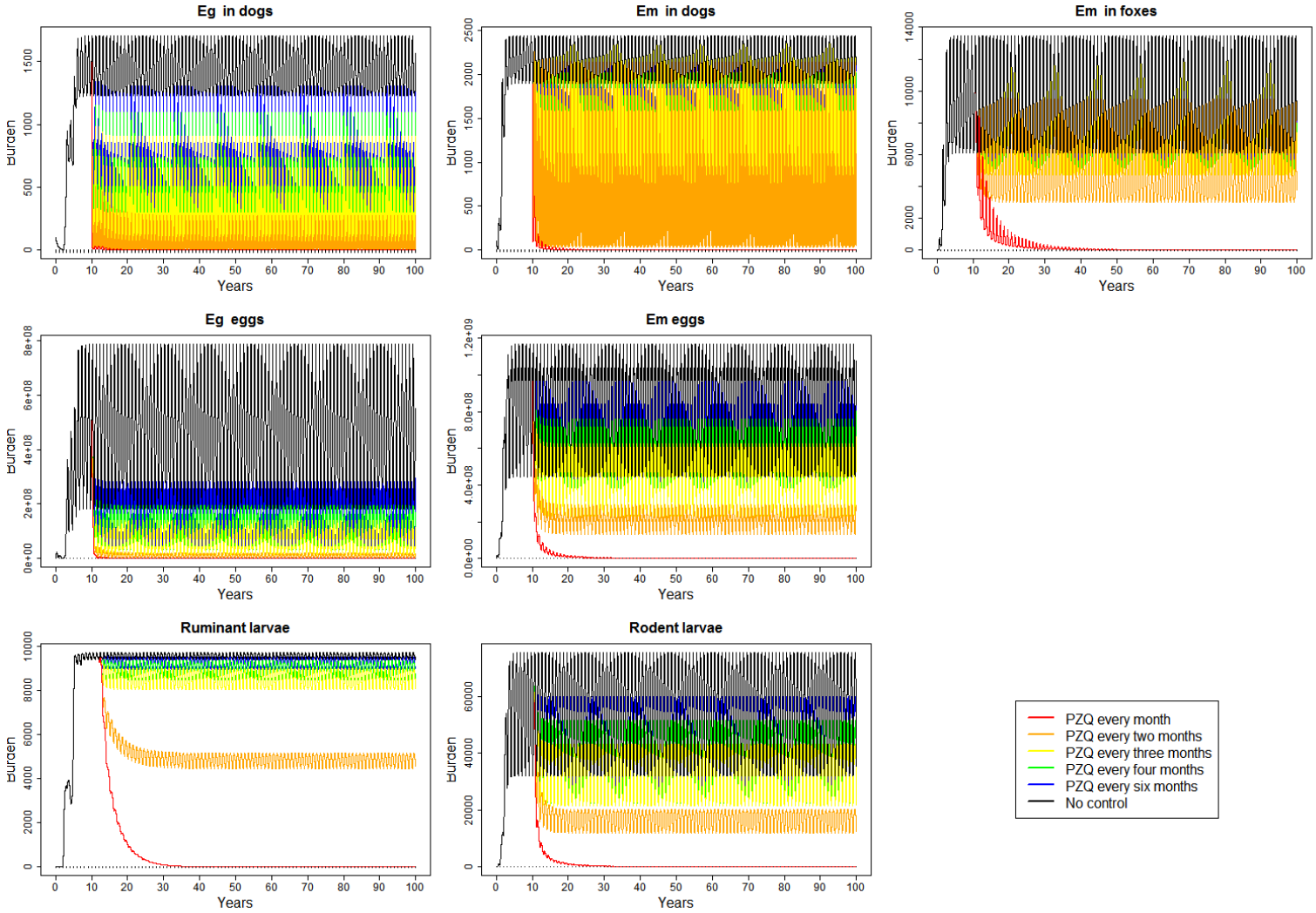


Figure 7.13. Effect of different random dosing frequencies on simulated mean *E. granulosus* (left column) and *E. multilocularis* (centre and right columns) burden/contamination over the course of 100 years

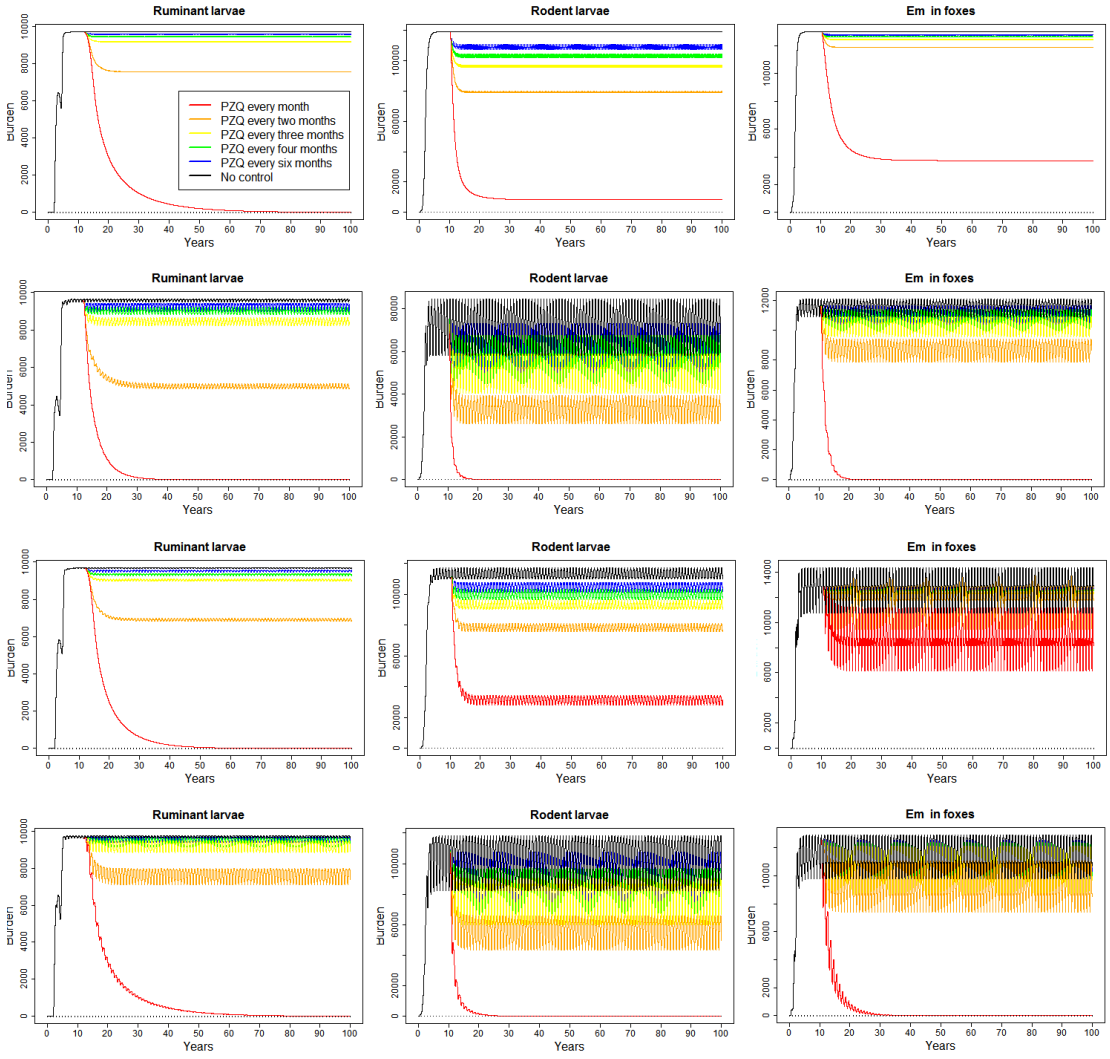


Figure 7.14. Effect of different seasonality assumptions on simulated mean protoscolex burden in ruminants and rodents, and mean adult worm burden in foxes, over the course of 100 years. The top row represents no seasonality; the second row represents seasonality in egg survival only; the third row represents seasonality in host population densities; and the bottom row represents seasonality in mortality-associated parameters (table 7.3).

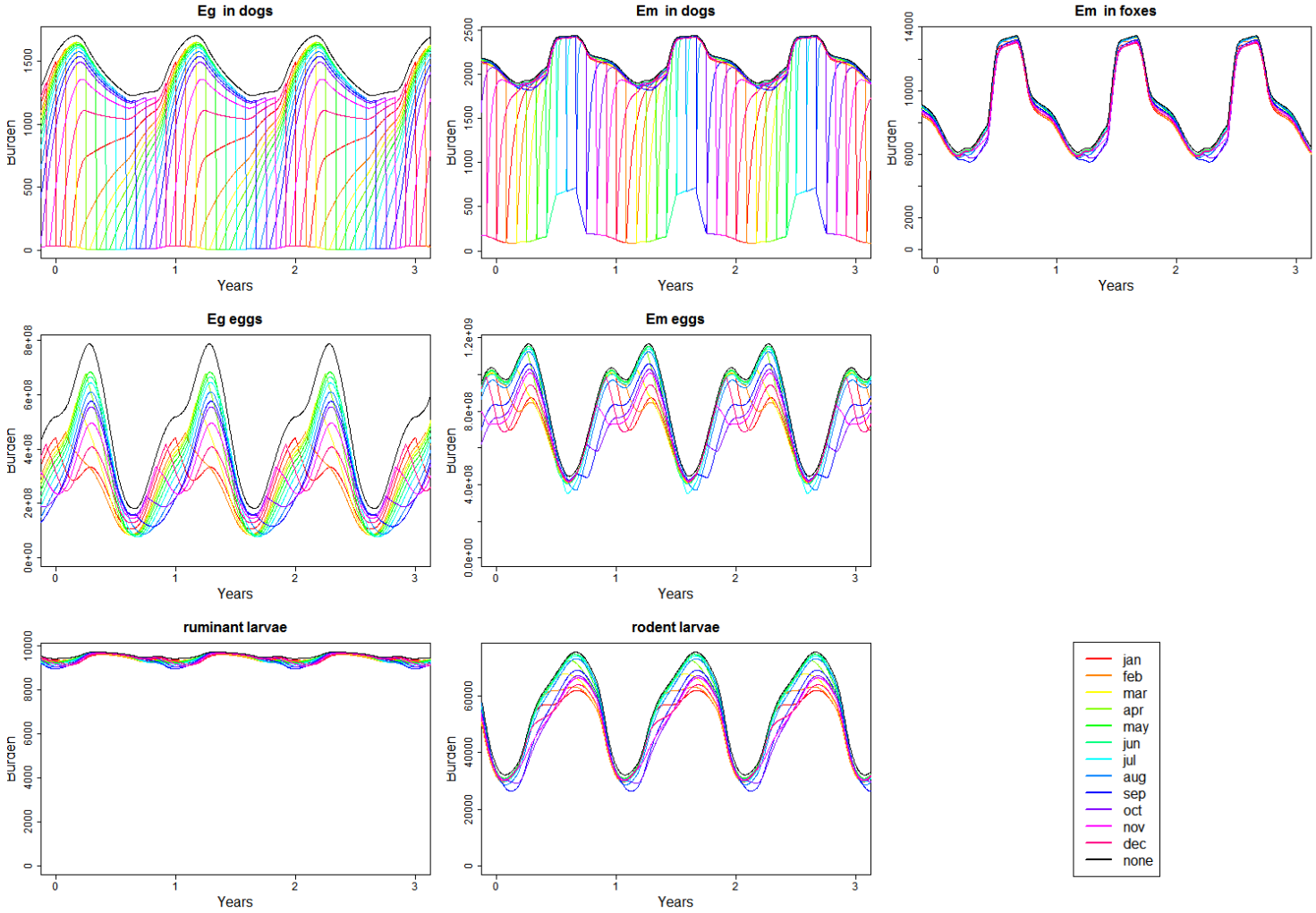


Figure 7.15. Effect of administering a single dose of praziquantel during specified months on the simulated mean *E. granulosus* (left column) and *E. multilocularis* (centre and right columns) burden/contamination over the course of three years, at the new steady state

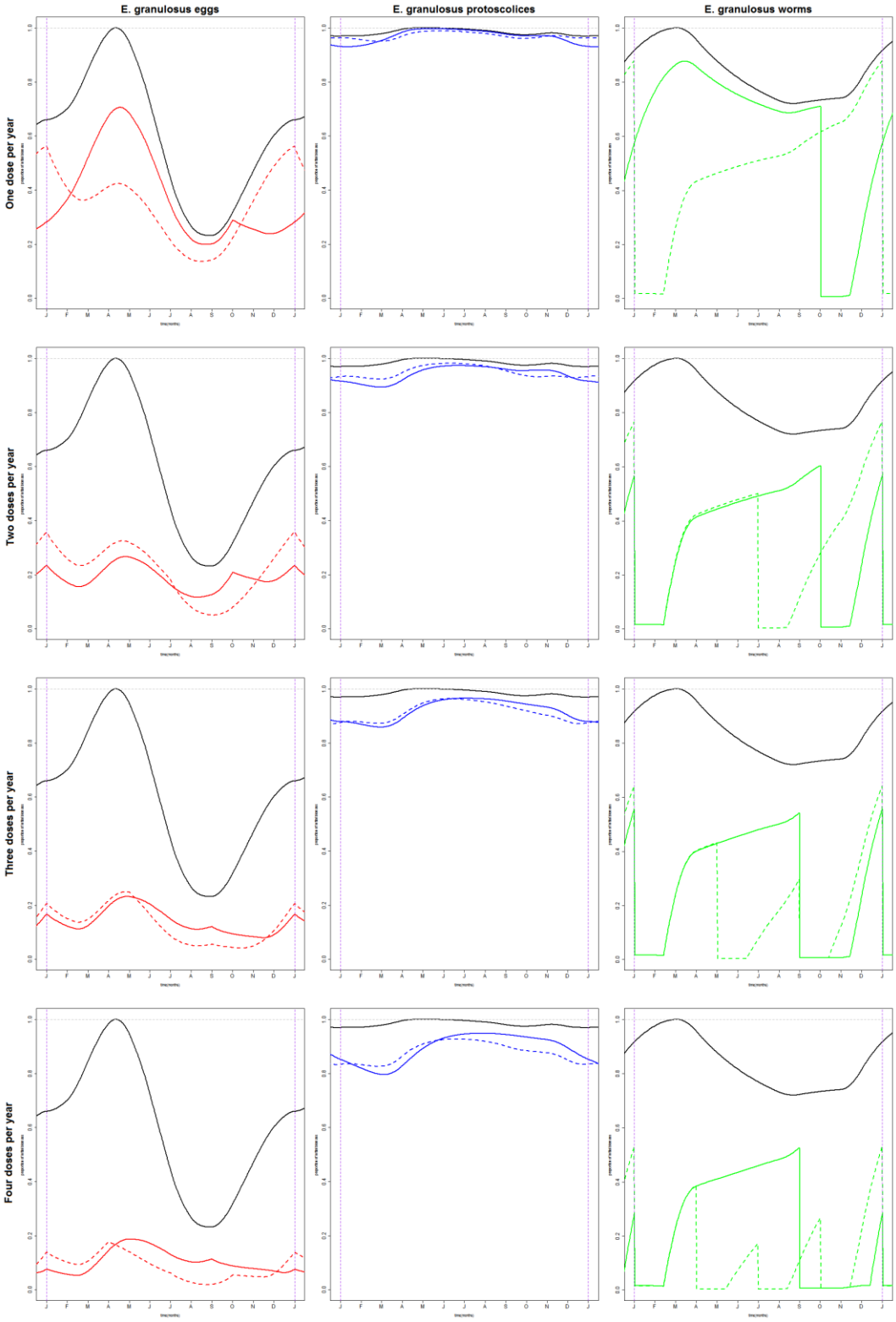


Figure 7.16. Effect of targeted (solid colour lines) and regular (dotted coloured lines) dosing on relative *E. granulosus* biomass over the course of one year at the new steady state. Rows indicate different dosing frequencies and columns indicate different parasite compartments. The black solid lines indicate the parasite dynamics in the absence of control

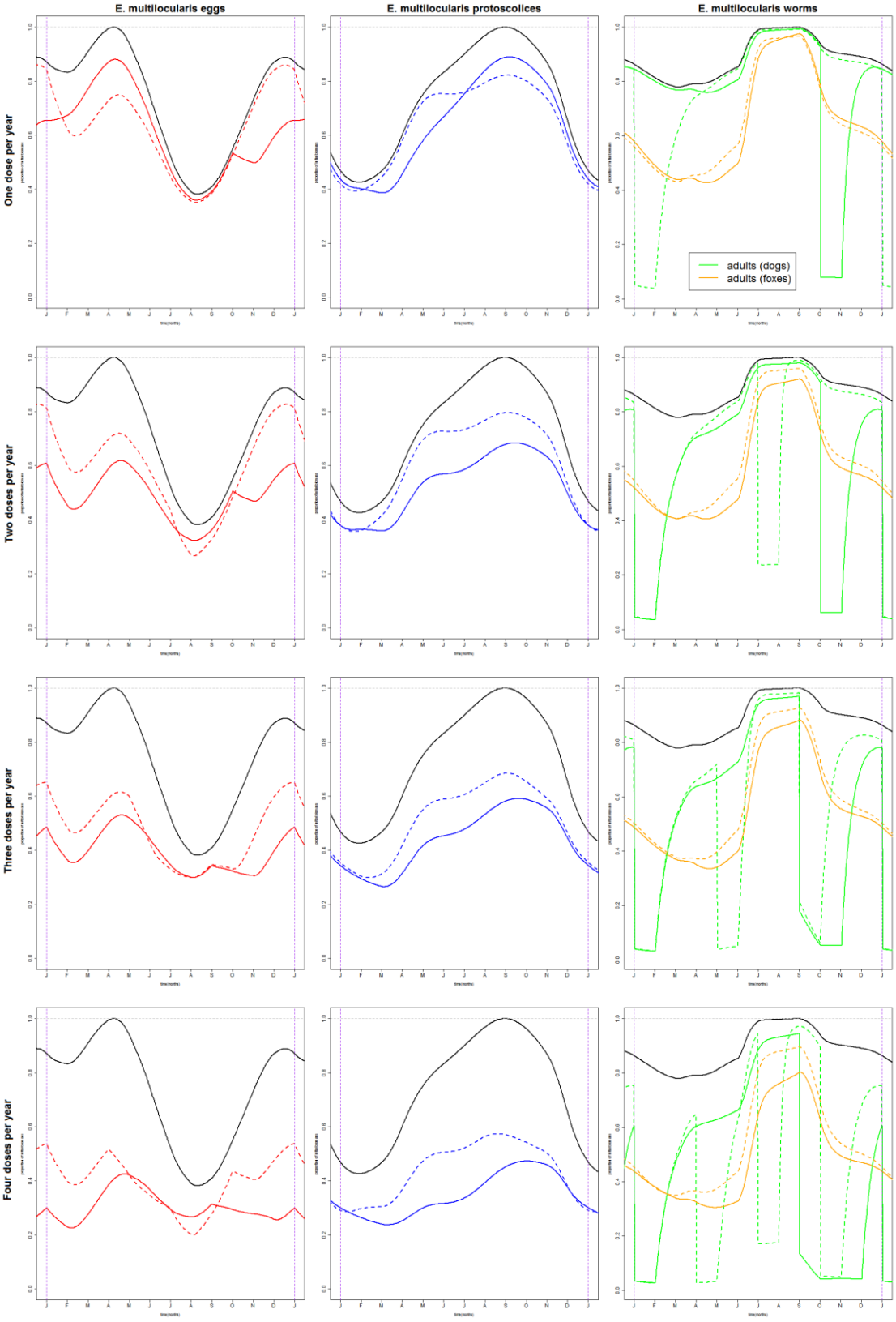


Figure 7.17. Effect of targeted (solid colour lines) and regular (dotted coloured lines) dosing on relative *E. multilocularis* biomass over the course of one year at the new steady state. Rows indicate different dosing frequencies and columns indicate different parasite compartments. The black solid lines indicate the parasite dynamics in the absence of control

7.4 Discussion

This chapter describes (above) and qualitatively evaluates (below) a framework for a deterministic compartmental differential equation model of transmission of *E. granulosus* (sensu lato) and *E. multilocularis* in areas where these two species coexist and are known to infect domestic dogs. To the author's knowledge, a full mathematical model of transmission in these coendemic areas has not been previously developed, despite these areas typically having the highest burden of human echinococcosis (Budke *et al.*, 2004). As should be apparent from the model parameterisation approach used, the current model should be considered to be complementary to mathematical models which attempt to model the force of definitive host infection with *Echinococcus* spp (and which have been developed and applied to coendemic areas) (Torgerson *et al.*, 2003c; Budke *et al.*, 2005b; Ziadinov *et al.*, 2008; Lewis *et al.*, 2014).

The large number of parameters and general lack of available field data for direct parameterisation places considerable constraints on the conclusions which can be drawn from this model as it increases the number of possible errors in parameterisation, and makes it more difficult to identify these when they do occur. However, as described by Ishikawa: "…it is unavoidable for the transmission model for *E. multilocularis* to have a somewhat complicated structure, and consequently to involve many ecological, as well as epidemiological, parameters because *E. multilocularis* has a complicated life cycle involving two kinds of hosts" (Ishikawa *et al.*, 2003). Incorporation of a dog cycle of *E. multilocularis* as well as an *E. granulosus* cycle necessarily increases the required number of parameters further in the current model. As a result of this, most attention in this discussion will be placed on the critical evaluation of the model structure and identification of areas worthy of further study, rather than placing too much focus on developing firm conclusions from the model output. With regards to parameterisation, attention will focus on major points of interest rather than attempting a comprehensive critique of each model parameter.

7.4.1 Model structure evaluation

7.4.1.1 Aggregation

The largest potential criticism of the general model structure relates to its inability to account for aggregation in parasite burdens within a community. The model represents changes in the total parasite biomass in a community rather than focussing on the prevalence of infection within the different host 'compartments', and assumes that this biomass is evenly distributed within the model compartments. As the model is not spatially explicit, it is implicitly assumed that egg contamination is homogenous throughout the study area. This is unlikely to be the case, and a number of studies have found that spatial heterogeneities in *E. multilocularis* contamination are of significance to transmission (Tackmann *et al.*, 1998; Giraudoux *et al.*, 2003; Milner-Gulland *et al.*, 2004; Hansen *et al.*, 2004)

In the case of those compartments which host the obligatory parasitic stages of *Echinococcus* (i.e. intermediate and definitive hosts), the total parasite biomass (protoscolices and adult worms, respectively) is estimated, and this is then divided amongst the available hosts. Whilst this gives an estimate of the "mean burden" per host, the concept of burdens within individual hosts is largely an artificial construct in this model, used primarily for ease of communication of the model predictions (and for incorporation of density dependence – see below). As alluded to earlier, if this were to be interpreted directly, it would assume that all individuals in the population are infected and all hold the same burden (that is, infection would be 'underdispersed' within the hosts (Anderson and Gordon, 1982)). However, it is well reported that the total parasite biomass is not divided equally between the different hosts. Instead, parasites are aggregated within hosts (and in the environment, as described above). This "overdispersion" is considered a key feature of macroparasitic infection (Crofton, 1971a), and is thought to have a particular role in parasite regulation and stability (Anderson and May, 1978; May and Anderson, 1978; Adler and Kretzschmar, 1992; Medley, 1992; Kretzschmar and Adler, 1993; Rosà and Pugliese, 2002; Rosà *et al.*, 2003).

Therefore, failure to account for it in the model could have considerable repercussions for the accuracy of model predictions – especially in the face of a control scheme.

Aggregation can be incorporated in deterministic models such as the current one by assuming that the individual burdens are distributed according to the negative binomial distribution (Anderson and May, 1978; May and Anderson, 1978). Although a number of studies have identified the negative binomial distribution to be an excellent predictor of the distribution of parasite burdens in a variety of settings (Crofton, 1971a; Anderson and May, 1978, 1985; May and Anderson, 1978; Guyatt *et al.*, 1990; Shaw and Dobson, 1995), it does not explicitly model the underlying generating process for this aggregation. As macroparasites generally do not multiply within the host, the generating process for aggregation can be broadly considered to result from heterogeneities in host acquisition (and/or loss) of parasites (Anderson and Gordon, 1982; Quinnell *et al.*, 1995). If it can be assumed that the rate of parasite acquisition, rather than being a fixed value for all hosts in a population, varied according to a Poisson distribution, then natural acquisition of parasites (which itself would be expected to be a Poisson process) would result in the overall distribution of parasites in the population being overdispersed (Anderson and Gordon, 1982). If the distribution of rate of acquisition followed a gamma distribution, then the resultant distribution of burdens would be negative binomial (Bundy and Medley, 1992; Medley *et al.*, 1993). However, there remains no biological support for this theory beyond the empirical observation that the negative binomial is generally a good fit to burden data.

A number of studies have attempted to further identify the exact generating processes underlying parasite aggregation (McCallum and Anderson, 1984; Anderson and May, 1985; Pacala and Dobson, 1988; Quinnell *et al.*, 1995; Grenfell *et al.*, 1995; Shaw *et al.*, 1998; Galvani, 2003; Churcher *et al.*, 2005). Heterogeneity in exposure can be considered to result from variation in host predisposition to infection (McCallum and Anderson, 1984), or from heterogeneities in exposure resulting from environmental, spatial and seasonal factors. Further identification of the exact mechanisms underlying these processes is an area of ongoing research – for example, with regards to the genetic basis for variation in susceptibility (Quinnell, 2003), the effect of

immunological processes on infection and parasite fecundity (Galvani, 2003), the effect of 'clumping' of transmissible stages in space or in time (Quinnell *et al.*, 1995; Heinzmann *et al.*, 2009, 2011b), spatial factors (Hansen *et al.*, 2003, 2004; Calabrese *et al.*, 2011), and the effect of increased susceptibility to predation in transmission systems based upon predator-prey relationships (Vervaeke *et al.*, 2006). Although the generating process responsible for aggregation will not be discussed any further in any detail here, it is of particular importance when the effect of control schemes is modelled since it is likely to have implications on parasite stability. It has been suggested that heterogeneity resulting from host effects (such as variability in susceptibility) would result in increased parasite stability compared to that resulting from clustering of infection (Rosà and Pugliese, 2002). Due to the potential impact of aggregation (and of density dependence and infection intensity, which are closely associated with aggregation (Churcher *et al.*, 2005)), incorporation of these processes is one of the main suggestions for future model development. This is likely to require the use of stochastic methods such as agent-based modelling (Hansen *et al.*, 2003; Heinzmann *et al.*, 2011b; Huang *et al.*, 2011), and will likely also require the incorporation of spatial modelling techniques due to the likely role of spatial heterogeneities in maintaining *Echinococcus* transmission in the face of a control scheme (Hansen *et al.*, 2003, 2004). These approaches will require higher quality data than that available currently. Since a major aim of the current model is the evaluation of possible control and surveillance strategies, high-quality longitudinal data collected during a control scheme would be of particular use (Churcher *et al.*, 2006; Basáñez *et al.*, 2012a; b).

7.4.1.2 Other features

In order to address other issues relating to the model structure, a number of key elements of *Echinococcus* models, identified in a recent review (Atkinson *et al.*, 2013), will be investigated. These are:

- Latency
- Age structure

- Density dependence
- Heterogeneities in risk of transmission
- Seasonality

The current model incorporates latency in the form of the lag period between infection and infectiousness for adult worms and protoscolices. This is particularly important for the modelling of control schemes focussed on the definitive host, as infection can persist in intermediate hosts even when infection in completely removed from the definitive host (Takumi and van der Giessen, 2005), and was therefore considered of considerable importance to the current model.

A crude attempt was also made in the current model to incorporate the effect of intermediate host age on the removal of protoscolex biomass through rodent predation and rodent and ruminant death (with ruminant death and rodent predation also resulting in definitive host infection), by incorporating κ parameters. It has been well reported that age is a major risk factor for intermediate host infection with *Echinococcus* spp, due to the lifelong nature of infection (reviewed in Otero-Abad and Torgerson, 2013). Age is also a major risk factor for the risk of death, with high mortality commonly observed in young and old animals. As a result, an association between infection status and risk of mortality (albeit indirect) would be expected. In the case of intermediate hosts, infected animals would be expected to be overrepresented amongst removed animals (although removal through predation may not be a linear process in the case of small rodents, due to behavioural differences between older and younger animals. This was considered to be of particular importance for ruminants, since older animals would be more likely to be slaughtered for human consumption (and possibly more likely to die, although the effects of harsh winter conditions may also affect young animals).

As well as this 'indirect' relationship between infection status and predation/ingestion, the increased predation of infected intermediate hosts may also be a natural feature of *Echinococcus* transmission. For example, a study of the degree of aggregation of *E. granulosus* cysts in moose in Canada has found that in areas of high predation

pressure, aggregation is lower – suggesting preferential removal of those animals with large numbers of cysts (Joly and Messier, 2004). A study of *Ellobius* rodents in Kyrgyzstan also suggested that infected individuals could be more easily caught (Afonso *et al.*, 2015). However, direct evidence of an increased susceptibility to predation of *Echinococcus*-infected intermediate hosts is currently lacking. A fuller discussion of this issue, with a focus on *E. multilocularis*, can be found in a recent modelling paper (Vervaeke *et al.*, 2006). Further work to better parameterise the κ parameters in the current model would be advised.

As mentioned above, density dependence is an important characteristic of macroparasite epidemiology, and is likely associated with aggregation as well as other processes (Churcher *et al.*, 2005, 2006). Density dependence was incorporated in the current model by assuming a logistic growth function for adult worms and protoscolices within their hosts. This was largely a phenomenological construct, and was not based upon any particular biological understanding of echinococcosis. This formulation also implies that the processes responsible for density dependence operate equally on all individuals in the population, which is unlikely to be the case.

Possible causes of density dependence are crowding, host immunological processes (Torgerson, 2006b; Zhang *et al.*, 2008), and competition between individual parasites. Constraints on the mean protoscolex or adult worm burden per host may also result from finite host lifespan (particularly in the case of intermediate host infection), due to the time required for cysts and protoscolices to develop. However, a mechanistic construct for including density dependence may be more useful in further work. In particular, immunity in the definitive host could have considerable repercussions for the efficacy of a control program, as control may be more difficult in the presence of immunity. There is some evidence of age-related immunity in the case of *E. granulosus* infection of dogs (Lahmar *et al.*, 2001; Torgerson *et al.*, 2003c; Budke *et al.*, 2005b; Moro *et al.*, 2005; Buishi *et al.*, 2005b), and *E. multilocularis* infection of foxes (Hofer *et al.*, 2000; Yimam *et al.*, 2002; Ziadinov *et al.*, 2010), but not *E. multilocularis* infection of dogs (Budke *et al.*, 2005b). This would be suggestive of a greater efficacy of dosing dogs with praziquantel on *E. multilocularis* than *E. granulosus* (Torgerson, 2006b),

although this may also be complicated by the presence of a sylvatic reservoir host in the case of the latter and not the former. As infection with *E. granulosus* and *E. multilocularis* was modelled separately, the constraints upon maximal infection intensity were modelled separately for each species. *E. granulosus* and *E. multilocularis* have been shown to inhabit different regions of the small intestine (Thompson and Eckert, 1983; Gemmell *et al.*, 1986c; Lymbery *et al.*, 1989; Morishima *et al.*, 1999a; Umhang *et al.*, 2011), with a clear demarcation in the case of coinfection, and it was therefore considered biologically reasonable to model the maximal burdens separately. However, there remains a possibility that interaction between the two species of interest here (or even between similar species of *E. granulosus* sensu lato) would impact upon transmission dynamics. Explicit investigation of coinfections in coendemic areas such as this would be an interesting avenue for further work, which could potentially inform further model development and refinement.

Non-temporal heterogeneities in transmission risk are also closely linked to the issue of parasite aggregation, as mentioned above, and as such can lead to spatial and temporal patterns of infection/aggregation (which may or may not be predictable) (Hansen *et al.*, 2004). As described above, these factors were generally not accounted for in the current model, which largely assumes homogenous mixing of hosts in the absence of any parasite aggregation. Identifying and extracting the variety of processes resulting in parasite aggregation will generally require the use of stochastic and spatial modelling approaches (Rosà *et al.*, 2003), which, as described above, would likely require the collection of higher quality data. Although collection of this sort of data is likely to be challenging in the poorer, more marginalised communities where *E. granulosus* and *E. multilocularis* are coendemic in domestic dogs, it is hoped that by continued development and refinement of the current model framework, data collection may be guided and targeted to those areas most relevant to investigation of these processes.

7.4.1.3 Seasonality

Seasonality is an important component of the current model, and was incorporated in a number of important parameters: egg survival on pasture, host population densities, μ parameters and availability and access to intermediate hosts. It has been argued that the inclusion of seasonal variables in mathematical models of *Echinococcus multilocularis* is not necessary, since they do not affect the general model conclusions (Roberts and Aubert, 1995). However, as demonstrated in the previous chapter and by reinfection studies on the eastern Tibetan plateau (Wang, 2011), seasonal trends in canine echinococcosis may exist, and indeed would be expected in the presence of varying levels of exposure to intermediate hosts. Previous models have incorporated seasonality in parasite and host survival/density (Ishikawa *et al.*, 2003; Ishikawa, 2006; Nishina and Ishikawa, 2008), and this has been suggested to be an area worthy of further exploration in other studies (Budke *et al.*, 2005b; Takumi and van der Giessen, 2005; Atkinson *et al.*, 2013). The incorporation of seasonality in the current model had an impact on certain important model outputs – including the ability of foxes to maintain infection in the absence of domestic dog infection (which was only predicted by the model in the absence of seasonality and is discussed in further detail below).

Seasonality in egg survival has been investigated in models of echinococcosis in Hokkaido, Japan, using a similar approach to that described here (Ishikawa *et al.*, 2003; Nishina and Ishikawa, 2008). Egg survival was also modelled seasonally in an agent-based model of *E. granulosus* parameterised using data from Kazakhstan (Heinzmann *et al.*, 2011b). This found little effect of seasonality in egg survival on the transmission process, due to the 'buffering' effect of intermediate host infection. Conversely, the original Takumi model suggested a considerable effect of egg survival on the time required to remove 50% of the parasite biomass during a control program (Takumi and van der Giessen, 2005). An agent-based model of reinfection of foxes with *E. multilocularis* following control suggested that spatial heterogeneity in egg survival rather than variation in egg survival *per se* had better support from field data collected during and after a control scheme (Hansen *et al.*, 2003). This spatial

heterogeneity in egg survival, ultimately leading to heterogeneity in fox infection, resulted in a suggestion of spatially targeted treatment of foxes – similar conceptually to the idea of temporal targeting explored in the current study. It also demonstrates the potential importance of considering spatial heterogeneities – which will be covered later.

One potential limitation in the current strategy of modelling egg survival based upon mean temperature was the failure to account for heterogeneities in temperature (for example, maximum temperatures during the summer months could have a greater impact on egg survival than the average temperature). As the effect of temperature on egg survival is also known to vary depending on whether eggs are suspended in water (Veit *et al.*, 1995; Matsumoto and Yagi, 2008; Federer *et al.*, 2015), relative humidity or spatial variation in ground moisture content could also have important implications for egg survival, which was not accounted for in the current model. A spatio-temporal study of risk factors for infection of water voles (*Arvicola terrestris*) with *E. multilocularis* in Switzerland found that the best models included mean temperature and mean precipitation, suggesting that these factors are of importance to infection risk (although these findings could also be due to confounding by other seasonal factors) (Burlet *et al.*, 2011). Similarly, a study in Germany found that infected foxes were more commonly found in areas of higher soil moisture content – which may indicate higher prevalences amongst intermediate hosts in these areas (Staubach *et al.*, 2001).

Seasonality in host population densities and mortality was largely based upon qualitative adjustments to 'average' estimates, partly because accurate data is just not available for some hosts, and partly because the aim of the model was not to perfectly predict the effect of these densities on the parasite burdens, but to identify general patterns and trends. The effect of an increase in host population density would be expected to initially reduce the mean burden through a 'dilution' effect. This is a considerable oversimplification of the situation, especially in the case of long-lived infections such as is seen in the intermediate host and where the true burden in any individual host at any one time would not be expected to be directly affected by the

total number of hosts. Most of the changes in population density resulted from changes in mortality, birth of new hosts, and movement of hosts out of the study area (dispersal in the case of sylvatic hosts, and seasonal movement to Jailoo in the case of domestic dogs). Again, the Japanese models investigated the effects of seasonality in host densities on the prevalence of host infection and the potential risk of human infection (due to egg contamination), and suggested a complex interplay between these factors (for example, changes in rodent population densities were considered to be responsible for changes in estimated intermediate host prevalence, but the estimated peak prevalence of fox infection did not occur at the time of the peak population density) (Ishikawa *et al.*, 2003; Nishina and Ishikawa, 2008).

Changes in host 'preference' for rodents (associated with availability of rodents) was also incorporated into the Japanese models by explicitly modelling a 'feeding function', which was incorporated the rodent population density and the depth of the snow (which would be expected to reduce access to rodents) (Ishikawa *et al.*, 2003; Nishina and Ishikawa, 2008). A much more crude parametrisation was used in the current model, and was based upon a relatively subjective reduction of the rate of rodent ingestion during the winter months, when access to rodents would be expected to be reduced. The effect of changes in rodent availability and rate of ingestion by foxes has been identified as a potentially important factor in the dynamics of *E. multilocularis* transmission (Hegglin *et al.*, 2007; Raoul *et al.*, 2010), and would be worthy of a more comprehensive evaluation in the study area – in particular, in relation to foxes living adjacent to areas of human habitation (the area of focus in the current study).

7.4.2 Model applicability to study area

7.4.2.1 Spatial context of model

Although not explicitly a spatial model, the spatial context of the model is of relevance to the interpretability of the model. For example, the transmission of *E. multilocularis* is thought to vary between rural, urban, and periurban settings (Stieger *et al.*, 2002; Robardet *et al.*, 2008), and has been found to vary between different provinces of

Hokkaido, Japan (Ishikawa *et al.*, 2003; Nishina and Ishikawa, 2008). Seasonality in reinfection may also vary between different Tibetan communities in China (Wang, 2011). As well as differences in the dynamics between different locations, one particular challenge faced when constructing a spatial model is deciding on which spatial scale the model should be developed. Intermediate hosts of *E. multilocularis* often have a smaller home range than definitive hosts, and a balance must therefore be reached between representing suitable numbers of each host type in order to capture complete transmission cycles, whilst not overgeneralising the transmission process (Morgan *et al.*, 2004). As domestic dogs are the host of primary interest in the current study, a decision was therefore made to focus primarily on the village setting, but to account for free roaming of foxes in the vicinity (which could have a role in sustaining *E. multilocularis* transmission cycles). Whilst it is unlikely that any foxes permanently inhabit the 1km^2 shown in figure 7.1, it is plausible that the home ranges of some foxes intersect this area (especially considering the potential availability of anthropogenic food). Despite these attempts to reduce the limitations associated with spatial constraints, it is likely that by setting any boundary on the area of interest, a lot of potentially relevant/vital processes are excluded.

This problem feeds naturally into the concept of metapopulations (Grenfell and Harwood, 1997; Hanski, 1998) – in particular, parasite stability in a metapopulation context (Giraudoux *et al.*, 2006). A metapopulation model of *E. multilocularis* in Kazakhstan has previously been developed (Milner-Gulland *et al.*, 2004), which shares a number of conceptual characteristics with the current model (i.e. modelling at the parasite, rather than the host, level). Expansion of the current model formulation into a metapopulation model would be a potentially useful area of further study which could reconcile some of the issues associated with setting strict geographical constraints on the model. However, this would require the collection of additional data (in particular, the locations and movements of potential sylvatic hosts, the movements of dogs and livestock around the village and between village and Jailoo, and the locations of rodent intermediate hosts). Whilst this approach would have some relevance for the study of the effects of a control scheme (in particular, by

incorporating more explicitly the role of foxes in the *E. multilocularis* transmission cycle amongst dogs), its main use would be to elucidate the transmission ecologies of the two species of *Echinococcus* of interest. From a logistical perspective, the Alay valley offers great potential for this sort of investigation – being a focus of *E. granulosus* and *E. multilocularis* endemicity which is relatively geographically and 'politically' accessible.

7.4.2.2 Host ecology

Although both the red fox (*Vulpes vulpes*) and the Corsac fox (*Vupes corsac*) have been found to be infected with *E. multilocularis* in neighbouring Kazakhstan (Shaikenov, 2006), the Corsac fox is not thought to be present in the Alay valley (Murdoch, 2014), and so was not considered in the current model. Estimating the density of red foxes was a challenge, since estimates of population densities of red foxes vary widely (MacDonald and Reynolds, 2008), and little information is available for Kyrgyzstan. It is also likely that the fox home range will vary seasonally, along with the availability of food, which would be expected to impact upon the effective fox density in the study area. In particular, it has been suggested that foxes may move into areas of human habitation during the autumn-winter period in search of food (Heptner and Naumov, 1992). This, along with the effects of breeding, is another potential source of seasonal variation in fox density within the village surroundings. The original Takumi model, based in the southern Netherlands, used expert opinion to estimate a population density of around four foxes per km^2 (Takumi and van der Giessen, 2005). Due to the low human population densities in the Alay valley, the fox density in this area was considered likely to be lower than this.

A similar challenge was faced when attempting to estimate intermediate host population densities. A large number of rodents are known to be susceptible to infection with *E. multilocularis*, and as such it is unrealistic to attempt to account for every potential host species in the model. Suggested hosts in the Alay valley are the cricetid rodents *Microtus gregalis* (narrow-headed vole), *Cricetulus migratorius* (grey dwarf hamster), *Ellobius tancrei* (eastern mole vole – possibly confused with the

northern mole vole, *Ellobius talpinus*, in some cases), *Microtus oeconomus* (root vole), *M. obscurus* (Altai vole), *M. carruthers* (archer vole), *Altikola argentalus* (silver mountain vole); the murid rodents *Apodemus sylvaticus* (wood mouse), *Mus musculus* (house mouse) and *Meriones erythrourus (Libyan jird)*; the glirid rodent *Dryromys nitedula* (forest dormouse); and the sciurid rodent *Marmota caudata* (red marmot) (Abdyjaparov and Kuttubaev, 2004; Giraudoux *et al.*, 2013b; Afonso *et al.*, 2015). Of these, *E. multilocularis* infection has been documented in *Marmota caudata* (11/256), *Microtus carruthers* (1/345), *Apodemus sylvaticus* (1/437) (Abdyjaparov and Kuttubaev, 2004), and *Ellobius tancrei* (1/42) (Afonso *et al.*, 2015) in the Alay valley. It should be noted that in the earlier study (Abdyjaparov and Kuttubaev, 2004), no *Ellobius* spp were caught, and these were therefore not evaluated. Whilst the highest prevalence was found in *Marmota caudate*, sciurid rodents are not known to be major hosts of *E. multilocularis* in any of the studied transmission ecosystems of *E. multilocularis* worldwide. Instead, it has been suggested that *E. tancrei* be considered a "flagship" intermediate host in the southern Kyrgyz transmission ecosystem (Giraudoux *et al.*, 2013b). Although this does not indicate that this species necessarily plays the largest role in the transmission cycle, it may be the most useful measure of the potential for transmission within the area (and as such could be a useful species for model validation). This species was therefore considered the intermediate host of main interest in the model. Little information is available on the population densities of these rodents, and as such any estimate of abundance will be very limited (Patrick Giraudoux, personal communication), but they were known to be present within and around the study villages, with increasing abundance as grassland biomass increased (Giraudoux *et al.*, 2013b; Afonso *et al.*, 2015). Another challenge with any estimate of population densities of these rodents is considerable interannual variation in population densities (as is also seen with *Ellobius talpinus* (Evdokimov, 2013)) and their tendency to undergo population 'outbreaks'. This tendency towards population outbreaks has been suggested to be a common feature of many *E. multilocularis* transmission foci (Giraudoux *et al.*, 2013b). Whilst this could play an important part in transmission, accounting for this in a mathematical model would be very complex; likely requiring the use of stochastic techniques.

7.4.2.3 Human culture and practices

A number of characteristics of human behaviour are of importance in the current model. Transhumance is still practiced to some degree in the Alay valley, meaning that many dogs were moved from the village to Jailoo in the summer months. As this was considered to be of importance to potential levels of egg contamination in this particular setting, this was also included in the model. Dogs were also almost invariably unrestrained within the communities, but were generally assumed to remain within the village surroundings during this time. This was substantiated by GPS monitoring of dog movements within the villages (van Kesteren *et al.*, 2013). However, this study also identified that some dogs travelled far from the village. Although this was relatively uncommon, it adds stochasticity to the process and could be of importance for the infection risk with *E. multilocularis*, and would be worthy of further investigation if the model was developed into a stochastic form as described above.

Although some data was available regarding ruminant ownership in Taldu-Suu, there remained difficulties in estimating the density of ruminants in the study area due to daily and seasonal movements of livestock to grazing land around the village and in the mountains (Jailoo), and the fact that some animals (generally when livestock ownership was very large) were kept at Jailoo permanently. A total of 2,000 sheep and 1,500 goats were reported to be owned by the residents of Taldu-Suu in May 2012. Goats have repeatedly been found to have lower prevalences of infection with *E. granulosus* G1 than sheep (Torgerson *et al.*, 1998; Cardona and Carmena, 2013) – which may be suggestive of differences in rates of ingestion due to feeding patterns, or host specificity issues. Goats have also repeatedly been identified as a host of *E. canadensis* (previously *E. granulosus* G6/7) (Soriano *et al.*, 2010; Romig *et al.*, 2011; Cardona and Carmena, 2013), which is known to occur in the Alay valley (van Kesteren *et al.*, 2013). Therefore, the selection of which livestock are responsible for transmission should take into account the possibility of two (likely overlapping) cycles: *E. granulosus* G1 between sheep and dogs, and *E. canadensis* G6 between goats and dogs. A decision was made in the current study to model these intermediate hosts

together, giving a total of 3,500 owned "small ruminants" throughout the village. Despite this high number, most animals are moved out of the village vicinity during the day (partly due to low quality pasture around the village resulting from previous overgrazing), and so cannot be considered to be either exposed continually to eggs on the pasture within the area of interest or a continual potential source of canine infection within the study area through scavenging. Additionally, most livestock owners travel to Jailoo during the summer months – further reducing the number of small ruminants within the village. For these reasons, the total number of small ruminants present in the 1km² area indicated in figure 7.1 during questionnaire surveys in spring and autumn was broadly estimated to be no more than around 500 animals, on average, throughout the year. Again, slight fluctuations in this on a seasonal and a daily basis (as animals return from grazing each evening during the spring and autumn months) could be of importance to *Echinococcus* transmission dynamics, and would be worthy of further study.

7.4.3 Force of infection

The β parameters in the current model describe the rate of ingestion of intermediate hosts by definitive hosts and the rate of ingestion of eggs by intermediate hosts, and therefore are central to the whole model. However, these parameters are also less tangible than many of the other parameters, since they relate to behavioural practices which are unlikely to be directly quantifiable. As a result, these parameters were estimated using data from studies conducted in Naryn province (with estimates for foxes extrapolated from a study in Switzerland). Although the Alay valley may differ epidemiologically from Naryn (as described above), animal ownership and management was similar in the two areas (with the exception of the presence of an abattoir system in Naryn).

The Takumi model estimated the rate of ingestion of *Arvicola* spp (the *E. multilocularis* intermediate host of main interest in the Netherlands) by foxes using data from a study which found an average of 22g of rodent material in the stomach of Dutch foxes, which was then converted into a rate of rodent ingestion. From this

estimate, the proportion of rodents which were *Arvicola* spp and the predicted population density of these species was incorporated to give a rate estimate of 2×10^{-7} (Takumi and van der Giessen, 2005). The estimate used in the current model is 30 times higher than this estimate - which, if correct would suggest that Kyrgyz foxes have an increased reliance on rodents than Dutch foxes. Foxes are generalist, opportunistic feeders (Calisti *et al.*, 1990), and therefore would be expected to adjust their feeding habits according to the availability of food. This may represent seasonal changes in rodent abundance (Ferrari and Weber, 1995), or the availability of anthropogenic food (Harris, 1981). This plasticity of predation behaviour has been considered to be of considerable importance to the transmission of *E. multilocularis* (Hegglin *et al.*, 2007; Robardet *et al.*, 2008; Raoul *et al.*, 2010), and it is therefore very unlikely to be accurately captured in the current parameter. However, given the lower human population densities in the Alay valley (which would be expected to result in less anthropogenic food for foxes), it is perhaps plausible that an increased reliance on rodents is seen in this area. Studies of foxes in the former USSR have suggested that rodents comprise the majority of the fox diet, although this would be expected to vary with rodent density and fox age. Although little information was available for mountain ecosystems, it was suggested that in the steppe, foxes feed almost exclusively on small rodents (especially *Microtus* spp) (Heptner and Naumov, 1992). However, a study in Mongolia suggested that insects comprised the majority of the diet of red foxes, with rodent remains only comprising 20% of scat volume (Murdoch *et al.*, 2010). Eating of insects in the Alay valley appeared to be common amongst dogs, based upon visual inspection of dog faeces in the area during the study, and therefore may be the case for foxes as well. Ongoing work by collaborators at Université de Franche-Comté will hopefully shed some light on the diet of foxes and dogs in the area.

Interestingly, the estimated rate of ingestion of rodents by dogs was estimated to be slightly higher than that of foxes (due to the method used for estimation of this parameter incorporating the probability of protoscolices developing into adult worms, which is thought to be lower for dogs than for foxes). Without further examination of

dog and fox faecal samples (which, as mentioned above, is currently being undertaken), it is not possible to further validate this. However, it should be noted that other studies of experimental infection of dogs have suggested higher proportions of protoscolices developing into worms than that used here (Matsumoto and Yagi, 2008), which would have repercussions for this parameter estimate. Due to the importance of this parameter on the transmission of *E. multilocularis*, sensitivity analysis of this parameter would be interesting, and will be conducted at a later date.

The rate of ingestion of ruminants was more difficult to parameterise due to the distinction made in the model between scavenging behaviour and direct feeding of offal at slaughter. This distinction was considered important, due to the potential control available over the two different routes (it would be easier to reduce feeding of offal at slaughter than prevent scavenging of carcasses). In order to estimate the rate of scavenging, an estimate of the rate of feeding of offal to dogs needed to be made. This would be expected to depend upon whether or not the offal was of use to humans. As livers were eaten by humans in the study communities, healthy livers would be unlikely to be thrown to dogs during slaughter. However, infected offal would be expected to be more commonly offered to dogs. When asked about what would be done with diseased offal in May 2012, 40% (275/692) of people reported that they would give it to dogs. However, given that only 55% of people reported ever finding diseased offal, the proportion of offal fed to dogs in the communities would be expected to be low (although the rate of feeding lungs to dogs may be higher, since these appeared to be less commonly eaten by humans). In the absence of further information, it was assumed that around of 5% of liver/lungs were offered to dogs. Based upon this, the probability of a dog scavenging a ruminant carcass was also estimated as around 5%. This appears low for the communities in question – indeed, it is likely that any animal dying in the vicinity of the village would be quickly scavenged by dogs (personal observation). This apparent error in parameterisation appears to result from an incompatibility of the estimated rate of canine infection (0.3 infections per year) with the other associated parameters (proportion of protoscolices resulting in infection following ingestion etc…), which would predict a higher rate of infection.

This problem is likely partly due to the current model's inability to model aggregation of infection within intermediate hosts, but further investigation of the levels of ruminant infection with *E. granulosus* (s.l.) in the study area would also be very useful to further clarify some of these issues (see below).

