



Probability of extinction for pathogens transmitted between ticks during co-feeding

A thesis submitted in fulfilment of the requirements for the degree of  
Bachelor of Science (Mathematics and Statistics) (Honours).

Jason Robert Thomas

Bachelor of Science (Mathematics and Statistics)

School of Science.

College of STEM.

RMIT University.

October 2025

## Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Signed: Jason Thomas

Date: 29th of October, 2025

## Acknowledgments

I thank:

- Associate Professor Stephen Davis for his weekly supervision, feedback and advice during this project. Much of the content is based directly on Professor Davis' ideas.
- Dr Kevin Bown for sharing the data used in this project.
- Dr Haydar Demirhan and Dr Simon Robertson-Johnstone for their infrequent but valuable advice.
- Daniel Longmuir, who helped me understand how to derive an important result.
- Sarah Cowen, Cooper Maher and Amy Lambalk for proofreading.

## Statement of contribution

All work presented in this thesis is my own, unless stated otherwise.

# Abstract

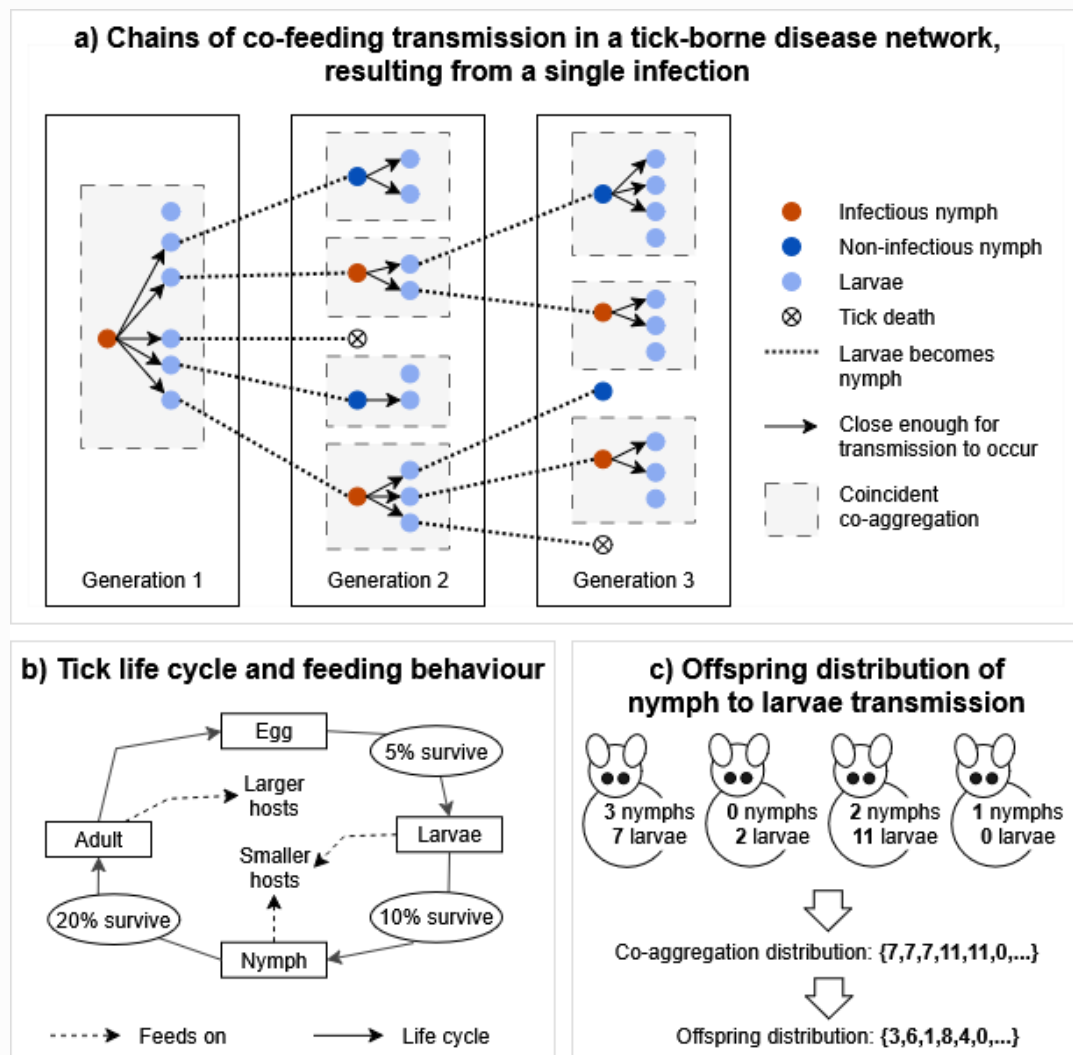
Tick-borne pathogens cause significant health issues worldwide. Ticks feed on vertebrate hosts, which allows pathogens to spread and to cause harmful diseases. Pathogens sometimes infect ticks via co-feeding transmission, where a feeding tick transmits a pathogen directly to another co-feeding tick.