In order to estimate the rate of ingestion of eggs on pasture by intermediate hosts, an adjustment was made to the equation used by Takumi and van der Giessen (2005) in order to model at the protoscolex level rather than the infection level. Although a central feature of macroparasitic transmission is the inability to reproduce inside the host, the production of protoscolices by *Echinococcus* in the intermediate host is an exception to this general rule, as a single ingested egg can produce a number of protoscolices. If it is assumed that the rate of protoscolex development would be equal to the rate of protoscolex removal at the steady state, then the number of protoscolices which are produced per ingested egg should be taken into consideration. Estimation of the number of protoscolices which develop per egg was challenging due to the extreme overdispersion seen in this parameter. This was therefore estimated in different ways for ruminants and rodents. For ruminants, it was assumed that each ingested egg would result in one cyst (albeit with a variable number of protoscolices). Therefore, the mean number of protoscolices per cyst was estimated, and the mean of this taken to indicate the mean number of protoscolices per ingested egg. This was not possible for *E. multilocularis*, due to the confluent, spreading nature of the cysts. Therefore, it was assumed that all rodents which were PCR positive for *E. multilocularis* were potentially infected, and this (rather than the presence of visible cysts, as used by Takumi and van der Giessen (2005)) was used as the denominator when estimating the mean number of protoscolices per infection. As a result of these adjustments, the rate of egg ingestion by rodents was an order of magnitude lower than that estimated in the Takumi model. The rate of egg ingestion by ruminants was higher than that of rodents, as would be expected due to the amount of grass ingested by a sheep or goat in comparison with a rodent, but may also be the cause of the particularly high estimated force of infection in ruminants (see below). As it is unlikely that the egg ingestion parameters will be able to be directly validated (because egg

contamination of the pasture is not likely to be directly quantifiable), identifying a suitable rate of ingestion would likely largely depend upon evaluation of the model outputs rather than attempting to directly quantify rates of egg ingestion.

7.4.4 Model validation

7.4.4.1 Data gaps

Full validation of the current model would require comparison of the model predictions with data obtained from the field. As field data is currently limited, attempts at model validation here will be largely qualitative (i.e. based largely on identification of general patterns and trends). However, it is hoped that further data collection will be conducted in order to improve upon this. It should be noted that, of the three general model compartments evaluated (adult worms, eggs, protoscolices), data is only likely to be easily available for the former (although even estimation of worm burdens is challenging due to the lack of an appropriate gold standard test for routine surveillance). Due to the difficulties in extracting and identifying eggs from soil samples (Craig *et al.*, 1988; Shaikenov *et al.*, 2004) and likely spatial aggregation in egg contamination (Hansen *et al.*, 2003, 2004), it is unlikely that the egg contamination compartment of the model will ever be able to be validated. Similarly, identification of rodent and ruminant infection is likely to be challenging (although not impossible) due to the low expected prevalences and spatial clustering of infection in the case of the former, and the lack of an abattoir system in these remote communities in the case of the latter. Despite this, some further investigation of infection in intermediate hosts (especially ruminants) would be very valuable for further work.

Although some data is already available on *E. multilocularis* infection in rodents in the study area (see above), the level of infection of ruminants with *E. granulosus* (or *E. canadensis*) in the Alay valley is currently completely unknown. When asked in May 2012 whether cysts were ever observed during home slaughter, a total of 43/95 (45%) of people reported that they had never observed any – suggesting that cysts are not commonly found in the area. This contrasts strongly with a report from an abattoir

survey in Naryn province, where the prevalence of cysts in young animals was around 50%; increasing to 100% as age increased (Torgerson *et al.*, 2009a). This may be suggestive of differences in endemicity between the two areas, or could result from information biases due to the reliance on self-reporting of cysts in the current study. Interestingly, all cases of human echinococcosis in the Alay valley identified to date through ultrasound scanning appear to be due to *E. multilocularis* (Paul Torgerson, personal communication), which would suggest a low endemicity of *E. granulosus* in this area, and would therefore corroborate the lack of reported cysts. However, both *E. granulosus* G1 and *E. canadensis* G6 are known to occur in the area (van Kesteren *et al.*, 2013). It is possible that *E. canadensis* G6 is the predominant cause of cystic echinococcosis in the area, which may have a predilection for pulmonary rather than hepatic sites in both humans and goats (the most common intermediate host for this species) (Varcasia *et al.*, 2007; Nikmanesh *et al.*, 2014; Alvarez Rojas *et al.*, 2014), and has also been found to be associated with brain lesions in humans (Sadjjadi et al., 2013). This would be expected to have repercussions for detection of infection, since pulmonary and cerebral lesions in humans would not tend to be diagnosed during ultrasound scanning, and lungs (although occasionally eaten in the area) may be more likely to be fed directly to dogs than inspected during slaughter. A better understanding of the species and strains of *E. granulosus* (s.l.) of epidemiological significance in the study area would allow better parameterisation of this component of the model, and could have considerable repercussions for control and surveillance.

The other main area of further investigation would be the possibility of using coproantigen data to parameterise the domestic dog compartments of the model – in particular, in the face of a control scheme (such as that described in the previous chapter). Coproantigen ELISA tests are relatively quick, easy and cheap to conduct, and as such are commonly used for surveillance. In these cases, large amounts of data may be obtained relatively easily. Through the use of the mixture model described in chapter 4, estimates of the 'burden score' for individual samples could be obtained. The exponent of these score estimates would be expected to broadly relate to the worm burden, and therefore could potentially be used to parameterise the model. As

there are a number of subtle differences between worm counts and coproELISA results, the model may need to be adjusted slightly in order to optimally represent transitions in infection status using coproantigen data. For example, it has been suggested that coproantigen levels are highest in the period just prior to patency and through to early patency, when the metabolism of the worm is highest (Kapel *et al.*, 2006). Therefore, the coproantigen levels would be expected to rise earlier than the corresponding burden, and fall in older infections (when egg output is also often reduced). A major constraint to application of this methodology to the current model is the lack of species specificity for current coproELISA tests (and the logistical difficulties associated with the use of PCR for surveillance). In a coendemic area such as the Alay valley, it is vital to be able to distinguish between *E. granulosus* and *E. multilocularis* if inferences on transmission pathways are to be made. Ultimately, it is hoped that coproELISA tests will be developed that are able to distinguish between *E. granulosus* and *E. multilocularis*. In the meantime, identification of methods of combining the results of coproantigen tests (conducted on all samples) with PCR tests (conducted on a selection of samples) in order to interpret broadly estimate the mean worm burdens for the different *Echinococcus* species would be useful, and would be worthy of further investigation.

7.4.4.2 Potential model faults

Two aspects of the current model were unexpected: firstly, the high predicted protoscolex burdens in ruminants (which reached the maximum level specified by the K_s parameter immediately after the lag period), and secondly the finding that complete control of *E. multilocularis* infection in dogs was also predicted to ultimately remove infection from rodents and foxes (suggesting that the foxes in the current model were acting solely as an overspill host from the domestic dog cycle). These are worthy of further scrutiny, as they do not necessarily agree with the expected transmission dynamics of *E. granulosus* and *E. multilocularis*.

As described above, one possible cause of the high rate of *E. granulosus* infection of ruminants is the high rate of egg ingestion. This issue would be further accentuated by

the high rate of egg production from dogs. Although egg production for both *E. multilocularis* and *E. granulosus* has been estimated to be in the region of 42 eggs per day (Gemmell *et al.*, 1986c; Matsudo *et al.*, 2003; Hansen *et al.*, 2003; Torgerson and Heath, 2003; Takumi and van der Giessen, 2005), this is highly variable. Experimental studies of *E. multilocularis* infection of dogs and foxes has suggested that the mean egg production per worm in each of these hosts was 114 over 77 days and 27 over 30 days, respectively – indicating a daily egg production of around 1 per worm for each host (Kapel *et al.*, 2006). Another report of experimental infection of dogs with *E. multilocularis* found very high variability in egg production, with one dog releasing a total of around ten million eggs over 132 days. Although the total number of adult worms in this case was not reported, using the data available this would be suggestive of an egg output of 1-2 eggs $worm^{-1}$ day^{-1}. Interestingly, in this experiment, egg production also appeared to come in synchronous waves: suggesting that proglottid release was synchronised between different individual worms (Matsumoto and Yagi, 2008). This yet again demonstrates a form of aggregation which was not accounted for in the current model: in this case, temporal aggregation in egg production (which would be expected to result in some degree of spatial aggregation of contamination). As a result of these issues, estimates of 'average' egg production are likely to generally be overestimates of the true egg contamination in most locations (and an underestimate in a minority of locations). Adopting a stochastic (and, ideally, spatially explicit) modelling approach which can account for these heterogeneities, as suggested earlier, would help remedy this problem.

The lack of predicted sustained transmission of *E. multilocularis* in foxes in the absence of canine infection is of particular interest, since the relative roles of dogs and foxes in the transmission of *E. multilocularis* is of potential public health importance. Domestic dogs are considered to be the main source of human infection with *E. multilocularis* is many areas (Craig and The Echinococcosis Working Group in China, 2006)), and the involvement of foxes in the canine *E. multilocularis* transmission ecosystem would be expected to impact upon the effect of control schemes. Interestingly, when all seasonality was removed from the model, vulpine

infection persisted even when the dog-rodent cycle was prevented (albeit at a lower level than that in the presence of canine infection) – as shown in figures 7.9 and 7.14. This predicted persistence in foxes remained when seasonality in host population densities alone was included (figures 7.11 and 7.14), but was removed when either seasonality in egg survival (figures 7.10 and 7.14) or mortality (figures 7.12 and 7.14) was included. Full interpretation of these effects is difficult without further investigation, but this demonstrates that the fox transmission cycle in the current model is sensitive to relatively small changes in parameter estimates. With the current parameterisation, reduced survival of eggs during the summer months appears to sufficiently reduce intermediate host burden to prevent effective transmission to foxes, and reduced access to rodents in the winter months also prevents effective transmission. This finding both demonstrates the potential impact of seasonality on the model predictions, and therefore the need to ensure that model parameterisation is as accurate as possible.

As the transmission parameters for vulpine infection were generally higher than those for canine infection, one likely cause of this perceived instability is the low number of foxes in the area of interest. One way to conceptualise the relative role of different host species in multi-host transmission ecosystems such as this one is in relation to the number of susceptible individuals of each species (Holt *et al.*, 2003) (although these effects become complex for parasites with indirect lifecycles, where the parasite may remain endemic despite low host population densities). According to this framework, the current model suggests that foxes and dogs are acting as 'substitutable' hosts, but that the low density of foxes precludes persistence in the absence of dogs (in this case, due to the reduction in the effective density of susceptible dogs due to praziquantel dosing), as schematically shown in figure 7.18.

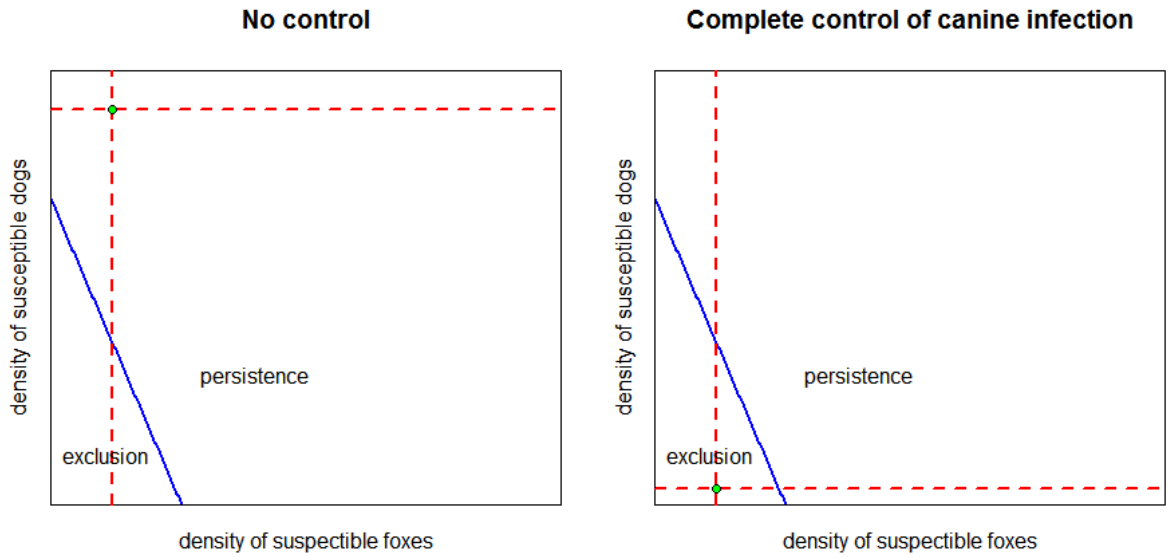


Figure 7.18. Schematic representation of the reason for the lack of *E. multilocularis* persistence in foxes predicted from the model. The vertical dotted line represents the density of foxes, and the horizontal dotted line represents the density of susceptible dogs in the presence and absence of control. The blue line indicates the threshold for parasite persistence. Based upon Holt *et al.*, 2003.

The schematic representation shown in figure 7.18 assumes that either foxes or dogs can maintain infection in the absence of the other (given that the fox density is sufficiently increased in the absence of dog infection, in this particular case). Although this generally agrees with the model predictions, it may not be the case in reality. A true rodent-dog cycle of *E. multilocularis* has been postulated to be present in some locations in China (Craig *et al.*, 2000; Pleydell *et al.*, 2008; Moss *et al.*, 2013), but has not yet been definitely proven. Therefore, the presence of an active dog-rodent transmission cycle which could persist in the absence of foxes (as would be predicted by the current model) is plausible, but requires further investigation. The other question is whether foxes would be able to maintain transmission in these settings in the absence of canine infection, since the current model predicts that complete control of canine infection (in the presence of seasonality) will also ultimately remove vulpine infection. Foxes associated with human habitation have been found to have lower levels of infection with *E. multilocularis* than those in rural settings (Stieger *et al.*, 2002; Hegglin *et al.*, 2007; Robardet *et al.*, 2008; Raoul *et al.*, 2010), which may suggest that vulpine transmission is less stable in these settings (likely due to reduced

ingestion of rodents). In this case, control in dogs would be expected to be more effective at reducing the infection pressure to rodents, and therefore would ultimately be expected to be more effective at removing infection from a community (see figure 7.13). However, this is unlikely to be the case – and, as described above, this model finding more likely represents limitations in the model structure and parameterisation. Other complexities in transmission are also likely to be present – for example, a low force of infection could be counterbalanced to some degree by a reduction in host immunity (Yimam *et al.*, 2002; Hansen *et al.*, 2003) and therefore a higher probability of infection, which would be expected to have a stabilising effect on the parasite. This was not accounted for in the current model, but again would be worthy of further investigation. In particular, the presence of in immunity in dogs (with the possibility that different species of *Echinococcus* will provoke different levels of immunity (Budke *et al.*, 2005b)) could have particular implications for the efficacy of control (Torgerson, 2006b).

The development of methods of quantifying the relative roles of dogs and foxes in transmission of *E. multilocularis* would be a useful area of further study. One method of achieving this is to estimate the R_0 (or Q_0 – see chapter 1) for each of the hosts. Reservoir hosts would be expected to have an R_0 of greater than 1, whereas spillover hosts would be expected to have an R_0 of less than 1. This general framework has been used to investigate the transmission ecology of *Schistosoma japonicum* in different settings in China (Rudge *et al.*, 2013), and would be very interesting approach for application to *E. multilocularis*. However, estimation of R_0 for each host species would be challenging in coendemic areas, since (unlike in the case of the *S. japonicum* model cited above) data on the source of intermediate host infection is currently not available for *E. multilocularis*. Approaches for estimating the relative host contributions in *E. multilocularis* transmission are therefore likely to be based upon measurement of definitive host infection either at equilibrium, or during control schemes (in a similar manner to reinfection studies). Estimation of parameters from these studies will generally require the fitting of data to a model framework, although

this to date has rarely been conducted in the face of a control scheme (Basáñez *et al.*, 2012a; b), and would be worthy of further study if more field data were collected.

7.4.4.3 Temporal trends

Two trends in particular were apparent both in the field data and in the model output: temporal trends over time during a control scheme, and seasonal trends for *E. multilocularis* and *E. granulosus*. As detailed in the previous chapter, differences in canine test prevalence were apparent between the spring and autumn visits. The coproantigen and *E. multilocularis* test prevalences were found to be higher in the autumn than the spring, and the reverse trend was true for *E. granulosus* and *E. canadensis*. These seasonal trends were also apparent in the output of the mathematical model – with peaks of definitive host infection with *E. multilocularis* in the autumn months, and peaks of infection with *E. granulosus* in the spring months (figure 7.7). These patterns are a result of the seasonal parameters included in the model. In the case of *E. granulosus* infection, the seasonal trends were associated with seasonality in intermediate host mortality, whereas in the case of *E. multilocularis*, trends in host population densities were also of importance (figures 7.11 and 7.12). Seasonality in egg survival alone appeared to have little effect on definitive host infection (figure 7.10). A mathematical model of fox infection with *E. multilocularis* in Zurich suggested that the force of infection for foxes was greatest during the winter months (Lewis *et al.*, 2014), which does not completely agree with the current results. This difference may be partly due to the current model assuming low access to rodents during the winter months due to heavy snowfall, which may not be the case in and around Zurich. However, the current model also predicts a reduction in the prevalence of rodent infection during the winter months, which was partly a result of a predicted increase in rodent mortality during the winter months (which, due to the κ parameters, would disproportionally reduce the total protoscolex biomass). It has been suggested that the proportion of infected rodents may actually increase during the winter months due to a lack of breeding (Burlet *et al.*, 2011). This possible effect was not incorporated in the current model due to the assumption of a balanced birth and death rate. More work is needed to better understand the dynamics of

intermediate host infection (and dog/fox access to these hosts) during the winter months.

As expected, most of the seasonality in egg contamination was associated with seasonality in egg survival (figure 7.10), with peaks in egg contamination in November and March corresponding to peaks in egg survival at these times, and lowest egg contamination during the hot months of July and August (see table 7.1). Seasonality in rodent infection appeared to also be associated with seasonality in egg survival (although the actual seasonal patterns observed as a result of this would depend upon the length of the prepatent period). Rodent mortality also impacted upon rodent infection, with lower mean burdens during the winter months (figures 7.10 and 7.12). As described above, the opposite trend may in fact be present during the winter months, due to a lack of 'dilution' of the mean protoscolex burden by young rodents. Very little seasonality in ruminant infection was observed, likely due to the high estimated infection pressure in the current model (described above).

As described in the previous chapter, the *Echinococcus* coproELISA prevalence and PCR prevalence for *E. canadensis* and *E. multilocularis* (and, to a lesser extent, *E. granulosus*) in the Alay valley appears to have been reduced during a praziquantel dosing scheme. Although the current model does not estimate the prevalence of infection (which is itself a problematic measure for overdispersed parasites such as *Echinococcus* (Hofer *et al.*, 2000)), the general trend of a decrease in the level of infection was also observed when praziquantel dosing was simulated in the current model. The initial trend reflected a simple removal of worm biomass from dogs. However, as time progressed, this was reinforced by reductions in the level of intermediate host infection and therefore the force of infection to dogs (and foxes, in the case of *E. multilocularis*).

7.4.5 Simulation of dosing strategies

The dosing strategies simulated in the current model assume that dosing coverage in the dog population was 100%. This is very unlikely to ever be achieved, but was

included in order to identify general trends and patterns. Further work to investigate the effect of different dosing coverages on the reduction in level of infection would be interesting, and could have practical advantages in terms of planning control campaigns (especially when unowned dogs are present in a community). It is also assumed that dosing with praziquantel immediately removes all eggs (and all adult worms) for the duration of the prepatent period, which meant that the model predicted that dosing dogs every month with praziquantel ultimately removed all infection from the study area. As well as the issues associated with the potential reservoir of *E. multilocularis* in foxes discussed above, this prediction does not account for natural stochasticity in prepatent period and onset of egg production. If egg production commenced before the next praziquantel dose, as has been observed following experimental infection of dogs with *E. multilocularis* (Matsumoto and Yagi, 2008) (where egg production was recorded as early as 26 days after infection), the transmission cycle could theoretically be maintained even in the presence of monthly dosing. Another potential issue is the fact that praziquantel does not inactivate eggs, and therefore it would be expected that following dosing of a dog, a large amount of egg contamination would take place (as dead worms containing eggs are expelled). This was not considered a particular problem for the current model, which focussed more with the long-term effects of dosing on the transmission dynamics, but could potentially be of epidemiological importance in the early stages of a control scheme.

The model predicted that administration of a single dose of praziquantel per year had no effect on the levels of infection of foxes or ruminants, but did reduce the mean egg contamination (especially for *E. granulosus* – see figure 7.4) and the rodent burden (figure 7.15). The greatest effect on reducing *E. multilocularis* egg contamination was observed when dosing was implemented in October, at the point where definitive host burdens were reducing but egg survival was increasing. Dosing during the summer months was less effective at reducing egg contamination, likely due to the reduced survival of eggs during this time even in the absence of control. This is a potentially useful finding, and incidentally has particular relevance to the study site due to the tendency of dog owners to travel to summer pasture during these months (meaning

that dosing frequency tended to decrease during these months, as detailed in the previous chapter). The results of the model suggest that targeting of control strategies during the autumn and early winter months could have the greatest impact upon egg contamination. The relative availability of dog owners during this time would also suggest that this strategy is logistically feasible. The general strategy of focussing control efforts just before and during the winter months has been suggested by the authors of a study of risk factors for water vole infection with *E. multilocularis* (Burlet *et al.*, 2011). This study suggested that living during periods of cold conditions was associated with a higher risk of infection with *E. multilocularis*, and therefore was suggestive of higher egg survival during these times.

Evaluation of the benefits of targeted dosing in comparison to random/regular dosing was complicated by the potential for random dosing to be implemented at optimal times by chance alone. For example, figure 7.4 suggests that the optimal timing of an annual dose of praziquantel to reduce *E. granulosus* would be January – yet, this month was also selected as the first month for the random dosing strategy. Therefore, a comparison of random dosing to targeted dosing would suggest no benefit of the latter even if there in fact was a benefit (this may explain some of the apparent lack of targeted dosing effect in comparison to random dosing in figures 7.16 and 7.17). To avoid this problem, the average effect of random dosing starting in each month could be used as the baseline effect of random dosing. This was not performed in the current analysis because of the difficulties in visualising the model output. The results of the current analysis suggest that targeted dosing has a relatively small impact on reducing the *E. multilocularis* biomass in a community, in comparison to regular dosing (figures 7.16 and 7.17). However, identification of optimal strategies of praziquantel dosing remains an area worthy of further exploration, as it could improve the cost effectiveness and therefore the sustainability of a control scheme.

7.4.6 Further work

Many specific avenues for further work have already been described or mentioned in the discussion, and will not be covered again here. It is clear that further work is

needed to better parameterise the current model, although it is unclear at the current stage how many of these issues result from flawed parameterisation, and how many from the effects of the limitations inherent in the representation of a stochastic system in a deterministic model (in particular, in relation to parasite aggregation, although some semi-deterministic models have incorporated overdispersion (Anderson and May, 1978; May and Anderson, 1978)). Sensitivity analysis of those parameters which are most questionable (especially the β parameters indicating the force of infection) would be useful for gaining a better understanding of some of these limitations, and their impacts upon model predictions. Despite the issues with setting a strict spatial focus of a complex system, continued surveillance in Taldu-Suu (and ideally the other three villages studied during the current project) would be useful for further characterisation of the effect of the ongoing praziquantel dosing scheme, and may provide useful data for future model parameterisation. As discussed above, further fieldwork in the area to identify the prevalence and intensity of ruminant infection with *E. granulosus* (and *E. canadensis*) would also be advised, especially in the light of low observed prevalences of cystic echinococcosis in people.

Given that suitable data for parameterisation become available, the ultimate aim will be to develop the current model into a more detailed model (such as an agent-based model or a metapopulation model). However, a 'bridging' model could also be developed from the current model framework by subdividing the host compartments according to risk factors for higher burdens (such as age or access to intermediate hosts), or in a spatial context (such as areas of pasture with higher moisture content and greater persistence of eggs) and modelling the parasite biomass within these, according to their relative frequencies in the population (Gurarie and King, 2005; Yakob *et al.*, 2014). These strategies could allow some of the heterogeneity in the system to be captured, and could also have potential benefits for modelling the effect of risk-based surveillance or control schemes targeting high risk/high burden individuals or areas. An further expansion of this idea would be to adapt the mean worm burden formulation used here (Macdonald, 1965) into a stratified worm burden (SWB) model (Gurarie *et al.*, 2010; Gurarie and King, 2014). The difficulty faced with

adopting the latter approach in the case of echinococcosis is the extreme degree of aggregation – where the parasite burden amongst infected hosts can range from individual worms up to tens of thousands, whereas for *Schistosoma* spp, the maximum burden considered was 150 (Gurarie *et al.*, 2010).

Another useful extension of the current model would be to combine it with an explicit economic model in order to evaluate the economic viability of different dosing and surveillance strategies, as is described in (Kato *et al.*, 2010). As well as improving the economic viability (and therefore the sustainability) of an echinococcosis control scheme, the output of such a model could aid the dissemination of information to risk managers and funding sources. Further investigation of other control approaches, such as evaluation of the effect of culling of definite hosts (Takumi *et al.*, 2008) or vaccination of sheep (Torgerson, 2006a) would be useful additions to the model, and could also be evaluated in an economic framework. Finally, as the current model explicitly models egg contamination of pasture, it could be expanded in order to explicitly model the potential risk of human infection through contact with eggs, as has been attempted in the Netherlands (Takumi *et al.*, 2012). Despite the relatively large number of models of *E. granulosus* and *E. multilocularis*, this has been rarely attempted (Atkinson *et al.*, 2013). A recent study has suggested that areas with high prevalences of human infection did not necessarily also have high levels of egg contamination in dog faeces (Chaâbane-Banaoues *et al.*, 2015), suggesting that other factors beyond levels of egg contamination are important in the acquisition of human infection, and that additional factors should be considered when modelling human infection.

Finally, as repeatedly mentioned above, the ultimate aim would be to develop the model into a stochastic, agent-based model, and/or a metapopulation model. These approaches would allow the explicit modelling of overdispersion, or spatial connectivity between the different model compartments. However, this is likely to require the collection of higher quality data than that currently available, and the development of this sort of model should therefore not be considered a primary aim until this data is forthcoming. The more basic models described above could be of

particular use in guiding data collection in preparation for the development of more complex models.

7.5 Conclusions

A deterministic mathematical model of transmission of *E. granulosus* (s.l.) and *E. multilocularis* in a coendemic area in the Alay valley, Kyrgyzstan, was developed and parameterised using a combination of published data from experimental studies, output of previous mathematical models, and general knowledge of the area. The model explicitly accounted for the lag period between exposure and patent/fertile infection in intermediate and definitive hosts, along with seasonal fluctuations in host population densities and mortality rates, and egg survival on pasture. Although the parameterisation is likely to be imperfect, and the model was not validated with field data, general trends observed in the course of a praziquantel dosing scheme in the study area were also apparent in the model output. The model was used to investigate the potential effects on *E. granulosus* (s.l.) and *E. multilocularis* biomass of targeted dosing of dogs with praziquantel instead of regular dosing. This suggested that when dosing is infrequent, then dosing during periods of higher egg production and higher egg survival resulted in a greater reduction in contamination of pasture with *E. multilocularis* eggs. However, this effect was reduced as the frequency of dosing increased. As the current model does not account for aggregation of parasites within hosts or in a spatial context, it is hoped that further developments to the model will ultimately incorporate this vital feature of parasite ecology and therefore better represent the transmission dynamics of these parasites in endemic areas and during control schemes.

Chapter 8: Canine echinococcosis in Kyrgyzstan: detection, diagnosis, and dynamics.

"Multa novit canis, verum echinus unum magnum"
"The dog knows many things, but the hedgehog knows one big thing"

(modified from Desiderius Erasmus Roterodamus (1466-1536); itself from Archilochus (680-645 BC))

8.1 Introduction

Zoonotic cestodes of the genus *Echinococcus* are distributed throughout the world, and in many areas are an important public health issue (Jenkins *et al.*, 2005). In particular, the number of cases of human echinococcosis is increasing in Central Asia (Torgerson, 2013), and the levels (and possibly the geographical distribution) of *E. multilocularis* infection amongst wild hosts in central Europe is increasing (Romig *et al.*, 2006). These changes are the result of a range of ecological and sociological effects, and management of them will generally benefit from a multidisciplinary approach (Giraudoux *et al.*, 2008). Despite this global distribution of *Echinococcus* spp, the burden of human disease (especially that resulting from *E. multilocularis* infection, which is commonly fatal in the absence of treatment) is aggregated within certain communities. Therefore, although the prevalence of alveolar echinococcosis is relatively low on a global scale, the burden of disease is high, and worthy of special attention in endemic areas (Torgerson *et al.*, 2010). The current thesis focusses on an area thought to be highly endemic for echinococcosis, and uses a variety of strategies to investigate infection in the definitive host of particular importance in relation to human infection – domestic dogs. Whilst most attention is focussed on surveillance strategies, it is hoped that further work with ecologists, epidemiologists, economists and sociologists may lead to improvements in our understanding of *Echinococcus* transmission in highly endemic areas, and to the development of sustainable control strategies.

The approach taken during the current thesis can be viewed from a variety of perspectives – the practical applications of statistical and mathematical methods in a surveillance context; the investigation of parasite ecology at a variety of levels; and from a more theoretical perspective regarding the pursuit of knowledge and the application of this within a transdisciplinary/interdisciplinary framework. At risk of appearing self-indulgent, each of these will be briefly described below, before the general conclusions arising from the current thesis are described. Finally, some possible areas worthy of further work will be described.

8.2 Practical framework of thesis: explaining the thesis motivation

Human echinococcosis is an increasing public health problem in Kyrgyzstan, especially in the south of the country. The reasons for this are associated with a range of cultural and socioeconomic factors which are difficult to address (although some efforts towards this aim are currently in place). As the most effective method of controlling human infection in an endemic area is to control infection in the definitive host (domestic dogs and foxes), a control scheme for echinococcosis based upon praziquantel treatment of dogs was recently instigated in Kyrgyzstan (WHO, 2011). Despite the availability of suitable tests for diagnosis of canine infection and effective drugs to remove canine infection, echinococcosis control is challenging due to the long duration of intervention required – which in many cases may be indefinite. As such, careful consideration of the control and surveillance strategy to be employed is essential to the sustainability, and the ultimate impact, of a control scheme. Despite this, in many highly endemic areas, this aspect of control is frequently overlooked in favour of generic rules and recommendations.

The current study has focussed on the identification of strategies which could be applied to echinococcosis surveillance and which may improve the effectiveness and sustainability of a control scheme. Aspects addressed have been methods of interpretation of diagnostic tests (chapters 3 and 4); approaches for classification of dog ownership patterns in communities and identification of risk factors for infection (chapters 5 and 6). The thesis work concludes with chapter 7, which describes a modelling framework which offers the potential for evaluating seasonality and the impact of control strategies in areas coendemic for *Echinococcus granulosus* (sensu lato) and *Echinococcus multilocularis*. In order to evaluate many of these strategies, faecal samples were collected and tested from four villages in the Alay valley of southern Kyrgyzstan over a period of 28 months during a control scheme based upon supervised dosing of dogs with praziquantel. Whilst useful results regarding patterns of dog ownership and seasonal and temporal trends have been obtained from this study, it is hoped that further work will further develop these ideas in order to

improve surveillance and control in the remote, resource-poor settings most commonly affected by echinococcosis.

8.3 Conceptual framework of thesis: explaining the thesis title

Just as the dynamics of echinococcosis operates on a variety of spatial and temporal scales (Giraudoux *et al.*, 2002, 2008), the current work has approached the investigation of echinococcosis on a variety of scales: from the individual parasite to the community of parasites in an area. These have been categorised into three groups (admittedly selected originally for their alliterative characteristics, but which have grown to represent some of the challenges facing *Echinococcus* control):

Detection. This relates to the identification of the presence of *Echinococcus* at the individual dog level, and was predominantly featured in chapter 4, where Bayesian mixture modelling was used to attribute a score to individual dogs based upon coproantigen ELISA data. Some of the limitations inherent in the mathematical model described in chapter 7 relate back to this issue, due to the effect of overdispersion on transmission dynamics (meaning that relatively few individuals would be expected to carry most of the parasite biomass). As such, detection of those individuals with the highest parasite burdens may be of use for targeted surveillance and control.

Diagnosis. This was selected to relate to the identification of echinococcosis at the population level. This was a central concept throughout the thesis – with aspects in each chapter. Chapters 3 and 4 identified methods of interpreting the results of coproantigen testing at the population level. Chapter 5 identified a novel approach for community-level data exploration prior to the implementation of a control scheme, in order to identify potentially important associations with *Echinococcus* test output. Chapter 6 described a temporal evaluation of an echinococcosis control scheme, with a novel approach to interpretation of data by aggregating at the household level rather than at the individual dog level. And finally, chapter 7 developed a framework for the mathematical modelling of echinococcosis in coendemic areas, by modelling parasite transmission using average measures for the community in question.

Dynamics. This was predominantly dealt with in chapters 6 and 7, which focussed on the identification and interpretation of seasonality in echinococcosis, the identification of temporal dynamics in the face of a control scheme, and potential methods of maximising the effect of praziquantel dosing strategies.

These three scales of investigation (individual dog, population, and temporal trends) are closely interrelated. As described above, the aggregated nature of infection within individual hosts would be expected to increase the effect of individual-level variation in parasite burden on the transmission process at the population level, leading to temporal and seasonal trends in infection, and so on. As such, any investigation of echinococcosis should concentrate on the whole system rather than any individual scale of interpretation.

One aspect of echinococcosis which was not explicitly covered in the current thesis was the spatial distribution of *Echinococcus* spp. *E. multilocularis* has been found to exist on a range of spatial scales beyond those defined by individual hosts: from the "patch" scale, to the "local" scale, the "regional" scale, and beyond (Tackmann *et al.*, 1998; Danson *et al.*, 2003; Giraudoux *et al.*, 2003; Shaikenov, 2006). The fact that transmission processes are clearly operating on a variety of different scales suggests that a "systems approach" to the investigation of echinococcosis would be beneficial (Giraudoux *et al.*, 2008), integrating analysis across these scales rather than focussing on only one scale in isolation. In particular, adopting a complex systems approach (Gisiger, 2001; Horwitz and Wilcox, 2005; Diez Roux, 2011) may allow a better understanding of *Echinococcus* transmission dynamics to be gained, with implications for our understanding of the transmission ecology of these parasites, as well as practical approaches for control and surveillance.

8.4 Philosophical framework of thesis: explaining the thesis subtitle

It is maybe unconventional to address the philosophical underpinnings of a PhD thesis directly within the thesis, but since this had a considerable impact on the direction the thesis took, it is worthy of mention here. The subtitle of this thesis, *"multa novit canis, verum echinus unum magnum"* is derived from the phrase *"multa novit vulpes, verum echinus unum magnum"* attributed to the ancient Greek poet Archilochus in Erasmus' *'Adagia'*. This phrase translates as "the fox knows many things, but the hedgehog knows one big thing". Despite the fortuitous coincidence that the etymological origin of *Echinococcus* is based upon the same Ancient Greek root as the Latin word for hedgehog (bioetymology.blogspot.co.uk, 2012), this phrase was selected because of its relation to the philosophy underlying the current thesis. The quote describes two main approaches for understanding the world, and was described further by Isaiah Berlin as such: "…there exists a great chasm between those, on one side, who relate everything to a single central vision, one system less or more coherent or articulate, in terms of which they understand, think and feel … and, on the other side, those who pursue many ends, often unrelated and even contradictory, connected, if at all, only in some *de facto* way, for some psychological or physiological cause, related by no moral or aesthetic principle" (Berlin, 1953). Whilst this is largely an artificial construct (as with most dichotomous classifications – see chapter 4), it does address two conflicting concepts of research. Namely, whether to follow the path of the hedgehog into perfecting skills within a single, clearly demarcated discipline, or to adopt the strategy of the fox, and attempt a variety of novel approaches to the problem in question. As described above, echinococcosis is a complex problem, and one which will (in most cases) require a multidisciplinary approach for effective control and management – and indeed, this will require effective collaboration between "foxes" and "hedgehogs". However, a decision was made during the current thesis to explore the application of a range of novel methodologies to the problem of *Echinococcus* surveillance (and, to a lesser degree, ecology and control). As a result, many of the conclusions made in the current thesis are intended to be merely stepping stones towards further work and development – as discussed in the respective chapters.

8.5 Description of study output

The practical (i.e. surveillance and control), conceptual (i.e. parasite ecology) and philosophical (i.e. transdisciplinary) frameworks described above led to the use of a variety of different analytic and investigative strategies in the current study. Some of these were contradictory – for example, despite arguing against dichotomisation of coproELISA data in chapter 4, coproantigen ELISA data was dichotomised in chapter 6; and despite discussing the importance of overdispersion in chapter 1, this was ignored in the mathematical model in chapter 7. However, all of these strategies help to build up a picture of the general aim of the thesis – to develop strategies for echinococcosis surveillance in remote areas where high-quality data may be difficult to obtain. These concepts will now be briefly detailed independently, before being summarised in aggregate.

8.5.1 Diagnostic test interpretation

An important component of the thesis was the interpretation of diagnostic tests – in particular, the *Echinococcus* coproantigen test (Deplazes *et al.*, 1992; Allan *et al.*, 1992). Since these were first developed in the early 1990s, the general approach to interpretation has been to specify a single cut-off, and interpret the OD value of samples in relation to this. The selection of this cut-off has generally been based upon a panel of known negative samples taken from a non-endemic area. This stratrgy was found to perform poorly in chapter 3, which also investigated alternative methods of determining a cut-off, including those based upon ROC curve analysis and mixture modelling. ROC curve strategies offered a number of advantages for tailoring a cut-off to the requirements of the analysis, but required suitable panels of known status; these were not required for the mixture modelling approach, but this made distributional assumptions which may not be appropriate. Chapter 3 also detailed the effect of variations in test performance (estimated sensitivity and specificity) on the estimated prevalence and the predicted true prevalence for a number of samples collected in Xinjiang province, China, and made some suggestions for accounting for this in interpretation.

Based upon the findings in chapter 3, chapter 4 developed a novel alternative strategy for the interpretation of coproantigen data by placing the mixture modelling strategy into a Bayesian context, and by incorporating logistic regression analysis. One important output from this strategy is the suggestion that a full panel of samples of "known" status (ideally endemic field samples classified using a gold standard diagnostic test) should be tested alongside the samples under investigation. This panel could have particular use for aiding the interpretation of ELISA data using a variety of methods, as well as offering the potential for sample standardisation (which is currently not routinely performed during the coproantigen ELISA test, and which may have particular implications for analysis of longitudinal data). The mixture model itself allowed the estimation of the probability that any particular sample is positive, but also allowed interpretation at the population-level (both in terms of the prevalence of infection and the distribution of positive samples). The prevalence estimate obtained from this model would be expected to have a higher sensitivity and specificity than prevalence estimates obtained from dichotomisation of individual samples, but further work is required to validate this suggestion. In particular, when the mixture model was applied to data collected from the Alay valley in Kyrgyzstan, the estimated prevalence was much lower than that based upon a single cut-off (chapter 5). Finally, by combining the individual estimates of probability of infection with the predicted log worm burden from a logistic regression model, an estimate of the log burden could be obtained, which may be useful for identifying and quantifying patterns of infection both at the individual and population level in the communities, both prior to and during a control scheme.

8.5.2 Classification of dog ownership in the Alay valley

Chapter 5 investigated the application of a multivariate technique, multiple correspondence analysis, to dog owners' responses to a number of questions prior to the implementation of a control scheme. It was hoped that this strategy could identify patterns of dog ownership in these communities, and may allow some form of community-level risk profiling prior to the instigation of control. Using this approach, a general picture of dog ownership in the study villages was able to be constructed,

which described patterns in livestock ownership, dog type, dog demographics and owner knowledge/behaviour. These dog ownership patterns appeared to differ between the four different villages, which could have implications for *Echinococcus* transmission ecosystems or control strategies in these communities (although no evidence of this was found in the data). Also, some associations between patterns of dog ownership and the results of diagnostic testing of faecal samples were obtained. Dogs from households which did not own livestock or visit Jailoo, and those from households which appeared to have a better understanding of echinococcosis, appeared to have a higher probability of *E. granulosus* PCR positivity. Identification of the relative role of dog ownership patterns and village-specific differences could be an area of further work. Also, use of hierarchical clustering methods to group individual dogs with respect to the identified patterns would allow more formal investigation of risk factors for infection and therefore this approach could offer benefits for quickly classifying households and villages with regards to potential risk factors of importance prior to the implementation of a control scheme. It is hoped that techniques such as these will be used in combination and in parallel with traditional regression techniques, such as that described in a recent analysis accepted for publication in the Journal of Helminthology (section A7 of the appendix), in order to maximise the information obtained during surveillance activities.

8.5.3 Longitudinal evaluation of a control scheme in the Alay valley

In chapter 6, information theoretic and model averaging approaches were used to develop a predictive model of seasonal and temporal trends in coproantigen and coproPCR positivity over the course of 28 months, in the face of a praziquantel dosing scheme. Due to the inability to reconcile differences between the ELISA and the PCR tests, separate models were created for each of these tests. However, further work to incorporate these different outcomes in a single modelling framework would be strongly advised, and would offer considerable benefits to surveillance in coendemic areas. In order to investigate risk factors for infection in situations where linking individual faecal samples to individual dogs is not possible (which would be expected to be a common occurrence in areas with aggressive dogs, or dogs which are free to

roam), analysis was conducted at the household level. Since many control strategies for echinococcosis are focussed at this level (in the current case, the local veterinarian was instructed to travel from household to household to dose dogs), it is hoped that information loss through aggregation will be minimal. However, care should be taken when attempting to interpret dog-level risk factors (including reported history of praziquantel dosing) based upon household-level predictions. Also, this approach reduces the ability to identify age-related patterns of infection, which are generally required for identification of the force of infection and evaluation of possible immunity. Therefore, when collecting samples prior to an intervention scheme in particular, it may be beneficial to focus more on attributing samples to individual dogs. Despite these caveats, the interpretation of results at the household level may allow the identification of household-level factors associated with test positivity, and therefore may be useful for identification of control scheme failures, or areas worthy of additional attention.

This modelling strategy identified a general reduction in test prevalence over time, with seasonal increases in test prevalence. However, the only significant factors for *E. granulosus* PCR positivity was an increase in test prevalence during the September 2013 visit. The cause of this is unknown, and may be worthy of further investigation. Coproantigen ELISA and *E. multilocularis* PCR prevalence were found to be higher in the autumn months than the spring months, whereas the reverse was true for *E. canadensis* (and, qualitatively, *E. granulosus*) PCR prevalence. This may indicate different patterns in access to infected intermediate hosts. Also, there appeared to be an effect of the presence of small dogs and young dogs in the household on the probability of coproantigen ELISA and *E. canadensis* test positivity, which may be indicative of some dogs not being dosed, or may represent age-related immunological effects. Further development of this strategy will focus on interpretation of coproantigen results in a continuous fashion (either by modelling the OD value directly, or by using the output of the mixture model described in chapter 4).

8.5.4 Development of a novel mathematical model of *Echinococcus* transmission

Chapter 7 detailed a novel mathematical model of transmission of *E. granulosus* (sensu lato) and *E. multilocularis* in areas where both are coendemic and both infect domestic dogs. Previous mathematical modelling strategies applied to coendemic areas have generally focussed on estimating transmission parameters rather than explicitly modelling transmission in the whole system (Budke *et al.*, 2005b; Ziadinov *et al.*, 2008). The current model aimed to represent the complete transmission system, and due to the complexities associated with this focussed on a relatively small area (based around one of the study villages). As the parasite biomass was modelled rather than host infection, egg contamination of pasture was able to be explicitly modelled, which offers the potential for model adjustment in order to explicitly model human infection. Although model parameterisation was very limited, the model may have some potential for the broad evaluation of the impact of control schemes and seasonality on transmission, and in particular may be useful for the identification of data gaps worthy of further attention. Simulation of the effect of dosing dogs with praziquantel at set points during the year on egg contamination of pasture suggested that dosing during the autumn and early winter months was most effective at reducing contamination. Comparison of the results of this with those of regular dosing suggested a relatively small benefit of targeted dosing, but as targeted dosing over the winter months may be more logistically feasible in these communities, may be an area worthy of further exploration. It is hoped that further model development will allow the evaluation of cost-effective dosing campaigns, which may be more sustainable in the long-term.

One major constraint to the model was the lack of incorporation of overdispersion in parasite burdens within individual hosts or in the environment. As these features would be expected to have implications for the stability and transmission ecology of *Echinococcus* spp, it is hoped that as more data become available, the model will be expanded in order to account for heterogeneities in transmission in a spatiotemporal context. Although deterministic models have been developed which incorporate

overdispersion using the negative binomial distribution (Anderson and May, 1978), this may be problematic for the evaluation of the effect of control schemes, which would disrupt the distribution of parasites within a community. Stratification of model compartments in order to capture different transmission patterns within the mean worm burden framework (Gurarie and King, 2005) may be useful, but the ultimate aim would be the development of the model framework into an agent-based modelling structure, and/or a metapopulation model. A major question in need of answering is the relative roles of domestic and dogs and foxes in the transmission cycle of *E. multilocularis*. In the absence of molecular tests which are able to differentiate between strains of *E. multilocularis* cycling between the different hosts, investigation of rates of (ideally, canine and vulpine) reinfection in the face of a control scheme may offer the best chance of further clarifying these transmission pathways and dynamics.

8.6 Areas of further study

As described above, the overarching theme throughout this thesis has been the investigation of approaches for surveillance of canine echinococcosis with a view towards improving the sustainability and effectiveness of control strategies, as effective control of domestic dog infection is likely to have the greatest impact on the risk of human infection. Although a variety of concepts and approaches are identified and evaluated in relation to each of the current study areas described above, further work is needed in order to fully realise these methods. Many specific areas for further study are described in the individual chapters themselves and in the brief summaries above. However, a number of areas of further study which incorporate aspects of all of these study areas can also be identified by looking at the system as a whole.

The three conceptual themes in the study (individual level interpretation, population level interpretation, and temporal trends) can be related back to the improvement of echinococcosis surveillance and control in the following way:

Detection (individual level interpretation): identification of high burden dogs (for example, using the mixture model described in chapter 4, or by using ROC curve analysis as described in chapter 6).

Diagnosis (population-level interpretation): methods of describing and "risk profiling" dog ownership patterns in a community (as described in chapter 5) and summarising the OD distribution amongst positive individuals (chapter 4).

Dynamics (temporal trends): the identification of the seasonal effects and the evaluation of praziquantel dosing strategies at the individual and population level (as identified in chapters 6 and 7).

The main constraint to further development of this work is the limited availability of high quality data – in particular, for validation of the mixture model and the mathematical model. Since the gold standard test for canine echinococcosis is based upon necropsy, collection of gold standard test data from domestic dogs is rarely possible. However, by coordinating with local fox hunters (Ziadinov *et al.*, 2010), and those responsible for dog culling in communities undergoing culling campaigns, some data may be obtainable. It is important that this is done with the full knowledge of the communities involved, since dog culling in particular is a contentious issue (Johansen and Penrith, 2009; Atema and Hiby, 2015). Alternatively, arecoline purgation has a high specificity (in the hands of trained operatives), and could be used in these communities. For further work, this would be particularly useful if conducted prior to the implementation of a control scheme, when the parasitological system is at a 'steady state' – although investigation of trends in infection in the face of a control scheme would also be of interest (Basáñez *et al.*, 2012a; b). However, this strategy is unlikely to be feasible for ongoing surveillance.

Another aspect of diagnostic testing which requires further attention is the incorporation of PCR testing with coproantigen ELISA results in a surveillance context. Until species-specific coproantigen tests are available (WHO/OIE, 2001d), coproPCR is required for diagnosis of canine echinococcosis to the species level.

However, as PCR is generally considered inappropriate for surveillance (Deplazes *et al.*, 2003; Torgerson and Deplazes, 2009), it is commonly used as a confirmatory test in surveillance settings (Eckert and Deplazes, 2001). This approach was not adopted in the current study, due to the apparent independence of ELISA and PCR test results, and instead a random selection of samples underwent PCR testing regardless of PCR status. Although this caused difficulties in the current study, the apparent independence of the two tests offers potential for the development of a method of incorporating both ELISA and PCR tests, for example based upon latent class methods (Hartnack *et al.*, 2013). Alternatively, the development of molecular methods of identification to the species or strain level which can be applied in a surveillance context would be worthy of further exploration. For example, loop-mediated isothermal amplification (LAMP), which has already been developed for *E granulosus* (s.s.) (Salant *et al.*, 2012; Ni *et al.*, 2014a) and *E. multilocularis* (Ni *et al.*, 2014b); or recombinase DNA polymerase amplification (RPA) (Piepenburg *et al.*, 2006). As these approaches do not require expensive equipment such as thermocyclers, they may also be more appropriate than PCR in resource-poor communities in developing countries.

Based upon the themes identified above, one particular area worthy of further study is the identification of high burden individuals/households, in order to concentrate surveillance and control campaigns upon these. Due to the potentially high *Echinococcus* burdens which can be reached in definitive hosts (measured in thousands), *Echinococcus* burdens have been found to approximate a power law distribution (Jenkins and Morris, 1991; Hofer *et al.*, 2000; Stieger *et al.*, 2002). As a result of this aggregation (as is seen for many diseases), the majority of transmission would be expected to occur through a minority of the individuals (Anderson and May, 1991c; Woolhouse *et al.*, 1997; Perkins *et al.*, 2003). Whilst aggregation in a natural setting is thought to stabilise transmission dynamics overall (Anderson and May, 1978) (although this may be dependent upon the exact generating process for the aggregation (Rosà and Pugliese, 2002)), it has been argued that control schemes may be more effective when the outcome of interest follows a power law distribution (Woolhouse *et al.*, 1997; Gladwell, 2006), due to the effect of the few high impact

individuals on driving the process. By focussing control on these most influential individuals (in the case of echinococcosis, those few dogs containing the majority of the parasite biomass), a considerable effect on the levels of the outcome could be achieved (Woolhouse *et al.*, 1997). Identification of high burden individuals could be achieved through the mixture modelling framework described in chapter 4, and could also be incorporated into a longitudinal model such as that described in chapter 6. Particular attention could be focussed on identifying risk factors for high burdens (including at the household level), since this may offer a method of specifically targeting these individuals. Conversely, if these dogs cannot be identified based upon their exposure history, control efforts may be better focussed on methods of maximising overall praziquantel coverage in a community and minimising the risk of any individual being missed. In the same way that controlling infection in the few dogs harbouring the majority of the parasite biomass in the community would be expected to be rapidly reduce the levels of infection, if these individuals were missed, the control scheme would be expected to not be effective.

Ultimately, it is hoped that mathematical modelling approaches will become increasingly useful in the planning and monitoring of control schemes and surveillance strategies for echinococcosis. As described earlier, parameterisation of complex models such as this require the collection of higher quality data than that available currently from the current field site, and it is hoped that data collection in the Alay valley will be continued to some degree as the current control scheme progresses. Incorporation of data collected during helminth control schemes into a mathematical modelling framework can offer valuable insights into transmission dynamics, and can allow control strategies to be tailored accordingly (Churcher *et al.*, 2006; Basáñez *et al.*, 2012b). However, very little is currently known about the dynamics of echinococcosis in the face of praziquantel dosing, since longitudinal data is rarely collected during echinococcosis control – instead, data is commonly collected prior to a control scheme, and then after a number of years have passed. Although imperfect, the interpretation of coproantigen ELISA-based longitudinal data (such as that described in chapter 6) in a semiquantiative manner through use of the mixture

modelling strategy described in chapter 4, in a mathematical modelling framework such as that described in chapter 7, could offer great benefits to our understanding of echinococcosis control.

As a final note, following on from the discussion above, the concept of power laws and fractal relationships could also be investigated in relation to the ecology and epidemiology of echinococcosis itself, by adopting a complex systems approach to modelling. As described in chapter 1, the concept of 'criticality' shares a number of characteristics with the distribution and transmission patterns of *Echinococcus* spp. As described earlier, the form of criticality can have considerable implications on the optimal control scheme (Pascual and Guichard, 2005), but another important issue is that, the timing of events in these systems is inherently unpredictable, even if the distribution of their relative frequency and magnitudes can be predicted and quantified. A good example is that of earthquakes, which have been found to follow a power-law distribution (the Gutenberg–Richter law: basically, small earthquakes are relatively common whereas large ones increasingly rare), but which can still generally not be predicted in advance (Bak *et al.*, 1994)). Work to further identify and clarify these would be interesting and could lead to both an improvement in control strategies and a better appreciation of the unpredictability inherent in these systems (especially in combination with the population explosions of intermediate hosts of *E. multilocularis*, which are known to occur in many endemic areas (Giraudoux *et al.*, 2013b)). Accounting for this natural stochasticity may allow a better understanding of the transmission dynamics of *Echinococcus* spp in endemic areas.

Appendix

"Science is not finished until it's communicated"

Sir Mark Walport (1953 –)

A1. The Reproduction Ratio

The basic reproduction ratio, R_0, in microparasite epidemiology (e.g. viruses and bacteria) is usually used to describe the average number of secondary cases resulting from each infection in a totally susceptible population (Anderson and May, 1982). Although the concept of R_0 has been in use in epidemiology since the early 20th century (Kermack and McKendrick, 1927; Heesterbeek and Dietz, 1996), it was not formally named (as 'Z_o') until the 1950s (MacDonald, 1957), and it was not popularised until the 1980s (Anderson and May, 1979; May and Anderson, 1979; Heesterbeek and Dietz, 1996). Following entry of a pathogen into an immunologically naive (i.e. susceptible) population, two possible 'steady states' can be considered. If R_0 is less than 1, the 'extinction steady state' would be expected, where the asymptotic prevalence of infection and the infection pressure are both zero. However, if R_0 is greater than 1, a nonzero proportion of susceptible animals will remain at the steady state (known as the 'asymptotic proportion'). In this steady state, the presence of immunity prevents further increases in infection prevalence regardless of the true value of R_0. In order to understand this, a related parameter, R, has been devised which describes the average number of secondary cases resulting from each infection in the presence of immunity. Since it can be assumed that sustained transmission will only take place when R is greater than 1, R has potential use as a 'threshold quantity' when considering the likely effect of vaccination schemes on pathogen transmission (which will aim to reduce R to less than 1 by stimulating immunity in a population). Despite this, R_0 (and R) is not a pancea: pathogens with an R_0 of less than 1 can spread in a population, and those with an R greater than 1 can become extinct (Roberts, 2007; Li *et al.*, 2011).

Three general methods of estimation of R_0 have been described. At the basic level, as described here, most of these rely on the assumption of a homogenously mixing population with no births or deaths (although methods are available to incorporate more complex structures, such as heterogeneous mixing). The first approach is based upon parameterisation of a model of transmission processes. For example, R_0 can be estimated as the product of the number of contacts an infectious individual has per

unit of time, the conditional probability of transmission given this contact, and the duration of infectiousness in an infected individual (D). However, as these parameters are often not directly known, alternative approaches to the estimation of R_0 have been developed.

An alternative approach is to estimate R_0 using data on the prevalence of infection in an endemic population in a 'steady state' of transmission (where the proportion of susceptible animals is no longer changing over time). In this situation of an unchanging prevalence over time, R will be equal to 1, and as R is directly related to R_0, R_0 can be estimated. R can be calculated as the product of the proportion of susceptible individuals in a population and R_0, and therefore this equation can be reformulated to allow R_0 to be estimated from the proportion of susceptible individuals in the population (S):

$$R_0 = \frac{1}{S}$$

The final method for estimation of R_0 is based upon the rate of increase in number of infected individuals (Λ) following introduction of a pathogen into a completely naive population. This approach assumes that the measurements are taken during the 'exponential phase' of transmission, meaning that the population must be large enough for there to be effectively unlimited susceptible individuals over the period of measurement (Dohoo and Medley, 2009). As the rate of increase in Λ can be estimated as the product of the rate of loss of infectiousness and $(R_0 - 1)$, this equation can be reconstructed to give $(R_0 - 1)$ as the product of Λ and the duration of infectiousness, D ($= {}^{1}/_{rate\ of\ recovery}$):

$$R_0 = \Lambda \mathrm{D} + 1$$

A similar approach to this was been used for the estimation of the R_0 for Severe Acute Respiratory Syndrome (SARS), where no endemic steady state was reached and insufficient parameter data were available to use the alternate methods (Lipsitch *et al.*, 2009).

These approaches for estimating of R_0 have been used for cestodes (Roberts *et al.*, 1987), although there are some caveats to interpretation as described below. If R_0 is greater than 1, it can be estimated as the reciprocal of the proportion of susceptible animals at the steady state (the 'asymptotic proportion'). The asymptotic infection pressure, h, can then be estimated as:

$$h = g(h)(R_0 - 1)$$

Where $g(h)$ is a function which denotes the proportional reduction in the asymptotic infection pressure due to development of immunity resulting from re-exposure. This can be reformulated into the following equation:

$$R_0 = 1 + \frac{mean\ duration\ of\ immunity}{mean\ time\ to\ immunity}$$

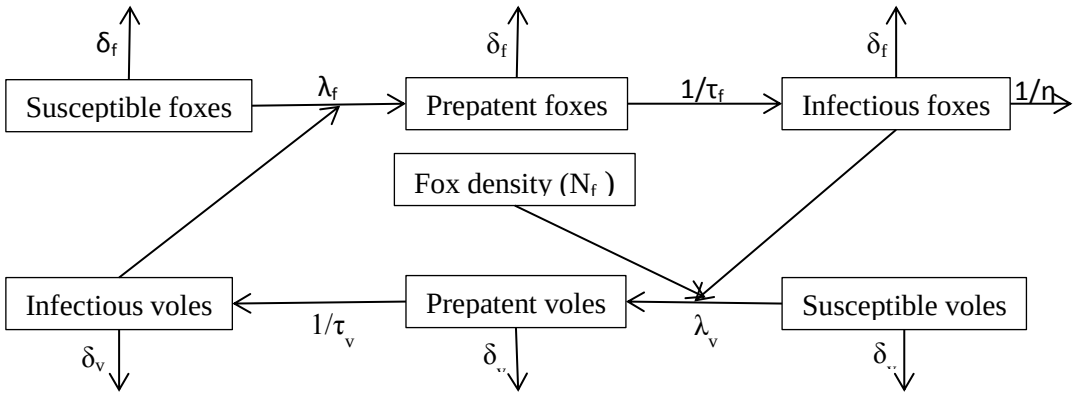
And if it can be assumed that infection pressure is sufficiently high so as to result in lifelong immunity following exposure (i.e. immunity resulting from repeated exposure, in the case of cestodes), this formula can be reformulated as (Roberts *et al.*, 1987):

$$R_0 = \frac{mean\ life\ expectancy}{mean\ age\ of\ immunity}$$

This equation is very similar to the estimation of R_0 as the inverse of the susceptible proportion at the steady state as described above, since in a population with equal numbers of animals of different ages, the ratio ($mean\ age\ of\ immunity/mean\ life\ expectancy$) represents the susceptible proportion. Therefore, inverting this gives the inverse of the susceptible proportion. Note, however, that it is unlikely that the population will have even representation of all ages (since older sheep may be more likely to be removed through slaughter). In the case of these 'negative exponential' populations (where most individuals are young), the formula can be adjusted by adding 1 to the estimate, as was used in a study of arboviruses (Dietz, 1974).

Using these equations on worm burden data collected from infected sheep, and assuming that a lack of increasing burden with age was suggestive of immunity, R_0 was estimated as between 3.0 and 3.5 for *Taenia.hydatigena* (with a higher estimate for female sheep than male sheep, due to their longer life expectancy), and between 1.0 and 1.5 for *E. granulosus* (Roberts *et al.*, 1987) – although as the timing of infection (especially with *T. hydatigena*) in these cases was left-censored, and so the true values of R_0 may in fact be higher. The impact of this difference between estimated R_0 for *T. hydatigena* and *E. granulosus* was that a dog dosing campaign was more effective against *E. granulosus* than against *T. hydatigena* (in fact, there was evidence of an increase in levels of *T. hydatigena* infection in older sheep in the face of the dosing campaign, due to the effects on acquired immunity, which could have repercussions for canine infection with this cestode) (Gemmell *et al.*, 1986b; Roberts *et al.*, 1987).

An estimate of R_0 has also been suggested for *E. multilocularis*, which accounts for both definitive and intermediate hosts simultaneously (Roberts and Aubert, 1995). Fuller details of the model structure and parameterisation are given in chapter 1, but the structure will be repeated here for clarity:



In order for the number of worms in the second generation to be greater than that in the current generation, the product of the sequential processes leading to transmission (excluding the size of the compartments) must be greater than the product of the processes resulting in loss of worms from the cycle:

$$\lambda_f . \frac{1}{\tau_f} . N_f . \lambda_v . \frac{1}{\tau_v} > \left(\delta_f + \frac{1}{\tau_f}\right) . \left(\delta_f + \frac{1}{\eta}\right) . \left(\delta_v + \frac{1}{\tau_v}\right) . \delta_v$$

This results in the following formula for R_0, which has been restructured by Ishikawa to give the equation on the right (Ishikawa, 2006):

$$R_0 = \frac{\lambda_f \lambda_v N_f}{\tau_f \tau_v (\delta_f + {}^1/_{\tau_f})(\delta_f + {}^1/_{\eta})(\delta_v + {}^1/_{\tau_v})\delta_v} = \frac{\lambda_f \lambda_v N_f}{\delta_v (1 + \delta_v \tau_v)(1 + \delta_f \tau_f)(1 + \delta_f \eta)} \eta$$

This can be shown to be based on the same concept introduced in the first equation (i.e. the inverse of the susceptible proportion of animals) once parameterised (Roberts and Aubert, 1995):

$$R_0 = \frac{2}{(1 - P_{fox})(2 - 3P_{vole})} \approx \frac{1}{(1 - P_{fox})} = \frac{1}{S_{fox}}$$

Where P_{fox} and P_{vole} represent the prevalence of infection in foxes and voles, respectively (the latter of which can be assumed to be approaching zero, hence its removal from the equation), and S_{fox} is the proportion of susceptible foxes.

The estimate of R_0 proposed by Ishikawa is similar to the Roberts equation above, but includes additional parameters in the estimation of the transmission to foxes and the rate of egg production:

$$\frac{s_f f \lambda_v N_f}{\delta_v (1 + \delta_v \tau_v)(1 + \delta_f \tau_f)(1 + \delta_f \eta_h)} \left(\rho \eta_h + \frac{\eta_l}{1 + \delta_f \eta_l}\right)$$

Where s_f indicates the probability of an ingested cyst developing to a mature worm in a fox, f is the average number of voles ingested per fox per day, η_l is the duration of low egg production, η_h is the duration of high egg production and ρ is the multiplicative factor for high egg production (cf low).

As mentioned above, the approaches described here assume homogenous mixing of a population, with no differences in the rate of transmission between different groups within the population. This is rarely the case, and in situations where transmission

(and therefore R_0) differs between groups within a population, the interpretation of an 'average' value can be challenging. In order to tackle this problem, the conceptual mathematical foundations of R_0 were clarified, leading to the definition of R_0 as the 'dominant eigenvalue of a positive linear operator' relating the number of infected hosts in one generation to that in the next generation (Diekmann *et al.*, 1990). This relationship can be used to estimate R_0 in heterogeneously mixing populations where the infection pressure may differ between different groups of individuals.

As alluded to earlier, the techniques described above have been derived from those used for the investigation of the R_0 for microparasites such as bacteria and viruses. Applying the concept of R_0 to macroparasitic infections is more complicated, due to the need to account for parasite numbers (rather than just classifying individuals as 'infected'), and due to the fact that increases in the parasite burden generally result from reinfection from the environment or intermediate hosts, which can exhibit considerable variability (Heesterbeek and Roberts, 1995) (rather than through multiplication within the host, as is generally the case with microparasite infections). Anderson and May defined R_0 for macroparasites as 'the average number of offspring … produced throughout the reproductive life span of a mature parasite that themselves survive to reproductive maturity in the absence of density-dependent constraints on population growth' (Anderson and May, 1991b).

Anderson and May attempted to incorporate an indirect parasite lifecycle into an estimate of R_0 by estimating 'transmission factors', T, which describe the transmission from definitive hosts to intermediate hosts (T_1), and from intermediate hosts to definitive hosts (T_2). If these are known, the overall R_0 can then be calculated as some form of the product of these factors (T_1T_2) (Anderson and May, 1991a; c). By accounting for both hosts in the estimation of R_0, interesting characteristics of the epidemiology of the pathogen can be identified and investigated. For example, it has been shown that the R_0 of directly transmitted pathogens is positively associated with the population density of (susceptible) hosts (Anderson and May, 1991c). However, a 'reservoir' of infection in intermediate hosts can result in a macroparasite remaining stable even if the density of definitive hosts is low (Anderson and May, 1991c). This

basic methodology was also used to estimate a threshold quantity, titled Q or Q_0, for nematode infections with a free-living stage, assuming no heterogeneity in host or parasite populations (Roberts and Grenfell, 1991; Heesterbeek and Roberts, 1995). This quantity describes the number of parasites (or female parasites, in the case of dioecious species) reaching reproductive maturity which are produced by an individual adult parasite over its whole reproductive life span in the absence of density-dependent constraints (Heesterbeek and Roberts, 1995).

One other consideration which arises from the indirect lifecycle is that the R_0 would be expected to be affected by the wide variety of factors which impact upon the different hosts, the parasites and the rates of contact between these. This variability ('periodicity') in intermediate host (or environmental) conditions, which are required for completion of the cycle, can make the estimation of an 'average' number of offspring difficult, and therefore has repercussions for the calculation and interpretation of any single value of R_0. Instead of estimating an R_0 value for a 'typical' individual, a 'time-averaged' value may be more appropriate (Roberts and Grenfell, 1992; Mollison *et al.*, 1994). As for the situation with heterogenous mixing of populations, a clear mathematical definition of an R_0-like parameter would be useful (Diekmann *et al.*, 1990; Heesterbeek and Roberts, 1995). A number of alternative measures with a similar 'threshold' characteristic, such as Q_0, $\lambda_d(E)$, and P, have been defined (Heesterbeek and Roberts, 1995). which have a mathematical foundation and allow for investigation of thresholds in the same way as R_0 (Roberts and Grenfell, 1991; Heesterbeek and Roberts, 1995; Roberts and Heesterbeek, 1995). Although Q_0 has direct biological relevance (being based upon the product of the transmission factors describing transitions between parasite stages), it lacks a clear mathematical foundation. In order to remedy this, it has been described as the dominant eigenvalue of a matrix (K) of transmission functions, raised to the power k, where k is the number of stages in the model. As well as retaining the threshold properties of R_0, this approach allows the incorporation of different stages of parasite, different types of hosts and/or different parasites (e.g. competition between similar species), as required, in the model (Heesterbeek and Roberts, 1995; Roberts and Heesterbeek,

1995). Whilst both the 'traditional' R_0 (relating to the number of secondary infections per primary infection) and Q_0 (relating to the number of individual parasites resulting from infection a single parasite) interpretations share a "threshold" quality relating to the ability of infection spread in a community (with a value of 1 being the threshold between eventual persistence and decay), the magnitude of their estimates would be expected to differ.

A2. Healthcare and pastoralism in Kyrgyzstan

A2.1 Public health in Kyrgyzstan

Under Soviet control, public health in Kyrgyzstan was prioritised and was served through a well-developed infrastructure of sanitary-epidemiological ('san-epid') stations, which were mainly focused on the surveillance and control of infectious diseases through vaccination campaigns and health education (Glass, 1976). Despite this, education for medical professionals in public health under the Soviet regime was limited (Ibraimova *et al.*, 2011). This system was effective at reducing or eliminating a large number of infectious diseases by the 1980s, including cholera, plague, polio, pertussis, typhoid, measles, rabies, anthrax and tuberculosis (Meimanaliev *et al.*, 2005). Following independence, the healthcare system was restructured, although a centralised structure was retained (Gotsadze *et al.*, 2010), resulting in the formation of the Republican Centre for Immunoprophylaxis in 1994 (which, along with assistance from international donors, has ensured high immunisation levels have been achieved), the Department of State Sanitary-Epidemiological Surveillance (DSSES) in 1997 (which is responsible for surveillance of infectious diseases and sanitary inspection), and the public health unit in the Ministry of Health in 2006 (which is responsible for public health policy) (Meimanaliev *et al.*, 2005; Ibraimova *et al.*, 2011). However, the public health service has been relatively slow to improve, compared to other components of the health sector (see below), and has experienced problems with poorly equipped laboratories, limited training of workers and inefficient flow of data. Public health services are currently in the process of reform, led primarily by the State Sanitary Epidemiological Surveillance (SSES) service (using both top-down [state-organised] and bottom-up [district-level] approaches), although the Republican Health Promotion Centre (which was created in 2001) is also involved (Ibraimova *et al.*, 2011). Laws are currently in place to ensure that sanitary inspection of food and water is conducted appropriately, and procedures for epidemiological surveillance are currently being developed (Ibraimova *et al.*, 2011). In villages throughout the country, health committees are in the process of being established (following on from a recent pilot programme, organised by the Kyrgyz-Swiss Health Reform Project), which aim to

provide community-based health promotion (Ibraimova *et al.*, 2011). A recent study has suggested that control of soil-transmitted helminths in Kyrgyzstan should be 'feasible', partly due to strong health and education systems (although political and financial obstacles may be present) (Brooker *et al.*, 2015).

A2.2 The Kyrgyz healthcare system

Under the Soviet regime, healthcare was provided solely by the state, and was free at the point of use. However, much of the human medical provision was hospital-based, with little attention given to primary healthcare or disease prevention (Belli, 2001) (which in urban areas was generally provided by polyclinics, and in rural areas was provided by health workers) (Rechel *et al.*, 2013). Additionally, as with much of the Soviet regime, more attention was given to inputs (such as staffing) and infrastructure than to outputs (such as quality of care), and the healthcare system as a whole was relatively fragmented between different departments and regions (Sargaldakova *et al.*, 2000; Rechel *et al.*, 2013). Following independence there was a rapid fall in government revenue (due to a loss of subsidies from Moscow, along with general economic collapse), with a considerable impact on the healthcare system, which foundered (partly due to the considerable infrastructure present, which depleted the available healthcare budget) (Jakab and Manjieva, 2008). Although Kyrgyz citizens were nominally entitled to free healthcare, this was not economically viable, and so a system based on informal, unrecorded, 'out-of-pocket' payments developed (Belli, 2001; Jakab and Manjieva, 2008), with additional funding from taxation at the national, regional, city or municipal level. This system resulted in lower use of the medical services by poorer people, and particularly impacted upon rural areas, who were generally excluded from access to the more highly funded services at the central or regional level and who had more difficulty affording the additional payments (Abel-Smith and Falkingham, 1996; Jakab and Manjieva, 2008). In an attempt to solve this problem, healthcare reforms were commenced in 1997. These first focussed on improving efficiency, by allowing the private sector to provide healthcare, and through mandatory health insurance from payroll, pension and land taxation. These contributions were pooled and used to develop contracts with providers such as

hospitals. Following this, fragmentation within the provision of healthcare was addressed through the pooling of funds within regions, and in doing so allowed monitoring of outputs, and adjustment according to these. Attempts were also made to improve the efficiency of the healthcare system by developing primary healthcare provision (Jakab and Manjieva, 2008) and therefore discourage the overuse of hospitals for trivial complaints. It is hoped that these latter changes will improve the equity of the healthcare system. A recent study of use of alternative and complementary medicine (ACM) amongst countries of the former Soviet Union has suggested that ACM practitioners are significantly more commonly consulted in Kyrgyzstan than in other countries (Stickley *et al.*, 2013), possibly as a result in the costs of traditional healthcare.

A2.3 Livestock management in Kyrgyzstan

As alluded to earlier, there have been considerable changes in pastoralism and livestock management in Kyrgyzstan over the last 150 years. Historically, Kyrgyzstan was a country of seminomadic pastoralists who herded sheep on horseback (although yak herding was also practiced at higher altitudes): moving between the mountains in the summer and the lowlands and foothills in the winter, when climatic conditions made it unsuitable to remain at high altitudes (Schillborn-van Veen, 1995). The arrival of Russian settlers in the late 19th century resulted in a loss of lowland pastures from the Kyrgyz, and initiated the process of settlement amongst local people (Farrington, 2005). This process was accelerated under Soviet control, due to the attitude that the practices of nomadism and pastoralism were archaic, and the belief that the Kyrgyz pastoral lifestyle was in fact solely a 'transitional' stage between nomadism and settled agriculture (Kerven *et al.*, 1996). As a result, forced collectivisation was undertaken in the early 20th century, with management of land and grazing by local Soviet councils and the development of sovkhozes (state- owned farms) and kolkhozes (collective farms) (although some private ownership of a few animals was still permitted (Fitzherbert, 2000)). A number of strategies, including pasture irrigation and fertilisation, and structured movement of livestock, were employed to reduce overgrazing (Dörre and Borchardt, 2012). This was also combined with a move away

from traditional breeds of sheep towards 'improved' breeds, bred for wool production. However, these new breeds were less suited to the harsh climates and required both supplementation with fodder, hay or grain, and provision of winter shelter. Transhumance continued, although the new breeds of sheep were less able to move the large distances required, and so had to be moved by truck or train, and many animals did not survive the winter due to shearing in order to meet wool production targets (Farrington, 2005). Combined with this process of collectivisation was the construction of permanent settlements around the country, in the locations of the winter camps. Following World War II, production targets were rapidly raised and farm sizes were increased. By the 1960s, the numbers of animals exceeded the stocking capacity of the grassland (much of which had been lost in order to produce the fodder needed for the herds), leading to pasture degradation (Fitzherbert, 2000; Farrington, 2005). Despite this, this trend continued until the years just prior to independence.

Following independence in 1991, the agricultural sector of Kyrgyzstan was gradually privatised and animals and other assets were allocated amongst the collective members (with an increased reliance on natural resources due to economic uncertainty and loss of jobs), with pasture land kept in state control but leased to local councils (Schillborn-van Veen, 1995; Lerman and Sedik, 2009; Dörre and Borchardt, 2012). However, at the same time, state-operated services (such as veterinary services) were abruptly removed, as were the guaranteed markets for products which were offered under Soviet control. As many of the livestock owners had little experience of farming without state support, most animals were either sold or slaughtered for meat (with the exception of horses, which increased in numbers and were used as draft animals due to the lack of agricultural machinery), although the exact numbers of animals in this period is unknown (Kerven *et al.*, 1996). Those livestock farms remaining either remained in a form of collective ownership (with greatly reduced numbers), or as individually owned farms used for a mixture of subsistence and commercial management (Lerman and Sedik, 2009). Due to a decline in wool prices, there was a movement towards the use of meat breeds (Schillborn-van Veen, 1995; Farrington, 2005; Kerven, 2006). It has been reported that the smaller farms in the

south of the country have increased in size faster than those in the north over the last 15 years (Kerven, 2006). Currently, livestock in Kyrgyzstan comprise around 10% of GDP (World Bank, 2005).

Dogs have played a part in the Kyrgyz lifestyle for centuries, and a local breed of sighthound, the Taigan, is mentioned multiple times in the Kyrgyz epic 'Manas'. These dogs were used for coursing and game hunting, and as a deterrent to wolves (with an experienced Taigan being reportedly able to kill a wolf). However, dog control campaigns, the increasing availability of guns for hunting and cross breeding with introduced breeds of dog has had a considerable impact upon the numbers of purebred Taigans in the country, and they are currently rarely found (Dubinina, 2005). Despite this, dogs retain an important role in the rural Kyrgyz communities, from a cultural and a practical perspective: being commonly used for guarding livestock and possessions (van Kesteren *et al.*, 2013).

A3. Questionnaire forms used throughout the study

A3.1 Household questionnaire, May 2012

Questionnaire for households *(анкета для домашнего хозяйства)*

Date *(Дата)* _____ **Grid point/GPS location** *(Расположение)*_____

Village code:_____ *Код села*	**Name:**_____ *ФИО*
Household number/№_____ *Номер дома*	**Address:**_____ *Адреси*

General information *(Общие сведения)*

1. Age *(Возраст)*:_____

2. Sex *(Пол)*: **M** *(М)*☐ **F** *(Ж)*☐

3a. Number of people in household *Количество людей в семье:*_____

3b. Ages of other people in household (Приблизительный возраст других людей вашей семье):

_____

4. Nationality *(Национальность)*: **Kyrgyz** *(кыргыз)*☐ **Other** *(друга)* ☐

5. Head of household *(хозяин дома)*: **Yes** *(Да)*☐ **No** *(нет)*☐

6. How long have you lived here?_____ *как долго живете здесь?*

7. Occupation *(Профессия)*:_____

8. Do you travel to Summer pastures? *(Едете ли летом на пастбища?)*: **Yes** *(Да)*☐ **No** *(нет)*☐

9. What is your living standard? *Каково ваше материальное положение?* **Good** *(хорошее)*☐ **Middle** *(среднее)*☐ **Poor** *(плохое)*☐

Dog information *(Информация о собаках и кошках)*

10. Total number of dogs in household *(Общее количество собак)*:_____

11. If you have no dogs, why is this?:_____
Если у вас нету собаки, почему не держите

12. Approximately how many dogs have you owned in the last ten years?_____
Общее количество собак в вашем доме в течение последних 10 лет?

13. How many dogs did you acquire in the last 12 months?_____
Сколько собак Вы приобрели за последние 12 месяцев?

14. How many dogs did you lose in the last 12 months?_____
Сколько собак вы выгнали, отдали и т.д. за последние 12 месяцев?)

15. Do you ever use your dogs for hunting marmots, gophers, pika etc...? *(когда нибудь использовали вашу собаку на охоту суркам,сусликам, пищухам и другим?)* **Yes** *(Да)*☐ **No** *(нет)*☐

16. Are there unowned dogs in your community? **Yes** ☐ **No** ☐ **Don't know** ☐
Имеется в вашем районе бесхозные собаки *Да* *нет* *не знаю*

17. Approximately how many of these dogs are always in the community and how many are unknown 'strange' dogs? Persistent *(устойчиво)*____ Unknown *(неизвестно)*____
Примерно сколько собак в вашей деревне и сколько из них бесхозных

Livestock information *(Информация о домашних животных)*

18. How many animals do you own? *(Сколько имеете голов животных)?*
Sheep *(Овцы)*____ **Goats** *(козы)*____ **Cattle** *(коровы)*____ **Horses** *(лошади)*____ **Other** *(другое)*_____

19. Where are the animals kept (approximate distance and location description)?
Где содержатся животные (примерное расстояние и расположение)_____

20. Do you slaughter your own animals? *Yes* ☐ *No* ☐
Режете дома животных? *Да* *нет*

21. Do you slaughter other people's animals? *Yes* ☐ *No* ☐
Режете ли вы животных у других людей *Да* *нет*

22. What do you with unhealthy organs of these animals? (Что делаете с пораженными органами животных?
Bury *(закапываем)*☐ **Give to dogs** *(даем собакам)*☐ **Burn** *(сжигаем)*☐
Throw away *(выбрасываем)*☐ **Other** *(другое)*☐_____

A1

A3.2 Dog questionnaire, May 2012

Questionnaire for individual dogs *(Анкета для собак)*

Date *(Дата)* _____ **Grid point/GPS location** *(Расположение)*_____

Village code:_____ *Код села* **Household number/№**_____ *Номер дома*	**Name:**_____ *ФИО* **Address:**_____ *Адреси*

Dog data *(Данные собаки)*

1. Dog No. (in household):_____
Номер собаки

2. Dog ID (photo☐/microchip☐):_____
ИД номер собаки (фотография/микрочипа)

3. Dog name *(Кличка)*: _____

4. Dog description: _____
Описание собаки

5. Age of dog *(Возраст)*:_____yrs

Exact☐ **estimate**☐ **unknown**☐
точно около неизвестно

6. Sex: **male** ☐ **female** ☐
Пол самец самка

7. Neutered: **No** ☐ **Yes** ☐
Кастрирован нет да

8. Breed *(Порода)*:_____

9. Size of dog: small☐ **medium**☐ **large**☐
Габарит собаки: малый средний большой

10. Approximate weight of dog:_____kg
примерный вес собаки

11. If entire female is she pregnant?
если самка, беременная?
Yes *(да)*☐ **No** *(нет)*☐

12. Reproductive data (if female)
(Репродуктивные данные если собака сука):
a. Number of litters in the last year
(сколько раз у вашей суки были беременности в течение последних 12 месяцев?) _____
b. Total number of puppies born: _____
Сколько родились щенки?
c. Number survived: _____
Сколько из них выжили?
d. Number kept by household: _____
Сколько щенят оставили себе?
e. Number given away: _____
Сколько щенят отдали?
f. Number of litters produced over lifetime: _____
сколько раз может быть у суки беременности за всю жизнь?

13. Source of dog *(Происхождение собаки)*
a.Offspring of own bitch ☐
от собственной собаки
b. Sourced/traded with neighbours ☐
купил или приобрел от соседей
c. Bought/traded from outside neighbourhood ☐
купил или приобрел в базаре или другого села

14. Type of dog*(тип собаки)*: pet*(домашний)*☐ guard dog(*сторожевой*)☐
sheep dog(*отарный*)☐ hunting dog(*охотничий*)☐ other(*другое*)_____

15. If owners travel to Summer pasture*(если хозяин идет в пастбища)*

a. Did dog travel to Summer pasture last year? No ☐ Yes ☐
последний год, брали вашу собаку собой в пастбищу нет да

b. Will dog travel to Summer pasture this year? No☐ Yes ☐
возьмете вашу собаку в этом году собой в пастбищу нет да

16. Dog is fed by : People in household☐ Neighbours☐ Forage for themselves☐
кто кормит вашу собаку члены семьи соседи собака сама находит

B1

17. Dogs are fed: *питание*	never/very rarely *никогда/очень редко*	monthly *ежемесячно*	weekly *недельный*	almost every day *каждый день*
a. Uncooked meat *(сырое мясо и потроха)*	☐	☐	☐	☐
b. Offal *(мусор)*	☐	☐	☐	☐
c. Table scraps *(остатками со стола)*	☐	☐	☐	☐
d. Commercial feed *(специальный корм,для собак)*	☐	☐	☐	☐
e. Seen eating rodents *Питается ваша собака с грызунами?*	☐	☐	☐	☐

18. How is the dog kept? *(Как держите вашу собаку?)*
chained at night ☐ *ночью в цепи* chained during day ☐ *днем в цепи* always chained ☐ *всегда в цепи* never chained ☐ *никогда в цепи*

19a. % of time dog is indoors *(% времени собака в помещении конуре):* _____
19b. % of time dog is outside leashed *(% времени собака в поводке):* _____
19c. % of time dog is outside unleashed *(% Времени собаки без поводка):* _____

20. Persons who handle or play with dog *(кто отвествен для собаки дома или кто играет с собакой):*
Adults of HH☐ *Взрослые* **Children of HH**☐ *Дети* **Friends&neighbours**☐ *Друзья и соседи* **Strangers**☐ *чужие* **Nobody**☐ *Никто*

21. Has the dog had any worming treatment in the last 6 months? **No** ☐ *нет* **Yes** ☐ *да*
Лечили собаку последние 6 месяцев?

22. If yes, by whom? **Vet**☐ **People in HH**☐ **Government worker**☐ **Other**_____
Если да, то кто делал? ветеринара члены семьи правительство работника другое

23. What drug(s) were used? *(какой препарат (ы) был/были использованы)*_____

24. How many tablets were given? *(Сколько таблеток давали)*_____

25. Approximately when was the dog last dosed _____
Примерно, когда была последняя дегельминтизация

26. Has dog been vaccinated against *(собака когда нибудь вакцинирована против):*
Rabies☐ *Бешенство* Distemper☐ *чума собак* Canine hepatitis☐ *гепатит собак* Leptospirosis☐ *Лептоспироз*

27. If dog has been vacc against rabies, date of last vaccination and vaccine type
Если собака была привита против бешенства укажите дату вакцинации и типа вакцины:

28. Dog visible outside house *(собака видна снаружи дома):*
No *(нет)* ☐ **In garden** *(в саду)* ☐ **Outside garden** *(за пределами сада)* ☐

Faecal sample taken? *взята кот проба?* **No**☐ *нет* **Yes**☐ *да*

PBS ☐ **Alcohol** *(спирт)* ☐

If not collected, why not?
*если проба не взята почему*_____

Sample ID *(ид номер пробы)* _____

Given arecoline? *(давали ареколин)* **No**☐ *нет* **Yes**☐ *да*

Purged? *(понос был)* **No** *(нет)*☐ **Yes** *(да)*☐

Frozen *(заморозен)* ☐

Alcohol *(спирт)* ☐

Adverse reaction to arecoline? ☐
различные э ффекты к ареколину?

Dosed with praziquantel: No☐ **Yes**☐
давали празиквантел нет да

Date of dosing:_____**20**____
Дата дачи таблетки

Dose min 1 day after arecoline if purged
минимум 1день после ареколина если очищен

B2

A3.3 Healthcare/economic questionnaire, May 2012

Healthcare/economic questionnaire ***(вопросник здоровья и экономики)***

Date *(Дата)* _____ **Grid point/GPS location** *(Расположение)*_____

Village code:_____ *Код села*	**Name:**_____ *ФИО*
Household number/№_____ *Номер дома*	**Address:**_____ *Адреси*

GENERAL INFORMATION *(Общие сведения)*

1. Age *(Возраст)*:_____

2. Sex *(Пол)*: **M** *(М)*☐ **F** *(Ж)*☐

3a. Number of people in household *Количество людей в семье:*_____

3b. Ages of other people in household (Приблизительный возраст других людей вашей семье):

_____

4. Nationality *(Национальность)*: **Kyrgyz** *(кыргыз)*☐ **Other** *(друга)* ☐

5. Head of household *(хозяин дома)*: **Yes** *(Да)*☐ **No** *(нет)*☐

6a. How long have you lived here?_____ *как долго живете здесь?*

6b. If you have lived elsewhere, where was the last place you lived?_____ *если вы раньше другом месте жили, где и как долго были там?*

7. Occupation *(Профессия)*:_____

8. Do you travel to Summer pastures? *(Едете ли летом на пастбища?)*: **Yes** *(Да)*☐ **No** *(нет)*☐

9. What is your living standard? *Каково ваше материальное положение?*
Good *(хорошее)*☐ **Middle** *(среднее)*☐ **Poor** *(плохое)*☐

LIVESTOCK INFORMATION ***(Информация о домашних животных)***

10. How many animals do you own? *(Сколько имеете голов животных)?*
Sheep *(Овцы)*____ **Goats** *(козы)*____ **Cattle** *(коровы)*____ **Horses** *(лошади)*____ **Other** *(другое)*_____

11. Approximately how much do you sell the following products for? (Insert 'x' if do not sell product, or leave blank if do not own animal):

	Sheep *(Овцы)*	**Goats** *(козы)*	**Cattle** *(коровы)*
a.Live animal			
b.Carcass (with contents) *туша с внутренними органами*			
c.Carcass (without contents) *туша без внутренних органов*			
c.Liver			
d.Lungs			
e.Skin/wool			

12a. What do you with livers and lungs containing cysts?
Bury *(закапываем)*☐ **Give to dogs** *(даем собакам)*☐ **Burn** *(сжигаем)*☐
Throw away *(выбрасываем)*☐ **Sell** *(продам)*☐ **Other** *(другое)*☐_____

12b. If you sell, approximately how much do you sell them for?

	Sheep *(Овцы)*	**Goats** *(козы)*	**Cattle** *(коровы)*
Liver			
Lungs			

13a. What do you do with meat containing cysts?
Bury *(закапываем)*☐ **Give to dogs** *(даем собакам)*☐ **Burn** *(сжигаем)*☐
Throw away *(выбрасываем)*☐ **Sell** *(продам)*☐ **Other** *(другое)*☐_____

13b. If you sell, approximately how much do you sell it for?

Sheep *(Овцы)*	**Goats** *(козы)*	**Cattle** *(коровы)*

C1

14. What are your main disease concerns for your livestock?
Как думаете, какие болезни основные для скота?

Sheep *(Овцы)*	**Goats** *(козы)*	**Cattle** *(коровы)*	**Horses** *(лошади)*

15a. Have you ever called out a veterinarian to your livestock? Yes ☐ No ☐
когда нибудь звали ветеринара посмотреть ваших животных Да нет

15b. If so, for which animals and why? *(Если да для которых животных и почему?):*

_____

DOG INFORMATION *(Информация о собаках и кошках)*

16. Total number of dogs in household (Общее количество собак): _____

17. If you have no dogs, why is this?: _____
Если у вас нету собаки, почему не держите?

18. Approximately how many dogs have you owned in the last ten years? _____
Общее количество собак в вашем доме в течение последних 10 лет?

19. How many dogs did you acquire in the last 12 months? _____
Сколько собак Вы приобрели за последние 12 месяцев?

20a. How many dogs did you turn out, give away etc in the last 12 months?
Сколько собак вы выгнали, отдали и т.д. за последние 12 месяцев?

	Adults *(Взрослые)*	**Puppies** *(щенки)*
a.Given away *(отданный)*		
b.Sold/exchanged *(Продано/обмен)*		
c.Lost *(потерянно)*		
d.Killed by person *(Убит людьи)*		
e.Died from disease *(Умер от болезни)*		
f.Died from old age *(Умер от старости)*		

20b. If any dogs died from disease, can you name what diseases were suspected and how this was confirmed? _____
Если какой нибудь собака умерла от болезни, можете ли вы назвать, имя от какой болезни и как это было подтверждено?

21. Are there unowned dogs in your community? **Yes** ☐ **No** ☐ **Don't know** ☐
Имеется в вашем районе бесхозные собаки Да нет не знаю

22. Approximately how many of these dogs are always in the community and how many are unknown 'strange' dogs? Persistent *(устойчиво)* ____ **Unknown** *(неизвестно)* ____
Примерно сколько собак в вашей деревне и сколько из них бесхозных

23. Have members of your family been bitten by dogs in the last 12 months?
У членов вашей семьи были ли укушены собаками в течение последних 12 месяцев?

a.Your dogs *(своя собака)*☐ **b.Other owned dogs** *(соседская собака)*☐
c.Community dogs* *(деревенская собака)*☐ **d.Strange dogs*** *(уличная собака)*☐
e.No *(нет)*☐

*** 'community dogs' are recognised and known within the village, whereas 'strange dogs' are generally unknown.** *(собаки сел известны в деревне, в то время как "уличные собаки, как правило, неизвестны).*

If you own/have owned dogs:

24. What are/were your main disease concerns for your dog(s)? *(Какие основные болезни вашей собаки?)* _____

25a. Have you ever taken your dog(s) to a veterinarian? *Вы когда-нибудь были у ветеринара насчет вашей собаки?* **Yes** *(Да)*☐ **No** *(нет)*☐

25b. If so, why? _____

C2

HUMAN HEALTH ISSUES

26. What do you perceive to be the main health concerns for you and your family?
Как вы считаете, какие основные проблемы со здоровьем у вас и вашей семьи?

27a. Where do you/your family most commonly go for medical treatment if you feel unwell? *Где вы или ваша семья может пойти на лечение в основном, если вы плохо себя чувствуете?*

doctor's surgery ☐ **health centre** ☐ **hospital** ☐ **other** ☐_____
хирургия врача медицинского центра больницы *другое*

27b. Where is this? _____

27c. How often have you been to this place in the last 12 months? _____
Как часто Вы были в больнице или у врача в течение последних 12 месяцев?

28. How many days were you sick in the last:
Сколько дней Вы были больными последний:
Year *(год)*?_____ **Month** *(месяц)*?_____ **Week** *(неделю)*?_____

29. Have you heard of hydatid disease? **Yes** ☐ **No** ☐
Знаете ли Вы о болезни эхинококкоза? *Да* *нет*

30. How do you become ill with hydatid disease? (Can tick more than one)
Как люди могут заразится этой болезнью? (можно несколько ответов)

a.Contact with dogs *(Контакт с собаками)*	
b.Contact with cats *(контакт с кошками)*	
c.Contact with sick people *(контакт с больными людьми)*	
d.Contact with livestock *(контакт с животными)*	
e.From food/drink *(от еды, воды)*	
f.Other *(другой)*	
g.Don't know *(не знаю)*	

31a. Have you or your family members ever had hydatid disease?
Диагностировали у Вас или членов вашей семьи на эхинококкоза?
No *(Нет)*☐ **Yourself** *(сам)*☐ **Spouse** *(жубайым)*☐ **Children** *(Дети)*☐ **Other** *(другой)*☐_____

31b. If so, how was this confirmed? *Если это так, как это подтверждено?*
_____

31c. What type of hydatid disease was it? *Знаете ли вы, типы эхинококкоза?*
CE ☐ **AE** ☐ **Don't know** *(не знаю)* ☐

32. Have you heard of brucellosis? **Yes** ☐ **No** ☐
Вы когда-нибудь слышали о бруцеллезе? *Да* *нет*

33. How do you become ill with brucellosis? (Can tick more than one)
Как люди могут заразится этой болезнью? *(можно несколько ответов)*

a.Contact with dogs *(Контакт с собаками)*	
b.Contact with cats *(контакт с кошками)*	
c.Contact with sick people *(контакт с больными людьми)*	
d.Contact with livestock *(контакт с животными)*	
e.From food/drink *(от еды, воды)*	
f.Other *(другой)*	
g.Don't know *(не знаю)*	

34a. Have you or your family members ever had brucellosis?
были ли у Вас или членов Вашей семьи бруцеллез?
No *(Нет)*☐ **Yourself** *(сам)*☐ **Spouse** *(жубайым)*☐ **Children** *(Дети)*☐ **Other** *(другой)*☐_____

C3

34b. If so, how was this confirmed? *Если это так, как это подтверждено?*

_____

35. Have you heard of anthrax? **Yes** ☐ **No** ☐
Вы когда-нибудь слышали о сибирской язве? Да нет

36. How do you become ill with anthrax? (Can tick more than one)
Как люди могут заразится этой болезнью? *(можно несколько ответов)*

a.Contact with dogs *(Контакт с собаками)*	
b.Contact with cats *(контакт с кошками)*	
c.Contact with sick people *(контакт с больными людьми)*	
d.Contact with livestock *(контакт с животными)*	
e.From food/drink *(от еды, воды)*	
f.Other *(другой)*	
g.Don't know *(не знаю)*	

37a. Have you or your family members ever had anthrax?
были и у Вас или членов Вашей семьи сибирской язвы?
No *(Нет)*☐ **Yourself** *(сам)*☐ **Spouse** *(жубайым)*☐ **Children** *(Дети)*☐ **Other** *(другой)*☐_____

37b. If so, how was this confirmed? *Если это так, как это подтверждено?*

_____

38. Have you heard of rabies? **Yes** ☐ **No** ☐
когда-нибудь слышали об бешенстве? *Да нет*

39. How do you become ill with rabies? (Can tick more than one)
Как люди могут заразится этой болезнью? *(можно несколько ответов)*

a.Contact with dogs *(Контакт с собаками)*	
b.Contact with cats *(контакт с кошками)*	
c.Contact with sick people *(контакт с больными людьми)*	
d.Contact with livestock *(контакт с животными)*	
e.From food/drink *(от еды, воды)*	
f.Other *(другой)*	
g.Don't know *(не знаю)*	

40a. Have there been any cases of rabies in people in your village in your living memory? *(Были ли случаи бешенства у людей в вашей деревне в вашей памяти?)*
Yes *(Да)*☐ **No** *(нет)*☐

40b. If so, how was this confirmed? *Если это так, как это подтверждено?*

_____

41. Have you heard of toxoplasmosis? **Yes** ☐ **No** ☐
слышали когда нибудь об токсоплазме? Да нет

42. How do you become ill with toxoplasmosis? (Can tick more than one)
как инфицируются с токсоплазмой? (можно несколько ответов)

a.Contact with dogs *(Контакт с собаками)*	
b.Contact with cats *(контакт с кошками)*	
c.Contact with sick people *(контакт с больными людьми)*	
d.Contact with livestock *(контакт с животными)*	
e.From food/drink *(от еды, воды)*	
f.Other *(другой)*	
g.Don't know *(не знаю)*	

43a. Have you or your family members ever had toxoplasmosis?
вы или кто-то другой член вашей семьи болел когда нибудь токсоплазмой?
No *(Нет)*☐ **Yourself** *(сам)*☐ **Spouse** *(жубайым)*☐ **Children** *(Дети)*☐ **Other** *(другой)*☐_____

43b. If so, how was this confirmed? *(Если это так, как это подтверждено?)*

_____

C4

A3.4 Questionnaire, September 2012 and April 2013

Questionnaire for dogs (Иттер үчүн анкета/Анкета для собак) B1-0912

Date (Датасы/Дата) /2012	Coordinates (Координаттары/Расположение)
ID code (Айылдын коду / ИД Код): _____	Name (Аты-жөнү/ФИО): _____
Ultrasound scanned: Yes ☐ No ☐ Узиге текшерилдиңизби: ооба жок прошли узи обследование: да нет	Address (Адреси/Адрес): SM☐ SM(W)☐ TS☐ KS☐ KK☐ с-м с-м(запад) т-с к'с к-к
1. Number of dogs (Иттин номери/Номер собаки): ___	2. Name (Иттин аты/Кличка): _____
3. Age (Иттин жашы/Возраст): _____	4. Weight (Иттин салмагы/Вес): _____
5. Sex (Иттин жынысы/Пол): Male ☐ Female ☐ Эркек/Самец канчык/Самка	6. Visited pasture (Жайлоого чыктыбы / посетил пастбищу): Yes (ооба/да) ☐ No (жок/нет) ☐
7. Pregnancy status: Pregnant☐ Lactating☐ Бооздугунун абалы Бооз эмизүүдө Положение беременности беременная кормящая	8. Dog present (ит барбы/собака присуствует): No (жок/нет) ☐ On arrival (келгенде/по прибытии) ☐ When called (эгер чакырса/если позвали) ☐
9. Last dosed (акыркы жолу качан дегельминтизация жүргүздүңүз/когда была последняя дегельминтизация)? Never (эч качан/никогда) ☐ Last visit (May) (акыркы келгенде (Бугу)/последний визит (май)) ☐ Other (башка/другой)	
11. Faecal sample (иттин заңы/кал собаки): No (жок/нет) ☐ Floor (сырттан/снаружи) ☐ Rectal (көтөнөн/с прямой кишки) ☐	12. PZQ given: Supervised☐ Unsupervised☐ празиквантел берилгенби: байкап турулган байкоосуз давали празиквантел? под наблюдением без наблюдением
13. Marked (белгиленгенби/отмеченный): Yes (ооба/да) ☐ No (жок/нет) ☐	14. GPS collared (мойн жип GPS менен/ошейник с GPS): No (жок/нет) ☐ Collar number (номери/Номер) _____
15. Other notes (башка маалыматтар/ другие данные):	

A3.5 Questionnaire, September 2013

Questionnaire for dogs (Иттер үчүн анкета/Анкета для собак) B1-0913

Date (Датасы/Дата) /2013 Team: F A	Coordinates (Координаттары/Расположение)
ID code (Айылдын коду / ИД Код): _____	Name (Аты-жөнү/ФИО): _____
Ultrasound scanned: Yes ☐ No ☐ Узиге текшерилдиңизби: ооба жок прошли узи обследование: да нет	Address (Адреси/Адрес): SM☐ SM(W)☐ TS☐ KS☐ KK☐ с-м с-м(запад) т-с к'с к-к
1. Number of dogs lost in last 12 months: _____ (12 айдын ичинде канча ит жоготтуңуз/Количество собак потерявшихся за последние 12 месяцев)	2. Killed(hunter)__ Killed(vet)__ Died__ Missing__ Мергенчи атып кетти Мал доктур өлтүрдү Өлдү Жоголуп кетти Убит охотником Убит ветеринаром Умер Отсутствует
3. Number of dogs (Иттин номери/Номер собаки): ___	4. Name (Иттин аты/Кличка): _____
5. Age (Иттин жашы/Возраст): _____	6. Weight (Иттин салмагы/Вес): _____
7. Sex (Иттин жынысы/Пол): Male ☐ Female ☐ Эркек/Самец канчык/Самка	8. Visited pasture (Жайлоого чыктыбы / посетил пастбищу): Yes (ооба/да) ☐ No (жок/нет) ☐
9. Last dosed (акыркы жолу качан дегельминтизация жүргүздүңүз/когда была последняя дегельминтизация)? Never (эч качан/никогда) ☐ Last October (Тогуздун айы/октябрь) ☐ Other (башка/другой) _____ Number of doses in last 12 months ____ (12 айдын ичинде иттиниизге канча жолу дары берилген?/Сколько раз было дегельминтизировано ваша собака за последние 12 месяцев?)	
10. Where was tablet obtained from? (Дарыны кимден алдыңыз?/От кого брали таблетки?) Vet (Мал доктурдан/От ветеринара) ☐ Dispensary (Дарыканадан/В аптеке) ☐ Researcher (Окумуштуулардан/От ученных) ☐ Other (башка/другой)	11. Who administered tablet? (Дарыны ким берди?/Кто давал таблетки?) Yourself (Өзүм/Сам) ☐ Other family (Үй-бүлөдөн бирөө/Другие из семьи) ☐ Vet (Мал доктурдан /От ветеринара) ☐ Other (башка/другой)
11. Faecal sample (иттин заңы/кал собаки): No (жок/нет) ☐ Floor (сырттан/снаружи) ☐ Rectal (көтөнөн/с прямой кишки) ☐	15. Other notes (башка маалыматтар/другие данные):

A3.6 Questionnaire, April 2014

Questionnaire for dogs (Иттер үчүн анкета/Анкета для собак) B1-0414

Date (Датасы/Дата) _____/2014 Team: F A	Coordinates (Координаттары/Расположение)_____
ID code (Айылдын коду / ИД Код): _____	Name (Аты-жөнү/ФИО):
Address (Адреси/Адрес): SM(с-м)☐ SM(W) (с-м запад)☐ TS(т-с)☐ KS(к'с)☐ KK(к-к)☐	
A. Have you heard of hydatid disease? Эхинококкоз оорусун билесизби Знаете ли Вы о болезни эхинококкоза? Yes (оба/да)☐ No (жок/нет) ☐	B. How do people become ill with hydatid disease? Адамдарга эхинококкоз оорусу кандай жугаарын билесизби? Как люди могут заразится этой болезнью? Don't know (билбейм/не знаю)☐
1. Number of dogs lost in last 6 months: _____ (6 айдын ичинде канча ит жоготтуңуз/Количество собак потерявшихся за последние 6 месяцев)	2. Killed(hunter)__ Killed(vet)__ Died__ Missing__ Мергенчи атып кетти Мал доктур өлтүрдү Өлдү Жоголуп кетти Убит охотником Убит ветеринаром Умер Отсутствует
3. Number of dogs (Иттин номери/Номер собаки):	4. Name (Иттин аты/Кличка):
5. Age (Иттин жашы/Возраст): _____	6. Weight (Иттин салмагы/Вес): _____
7. Sex (Иттин жынысы/Пол): Male ☐ Female ☐ Эркек/Самец канчык/Самка	8. Visited pasture (Жайлоого чыктыбы / посетил пастбищу): Yes (оба/да) ☐ No (жок/нет) ☐
9. Last dosed: _____ акыркы жолу качан дегельминтизация жүргүздүңүз когда была последняя дегельминтизация Never dosed (эч качан/никогда) ☐	10. Number of doses in last 12 months_____ 12 айдын ичинде иттинизге канча жолу дары берилген? Сколька раз было дегельминтизировано ваша собака за последние 12 месяцев?
11. Where was tablet obtained from? (Дарыны кимден алдыңыз?/От кого брали таблетки?) Vet (Мал доктурдан/От ветеринара)☐ Dispensary (Дарыканадан/В аптеке)☐ Researcher (Окумуштуулардан/От ученных) ☐ Other (башка/другой)_____	12. Who administered tablet? (Дарыны ким берди?/Кто давал таблетки?) Yourself (Өзүм/Сам)☐ Other family (Үй-бүлөдөн бирөө/Другие из семьи) ☐ Vet (Мал доктурдан /От ветеринара)☐ Other (башка/другой)_____
13. Faecal sample (иттин заңы/кал собаки): No (жок/нет) ☐ Floor (сырттан/снаружи) ☐ Rectal (көтөндөн/с прямой кишки) ☐	14. Other notes (башка маалыматтар/другие данные):