This project makes use of a branching process technique to determine the probability that a chain of co-feeding transmissions becomes extinct. We present a technique, which we believe is novel, to use commonly-collected tick burden data to find an offspring distribution, which is required to perform a branching process analysis. We use numerical techniques and simulations to find ranges of extinction probabilities.

This project uses 2004-2005 field data from the Kielder Forest in England, for hard-bodied tick species *Ixodes ricinus* and *Ixodes trianguliceps*, and two species of vertebrate host. We show that in some years in the Kielder Forest some pathogens would have a chance of persisting via co-feeding transmission alone.

This project uses a novel combination of Lorenz curves and a recent result for estimating  $R_0$  when pathogens are spread by co-feeding transmission alone. For some combinations of tick and host species, just 1% of vertebrate hosts are responsible for all co-feeding transmission potential.

## Graphical abstract



**Figure 1: a) Branching process.** Co-feeding transmission occurs when susceptible ticks (usually larvae) attach and feed in close proximity to one or more infectious ticks (usually a nymph) and acquire a pathogen directly. This project introduces simplifying assumptions about the situation in nature: i) all larvae are susceptible; ii) larvae and nymphs might co-feed on one host, but they might not be close enough for transmission to occur; iii) there are no other routes of transmission; iv) two larvae that become infectious nymphs will not co-feed on the same vertebrate.

**b) Life cycle.** This life cycle of ticks is: immature ticks are more plentiful, and that they tend to feed on small vertebrates (Randolph, 1998). This implies that immature ticks are largely responsible for co-feeding transmission.

**c) Offspring distribution.** An offspring distribution is needed when using a branching process simulation. We use this simulation to calculate the probability that a chain of transmissions, starting with just a single infectious tick, goes extinct. We introduce a method to derive the offspring distribution from commonly-collected tick burden data.

## Summary

Ticks are parasitic arachnids that take blood meals from vertebrate hosts. That feeding behaviour is an opportunity for pathogens to spread to and from those hosts. Often pathogens will spread directly between closely feeding ticks.

This project implements what we believe is a novel technique to make use of common tick count data, combined with well-established numerical and stochastic techniques, to determine the probability that a chain of transmissions following a single infected tick becomes extinct. However, we restrict our analysis to transmission directly between ticks, which is known as co-feeding transmission. This project also uses a novel technique to measure the inequality in vertebrate hosts' potentials to act as superspreading hosts in co-feeding transmission, in the sense that some hosts attract a disproportionate number of ticks.

To make our analysis, we focus on data from the Kielder Forest in England, which concerns two species of ticks and two species of vertebrate hosts. This project shows that some combinations of tick, host and pathogen could allow for pathogens to persist in nature if several conditions are met. Future researchers could use the method presented here to approximate the probability of extinction for co-feeding transmissions.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Tick biology . . . . .	1
1.2	Tick-borne diseases . . . . .	2
1.3	Tick seasonal activity and distribution . . . . .	4
1.4	The definition and importance of $R_0$ . . . . .	4
1.5	Heterogeneity in disease networks . . . . .	5
1.6	Using the negative binomial distribution . . . . .	6
1.7	Thinking about tick-borne transmission using contact networks . . . . .	7
1.8	Galton-Watson Branching Processes . . . . .	8
<b>2</b>	<b>Summaries of Kielder Forest data</b>	<b>10</b>
2.1	Evidence of co-aggregation . . . . .	10
2.2	Tick burden heterogeneity . . . . .	13
<b>3</b>	<b>Reproducing the results by Lloyd-Smith et al, 2005</b>	<b>17</b>
<b>4</b>	<b>Reparameterisation of the negative binomial (NB) distribution</b>	<b>20</b>
<b>5</b>	<b>Determining the offspring distribution's parameters</b>	<b>22</b>
5.1	A process to find parameters $R_0$ and $k$ . . . . .	22
5.2	Avoiding simulation of the offspring distribution to find $R_0$ . . . . .	23
5.3	Goodness of fit tests for distributions fit to co-aggregation data $X$ . . . . .	25
5.4	Results of fitting the negative binomial (NB) distribution to co-aggregation data $X$ . . . . .	29
5.5	Estimates for $\alpha$ and its effect on $R_0$ . . . . .	31
<b>6</b>	<b>Estimates for the probability that a chain of transmissions becomes extinct</b>	<b>32</b>
<b>7</b>	<b>Discussion</b>	<b>36</b>
<b>8</b>	<b>Conclusion</b>	<b>38</b>
<b>9</b>	<b>References</b>	<b>39</b>
<b>10</b>	<b>Appendix</b>	<b>43</b>

# 1 Introduction

Worldwide, tick-borne diseases (TBD) pose a significant health burden for humans and livestock. When ticks take blood meals from vertebrate hosts, pathogens living inside the ticks have an opportunity to infect the hosts. Those pathogens can cause diseases, with sometimes severe consequences for the individual vertebrate hosts that ticks feed on (Johnson, 2023c). This feeding behaviour allows for pathogens to spread either from tick-to-vertebrate-to-tick, or sometimes directly from tick-to-tick.