A3.7 Questionnaire, September 2014

Questionnaire for dogs (Иттер үчүн анкета/Анкета для собак) B1-0914

Date (Датасы/Дата) _____/2014 Team: F A	Coordinates (Координаттары/Расположение)_____
ID code (Айылдын коду / ИД Код): _____	Name (Аты-жөнү/ФИО):
Address (Адреси/Адрес): SM(с-м)☐ SM(W) (с-м запад)☐ TS(т-с)☐ KS(к'с)☐ KK(к-к)☐	
1. Number of dogs (Иттин номери/Номер собаки):	2. Names (Иттин аты/Кличка):
3. Age (Иттин жашы/Возраст):	4. Weight (Иттин салмагы/Вес):
5. Sex (Иттин жынысы/Пол): Male ☐ Female ☐ Эркек/Самец канчык/Самка	6. Visited pasture (Жайлоого чыктыбы / посетил пастбищу): Yes (оба/да) ☐ No (жок/нет) ☐
7. Age of dog when obtained (возраст собаки, когда получены): Puppy (күчүк/щенок)☐ Adult (чоңдор/взрослая)☐	
8. Last dosed: _____ акыркы жолу качан дегельминтизация жүргүздүңүз когда была последняя дегельминтизация Never dosed (эч качан/никогда) ☐	9. Number of doses in last 12 months_____ 12 айдын ичинде иттинизге канча жолу дары берилген? Сколька раз было дегельминтизировано ваша собака за последние 12 месяцев?
10. Where was tablet obtained from? (Дарыны кимден алдыңыз?/От кого брали таблетки?) Vet (Мал доктурдан/От ветеринара)☐ Dispensary (Дарыканадан/В аптеке)☐ Researcher (Окумуштуулардан/От ученных) ☐ Other (башка/другой)_____	11. Who administered tablet? (Дарыны ким берди?/Кто давал таблетки?) Yourself (Өзүм/Сам)☐ Other family (Үй-бүлөдөн бирөө/Другие из семьи) ☐ Vet (Мал доктурдан /От ветеринара)☐ Other (башка/другой)_____
12. Date of last rabies vaccination (Последнее прививки против бешенства): _____ Never (эч качан/никогда)☐	
13. Faecal sample (иттин заңы/кал собаки): No (жок/нет) ☐ Floor (сырттан/снаружи) ☐ Rectal (көтөндөн/с прямой кишки) ☐	14. Faecal sample condition Dry ☐ Medium ☐ Fresh ☐
15. Faecal sample for diet analysis: ☐ Dry ☐ Medium ☐ Fresh ☐	16. Other notes (башка маалыматтар/другие данные):

A4. R code for Bayesian mixture model (using Xinjiang data)

```
require(rjags)
require(coda)
source("plotPost.R")

modelstring = "
data{
for(i in 1:N.neg) {
  zerosneg[i] <- 0}
for(i in 1:N.field) {
  zeros[i] <- 0
  ones.fieldpos[i] <- 1}
}

model{

##########'Known' negative samples (excluding outliers according to
z distribution)
for(i in 1:N.neg) {
      zerosneg[i] ~ dpois(negpanel[i])
      negpanel[i] <- -retain[i]*log(neglike[i])+C
      neglike[i] <- sqrt(tau.neg/(2*3.14159))*
               exp(-(tau.neg/2)*pow((y.neg[i]-mu.neg),2))
      y.neg[i] ~ dnorm(mu.negdata,tau.negdata) #training data (neg)
      stres[i]<-(y.neg[i]-mu.negdata)/sigma.negdata #stand. res.
      outlier[i]<-step(stres[i]-zlevel) #identifying outliers
      retain[i]<-(1-outlier[i]) #selecting which obs to retain
}

##########Linear regression of log burden on OD
for( i in 1 : N.pos ) {
    zlogburden[i] ~ dnorm( muod[i] , tauod )
    muod[i] <- zbetaod0 + zbetaod1 * zod[i]
}

betaod1<-zbetaod1*logburdensd  / odsd
betaod0<-zbetaod0*logburdensd+logburdenm-
zbetaod1*logburdensd*odm/odsd

##########Combining field distributions (using 'zeros trick' to
model combined distribution)
for(i in 1:N.field) {
      z[i] ~ dbern(p)
      zeros[i] ~ dpois(phi[i])
      phi[i] <- -z[i]*log(like.fieldpos[i])-(1-
      z[i])*log(like.fieldneg[i])+C

##########Distribution of field positives (Polya tree)
      z.fieldpos[i] <- (y.field[i]-mu.fieldpos)/sigma.fieldpos
      ind.fieldpos[i] <- trunc(sets.pos*phi(z.fieldpos[i])+1)
      like.fieldpos[i] <- coi*exp(-0.5*z.fieldpos[i]*z.fieldpos[i])*
                polprop.fieldpos[ind.fieldpos[i]]/sigma.fieldpos
```

```
      cpoinv.fieldpos[i] <- 1/(like.fieldpos[i]*sets.pos)
      l.fieldpos[i] <- like.fieldpos[i]*sets.pos
      ones.fieldpos[i]~dbern(like.fieldpos[i])

##########Distribtion of field negatives (Gaussian)
      like.fieldneg[i]<- sqrt(tau.neg/(2*3.14159))*
                  exp(-(tau.neg/2)*pow((y.field[i]-mu.neg),2))

##########Estimating probability of positivity, burden and score
      probsamppos[i] <-
            p*l.fieldpos[i]/(p*l.fieldpos[i]+(1-p)*like.fieldneg[i])
      burdensamp[i]<-(y.field[i]*betaod1)+betaod0
      scoresamp[i]<-probsamppos[i]*burdensamp[i]

}

#####Fitting Polya tree to field data
x.fieldpos[1,1] <- 0.5; x.fieldpos[1,2] <- 0.5

for(k in 1:j.int[Jt]){
      m.fieldpos[1,k] <- log(x.fieldpos[1,1])
      m.fieldpos[1,j.int[Jt]+k] <- log(x.fieldpos[1,2])
}

for(i in 2:Jt){
par.lev.fieldpos[i] <- c.fieldpos*i*i
      for(j in 1:j.int[i]){
x.fieldpos[i,2*j-1]~dbeta(par.lev.fieldpos[i],par.lev.fieldpos[i]);
x.fieldpos[i,2*j] <- 1-x.fieldpos[i,2*j-1]
            for(k in 1:j.int[Jt-i+1]){
m.fieldpos[i,(2*j-2)*j.int[Jt-i+1]+k] <- log(x.fieldpos[i,2*j-1])
m.fieldpos[i,(2*j-1)*j.int[Jt-i+1]+k] <- log(x.fieldpos[i,2*j])
            }
      }
}

for(i in 1:sets.pos){
      polprop.fieldpos[i] <- exp(sum(m.fieldpos[1:Jt,i]))
}

##########Estimating distributions over grid of OD values
for(i in 1:100){
      grid[i]<-i*0.01

      z.fieldf[i] <- (grid[i]-mu.fieldpos)/sigma.fieldpos
      ind.fieldf[i] <- trunc(sets.pos*phi(z.fieldf[i])+1)
      f.fieldpos[i]<- sets.pos*coi*exp(-0.5*z.fieldf[i]*z.fieldf[i])
               *polprop.fieldpos[ind.fieldf[i]]/sigma.fieldpos

      f.neg[i] <- sqrt(tau.neg/(2*3.14159))*exp(-(tau.neg/2)
               *pow((grid[i]-mu.neg),2))

      cdffieldpos[i]<-sum(f.fieldpos[1:i])
      cdfneg[i]<-sum(f.neg[1:i])
```

```
      probpos[i] <- p*f.fieldpos[i]/(p*f.fieldpos[i]+(1-p)*f.neg[i])
      burden[i]<-(grid[i]*betaod1)+betaod0
      score[i]<-probpos[i]*burden[i]

      se[i]<-1-(cdffieldpos[i]/totalpos)
      sp[i]<-cdfneg[i]/totalneg
      fpr[i]<-1-sp[i]
}

totalpos<-sum(f.fieldpos[1:100])
totalneg<-sum(f.neg[1:100])


#####Priors
#positives
mu.fieldpos~dunif(0,1)
sigma.fieldpos~dunif(0,1)
c.fieldpos~dgamma(5,1)

#negatives
mu.neg~dunif(0,1)
sigma.neg~dunif(0,1)
mu.negdata~dunif(0,1)
sigma.negdata~dunif(0,1)
zlevel~dunif(0,4)
sigma2.negdata<-pow(sigma.negdata,2)
tau.negdata<-1/sigma2.negdata
sigma2.neg<-pow(sigma.neg,2)
tau.neg<-1/sigma2.neg

#combination
p~dunif(0,1)
coi <- 0.3989422804
C<-10000

#linear regression
zbetaod0 ~ dnorm( 0 , 1.0E-12 )
zbetaod1 ~ dnorm( 0 , 1.0E-12 )
tauod ~ dgamma( 0.001 , 0.001 )
sigmaod<-1/sqrt(tauod)

}
"
writeLines(modelstring,con="model.txt")

#####Data
#main data
y.pos=c(0.0877,0.2524,0.31955,0.7928,0.6646,0.171,0.11685,0.23965,0.
57075,0.46145,0.17555,0.39625,0.1546,0.462,0.68,0.37345)
y.neg=c(0.05285,0.0846,0.27815,0.10385,0.05675,0.05475,0.1116,0.2055
,0.0728,0.25945,0.11585,0.08645,0.06905,0.0394,0.0753,0.2478,0.05385
,0.16555,0.04705,0.05725,0.10625,0.0475)
```

```
y.field=c(0.05285,0.0846,0.27815,0.10385,0.05675,0.05475,0.1116,0.20
55,0.0728,0.25945,0.11585,0.08645,0.06905,0.0394,0.0753,0.2478,0.053
85,0.16555,0.04705,0.05725,0.10625,0.0475,0.0877,0.2524,0.31955,0.79
28,0.6646,0.171,0.11685,0.23965,0.57075,0.46145,0.17555,0.39625,0.15
46,0.462,0.68,0.37345)

logburden=c(1.098612289,3.912023005,NA,9.210340372,9.210340372,3.912
023005,0.693147181,4.605170186,8.517193191,4.605170186,2.995732274,5
.703782475,2.302585093,6.214608098,9.210340372,4.605170186)

#polya tree parameters
Jt=2
sets.pos=2^Jt
j.int=c(1,2)

#data processing
N.pos=length(y.pos)
N.neg=length(y.neg)
N.field=length(y.field)
odm = mean( y.pos ) ; odsd = sd( y.pos )
logburdenm = mean( logburden , na.rm=TRUE) ; logburdensd = sd(
logburden , na.rm=TRUE)
zod = ( y.pos - odm ) / odsd
zlogburden = ( logburden - logburdenm ) / logburdensd

#####Load data
datalist = list(N.pos=N.pos, N.neg=N.neg, N.field=N.field,
                y.neg=y.neg, y.field=y.field,
                sets.pos=sets.pos, Jt=Jt, j.int=j.int,
                zod=zod, zlogburden=zlogburden,
                odm=odm, odsd=odsd,
                 logburdensd=logburdensd, logburdenm=logburdenm)

#####Inits
initslist = list(
     mu.fieldpos=mean(y.pos),
     sigma.fieldpos=sqrt(var(y.pos)) ,
     mu.neg=mean(datalist$y.neg) ,
     sigma.neg=sqrt(var(datalist$y.neg)) ,
     c.fieldpos=1 , p=1, zlevel=2.5
     )

#####Model set-up
parameters=c("mu.fieldpos", "sigma.fieldpos", "c.fieldpos",
             "mu.neg", "sigma.neg", "zlevel",
             "mu.negdata", "sigma.negdata",
             "p", "f.neg", "f.fieldpos", "probpos",
             "se", "sp", "fpr",
             "z", "phi",
             "l.fieldpos", "neglike", "like.fieldneg",
             "outlier", "retain", "probsamppos",
             "betaod0", "betaod1", "zbetaod0", "zbetaod1",
             "muod", "sigmaod",
             "burdensamp", "scoresamp", "burden", "score")
```

```
adapt = 1000
burnin = 5000
chains = 1
numsaved=10000
thin=10
niter = ceiling( ( numsaved * thin ) / chains )

#####Running the model
jagsmodel = jags.model( "model.txt" , data=datalist ,
inits=initslist , n.chains=chains , n.adapt=adapt)

# Burn-in:
cat( "Burning in the MCMC chain...\n" )
update( jagsmodel , n.iter=burnin )

# The saved MCMC chain:
cat( "Sampling final MCMC chain...\n" )
codaSamples = coda.samples( jagsmodel , variable.names=parameters ,
                            n.iter=niter , thin=thin )

#####Extracting all MCMC results
mcmcChain = as.matrix( codaSamples )

# Extract individual parameters:

mufieldposSample = mcmcChain[,"mu.fieldpos"]
sigmafieldposSample = mcmcChain[,"sigma.fieldpos"]
cfieldposSample = mcmcChain[,"c.fieldpos"]

munegSample = mcmcChain[,"mu.neg"]
sigmanegSample = mcmcChain[,"sigma.neg"]
zlevelSample = mcmcChain[,"zlevel"]

pSample = mcmcChain[,"p"]

od0Sample = mcmcChain[,"betaod0"]
od1Sample = mcmcChain[,"betaod1"]
zsigmaodSample = mcmcChain[,"sigmaod"]

# Extract grid samples
gridnumbers<-seq(1,100,1)

f.fieldposs<-paste(rep("f.fieldpos[",100),gridnumbers,"]", sep="")
ffieldposSample = mcmcChain[,f.fieldposs]

f.negs<-paste(rep("f.neg[",100),gridnumbers,"]", sep="")
fnegSample = mcmcChain[,f.negs]

ses<-paste(rep("se[",100),gridnumbers,"]", sep="")
seSample = mcmcChain[,ses]
sps<-paste(rep("sp[",100),gridnumbers,"]", sep="")
spSample = mcmcChain[,sps]
fprs<-paste(rep("fpr[",100),gridnumbers,"]", sep="")
fprSample = mcmcChain[,fprs]
```

```
probposs<-paste(rep("probpos[",100),gridnumbers,"]", sep="")
probposSample = mcmcChain[,probposs]
burdengrids<-paste(rep("burden[",100),gridnumbers,"]", sep="")
burdengrid = mcmcChain[,burdengrids]
scoregrids<-paste(rep("score[",100),gridnumbers,"]", sep="")
scoregrid = mcmcChain[,scoregrids]

# Extract neg panel samples
numbernegs<-seq(1,N.neg,1)

retained<-paste(rep("retain[",N.neg),numbernegs,"]", sep="")
retainedSample = mcmcChain[,retained]
outliers<-paste(rep("outlier[",N.neg),numbernegs,"]", sep="")
outlierSample = mcmcChain[,outliers]

# Extract pos samples (burden)
numberposs<-seq(1,N.pos,1)

muods<-paste(rep("muod[",N.pos),numberposs,"]",sep="")
muodSample = mcmcChain[,muods]

#Individual sample assessment
numberfield<-seq(1,N.field,1)
probsampleposs<-paste(rep("probsamppos[",N.field),numberfield,"]",
sep="")
probsamplepos = mcmcChain[,probsampleposs]
burdensamples<-paste(rep("burdensamp[",N.field),numberfield,"]",
sep="")
burdensample = mcmcChain[,burdensamples]
scoresamples<-paste(rep("scoresamp[",N.field),numberfield,"]",
sep="")
scoresample = mcmcChain[,scoresamples]
```

A5. R code for Bayesian mixture model (May 2012 Kyrgyzstan data)

Note: code basically the same as described in A2, but model positive (Polya tree) component separately for each village.

```
require(rjags)
require(ggplot2)

memory.limit(size=4000)

modelstring = "
data{
for(i in 1:N.neg) {
  zerosneg[i] <- 0}
for(i in 1:N.fieldxsm) {
  zerosxsm[i] <- 0
  ones.fieldposxsm[i] <- 1}
for(i in 1:N.fieldxts) {
  zerosxts[i] <- 0
  ones.fieldposxts[i] <- 1}
for(i in 1:N.fieldxks) {
  zerosxks[i] <- 0
  ones.fieldposxks[i] <- 1}
for(i in 1:N.fieldxkk) {
  zerosxkk[i] <- 0
  ones.fieldposxkk[i] <- 1}
}

model{

##########'Known' negative samples, regardless of village of
original (excluding outliers according to z distribution)

for(i in 1:N.neg) {
      zerosneg[i] ~ dpois(negpanel[i])
      negpanel[i] <- -retain[i]*log(neglike[i])+C
      neglike[i] <- sqrt(tau.neg/(2*3.14159))*
                exp(-(tau.neg/2)*pow((y.neg[i]-mu.neg),2))
      y.neg[i] ~ dnorm(mu.negdata,tau.negdata) #training data (neg)
      stres[i]<-(y.neg[i]-mu.negdata)/sigma.negdata #stand. res.
      outlier[i]<-step(stres[i]-zlevel)     #identifying outliers
      retain[i]<-(1-outlier[i]) #selecting which obs to retain
}

##########Linear regression of log burden on OD
for( i in 1 : N.pos ) {
    zlogburden[i] ~ dnorm( muod[i] , tauod )
    muod[i] <- zbetaod0 + zbetaod1 * zod[i]
}

betaod1<-zbetaod1*logburdensd / odsd
betaod0<-zbetaod0*logburdensd+logburdenm-
zbetaod1*logburdensd*odm/odsd
```

```
####################Combining model parameters for each village

####################SM
##########Combining field distributions using zeros trick
for(i in 1:N.fieldxsm) {
zxsm[i] ~ dbern(pxsm)
zerosxsm[i] ~ dpois(phixsm[i])
phixsm[i] <- -zxsm[i]*log(like.fieldposxsm[i])-
                (1-zxsm[i])*log(like.fieldnegxsm[i])+C

##########Distribution of field positives (Polya trees)

z.fieldposxsm[i] <- (y.fieldxsm[i]-mu.fieldposxsm)/sigma.fieldposxsm
ind.fieldposxsm[i] <- trunc(sets.pos*phi(z.fieldposxsm[i])+1)
like.fieldposxsm[i]
          <- coi*exp(-0.5*z.fieldposxsm[i]*z.fieldposxsm[i])*
          polprop.fieldposxsm[ind.fieldposxsm[i]]/sigma.fieldposxsm
cpoinv.fieldposxsm[i] <- 1/(like.fieldposxsm[i]*sets.pos)
l.fieldposxsm[i]<-like.fieldposxsm[i]*sets.pos
ones.fieldposxsm[i]~dbern(like.fieldposxsm[i])

##########Distribution of field negatives (Gaussian)
like.fieldnegxsm[i]<- sqrt(tau.neg/(2*3.14159))*
                 exp(-(tau.neg/2)*pow((y.fieldxsm[i]-mu.neg),2))

##########estimating prob of positivity, burden and score
burdensampxsm[i]<-(y.fieldxsm[i]*betaod1)+betaod0
probsampposxsm[i] <-
pxsm*l.fieldposxsm[i]/(pxsm*l.fieldposxsm[i]+(1-
pxsm)*like.fieldnegxsm[i])
scoresampxsm[i]<-probsampposxsm[i]*burdensampxsm[i]
}

#####Further field Polya tree code
x.fieldposxsm[1,1] <- 0.5; x.fieldposxsm[1,2] <- 0.5
for(k in 1:j.int[Jt]){
m.fieldposxsm[1,k] <- log(x.fieldposxsm[1,1])
m.fieldposxsm[1,j.int[Jt]+k] <- log(x.fieldposxsm[1,2])
}

for(i in 2:Jt){
par.lev.fieldposxsm[i] <- c.fieldposxsm*i*i
     for(j in 1:j.int[i]){
x.fieldposxsm[i,2*j-
1]~dbeta(par.lev.fieldposxsm[i],par.lev.fieldposxsm[i]);
x.fieldposxsm[i,2*j] <- 1-x.fieldposxsm[i,2*j-1]
          for(k in 1:j.int[Jt-i+1]){
m.fieldposxsm[i,(2*j-2)*j.int[Jt-i+1]+k] <- log(x.fieldposxsm[i,2*j-
1])
m.fieldposxsm[i,(2*j-1)*j.int[Jt-i+1]+k]<-log(x.fieldposxsm[i,2*j])
          }
     }
}
```

```
for(i in 1:sets.pos){
     polprop.fieldposxsm[i] <- exp(sum(m.fieldposxsm[1:Jt,i]))
}

####################TS
##########Combining field distributions using zeros trick
for(i in 1:N.fieldxts) {
zxts[i] ~ dbern(pxts)
zerosxts[i] ~ dpois(phixts[i])
phixts[i] <- -zxts[i]*log(like.fieldposxts[i])-
               (1-zxts[i])*log(like.fieldnegxts[i])+C

##########Distribution of field positives (Polya trees)

z.fieldposxts[i] <- (y.fieldxts[i]-mu.fieldposxts)/sigma.fieldposxts
ind.fieldposxts[i] <- trunc(sets.pos*phi(z.fieldposxts[i])+1)
like.fieldposxts[i]
          <- coi*exp(-0.5*z.fieldposxts[i]*z.fieldposxts[i])*
          polprop.fieldposxts[ind.fieldposxts[i]]/sigma.fieldposxts
cpoinv.fieldposxts[i] <- 1/(like.fieldposxts[i]*sets.pos)
l.fieldposxts[i]<-like.fieldposxts[i]*sets.pos
ones.fieldposxts[i]~dbern(like.fieldposxts[i])

##########Distribution of field negatives (Gaussian)
like.fieldnegxts[i]<- sqrt(tau.neg/(2*3.14159))*
               exp(-(tau.neg/2)*pow((y.fieldxts[i]-mu.neg),2))

##########estimating prob of positivity, burden and score
#####Need to decide whether to use l.fieldpos or like.fieldpos.
#####Using l.fieldpos here as suspect is best
burdensampxts[i]<-(y.fieldxts[i]*betaod1)+betaod0
probsampposxts[i] <-
pxts*l.fieldposxts[i]/(pxts*l.fieldposxts[i]+(1-
pxts)*like.fieldnegxts[i])
scoresampxts[i]<-probsampposxts[i]*burdensampxts[i]
}

#####Further field Polya tree code
x.fieldposxts[1,1] <- 0.5; x.fieldposxts[1,2] <- 0.5
for(k in 1:j.int[Jt]){
m.fieldposxts[1,k] <- log(x.fieldposxts[1,1])
m.fieldposxts[1,j.int[Jt]+k] <- log(x.fieldposxts[1,2])
}

for(i in 2:Jt){
par.lev.fieldposxts[i] <- c.fieldposxts*i*i
     for(j in 1:j.int[i]){
x.fieldposxts[i,2*j-
1]~dbeta(par.lev.fieldposxts[i],par.lev.fieldposxts[i]);
x.fieldposxts[i,2*j] <- 1-x.fieldposxts[i,2*j-1]
          for(k in 1:j.int[Jt-i+1]){
m.fieldposxts[i,(2*j-2)*j.int[Jt-i+1]+k] <- log(x.fieldposxts[i,2*j-
1])
```

```
m.fieldposxts[i,(2*j-1)*j.int[Jt-i+1]+k] <-
log(x.fieldposxts[i,2*j])
           }
     }
}

for(i in 1:sets.pos){
     polprop.fieldposxts[i] <- exp(sum(m.fieldposxts[1:Jt,i]))
}


####################KS
##########Combining field distributions using zeros trick
for(i in 1:N.fieldxks) {
zxks[i] ~ dbern(pxks)
zerosxks[i] ~ dpois(phixks[i])
phixks[i] <- -zxks[i]*log(like.fieldposxks[i])-
               (1-zxks[i])*log(like.fieldnegxks[i])+C

##########Distribution of field positives (Polya trees)

z.fieldposxks[i] <- (y.fieldxks[i]-mu.fieldposxks)/sigma.fieldposxks
ind.fieldposxks[i] <- trunc(sets.pos*phi(z.fieldposxks[i])+1)
like.fieldposxks[i]
          <- coi*exp(-0.5*z.fieldposxks[i]*z.fieldposxks[i])*
          polprop.fieldposxks[ind.fieldposxks[i]]/sigma.fieldposxks
cpoinv.fieldposxks[i] <- 1/(like.fieldposxks[i]*sets.pos)
l.fieldposxks[i]<-like.fieldposxks[i]*sets.pos
ones.fieldposxks[i]~dbern(like.fieldposxks[i])

##########Distribution of field negatives (Gaussian)
like.fieldnegxks[i]<- sqrt(tau.neg/(2*3.14159))*
               exp(-(tau.neg/2)*pow((y.fieldxks[i]-mu.neg),2))

##########estimating prob of positivity, burden and score
burdensampxks[i]<-(y.fieldxks[i]*betaod1)+betaod0
probsampposxks[i] <-
pxks*l.fieldposxks[i]/(pxks*l.fieldposxks[i]+(1-
pxks)*like.fieldnegxks[i])
scoresampxks[i]<-probsampposxks[i]*burdensampxks[i]
}

#####Further field Polya tree code
x.fieldposxks[1,1] <- 0.5; x.fieldposxks[1,2] <- 0.5
for(k in 1:j.int[Jt]){
m.fieldposxks[1,k] <- log(x.fieldposxks[1,1])
m.fieldposxks[1,j.int[Jt]+k] <- log(x.fieldposxks[1,2])
}

for(i in 2:Jt){
par.lev.fieldposxks[i] <- c.fieldposxks*i*i
     for(j in 1:j.int[i]){
x.fieldposxks[i,2*j-
1]~dbeta(par.lev.fieldposxks[i],par.lev.fieldposxks[i]);
```

```
x.fieldposxks[i,2*j] <- 1-x.fieldposxks[i,2*j-1]
          for(k in 1:j.int[Jt-i+1]){
m.fieldposxks[i,(2*j-2)*j.int[Jt-i+1]+k] <- log(x.fieldposxks[i,2*j-
1])
m.fieldposxks[i,(2*j-1)*j.int[Jt-i+1]+k] <-
log(x.fieldposxks[i,2*j])
          }
     }
}

for(i in 1:sets.pos){
     polprop.fieldposxks[i] <- exp(sum(m.fieldposxks[1:Jt,i]))
}


####################KK
##########Combining field distributions using zeros trick
for(i in 1:N.fieldxkk) {
zxkk[i] ~ dbern(pxkk)
zerosxkk[i] ~ dpois(phixkk[i])
phixkk[i] <- -zxkk[i]*log(like.fieldposxkk[i])-
                (1-zxkk[i])*log(like.fieldnegxkk[i])+C

##########Distribution of field positives (Polya trees)

z.fieldposxkk[i] <- (y.fieldxkk[i]-mu.fieldposxkk)/sigma.fieldposxkk
ind.fieldposxkk[i] <- trunc(sets.pos*phi(z.fieldposxkk[i])+1)
like.fieldposxkk[i]
          <- coi*exp(-0.5*z.fieldposxkk[i]*z.fieldposxkk[i])*
          polprop.fieldposxkk[ind.fieldposxkk[i]]/sigma.fieldposxkk
cpoinv.fieldposxkk[i] <- 1/(like.fieldposxkk[i]*sets.pos)
l.fieldposxkk[i]<-like.fieldposxkk[i]*sets.pos
ones.fieldposxkk[i]~dbern(like.fieldposxkk[i])

##########Distribution of field negatives (Gaussian)
like.fieldnegxkk[i]<- sqrt(tau.neg/(2*3.14159))*
exp(-(tau.neg/2)*pow((y.fieldxkk[i]-mu.neg),2))

##########estimating prob of positivity, burden and score
#####Need to decide whether to use l.fieldpos or like.fieldpos.
#####Using l.fieldpos here as suspect is best
burdensampxkk[i]<-(y.fieldxkk[i]*betaod1)+betaod0
probsampposxkk[i] <-
pxkk*l.fieldposxkk[i]/(pxkk*l.fieldposxkk[i]+(1-
pxkk)*like.fieldnegxkk[i])
scoresampxkk[i]<-probsampposxkk[i]*burdensampxkk[i]
}

#####Further field Polya tree code
x.fieldposxkk[1,1] <- 0.5; x.fieldposxkk[1,2] <- 0.5
for(k in 1:j.int[Jt]){
m.fieldposxkk[1,k] <- log(x.fieldposxkk[1,1])
m.fieldposxkk[1,j.int[Jt]+k] <- log(x.fieldposxkk[1,2])
}
```

```
for(i in 2:Jt){
par.lev.fieldposxkk[i] <- c.fieldposxkk*i*i
     for(j in 1:j.int[i]){
x.fieldposxkk[i,2*j-
1]~dbeta(par.lev.fieldposxkk[i],par.lev.fieldposxkk[i]);
x.fieldposxkk[i,2*j] <- 1-x.fieldposxkk[i,2*j-1]
          for(k in 1:j.int[Jt-i+1]){
m.fieldposxkk[i,(2*j-2)*j.int[Jt-i+1]+k] <- log(x.fieldposxkk[i,2*j-
1])
m.fieldposxkk[i,(2*j-1)*j.int[Jt-i+1]+k] <-
log(x.fieldposxkk[i,2*j])
          }
     }
}

for(i in 1:sets.pos){
     polprop.fieldposxkk[i] <- exp(sum(m.fieldposxkk[1:Jt,i]))
}


##########Estimating distributions over grid of OD values
for(i in 1:totalgrid){
     grid[i]<-i*0.01
     f.neg[i] <- sqrt(tau.neg/(2*3.14159))*exp(-(tau.neg/2)
              *pow((grid[i]-mu.neg),2))
     cdfneg[i]<-sum(f.neg[1:i])
     sp[i]<-cdfneg[i]/totalneg
     fpr[i]<-1-sp[i]

burdenest[i]<-(grid[i]*betaod1)+betaod0

z.fieldfxsm[i] <- (grid[i]-mu.fieldposxsm)/sigma.fieldposxsm
ind.fieldfxsm[i] <- trunc(sets.pos*phi(z.fieldfxsm[i])+1)
f.fieldposxsm[i]<-
     sets.pos*coi*exp(-0.5*z.fieldfxsm[i]*z.fieldfxsm[i])
     *polprop.fieldposxsm[ind.fieldfxsm[i]]/sigma.fieldposxsm
cdffieldposxsm[i]<-sum(f.fieldposxsm[1:i])

probposxsm[i] <- pxsm*f.fieldposxsm[i]/(pxsm*f.fieldposxsm[i]+(1-
pxsm)*f.neg[i])
scorexsm[i]<-probposxsm[i]*burdenest[i]

sexsm[i]<-1-(cdffieldposxsm[i]/totalposxsm)

z.fieldfxts[i] <- (grid[i]-mu.fieldposxts)/sigma.fieldposxts
ind.fieldfxts[i] <- trunc(sets.pos*phi(z.fieldfxts[i])+1)
f.fieldposxts[i]<-
     sets.pos*coi*exp(-0.5*z.fieldfxts[i]*z.fieldfxts[i])
     *polprop.fieldposxts[ind.fieldfxts[i]]/sigma.fieldposxts
cdffieldposxts[i]<-sum(f.fieldposxts[1:i])

probposxts[i] <- pxts*f.fieldposxts[i]/(pxts*f.fieldposxts[i]+(1-
pxts)*f.neg[i])
```

```
scorexts[i]<-probposxts[i]*burdenest[i]

sexts[i]<-1-(cdffieldposxts[i]/totalposxts)

z.fieldfxks[i] <- (grid[i]-mu.fieldposxks)/sigma.fieldposxks
ind.fieldfxks[i] <- trunc(sets.pos*phi(z.fieldfxks[i])+1)
f.fieldposxks[i]<-
      sets.pos*coi*exp(-0.5*z.fieldfxks[i]*z.fieldfxks[i])
      *polprop.fieldposxks[ind.fieldfxks[i]]/sigma.fieldposxks
cdffieldposxks[i]<-sum(f.fieldposxks[1:i])

probposxks[i] <- pxks*f.fieldposxks[i]/(pxks*f.fieldposxks[i]+(1-
pxks)*f.neg[i])
scorexks[i]<-probposxks[i]*burdenest[i]

sexks[i]<-1-(cdffieldposxks[i]/totalposxks)

z.fieldfxkk[i] <- (grid[i]-mu.fieldposxkk)/sigma.fieldposxkk
ind.fieldfxkk[i] <- trunc(sets.pos*phi(z.fieldfxkk[i])+1)
f.fieldposxkk[i]<-
      sets.pos*coi*exp(-0.5*z.fieldfxkk[i]*z.fieldfxkk[i])
      *polprop.fieldposxkk[ind.fieldfxkk[i]]/sigma.fieldposxkk
cdffieldposxkk[i]<-sum(f.fieldposxkk[1:i])

probposxkk[i] <- pxkk*f.fieldposxkk[i]/(pxkk*f.fieldposxkk[i]+(1-
pxkk)*f.neg[i])
scorexkk[i]<-probposxkk[i]*burdenest[i]

sexkk[i]<-1-(cdffieldposxkk[i]/totalposxkk)

}

totalneg<-sum(f.neg[1:totalgrid])

totalposxsm<-sum(f.fieldposxsm[1:totalgrid])
totalposxts<-sum(f.fieldposxts[1:totalgrid])
totalposxks<-sum(f.fieldposxks[1:totalgrid])
totalposxkk<-sum(f.fieldposxkk[1:totalgrid])


########################################Priors
#positives
mu.fieldposxsm~dunif(minpos,maxpos)
sigma.fieldposxsm~dunif(0.001,1)
c.fieldposxsm~dgamma(5,1)
mu.fieldposxts~dunif(minpos,maxpos)
sigma.fieldposxts~dunif(0.001,1)
c.fieldposxts~dgamma(5,1)
mu.fieldposxks~dunif(minpos,maxpos)
sigma.fieldposxks~dunif(0.001,1)
c.fieldposxks~dgamma(5,1)
mu.fieldposxkk~dunif(minpos,maxpos)
sigma.fieldposxkk~dunif(0.001,1)
c.fieldposxkk~dgamma(5,1)
```

```
#negatives
mu.neg~dunif(minneg,maxneg)
sigma.neg~dunif(0.001,1)
mu.negdata~dunif(minneg,maxneg)
sigma.negdata~dunif(0.001,1)
zlevel~dunif(0,4)
sigma2.negdata<-pow(sigma.negdata,2)
tau.negdata<-1/sigma2.negdata
sigma2.neg<-pow(sigma.neg,2)
tau.neg<-1/sigma2.neg

#combination
pxsm~dunif(0,1)
pxts~dunif(0,1)
pxks~dunif(0,1)
pxkk~dunif(0,1)
coi <- 0.3989422804
C<-10000

#linear regression
zbetaod0 ~ dnorm( 0 , 1.0E-12 )
zbetaod1 ~ dnorm( 0 , 1.0E-12 )
tauod ~ dgamma( 0.001 , 0.001 )
sigmaod<-1/sqrt(tauod)

}
"
writeLines(modelstring,con="model.txt")

#####Fitting to May 2012 Data
maydata<-read.csv("may2012database.csv")

str(maydata)
maydata$village<-as.factor(maydata$village)

maypaneldata<-subset(maydata,panel==1)
maypaneldata<-subset(maypaneldata,repfailureinccont!=1)

mayfielddata<-subset(maydata,use==1)
may.pos<-maypaneldata$od[which(maypaneldata$knownstatus==1 &
is.na(maypaneldata$od)==FALSE)]
may.neg<-maypaneldata$od[which(maypaneldata$knownstatus==0 &
is.na(maypaneldata$od)==FALSE)]
may.field<-mayfielddata$od
maysm<-mayfielddata$od[mayfielddata$village=="SM"]
mayts<-mayfielddata$od[mayfielddata$village=="TS"]
mayks<-mayfielddata$od[mayfielddata$village=="KS"]
maykk<-mayfielddata$od[mayfielddata$village=="KK"]
may.burden<-maypaneldata$burden[which(maypaneldata$knownstatus==1 &
is.na(maypaneldata$od)==FALSE)]
may.logburden<-log(may.burden)

##########Running code
```

```
fileNameRoot="may"

#polya tree
Jt=2
sets.pos=2^Jt
j.int=c(1,2)

#selecting parameters to use for model
y.pos=may.pos
y.neg=may.neg
y.fieldxsm<-maysm
y.fieldxts<-mayts
y.fieldxks<-mayks
y.fieldxkk<-maykk
burden=may.burden
logburden=may.logburden

minimum<-min(y.fieldxsm, y.fieldxts, y.fieldxks, y.fieldxkk, y.pos,
y.neg)
maximum<-max(y.fieldxsm, y.fieldxts, y.fieldxks, y.fieldxkk, y.pos,
y.neg)
maxfield<-max(y.fieldxsm, y.fieldxts, y.fieldxks, y.fieldxkk)
minfield<-min(y.fieldxsm, y.fieldxts, y.fieldxks, y.fieldxkk)
totalgrid=max(60,maximum,maxfield)

#data processing
N.pos=length(y.pos)
N.neg=length(y.neg)
N.fieldxsm=length(y.fieldxsm)
N.fieldxts=length(y.fieldxts)
N.fieldxks=length(y.fieldxks)
N.fieldxkk=length(y.fieldxkk)
odm = mean( y.pos ,na.rm=TRUE) ; odsd = sd( y.pos ,na.rm=TRUE)
zod = ( y.pos - odm ) / odsd
logburdenm = mean( logburden , na.rm=TRUE) ; logburdensd = sd(
logburden , na.rm=TRUE)
zlogburden = ( logburden - logburdenm ) / logburdensd
minneg=min(y.neg)
maxneg=max(y.neg)
minpos=min(y.pos)
maxpos=max(y.pos)

#####Load data
datalist = list( N.pos=N.pos, N.neg=N.neg, y.neg=y.neg,
                 N.fieldxsm=N.fieldxsm, N.fieldxts=N.fieldxts,
N.fieldxks=N.fieldxks, N.fieldxkk=N.fieldxkk,
                 y.fieldxsm=y.fieldxsm, y.fieldxts=y.fieldxts,
y.fieldxks=y.fieldxks, y.fieldxkk=y.fieldxkk,
                 sets.pos=sets.pos, Jt=Jt, j.int=j.int,
                 odm=odm, odsd=odsd, zod=zod,
                 zlogburden=zlogburden, logburdensd=logburdensd,
logburdenm=logburdenm,
                 minneg=minneg, maxneg=maxneg, minpos=minpos,
maxpos=maxpos, totalgrid=totalgrid)
```

```
#####Inits
initslist = list(
  mu.fieldposxsm=mean(y.pos, na.rm=TRUE), mu.fieldposxts=mean(y.pos,
na.rm=TRUE),
  mu.fieldposxks=mean(y.pos, na.rm=TRUE), mu.fieldposxkk=mean(y.pos,
na.rm=TRUE),
  sigma.fieldposxsm=sqrt(var(y.pos, na.rm=TRUE)),
sigma.fieldposxts=sqrt(var(y.pos, na.rm=TRUE)),
  sigma.fieldposxks=sqrt(var(y.pos, na.rm=TRUE)),
sigma.fieldposxkk=sqrt(var(y.pos, na.rm=TRUE)),
  mu.neg=mean(datalist$y.neg, na.rm=TRUE) ,
sigma.neg=sqrt(var(datalist$y.neg, na.rm=TRUE)) ,
  c.fieldposxsm=1, c.fieldposxts=1, c.fieldposxks=1,
c.fieldposxkk=1,
  pxsm=1, pxts=1, pxks=1, pxkk=1, zlevel=2.5
  )

#####Model set-up
adapt = 1000
burnin = 2000
chains = 1
numsaved=5000
thin=1
niter = ceiling( ( numsaved * thin ) / chains )

parameters=c("mu.fieldposxsm","sigma.fieldposxsm","c.fieldposxsm",
             "mu.fieldposxts","sigma.fieldposxts","c.fieldposxts",
             "mu.fieldposxks","sigma.fieldposxks","c.fieldposxks",
             "mu.fieldposxkk","sigma.fieldposxkk","c.fieldposxkk",
             "mu.neg", "sigma.neg", "zlevel",
             "mu.negdata", "sigma.negdata",
             "pxsm", "pxts", "pxks", "pxkk",
             "f.neg",
             "f.fieldposxsm", "f.fieldposxts",
             "f.fieldposxks", "f.fieldposxkk",
             "probposxsm", "probposxts", "probposxks", "probposxkk",
             "sexsm", "sexts", "sexks", "sexkk",
             "sp",
             "fpr",
             "zxsm", "zxts", "zxks", "zxkk",
             "phixsm", "phixts", "phixks", "phixkk",
             "l.fieldposxsm", "l.fieldposxts",
             "l.fieldposxks", "l.fieldposxkk",
             "neglike", "outlier", "retain",
             "like.fieldnegxsm", "like.fieldnegxts",
             "like.fieldnegxks", "like.fieldnegxkk",
             "probsampposxsm", "probsampposxts",
             "probsampposxks", "probsampposxkk",
             "betaod0", "betaod1", "zbetaod0", "zbetaod1",
             "muod", "sigmaod",
             "burdensampxsm", "burdensampxts",
             "burdensampxks", "burdensampxkk",
             "scoresampxsm", "scoresampxts",
```

```
              "scoresampxks", "scoresampxkk",
              "burdenest",
              "scorexsm", "scorexts", "scorexks", "scorexkk")


##############################################Running the model
jagsmodel = jags.model( "model.txt" , data=datalist ,
inits=initslist , n.chains=chains , n.adapt=adapt)

# Burn-in:
cat( "Burning in the MCMC chain...\n" )
update( jagsmodel , n.iter=burnin )

# The saved MCMC chain:
cat( "Sampling final MCMC chain...\n" )
codaSamples = coda.samples( jagsmodel , variable.names=parameters ,
n.iter=niter , thin=thin )

#####Extracting all MCMC results
mcmcChain = as.matrix( codaSamples )
write.csv(mcmcChain, file=paste(fileNameRoot,"mcmc.csv"))

# Extract individual parameters:

mufieldposxsmSample = mcmcChain[,"mu.fieldposxsm"]
sigmafieldposxsmSample = mcmcChain[,"sigma.fieldposxsm"]
cfieldposxsmSample = mcmcChain[,"c.fieldposxsm"]
mufieldposxtsSample = mcmcChain[,"mu.fieldposxts"]
sigmafieldposxtsSample = mcmcChain[,"sigma.fieldposxts"]
cfieldposxtsSample = mcmcChain[,"c.fieldposxts"]
mufieldposxksSample = mcmcChain[,"mu.fieldposxks"]
sigmafieldposxksSample = mcmcChain[,"sigma.fieldposxks"]
cfieldposxksSample = mcmcChain[,"c.fieldposxks"]
mufieldposxkkSample = mcmcChain[,"mu.fieldposxkk"]
sigmafieldposxkkSample = mcmcChain[,"sigma.fieldposxkk"]
cfieldposxkkSample = mcmcChain[,"c.fieldposxkk"]

munegSample = mcmcChain[,"mu.neg"]
sigmanegSample = mcmcChain[,"sigma.neg"]
zlevelSample = mcmcChain[,"zlevel"]

pxsmSample = mcmcChain[,"pxsm"]
pxtsSample = mcmcChain[,"pxts"]
pxksSample = mcmcChain[,"pxks"]
pxkkSample = mcmcChain[,"pxkk"]

od0Sample = mcmcChain[,"betaod0"]
od1Sample = mcmcChain[,"betaod1"]
zsigmaodSample = mcmcChain[,"sigmaod"]

# Extract grid samples
gridnumbers<-seq(1,totalgrid,1)
```

```
f.fieldpossxsm<-
paste(rep("f.fieldposxsm[",totalgrid),gridnumbers,"]", sep="")
ffieldposxsmSample = mcmcChain[,f.fieldpossxsm]
f.fieldpossxts<-
paste(rep("f.fieldposxts[",totalgrid),gridnumbers,"]", sep="")
ffieldposxtsSample = mcmcChain[,f.fieldpossxts]
f.fieldpossxks<-
paste(rep("f.fieldposxks[",totalgrid),gridnumbers,"]", sep="")
ffieldposxksSample = mcmcChain[,f.fieldpossxks]
f.fieldpossxkk<-
paste(rep("f.fieldposxkk[",totalgrid),gridnumbers,"]", sep="")
ffieldposxkkSample = mcmcChain[,f.fieldpossxkk]

f.negs<-paste(rep("f.neg[",totalgrid),gridnumbers,"]", sep="")
fnegSample = mcmcChain[,f.negs]

sexsms<-paste(rep("sexsm[",totalgrid),gridnumbers,"]", sep="")
sexsmSample = mcmcChain[,sexsms]
sextss<-paste(rep("sexts[",totalgrid),gridnumbers,"]", sep="")
sextsSample = mcmcChain[,sextss]
sexkss<-paste(rep("sexks[",totalgrid),gridnumbers,"]", sep="")
sexksSample = mcmcChain[,sexkss]
sexkks<-paste(rep("sexkk[",totalgrid),gridnumbers,"]", sep="")
sexkkSample = mcmcChain[,sexkks]
sps<-paste(rep("sp[",totalgrid),gridnumbers,"]", sep="")
spSample = mcmcChain[,sps]
fprs<-paste(rep("fpr[",totalgrid),gridnumbers,"]", sep="")
fprSample = mcmcChain[,fprs]

probposxsms<-paste(rep("probposxsm[",totalgrid),gridnumbers,"]",
sep="")
probposxsmSample = mcmcChain[,probposxsms]
probposxtss<-paste(rep("probposxts[",totalgrid),gridnumbers,"]",
sep="")
probposxtsSample = mcmcChain[,probposxtss]
probposxkss<-paste(rep("probposxks[",totalgrid),gridnumbers,"]",
sep="")
probposxksSample = mcmcChain[,probposxkss]
probposxkks<-paste(rep("probposxkk[",totalgrid),gridnumbers,"]",
sep="")
probposxkkSample = mcmcChain[,probposxkks]

burdengrids<-paste(rep("burdenest[",totalgrid),gridnumbers,"]",
sep="")
burdengrid = mcmcChain[,burdengrids]

scoregridxsms<-paste(rep("scorexsm[",totalgrid),gridnumbers,"]",
sep="")
scoregridxsm = mcmcChain[,scoregridxsms]
scoregridxtss<-paste(rep("scorexts[",totalgrid),gridnumbers,"]",
sep="")
scoregridxts = mcmcChain[,scoregridxtss]
scoregridxkss<-paste(rep("scorexks[",totalgrid),gridnumbers,"]",
sep="")
```

```
scoregridxks = mcmcChain[,scoregridxkss]
scoregridxkks<-paste(rep("scorexkk[",totalgrid),gridnumbers,"]",
sep="")
scoregridxkk = mcmcChain[,scoregridxkks]

# Extract neg panel samples
numbernegs<-seq(1,N.neg,1)

retained<-paste(rep("retain[",N.neg),numbernegs,"]", sep="")
retainedSample = mcmcChain[,retained]
outliers<-paste(rep("outlier[",N.neg),numbernegs,"]", sep="")
outlierSample = mcmcChain[,outliers]

# Extract pos samples (burden)
numberposs<-seq(1,N.pos,1)

muods<-paste(rep("muod[",N.pos),numberposs,"]",sep="")
muodSample = mcmcChain[,muods]

# Extract field data samples
numberfieldxsm<-seq(1,N.fieldxsm,1)
numberfieldxts<-seq(1,N.fieldxts,1)
numberfieldxks<-seq(1,N.fieldxks,1)
numberfieldxkk<-seq(1,N.fieldxkk,1)

probsampleposxsms<-
paste(rep("probsampposxsm[",N.fieldxsm),numberfieldxsm,"]", sep="")
probsampleposxsm = mcmcChain[,probsampleposxsms]
probsampleposxtss<-
paste(rep("probsampposxts[",N.fieldxts),numberfieldxts,"]", sep="")
probsampleposxts = mcmcChain[,probsampleposxtss]
probsampleposxkss<-
paste(rep("probsampposxks[",N.fieldxks),numberfieldxks,"]", sep="")
probsampleposxks = mcmcChain[,probsampleposxkss]
probsampleposxkks<-
paste(rep("probsampposxkk[",N.fieldxkk),numberfieldxkk,"]", sep="")
probsampleposxkk = mcmcChain[,probsampleposxkks]

burdensamplexsms<-
paste(rep("burdensampxsm[",N.fieldxsm),numberfieldxsm,"]", sep="")
burdensamplexsm = mcmcChain[,burdensamplexsms]
burdensamplextss<-
paste(rep("burdensampxts[",N.fieldxts),numberfieldxts,"]", sep="")
burdensamplexts = mcmcChain[,burdensamplextss]
burdensamplexkss<-
paste(rep("burdensampxks[",N.fieldxks),numberfieldxks,"]", sep="")
burdensamplexks = mcmcChain[,burdensamplexkss]
burdensamplexkks<-
paste(rep("burdensampxkk[",N.fieldxkk),numberfieldxkk,"]", sep="")
burdensamplexkk = mcmcChain[,burdensamplexkks]

scoresamplexsms<-
paste(rep("scoresampxsm[",N.fieldxsm),numberfieldxsm,"]", sep="")
scoresamplexsm = mcmcChain[,scoresamplexsms]
```

```
scoresamplextss<-
paste(rep("scoresampxts[",N.fieldxts),numberfieldxts,"]", sep="")
scoresamplexts = mcmcChain[,scoresamplextss]
scoresamplexkss<-
paste(rep("scoresampxks[",N.fieldxks),numberfieldxks,"]", sep="")
scoresamplexks = mcmcChain[,scoresamplexkss]
scoresamplexkks<-
paste(rep("scoresampxkk[",N.fieldxkk),numberfieldxkk,"]", sep="")
scoresamplexkk = mcmcChain[,scoresamplexkks]
```

A6. R code for mathematical model

```
library(deSolve)

#####Parameters
nd = 100              # peak dog population density (per km^2)
nf = 2               # peak fox population density (per km^2)
nv = 5000            # peak rodent population density (per km^2)
ns = 500              # average ruminant population density (per
km^2)
muf = 0.0022          # rate of fox death (foxes per day)
mud = 1/(5*365)       # rate of dog death (dogs per day)
mumf = 1/30           # rate of natural Em death in foxes (worms per
day)
mumd = 1/((365/3.4)-30) # rate of natural Em death in dogs (worms/d)
mucd = 1/((365/1.3)-40) # rate of natural Eg death in dogs (worms/d)
muv = 0.0055          # baseline rate of rodent death (rodents/d)
mus = 1/(5*365) # baseline rate of ruminant natural death (per day)
mu_s = 5/(100*365) # baseline rate of ruminant slaughter (per day)
sigmav = (926239/81) # average number of Em protoscolices (per egg)
sigmas =
mean(c(21/2.6,99/3.1,711/2.5,1726/4.61,6899/7.1,9774/10.28))
# average number of Eg protoscolices (per egg)
pidv = 0.1            # probability of dog infection from rodent
pifv = 0.4            # probability of fox infection from rodent
pids = 0.05           # probability of dog infection from ruminant
pivg = 0.007          # probability of rodent infection from eggs
pisg = 0.003          # probability of ruminant infection from eggs
lambdamf = 42 # rate of egg release from foxes (per worm per day)
lambdamd = 42 # rate of Em egg release from dogs (per worm per day)
lambdacd = 42 # rate of Eg egg release from dogs (per worm per day)
taumd = 30            # Em maturation time in dogs (days)
taucd = 42            # Eg maturation time in dogs (days)
taumf = 30            # Em maturation time in foxes (days)
tauv = 112            # Em maturation time in rodents (days)
taus = 365*2          # Eg maturation time in ruminants (days)
Kcd =  2500           # MWB carrying capacity in dogs (worms)
Kmd =  2534           # MWB carrying capacity in dogs (worms)
Kmf =  16000          # MWB carrying capacity in foxes (worms)
Kv =  244400          # MPB carrying capacity in rodents
Ks =  9774            # MPB carrying capacity in ruminants
kappad = 1            # relative mortality in infected dogs
kappaf = 1            # relative mortality in infected foxes
kappav = 2            # relative mortality in infected rodents
kappavp = 2           # relative predation of infected rodents
kappas = 2  # relative natural mortality in infected ruminants
kappa_s = 2 # relative slaughter of infected ruminants

#beta parameters
betadv = ((1.1/365)/(pidv*nv))  # max rate of rodent hunting by dogs
(prop of total rodents (in 1 km^2 area) per dog per day)
betafv = ((3.8/365)/(pifv*nv)) # max rate of rodent hunting by foxes
(prop of total rodents (in 1 km^2 area) per dog per day)
```

```
beta_ds = 0.05 # proportion of slaughtered ruminants fed to dogs
(per dog per day)
betads = (((0.3/365)/(pids*ns))-(beta_ds*mu_s))/(mus)  # proportion
of dead ruminants scavenged by dogs (dead ruminants per dog per day)
betavg = (betadv*nd+betafv*nf+muv)/(pivg*nv*sigmav) # rate of
ingestion of eggs by rodents (per rodent per day)
betasg<-(mus+mu_s)/(pisg*ns*sigmas) # rate of ingestion of eggs by
ruminants (per ruminant per day)

#####Creating vector of parameters
parms <- c(mud = mud, muf = muf,
           mumd = mumd, mucd = mucd, mumf = mumf,
           muv = muv, mus = mus, mu_s = mu_s,
           mugm = mugm, mugc = mugc,
           sigmav = sigmav, sigmas = sigmas,
           pidv = pidv, pifv = pifv,
           pids = pids,
           pivg = pivg, pisg = pisg,
           lambdamf = lambdamf, lambdamd = lambdamd,
           lambdacd = lambdacd,
           betadv = betadv, betafv = betafv,
           betads = betads, beta_ds = beta_ds,
           betavg = betavg, betasg = betasg,
           nf = nf, nd = nd,
           nv = nv, ns = ns,
           taumd = taumd, taucd = taucd, taumf = taumf,
           tauv = tauv, taus = taus,
           Kmd = Kmd, Kmf = Kmf, Kv = Kv, Ks = Ks,
           kappad = kappad, kappaf = kappaf,
           kappav = kappav, kappavp = kappavp,
           kappas = kappas, kappa_s = kappa_s)


#####Creating vector of timesteps
years<-100
times  <- seq(0, 365*years, by=1)

#####Start values - starting with 100 worms in dogs
y <- xstart <- c(Gm = 0, Gc = 0, Lv = 0, Ls=0, Mmd = 100, Mcd = 100,
Mmf = 0)

#####Modelling seasonal effects

fulldates1<-seq(0,max(times),365/12)
maxdates<-length(fulldates1)
fulldates<-fulldates1[-maxdates]

fulldates1<-round(fulldates1,0)
fulldates<-round(fulldates,0)

#1) Temperature effect on eggs
temps<-as.numeric(c(-11,-9,-2.2,5.5,9.5,13.8,17,16.4,11.6,5.2,-1.4,-
7.1))
```

```
emsurvival<-exp(-0.135*(abs(temps)-43.49))
egsurvival<-exp(-0.135*(abs(temps)-43.49))

emeggsurvapprox <-
as.data.frame(list(times=fulldates,import=rep(emsurvival,years)))
egeggsurvapprox <-
as.data.frame(list(times=fulldates,import=rep(egsurvival,years)))

emeggsurv <- approxfun(emeggsurvapprox, rule = 2)
egeggsurv <- approxfun(egeggsurvapprox, rule = 2)

#2) Dog density
doghigh<-nd
doglow<-nd/2

dogdynamics<-c(rep(doghigh,4),rep(doglow,5),rep(doghigh,3))
dogapprox <-
as.data.frame(list(times=fulldates,import=rep(dogdynamics,years)))
dogdens <- approxfun(dogapprox, rule = 2)

#3) Fox density
numpups<-5
foxnopup<-nf
foxpup<-foxnopup+((foxnopup/2)*numpups)

foxdynamics<-c(rep(foxnopup,5),rep(foxpup,4),rep(foxnopup,3))
foxapprox <-
as.data.frame(list(times=fulldates,import=rep(foxdynamics,years)))
foxdens <- approxfun(foxapprox, rule = 2)

#4) Rodent density
volehigh<-nv*2
volemed<-nv
volelow<-nv/2

voledynamics<-
c(rep(volemed,3),rep(volelow,3),rep(volehigh,3),rep(volemed,3))
voleapprox <-
as.data.frame(list(times=fulldates,import=rep(voledynamics,years)))
voledens <- approxfun(voleapprox, rule = 2)

#5) Ruminant density (no seasonality included here)
sheephigh<-ns+(ns/2)
sheepmed<-ns
sheeplow<-ns/2

sheepdynamics<-
c(rep(sheepmed,1),rep(sheepmed,3),rep(sheepmed,3),rep(sheepmed,3),re
p(sheepmed,2))
sheepapprox <-
as.data.frame(list(times=fulldates,import=rep(sheepdynamics,years)))
sheepdens <- approxfun(sheepapprox, rule = 2)

#6) Rodent mortality
```

```
voledeathhigh<-muv*2
voledeathmed<-muv
voledeathlow<-muv/2

voledeathdynamics<-
c(rep(voledeathhigh,3),rep(voledeathmed,8),rep(voledeathhigh,1))
voledeathapprox <-
as.data.frame(list(times=fulldates,import=rep(voledeathdynamics,year
s)))
voledeath <- approxfun(voledeathapprox, rule = 2)

#7) Ruminant mortality
rumdeathlow<-mus/2
rumdeathhigh<-mus*2

rumdeathdynamics<-
c(rep(rumdeathhigh,3),rep(rumdeathlow,8),rep(rumdeathhigh,1))
rumdeathapprox <-
as.data.frame(list(times=fulldates,import=rep(rumdeathdynamics,years
)))
rumdeath <- approxfun(rumdeathapprox, rule = 2)

#8) Ruminant slaughter
rumslaughtlow<-mu_s/2
rumslaughthigh<-mu_s*2

rumslaughtdynamics<-
c(rep(rumslaughthigh,2),rep(rumslaughtlow,6),rep(rumslaughthigh,4))
rumslaughtapprox <-
as.data.frame(list(times=fulldates,import=rep(rumslaughtdynamics,yea
rs)))
rumslaught <- approxfun(rumslaughtapprox, rule = 2)

#9) Fox and dog reliance on rodents
betafvhigh<-betafv
betafvlow<-betafv/2

foxingestvole<-
c(rep(betafvlow,3),rep(betafvhigh,6),rep(betafvlow,3))
foxingestvoleapprox <-
as.data.frame(list(times=fulldates,import=rep(foxingestvole,years)))
betafvdynamic <- approxfun(foxingestvoleapprox, rule = 2)

betadvhigh<-betadv
betadvlow<-betadv/2

dogingestvole<-
c(rep(betadvlow,3),rep(betadvhigh,6),rep(betadvlow,3))
dogingestvoleapprox <-
as.data.frame(list(times=fulldates,import=rep(dogingestvole,years)))
betadvdynamic <- approxfun(dogingestvoleapprox, rule = 2)
```

```
#####Final model code
alaymodel <- function(t, x, parms) {
  with(as.list(c(parms, x)), {

    emeggsurvival<-emeggsurv(t)
    egeggsurvival<-egeggsurv(t)
    volepop<-voledens(t)
    sheeppop<-sheepdens(t)
    foxpop<-foxdens(t)
    dogpop<-dogdens(t)
    volemort<-voledeath(t)
    rummort<-rumdeath(t)
    rumkill<-rumslaught(t)
    betafvtemp<-betafvdynamic(t)
    betadvtemp<-betadvdynamic(t)

    Gm=x[1]; Gc=x[2];
    Lv=x[3]; Ls=x[4];
    Mmd=x[5]; Mcd=x[6]; Mmf=x[7];

    #####temporal lags
    if (t<=tauv)
      lagGm<-y[1]
    else                              # (Em eggs)
      lagGm<-lagvalue(t-tauv, 1) # lag for maturation in rodents

    if (t<=taus)
      lagGc<-y[2]
    else                              # (Eg eggs)
      lagGc<-lagvalue(t-taus, 2) # lag for maturation in ruminants

    if(t<=taumd)
      lagLvd<-y[3]
    else                              # (Em protoscolices)
      lagLvd<-lagvalue(t-taumd, 3) # lag for Em maturation in dogs

    if(t<=taucd)
      lagLs<-y[4]
    else                              # (Eg protoscolices)
      lagLs<-lagvalue(t-taucd, 4) # lag for Eg maturation in dogs

    if(t<=taumf)
      lagLvf<-y[3]
    else                              # (Em protoscolices)
      lagLvf<-lagvalue(t-taumf, 3) # lag for Em maturation in foxes

    #####model compartments

    #eggs per km2
    dGm <- lambdamd*dogpop*Mmd
                + lambdamf*foxpop*Mmf
                - (1/emeggsurvival)*Gm
    dGc <- lambdacd*dogpop*Mcd - (1/egeggsurvival)*Gc
```

```
    # protoscolices per rodent
    dLv <- betavg*pivg*sigmav*lagGm*(1-(Lv/Kv))
           - volemort*kappav*Lv
           - betadvtemp*dogpop*kappavp*Lv
           - betafvtemp*foxpop*kappavp*Lv

    # protoscolices per ruminant
    dLs <- betasg*pisg*sigmas*lagGc*(1-(Ls/Ks))
           - (rummort*kappas+rumkill*kappa_s)*Ls

    # adult worms per dog
    dMmd <- betadvtemp*pidv*volepop*kappavp*lagLvd*(1-(Mmd/Kmd))
                - mud*kappad*Mmd - mumd*Mmd
    dMcd <-
(betads*rummort*kappas+beta_ds*rumkill*kappa_s)*pids*sheeppop*lagLs*
(1-(Mcd/Kcd)) - mud*kappad*Mcd - mucd*Mcd

    # worms per fox
    dMmf <- betafvtemp*pifv*volepop*kappavp*lagLvf*(1-(Mmf/Kmf))
                - muf*kappaf*Mmf - mumf*Mmf


    #####output

    res <- c(dGm, dGc, dLv, dLs, dMmd, dMcd, dMmf)
    list(res)
  })
}

####################Running no intervention model
out <-  dede(y=xstart, times=times, func=alaymodel, parms=parms)

####################Looking into effect of regular dosing
startdose=10
steadystate<-startdose*365

#creating functions to account for effective suppression of adult
worms following dosing due to PPP
dosingfunction_eg<-function(dosetimes){
  times_eg<-NA
  for (i in 1:length(dosetimes)){
    times_egx<-seq(dosetimes[i],dosetimes[i]+taucd,1)
    times_eg<-c(times_eg,times_egx)
  }
  times_eg<-times_eg[times_eg<=max(times)]
  times_eg<-times_eg[-1]
  print(times_eg)
}

dosingfunction_em<-function(dosetimes){
  times_em<-NA
  for (i in 1:length(dosetimes)){
    times_emx<-seq(dosetimes[i],dosetimes[i]+taumd,1)
    times_em<-c(times_em,times_emx)
```

```
  }
  times_em<-times_em[times_em<=max(times)]
  times_em<-times_em[-1]
  print(times_em)
}

##### timing of dosing with PZQ
times1m <- round(seq(steadystate,max(times),(365/12)),0) #q1m
times2m <- round(seq(steadystate,max(times),((365/12)*2)),0) #q2m
times3m <- round(seq(steadystate,max(times),((365/12)*3)),0) #q3m
times4m <- round(seq(steadystate,max(times),((365/12)*4)),0) #q4m
times6m = round(seq(steadystate,max(times),((365/12)*6)),0) #q6m
times12m = round(seq(steadystate,max(times),365),0) #q12m

times1m_eg<-dosingfunction_eg(times1m)
times1m_em<-dosingfunction_em(times1m)
pzqdoses1meg<-length(times1m_eg)
pzqdoses1mem<-length(times1m_em)

times2m_eg<-dosingfunction_eg(times2m)
times2m_em<-dosingfunction_em(times2m)
pzqdoses2meg<-length(times2m_eg)
pzqdoses2mem<-length(times2m_em)

times3m_eg<-dosingfunction_eg(times3m)
times3m_em<-dosingfunction_em(times3m)
pzqdoses3meg<-length(times3m_eg)
pzqdoses3mem<-length(times3m_em)

times4m_eg<-dosingfunction_eg(times4m)
times4m_em<-dosingfunction_em(times4m)
pzqdoses4meg<-length(times4m_eg)
pzqdoses4mem<-length(times4m_em)

times6m_eg<-dosingfunction_eg(times6m)
times6m_em<-dosingfunction_em(times6m)
pzqdoses6meg<-length(times6m_eg)
pzqdoses6mem<-length(times6m_em)

times12m_eg<-dosingfunction_eg(times12m)
times12m_em<-dosingfunction_em(times12m)
pzqdoses12meg<-length(times12m_eg)
pzqdoses12mem<-length(times12m_em)


#####Creating dataframes of dosing events
alpha = 1 # PZQ coverage

events1m <- data.frame(
  var = c(rep("Mmd", pzqdoses1mem), rep("Mcd", pzqdoses1meg)),
  time = c(times1m_em, times1m_eg),
  value = c(rep((1-alpha),pzqdoses1mem+pzqdoses1meg)),
  method = rep("mult", pzqdoses1mem+pzqdoses1meg))
events1m<-events1m[!duplicated(events1m), ]
```

```
events2m <- data.frame(
  var = c(rep("Mmd", pzqdoses2mem), rep("Mcd", pzqdoses2meg)),
  time = c(times2m_em, times2m_eg),
  value = c(rep((1-alpha),pzqdoses2mem+pzqdoses2meg)),
  method = rep("mult", pzqdoses2mem+pzqdoses2meg))
events2m<-events2m[!duplicated(events2m), ]

events3m <- data.frame(
  var = c(rep("Mmd", pzqdoses3mem), rep("Mcd", pzqdoses3meg)),
  time = c(times3m_em, times3m_eg),
  value = c(rep((1-alpha),pzqdoses3mem+pzqdoses3meg)),
  method = rep("mult", pzqdoses3mem+pzqdoses3meg))
events3m<-events3m[!duplicated(events3m), ]

events4m <- data.frame(
  var = c(rep("Mmd", pzqdoses4mem), rep("Mcd", pzqdoses4meg)),
  time = c(times4m_em, times4m_eg),
  value = c(rep((1-alpha),pzqdoses4mem+pzqdoses4meg)),
  method = rep("mult", pzqdoses4mem+pzqdoses4meg))
events4m<-events4m[!duplicated(events4m), ]

events6m <- data.frame(
  var = c(rep("Mmd", pzqdoses6mem), rep("Mcd", pzqdoses6meg)),
  time = c(times6m_em, times6m_eg),
  value = c(rep((1-alpha),pzqdoses6mem+pzqdoses6meg)),
  method = rep("mult", pzqdoses6mem+pzqdoses6meg))
events6m<-events6m[!duplicated(events6m), ]

events12m <- data.frame(
  var = c(rep("Mmd", pzqdoses12mem), rep("Mcd", pzqdoses12meg)),
  time = c(times12m_em, times12m_eg),
  value = c(rep((1-alpha),pzqdoses12mem+pzqdoses12meg)),
  method = rep("mult", pzqdoses12mem+pzqdoses12meg))
events12m<-events12m[!duplicated(events12m), ]

##### Solving
out1m <-  dede(y=xstart, times=times, func=alaymodel, parms=parms,
events=list(data=events1m))
out2m <-  dede(y=xstart, times=times, func=alaymodel, parms=parms,
events=list(data=events2m))
out3m <-  dede(y=xstart, times=times, func=alaymodel, parms=parms,
events=list(data=events3m))
out4m <-  dede(y=xstart, times=times, func=alaymodel, parms=parms,
events=list(data=events4m))
out6m <-  dede(y=xstart, times=times, func=alaymodel, parms=parms,
events=list(data=events6m))
out12m <-  dede(y=xstart, times=times, func=alaymodel, parms=parms,
events=list(data=events12m))



####################Investigation of seasonal targeting of dosing
```

```
#Timing of monthly dosing
jandosepoints_seq<-fulldates1[seq(1,length(fulldates1),12)]
jandosepoints<-steadystate+jandosepoints_seq
jandosepoints<-jandosepoints[jandosepoints<=max(times)]
jandosepoints<-jandosepoints[-length(jandosepoints)]
febdosepoints_seq<-fulldates1[seq(2,length(fulldates1),12)]
febdosepoints<-steadystate+febdosepoints_seq
febdosepoints<-febdosepoints[febdosepoints<=max(times)]
mardosepoints_seq<-fulldates1[seq(3,length(fulldates1),12)]
mardosepoints<-steadystate+mardosepoints_seq
mardosepoints<-mardosepoints[mardosepoints<=max(times)]
aprdosepoints_seq<-fulldates1[seq(4,length(fulldates1),12)]
aprdosepoints<-steadystate+aprdosepoints_seq
aprdosepoints<-aprdosepoints[aprdosepoints<=max(times)]
maydosepoints_seq<-fulldates1[seq(5,length(fulldates1),12)]
maydosepoints<-steadystate+maydosepoints_seq
maydosepoints<-maydosepoints[maydosepoints<=max(times)]
jundosepoints_seq<-fulldates1[seq(6,length(fulldates1),12)]
jundosepoints<-steadystate+jundosepoints_seq
jundosepoints<-jundosepoints[jundosepoints<=max(times)]
juldosepoints_seq<-fulldates1[seq(7,length(fulldates1),12)]
juldosepoints<-steadystate+juldosepoints_seq
juldosepoints<-juldosepoints[juldosepoints<=max(times)]
augdosepoints_seq<-fulldates1[seq(8,length(fulldates1),12)]
augdosepoints<-steadystate+augdosepoints_seq
augdosepoints<-augdosepoints[augdosepoints<=max(times)]
sepdosepoints_seq<-fulldates1[seq(9,length(fulldates1),12)]
sepdosepoints<-steadystate+sepdosepoints_seq
sepdosepoints<-sepdosepoints[sepdosepoints<=max(times)]
octdosepoints_seq<-fulldates1[seq(10,length(fulldates1),12)]
octdosepoints<-steadystate+octdosepoints_seq
octdosepoints<-octdosepoints[octdosepoints<=max(times)]
novdosepoints_seq<-fulldates1[seq(11,length(fulldates1),12)]
novdosepoints<-steadystate+novdosepoints_seq
novdosepoints<-novdosepoints[novdosepoints<=max(times)]
decdosepoints_seq<-fulldates1[seq(12,length(fulldates1),12)]
decdosepoints<-steadystate+decdosepoints_seq
decdosepoints<-decdosepoints[decdosepoints<=max(times)]

#Creating dosing functions
jandose_eg<-dosingfunction_eg(jandosepoints)
jandose_em<-dosingfunction_em(jandosepoints)
febdose_eg<-dosingfunction_eg(febdosepoints)
febdose_em<-dosingfunction_em(febdosepoints)
mardose_eg<-dosingfunction_eg(mardosepoints)
mardose_em<-dosingfunction_em(mardosepoints)
aprdose_eg<-dosingfunction_eg(aprdosepoints)
aprdose_em<-dosingfunction_em(aprdosepoints)
maydose_eg<-dosingfunction_eg(maydosepoints)
maydose_em<-dosingfunction_em(maydosepoints)
jundose_eg<-dosingfunction_eg(jundosepoints)
jundose_em<-dosingfunction_em(jundosepoints)
juldose_eg<-dosingfunction_eg(juldosepoints)
juldose_em<-dosingfunction_em(juldosepoints)
```

```
augdose_eg<-dosingfunction_eg(augdosepoints)
augdose_em<-dosingfunction_em(augdosepoints)
sepdose_eg<-dosingfunction_eg(sepdosepoints)
sepdose_em<-dosingfunction_em(sepdosepoints)
octdose_eg<-dosingfunction_eg(octdosepoints)
octdose_em<-dosingfunction_em(octdosepoints)
novdose_eg<-dosingfunction_eg(novdosepoints)
novdose_em<-dosingfunction_em(novdosepoints)
decdose_eg<-dosingfunction_eg(decdosepoints)
decdose_em<-dosingfunction_em(decdosepoints)


monthdose_eg<-length(jandose_eg)
monthdose_em<-length(jandose_em)
monthdosedec_eg<-length(decdose_eg) # dec length shorter for Eg due
to PPP of 42 days. 12 days removed from end to keep within time
limits
monthdosedec_em<-length(decdose_em) #keep this in case change PPP
for Em in future simulations

#Creating event dataframes
janevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(jandose_em, jandose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
febevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(febdose_em, febdose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
marevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(mardose_em, mardose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
aprevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(aprdose_em, aprdose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
mayevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(maydose_em, maydose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
junevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(jundose_em, jundose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
julevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(juldose_em, juldose_eg),
```

```
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
augevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(augdose_em, augdose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
sepevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(sepdose_em, sepdose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
octevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(octdose_em, octdose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
novevents <- data.frame(
  var = c(rep("Mmd", monthdose_em), rep("Mcd", monthdose_eg)),
  time = c(novdose_em, novdose_eg),
  value = rep((1-alpha),monthdose_em+monthdose_eg),
  method = rep("mult", monthdose_em+monthdose_eg))
decevents <- data.frame(
  var = c(rep("Mmd", monthdosedec_em), rep("Mcd", monthdosedec_eg)),
  time = c(decdose_em, decdose_eg),
  value = rep((1-alpha),monthdosedec_em+monthdosedec_eg),
  method = rep("mult", monthdosedec_em+monthdosedec_eg))


#Solving for each dosing point (once yearly dosing)
outjandose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=janevents))
outfebdose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=febevents))
outmardose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=marevents))
outaprdose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=aprevents))
outmaydose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=mayevents))
outjundose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=junevents))
outjuldose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=julevents))
outaugdose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=augevents))
outsepdose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=sepevents))
outoctdose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=octevents))
outnovdose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=novevents))
outdecdose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=decevents))
```

```
#####Investigating the mean worm burden and egg density over one
year for different monthly dosing points
startpoint<-steadystate*2
endpoint<-steadystate*2+365

#Model with no dosing
out_year<-as.data.frame(out)
out_year<-subset(out_year,time>startpoint & time<endpoint)
out_emeggs<-sum(out_year$Gm)/length(out_year$Gm)
out_egeggs<-sum(out_year$Gc)/length(out_year$Gc)
out_emdogs<-sum(out_year$Mmd)/length(out_year$Mmd)
out_egdogs<-sum(out_year$Mcd)/length(out_year$Mcd)

#Models with once annual dosing
outjandose_year<-as.data.frame(outjandose)
outjandose_year<-subset(outjandose_year,time>startpoint &
time<endpoint)
jandose_emeggs<-sum(outjandose_year$Gm)/length(outjandose_year$Gm)
jandose_egeggs<-sum(outjandose_year$Gc)/length(outjandose_year$Gc)
jandose_emdogs<-sum(outjandose_year$Mmd)/length(outjandose_year$Mmd)
jandose_egdogs<-sum(outjandose_year$Mcd)/length(outjandose_year$Mcd)

outfebdose_year<-as.data.frame(outfebdose)
outfebdose_year<-subset(outfebdose_year,time>startpoint &
time<endpoint)
febdose_emeggs<-sum(outfebdose_year$Gm)/length(outfebdose_year$Gm)
febdose_egeggs<-sum(outfebdose_year$Gc)/length(outfebdose_year$Gc)
febdose_emdogs<-sum(outfebdose_year$Mmd)/length(outfebdose_year$Mmd)
febdose_egdogs<-sum(outfebdose_year$Mcd)/length(outfebdose_year$Mcd)

outmardose_year<-as.data.frame(outmardose)
outmardose_year<-subset(outmardose_year,time>startpoint &
time<endpoint)
mardose_emeggs<-sum(outmardose_year$Gm)/length(outmardose_year$Gm)
mardose_egeggs<-sum(outmardose_year$Gc)/length(outmardose_year$Gc)
mardose_emdogs<-sum(outmardose_year$Mmd)/length(outmardose_year$Mmd)
mardose_egdogs<-sum(outmardose_year$Mcd)/length(outmardose_year$Mcd)

outaprdose_year<-as.data.frame(outaprdose)
outaprdose_year<-subset(outaprdose_year,time>startpoint &
time<endpoint)
aprdose_emeggs<-sum(outaprdose_year$Gm)/length(outaprdose_year$Gm)
aprdose_egeggs<-sum(outaprdose_year$Gc)/length(outaprdose_year$Gc)
aprdose_emdogs<-sum(outaprdose_year$Mmd)/length(outaprdose_year$Mmd)
aprdose_egdogs<-sum(outaprdose_year$Mcd)/length(outaprdose_year$Mcd)

outmaydose_year<-as.data.frame(outmaydose)
outmaydose_year<-subset(outmaydose_year,time>startpoint &
time<endpoint)
maydose_emeggs<-sum(outmaydose_year$Gm)/length(outmaydose_year$Gm)
maydose_egeggs<-sum(outmaydose_year$Gc)/length(outmaydose_year$Gc)
maydose_emdogs<-sum(outmaydose_year$Mmd)/length(outmaydose_year$Mmd)
```

```
maydose_egdogs<-sum(outmaydose_year$Mcd)/length(outmaydose_year$Mcd)

outjundose_year<-as.data.frame(outjundose)
outjundose_year<-subset(outjundose_year,time>startpoint &
time<endpoint)
jundose_emeggs<-sum(outjundose_year$Gm)/length(outjundose_year$Gm)
jundose_egeggs<-sum(outjundose_year$Gc)/length(outjundose_year$Gc)
jundose_emdogs<-sum(outjundose_year$Mmd)/length(outjundose_year$Mmd)
jundose_egdogs<-sum(outjundose_year$Mcd)/length(outjundose_year$Mcd)

outjuldose_year<-as.data.frame(outjuldose)
outjuldose_year<-subset(outjuldose_year,time>startpoint &
time<endpoint)
juldose_emeggs<-sum(outjuldose_year$Gm)/length(outjuldose_year$Gm)
juldose_egeggs<-sum(outjuldose_year$Gc)/length(outjuldose_year$Gc)
juldose_emdogs<-sum(outjuldose_year$Mmd)/length(outjuldose_year$Mmd)
juldose_egdogs<-sum(outjuldose_year$Mcd)/length(outjuldose_year$Mcd)

outaugdose_year<-as.data.frame(outaugdose)
outaugdose_year<-subset(outaugdose_year,time>startpoint &
time<endpoint)
augdose_emeggs<-sum(outaugdose_year$Gm)/length(outaugdose_year$Gm)
augdose_egeggs<-sum(outaugdose_year$Gc)/length(outaugdose_year$Gc)
augdose_emdogs<-sum(outaugdose_year$Mmd)/length(outaugdose_year$Mmd)
augdose_egdogs<-sum(outaugdose_year$Mcd)/length(outaugdose_year$Mcd)

outsepdose_year<-as.data.frame(outsepdose)
outsepdose_year<-subset(outsepdose_year,time>startpoint &
time<endpoint)
sepdose_emeggs<-sum(outsepdose_year$Gm)/length(outsepdose_year$Gm)
sepdose_egeggs<-sum(outsepdose_year$Gc)/length(outsepdose_year$Gc)
sepdose_emdogs<-sum(outsepdose_year$Mmd)/length(outsepdose_year$Mmd)
sepdose_egdogs<-sum(outsepdose_year$Mcd)/length(outsepdose_year$Mcd)

outoctdose_year<-as.data.frame(outoctdose)
outoctdose_year<-subset(outoctdose_year,time>startpoint &
time<endpoint)
octdose_emeggs<-sum(outoctdose_year$Gm)/length(outoctdose_year$Gm)
octdose_egeggs<-sum(outoctdose_year$Gc)/length(outoctdose_year$Gc)
octdose_emdogs<-sum(outoctdose_year$Mmd)/length(outoctdose_year$Mmd)
octdose_egdogs<-sum(outoctdose_year$Mcd)/length(outoctdose_year$Mcd)

outnovdose_year<-as.data.frame(outnovdose)
outnovdose_year<-subset(outnovdose_year,time>startpoint &
time<endpoint)
novdose_emeggs<-sum(outnovdose_year$Gm)/length(outnovdose_year$Gm)
novdose_egeggs<-sum(outnovdose_year$Gc)/length(outnovdose_year$Gc)
novdose_emdogs<-sum(outnovdose_year$Mmd)/length(outnovdose_year$Mmd)
novdose_egdogs<-sum(outnovdose_year$Mcd)/length(outnovdose_year$Mcd)

outdecdose_year<-as.data.frame(outdecdose)
outdecdose_year<-subset(outdecdose_year,time>startpoint &
time<endpoint)
decdose_emeggs<-sum(outdecdose_year$Gm)/length(outdecdose_year$Gm)
```

```
decdose_egeggs<-sum(outdecdose_year$Gc)/length(outdecdose_year$Gc)
decdose_emdogs<-sum(outdecdose_year$Mmd)/length(outdecdose_year$Mmd)
decdose_egdogs<-sum(outdecdose_year$Mcd)/length(outdecdose_year$Mcd)

#Creating matrices
emeggs<-
data.frame(jandose_emeggs,febdose_emeggs,mardose_emeggs,aprdose_emeg
gs,maydose_emeggs,jundose_emeggs,juldose_emeggs,augdose_emeggs,sepdo
se_emeggs,octdose_emeggs,novdose_emeggs,decdose_emeggs)
emeggs<-as.matrix(emeggs)
colnames(emeggs)<-
c("jan","feb","mar","apr","may","jun","jul","aug","sep","oct","nov",
"dec")
egeggs<-
data.frame(jandose_egeggs,febdose_egeggs,mardose_egeggs,aprdose_egeg
gs,maydose_egeggs,jundose_egeggs,juldose_egeggs,augdose_egeggs,sepdo
se_egeggs,octdose_egeggs,novdose_egeggs,decdose_egeggs)
egeggs<-as.matrix(egeggs)
colnames(egeggs)<-
c("jan","feb","mar","apr","may","jun","jul","aug","sep","oct","nov",
"dec")
emdogs<-
data.frame(jandose_emdogs,febdose_emdogs,mardose_emdogs,aprdose_emdo
gs,maydose_emdogs,jundose_emdogs,juldose_emdogs,augdose_emdogs,sepdo
se_emdogs,octdose_emdogs,novdose_emdogs,decdose_emdogs)
emdogs<-as.matrix(emdogs)
colnames(emdogs)<-
c("jan","feb","mar","apr","may","jun","jul","aug","sep","oct","nov",
"dec")
egdogs<-
data.frame(jandose_egdogs,febdose_egdogs,mardose_egdogs,aprdose_egdo
gs,maydose_egdogs,jundose_egdogs,juldose_egdogs,augdose_egdogs,sepdo
se_egdogs,octdose_egdogs,novdose_egdogs,decdose_egdogs)
egdogs<-as.matrix(egdogs)
colnames(egdogs)<-
c("jan","feb","mar","apr","may","jun","jul","aug","sep","oct","nov",
"dec")

#Matrix subtraction
rel_emeggs<-out_emeggs-emeggs
rel_egeggs<-out_egeggs-egeggs
rel_emdogs<-out_emdogs-emdogs
rel_egdogs<-out_egdogs-egdogs

#Best first month is October
```

#The same approach is now run again, including a second month of dosing, and then again two more times – up to four doses per year. The code is not included here due to space constraints.
#This investigation found that October, January, September and November dosing was the best method of minimising E.m egg contamination

```
####################Running targeted intervention
```

```
onedosedates_eg<-octdose_eg
onedosedates_em<-octdose_em
twodosedates_eg<-c(jandose_eg,octdose_eg)
twodosedates_eg<-sort(twodosedates_eg)
twodosedates_em<-c(jandose_em,octdose_em)
twodosedates_em<-sort(twodosedates_em)
threedosedates_eg<-c(jandose_eg,sepdose_eg,octdose_eg)
threedosedates_eg<-sort(threedosedates_eg)
threedosedates_em<-c(jandose_em,sepdose_em,octdose_em)
threedosedates_em<-sort(threedosedates_em)
fourdosedates_eg<-c(jandose_eg,sepdose_eg,octdose_eg,novdose_eg)
fourdosedates_eg<-sort(fourdosedates_eg)
fourdosedates_em<-c(jandose_em,sepdose_em,octdose_em,novdose_em)
fourdosedates_em<-sort(fourdosedates_em)


onedoselength_eg<-length(onedosedates_eg)
onedoselength_em<-length(onedosedates_em)
onedoseevents <- data.frame(
  var = c(rep("Mmd", onedoselength_em), rep("Mcd",
onedoselength_eg)),
  time = c(onedosedates_em, onedosedates_eg),
  value = rep((1-alpha),onedoselength_em+onedoselength_eg),
  method = rep("mult", onedoselength_em+onedoselength_eg))

twodoselength_eg<-length(twodosedates_eg)
twodoselength_em<-length(twodosedates_em)
twodoseevents <- data.frame(
  var = c(rep("Mmd", twodoselength_em), rep("Mcd",
twodoselength_eg)),
  time = c(twodosedates_em, twodosedates_eg),
  value = rep((1-alpha),twodoselength_em+twodoselength_eg),
  method = rep("mult", twodoselength_em+twodoselength_eg))

threedoselength_eg<-length(threedosedates_eg)
threedoselength_em<-length(threedosedates_em)
threedoseevents <- data.frame(
  var = c(rep("Mmd", threedoselength_em), rep("Mcd",
threedoselength_eg)),
  time = c(threedosedates_em, threedosedates_eg),
  value = rep((1-alpha),threedoselength_em+threedoselength_eg),
  method = rep("mult", threedoselength_em+threedoselength_eg))

fourdoselength_eg<-length(fourdosedates_eg)
fourdoselength_em<-length(fourdosedates_em)
fourdoseevents <- data.frame(
  var = c(rep("Mmd", fourdoselength_em), rep("Mcd",
fourdoselength_eg)),
  time = c(fourdosedates_em, fourdosedates_eg),
  value = rep((1-alpha),fourdoselength_em+fourdoselength_eg),
  method = rep("mult", fourdoselength_em+fourdoselength_eg))
```

```
#####Running targeted dosing strategies
out_onedose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=onedoseevents))
out_twodose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=twodoseevents))
out_threedose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=threedoseevents))
out_fourdose <-  dede(y=xstart, times=times, func=alaymodel,
parms=parms, events=list(data=fourdoseevents))
```

A7. Manuscript submitted to the "Journal of Helminthology"

Journal of Helminthology, page 1 of 9 doi:10.1017/S0022149X15000590

Risk factors for *Echinococcus* coproantigen positivity in dogs from the Alay valley, Kyrgyzstan

A. Mastin[1]*, F. van Kesteren[1,2], P.R. Torgerson[3], I. Ziadinov[3], B. Mytynova[4], M.T. Rogan[1], T. Tursunov[4] and P.S. Craig[1]

[1]Cestode Zoonoses Research Group, School of Environment and Life Sciences, University of Salford, Salford, M5 4WT, UK: [2]Department of Psychology, University of Michigan, 1012 East Hall, 530 Church Street, Ann Arbor, MI, 48109-1043, USA: [3]Section of Epidemiology, Vetsuisse Faculty, University of Zürich, Switzerland: [4]Parasitology group, Kyrgyz Veterinary Research Institute, Togolok Moldo 60, Bishkek, Kyrgyzstan

(Received 6 March 2015; Accepted 6 June 2015)

Abstract

Echinococcosis, caused by the zoonotic cestodes *Echinococcus granulosus* (sensu lato) and *Echinococcus multilocularis*, is highly endemic in the Central Asian Republic of Kyrgyzstan, and is being identified increasingly as a public health problem, especially amongst pastoral communities. As domestic dogs are considered to be the main source of human infection, the identification of potential transmission pathways is of relevance when considering implementing an echinococcosis control scheme. The current report describes the results of an analytical study of canine *Echinococcus* coproantigen enzyme-linked immunosorbent assay (ELISA) prevalence in the Alay valley of southern Kyrgyzstan prior to the commencement of regular praziquantel dosing of dogs. A logistic regression model using a form of Bayes modal estimation was used to identify possible risk factors for coproantigen positivity, and the output was interpreted in a Bayesian context (posterior distributions of the coefficients of interest). The study found that sheepdogs had lower odds of coproantigen positivity, as did dogs in households with donkeys, where owners had knowledge of echinococcosis, and households which engaged in home slaughtering. Surprisingly, there was no evidence of an association between free roaming or previous praziquantel dosing and coproantigen positivity, as has been found in previous studies. Possible reasons for these findings are discussed in the context of the epidemiology of echinococcosis and potential intervention approaches.

Introduction

Human echinococcosis, caused by infection with the metacestode stage of cestodes of the genus *Echinococcus*, is an important public health concern in various parts of the world. Due to the parasite's complex life cycle and long period between infection and clinical signs in human hosts, accurate investigation of risk factors for human infection can be challenging. However, where domestic dogs act as a definitive host (most areas of *Echinococcus granulosus* endemicity and some areas of *E. multilocularis* endemicity (Craig *et al.*, 1992)), identification of risk factors for canine infection can provide useful information on potential human risk, and can be useful for designing and monitoring *Echinococcus* control schemes based on treatment of infection in dogs. Although *Echinococcus* spp. infection in dogs is asymptomatic, a number of diagnostic tools are available for diagnosis of current infection. Detection of coproantigens is of

*E-mail: a.mastin@salford.ac.uk

particular use and has been advised as the mainstay of surveillance of echinococcosis in endemic areas by the World Health Organization (WHO), World Organization for Animal Health (OIE) and Pan American Health Organization (PAHO) (WHO/OIE, 2001; Morel *et al.*, 2013).

As the transmission cycle of *Echinococcus* spp. may vary between locations, it is useful to identify risk factors specific to the particular transmission system in question. It is also useful to evaluate commonly identified risk factors from the wide range of studies that have been conducted worldwide (Otero-Abad & Torgerson, 2013). Coproantigen test results are often used to approximate canine infection status with *Echinococcus* spp. – classifying samples as coproantigen 'negative' or 'positive', according to their enzyme-linked immunosorbent assay (ELISA) optical density (OD) value in relation to a defined cut-off value. This will lead to some misclassification, which has been addressed in some studies by combining the results of purgation and coproPCR (Ziadinov *et al.*, 2008). Potential risk factors for infection can be classified according to a number of general transmission processes: factors associated with access to infected material (infected offal, in the case of *E. granulosus*, or small mammal intermediate hosts in the case of *E. multilocularis*); factors associated with variability in infection after ingestion of infectious material; and factors associated with removal of infection, such as a history of anthelmintic treatment. The most commonly identified risk factors are those relating to access to infected material, including access to offal or infected rodents (Bchir *et al.*, 1987; Parada *et al.*, 1995; Moro *et al.*, 1999; Y.H. Wang *et al.*, 2001; Shaikenov *et al.*, 2003; Budke *et al.*, 2005; Buishi *et al.*, 2005b, 2006; Q. Wang *et al.*, 2007, 2010; Dyachenko *et al.*, 2008; Guzel *et al.*, 2008; Huang *et al.*, 2008; Ziadinov *et al.*, 2008; Antolová *et al.*, 2009; Acosta-Jamett *et al.*, 2010; Mastin *et al.*, 2011; Reyes *et al.*, 2012).

Risk factors have also been identified at the dog level: in particular, dog type and age. Working dogs such as sheepdogs and farm dogs (Moro *et al.*, 1999; Shaikenov *et al.*, 2003; Buishi *et al.*, 2005a) have been found repeatedly to have higher odds of infection with *E. granulosus*, which likely relates to increased availability of, and access to, potentially infectious material. Younger dogs have also been found repeatedly to have higher odds of infection than older dogs (Sharifi & Zia-Ali, 1996; Buishi *et al.*, 2005a, 2006; Acosta-Jamett *et al.*, 2010; Inangolet *et al.*, 2010). This could result from differences in feeding behaviour between younger and older dogs. Alternatively, modelling approaches have suggested that this could represent immunity acquired in young dogs preventing parasite acquisition as the dog ages, which may be more apparent for *E. granulosus* than for *E. multilocularis* infection (Torgerson *et al.*, 2003; Torgerson, 2006). Finally, a lack of recent anthelmintic dosing has been commonly identified as a risk factor for canid infection (Parada *et al.*, 1995; Buishi *et al.*, 2005a; Huang *et al.*, 2008; Acosta-Jamett *et al.*, 2010).

It should be noted here that studies of risk factors for infection of domestic dogs (rather than foxes) with *E. multilocularis* are relatively uncommon. Domestic dog infection with *E. multilocularis* is most commonly only identified in particular pastoral communities, such as Tibetan communities in China or Kyrgyz communities in Kyrgyzstan (Budke *et al.*, 2005; Q. Wang *et al.*, 2007, 2010; Ziadinov *et al.*, 2008). However, recent work has also identified infection in domestic dogs (albeit often at very low levels) in central and eastern Europe (Deplazes *et al.*, 1999; Dyachenko *et al.*, 2008; Antolová *et al.*, 2009), and so evaluation of risk factors for *E. multilocularis* in domestic dogs may become increasingly common.

The current study investigates risk factors for canine *Echinococcus* spp. coproantigen positivity in the Alay valley in the Osh Oblast of Kyrgyzstan – an area of known high endemicity of human alveolar echinococcosis (Usubalieva *et al.*, 2013). A combination of Bayesian and frequentist strategies were utilized in order to identify and describe these risk factors.

Materials and methods

Study sites

In May 2012, four communities in the Alay valley of southern Kyrgyzstan were visited (Sary Mogol (39.68°N, 72.89°E), Taldu Suu (39.70°N, 72.98°E), Kashka Suu (39.64°N, 72.67°E) and Kara Kavak (39.66°N, 72.72°E)), prior to the commencement of a praziquantel-based pilot intervention for canine echinococcosis. A more detailed description of the study site can be found elsewhere (van Kesteren *et al.*, 2013). All occupied households in Sary Mogol, Taldu Suu and Kara Kavak, and a random selection of approximately 25% of the households in Kashka Suu were visited. For each household visited, a questionnaire was administered relating to details such as general demographics (age, sex, occupation of interviewee), dog ownership (number of dogs currently owned, management of these dogs), dog demographics (dog age, dog sex, dog weight) and perception of echinococcosis (recent administration of praziquantel to dogs, understanding of source of human echinococcosis). However, not all questions were answered by all interviewees. Of 692 households registered, a total of 329 individuals reported owning dogs, and a total of 388 dogs in total were registered. All questionnaire data were entered into Microsoft® Access.

Collection and examination of faecal samples

Faecal samples were taken from the available owned dogs. Samples were collected rectally when possible, but otherwise were taken from the ground around the homestead, with attempts made to match individual samples to individual dogs. Each sample was divided and each part stored in either saline buffer (phosphate-buffered saline (PBS) with Tween) or ethanol before transportation to the University of Salford, England, where they were stored at −80°C for a minimum of 5 days prior to testing (WHO/OIE, 2001). A total of 318 faecal samples were available. These were tested using a standardized sandwich coproantigen ELISA (Allan *et al.*, 1992), with modifications in that the capture and conjugate antibodies were raised from two different hyperimmune rabbit sera (van Kesteren *et al.*, 2015). All samples were tested using the same reagents in the same 'batch' period of no more than 4 days, with each sample tested in duplicate in adjacent wells. For

controls, a parasitologically defined faecal panel of necropsy dog samples was available as described in van Kesteren *et al.* (2015).

Data analysis

Initial data processing was conducted using Microsoft® Access 2010, and all further data processing and analysis was conducted using R version 3.1.1 (R Development Core Team, 2014). The difference in coproantigen ELISA OD between the two duplicates for each sample was calculated and the Studentized residuals of an intercept-only linear regression were inspected for outliers. A Bonferroni-corrected *t*-test was conducted using the 'outlierTest' function in the 'car' package for R (Fox & Weisberg, 2011), and any samples that gave a *P* value of 0.05 or less were removed from further analysis as possible failures of replication. Of the 318 faecal samples, 23 could not be matched to an individual questionnaire (due to illegible or damaged sample labels), but were retained in the model as the village of origin was known. Receiver-operator characteristic (ROC) curve analysis (Zweig & Campbell, 1993; Greiner, Pfeiffer & Smith, 2000) was used on a panel of parasitologically defined dog faecal samples taken from Xinjiang province in China (van Kesteren *et al.*, 2015). The Youden index approach, i.e. maximization of both test sensitivity and specificity (Youden, 1950) was used to determine the optimal cut-off point. The resultant cut-off point (OD 0.07635) gave an estimated test sensitivity of 96% and specificity of 83%, based upon the panel evaluated.

Prior to analysis, the number of variables with missing data was assessed. All variables with more than 250 missing data points were removed from further analysis, as were those categorical variables with fewer than five outcomes in any single category. This process left a total of 41 variables to be investigated (table 1).

Initial analysis utilized simple non-parametric univariable methods (Fisher's exact test, chi-square test or Mann–Whitney *U*-test) to identify those variables with some evidence of association with coproantigen status (using a *P* value of less than 0.3 to suggest some association). Any collinear parameters identified at this stage were reduced to one parameter based on the *P* value obtained. Twelve variables were selected for inclusion in the preliminary regression model.

All variables identified in the previous stage were added to a Bayesian logistic regression model, using the 'bayesglm' procedure in the 'arm' package (Gay *et al.*, 2008; Gay & Su, 2014) for R. This procedure incorporates 'vague' priors based upon Cauchy distributions into the regression model using data augmentation techniques (Cole *et al.*, 2012). The log posterior density [$\log p(\beta, \sigma|y)$] is maximized using an iterative process combining weighted least squares (Nelder & Wedderburn, 1972; Fox, 2008) and an expectation-maximization (EM) algorithm (Dempster *et al.*, 1977) in order to obtain parameter estimates. In line with the output of the 'glm' procedure upon which it is based (R Development Core Team, 2014), coefficient estimates are provided as point estimates along with their standard errors.

Model selection was based upon a manual stepwise removal process according to their Wald test *P* values, and a likelihood ratio test was used to identify possible contribution to the model (with a *P* value of 0.1 or less suggesting some contribution). Confounding was assessed by monitoring coefficients of other variables before and after variable removal, with a change of 30% or more suggestive of possible confounding. Where coefficients were less than 0.001 in magnitude, an absolute change in the magnitude of the coefficient of 0.001 or more was used to indicate a potentially confounding effect. Following this process, all plausible interactions between the remaining variables, and any quadratic and cubic trends in any continuous variables, were assessed using a likelihood ratio test, with a *P* value of 0.05 or less suggestive of a significant effect. Model diagnostics were conducted using residual plots and influence plots, and observations removed as appropriate. Variables were

Table 1. Variables considered in the risk factor modelling process; livestock ownership was evaluated using both a dichotomous variable (presence/absence) and a continuous variable (number of animals owned).

Variable type	Variables
Village	Village
Animal ownership	Current number of dogs
	Number of dogs owned in past 10 years
	Sheep
	Goats
	Cattle
	Horses
	Yaks
	Donkeys
Dog demographics	Age
	Size (small/medium/large)
	Weight
	Sex
	Used for hunting
	Guard dog
	Pet dog
	Sheepdog
Dog management/behaviour	Wormed in past 6 months
	Percentage of time spent free roaming
	Known to eat rodents
	Fed meat
	Fed offal
	Chained at all
	Handled by adults from the household
	Handled by children from the household
	Handled by friends of the family
	Not handled
	Visited pasture previous year
	Will visit pasture this year
Animal slaughter	Home slaughter, own
	Home slaughter, others
	Organs thrown away
	Organs given to dogs
	Organs buried
Perceived source of human echinococcosis	Dogs
	Cats
	Livestock
	Unknown

then removed sequentially from the final model, using a likelihood ratio test of 0.05 or less to suggest model contribution. The fit of the final model was assessed using a likelihood ratio goodness-of-fit test.

Posterior simulation using the 'sim' procedure in the 'arm' package was used with 10,000 iterations in order to approximate Bayesian posterior estimates of coefficients in the final model. The simulation output was exponentiated in order to obtain estimates of the posterior distribution of odds ratio estimates. These distributions were then summarized using the mode (as estimated from the kernel density, according to Parzen (1962), using the 'modeest' package in R (Poncet, 2012)) and the highest density intervals (HDI) (using the 'HPDinterval' procedure in the 'coda' package (Plummer *et al.*, 2006)).

Results

A total of 78 out of 318 canine faecal samples (25%) tested coproantigen positive using the cut-off as calculated from ROC curves (table 2). The distribution of OD values from these samples showed a clear right skew, as shown in fig. 1. There was no evidence of a difference in coproantigen prevalences between villages ($P = 0.5$) (table 2).

Ten categorical variables were found to be associated with coproantigen status at the end of the first stage of analysis (P value < 0.3). None of the continuous variables were found to be associated. Of the categorical variables, two variables relating to donkey ownership were identified as associated with coproantigen status: one based upon a dichotomous classification of donkey ownership, and one where donkey ownership was categorized according to the number of donkeys owned. As the latter variable was found to have a higher P value than the former, this was removed from further analysis. The variables found to be associated with coproantigen status in the first stage of analysis are shown in table 3, along with the associated P values.

At the end of the second stage of analysis, home slaughter, knowledge of hydatid source in dog, sheepdog ownership and donkey ownership variables were found to be associated with coproantigen status, with no evidence of interaction between variables. In the final stage of model selection, likelihood ratio tests found all four remaining variables to be significant at $P < 0.05$. The likelihood ratio goodness-of-fit test gave a P value of 0.27, suggesting a reasonable model fit. The exponents of the simulated posterior estimates (which describe the change in the log odds of infection associated with each variable) were calculated and the resultant distributions summarized in order to estimate the odds ratios of the effect of the different variables on coproantigen positivity (table 4).

Table 2. Numbers of canine faecal samples analysed from the four study villages in the Alay valley together with point estimates of the *Echinococcus* coproprevalence (%). Confidence intervals are not shown as the data were collected by census from all villages with the exception of Kashka Suu.

Village	Proportion of total samples	Number of samples	Coproprevalence (%)
Sary Mogol	0.49	155	27
Taldu Suu	0.27	86	19
Kara Kabak	0.13	42	24
Kashka Suu	0.11	35	29

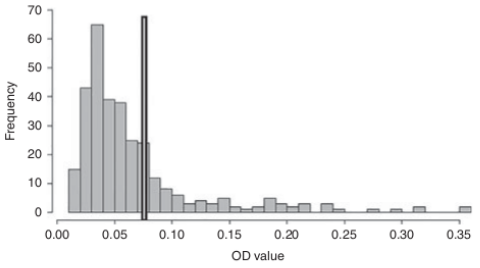


Fig. 1. The distribution of coproELISA OD values for all 318 dog faecal samples tested; cut-off for positivity at 0.07635 (bold vertical bar).

Discussion

Using the *Echinococcus* ELISA coproantigen cut-off described here, the overall canine coproantigen prevalence in the study area was estimated as 25%. Due to the limitations in the coproantigen test, especially in the case of low worm burdens (Allan & Craig, 2006), the true prevalence is likely to differ from the coproantigen prevalence, and it is for this reason that coproantigen prevalence estimates are not presented to any higher precision here. The aim of the current study was not to estimate the prevalence of infection, but to identify possible risk factors for infection. Four risk factors were found to be associated with reduced odds of coproantigen positivity: ownership of donkeys, description of the dog as a sheepdog, knowledge that dogs are a source of human echinococcosis and a lack of home slaughtering in the household.

The logistic regression modelling framework used for the current study utilized a combination of frequentist and Bayesian methodologies: a Bayesian prior was incorporated into the model in order to ensure model identifiability even when data were sparse, as was the case with home slaughter and owner knowledge of risk factors for human echinococcosis. Model selection was based on frequentist interpretation of coefficient estimates, but as the model likelihoods incorporated the prior, the likelihood ratio test used for model comparison could be considered partly Bayesian. Initial interpretation of the final model used both approaches, but the final conclusions were based upon Bayesian posterior estimates. The distinction between these approaches is of importance in terms of the communication of model selection and the final conclusions. The use of well-known frequentist strategies, such as likelihood ratio testing, ensures that the model selection process can be understood by people not familiar with Bayesian methods, but the final interpretation of the model output

Table 3. *Echinococcus* coproprevalences (%) in dogs relative to variables identified during univariable analysis; $N = 295$ respondents.

Variable	Negative respondents	Positive respondents	Coproprevalence amongst dogs of negative respondents (%)	Coproprevalence amongst dogs of positive respondents (%)	*P* value (all <0.3)
Hunting dog	227	68	26	18	0.21
Home slaughter practised	14	281	7	25	0.23
Organs thrown away	200	95	22	28	0.29
Dogs perceived source of hydatid	283	12	25	0	0.10
Cats perceived source of hydatid	288	7	24	0	0.29
Sheepdog	251	44	26	11	0.05
Dog handled by adults	139	156	21	27	0.28
Owns donkeys	165	130	29	18	0.03

in a Bayesian setting makes the model output conceptually easier to understand.

Donkey ownership was found to be associated with reduced odds of owned dog coproantigen positivity. This is an unexpected finding, and has not been reported in any previous studies. Donkey ownership may reflect a socio-economic factor: for example, people with donkeys may have more disposable income than those without, which could affect the risk of canine infection as dogs from higher-income households may be better fed and less likely to scavenge. Alternatively, donkey ownership may relate to spatial factors. Donkeys were commonly used for water transportation, and therefore households with donkeys may tend to be located further from a water source than those without donkeys. Proximity to water has been suggested repeatedly to be a factor of importance in the sylvatic *E. multilocularis* transmission cycle. A study of vulpine infection with *E. multilocularis* found that infected foxes tended to be located near water sources (Staubach *et al.*, 2001), which may be suggestive of differences in suitable intermediate host habitats or may indicate spatial aggregation of infection in intermediate hosts due to prolonged egg survival in water (Hansen *et al.*, 2003, 2004). Also, a study of *E. granulosus* infection of dogs living around eight abattoirs in Lima, Peru, found that dogs from abattoirs located close to the river were more likely to be infected with *E. granulosus* (Reyes *et al.*, 2012), which was suggested to potentially result from infected offal being discarded into the river. Further work to investigate spatial patterns of coproantigen positivity within the communities would be interesting – with a particular focus on *E. multilocularis* infection.

Sheepdogs were found to have lower odds of coproantigen positivity than other dogs. This is contrary to previous studies, which have routinely identified sheepdogs as having a higher probability of coproantigen positivity or infection than non-sheepdogs (Moro *et al.*, 1999; Buishi *et al.*, 2005a). Similarly, farm dogs have also been identified as having a higher probability of infection than village dogs in Kazakhstan (Shaikenov *et al.*, 2003; Torgerson *et al.*, 2003). In the study area, there was a clear distinction between 'sheepdogs' and 'pet' or 'guard' dogs (with the latter two classifications apparently being used interchangeably) (van Kesteren *et al.*, 2013). It is likely that sheepdogs were used for herding and guarding livestock during their seasonal and daily movements to pasture, whereas pet and guard dogs likely represent those dogs that remained in the village. In studies where sheepdogs were found to have higher odds of infection, this is usually considered to be due to potential access to infected offal, and therefore exposure to *E. granulosus*. However, *E. multilocularis* infection may be of particular importance in the current study area, for which contact with livestock would not be expected to be of importance. Therefore, it is possible that this association with dog type represents a spatial risk factor, with dogs based mostly within the village having a higher probability of infection than those sheepdogs that spend more time outside the village.

Both knowledge that dogs were a risk factor for human echinococcosis and lack of home slaughtering were found to be associated with a reduced probability of canine infection when assessed using the likelihood ratio test. However, the confidence intervals of the unadjusted model coefficients crossed the threshold of zero in both cases. In the case of the home slaughtering variable, this 'non-significant' effect persisted in the final HDI estimates (table 4). These issues are likely the result of a scarcity of positive or negative responses for these variables (table 3), and as these associations have been reported previously in the literature, they will be discussed further here as potential risk factors.

In previous studies, knowledge of cystic echinococcosis has been identified to be a significant risk factor for canine coproantigen status (Buishi *et al.*, 2005a; Huang *et al.*, 2008). Although no explicit distinction was made between cystic echinococcosis and alveolar echinococcosis in the question asked in the current study, it is likely that people with knowledge of echinococcosis are less likely to

Table 4. Odds ratios of the variables included in the final regression model.

Variable	Odds ratio (mode)	95% highest density interval
Home slaughter practised	2.04	0.18–18.46
Dogs perceived source of human hydatid disease	0.03	0.0005–0.95
Sheepdog	0.27	0.09–0.77
Owns donkeys	0.46	0.24–0.77

engage in practices such as feeding of infected offal which could facilitate transmission to dogs. This finding therefore may demonstrate some potential benefits of education campaigns as an adjunct to an echinococcosis control scheme.

Home slaughter has been found, in previous studies, to be positively associated with coproantigen positivity (Buishi *et al*., 2005b; Acosta-Jamett *et al.*, 2010, 2014), and although it is likely that almost all households slaughtered animals at some point, this association is plausible. As home slaughter likely increases the risk of feeding unwanted infected offal to dogs, this association would be expected to represent *E. granulosus* rather than *E. multilocularis* infection.

The lack of any identified association between reported dog dosing history and current status in the current study may result from the fact that information on dosing history was aggregated over a period of 6 months. Since praziquantel has no residual effect after administration, dosed dogs can become re-infected immediately after dosing. Recall bias amongst owners is also likely to be present: people who have not dosed their dogs recently may report they have, and people who have dosed recently may report that they have not, which would tend to reduce any coefficient estimates towards zero. Free roaming, which is probably the most commonly identified risk factor for echinococcosis in dogs (Parada *et al*., 1995; Budke *et al*., 2005; Buishi *et al.*, 2005b, 2006; Guzel *et al*., 2008; Huang *et al.*, 2008; Ziadinov *et al.*, 2008; Antolová *et al.*, 2009; Mastin *et al*., 2011), was also not found to be associated with test status. In the Alay valley, most dogs were free to roam throughout the village, with only 28/288 dogs (10%) reported to be chained at all, and therefore a similar lack of power to that described above would be expected for this variable. However, these results are suggestive that even chained dogs are gaining access to infected material, possibly through purposeful feeding of infected offal or resulting from occasional release from restraint.

One issue with any risk factor study based on identification of 'significant' risk factors from a large number of possible variables is that as the number of variables considered is increased, the probability of type I errors (i.e. finding a 'significant' association when this is not truly the case) also increases. In total, 41 variables were assessed in the current study, meaning that with an alpha error of 0.05, approximately two associations would be expected to be identified as 'significant' due to random variation alone. Model selection and evaluation strategies based upon information theoretic measures may reduce this problem, and would be a useful avenue for further exploration (Burnham & Anderson, 2002). Three other major considerations are particular to this study, and should be considered when interpreting the conclusions. These are the limitations in coproantigen test sensitivity and specificity, the lack of any differentiation between *E. granulosus* (sensu lato) and *E. multilocularis*, and the fact that relatively few faecal samples were matched conclusively to individual dogs.

As alluded to above, most studies of echinococcosis based upon coproantigen data classify all samples as 'negative' or 'positive' based upon a single ELISA optical density (OD) cut-off. This strategy will generally result in some misclassification: in particular, in the case of animals with low *Echinococcus* burdens (i.e. imperfect sensitivity), and animals infected with other taeniid cestodes (i.e. imperfect specificity). Therefore, estimates of the coproantigen prevalence are likely to differ from the true prevalence of infection. While this will tend to affect the accuracy of any prevalence estimates, it may be less of an issue in the case of analytical studies, where the intention is to identify risk factors for infection. Despite this, further work is planned to reduce the reliance on coproantigen status and instead attempt to model the true infection status. This has been achieved in recent studies by explicitly accounting for diagnostic test limitations (Ziadinov *et al*., 2008), or by avoiding the dichotomization of ELISA results completely (Choi *et al.*, 2006).

The major limitation in the current study is the lack of *Echinococcus* species discrimination. Previous work has shown that both *E. granulosus* sensu lato (*E. granulosus* G1 and *Echinococcus canadensis* G6) and *E. multilocularis* are present in dogs in the Alay valley (van Kesteren *et al.*, 2013), although the human health problem, to date, appears to be due mainly to *E. multilocularis* (Usubalieva *et al.*, 2013). While all faecal samples in the current study were collected and tested for faecal *Echinococcus* DNA, these results were not included here due to the difficulties in combining these results with the coproantigen ELISA results in a useful way. Further work will be undertaken to investigate risk factors for infection as identified by polymerase chain reaction (PCR), and it is hoped that methods of combining results obtained from these different testing methodologies, as has been achieved in other studies (Ziadinov *et al.*, 2008; Hartnack *et al*., 2013), will be developed for co-endemic situations in due course. Development of coproantigen ELISA tests that are specific for a variety of different species and strains of *Echinococcus* (WHO/OIE, 2001), or the development of alternative tests that would allow species discrimination in a surveillance setting, would be of great use. Examples of the latter are single-tube, isothermal DNA amplification techniques, such as loop-mediated isothermal amplification (LAMP) – which has already been developed for *E. granulosus* G1 and *E. multilocularis* (Salant *et al.*, 2012; Ni *et al*., 2014a, b) – or recombinase DNA polymerase amplification (RPA) (Piepenburg *et al*., 2006).

Finally, despite efforts to sample dogs per rectum whenever possible, most of the samples were collected from the ground around the household, and therefore cannot be definitively matched to individual dogs (or even individual households), due to the free-roaming behaviour of the dogs. Attempts were always made to involve the owners in order to identify faeces passed by the dog in question, but this was not always possible. Therefore, it is highly likely that some of the samples analysed in the current study are not from the dogs for which questionnaire data were collected. As the correct identification of an individual dog is unlikely to be associated with the coproantigen status of that dog, there is no reason to believe that this will result in directional bias, but this sampling strategy would be expected to reduce the study power. This problem of identifying samples from individual dogs would also be expected to be a problem in *Echinococcus* surveillance schemes, where ground samples are collected from free-roaming dogs.

Further work to determine an optimal strategy to deal with this problem is planned.

In conclusion, the unexpected finding that sheepdogs and dogs from households that owned donkeys appeared to have lower odds of coproantigen positivity may be suggestive of a spatial component to transmission in these communities, and will be explored further in future work. A lack of owner knowledge of echinococcosis was found to be associated with higher odds of coproantigen positivity, as was home slaughter. Although it was not possible to distinguish between *E. granulosus*, *E. canadensis* and *E. multilocularis* infection (all three of which appear to be co-endemic in the study area), these risk factors have previously only been found to be associated with *E. granulosus* infection. As well as investigation of potential spatial factors associated with the risk of infection, further work will attempt to identify the species of *Echinococcus* present and evaluate risk factors for these different species. It is also hoped that the results of this and other studies will assist in the development of a comprehensive surveillance strategy including aspects of sampling, coproantigen testing and coproPCR testing, which facilitate the implementation and evaluation of echinococcosis control schemes in Kyrgyzstan and similar areas.

Acknowledgements

The authors would like to thank Akjol Tagaibekov and Turdumamat Sultanov for fieldwork assistance and hospitality, and Belgees Boufana for laboratory assistance.

Financial support

This work was supported by the Wellcome Trust (grant number 094325/Z/10/Z).

Conflict of interest

None.

References

Acosta-Jamett, G., Cleaveland, S., Bronsvoort, B.M.D., Cunningham, A.A., Bradshaw, H. & Craig, P.S. (2010) *Echinococcus granulosus* infection in domestic dogs in urban and rural areas of the Coquimbo region, north-central Chile. *Veterinary Parasitology* **169**, 117–122.

Acosta-Jamett, G., Weitzel, T., Boufana, B., Adones, C., Bahamonde, A., Abarca, K., Craig, P.S. & Reiter-Owona, I. (2014) Prevalence and risk factors for echinococcal infection in a rural area of northern Chile: a household-based cross-sectional study. *PLoS Neglected Tropical Diseases* **8**, e3090.

Allan, J.C. & Craig, P.S. (2006) Coproantigens in taeniasis and echinococcosis. *Parasitology International* **55**, S75–S80.

Allan, J.C., Craig, P.S., Garcia Noval, J., Mencos, F., Liu, D., Wang, Y., Wen, H., Zhou, P., Stringer, R., Rogan, M.T. & Zeyhle, E. (1992) Coproantigen detection for immunodiagnosis of echinococcosis and taeniasis in dogs and humans. *Parasitology* **104**, 347–355.

Antolová, D., Reiterová, K., Miterpáková, M., Dinkel, A. & Dubinsky, P. (2009) The first finding of *Echinococcus multilocularis* in dogs in Slovakia: an emerging risk for spreading of infection. *Zoonoses and Public Health* **56**, 53–58.

Bchir, A., Jaiem, A., Jemmali, M., Rousset, J.J., Gaudebout, C. & Larouze, B. (1987) Possible existence of an urban cycle of *Echinococcus granulosus* in central Tunisia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **81**, 650.

Budke, C.M., Campos-Ponce, M., Qian, W. & Torgerson, P.R. (2005) A canine purgation study and risk factor analysis for echinococcosis in a high endemic region of the Tibetan plateau. *Veterinary Parasitology* **127**, 43–49.

Buishi, I.E., Njoroge, E.M., Bouamra, O. & Craig, P.S. (2005a) Canine echinococcosis in northwest Libya: assessment of coproantigen ELISA, and a survey of infection with analysis of risk-factors. *Veterinary Parasitology* **130**, 223–232.

Buishi, I.E., Walters, T., Guildea, Z., Craig, P.S. & Palmer, S. (2005b) Reemergence of canine *Echinococcus granulosus* infection, Wales. *Emerging Infectious Diseases* **11**, 568–571.

Buishi, I.E., Njoroge, E., Zeyhle, E., Rogan, M.T. & Craig, P.S. (2006) Canine echinococcosis in Turkana (north-western Kenya): a coproantigen survey in the previous hydatid-control area and an analysis of risk factors. *Annals of Tropical Medicine and Parasitology* **100**, 601–610.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: A practical information-theoretic approach*. 2nd edn. New York, Springer.

Choi, Y.-K., Johnson, W.O. & Thurmond, M.C. (2006) Diagnosis using predictive probabilities without cut-offs. *Statistics in Medicine* **25**, 699–717.

Cole, S.R., Chu, H., Greenland, S., Hamra, G. & Richardson, D.B. (2012) Bayesian posterior distributions without Markov chains. *American Journal of Epidemiology* **175**, 368–375.

Craig, P.S., Liu, D., Shi, D., Macpherson, C.N.L., Barnish, G., Reynolds, D., Gottstein, B. & Wang, Z. (1992) A large focus of alveolar echinococcosis in central China. *Lancet* **340**, 826–831.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* **39**, 1–38.

Deplazes, P., Alther, P., Tanner, I., Thompson, R.C. & Eckert, J. (1999) *Echinococcus multilocularis* coproantigen detection by enzyme-linked immunosorbent assay in fox, dog, and cat populations. *Journal of Parasitology* **85**, 115–121.

Dyachenko, V., Pantchev, N., Gawlowska, S., Vrhovec, M.G. & Bauer, C. (2008) *Echinococcus multilocularis* infections in domestic dogs and cats from Germany and other European countries. *Veterinary Parasitology* **157**, 244–253.

Fox, J. (2008) Generalized linear models. *Applied regression analysis and generalized linear models*. 2nd edn.

pp. 379–424. Thousand Oaks, California, Sage Publications.

Fox, J. & Weisberg, S. (2011) *An R companion to applied regression*. 2nd edn. Thousand Oaks CA, Sage.

Gay, N.J. & Su, Y.-S. (2014) arm: data analysis using regression and multilevel/hierarchical models. R package version 1.7-07. Available at http://cran.r-project.org/package=arm (accessed 2 March 2015).

Gay, N.J., Jakulin, A., Pittau, M.G. & Su, Y.-S. (2008) A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* **2**, 1360–1383.

Greiner, M., Pfeiffer, D.U. & Smith, R.D. (2000) Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine* **45**, 23–41.

Guzel, M., Yaman, M., Koltas, I., Demirkazik, M. & Aktas, H. (2008) Detection of *Echinococcus granulosus* coproantigens in dogs from Antakya Province, Turkey. *Helminthologia* **45**, 150–153.

Hansen, F., Tackmann, K., Jeltsch, F., Wissel, C. & Thulke, H.H. (2003) Controlling *Echinococcus multilocularis* – ecological implications of field trials. *Preventive Veterinary Medicine* **60**, 91–105.

Hansen, F., Jeltsch, F., Tackmann, K., Staubach, C. & Thulke, H.H. (2004) Processes leading to a spatial aggregation of *Echinococcus multilocularis* in its natural intermediate host *Microtus arvalis*. *International Journal for Parasitology* **34**, 37–44.

Hartnack, S., Budke, C.M., Craig, P.S., Jiamin, Q., Boufana, B., Campos-Ponce, M. & Torgerson, P.R. (2013) Latent-class methods to evaluate diagnostics tests for *Echinococcus* infections in dogs. *PLoS Neglected Tropical Diseases* **7**, e2068.

Huang, Y., Heath, D.D., Yang, W., Qiu, J., Chen, X., Yang, Y., Wang, Q., Li, T.Y., Xiao, Y., Qiu, D.C., Xiao, N., Chen, F., Ge, S. & Se, D. (2008) Epidemiology and risk factor analysis for canine echinococcosis in a Tibetan pastoral area of Sichuan. *Chinese Journal of Parasitology and Parasitic Diseases* **26**, 245–252.

Inangolet, F., Biffa, D., Opuda-Asibo, J., Oloya, J. & Skjerve, E. (2010) Distribution and intensity of *Echinococcus granulosus* infections in dogs in Moroto District, Uganda. *Tropical Animal Health and Production* **42**, 1451–1457.

Mastin, A., Brouwer, A., Fox, M., Craig, P.S., Guitián, J., Li, W. & Stevens, K. (2011) Spatial and temporal investigation of *Echinococcus granulosus* coproantigen prevalence in farm dogs in South Powys, Wales. *Veterinary Parasitology* **178**, 100–107.

Morel, N., Lassabe, G., Elola, S., Bondad, M., Herrera, S., Mari, C., Last, J.A., Jensen, O. & Gonzalez-Sapienza, G. (2013) A monoclonal antibody-based copro-ELISA kit for canine echinococcosis to support the PAHO effort for hydatid disease control in South America. *PLoS Neglected Tropical Diseases* **7**, e1967.

Moro, P.L., Bonifacio, N., Gilman, R.H., Lopera, L., Silva, B., Takumoto, R., Verastegui, M. & Cabrera, L. (1999) Field diagnosis of *Echinococcus granulosus* infection among intermediate and definitive hosts in an endemic focus of human cystic echinococcosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93**, 611–615.

Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of the Royal Statistical Society A* **135**, 370–384.

Ni, X.-W., McManus, D.P., Lou, Z.-Z., Yang, J.-F., Yan, H.-B., Li, L., Li, H.-M., Liu, Q.-Y., Li, C.-H., Shi, W.-G., Fan, Y.-L., Liu, X., Cai, J.-Z., Lei, M.-T., Fu, B.-Q., Yang, Y.-R. & Jia, W.-Z. (2014a) A comparison of loop-mediated isothermal amplification (LAMP) with other surveillance tools for *Echinococcus granulosus* diagnosis in canine definitive hosts. *PLoS One* **9**, e100877.

Ni, X.-W., McManus, D.P., Yan, H., Yang, J., Lou, Z.-Z., Li, H.-M., Li, L., Lei, M.-T., Cai, J.-Z., Fan, Y.-L., Li, C.-H., Liu, Q.-Y., Shi, W.-G., Liu, X., Zheng, Y., Fu, B.-Q., Yang, Y. & Jia, W.-Z. (2014b) Loop-mediated isothermal amplification (LAMP) assay for the identification of *Echinococcus multilocularis* infections in canine definitive hosts. *Parasites & Vectors* **7**, 254.

Otero-Abad, B. & Torgerson, P.R. (2013) A systematic review of the epidemiology of echinococcosis in domestic and wild animals. *PLoS Neglected Tropical Diseases* **7**, e2249.

Parada, L., Cabrera, P., Burges, C., Acuna, A., Barcelona, C., Laurenson, M.K., Gulland, F.M., Aqulla, J., Parietti, S., Paolillo, E. & Botta, B. (1995) *Echinococcus granulosus* infections of dogs in the Durazno region of Uruguay. *Veterinary Record* **136**, 389–391.

Parzen, E. (1962) On estimation of a probability density function and mode. *Annals of Mathematical Statistics* **33**, 1065–1076.

Piepenburg, O., Williams, C.H., Stemple, D.L. & Armes, N.A. (2006) DNA detection using recombination proteins. *PLoS Biology* **4**, 1115–1121.

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

Poncet, P. (2012) modeest: mode estimation. R package version 2.1. Available at http://cran.r-project.org/package=modeest (accessed 2 March 2015).

R Development Core Team. (2014) *R: a language and environment for statistical computing*. Vienna, Austria, R Foundation for Statistical Computing. Available at http://www.r-project.org/ (accessed 2 March 2015).

Reyes, M.M., Taramona, C.P., Saire-Mendoza, M., Gavidia, C.M., Barron, E., Boufana, B., Craig, P.S., Tello, L., García, H.H. & Santivañez, S.J. (2012) Human and canine echinococcosis infection in informal, unlicensed abattoirs in Lima, Peru. *PLoS Neglected Tropical Diseases* **6**, e1462.

Salant, H., Abbasi, I. & Hamburger, J. (2012) The development of a loop-mediated isothermal amplification method (LAMP) for *Echinococcus granulosus* coprodetection. *American Journal of Tropical Medicine and Hygiene* **87**, 883–887.

Shaikenov, B.S., Torgerson, P.R., Usenbayev, A.E., Baitursinov, K.K., Rysmukhambetova, A.T., Abdybekova, A.M. & Karamendin, K.O. (2003) The changing epidemiology of echinococcosis in Kazakhstan due to transformation of farming practices. *Acta Tropica* **85**, 287–293.

Sharifi, I. & Zia-Ali, N. (1996) The present status and intensity of *Echinococcus granulosus* infection in 391 stray

dogs in rural and urban areas of the city of Kerman, Iran. *Iranian Journal of Public Health* **25**, 13–20.

Staubach, C., Thulke, H.H., Tackmann, K., Hugh-Jones, M. & Conraths, F.J. (2001) Geographic information system-aided analysis of factors associated with the spatial distribution of *Echinococcus multilocularis* infections of foxes. *American Journal of Tropical Medicine and Hygiene* **65**, 943–948.

Torgerson, P.R. (2006) Canid immunity to *Echinococcus* spp.: impact on transmission. *Parasite Immunology* **28**, 295–303.

Torgerson, P.R., Shaikenov, B.S., Rysmukhambetova, A.T., Ussenbayev, A.E., Abdybekova, A.M. & Burtisurnov, K.K. (2003) Modelling the transmission dynamics of *Echinococcus granulosus* in dogs in rural Kazakhstan. *Parasitology* **126**, 417–424.

Usubalieva, J., Minbaeva, G., Ziadinov, I., Deplazes, P. & Torgerson, P.R. (2013) Human alveolar echinococcosis in Kyrgyzstan. *Emerging Infectious Diseases* **19**, 1095–1097.

van Kesteren, F., Mastin, A., Mytynova, B., Ziadinov, I., Boufana, B., Torgerson, P.R., Rogan, M.T. & Craig, P.S. (2013) Dog ownership, dog behaviour and transmission of *Echinococcus* spp. in the Alay Valley, southern Kyrgyzstan. *Parasitology* **140**, 1674–1684.

van Kesteren, F., Qi, X., Jiang, T., Feng, X., Mastin, A., Craig, P.S., Vuitton, D.A., Xinyu, D., Chu, X., Jinlong, Z. & Hao, W. (2015) Independent evaluation of a canine echinococcosis control programme in Hobukesar County, Xinjiang, China. *Acta Tropica* **145**, 1–7.

Wang, Q., Yong-fu, X., Vuitton, D.A., Schantz, P.M., Raoul, F., Budke, C.M., Campos-Ponce, M., Craig, P.S. & Giraudoux, P. (2007) Impact of overgrazing on the transmission of *Echinococcus multilocularis* in Tibetan pastoral communities of Sichuan Province, China. *Chinese Medical Journal* **120**, 237–242.

Wang, Q., Raoul, F., Budke, C., Craig, P.S., Xiao, Y.F., Vuitton, D.A., Campos-Ponce, M., Qiu, D.C., Pleydell, D.R.J. & Giraudoux, P. (2010) Grass height and transmission ecology of *Echinococcus multilocularis*. *Chinese Medical Journal* **123**, 61–67.

Wang, Y.H., Rogan, M.T., Vuitton, D.A., Wen, H., Bartholomot, B., Macpherson, C.N.L., Zou, P.F., Ding, Z.X., Zhou, H.X., Zhang, X.F., Luo, J., Xiong, H.B., Fu, Y., McVie, A., Giraudoux, P., Yang, W.G. & Craig, P.S. (2001) Cystic echinococcosis in semi-nomadic pastoral communities in north-west China. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **95**, 153–158.

WHO/OIE (World Health Organization/World Organization for Animal Health) (2001) Echinococcosis in animals: clinical aspects, diagnosis and treatment. *WHO/OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern*. Chapter 3. Paris, France, WHO/OIE.

Youden, W.J. (1950) Index for rating diagnostic tests. *Cancer* **3**, 32–35.

Ziadinov, I., Mathis, A., Trachsel, D., Rysmukhambetova, A., Abdyjaparov, T.A., Kuttubaev, O.T., Deplazes, P. & Torgerson, P.R. (2008) Canine echinococcosis in Kyrgyzstan: using prevalence data adjusted for measurement error to develop transmission dynamics models. *International Journal for Parasitology* **38**, 1179–1190.

Zweig, M.H. & Campbell, G. (1993) Receiver-operating clinical medicine (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* **39**, 561–577.

References

"One doesn't discover new lands without consenting to lose sight of the shore for a very long time"

Andre Gide (1869 – 1951)

Abbasi, I., Branzburg, A., Campos-Ponce, M., Abdel-Hafez, S. K., Raoul, F., Craig, P. S. and Hamburger, J. (2003). Copro-diagnosis of *Echinococcus granulosus* infection in dogs by amplification of a newly identified repeated DNA sequence. *American Journal of Tropical Medicine and Hygiene* **69**, 324–30.

Abdi, H. and Valentin, D. (2007). Multiple Correspondence Analysis. In *Encyclopedia of Measurement and Statistics* (ed. Salkind, N.), Sage, Thousand Oaks CA.

Abdyjaparov, T. A. and Kuttubaev, O. T. (2004). Alveolar Echinococcosis in Rodents of Mountainous Pastures of Kyrgyzstan. In *Echinococcosis in Central Asia: Problems and Solutions*, pp. 253–262.

Abel-Smith, B. and Falkingham, J. (1996). *Financing Health Services in Krygyzstan: The Extent of Private Payments.*

Acosta-Jamett, G., Cleaveland, S., Bronsvoort, B. M. D., Cunningham, A. A., Bradshaw, H. and Craig, P. S. (2010). *Echinococcus granulosus* infection in domestic dogs in urban and rural areas of the Coquimbo region, north-central Chile. *Veterinary Parasitology* **169**, 117–22.

Adler, F. R. and Kretzschmar, M. (1992). Aggregation and stability in parasite-host models. *Parasitology* **104**, 199–205.

Afonso, E., Knapp, J., Tête, N., Umhang, G., Rieffel, D., van Kesteren, F., Ziadinov, I., Craig, P. S., Torgerson, P. R. and Giraudoux, P. (2015). *Echinococcus multilocularis* in Kyrgyzstan: Similar Asian EmsB genotypic profiles found in village populations of Zaisan mole voles (*Ellobius tancrei*) and dogs in the Alay valley. *Journal of Helminthology (in press)*.

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. 2nd ed. Wiley-Interscience.

Ahmad, G. and Nizami, W. A. (1998). Coproantigens: early detection and suitability of an immunodiagnostic method for echinococcosis in dogs. *Veterinary Parasitology* **77**, 237–244.

Aitkin, M. A. and Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society B* **47**, 67–75.

Aitkin, M. A., Anderson, D. and Hinde, J. (1981). Statistical modelling of data on teaching styles (with discussion). *Journal of the Royal Statistical Society A* **144**, 419–461.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory* (ed. Petrov, B. N. and Caski, F.), pp. 267–281.

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics* **16**, 3–14.

Allan, J. C. and Craig, P. S. (2006). Coproantigens in taeniasis and echinococcosis. *Parasitology International* **55**, S75–S80.

Allan, J. C., Avila, G., Garcia Noval, J., Flisser, A. and Craig, P. S. (1990). Immunodiagnosis of taeniasis by coproantigen detection. *Parasitology* **101**, 473–7.

Allan, J. C., Craig, P. S., Garcia Noval, J., Mencos, F., Liu, D., Wang, Y., Wen, H., Zhou, P., Stringer, R., Rogan, M. T. and Zeyhle, E. (1992). Coproantigen detection for immunodiagnosis of echinococcosis and taeniasis in dogs and humans. *Parasitology* **104**, 347–355.

Al-Sabi', M. N. S., Kapel, C. M. O., Deplazes, P. and Mathis, A. (2007). Comparative copro-diagnosis of *Echinococcus multilocularis* in experimentally infected foxes. *Parasitology Research* **101**, 731–736.

Altman, D. G. and Bland, J. M. (1994a). Diagnostic tests 1: sensitivity and specificity. *BMJ* **308**, 1552.

Altman, D. G. and Bland, J. M. (1994b). Diagnostic tests 2: predictive values. *BMJ* **309**, 102.

Altman, D. G. and Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ* **332**, 1080.

Alvarez Rojas, C. A., Romig, T. and Lightowlers, M. W. (2014). *Echinococcus granulosus* sensu lato genotypes infecting humans--review of current knowledge. *International Journal for Parasitology* **44**, 9–18.

Anderson, R. M. (1994). The Croonian Lecture, 1994. Populations, infectious disease and immunity: a very nonlinear world. *Philosophical Transactions of the Royal Society B* **346**, 457–505.

Anderson, R. M. and Gordon, D. M. (1982). Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. *Parasitology* **85**, 373–398.

Anderson, R. M. and May, R. M. (1978). Regulation and stability of host-parasite population interactions: I. Regulatory processes. *The Journal of Animal Ecology* **47**, 219–247.

Anderson, R. M. and May, R. M. (1979). Population biology of infectious diseases: part I. *Nature* **280**, 455–461.

Anderson, R. M. and May, R. M. (1982). *The Population Biology of Infectious Diseases*. Springer-Verlag, Berlin.

Anderson, R. M. and May, R. M. (1985). Helminth infections of humans: mathematical models, population dynamics, and control. *Advances in Parasitology* **24**, 1–101.

Anderson, R. M. and May, R. M. (1991a). Indirectly transmitted helminths. In *Infectious Diseases of Humans*, pp. 550–589. Oxford University Press.

Anderson, R. M. and May, R. M. (1991b). Biology of host-macroparasite interactions. In *Infectious Diseases of Humans*, pp. 433–466. Oxford University Press.

Anderson, R. M. and May, R. M. (1991c). A framework for discussing the population biology of infectious diseases. In *Infectious Diseases of Humans*, pp. 13–23. Oxford University Press.

Anderson, R. M. and May, R. M. (1991d). The basic model: statics. In *Infectious Diseases of Humans*, pp. 467–506. Oxford University Press.

Anderson, R. M., Whitfield, P. J. and Dobson, A. P. (1978). Experimental studies of infection dynamics: infection of the definitive host by the cercariae of *Transversotrema patialense*. *Parasitology* **77**, 189–200.

Anderson, D. R., Burnham, K. P. and Thompson, W. L. (2000). Null hypothesis testing: problems, prevalence and an alternative. *Journal of Wildlife Management* **64**, 912–923.

Antolová, D., Reiterová, K., Miterpáková, M., Dinkel, A. and Dubinsky, P. (2009). The first finding of *Echinococcus multilocularis* in dogs in Slovakia: an emerging risk for spreading of infection. *Zoonoses and Public Health* **56**, 53–58.

Atema, K. and Hiby, E. (2015). Is culling dogs really necessary for echinococcosis control? *Acta Tropica* **143**, 77–78.

Atkinson, J. M., Williams, G. M., Yakob, L., Clements, A. C., Barnes, T. S., McManus, D. P., Yang, Y. R. and Gray, D. J. (2013). Synthesising 30 years of mathematical modelling of *Echinococcus* transmission. *PLoS Neglected Tropical Diseases* **7**, e2386.

Azlaf, R., Dakkak, A., Chentoufi, A. and El-Berrahmani, M. (2007). Modelling the transmission of *Echinococcus granulosus* in dogs in the northwest and in the southwest of Morocco. *Veterinary Parasitology* **145**, 297–303.

Azzalini, A. (1985). A class of distributions which includes the Normal ones. *Scandinavian Journal of Statistics* **12**, 171–178.

Bak, P. and Chen, K. (1991). Self-organized criticality. *Scientific American*.

Bak, P., Tang, C. and Wiesenfeld, K. (1987). Self-organized criticality: an explanation of the 1/f noise. *Physical Review Letters* **59**, 381–384.

Bak, P., Christensen, K. and Olami, Z. (1994). Self-organized criticality: consequences for statistics and predictability of earthquakes. In *Nonlinear Dynamics and Predictability of Geophysical Phenomena*, American Geophysical Union.

Barbour, A. D. and Kafetzaki, M. (1993). A host-parasite model yielding heterogeneous parasite loads. *Journal of Mathematical Biology* **31**, 157–176.

Barnes, T. S., Deplazes, P., Gottstein, B., Jenkins, D. J., Mathis, A., Siles-Lucas, M., Torgerson, P. R., Ziadinov, I. and Heath, D. D. (2012). Challenges for diagnosis and control of cystic hydatid disease. *Acta Tropica* **123**, 1–7.

Barr, D. J., Levy, R., Scheepers, C. and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* **68**, 255–278.

Barton, K. (2014). MuMIn: Multi-Model Inference. R package version 1.12.1.

Basáñez, M.-G., McCarthy, J. S., French, M. D., Yang, G.-J., Walker, M., Gambhir, M., Prichard, R. K. and Churcher, T. S. (2012a). A research agenda for helminth diseases of humans: modelling for control and elimination. *PLoS Neglected Tropical Diseases* **6**, e1548.

Basáñez, M.-G., French, M. D., Walker, M. and Churcher, T. S. (2012b). Paradigm lost: how parasite control may alter pattern and process in human helminthiases. *Trends in Parasitology* **28**, 161–71.

Baughman, A. L., Bisgard, K. M., Lynn, F. and Meade, B. D. (2006). Mixture model analysis for establishing a diagnostic cut-off point for pertussis antibody levels. *Statistics in Medicine* **25**, 2994–3010.

Bchir, A., Jaiem, A., Jemmali, M., Rousset, J. J., Gaudebout, C. and Larouze, B. (1987). Possible existence of an urban cycle of *Echinococcus granulosus* in central Tunisia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **81**, 650.

Belli, P. (2001). *Ten years of health reforms in former socialist economies: lessons learned and options for the future*. Cambridge, MA.

Benaglia, T., Chauveau, D., Hunter, D. R. and Young, D. S. (2009). mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software* **32**, 1–29.

Berlin, I. (1953). *The Hedgehog and the Fox*. Simon & Schuster, New York.

bioetymology.blogspot.co.uk (2012). http://bioetymology.blogspot.co.uk/2012/02/*Echinococcus*.html.

Birkes, D. and Dodge, Y. (1993). Constructing and checking the model. In *Alternative Methods of Regression*, pp. 13–28. John Wiley & Sons.

Black, M. A. and Craig, B. A. (2002). Estimating disease prevalence in the absence of a gold standard. *Statistics in Medicine* **21**, 2653–2669.

Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. Canadian Journal of Statistics **28**, 783–798.

Bliss, C. I. and Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics* **9**, 176–200.

Boatin, B. A., Basáñez, M.-G., Prichard, R. K., Awadzi, K., Barakat, R. M., García, H. H., Gazzinelli, A., Grant, W. N., McCarthy, J. S., N'Goran, E. K., Osei-Atweneboana, M. Y., Sripa, B., Yang, G.-J. and Lustigman, S. (2012). A research agenda for helminth diseases of humans: towards control and elimination. *PLoS Neglected Tropical Diseases* **6**, e1547.

Bolker, B. M. and Grenfell, B. T. (1993). Chaos and biological complexity in measles dynamics. *Proceedings of the Royal Society B* **251**, 75–81.

Bondareva, V. I., Boev, S. N., Sokolova, I. B. and Tazieva, E. H. (1975). Alveococcus in wild and domestic animals in the Balhash region. In *Questions of the Natural Foci of Disease.*, pp. 97–106. Alma-Ata.

Bonsall, M. B., French, D. R. and Hassell, M. P. (2002). Metapopulation structures affect persistence of predator – prey interactions. *Journal of Animal Ecology* **71**, 1075–1084.

Borrie, J., Gemmell, M. A. and Manktelow, B. W. (1965). An experimental approach to evaluate the potential risk of hydatid disease from inhalation of *Echinococcus* ova. *British Journal of Surgery* **52**, 876–8.

Boswell, M. T. and Patil, G. P. (1970). Chance mechanisms generating the negative binomial distribution. In *Random Counts in Scientific Work* (ed. Patil, G.), pp. 3–22. Pennsylvania State University Press, University Park, PA.

Boubaker, G., Macchiaroli, N., Prada, L., Cucher, M. A., Rosenzvit, M. C., Ziadinov, I., Deplazes, P., Saarma, U., Babba, H., Gottstein, B. and Spiliotis, M. (2013). A multiplex PCR for the simultaneous detection and genotyping of the *Echinococcus granulosus* complex. *PLoS Neglected Tropical Diseases* **7**, e2017.

Boufana, B., Campos-Ponce, M., Naidich, A., Buishi, I. E., Lahmar, S., Zeyhle, E., Jenkins, D. J., Combes, B., Wen, H., Xiao, N., Nakao, M., Ito, A., Qiu, J. and Craig, P. S. (2008). Evaluation of three PCR assays for the identification of the sheep strain (genotype 1) of *Echinococcus granulosus* in canid feces and parasite tissues. *American Journal of Tropical Medicine and Hygiene* **78**, 777–783.

Boufana, B., Umhang, G., Qiu, J., Chen, X., Lahmar, S., Boué, F., Jenkins, D. J. and Craig, P. S. (2013). Development of three PCR assays for the differentiation between *Echinococcus shiquicus, E. granulosus* (G1 genotype), and *E. multilocularis* DNA in the co-endemic region of Qinghai-Tibet plateau, China. *American Journal of Tropical Medicine and Hygiene* **88**, 795–802.

Bowles, J., Blair, D. and McManus, D. P. (1992). Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Molecular and Biochemical Parasitology* **54**, 165–173.

Bowles, J., Blair, D. and McManus, D. P. (1995). A molecular phylogeny of the genus *Echinococcus*. *Parasitology* **110 (Pt 3**, 317–328.

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model Building and Response Surfaces*. John Wiley & Sons, New York, NY.

Branscum, A. J., Johnson, W. O., Hanson, T. E. and Gardner, I. A. (2008). Bayesian semiparametric ROC curve estimation and disease diagnosis. *Statistics in Medicine* **27**, 2474–2496.

Bretagne, S., Guillou, J. P., Morand, M. and Houin, R. (1993). Detection of *Echinococcus multilocularis* DNA in fox faeces using DNA amplification. *Parasitology* **106**, 193–199.

Brisson, D., Brinkley, C., Humphrey, P. T., Kemps, B. D. and Ostfeld, R. S. (2011). It takes a community to raise the prevalence of a zoonotic pathogen. *Interdisciplinary Perspectives on Infectious Diseases* 741406.

Brockhurst, M. A., Buckling, A. and Rainey, P. B. (2006). Spatial heterogeneity and the stability of host-parasite coexistence. *Journal of Evolutionary Biology* **19**, 374–9.

Brooker, S. J., Nikolay, B., Balabanova, D. and Pullan, R. L. (2015). Global feasibility assessment of interrupting the transmission of soil-transmitted helminths: a statistical modelling study. *The Lancet Infectious Diseases* **3099**, 1–10.

Brossard, M., Andreutti, C. and Siegenthaler, M. (2007). Infection of red foxes with *Echinococcus multilocularis* in western Switzerland. *Journal of Helminthology* **81**, 369–376.

Browne, W. J., Subramanian, S. V., Jones, K. and Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society A* **168**, 599–613.

Brunetti, E., Kern, P. and Vuitton, D. A. (2010). Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. *Acta Tropica* **114**, 1–16.

Buckland, S. T., Burnham, K. P. and Augustin, N. H. (1997). Model Selection: An Integral Part of Inference. *Biometrics* **53**, 603–618.

Budke, C. M., Jiamin, Q., Zinsstag, J., Qian, W. and Torgerson, P. R. (2004). Use of disability adjusted life years in the estimation of the disease burden of Echinococcosis for a high endemic region of the Tibetan Plateau. *American Journal of Tropical Medicine and Hygiene* **71**, 56–64.

Budke, C. M., Campos-Ponce, M., Qian, W. and Torgerson, P. R. (2005a). A canine purgation study and risk factor analysis for echinococcosis in a high endemic region of the Tibetan plateau. *Veterinary Parasitology* **127**, 43–49.

Budke, C. M., Jiamin, Q., Craig, P. S. and Torgerson, P. R. (2005b). Modeling the transmission of *Echinococcus granulosus* and *Echinococcus multilocularis* in dogs for a high endemic region of the Tibetan plateau. *International Journal for Parasitology* **35**, 163–70.

Budke, C. M., Jiamin, Q., Qian, W. and Torgerson, P. R. (2005c). Economic effects of echinococcosis in a disease-endemic region of the Tibetan Plateau. *American Journal of Tropical Medicine and Hygiene* **73**, 2–10.

Budke, C. M., Deplazes, P. and Torgerson, P. R. (2006). Global socioeconomic impact of cystic echinococcosis. *Emerging Infectious Diseases* **12**, 296–303.

Buishi, I. E., Walters, T., Guildea, Z., Craig, P. S. and Palmer, S. (2005a). Reemergence of canine *Echinococcus granulosus* infection, Wales. *Emerging Infectious Diseases* **11**, 568–571.

Buishi, I. E., Njoroge, E. M., Bouamra, O. and Craig, P. S. (2005b). Canine echinococcosis in northwest Libya: assessment of coproantigen ELISA, and a survey of infection with analysis of risk-factors. *Veterinary Parasitology* **130**, 223–32.

Buishi, I. E., Njoroge, E., Zeyhle, E., Rogan, M. T. and Craig, P. S. (2006). Canine echinococcosis in Turkana (north-western Kenya): a coproantigen survey in the previous hydatid-control area and an analysis of risk factors. *Annals of Tropical Medicine and Parasitology* **100**, 601–610.

Bundy, D. A. and Medley, G. F. (1992). Immuno-epidemiology of human geohelminthiasis: ecological and immunological determinants of worm burden. *Parasitology* **104 Suppl**, S105–S119.

Burlet, P., Deplazes, P. and Hegglin, D. (2011). Age, season and spatio-temporal factors affecting the prevalence of *Echinococcus multilocularis* and *Taenia taeniaeformis* in *Arvicola terrestris*. *Parasites & Vectors* **4**, 6.

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

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research* **33**, 261–304.

Calabrese, J. M., Brunner, J. L. and Ostfeld, R. S. (2011). Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-gamma mixture model. *PLoS One* **6**, e29215.

Calisti, M., Ciampalini, B., Lovari, S. and Lucherini, M. (1990). Food habits and trophic niche variation of the red fox Vulpes vulpes (L., 1758) in a Mediterranean coastal area. *Revue D'Ecologie - La Terre Et La Vie* **45**, 309–320.

Cameron, A. and Baldock, F. C. (1998). A new probability formula for surveys to substantiate freedom from disease. *Preventive Veterinary Medicine* **34**, 1–17.

Cameron, T. W. and Webster, G. A. (1969). The histogenesis of the hydatid cyst (*Echinococcus* spp.) Part 1. Liver cysts in large mammals. *Canadian Journal of Veterinary Research* **47**, 1405–1410.

Campos-Bueno, A., Lopez-Abente, G. and Andres-Cercadillo, A. M. (2000). Risk factors for *Echinococcus granulosus* infection: a case-control study. *American Journal of Tropical Medicine and Hygiene* **62**, 329–34.

Canning, D. (2006). Priority setting and the "neglected" tropical diseases. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **100**, 499–504.

Cardona, G. A. and Carmena, D. (2013). A review of the global prevalence, molecular epidemiology and economics of cystic echinococcosis in production animals. *Veterinary Parasitology* **192**, 10–32.

Carmena, D., Benito, A. and Eraso, E. (2006). Antigens for the immunodiagnosis of *Echinococcus granulosus* infection: An update. *Acta Tropica* **98**, 74–86.

Carmona, C., Perdomo, R., Carbo, A., Alvarez, C., Monti, J., Grauert, R., Stern, D., Perera, G., Lloyd, S., Bazini, R., Gemmell, M. A. and Yarzabal, L. (1998). Risk factors associated with human cystic echinococcosis in Florida, Uruguay:

results of a mass screening study using ultrasound and serology. *American Journal of Tropical Medicine and Hygiene* **58**, 599–605.

Casaravilla, C., Malgor, R., Rossi, A., Sakai, H., Nonaka, N., Kamiya, M. and Carmona, C. (2005). Production and characterization of monoclonal antibodies against excretory/secretory products of adult *Echinococcus granulosus*, and their application to coproantigen detection. *Parasitology International* **54**, 43–49.

Cassini, R., Mulatti, P., Zanardello, C., Simonato, G., Signorini, M., Cazzin, S., Tambalo, P., Cobianchi, M., Pietrobelli, M. and Capelli, G. (2014). Retrospective and spatial analysis tools for integrated surveillance of cystic echinococcosis and bovine cysticercosis in hypo-endemic areas. *Geospatial Health* **8**, 509–515.

Chaâbane-Banaoues, R., Oudni-M'rad, M., Cabaret, J., M'rad, S., Mezhoud, H. and Babba, H. (2015). Infection of dogs with *Echinococcus granulosus*: causes and consequences in an hyperendemic area. *Parasites & Vectors* **8**, 231.

Chamberlin, T. C. (1965). The Method of Multiple Working Hypotheses [1890]. *Science* **148**, 754–759.

Cheng, L., Lei, L. and Guo, S. (2010). In vitro and in vivo evaluation of praziquantel loaded implants based on PEG/PCL blends. *International Journal of Pharmaceutics* **387**, 129–138.

Choi, Y.-K., Johnson, W. O., Collins, M. T. and Gardner, I. A. (2006a). Bayesian inferences for receiver operating characteristic curves in the absence of a gold standard. *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 210–229.

Choi, Y.-K., Johnson, W. O. and Thurmond, M. C. (2006b). Diagnosis using predictive probabilities without cut-offs. *Statistics in Medicine* **25**, 699–717.

Christensen, R., Hanson, T. E. and Jara, A. (2008). Parametric nonparametric statistics. An introduction to mixtures of finite Polya trees. *The American Statistician* **62**, 296–306.

Christensen, R., Johnson, W. O., Branscum, A. J. and Hanson, T. E. (2011). Nonparametric models. In *Bayesian ideas and data analysis: an introduction for scientists and statisticians*, pp. 385–418. CRC press, Boca Raton, FL.

Christofi, G., Deplazes, P., Christofi, N., Tanner, I., Economides, P. and Eckert, J. (2002). Screening of dogs for *Echinococcus granulosus* coproantigen in a low endemic situation in Cyprus. *Veterinary Parasitology* **104**, 299–306.

Churcher, T. S., Ferguson, N. M. and Basáñez, M.-G. (2005). Density dependence and overdispersion in the transmission of helminth parasites. *Parasitology* **131**, 121–132.

Churcher, T. S., Filipe, J. A. N. and Basáñez, M.-G. (2006). Density dependence and the control of helminth parasites. *Journal of Animal Ecology* **75**, 1313–1320.

Clopper, C. J. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404–413.

Cohen, H., Paolillo, E., Bonifacino, R., Botta, B., Parada, L., Cabrera, P., Snowden, K., Gasser, R., Tessier, R., Dibarboure, L., Wen, H., Allan, J. C., de Alfaro, H. S., Rogan, M. T. and Craig, P. S. (1998). Human cystic echinococcosis in a Uruguayan community: a sonographic, serologic, and epidemiologic study. *American Journal of Tropical Medicine and Hygiene* **59**, 620–627.

Coleman, P. G., Perry, B. D. and Woolhouse, M. E. J. (2001). Endemic stability--a veterinary idea applied to human public health. *The Lancet* **357**, 1284–1286.

Conder, G. A., Marchiondo, A. A. and Andersen, F. L. (1981). Effect of praziquantel on adult *Echinococcus granulosus* in vitro: Scanning electron microscopy. *Zeitschrift fur Parasitenkunde* **66**, 191–199.

Craig, P. S. (1986). Detection of specific circulating antigen, immune complexes and antibodies in human hydatidosis from Turkana (Kenya) and Great Britain, by enzyme-immunoassay. *Parasite Immunology* **8**, 171–188.

Craig, P. S. (1993). Immunodiagnosis of *Echinococcus granulosus*. In *Compendium on cystic echinococcosis with special reference to the Xinjiang Uygur Autonomous Region, The People`s Republic of China.* (ed. Andersen, F. L., Chai, J., and Liu, C. H.), Brigham Young University Print Services, Provo, Utah, USA.

Craig, P. S. (1997). Immunodiagnosis of *Echinococcus granulosus* and a comparison of techniques for diagnosis of canine echinococcosis. In *Compendium on Cystic Echinococcosis in Africa and in Middle Eastern Countries with special reference to Morocco* (ed. Andersen, F., Ouhelli, H., and Kachani, M.), pp. 85–118. Brigham Young University Print Services, Provo, Utah.

Craig, P. S. and Larrieu, E. J. (2006). Control of cystic echinococcosis/hydatidosis: 1863-2002. *Advances in Parasitology* **61**, 443–508.

Craig, P. S. and The Echinococcosis Working Group in China (2006). Epidemiology of human alveolar echinococcosis in China. *Parasitology International* **55**, S221–S225.

Craig, P. S., Macpherson, C. N. L., Watson-Jones, D. L. and Nelson, G. S. (1988). Immunodetection of *Echinococcus* eggs from naturally infected dogs and from

environmental contamination sites in settlements in Turkana, Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **82**, 268–274.

Craig, P. S., Liu, D., Shi, D., Macpherson, C. N. L., Barnish, G., Reynolds, D., Gottstein, B. and Wang, Z. (1992). A large focus of alveolar echinococcosis in central China. *The Lancet* **340**, 826–831.

Craig, P. S., Gasser, R. B., Parada, L., Cabrera, P., Parietti, S., Borgues, C., Acuttis, A., Agulla, J., Snowden, K. and Paolillo, E. (1995). Diagnosis of canine echinococcosis: comparison of coproantigen and serum antibody tests with arecoline purgation in Uruguay. *Veterinary Parasitology* **56**, 293–301.

Craig, P. S., Giraudoux, P., Shi, D., Bartholomot, B., Barnish, G., Delattre, P., Quere, J. P., Harraga, S., Bao, G., Wang, Y., Lu, F., Ito, A. and Vuitton, D. A. (2000). An epidemiological and ecological study of human alveolar echinococcosis transmission in south Gansu, China. *Acta Tropica* **77**, 167–177.

Craig, P. S., Rogan, M. T. and Campos-Ponce, M. (2003). Echinococcosis: disease, detection and transmission. *Parasitology* **127**, S1–S16.

Craig, P. S., Budke, C. M., Schantz, P. M., Li, T., Qiu, J. M., Yang, Y., Zeyhle, E., Rogan, M. T. and Ito, A. (2007a). Human echinococcosis: a neglected disease? *Tropical Medicine and Health* **35**, 283–292.

Craig, P. S., McManus, D. P., Lightowlers, M. W., Chabalgoity, J. A., García, H. H., Gavidia, C. M., Gilman, R. H., Gonzalez, A. E., Lorca, M., Naquira, C., Nieto, A. and Schantz, P. M. (2007b). Prevention and control of cystic echinococcosis. *The Lancet Infectious Diseases* **7**, 385–394.

Craig, P. S., Li, T. Y., Qiu, J. M., Zhen, R., Wang, Q., Giraudoux, P., Ito, A., Heath, D. D., Warnock, B., Schantz, P. and Yang, W. (2008). Echinococcoses and Tibetan Communities. *Emerging Infectious Diseases* **14**, 1674–1675.

Craig, P. S., Mastin, A., van Kesteren, F. and Boufana, B. (2015). *Echinococcus granulosus*: epidemiology and state-of-the-art of diagnostics in animals. *Veterinary Parasitology (in press)*.

Cringoli, G., Rinaldi, L., Musella, V., Veneziano, V., Maurelli, M. P., Di Pietro, F., Frisiello, M. and Di Pietro, S. (2007). Geo-referencing livestock farms as tool for studying cystic echinococcosis epidemiology in cattle and water buffaloes from southern Italy. *Geospatial Health* **2**, 105–111.

Crofton, H. D. (1971a). A quantitative approach to parasitism. *Parasitology* **62**, 179–193.

Crofton, H. D. (1971b). A model of host-parasite relationships. *Parasitology* **63**, 343–364.

Danson, F. M., Graham, A. J., Pleydell, D. R. J., Campos-Ponce, M., Giraudoux, P. and Craig, P. S. (2003). Multi-scale spatial analysis of human alveolar echinococcosis risk in China. *Parasitology* **127**, S131–S139.

Danson, F. M., Giraudoux, P. and Craig, P. S. (2006). Spatial modelling and ecology of *Echinococcus multilocularis* transmission in China. *Parasitology International* **55**, S227–S231.

Deeks, J. J. and Altman, D. G. (2004). Diagnostic tests 4: likelihood ratios. *BMJ* **329**, 168–169.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* **39**, 1–38.

Denwood, M. J., Stear, M. J., Matthews, L., Reid, S. W. J., Toft, N. and Innocent, G. T. (2008). The distribution of the pathogenic nematode *Nematodirus battus* in lambs is zero-inflated. *Parasitology* **135**, 1225–1235.

Deplazes, P. and Hegglin, D. (2004). Control of *Echinococcus multilocularis* in definitive host populations. In *Echinococcosis in Central Asia: Problems and Solutions*, pp. 263–272.

Deplazes, P., Gottstein, B., Stingelin, Y. and Eckert, J. (1990). Detection of *Taenia hydatigena* coproantigens by ELISA in dogs. *Veterinary Parasitology* **36**, 91–103.

Deplazes, P., Gottstein, B., Eckert, J., Jenkins, D. J., Ewald, D. and Jimenez-Palacios, S. (1992). Detection of *Echinococcus* coproantigens by enzyme-linked immunosorbent assay in dogs, dingoes and foxes. *Parasitology Research* **78**, 303–308.

Deplazes, P., Jimenez-Palacios, S., Gottstein, B., Skaggs, J. and Eckert, J. (1994). Detection of *Echinococcus* coproantigens in stray dogs of northern Spain. *Applied Parasitology* **35**, 297–301.

Deplazes, P., Alther, P., Tanner, I., Thompson, R. C. and Eckert, J. (1999). *Echinococcus multilocularis* coproantigen detection by enzyme-linked immunosorbent assay in fox, dog, and cat populations. *Journal of Parasitology* **85**, 115–121.

Deplazes, P., Dinkel, A. and Mathis, A. (2003). Molecular tools for studies on the transmission biology of *Echinococcus multilocularis*. *Parasitology* **127**, S53–S61.

Dévé, F. (1949). Léchinococcose primitive (maladie hydatique). *Masson & Cie, Paris* 362.

Diekmann, O., Heesterbeek, J. A. P. and Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio Ro in models for infectious diseases in heterogenous populations. *Journal of Mathematical Biology* **28**, 365–382.

Dietz, K. (1974). Transmission and control of arbovirus diseases. In *Epidemiology. Proceedings of a SIMS Conference on Epidemiology, Alta, Utah* (ed. Ludwig, D. and Cooke, K. L.), pp. 104–121. Society for Industrial and Applied Mathematics, Philadelphia.

Diez Roux, A. V. (2011). Complex systems thinking and current impasses in health disparities research. *American Journal of Public Health* **101**, 1627–1634.

Dinkel, A., von Nickisch-Rosenegk, M., Bilger, B., Merli, M., Lucius, R. and Romig, T. (1998). Detection of *Echinococcus multilocularis* in the definitive host: coprodiagnosis by PCR as an alternative to necropsy. *Journal of Clinical Microbiology* **36**, 1871–1876.

Dinkel, A., Njoroge, E. M., Zimmermann, A., Wälz, M., Zeyhle, E., Elmahdi, I. E., Mackenstedt, U. and Romig, T. (2004). A PCR system for detection of species and genotypes of the *Echinococcus granulosus*-complex, with reference to the epidemiological situation in eastern Africa. *International Journal for Parasitology* **34**, 645–653.

Dohoo, I. R. and Medley, G. F. (2009). Concepts of infectious disease epidemiology. In *Veterinary Epidemiologic Research* (ed. Dohoo, I. R., Martin, W., and Stryhn, H.), pp. 715–737. VER Inc.

Dore, F., Varcasia, A., Pipia, A. P., Sanna, G., Pinna Parpaglia, M. L., Corda, A., Romig, T. and Scala, A. (2014). Ultrasound as a monitoring tool for cystic echinococcosis in sheep. *Veterinary Parasitology* **203**, 59–64.

Dörre, A. and Borchardt, P. (2012). Changing systems, changing effects—pasture utilization in the post-soviet transition. *Mountain Research and Development* **32**, 313–323.

Dowling, P. M. and Torgerson, P. R. (2000). A cross-sectional survey to analyse the risk factors associated with human cystic echinococcosis in an endemic area of mid-Wales. *Annals of Tropical Medicine and Parasitology* **94**, 241–245.

Dowling, P. M., Abo-Shehada, M. N. and Torgerson, P. R. (2000). Risk factors associated with human cystic echinococcosis in Jordan: results of a case–control study. *Annals of Tropical Medicine and Parasitology* **94**, 69–75.

Drolette, M. E. (1974). The vanishing 2x2 table: linking the hypergeometric, binomial and Poisson. *The American Statistician* **28**, 102–103.

Dubinina, T. (2005). Living legend of Tyan-Shan mountains. *R-PADS Newsletter, No. 7* 2–5.

Dyachenko, V., Pantchev, N., Gawlowska, S., Vrhovec, M. G. and Bauer, C. (2008). *Echinococcus multilocularis* infections in domestic dogs and cats from Germany and other European countries. *Veterinary Parasitology* **157**, 244–253.

Eckert, J. (2003). Predictive values and quality control of techniques for the diagnosis of *Echinococcus multilocularis* in definitive hosts. *Acta Tropica* **85**, 157–163.

Eckert, J. and Deplazes, P. (2001). Immunological and molecular techniques for diagnosing the *Echinococcus multilocularis* infection in definitive and intermediate hosts. *Acta Parasitologica* **46**, 1–7.

Eckert, J. and Deplazes, P. (2004). Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. *Clinical Microbiology Reviews* **17**, 107–133.

Eckert, J., Gemmell, M. A. and Soulsby, E. J. (1982). Surveys on prevalence and geographic distribution of echinococcosis. In *Echinococcosis/Hydatidosis Surveillance, Prevention and Control: FAO/UNEP/WHO Guidelines*, pp. 36–45.

Economides, P. and Christofi, G. (2000). Evaluation of control programmes for echinococcosis/hydatidosis in Cyprus. *Revue Scientifique et Technique* **19**, 784–792.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability)*. Chapman and Hall / CRC.

Elayoubi, F. A. and Craig, P. S. (2004). *Echinococcus granulosus* coproantigens: chromatographic fractionation and characterization. *Parasitology* **128**, 455–465.

Elayoubi, F. A., Fraser, A., Jenkins, D. J. and Craig, P. S. (2003). Partial characterisation of carbohydrate-rich *Echinococcus granulosus* coproantigens. *International Journal for Parasitology* **33**, 1553–1559.

Elshazly, A. M., Awad, S. E., Abdel Tawab, A. H., Haridy, F. M. and Morsy, T. A. (2007). Echinococcosis (zoonotic hydatidosis) in street dogs in urban and rural areas, Dakahlia Governorate, Egypt. *Journal of the Egyptian Society of Parasitology* **37**, 287–298.

El-Shehabi, F. S., Kamhawi, S. A., Schantz, P. M., Craig, P. S. and Abdel-Hafez, S. K. (2000). Diagnosis of canine echinococcosis: comparison of coproantigen detection with necropsy in stray dogs and red foxes from northern Jordan. *Parasite* **7**, 83–90.

Elsheikha, H. M., McOrist, S. and Geary, T. G. (2011). Antiparasitic drugs: mechanisms of action and resistance. In *Essentials of Veterinary Parasitology* (ed. Elsheikha, H. M. and Ahmed Khan, N.), pp. 187–200. Caister Academic Press.

Elston, D. A., Moss, R., Boulinier, T., Arrowsmith, C. and Lambin, X. (2001). Analysis of aggregation, a worked example: numbers of ticks on red grouse chicks. *Parasitology* **122**, 563–569.

Engen, S., Bakke, O. and Islam, A. (1998). Demographic and environmental stochasticity - concepts and definitions. *Biometrics* **54**, 840–846.

Enøe, C., Georgiadis, M. P. and Johnson, W. O. (2000). Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive Veterinary Medicine* **45**, 61–81.

Erkanli, A., Sung, M., Costello, E. J. and Angold, A. (2006). Bayesian semi-parametric ROC analysis. *Statistics in Medicine* **25**, 3905–3928.

Eslami, A. and Hosseini, S. H. (1998). *Echinococcus granulosus* infection of farm dogs of Iran. *Parasitology Research* **84**, 205–207.

Evdokimov, N. G. (2013). Structure of colonies in the northern mole vole, *Ellobius talpinus* (Rodentia, Cricetidae). *Biology Bulletin* **40**, 768–778.

Farrington, J. D. (2005). De-development in eastern Kyrgyzstan and persistence of semi- nomadic livestock herding. *Nomadic Peoples* **9**, 171–197.

Fay, F. H. and Rausch, R. L. (1964). The seasonal cycle of abundance of *Echinococcus multilocularis* in naturally infected arctic foxes. *Faculty Publications from the Harold W Manter Laboratory of Parasitology* **596**, 765–766.

Federer, K., Armua-Fernandez, M. T., Hoby, S., Wenker, C. and Deplazes, P. (2015). In vivo viability of *Echinococcus multilocularis* eggs in a rodent model after different thermo-treatments. *Experimental Parasitology* **41**, 1–6.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics* **2**, 615–629.

Ferrari, N. and Weber, J.-M. (1995). Influence of the abundance of food resources on the feeding habits of the red fox, *Vulpes vulpes*, in western Switzerland. *Journal of Zoology* **236**, 117–129.

Fisher, R. A. (1941). The negative binomial distribution. *Annals of Eugenics* **11**, 182–187.

Fitzherbert, A. (2000). *Country Pasture/Forage Resource Profiles: Kyrgyzstan*.

Flom, P. L. and Cassell, D. L. (2007). Stopping stepwise : Why stepwise and similar selection methods are bad , and what you should use. In *NESUG 2007 Proceedings*, pp. 1–7.

Forster, M. R. (2000). Key concepts in model selection: performance and generalizability. *Journal of Mathematical Psychology* **44**, 205–231.

Fox, J. and Weisberg, S. (2011). *An R companion to applied regression*. Second. Sage, Thousand Oaks CA.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**, 611–631.

Galvani, A. P. (2003). Immunity, antigenic heterogeneity, and aggregation of helminth parasites. *The Journal of Parasitology* **89**, 232–241.

Gasser, R. B., Lightowlers, M. W., Obendorf, D. L., Jenkins, D. J. and Rickard, M. D. (1988). Evaluation of a serological test system for the diagnosis of natural *Echinococcus granulosus* infection in dogs using *E. granulosus* protoscolex and oncosphere antigens. *Australian Veterinary Journal* **65**, 369–373.

Gasser, R. B., Jenkins, D. J., Paolillo, E., Parada, L., Cabrera, P. and Craig, P. S. (1993). Serum antibodies in canine echinococcosis. *International Journal for Parasitology* **23**, 579–586.

Gasser, R. B., Parada, L., Acuna, A., Burges, C., Laurenson, M. K., Gulland, F. M., Reichel, M. P. and Paolillo, E. (1994). Immunological assessment of exposure to *Echinococcus granulosus* in a rural dog population in Uruguay. *Acta Tropica* **58**, 179–185.

Gatti, A., Alvarez, A. R., Araya, D., Mancini, S., Herrero, E., Santillan, G. and Larrieu, E. J. (2007). Ovine echinococcosis I. Immunological diagnosis by enzyme immunoassay. *Veterinary Parasitology* **143**, 112–121.

Gay, N. J. (1996). Analysis of serological surveys using mixture models: application to a survey of parvovirus B19. *Statistics in Medicine* **15**, 1567–1573.

Gay, N. J. and Su, Y.-S. (2014). arm: data analysis using regression and multilevel/hierarchical models. R package version 1.7-07.

Gelman, A. G. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* **27**, 2865–2873.

Gelman, A. G. (2014). Why we hate stepwise regression. *http://andrewgelman.com/2014/06/02/hate-stepwise-regression/*.

Gelman, A. and Hill, J. (2006a). Generalized linear models. In *Data Analysis using Regression and Multilevel/Hierarchical Models*, pp. 109–134. Cambridge University Press.

Gelman, A. and Hill, J. (2006b). *Data Analysis using Regression and Multilevel/Hierarchical Models*. 1st ed. Cambridge University Press.

Gemmell, M. A. (1978). The Styx Field Trial. *Bulletin of the World Health Organization* **56**, 433–443.

Gemmell, M. A. and Johnstone, P. D. (1977). Experimental epidemiology of hydatidosis and cysticercosis. *Advances in Parasitology* **15**, 311–369.

Gemmell, M. A. and Lawson, J. R. (1986). Epidemiology and control of hydatid disease. In *The Biology of Echinococcosis and Hydatid Disease*, pp. 189–216.

Gemmell, M. A. and Roberts, M. G. (1995). Modelling *Echinococcus* life cycles. In *Echinococcus and Hydatid Disease*, pp. 333–354.

Gemmell, M. A. and Schantz, P. M. (1997). Formulating policies for control of *Echinococcus granulosus*. In *Compendium on Cystic Echinococcosis in Africa and in Middle Eastern Countries with special reference to Morocco* (ed. Andersen, F. L., Ouhelli, H., and Kachani, M.), pp. 329–345. Brigham Young University Press, Provo, Utah.

Gemmell, M. A., Lawson, J. R. and Roberts, M. G. (1986a). Control of echinococcosis/hydatidosis: present status of worldwide progress. *Bulletin of the World Health Organization* **64**, 333–339.

Gemmell, M. A., Lawson, J. R., Roberts, M. G., Kerin, B. R. and Mason, C. J. (1986b). Population dynamics in echinococcosis and cysticercosis : comparison of the response of *Echinococcus granulosus*, *Taenia hydatigena* and *T. ovis* to control. *Parasitology* **93**, 357–369.

Gemmell, M. A., Lawson, J. R. and Roberts, M. G. (1986c). Population dynamics in echinococcosis and cysticercosis: biological parameters of *Echinococcus granulosus* in dogs and sheep. *Parasitology* **92**, 599–620.

Gemmell, M. A., Lawson, J. R., Roberts, M. G. and Griffin, J. F. T. (1990). Population dynamics in echinococcosis and cysticercosis: regulation of *Taenia hydatigena* and *T. ovis* in lambs through passively transferred immunity. *Parasitology* **101**, 145–151.

Gemmell, M. A., Roberts, M. G., Beard, T. C., Campano Diaz, S., Lawson, J. R. and Nonnemaker, J. M. (2001). Formulating effective and cost-effective policies in the planning phase for permanent control of *Echinococcus granulosus*. In

WHO / OIE Manual on Echinococcosis in Humans and Animals : a Public Health Problem of Global Concern, pp. 209–218.

Giraudoux, P., Delattre, P., Takahashi, K., Raoul, F., Quere, J. P., Craig, P. S. and Vuitton, D. (2002). Transmission ecology of *Echinococcus multilocularis* in wildlife: what can be learned from comparative studies and multiscale approaches. In *Cestode Zoonoses: Echinococcosis and Cysticercosis - An Emergent and Global Problem* (ed. Craig, P. and Pawlowski, Z.), pp. 251–266. IOS Press, Amsterdam.

Giraudoux, P., Craig, P. S., Delattre, P., Bao, G., Bartholomot, B., Harraga, S., Qur, J.-P., Raoul, F., Wang, Y., Shi, D. and Vuitton, D. A. (2003). Interactions between landscape changes and host communities can regulate *Echinococcus multilocularis* transmission. *Parasitology* **127**, S119–S129.

Giraudoux, P., Pleydell, D. R. J., Raoul, F., Quéré, J.-P., Wang, Q., Yang, Y., Vuitton, D. A., Qiu, J., Yang, W. and Craig, P. S. (2006). Transmission ecology of *Echinococcus multilocularis*: what are the ranges of parasite stability among various host communities in China? *Parasitology International* **55**, S237–46.

Giraudoux, P., Pleydell, D., Raoul, F., Vaniscotte, A., Ito, A. and Craig, P. S. (2007). *Echinococcus multilocularis*: Why are multidisciplinary and multiscale approaches essential in infectious disease ecology? *Tropical Medicine and Health* **35**, 293–299.

Giraudoux, P., Raoul, F., Pleydell, D. R. J. and Craig, P. S. (2008). Multidisciplinary studies, systems approaches and parasite eco-epidemiology: Something old, something new. *Parasite* **15**, 469–476.

Giraudoux, P., Raoul, F., Pleydell, D. R. J., Li, T. Y., Han, X. M., Qiu, J., Xie, Y., Wang, H., Ito, A. and Craig, P. S. (2013a). Drivers of *Echinococcus multilocularis* transmission in China: small mammal diversity, landscape or climate? *PLoS Neglected Tropical Diseases* **7**, e2045.

Giraudoux, P., Raoul, F., Afonso, E., Ziadinov, I., Yang, Y., Li, L., Li, T. Y., Quéré, J.-P., Feng, X., Wang, Q., Wen, H., Ito, A. and Craig, P. S. (2013b). Transmission ecosystems of *Echinococcus multilocularis* in China and Central Asia. *Parasitology* **140**, 1655–1666.

Gisiger, T. (2001). Scale invariance in biology: coincidence or footprint of a universal mechanism? *Biological Reviews of the Cambridge Philosophical Society* **76**, 161–209.

Gladwell, M. (2006). Million Dollar Murray. *The New Yorker* **82**, 96–107.

Glass, R. I. (1976). The SANEPID Service in the U.S.S.R. *Public Health Reports* **91**, 154–158.

Glattre, E. and Nygård, J. F. (2004). Fractal meta-analysis and "causality" embedded in complexity: advanced understanding of disease etiology. *Nonlinear Dynamics, Psychology, and Life Sciences* **8**, 315–344.

Glattre, E., Nygård, J. F. and Skjerve, E. (2008). *Fundamental Aspects of Fractal Epidemiology*.

Glattre, E., Nygård, J. F. and Aaseth, J. (2012). Selenium and cancer prevention: observations and complexity. *Journal of Trace Elements in Medicine and Biology* **26**, 168–169.

Goldberger, A. L. (1996). Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. *The Lancet* **347**, 1312–1314.

Goldenfeld, N. and Kadanoff, L. P. (1999). Simple lessons from complexity. *Science* **284**, 87–89.

Gotsadze, G., Chikovani, I., Goguadze, K., Balabanova, D. and McKee, M. (2010). Reforming sanitary-epidemiological service in Central and Eastern Europe and the former Soviet Union: an exploratory study. *BMC Public Health* **10**, 440.

Gottstein, B. (1984). An immunoassay for the detection of circulating antigens in human echinococcosis. *American Journal of Tropical Medicine and Hygiene* **33**, 1185–1191.

Gottstein, B. (1985). Purification and characterization of a specific antigen from *Echinococcus multilocularis*. *Parasite Immunology* **7**, 201–212.

Gottstein, B. (1992). Molecular and immunological diagnosis of echinococcosis. *Clinical Microbiology Reviews* **5**, 248–261.

Gottstein, B., Deplazes, P., Eckert, J., Müller, B., Schott, E., Helle, O., Boujon, P., Wolff, K., Wandeler, A. and Schwiete, U. (1991). Serological (Em2-ELISA) and parasitological examinations of fox populations for *Echinococcus multilocularis* infections. *Zentralbl Veterinarmed B* **38**, 161–168.

Gottstein, B., Jacquier, P., Bresson-Hadni, S. and Eckert, J. (1993). Improved primary immunodiagnosis of alveolar echinococcosis in humans by an enzyme-linked immunosorbent assay using the improved primary immunodiagnosis of alveolar echinococcosis in humans by an enzyme-linked immunosorbent assay using the Em2Plus antigen. *Journal of Clinical Microbiology* **31**, 373–376.

Gourbière, S., Morand, S. and Waxman, D. (2015). Fundamental factors determining the nature of parasite aggregation in hosts. *PLoS One* **10**, e0116893.

Graham, A. J., Danson, F. M. and Craig, P. S. (2005). Ecological epidemiology: the role of landscape structure in the transmission risk of the fox tapeworm

Echinococcus multilocularis (Leukart 1863) (Cestoda : Cyclophyllidea : Taeniidae). *Progress in Physical Geography* **29**, 77–92.

Grannis, G. F. and Lott, J. A. (1978). A technique for determining the probability of abnormality. *Clinical Chemistry* **24**, 640–651.

Greiner, M., Sohr, D. and Göbel, P. (1995). A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *Journal of Immunological Methods* **185**, 123–132.

Greiner, M., Pfeiffer, D. U. and Smith, R. D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine* **45**, 23–41.

Grenfell, B. T. and Harwood, J. (1997). (Meta)population dynamics of infectious diseases. *Trends in Ecology and Evolution* **12**, 395–399.

Grenfell, B. T., Wilson, K., Isham, V. S., Boyd, H. E. and Dietz, K. (1995). Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study. *Parasitology* **111**, S135–S151.

Grueber, C. E., Nakagawa, S., Laws, R. J. and Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology* **24**, 699–711.

Guezala, M. C., Rodriguez, S., Zamora, H., García, H. H., Gonzalez, A. E., Tembo, A., Allan, J. C. and Craig, P. S. (2009). Development of a species-specific coproantigen ELISA for human *Taenia solium* taeniasis. *American Journal of Tropical Medicine and Hygiene* **81**, 433–437.

Gurarie, D. and King, C. H. (2005). Heterogeneous model of schistosomiasis transmission and long-term control: the combined influence of spatial variation and age-dependent factors on optimal allocation of drug therapy. *Parasitology* **130**, 49–65.

Gurarie, D. and King, C. H. (2014). Population biology of *Schistosoma* mating, aggregation, and transmission breakpoints: more reliable model analysis for the end-game in communities at risk. *PloS one* **9**, e115875.

Gurarie, D., King, C. H. and Wang, X. (2010). A new approach to modelling schistosomiasis transmission based on stratified worm burden. *Parasitology* **137**, 1951–1965.

Guyatt, H. L., Bundy, D. A., Medley, G. F. and Grenfell, B. T. (1990). The relationship between the frequency distribution of *Ascaris lumbricoides* and the

prevalence and intensity of infection in human communities. *Parasitology* **101 Pt 1**, 139–143.

Guzel, M., Yaman, M., Koltas, I., Demirkazik, M. and Aktas, H. (2008). Detection of *Echinococcus granulosus* coproantigens in dogs from Antakya Province, Turkey. *Helminthologia* **45**, 150–153.

Hairston, N. G. (1962). On the mathematical analysis of schistosome populations. *Bulletin of the World Health Organization* 45–62.

Hairston, N. G. (1965). An analysis of age-prevalence data by catalytic models. A contribution to the study of bilharziasis. *Bulletin of the World Health Organization* **33**, 163–175.

Hansen, F., Jeltsch, F., Tackmann, K., Staubach, C. and Thulke, H. H. (2000). Controlling parasites: the impact of space. In *9th International Symposium on Veterinary Epidemiology and Economics*, pp. 1–3.

Hansen, F., Tackmann, K., Jeltsch, F., Wissel, C. and Thulke, H. H. (2003). Controlling *Echinococcus multilocularis*—ecological implications of field trials. *Preventive Veterinary Medicine* **60**, 91–105.

Hansen, F., Jeltsch, F., Tackmann, K., Staubach, C. and Thulke, H. H. (2004). Processes leading to a spatial aggregation of *Echinococcus multilocularis* in its natural intermediate host Microtus arvalis. *International Journal for Parasitology* **34**, 37–44.

Hanski, I. (1991). Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society* **42**, 3–16.

Hanski, I. (1998). Metapopulation dynamics. *Nature* **396**, 41–49.

Hanson, T. E. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association* **101**, 1548–1565.

Hanson, T. E., Branscum, A. J. and Gardner, I. A. (2008). Multivariate mixtures of Polya trees for modeling ROC data. *Statistical Modelling* **8**, 81–96.

Hardelid, P., Williams, D., Dezateux, C., Tookey, P. A., Peckham, C. S., Cubitt, W. D. and Cortina-Borja, M. (2008). Analysis of rubella antibody distribution from newborn dried blood spots using finite mixture models. *Epidemiology and Infection* **136**, 1698–706.

Harris, S. (1981). The food of suburban foxes (*Vulpes vulpes*), with special reference to London. *Mammal Review* **11**, 151–168.

Harris, R. E., Revfeim, K. J. and Heath, D. D. (1980). Simulating strategies for control of *Echinococcus granulosus, Taenia hydatigena* and *T. ovis*. *The Journal of Hygiene* **84**, 389–404.

Harris, A., Heath, D. D., Lawrence, S. B. and Shaw, R. J. (1989). *Echinococcus granulosus*: ultrastructure of epithelial changes during the first 8 days of metacestode development in vitro. *International Journal for Parasitology* **19**, 621–9.

Hartnack, S., Budke, C. M., Craig, P. S., Jiamin, Q., Boufana, B., Campos-Ponce, M. and Torgerson, P. R. (2013). Latent-class methods to evaluate diagnostics tests for *Echinococcus* infections in dogs. *PLoS Neglected Tropical Diseases* **7**, e2068.

Hassell, M. P., May, R. M., Pacala, S. W. and Chesson, P. L. (1991a). The persistence of host-parasitoid associations in patchy environments. I. A general criterion. *The American Naturalist* **138**, 568–583.

Hassell, M. P., Comins, H. N. and May, R. M. (1991b). Spatial structure and chaos in insect population dynamics. *Nature* **353**, 255–258.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Heath, D. D. (1971). The migration of oncospheres of *Taenia pisiformis, T. serialis* and *Echinococcus granulosus* within the intermediate host. *International Journal for Parasitology* **1**, 145–152.

Heath, D. D., Jensen, O. and Lightowlers, M. W. (2003). Progress in control of hydatidosis using vaccination - a review of formulation and delivery of the vaccine and recommendations for practical use in control programmes. *Acta Tropica* **85**, 133–143.

Heath, D. D., Lightowlers, M. W., Qiu, J., Yang, W., Zhang, L. H. and McManus, D. P. (2004). Vaccination against hydatidosis and the significance of parasite strain variation in designing a control scheme. In *Echinococcosis in Central Asia: Problems and Solutions*, pp. 61–69.

Hedt, B. L., Olives, C., Pagano, M. and Valadez, J. J. (2008). Large country-lot quality assurance sampling: a new method for rapid monitoring and evaluation of health, nutrition and population programs at sub-national levels.

Heesterbeek, J. A. P. and Dietz, K. (1996). The concept of Ro in epidemic theory. *Statistica Neerlandica* **50**, 89–110.

Heesterbeek, J. A. P. and Roberts, M. G. (1995). Threshold quantities for helminth infections. *Journal of Mathematical Biology* **33**, 415–434.

Hegglin, D. and Deplazes, P. (2013). Control of *Echinococcus multilocularis*: strategies, feasibility and cost-benefit analyses. *International Journal for Parasitology* **43**, 327–337.

Hegglin, D., Ward, P. I. and Deplazes, P. (2003). Anthelmintic baiting of foxes against urban contamination with *Echinococcus multilocularis*. *Emerging Infectious Diseases* **9**, 1266–1272.

Hegglin, D., Bontadina, F., Contesse, P., Gloor, S. and Deplazes, P. (2007). Plasticity of predation behaviour as a putative driving force for parasite life-cycle dynamics: the case of urban foxes and *Echinococcus multilocularis* tapeworm. *Functional Ecology* **21**, 552–560.

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* **36**, 531–547.

Heinzmann, D., Barbour, A. D. and Torgerson, P. R. (2009). Compound processes as models for clumped parasite data. *Mathematical Biosciences* **222**, 27–35.

Heinzmann, D., Barbour, A. D. and Torgerson, P. R. (2011a). Shot noise processes for clumped parasite infections with time-dependent decay dynamics. *Biostatistics, Bioinformatics and Biomathematics*.

Heinzmann, D., Barbour, A. D. and Torgerson, P. R. (2011b). A mechanistic individual-based two-host interaction model for the transmission of a parasitic disease. *International Journal of Biomathematics* **4**, 443–460.

Henderson, R. H., Davis, H., Eddins, D. L. and Foege, W. H. (1973). Assessment of vaccination coverage, vaccination scar rates, and smallpox scarring in five areas of West Africa. *Bulletin of the World Health Organization* **48**, 183–94.

Heptner, V. G. and Naumov, N. P. (1992). *Mammals of the Soviet Union, vol. 2, part 1A.* E.J. Brill, New York.

Hess, G. (1996). Disease in metapopulation models: implications for conservation. *Ecology* **77**, 1617–1632.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical Science* **14**, 382–417.

Hofer, S., Gloor, S., Müller, U., Mathis, A., Hegglin, D. and Deplazes, P. (2000). High prevalence of *Echinococcus multilocularis* in urban red foxes (*Vulpes vulpes*) and voles (*Arvicola terrestris*) in the city of Zürich, Switzerland. *Parasitology* **120**, 135–42.

Hollingsworth, T. D. (2009). Controlling infectious disease outbreaks: lessons from mathematical modelling. *Journal of Public Health Policy* **30**, 328–341.

Holt, R. D. and Hassell, M. P. (1993). Environmental heterogeneity and the stability of host-parasitoid interactions. *Journal of Animal Ecology* **62**, 89–100.

Holt, R. D., Dobson, A. P., Begon, M., Bowers, R. G. and Schauber, E. M. (2003). Parasite establishment in host communities. *Ecology Letters* **6**, 837–842.

Hope, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society B* **30**, 582–598.

Horwitz, P. and Wilcox, B. A. (2005). Parasites, ecosystems and sustainability: an ecological and complex systems perspective. *International Journal for Parasitology* **35**, 725–732.

http://glmm.wikidot.com/faq (2014). DRAFT r-sig-mixed-models FAQ. 1–10.

Huang, Y., Yang, W., Qiu, J., Chen, X., Yang, Y., Qiu, D. C., Xiao, N., Xiao, Y. and Heath, D. D. (2007). A modified coproantigen test used for surveillance of *Echinococcus* spp. in Tibetan dogs. *Veterinary Parasitology* **149**, 229–38.

Huang, Y., Heath, D. D., Yang, W., Qiu, J., Chen, X., Yang, Y., Wang, Q., Li, T. Y., Xiao, Y., Qiu, D. C., Xiao, N., Chen, F., Ge, S. and Se, D. (2008). Epidemiology and risk factor analysis for canine echinococcosis in a Tibetan pastoral area of Sichuan. *Chinese Journal of Parasitology and Parasitic Diseases* **26**, 245–252.

Huang, L., Huang, Y., Wang, Q., Xiao, N., Yi, D., Yu, W. and Qiu, D. C. (2011). An agent-based model for control strategies of *Echinococcus granulosus*. *Veterinary Parasitology* **179**, 84–91.

Huang, Y., Yi, D. Y., Liu, L. L., Huang, L., Yu, W. J., Wang, Q., Li, Y. Q., Han, X. M., Qiu, D. C., Wang, H., Xiao, N., Wu, W. P. and Heath, D. D. (2013). *Echinococcus* infections in Chinese dogs: a comparison of coproantigen kits. *Journal of Helminthology* 1–7.

Hui, S. L. and Walter, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics* **36**, 167–171.

Hurvich, C. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.

Husson, F. and Josse, J. (2014). missMDA: Handling missing values with/in multivariate data analysis (principal component methods). R package version 1.7.3.

Husson, F., Josse, J. and Pagès, J. (2010). Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? *Technical Report - Agrocampus* 1–17.

Husson, F., Lê, S. and Pagès, J. (2011). *Exploratory Multivariate Analysis by Example Using R.* Chapman & Hall / CRC Press.

Husson, F., Josse, J., Lê, S. and Mazet, J. (2015). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R package version 1.29.

Ibrahem, M. M., Craig, P. S., McVie, A., Ersfeld, K. and Rogan, M. T. (1996). *Echinococcus granulosus* antigen B and seroreactivity in natural ovine hydatidosis. *Research in Veterinary Science* **61**, 102–106.

Ibraimova, A., Akkazieva, B., Ibraimov, A., Manzhieva, E. and Rechel, B. (2011). *Kyrgyzstan: Health System Review*.

Inangolet, F., Biffa, D., Opuda-Asibo, J., Oloya, J. and Skjerve, E. (2010). Distribution and intensity of *Echinococcus granulosus* infections in dogs in Moroto District, Uganda. *Tropical Animal Health and Production* **42**, 1451–1457.

Ishikawa, H. (2006). Mathematical modeling of *Echinococcus multilocularis* transmission. *Parasitology International* **55 Suppl**, S259–S261.

Ishikawa, H., Ohga, Y. and Doi, R. (2003). A model for the transmission of *Echinococcus multilocularis* in Hokkaido, Japan. *Parasitology Research* **91**, 444–451.

Islam, A. W. (1980). *Echinococcus granulosus* in dogs in Bangladesh. *American Journal of Veterinary Research* **41**, 415–416.

Ito, A., Sako, Y., Yamasaki, H., Mamuti, W., Nakaya, K., Nakao, M. and Ishikawa, Y. (2003a). Development of Em18-immunoblot and Em18-ELISA for specific diagnosis of alveolar echinococcosis. *Acta Tropica* **85**, 173–182.

Ito, A., Urbani, C., Jiamin, Q., Vuitton, D. A., Qiu, D. C., Heath, D. D., Craig, P. S., Zheng, F. and Schantz, P. M. (2003b). Control of echinococcosis and cysticercosis: a public health challenge to international cooperation in China. *Acta Tropica* **86**, 3–17.

Jafarzadeh, S. R., Johnson, W. O., Utts, J. M. and Gardner, I. A. (2010). Bayesian estimation of the receiver operating characteristic curve for a diagnostic test with a limit of detection in the absence of a gold standard. *Statistics in Medicine* **29**, 2090–2106.

Jakab, M. and Manjieva, E. (2008). The Kyrgyz Republic: good practices in expanding health care coverage, 1991–2006. In *Good Practices in Health Financing* (ed. Gottret, P., Schieber, G. J., and Waters, Hugh, R.), The World Bank.

Jeff Wu, C. F. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics* **11**, 95–103.

Jenkins, D. J. (2005). Hydatid control in Australia: where it began, what we have achieved and where to from here. *International Journal for Parasitology* **35**, 733–740.

Jenkins, D. J. and Morris, B. (1991). Unusually heavy infections of *Echinococcus granulosus* in wild dogs in south-eastern Australia. *Australian Veterinary Journal* **68**, 36–37.

Jenkins, D. J., Gasser, R. B., Zeyhle, E., Romig, T. and Macpherson, C. N. L. (1990). Assessment of a serological test for the detection of *Echinococcus granulosus* infection in dogs in Kenya. *Acta Tropica* **47**, 245–248.

Jenkins, D. J., Fraser, A., Bradshaw, H. and Craig, P. S. (2000). Detection of *Echinococcus granulosus* coproantigens in Australian canids with natural or experimental infection. *Journal of Parasitology* **86**, 140–145.

Jenkins, D. J., Romig, T. and Thompson, R. C. A. (2005). Emergence/re-emergence of *Echinococcus* spp. - a global update. *International Journal for Parasitology* **35**, 1205–1219.

Johansen, M. V. and Penrith, M.-L. (2009). Has culling been properly assessed as a valid and justified control intervention measure for zoonotic diseases? *PLoS Neglected Tropical Diseases* **3**, e541.

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2 GLMM to random slopes models. *Methods in Ecology and Evolution* **5**, 944–946.

Johnson, W. O., Gastwirth, J. L. and Pearson, L. M. (2001). Screening without a "gold standard": the Hui-Walter paradigm revisited. *American Journal of Epidemiology* **153**, 921–924.

Joly, D. O. and Messier, F. (2004). The distribution of *Echinococcus granulosus* in moose: evidence for parasite-induced vulnerability to predation by wolves? *Oecologia* **140**, 586–590.

Jones, A. and Walters, T. M. (1992). A survey of taeniid cestodes in farm dogs in mid-Wales. *Annals of Tropical Medicine and Parasitology* **86**, 137–142.

Josse, J., Chavent, M., Liquet, B. and Husson, F. (2012). Handling missing values with regularized iterative multiple correspondence analysis. *Journal of Classification* **29**, 91–116.

Jung, T. and Wickrama, K. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass* **2**, 302–317.

Kachani, M. and Heath, D. (2014). Dog population management for the control of human echinococcosis. *Acta Tropica* **139**, 99–108.

Kachani, M., Macpherson, C. N. L., Lyagoubi, M., Berrada, M., Bouslikhane, M., Kachani, F. and El-Hasnaoui, M. (2003). Public health education/importance and experience from the field. Educational impact of community-based ultrasound screening surveys. *Acta Tropica* **85**, 263–269.

Kagan, I. G. (1968). A review of serological tests for the diagnosis of hydatid disease. *Bulletin of the World Health Organization* **39**, 25–37.

Kapel, C. M. O., Torgerson, P. R., Thompson, R. C. A. and Deplazes, P. (2006). Reproductive potential of *Echinococcus multilocularis* in experimentally infected foxes, dogs, raccoon dogs and cats. *International Journal for Parasitology* **36**, 79–86.

Kato, N., Kotani, K., Ueno, S. and Matsuda, H. (2010). Optimal risk management of human alveolar echinococcosis with vermifuge. *Journal of Theoretical Biology* **267**, 265–271.

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

Kerven, C. (2006). *Review of the Literature on Pastoral Economics and Marketing: Central Asia, China, Mongolia and Siberia*.

Kerven, C., Channon, J. and Behnke, R. (1996). *Planning and policies on extensive livestock development in central Asia (ODI Working Paper 91)*.

Kerven, C., Steimann, B., Dear, C. and Ashley, L. (2012). Researching the future of pastoralism in Central Asia's mountains: examining development orthodoxies. *Mountain Research and Development* **32**, 368–377.

Keymer, A. (1982). Density-dependent mechanisms in the regulation of intestinal helminth populations. *Parasitology* **84**, 573–587.

King, C. H. and Mahmoud, A. A. F. (1989). Drugs five years later: praziquantel. *Annals of Internal Medicine* **110**, 290–296.

Kittelberger, R., Reichel, M. P., Jenner, J., Heath, D. D., Lightowlers, M. W., Moro, P. L., Ibrahem, M. M., Craig, P. S. and O'Keefe, J. S. (2002). Evaluation of three enzyme-linked immunosorbent assays (ELISAs) for the detection of serum antibodies in sheep infected with *Echinococcus granulosus*. *Veterinary Parasitology* **110**, 57–76.

Knapp, J., Millon, L., Mouzon, L., Umhang, G., Raoul, F., Ali, Z. S., Combes, B., Comte, S., Gbaguidi-Haore, H., Grenouillet, F. and Giraudoux, P. (2014).

Real time PCR to detect the environmental faecal contamination by *Echinococcus multilocularis* from red fox stools. *Veterinary Parasitology* **201**, 40–47.

Kretzschmar, M. and Adler, F. R. (1993). Aggregated distributions in models for patchy populations. *Theoretical population biology* **43**, 1–30.

Kruschke, J. (2011). Overview of the Generalized Linear Model. In *Doing Bayesian Data Analysis*, pp. 357–388.

Kullbank, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics* **22**, 79–86.

Kumaratilake, L. M. and Thompson, R. C. A. (1982). A review of the taxonomy and speciation of the genus *Echinococcus* (Rudolphi 1801). *Zeitschrift fur Parasitenkunde* 121–146.

Kuttubaev, O. T., Kasimbekov, B. K., Omorov, R. A. and Karaeva, R. R. (2004). Echincoccosis in the Kyrgyz Republic. In *Echinococcosis in Central Asia: Problems and Solutions*, pp. 31–42.

Lahmar, S., Kilani, M., Torgerson, P. R. and Gemmell, M. A. (1999). *Echinococcus granulosus* larvae in the livers of sheep in Tunisia: the effects of host age. *Annals of Tropical Medicine and Parasitology* **93**, 75–81.

Lahmar, S., Kilani, M. and Torgerson, P. R. (2001). Frequency distributions of *Echinococcus granulosus* and other helminths in stray dogs in Tunisia. *Annals of Tropical Medicine and Parasitology* **95**, 69–76.

Lahmar, S., Chéhida, F. B., Pétavy, A. F., Hammou, A., Lahmar, J., Ghannay, A., Gharbi, H. A. and Sarciron, M. E. (2007a). Ultrasonographic screening for cystic echinococcosis in sheep in Tunisia. *Veterinary Parasitology* **143**, 42–49.

Lahmar, S., Boufana, B., Bradshaw, H. and Craig, P. S. (2007b). Screening for *Echinococcus granulosus* in dogs: Comparison between arecoline purgation, coproELISA and coproPCR with necropsy in pre-patent infections. *Veterinary Parasitology* **144**, 287–292.

Lambin, E. F., Tran, A., Vanwambeke, S. O., Linard, C. and Soti, V. (2010). Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. *International Journal of Health Geographics* **9**, 54.

Lang, Z. and Reiczigel, J. (2014). Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. *Preventive Veterinary Medicine* **113**, 13–22.

Larrieu, E. J. and Frider, B. (2001). Human cystic echinococcosis: contributions to the natural history of the disease. *Annals of Tropical Medicine and Parasitology* **95**, 679–687.

Larrieu, E. J. and Zanini, F. (2012). Critical analysis of cystic echinococcosis control programs and praziquantel use in South America, 1974-2010. *American Journal of Public Health* **31**, 81–87.

Larrieu, E. J., Costa, M. T., del Carpio, M., Moguillansky, S., Bianchi, G. and Yadon, Z. E. (2002). A case-control study of the risk factors for cystic echinococcosis among the children of Río Negro province, Argentina. *Annals of Tropical Medicine and Parasitology* **96**, 43–52.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics* **20**, 1222–1235.

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics* **22**, 1161–1176.

Lê, S., Josse, J. and Husson, F. (2008). FactoMineR : An R package for multivariate analysis. *Journal of Statistical Software* **25**, 1–18.

Lefcheck, J. and Casallas, J. S. (2014). https://github.com/jslefche/rsquared.glmm.

Lerman, Z. and Sedik, D. (2009). *Agrarian Reform in Kyrgyzstan : Achievements and the Unfinished Agenda (Policy Studies on Rural Transition No. 2009-1)*.

Lett, W. S. (2013). Detection of *Echinococcus granulosus* and *Echinococcus equinus* in dogs and epidemiology of canine echinococcosis in the UK.

Lewis, F., Sanchez-Vazquez, M. J. and Torgerson, P. R. (2012). Association between covariates and disease occurrence in the presence of diagnostic error. *Epidemiology and Infection* **140**, 1515–1524.

Lewis, F. I., Otero-Abad, B., Hegglin, D., Deplazes, P. and Torgerson, P. R. (2014). Dynamics of the force of infection: insights from *Echinococcus multilocularis* infection in foxes. *PLoS Neglected Tropical Diseases* **8**, e2731.

Li, T. Y., Qiu, J. M., Yang, W., Craig, P. S., Chen, X. W., Xiao, N., Ito, A., Giraudoux, P., Wulamu, M., Yu, W. and Schantz, P. M. (2005). Echinococcosis in Tibetan populations, western Sichuan Province, China. *Emerging Infectious Diseases* **11**, 1866–1873.

Li, J., Blakeley, D. and Smith?, R. J. (2011). The failure of Ro. *Computational and Mathematical Methods in Medicine* 527610.

Lidicker, W. Z. (2000). A food web / landscape interaction model for microtine rodent density cycles. *Oikos* **3**, 435–445.

Liechti, K. (2012). The meanings of pasture in resource degradation negotiations: evidence from post-socialist rural Kyrgyzstan. *Mountain Research and Development* **32**, 304–312.

Lightowlers, M. W. and Gottstein, B. (1995). Echinococcosis/hydatidosis: antigens, immunological and molecular diagnosis. In *Echinococcus and Hydatid Disease*, pp. 355–410.

Lightowlers, M. W., Lawrence, S. B., Gauci, C. G., Young, J., Ralston, M. J., Maas, D. and Heath, D. D. (1996). Vaccination against hydatidosis using a defined recombinant antigen. *Parasite Immunology* **18**, 457–462.

Lightowlers, M. W., Jensen, O., Fernandez, E., Iriarte, J. A., Woollard, D. J., Gauci, C. G., Jenkins, D. J. and Heath, D. D. (1999). Vaccination trials in Australia and Argentina confirm the effectiveness of the EG95 hydatid vaccine in sheep. *International Journal for Parasitology* **29**, 531–534.

Lin, T. I., Lee, J. C. and Yen, S. Y. (2007). Finite mixture modelling using the Skew Normal distribution. *Statistica Sinica* **17**, 909–927.

Lindsey, J. K. (1999). On the use of correction for overdispersion. *Applied Statistics* **48**, 553–561.

Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., Fisman, D. and Murray, M. (2009). Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966–1970.

Lloyd, S., Martin, S. C., Walters, T. M. and Soulsby, E. J. (1991). Use of sentinel lambs for early monitoring of the South Powys Hydatidosis Control Scheme: prevalence of *Echinococcus granulosus* and some other helminths. *Veterinary Record* **129**, 73–76.

Lloyd, S., Walters, T. M. and Craig, P. S. (1998). Use of sentinel lambs to survey the effect of an education programme on control of transmission of *Echinococcus granulosus* in South Powys, Wales. *Bulletin of the World Health Organization* **76**, 469–473.

Lukacs, P. M., Burnham, K. P. and Anderson, D. R. (2010). Model selection bias and Freedman's paradox. *Annals of the Institute of Statistical Mathematics* **62**, 117–125.

Lymbery, A. J., Hobbs, R. P. and Thompson, R. C. A. (1989). The dispersion of *Echinococcus granulosus* in the intestine of dogs. *Journal of Parasitology* **75**, 562–570.

MacDonald, G. (1957). *The epidemiology and control of malaria*. Oxford University Press.

Macdonald, G. (1965). The dynamics of helminth infections, with special reference to schistosomes. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **59**, 489–506.

MacDonald, D. W. and Reynolds, J. C. (2008). IUCN. *Vulpes vulpes. The IUCN Red List of Threatened Species. Version 2014.3.*

Machnicka, B., Dziemian, E., Rocki, B. and Kołodziej-Sobocińska, M. (2003). Detection of *Echinococcus multilocularis* antigens in faeces by ELISA. *Parasitology Research* **91**, 491–496.

Macpherson, C. N. L., Romig, T., Zeyhle, E., Rees, P. H. and Were, J. B. (1987). Portable ultrasound scanner versus serology in screening for hydatid cysts in a nomadic population. *The Lancet* **330**, 259–261.

Macpherson, C. N. L., Bartholomot, B. and Frider, B. (2003). Application of ultrasound in diagnosis, treatment, epidemiology, public health and control of *Echinococcus granulosus* and *E. multilocularis*. *Parasitology* **127**, S21–S35.

Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association* **89**, 1535–1546.

Malgor, R., Nonaka, N., Basmadjian, I., Sakai, H., Carambula, B., Oku, Y., Camona, C. and Kamiya, M. (1997). Coproantigen detection in dogs experimentally and naturally infected with *Echinococcus granulosus* by a monoclonal antibody-based enzyme-linked immunosorbent assay. *International Journal for Parasitology* **27**, 1605–1612.

Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* **156**, 636–638.

Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Tyre, A. J. and Possingham, H. P. (2005). Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters* **8**, 1235–1246.

Mastin, A., Brouwer, A., Fox, M., Craig, P. S., Guitián, J., Li, W. and Stevens, K. (2011). Spatial and temporal investigation of *Echinococcus granulosus*

coproantigen prevalence in farm dogs in South Powys, Wales. *Veterinary Parasitology* **178**, 100–107.

Mathis, A. and Deplazes, P. (2006). Copro-DNA tests for diagnosis of animal taeniid cestodes. *Parasitology International* **55**, S87–S90.

Mathis, A., Deplazes, P. and Eckert, J. (1996). An improved test system for PCR-based specific detection of *Echinococcus multilocularis* eggs. *Journal of Helminthology* **70**, 219–222.

Matsudo, K., Inaba, T. and Kamiya, H. (2003). Detection of *Echinococcus multilocularis* eggs by centrifugal flotation technique: preliminary survey of soil left in the ferryboats commuting between Hokkaido Island, where *E. multilocularis* is endemic, and mainland Japan. *Japanese Journal of Infectious Diseases* **56**, 118–119.

Matsumoto, J. and Yagi, K. (2008). Experimental studies on *Echinococcus multilocularis* in Japan, focusing on biohazardous stages of the parasite. *Experimental Parasitology* **119**, 534–541.

Matsumoto, J., Yagi, K., Nonaka, N., Oku, Y. and Kamiya, M. (1998). Time-course of antibody response in mice against oral infection with eggs of *Echinococcus multilocularis*. *Parasitology* **116**, 463–469.

Matsuo, K. and Kamiya, H. (2005). Modified sugar centrifugal flotation technique for recovering *Echinococcus multilocularis* eggs from soil. *The Journal of Parasitology* **91**, 208–209.

Maudlin, I., Eisler, M. C. and Welburn, S. C. (2009). Neglected and endemic zoonoses. *Philosophical Transactions of the Royal Society B* **364**, 2777–2787.

Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992). Polya trees and random distributions. *The Annals of Statistics* **20**, 1203–1221.

Maxon Sage, A., Wachira, T. M., Zeyhle, E. E., Weber, E. P., Njoroge, E. and Smith, G. (1998). Evaluation of diagnostic ultrasound as a mass screening technique for the detection of hydatid cysts in the liver and lung of sheep and goats. *International Journal for Parasitology* **28**, 349–353.

Maxson, A. D., Wachira, T. M., Zeyhle, E. E., Fine, A., Mwangi, T. W. and Smith, G. (1996). The use of ultrasound to study the prevalence of hydatid cysts in the right lung and liver of sheep and goats in Turkana, Kenya. *International Journal for Parasitology* **26**, 1335–1338.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature* **261**, 85–93.

May, R. M. (1977). Dynamical aspects of host-parasite associations: Crofton's model revisited. *Parasitology* **75**, 259–276.

May, R. M. (1978). Host-parasitoid systems in patchy environments: a phenomenological model. *Journal of Animal Ecology* **47**, 833–844.

May, R. M. and Anderson, R. M. (1978). Regulation and stability of host-parasite population interactions: II. Destabilizing processes. *Journal of Animal Ecology* **47**, 249–267.

May, R. M. and Anderson, R. M. (1979). Population biology of infectious diseases: part II. *Nature* **280**, 455–461.

McCallum, H. I. and Anderson, R. M. (1984). Systematic temporal changes in host susceptibility to infection: demographic mechanisms. *Parasitology* **89 (Pt 1)**, 195–208.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Journal of the Royal Statistical Society C* **36**, 318–324.

McLachlan, G. J. and Peel, D. (2000a). ML Fitting of Mixture Models. In *Finite mixture models (Wiley series in probability and statistics)*, pp. 40–80. Wiley-Blackwell.

McLachlan, G. J. and Peel, D. (2000b). Bayesian Approach to Mixture Analysis. In *Finite mixture models (Wiley series in probability and statistics)*, pp. 117–134. Wiley-Blackwell.

McLachlan, G. J. and Peel, D. (2000c). General Introduction. In *Finite mixture models (Wiley series in probability and statistics)*, pp. 1–39. Wiley-Blackwell.

McManus, D. P., Zhang, W., Li, J. and Bartley, P. B. (2003). Echinococcosis. *The Lancet* **362**, 1295–1304.

Medley, G. F. (1992). Which comes first in host-parasite systems: density dependence or parasite distribution? *Parasitology Today* **8**, 321–322.

Medley, G. F., Guyatt, H. L. and Bundy, D. A. (1993). A quantitative framework for evaluating the effect of community treatment on the morbidity due to ascariasis. *Parasitology* **106 (Pt 2**, 211–221.

Meimanaliev, A.-S., Ibraimova, A., Elebesov, B. and Rechel, B. (2005). *Health Care Systems in Transition: Kyrgyzstan*.

Menten, J., Boelaert, M. and Lesaffre, E. (2012). An application of Bayesian growth mixture modelling to estimate infection incidences from repeated serological tests. *Statistical Modelling* **12**, 551–578.

Milner-Gulland, E. J., Torgerson, P. R., Shaikenov, B. S. and Morgan, E. R. (2004). Transmission dynamics of *Echinococcus multilocularis* in a patchy environment. In *Species conservation and management. Case studies.*, pp. 179–189.

Ming, R., Tolley, H. D., Andersen, F. L., Chai, J. and Chang, Q. (1992). Frequency distribution of *Echinococcus granulosus* in dog populations in the Xinjiang Uygur Autonomous Region, China. *Veterinary Parasitology* **43**, 233–241.

Mitchell, S. and Pagano, M. (2012). Pooled testing for effective estimation of the prevalence of *Schistosoma mansoni*. *American Journal of Tropical Medicine and Hygiene* **87**, 850–861.

Moks, E., Jõgisalu, I., Valdmann, H. and Saarma, U. (2008). First report of *Echinococcus granulosus* G8 in Eurasia and a reappraisal of the phylogenetic relationships of "genotypes" G5-G10. *Parasitology* **135**, 647–654.

Mollison, D., Isham, V. S. and Grenfell, B. T. (1994). Epidemics: models and data. *Journal of the Royal Statistical Society A* **157**, 115–149.

Molyneux, D. H. (2012). The "Neglected Tropical Diseases": now a brand identity; responsibilities, context and promise. *Parasites & Vectors* **5**, 23.

Molyneux, D. H., Hallaj, Z., Keusch, G. T., McManus, D. P., Ngowi, H., Cleaveland, S., Ramos-Jimenez, P., Gotuzzo, E., Kar, K., Sanchez, A., Garba, A., Carabin, H., Bassili, A., Chaignat, C. L., Meslin, F.-X., Abushama, H. M., Willingham, A. L. and Kioy, D. (2011). Zoonoses and marginalised infectious diseases of poverty: Where do we stand? *Parasites & Vectors* **4**, 106.

Monnier, P., Cliquet, F., Aubert, M. F. and Bretagne, S. (1996). Improvement of a polymerase chain reaction assay for the detection of *Echinococcus multilocularis* DNA in faecal samples of foxes. *Veterinary Parasitology* **67**, 185–195.

Montresor, A., Crompton, D. W. T., Hall, A., Bundy, D. A. P. and Savioli, L. (1998). *Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at community level*. Geneva.

Morel, N., Lassabe, G., Elola, S., Bondad, M., Herrera, S., Mari, C., Last, J. A., Jensen, O. and Gonzalez-Sapienza, G. (2013). A monoclonal antibody-based copro-ELISA kit for canine echinococcosis to support the PAHO effort for hydatid disease control in South America. *PLoS Neglected Tropical Diseases* **7**, e1967.

Morgan, E. R., Milner-Gulland, E. J., Torgerson, P. R. and Medley, G. F. (2004). Ruminating on complexity: macroparasites of wildlife and livestock. *Trends in Ecology and Evolution* **19**, 181–188.

Morishima, Y., Tsukada, H., Nonaka, N. and Oku, Y. (1999a). Evaluation of coproantigen diagnosis for natural *Echinococcus multilocularis* infection in red foxes. *Japanese Journal of Veterinary Research* **46**, 185–189.

Morishima, Y., Tsukada, H., Nonaka, N., Oku, Y. and Kamiya, M. (1999b). Coproantigen survey for *Echinococcus multilocularis* prevalence of red foxes in Hokkaido , Japan. *Parasitology International* **48**, 121–134.

Moro, P. L., Bonifacio, N., Gilman, R. H., Lopera, L., Silva, B., Takumoto, R., Verastegui, M. and Cabrera, L. (1999). Field diagnosis of *Echinococcus granulosus* infection among intermediate and definitive hosts in an endemic focus of human cystic echinococcosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93**, 611–615.

Moro, P. L., Lopera, L., Bonifacio, N., Gonzales, A., Gilman, R. H. and Moro, M. H. (2005). Risk factors for canine echinococcosis in an endemic area of Peru. *Veterinary Parasitology* **130**, 99–104.

Moro, P. L., Cavero, C. A., Tambini, M., Briceño, Y., Jiménez, R. and Cabrera, L. (2008). Identification of risk factors for cystic echinococcosis in a peri-urban population of Peru. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **102**, 75–78.

Moss, J. E., Chen, X., Li, T., Qiu, J., Wang, Q., Giraudoux, P., Ito, A., Torgerson, P. R. and Craig, P. S. (2013). Reinfection studies of canine echinococcosis and role of dogs in transmission of *Echinococcus multilocularis* in Tibetan communities, Sichuan, China. *Parasitology* **140**, 1685–92.

Muench, H. (1959). *Catalytic models in epidemiology*. Harvard University Press.

Murdoch, J. D. (2014). *Vulpes corsac*. The IUCN Red List of Threatened Species. Version 2014.3.

Murdoch, J. D., Munkhzul, T., Buyandelger, S., Reading, R. P. and Sillero-Zubiri, C. (2010). Seasonal food habits of corsac and red foxes in Mongolia and the potential for competition. *Mammalian Biology* **75**, 36–44.

Muthén, B. O. (2002). Beyond SEM: general latent variable modeling. *Behaviormetrika* **29**, 81–117.

Muthén, B. and Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism, Clinical and Experimental Research* **24**, 882–891.

Nagin, D. S. and Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology* **31**, 327–362.

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133–142.

Nakao, M., McManus, D. P., Schantz, P. M., Craig, P. S. and Ito, A. (2007). A molecular phylogeny of the genus *Echinococcus* inferred from complete mitochondrial genomes. *Parasitology* **134**, 713–722.

Nakao, M., Xiao, N., Okamoto, M., Yanagida, T., Sako, Y. and Ito, A. (2009). Geographic pattern of genetic variation in the fox tapeworm *Echinococcus multilocularis*. *Parasitology International* **58**, 384–389.

Nakao, M., Lavikainen, A., Yanagida, T. and Ito, A. (2013). Phylogenetic systematics of the genus *Echinococcus* (Cestoda: Taeniidae). *International Journal for Parasitology* **43**, 1017–1029.

Neuenschwander, B. E., Zwahlen, M., Kim, S. J., Engel, R. R. and Rieder, H. L. (2000). Trends in the prevalence of infection with mycobacterium tuberculosis in Korea from 1965 to 1995: an analysis of seven surveys by mixture models. *The International Journal of Tuberculosis and Lung Disease* **4**, 719–729.

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics* **8**, 343–366.

Ni, X.-W., McManus, D. P., Lou, Z.-Z., Yang, J.-F., Yan, H.-B., Li, L., Li, H.-M., Liu, Q.-Y., Li, C.-H., Shi, W.-G., Fan, Y.-L., Liu, X., Cai, J.-Z., Lei, M.-T., Fu, B.-Q., Yang, Y.-R. and Jia, W.-Z. (2014a). A comparison of loop-mediated isothermal amplification (LAMP) with other surveillance tools for *Echinococcus granulosus* diagnosis in canine definitive hosts. *PLoS One* **9**, e100877.

Ni, X.-W., McManus, D. P., Yan, H., Yang, J., Lou, Z.-Z., Li, H.-M., Li, L., Lei, M.-T., Cai, J.-Z., Fan, Y.-L., Li, C.-H., Liu, Q.-Y., Shi, W.-G., Liu, X., Zheng, Y., Fu, B.-Q., Yang, Y. and Jia, W.-Z. (2014b). Loop-Mediated Isothermal Amplification (LAMP) assay for the identification of *Echinococcus multilocularis* infections in canine definitive hosts. *Parasites & Vectors* **7**, 254.

Nikmanesh, B., Mirhendi, H., Ghalavand, Z., Alebouyeh, M., Sharbatkhori, M., Kia, E., Mohebali, M. and Rokni, M. B. (2014). Genotyping of *Echinococcus granulosus* isolates from human clinical samples based on sequencing of mitochondrial genes in Iran, Tehran. *Iranian Journal of Parasitology* **9**, 20–27.

Nishina, T. and Ishikawa, H. (2008). A stochastic model of *Echinococcus multilocularis* transmission in Hokkaido, Japan, focusing on the infection process. *Parasitology Research* **102**, 465–479.

Nødtvedt, A., Dohoo, I., Sanchez, J., Conboy, G., DesCoteaux, L., Keefe, G., Leslie, K. and Campbell, J. (2002). The use of negative binomial modelling in a longitudinal study of gastrointestinal parasite burdens in Canadian dairy cows. *Canadian Journal of Veterinary Research* **66**, 249–257.

Nonaka, N., Iida, M., Yagi, K., Ito, T., Ooi, H. K., Oku, Y. and Kamiya, M. (1996). Time course of coproantigen excretion in *Echinococcus multilocularis* infections in foxes and an alternative definitive host, golden hamsters. *International Journal for Parasitology* **26**, 1271–1278.

O Carroll, F. M. (1962). Fitting a negative binomial distribution to coarsely grouped data by maximum likelihood. *Journal of the Royal Statistical Society C* **11**, 196–201.

OIE (2010). Stray dog population control. In *Terrestrial Animal Health Code*, pp. 1–17.

Oliveira-Brochado, A. and Martins, F. V. (2005). *Assessing the Number of Components in Mixture Models : A Review (FEP Working Papers 194)*.

Olives, C., Valadez, J. J., Brooker, S. J. and Pagano, M. (2012). Multiple category-lot quality assurance sampling: a new classification system with application to schistosomiasis control. *PLoS Neglected Tropical Diseases* **6**, e1806.

Opel, K. L., Chung, D. and McCord, B. R. (2010). A study of PCR inhibition mechanisms using real time PCR. *Journal of Forensic Sciences* **55**, 25–33.

Ostfeld, R. S., Glass, G. E. and Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology and Evolution* **20**, 328–336.

Otero-Abad, B. and Torgerson, P. R. (2013). A systematic review of the epidemiology of echinococcosis in domestic and wild animals. *PLoS Neglected Tropical Diseases* **7**, e2249.

Paarmann, B. (2009). Revisiting Sary Moghul (https://www.neweurasia.net/culture-and-history/revisiting-sary-moghul). *neweurasia*.

Pacala, S. W. and Dobson, A. P. (1988). The relation between the number of parasites/host and host age: population dynamic causes and maximum likelihood estimation. *Parasitology* **96**, 197–210.

Pacala, S. W., Hassell, M. P. and May, R. M. (1990). Host-parasitoid associations in patchy environments. *Nature* **344**, 150–153.

Paddock, S. M., Ruggeri, F., Lavine, M. and West, M. (2003). Randomized Polya Tree models for nonparametric Bayesian inference. *Statistica Sinica* **13**, 443–460.

Pagano, M. and Valadez, J. J. (2010). Commentary: Understanding practical lot quality assurance sampling. *International Journal of Epidemiology* **39**, 69–71.

Palmer, S. R., Biffin, A. H., Craig, P. S. and Walters, T. M. (1996). Control of hydatid disease in Wales. *British Medical Journal* 674–675.

Pappas, G., Papadimitriou, P., Akritidis, N., Christou, L. and Tsianos, E. V. (2006). The new global map of human brucellosis. *The Lancet Infectious Diseases* **6**, 91–99.

Parada, L., Cabrera, P., Burges, C., Acuna, A., Barcelona, C., Laurenson, M. K., Gulland, F. M., Aqulla, J., Parietti, S., Paolillo, E. and Botta, B. (1995). *Echinococcus granulosus* infections of dogs in the Durazno region of Uruguay. *Veterinary Record* **136**, 389–391.

Parker, R. A., Erdman, D. D. and Anderson, L. J. (1990). Use of mixture models in determining laboratory criterion for identification of seropositive individuals: application to parvovirus B19 serology. *Journal of Virological Methods* **27**, 135–144.

Pascual, M. and Guichard, F. (2005). Criticality and disturbance in spatial ecological systems. *Trends in Ecology and Evolution* **20**, 88-95.

Pavlovsky, E. N. (1966). *Natural Nidality of Transmissible Diseases, With Special Reference to the Landscape Epidemiology of Zooanthroponses*. University of Illinois Press.

Pearce, N. and Merletti, F. (2006). Complexity, simplicity, and epidemiology. *International Journal of Epidemiology* **35**, 515–519.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society A* **185**, 71–110.

Perkins, S. E., Cattadori, I. M., Tagliapietra, V., Rizzoli, A. P. and Hudson, P. J. (2003). Empirical evidence for key hosts in persistence of a tick-borne disease. *International Journal for Parasitology* **33**, 909–917.

Piepenburg, O., Williams, C. H., Stemple, D. L. and Armes, N. A. (2006). DNA detection using recombination proteins. *PLoS Biology* **4**, 1115–1121.

Pleydell, D. R. J., Raoul, F., Tourneux, F., Danson, F. M., Graham, A. J., Craig, P. S. and Giraudoux, P. (2004). Modelling the spatial distribution of *Echinococcus multilocularis* infection in foxes. *Acta Tropica* **91**, 253–265.

Pleydell, D. R. J., Yang, Y. R., Danson, F. M., Raoul, F., Craig, P. S., McManus, D. P., Vuitton, D. A., Wang, Q. and Giraudoux, P. (2008). Landscape composition and spatial prediction of alveolar echinococcosis in southern Ningxia, China. *PLoS Neglected Tropical Diseases* **2**, e287.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria.

Plummer, M. (2013). rjags: Bayesian graphical models using MCMC.

Poulin, R. (1995). "Adaptive" parasitized changes in the behaviour of animals: a critical review. *International Journal for Parasitology* **25**, 1371–1383.

Poulin, R. (2007). Are there general laws in parasite ecology? *Parasitology* **134**, 763–776.

Poulin, R. (2013). Explaining variability in parasite aggregation levels among host samples. *Parasitology* **140**, 541–546.

Praet, N., Verweij, J. J., Mwape, K. E., Phiri, I. K., Muma, J. B., Zulu, G., van Lieshout, L., Rodriguez-Hidalgo, R., Benitez-Ortiz, W., Dorny, P. and Gabriël, S. (2013). Bayesian modelling to estimate the test characteristics of coprology, coproantigen ELISA and a novel real-time PCR for the diagnosis of taeniasis. *Tropical Medicine & International Health* **18**, 608–614.

Priebe, C. E. (1994). Adaptive mixtures. *Journal of the American Statistical Association* **89**, 796–806.

Qiagen (2010). *QIAamp DNA Stool Handbook*. Second edi. Hilden, Germany.

Quinnell, R. J. (2003). Genetics of susceptibility to human helminth infection. *International Journal for Parasitology* **33**, 1219–1231.

Quinnell, R. J., Medley, G. F. and Keymer, A. E. (1990). The regulation of gastrointestinal helminth populations. *Philosophical Transactions of the Royal Society B* **330**, 191–201.

Quinnell, R. J., Grafen, A. and Woolhouse, M. E. J. (1995). Changes in parasite aggregation with age: a discrete infection model. *Parasitology* **111**, 635–644.

R Development Core Team (2014). R: a language and environment for statistical computing.

Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). In *Sociological Methodology* (ed. Marsden, P. V.), pp. 111–163. Blackwell, Cambridge, Mass.

Raoul, F., Deplazes, P., Nonaka, N., Piarroux, R., Vuitton, D. A. and Giraudoux, P. (2001). Assessment of the epidemiological status of *Echinococcus multilocularis* in foxes in France using ELISA coprotests on fox faeces collected in the field. *International Journal for Parasitology* **31**, 1579–1588.

Raoul, F., Deplazes, P., Rieffel, D., Lambert, J. C. and Giraudoux, P. (2010). Predator dietary response to prey density variation and consequences for cestode transmission. *Oecologia* **164**, 129–139.

Rausch, R. L. and Wilson, J. F. (1973). Rearing of the adult *Echinococcus multilocularis* (Leuckart, 1863) from sterile larvae from man. *American Journal of Tropical Medicine and Hygiene* **22**, 357–360.

Rechel, B., Roberts, B., Richardson, E., Shishkin, S., Shkolnikov, V. M., Leon, D. A., Bobak, M., Karanikolos, M. and McKee, M. (2013). Health and health systems in the Commonwealth of Independent States. *The Lancet* **381**, 1145–1155.

Reeve, J. D., Cronin, J. T. and Strong, D. R. (1994). Parasitoid aggregation and the stabilization of a salt marsh host-parasitoid system. *Ecology* **75**, 288–295.

Reiczigel, J., Foldi, J. and Ozsvari, L. (2010). Exact confidence limits for prevalence of a disease with an imperfect diagnostic test. *Epidemiology and Infection* **138**, 1674–1678.

Reiterová, K., Miterpáková, M., Turceková, L., Antolová, D. and Dubinsky, P. (2005). Field evaluation of an intravital diagnostic test of *Echinococcus multilocularis* infection in red foxes. *Veterinary Parasitology* **128**, 65–71.

Richardson, M. D., Turner, A., Warnock, D. W. and Llewellyn, P. A. (1983). Computer-assisted rapid enzyme-linked immunosorbent assay (ELISA) in the serological diagnosis of aspergillosis. *Journal of Immunological Methods* **56**, 201–207.

Rinaldi, L., Maurelli, M. P., Veneziano, V., Capuano, F., Perugini, A. G. and Cringoli, S. (2008). The role of cattle in the epidemiology of *Echinococcus granulosus* in an endemic area of southern Italy. *Parasitology Research* **103**, 175–179.

Robardet, E., Giraudoux, P., Caillot, C., Boué, F., Cliquet, F., Augot, D. and Barrat, J. (2008). Infection of foxes by *Echinococcus multilocularis* in urban and suburban areas of Nancy, France: influence of feeding habits and environment. *Parasite* **15**, 77–85.

Roberts, M. G. (2007). The pluses and minuses of Ro. *Journal of the Royal Society, Interface* **4**, 949–961.

Roberts, M. G. and Aubert, M. F. (1995). A model for the control of *Echinococcus multilocularis* in France. *Veterinary Parasitology* **56**, 67–74.

Roberts, M. G. and Grenfell, B. T. (1991). The population dynamics of nematode infections of ruminants: periodic perturbations as a model for management. *IMA Journal of Mathematics Applied in Medicine and Biology* **8**, 83–93.

Roberts, M. G. and Grenfell, B. T. (1992). The population dynamics of nematode infections of ruminants: the effect of seasonality in the free-living stages. *IMA Journal of Mathematics Applied in Medicine and Biology* **9**, 29–41.

Roberts, M. G. and Heesterbeek, J. A. P. (1995). The dynamics of nematode infections of farmed ruminants. *Parasitology* **110 (Pt 4**, 493–502.

Roberts, M. G., Lawson, J. R. and Gemmell, M. A. (1986). Population dynamics in echinococcosis and cysticercosis: mathematical model of the life-cycle of *Echinococcus granulosus*. *Parasitology* **92 (Pt 3)**, 621–641.

Roberts, M. G., Lawson, J. R. and Gemmell, M. A. (1987). Population dynamics in echinococcosis and cysticercosis: mathematical model of the life-cycles of Taenia hydatigena and T. ovis. *Parasitology* **94 (Pt 1)**, 181–197.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77.

Rogan, W. J. and Gladen, B. (1978). Estimating prevalence from the results of a screening test. *American Journal of Epidemiology* **107**, 71–76.

Romig, T., Dinkel, A. and Mackenstedt, U. (2006). The present situation of echinococcosis in Europe. *Parasitology International* **55**, 187–191.

Romig, T., Omer, R. A., Zeyhle, E., Huttner, M., Dinkel, A., Siefert, L., Elmahdi, I. E., Magambo, J., Ocaido, M., Menezes, C. N., Ahmed, M. E., Mbae, C., Grobusch, M. P. and Kern, P. (2011). Echinococcosis in sub-Saharan Africa: emerging complexity. *Veterinary Parasitology* **181**, 43–47.

Rosà, R. and Pugliese, A. (2002). Aggregation, stability, and oscillations in different models for host-macroparasite interactions. *Theoretical Population Biology* **61**, 319–334.

Rosà, R., Pugliese, A., Villani, A. and Rizzoli, A. (2003). Individual-based vs. deterministic models for macroparasites: host cycles and extinction. *Theoretical Population Biology* **63**, 295–307.

Roth, F., Zinsstag, J., Orkhon, D., Chimed-Ochir, G., Hutton, G., Cosivi, O., Carrin, G. and Otte, J. (2003). Human health benefits from livestock vaccination for brucellosis: case study. *Bulletin of the World Health Organization* **81**, 867–876.

Rothman, K. J. and Greenland, S. (2005). Causation and causal inference in epidemiology. *American Journal of Public Health* **95 Suppl 1**, S144–50.

Rudge, J. W., Webster, J. P., Lu, D.-B., Wang, T.-P., Fang, G.-R. and Basáñez, M.-G. (2013). Identifying host species driving transmission of schistosomiasis japonica, a multihost parasite system, in China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11457–11462.

Rushforth, N. B., Bennett, P. H., Steinberg, A. G., Burch, T. A. and Miller, M. (1971). Diabetes in the Pima Indians. Evidence of bimodality in glucose tolerance distributions. *Diabetes* **20**, 756–765.

Rysmukhambetova, A. T., Shaikenov, B. S. and Abdybekova, A. M. (2004). Dynamics of taeniid infections of dogs in southern Kazakhstan. In *Echinococcosis in Central Asia: Problems and Solutions*, pp. 76–91.

Saarma, U., Jogisalu, I., Moks, E., Varcasia, A., Lavikainen, A., Oksanen, A., Simsek, S., Andresiuk, V., Denegri, G., Gonzalez, L. M., Ferrer, E., Garate, T., Rinaldi, L. and Maravilla, P. (2009). A novel phylogeny for the genus *Echinococcus*, based on nuclear data, challenges relationships based on mitochondrial evidence. *Parasitology* **136**, 317–328.

Sadjjadi, S. M., Mikaeili, F., Karamian, M., Maraghi, S., Sadjjadi, F. S., Shariat-Torbaghan, S., Kia, E. B. (2013). Evidence that the *Echinococcus granulosus* G6 genotype has an affinity for the brain in humans. *International Journal for Parasitology* **43**, 875–877.

Sakamoto, T. and Sugimura, M. (1970). Studies on echinococcosis. 23. Electron microscopical observations on histogenesis of larval *Echinococcus multilocularis*. *Japanese Journal of Veterinary Research* **18**, 131–144.

Salant, H., Abbasi, I. and Hamburger, J. (2012). The development of a loop-mediated isothermal amplification method (LAMP) for *Echinococcus granulosus* coprodetection. *American Journal of Tropical Medicine and Hygiene* **87**, 883–887.

Sargaldakova, A., Healy, J., Kutzin, J. and Gedik, G. (2000). *Health Care Systems in Transition: Kyrgyzstan*.

Šarkūnas, M. and Deplazes, P. (2014). The effects of seasonal praziquantel dosing of dogs on the prevalence of *Echinococcus granulosus* (G7) in dogs and pigs in Lithuanian villages. In *European Scientific Counsel for Companion Animal Parasites (ESCCAP) Echinococcus*, Vilnius, Lithuania.

Schaefer, J. W. and Khan, M. Y. (1991). Echinococcosis (hydatid disease): lessons from experience with 59 patients. *Reviews of Infectious Diseases* **13**, 243–247.

Schantz, P. M. (1997). Sources and use of surveillance data for cystic echinococcosis. In *Compendium on Cystic Echinococcosis in Africa and in Middle Eastern Countries with special reference to Morocco* (ed. Andersen, F., Ouhelli, H., and Kachani, M.), pp. 72–84. Brigham Young University Print Services, Provo, Utah.

Schantz, P. M., Chai, J. J., Craig, P. S., Eckert, J., Jenkins, D. J., Macpherson, C. N. L. and Thakur, A. (1995). Epidemiology and control of hydatid disease. In *Echinococcus and Hydatid Disease*, pp. 233–331.

Schelling, U., Frank, W., Will, R., Romig, T. and Lucius, R. (1997). Chemotherapy with praziquantel has the potential to reduce the prevalence of *Echinococcus multilocularis* in wild foxes (Vulpes vulpes). *Annals of Tropical Medicine and Parasitology* **91**, 179–186.

Schillborn-van Veen, T. W. (1995). *The Kyrgyz Sheep herders at the Crossroads (Pastoral Development Network Papers 38d)*.

Schmidt, P. (2001). The scientific world and the farmer's reality: agricultural research and extension in Kyrgyzstan. *Mountain Research and Development* **21**, 109–112.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.

Shaikenov, B. S. (2004). Seasonal dynamics of the transmission of *Echinococcus multilocularis*. In *Echinococcosis in Central Asia: Problems and Solutions*, pp. 283–288.

Shaikenov, B. S. (2006). Distribution and ecology of *Echinococcus multilocularis* in Central Asia. *Parasitology International* **55**, S213–S219.

Shaikenov, B. S., Torgerson, P. R., Usenbayev, A. E., Baitursinov, K. K., Rysmukhambetova, A. T., Abdybekova, A. M. and Karamendin, K. O. (2003). The changing epidemiology of echinococcosis in Kazakhstan due to transformation of farming practices. *Acta Tropica* **85**, 287–293.

Shaikenov, B. S., Rysmukhambetova, A. T., Massenov, B., Deplazes, P., Mathis, A. and Torgerson, P. R. (2004). Short report: the use of a polymerase chain reaction to detect *Echinococcus granulosus* (G1 strain) eggs in soil samples. *American Journal of Tropical Medicine and Hygiene* **71**, 441–443.

Sharifi, I. and Zia-Ali, N. (1996). The present status and intensity of *Echinococcus granulosus* infection in 391 stray dogs in rural and urban areas of the city of Kerman, Iran. *Iranian Journal of Public Health* **25**, 13–20.

Shaw, D. J. and Dobson, A. P. (1995). Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* **111**, S111–S127.

Shaw, D. J., Grenfell, B. T. and Dobson, A. P. (1998). Patterns of macroparasite aggregation in wildlife host populations. *Parasitology* **117**, 597–610.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. (2005). ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941.

Skjerve, E. and Glattre, E. (2006). Fractal phenomena and fractal analysis in epidemiological studies. In *11th International Symposium on Veterinary Epidemiology and Economics*, .

Smyth, J. D. (1964). The biology of the hydatid organisms. *Advances in Parasitology* **7**, 169–219.

Soetaert, K., Petzoldt, T. and Setzer, R. W. (2010). Package deSolve : solving initial value differential equations in R. *Journal Of Statistical Software* **33**, 1–25.

Soriano, S. V., Pierangeli, N. B., Pianciola, L., Mazzeo, M., Lazzarini, L. E., Saiz, M. S., Kossman, A. V., Bergagna, H. F. J., Chartier, K. and Basualdo, J. A. (2010). Molecular characterization of *Echinococcus* isolates indicates goats as reservoir for *Echinococcus canadensis* G6 genotype in Neuquén, Patagonia Argentina. *Parasitology International* **59**, 626–8.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* **64**, 583–639.

Stammbach, M. (2009). Sheep in Kyrgyzstan: a comparative analysis of different management techniques to enhance the output of production.

Staubach, C., Thulke, H. H., Tackmann, K., Hugh-Jones, M. and Conraths, F. J. (2001). Geographic information system-aided analysis of factors associated with the spatial distribution of *Echinococcus multilocularis* infections of foxes. *American Journal of Tropical Medicine and Hygiene* **65**, 943–948.

Stefanić, S., Shaikenov, B. S., Deplazes, P., Dinkel, A., Torgerson, P. R. and Mathis, A. (2004). Polymerase chain reaction for detection of patent infections of *Echinococcus granulosus* ("sheep strain") in naturally infected dogs. *Parasitology Research* **92**, 347–351.

Sterne, T. E. (1954). Some remarks on confidence or fiducial limits. *Biometrika* **41**, 275–278.

Stevenson, M., Nunes, T., Sanchez, J., Thornton, R., Reiczigel, J., Robison-Cox, J., Sebastiani, P. and Solymos, P. (2013). epiR: An R package for the analysis of epidemiological data (version 0.9-48).

Stickley, A., Koyanagi, A., Richardson, E., Roberts, B., Balabanova, D. and McKee, M. (2013). Prevalence and factors associated with the use of alternative (folk) medicine practitioners in 8 countries of the former Soviet Union. *BMC Complementary and Alternative Medicine* **13**, 83.

Stieger, C., Hegglin, D., Schwarzenbach, G., Mathis, A. and Deplazes, P. (2002). Spatial and temporal aspects of urban transmission of *Echinococcus multilocularis*. *Parasitology* **124**, 631–640.

Subtil, F., Boussari, O., Bastard, M., Etard, J.-F., Ecochard, R. and Genolini, C. (2014). An alternative classification to mixture modeling for longitudinal counts or binary measures. *Statistical Methods in Medical Research* **September**.

Susser, M. and Susser, E. (1996a). Choosing a future for epidemiology: I. Eras and paradigms. *American Journal of Public Health* **86**, 668–673.

Susser, M. and Susser, E. (1996b). Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *American Journal of Public Health* **86**, 674–677.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science* **240**, 1285–1293.

Tackmann, K., Löschner, U., Mix, H., Staubach, C., Thulke, H. H. and Conraths, F. J. (1998). Spatial distribution patterns of *Echinococcus multilocularis* (Leuckart 1863) (Cestoda: Cyclophyllidea: Taeniidae) among red foxes in an endemic focus in Brandenburg, Germany. *Epidemiology and infection* **120**, 101–109.

Tackmann, K., Löschner, U., Mix, H., Staubach, C., Thulke, H. H., Ziller, M. and Conraths, F. J. (2001). A field study to control *Echinococcus multilocularis* - infections of the red fox (*Vulpes vulpes*) in an endemic focus. *Epidemiology and Infection* **127**, 577–587.

Takumi, K. and van der Giessen, J. W. B. (2005). Transmission dynamics of *Echinococcus multilocularis*; its reproduction number, persistence in an area of low rodent prevalence, and effectiveness of control. *Parasitology* **131**, 133–140.

Takumi, K., de Vries, A., Chu, M. L., Mulder, J., Teunis, P. and van der Giessen, J. W. B. (2008). Evidence for an increasing presence of *Echinococcus multilocularis* in foxes in The Netherlands. *International Journal for Parasitology* **38**, 571–578.

Takumi, K., Hegglin, D., Deplazes, P., Gottstein, B., Teunis, P. and van der Giessen, J. (2012). Mapping the increasing risk of human alveolar echinococcosis in Limburg, The Netherlands. *Epidemiology and Infection* **140**, 867–871.

Tasker, L. (2008). *Methods for the euthanasia of dogs and cats: comparison and recommendations*.

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature* **189**, 732 – 735.

Temple, H., Jones, P. H. and Brouwer, A. (2013). Distribution of *Echinococcus granulosus* in Wales and Great Britain using cattle as sentinel hosts. In *Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine*, pp. 57–64. Madrid.

Thompson, R. C. A. (1991). *Echinococcus* and *Giardia*: variation on a theme. *International Journal for Parasitology* **21**, 291–297.

Thompson, R. C. A. (1995). Biology and systematics of *Echinococcus*. In *Echinococcus and Hydatid Disease*, pp. 1–50.

Thompson, R. C. A. (2008). The taxonomy, phylogeny and transmission of *Echinococcus*. *Experimental Parasitology* **119**, 439–446.

Thompson, R. C. A. and Eckert, J. (1983). Observations on *Echinococcus multilocularis* in the definitive host. *Zeitschrift fur Parasitenkunde* **69**, 335–345.

Thompson, R. C. A. and McManus, D. P. (2002). Towards a taxonomic revision of the genus *Echinococcus*. *Trends in Parasitology* **18**, 452–457.

Thompson, R. C. A., Lymbery, A. J. and Constantine, C. C. (1995). Variation in *Echinococcus*: towards a taxonomic revision of the genus. *Advances in Parasitology* **35**, 145–176.

Thurmond, M. C., Johnson, W. O., Muñoz-Zanzi, C. A., Su, C.-L. and Hietala, S. K. (2002). A method of probability diagnostic assignment that applies Bayes theorem for use in serologic diagnostics, using an example of *Neospora caninum* infection in cattle. *American Journal of Veterinary Research* **63**, 318–325.

Toft, N., Jørgensen, E. and Højsgaard, S. (2005). Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine* **68**, 19–33.

Torgerson, P. R. (2003a). Economic effects of echinococcosis. *Acta Tropica* **85**, 113–118.

Torgerson, P. R. (2003b). The use of mathematical models to simulate control options for echinococcosis. *Acta Tropica* **85**, 211–221.

Torgerson, P. R. (2006a). Mathematical models for the control of cystic echinococcosis. *Parasitology International* **55**, S253–S258.

Torgerson, P. R. (2006b). Canid immunity to *Echinococcus* spp.: impact on transmission. *Parasite Immunology* **28**, 295–303.

Torgerson, P. R. (2013). The emergence of echinococcosis in Central Asia. *Parasitology* 1–7.

Torgerson, P. R. and Budke, C. M. (2003). Echinococcosis – an international public health challenge. *Research in Veterinary Science* **74**, 191–202.

Torgerson, P. R. and Deplazes, P. (2009). Echinococcosis: diagnosis and diagnostic interpretation in population studies. *Trends in Parasitology* **25**, 164–170.

Torgerson, P. R. and Dowling, P. M. (2001). Estimating the economic effects of cystic echinococcosis. Part 2: an endemic region in the United Kingdom, a wealthy, industrialized economy. *Annals of Tropical Medicine and Parasitology* **95**, 177–185.

Torgerson, P. R. and Heath, D. D. (2003). Transmission dynamics and control options for *Echinococcus granulosus*. *Parasitology* **127**, S143–S158.

Torgerson, P. R., Williams, D. H. and Abo-Shehada, M. N. (1998). Modelling the prevalence of *Echinococcus* and Taenia species in small ruminants of different ages in northern Jordan. *Veterinary Parasitology* **79**, 35–51.

Torgerson, P. R., Carmona, C. and Bonifacino, R. (2000). Estimating the economic effects of cystic echinococcosis: Uruguay, a developing country with upper-middle income. *Annals of Tropical Medicine and Parasitology* **94**, 703–713.

Torgerson, P. R., Dowling, P. M. and Abo-Shehada, M. N. (2001). Estimating the economic effects of cystic echinococcosis. Part 3: Jordan, a developing country with lower-middle income. *Annals of Tropical Medicine and Parasitology* **95**, 593–603.

Torgerson, P. R., Karaeva, R. R., Corkeri, N., Abdyjaparov, T. A., Kuttubaev, O. T. and Shaikenov, B. S. (2003a). Human cystic echinococcosis in Kyrgystan: an epidemiological study. *Acta Tropica* **85**, 51–61.

Torgerson, P. R., Burtisurnov, K. K., Shaikenov, B. S., Rysmukhambetova, A. T., Abdybekova, A. M. and Ussenbayev, A. E. (2003b). Modelling the transmission dynamics of *Echinococcus granulosus* in sheep and cattle in Kazakhstan. *Veterinary Parasitology* **114**, 143–153.

Torgerson, P. R., Shaikenov, B. S., Rysmukhambetova, A. T., Ussenbayev, A. E., Abdybekova, A. M. and Burtisurnov, K. K. (2003c). Modelling the transmission

dynamics of *Echinococcus granulosus* in dogs in rural Kazakhstan. *Parasitology* **126**, 417–424.

Torgerson, P. R., Schweiger, A., Deplazes, P., Pohar, M., Reichen, J., Ammann, R. W., Tarr, P. E., Halkik, N. and Mullhaupt, B. (2008). Alveolar echinococcosis: from a deadly disease to a well-controlled infection. Relative survival and economic analysis in Switzerland over the last 35 years. *Journal of Hepatology* **49**, 72–77.

Torgerson, P. R., Ziadinov, I., Aknazarov, D., Nurgaziev, R. and Deplazes, P. (2009a). Modelling the age variation of larval protoscoleces of *Echinococcus granulosus* in sheep. *International Journal for Parasitology* **39**, 1031–1035.

Torgerson, P. R., Rosenheim, K., Tanner, I., Ziadinov, I., Grimm, F., Brunner, M., Shaiken, S., Shaikenov, B. S., Rysmukhambetova, A. and Deplazes, P. (2009b). Echinococcosis, toxocarosis and toxoplasmosis screening in a rural community in eastern Kazakhstan. *Tropical Medicine & International Health* **14**, 341–348.

Torgerson, P. R., Keller, K., Magnotta, M. and Ragland, N. (2010). The global burden of alveolar echinococcosis. *PLoS Neglected Tropical Diseases* **4**, e722.

Torgerson, P. R., Raimkylov, K. M. and Kuttubaev, O. T. (2015). Epidemiological analysis of the distribution of cystic and alveolar echinococcosis in Osh Oblast in the Kyrgz republic, 2000-2013. *Journal of Helminthology (in press)*.

Trachsel, D., Deplazes, P. and Mathis, A. (2007). Identification of taeniid eggs in the faeces from carnivores based on multiplex PCR using targets in mitochondrial DNA. *Parasitology* **134**, 911–920.

Tsukada, H., Morishima, Y., Nonaka, N., Oku, Y. and Kamiya, M. (2000). Preliminary study of the role of red foxes in *Echinococcus multilocularis* transmission in the urban area of Sapporo, Japan. *Parasitology* **120**, 423–428.

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K. and Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications* **13**, 1790–1801.

Umhang, G., Woronoff-Rhen, N., Combes, B. and Boué, F. (2011). Segmental sedimentation and counting technique (SSCT): an adaptable method for qualitative diagnosis of *Echinococcus multilocularis* in fox intestines. *Experimental Parasitology* **128**, 57–60.

Usubalieva, J., Minbaeva, G., Ziadinov, I., Deplazes, P. and Torgerson, P. R. (2013). Human alveolar echinococcosis in Kyrgyzstan. *Emerging Infectious Diseases* **19**, 1095–1097.

Valadez, J. J., Weiss, W., Leburg, C. and Davis, R. (2002). Assessing Community Health Programs: A Participant's Manual and Workbook Using LQAS for Baseline Surveys and Regular Monitoring.

van der Giessen, J. W. B., Rombout, Y. B., Franchimont, J. H., Limper, L. P. and Homan, W. L. (1999). Detection of *Echinococcus multilocularis* in foxes in The Netherlands. *Veterinary Parasitology* **82**, 49–57.

van Kesteren, F. (2015). Canine echinococcosis in the Alay Valley, southern Kyrgyzstan (PhD thesis).

van Kesteren, F., Mastin, A., Mytynova, B., Ziadinov, I., Boufana, B., Torgerson, P. R., Rogan, M. T. and Craig, P. S. (2013). Dog ownership, dog behaviour and transmission of *Echinococcus* spp. in the Alay Valley, southern Kyrgyzstan. *Parasitology* **140**, 1674–1684.

van Kesteren, F., Qi, X., Jiang, T., Feng, X., Mastin, A., Craig, P. S., Vuitton, D. A., Xinyu, D., Chu, X., Jinlong, Z. and Hao, W. (2015). Independent evaluation of a canine echinococcosis control programme in Hobukesar County, Xinjiang, China. *Acta Tropica*.

Vaniscotte, A., Raoul, F., Poulle, M. L., Romig, T., Dinkel, A., Takahashi, K., Guislain, M. H., Moss, J., Tiaoying, L., Wang, Q., Qiu, J., Craig, P. S. and Giraudoux, P. (2011). Role of dog behaviour and environmental fecal contamination in transmission of *Echinococcus multilocularis* in Tibetan communities. *Parasitology* **138**, 1316–1329.

Varcasia, A., Canu, S., Kogkos, A., Pipia, A. P., Scala, A., Garippa, G. and Seimenis, A. (2007). Molecular characterization of *Echinococcus granulosus* in sheep and goats of Peloponnesus, Greece. *Parasitology Research* **101**, 1135–1139.

Veit, P., Bilger, B., Schad, V., Schäfer, J., Frank, W. and Lucius, R. (1995). Influence of environmental factors on the infectivity of *Echinococcus multilocularis* eggs. *Parasitology* **110**, 79–86.

Vervaeke, M., Davis, S., Leirs, H. and Verhagen, R. (2006). Implications of increased susceptibility to predation for managing the sylvatic cycle of *Echinococcus multilocularis*. *Parasitology* **132**, 893–901.

Viel, J. F., Giraudoux, P., Abrial, V. and Bresson-Hadni, S. (1999). Water vole (Arvicola terrestris scherman) density as risk factor for human alveolar echinococcosis. *American Journal of Tropical Medicine and Hygiene* **61**, 559–565.

Vyse, A. J., Gay, N. J., Hesketh, L. M., Pebody, R., Morgan-Capner, P. and Miller, E. (2006). Interpreting serological surveys using mixture models: the seroepidemiology of measles, mumps and rubella in England and Wales at the beginning of the 21st century. *Epidemiology and Infection* **134**, 1303–1312.

Wachira, T. M., Macpherson, C. N. L. and Gathuma, J. M. (1991). Release and survival of *Echinococcus* eggs in different environments in Turkana, and their possible impact on the incidence of hydatidosis in man and livestock. *Journal of Helminthology* **65**, 55–61.

Walters, T. M. H. (1978). Hydatid disease in Wales. 1. Epidemiology. *Veterinary Record* **102**, 257–259.

Wang, Q. (2011). Seasonal livestock slaughtering and the risk of *Echinococcus granulosus* infection in dogs in Tibetan communities, Sichuan, China. In *Proceedings of the 24th World Congress of Hydatidology*, Urumqi, China.

Wang, Y. H., Rogan, M. T., Vuitton, D. A., Wen, H., Bartholomot, B., Macpherson, C. N. L., Zou, P. F., Ding, Z. X., Zhou, H. X., Zhang, X. F., Luo, J., Xiong, H. B., Fu, Y., McVie, A., Giraudoux, P., Yang, W. G. and Craig, P. S. (2001). Cystic echinococcosis in semi-nomadic pastoral communities in north-west China. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **95**, 153–8.

Wang, Y., Zhang, X., Bartholomot, B., Liu, B., Luo, J., Li, T., Wen, X., Zheng, H., Zhou, H., Wen, H., Davaadorj, N., Gambolt, L., Mukhar, T., Al-Qaoud, K., Abdel-Hafez, S. K., Giraudoux, P., Vuitton, D. A., Fraser, A., Rogan, M. T. and Craig, P. S. (2003). Classification, follow-up and recurrence of hepatic cystic echinococcosis using ultrasound images. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **97**, 203–211.

Wang, Q., Qiu, J., Yang, W., Schantz, P. M., Raoul, F., Craig, P. S., Giraudoux, P. and Vuitton, D. A. (2006a). Socioeconomic and behaviour risk factors of human alveolar echinococcosis in Tibetan communities in Sichuan, People's Republic of China. *American Journal of Tropical Medicine and Hygiene* **74**, 856–862.

Wang, Q., Vuitton, D. A., Xiao, Y., Budke, C. M., Campos-Ponce, M., Schantz, P. M., Raoul, F., Yang, W., Craig, P. S. and Giraudoux, P. (2006b). Pasture types and *Echinococcus multilocularis*, Tibetan communities. *Emerging Infectious Diseases* **12**, 1008–1010.

Wang, Q., Yong-fu, X., Vuitton, D. A., Schantz, P. M., Raoul, F., Budke, C. M., Campos-Ponce, M., Craig, P. S. and Giraudoux, P. (2007). Impact of overgrazing on the transmission of *Echinococcus multilocularis* in Tibetan pastoral communities of Sichuan Province, China. *Chinese Medical Journal* **120**, 237–242.

Wang, Q., Raoul, F., Budke, C., Craig, P. S., Xiao, Y. F., Vuitton, D. A., Campos-Ponce, M., Qiu, D. C., Pleydell, D. R. J. and Giraudoux, P. (2010). Grass height and transmission ecology of *Echinococcus multilocularis*. *Chinese Medical Journal* **123**, 61–67.

Wang, Q., Budke, C. M., Liang, H., Giraudoux, P., Raoul, F., Vuitton, D. A., Craig, P. S. and Qiu, D. C. (2012). Spatial clustering of *E. multilocularis* infected dogs in pastoral Tibetan communities, Sichuan, China. *American Journal of Preventive Medicine* **28**, 323–328.

weatherbase.com (2015). Karakenja, Tajikistan weather averages (http://www.weatherbase.com/weather/weather.php3?s=605170).

Wei, J., Cheng, F., Qun, Q., Nurbek, Xu, S. D., Sun, L. F., Han, X. K., Muhan, Han, L. L., Irixiati, Jie, P., Zhang, K. J., Islayin and Chai, J. J. (2005). Epidemiological evaluations of the efficacy of slow-released praziquantel-medicated bars for dogs in the prevention and control of cystic echinococcosis in man and animals. *Parasitology International* **54**, 231–236.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B. and Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology* **75**, 1182–1189.

WHO (2009). Neglected zoonotic diseases (NZD).

WHO (2010a). *Working to overcome the global impact of neglected tropical diseases: First WHO report on neglected tropical diseases*.

WHO (2010b). *The 17 neglected tropical diseases*.

WHO (2011). *Report of the WHO informal working group on cystic and alveolar echinococcosis surveillance, prevention and control, with the participation of the Food and Agriculture Organization of the United Nations and the World Organization for Animal Health*. Geneva, Swizerland.

WHO (2013a). *Sustaining the drive to overcome the global impact of neglected tropical diseases: Second WHO report on neglected tropical diseases*.

WHO (2013b). Echinococcosis. *Fact Sheet 377*.

WHO Expert Committee (2002). Prevention and control of schistosomiasis and soil-transmitted helminthiasis. *World Health Organization Technical Report Series 912* 1– 57.

WHO Informal Working Group (2003). International classification of ultrasound images in cystic echinococcosis for application in clinical and field epidemiological settings. *Acta Tropica* **85**, 253–261.

WHO/OIE (2001a). Chapter 4: Geographic distribution and prevalence. In *WHO / OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern*, .

WHO/OIE (2001b). Chapter 1: Aetiology: parasites and life-cycles. In *WHO / OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern*, .

WHO/OIE (2001c). Chapter 5: Epidemiology. In *WHO / OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern*, .

WHO/OIE (2001d). Chapter 3: Echinococcosis in animals: clinical aspects, diagnosis and treatment. In *WHO / OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern*, .

WHO/OIE (2001e). Chapter 2: Echinococcosis in humans: clinical aspects, diagnosis and treatment. In *WHO / OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern*, .

WHO/OIE (2001f). Chapter 6: Control of echinococcosis. In *WHO / OIE Manual on Echinococcosis in Humans and Animals: A Public Health Problem of Global Concern*, .

Willeberg, P., Paisley, L. G. and Lind, P. (2011). Epidemiological models to support animal disease surveillance activities. *Revue Scientifique et Technique* **30**, 603–614.

Williams, C. B. (1964). *Patterns in the Balance of Nature*. Academic Press, London and New York.

Wilson, K., Grenfell, B. T. and Shaw, D. J. (1996). Analysis of aggregated parasite distributions: a comparison of methods. *Functional Ecology* **10**, 592–601.

Woolhouse, M. E. J. and Chandiwana, S. K. (1989). Spatial and temporal heterogeneity in the population dynamics of *Bulinus globosus* and *Biomphalaria pfeifferi* and in the epidemiology of their infection with schistosomes. *Parasitology* **98**, 21–34.

Woolhouse, M. E. J., Dye, C., Etard, J.-F., Smith, T., Charlwood, J. D., Garnett, G. P., Hagan, P., Hii, J. L. K., Ndhlovu, P. D., Quinnell, R. J., Watts, C. H., Chandiwana, S. K. and Anderson, R. M. (1997). Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 338–342.

World Bank (2005). Kyrgyz Republic: Country Economic Memorandum. Enhancing the Prospects for Growth and Trade. Volume I.

World Bank (2006). *Kyrgyz Republic, Livestock Sector Review: Embracing the New Challenges*.

World Bank (2010). Kyrgyz Republic - European Union (EU) Special Facility Support to Animal Health and Feeding Project (http://documents.worldbank.org/curated/en/2010/03/11973693/kyrgyz-republic-european-union-eu-special-facility-support-animal-health-feeding-project).

Wright, D. B. (1997). Extra-binomial variation in multilevel logistic models with sparse structures. *British Journal of Mathematical and Statistical Psychology* **50**, 21–29.

Wyn-Jones, G. and Clarkson, M. J. (1984). Radiologic detection of ovine hydatidosis. *Veterinary Radiology* **25**, 182–186.

Xiao, N., Nakao, M., Qiu, J., Budke, C. M., Giraudoux, P., Craig, P. S. and Ito, A. (2006a). Dual infection of animal hosts with different *Echinococcus* species in the Eastern Qinghai-Tibet plateau region of China. *American Journal of Tropical Medicine and Hygiene* **75**, 292–294.

Xiao, N., Qiu, J., Nakao, M., Li, T. Y., Yang, W., Chen, X., Schantz, P. M., Craig, P. S. and Ito, A. (2006b). *Echinococcus shiquicus*, a new species from the Qinghai-Tibet plateau region of China: discovery and epidemiological implications. *Parasitology International* **55**, S233–S236.

Yakob, L., Soares Magalhães, R. J., Gray, D. J., Milinovich, G., Wardrop, N., Dunning, R., Barendregt, J., Bieri, F., Williams, G. M. and Clements, A. C. A. (2014). Modelling parasite aggregation: disentangling statistical and ecological approaches. *International Journal for Parasitology* **44**, 339–342.

Yamamoto, N., Kishi, R., Katakura, Y. and Miyake, H. (2001). Risk factors for human alveolar echinococcosis: a case–control study in Hokkaido, Japan. *Annals of Tropical Medicine and Parasitology* **95**, 689–696.

Yang, Y. R., Sun, T., Li, Z. Z., Zhang, J. Z., Teng, J., Liu, X. Z., Liu, R. Q., Zhao, R., Jones, M. K., Wang, Y. H., Wen, H., Feng, X. H., Zhao, Q., Zhao, Y. M., Shi, D. Z., Bartholomot, B., Vuitton, D. A., Pleydell, D. R. J., Giraudoux, P., Ito, A., Danson, F. M., Boufana, B., Craig, P. S., Williams, G. M. and McManus, D. P. (2006). Community surveys and risk factor analysis of human alveolar and cystic echinococcosis in Ningxia Hui Autonomous Region, China. *Bulletin of the World Health Organization* **84**, 714–721.

Yang, Y. R., Clements, A. C. A., Gray, D. J., Atkinson, J.-A. M., Williams, G. M., Barnes, T. S. and McManus, D. P. (2012). Impact of anthropogenic and natural environmental changes on *Echinococcus* transmission in Ningxia Hui Autonomous Region, the People's Republic of China. *Parasites & Vectors* **5**, 146.

Yimam, A. E., Nonaka, N., Oku, Y. and Kamiya, M. (2002). Prevalence and intensity of *Echinococcus multilocularis* in red foxes (*Vulpes vulpes schrencki*) and raccoon

dogs (*Nyctereutes procyonoides albus*) in Otaru City, Hokkaido, Japan. *The Japanese Journal of Veterinary Research* **49**, 287–296.

Yong, W. K., Heath, D. D. and van Knapen, F. (1984). Comparison of cestode antigens in an enzyme-linked immunosorbent assay for the diagnosis of *Echinococcus granulosus*, *Taenia hydatigena* and *T. ovis* infections in sheep. *Research in Veterinary Science* **36**, 24–31.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer* **3**, 32–35.

Zhang, W. and McManus, D. P. (2008). Vaccination of dogs against *Echinococcus granulosus*: a means to control hydatid disease? *Trends in Parasitology* **24**, 419–424.

Zhang, W., Li, J. and McManus, D. P. (2003). Concepts in immunology and diagnosis of hydatid disease. *Clinical Microbiology Reviews* **16**, 18–36.

Zhang, W., Ross, A. G. and McManus, D. P. (2008). Mechanisms of immunity in hydatid disease: implications for vaccine development. *Journal of Immunology* **181**, 6679–6685.

Zhang, W., Zhang, Z., Yimit, T., Shi, B., Aili, H., Tulson, G., You, H., Li, J., Gray, D. J., McManus, D. P. and Wang, J. (2009). A pilot study for control of hyperendemic cystic hydatid disease in China. *PLoS Neglected Tropical Diseases* **3**, e534.

Zhang, W., Zhang, Z., Wu, W., Shi, B., Li, J., Zhou, X., Wen, H. and McManus, D. P. (2015). Epidemiology and control of echinococcosis in central Asia, with particular reference to the People's Republic of China. *Acta Tropica* **141**, 235–243.

Zhou, H. X., Chai, S. X., Craig, P. S., Delattre, P., Quere, J. P., Raoul, F., Vuitton, D. A., Wen, H. and Giraudoux, P. (2000). Epidemiology of alveolar echinococcosis in Xinjiang Uygur autonomous region, China: a preliminary analysis. *Annals of Tropical Medicine and Parasitology* **94**, 715–729.

Ziadinov, I., Mathis, A., Trachsel, D., Rysmukhambetova, A., Abdyjaparov, T. A., Kuttubaev, O. T., Deplazes, P. and Torgerson, P. R. (2008). Canine echinococcosis in Kyrgyzstan: using prevalence data adjusted for measurement error to develop transmission dynamics models. *International Journal for Parasitology* **38**, 1179–1190.

Ziadinov, I., Deplazes, P., Mathis, A., Mutunova, B., Abdykerimov, K., Nurgaziev, R. and Torgerson, P. R. (2010). Frequency distribution of *Echinococcus multilocularis* and other helminths of foxes in Kyrgyzstan. *Veterinary Parasitology* **171**, 286–292.

Zinck, R. D., Pascual, M. and Grimm, V. (2011). Understanding shifts in wildfire regimes as emergent threshold phenomena. *The American Naturalist* **178**, 149-161.

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A. and Smith, G. M. (2009). Zero-truncated and zero-inflated Models for Count Data. In *Mixed effects models and extensions in ecology with R*, pp. 261–293. Springer New York, New York, NY.

Zweig, M. H. and Campbell, G. (1993). Receiver-operating clinical medicine (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* **39**, 561–577.