We can characterise and understand an individual tick's expected behaviour by its life stage. All tick species have four life stages: egg, larva, nymph and adult. During a tick's juvenile life stages - larvae then nymph - it will take blood meals from vertebrate hosts so that it can moult into its next life stage. Female adults will take a blood meal in order to produce eggs, but it will no longer moult (Johnson, 2023b).

When most ticks of a particular life stage feed on a small number of vertebrate hosts, they are said to aggregate (Johnstone-Robertson et al., 2020). When ticks of different life stages aggregate on the same subset of hosts, then this is known as co-aggregation. When those ticks co-aggregate at the same time, then this is known as coincident co-aggregation. Field data frequently shows that aggregation disproportionately affects only a fraction of the vertebrate host population, and the 80:20 rule is often cited for many parasites (including ticks), where 80% of parasites can be found on just 20% of vertebrate hosts (Woolhouse et al., 1997).

## 1.1 Tick biology

Ticks are in the *Arachnida* class of animals, as are spiders and mites. Ticks are diverse and widespread; the diversity of ticks, and the vertebrate hosts that they feed on, is greater where the climate is more tropical, however, ticks are found on all continents except Antarctica (Johnson, 2023b). There are three families of ticks described in research: *Ixodidae* (Hard Ticks), *Argasidae* (Soft Ticks) and *Nuttalliellidae*. *Ixodidae* has the most known genera and about 700 species. *Nuttalliellidae* has just one known species (Nicholson et al., 2019).

A tick species' family is indicative of its behaviour. *Ixodidae* ticks take a single blood meal during each life stage, whereas *Argasidae* ticks take multiple blood meals during each life stage (Johnson, 2023a). Nidicolous ticks are said to spend their lives in burrows, caves and nests that their hosts live in. Many species are said to be specialists and will target few species of vertebrate hosts. Some nidicolous species of ticks are host-specific. Many species of *Ixodidae* ticks are said to be non-nidicolous and their typical habitats include forests, meadows and other clearings, grasslands, savannahs, and semi-desert or desert areas. Such species are often said to be generalists, meaning they take blood meals from a wide array of hosts (Nicholson et al., 2019). Most *Argasidae* are nidicolous (Vial, 2009), while only some *Ixodidae* are nidicolous.

This thesis will focus on species from the *Ixodidae* family of ticks, which pose a more significant health concern for humans (Parola & Raoult, 2001). Specifically, the analysis in this thesis will focus on *Ixodes ricinus* and *Ixodes trianguliceps*, for which we have data. *I. ricinus* is said to be non-nidicolous while *I. trianguliceps* is nidicolous (Nicholson et al., 2019).

The biology of *Ixodidae* make them particularly competent vectors for many diseases. When

Pathogen	Disease	<i>I. ricinus</i>	<i>I. trianguliceps</i>
<i>Anaplasma phagocytophilum</i>	Anaplasmosis	Confirmed	Potential
<i>Borrelia burgdorferi</i>	Lyme borreliosis	Confirmed	Potential
<i>Babesia microti</i>	Babesiosis	Confirmed	Confirmed

**Table 1:** Pathogens for which *I. ricinus* and *I. trianguliceps* are confirmed to be vectors or potentially vectors. Researchers have detected pathogens, or evidence of pathogens, in *I. trianguliceps* ticks but this is considered as insufficient evidence for vector competence (J. Gray et al., 2024; Rubel & Kahl, 2023)

they feed, they can stay attached for more than one week (Johnson, 2023d). Their extended period of feeding means that there is more time available for the transmission of pathogens (J. Gray et al., 2024). They lack digestive enzymes, meaning that their midgut is relatively benign, which increases the probability that a pathogen will survive inside the tick during moulting (this is defined later as transtadial transmission). When they feed, hard-bodied ticks inject saliva into their hosts, which increases the chance of successful transmission (J. Gray et al., 2024).

Usually the specialist Nidicolous ticks are not directly a threat to humans or livestock. Their sheltered lifestyle imposes limitations on their potential as vectors. The main role of nidicolous *Ixodidae* ticks, in their relation to zoonotic diseases, appears to be the maintenance of pathogens in nature (J. S. Gray et al., 2014). Non-nidicolous ticks, however, target a wider range of vertebrate hosts and therefore are more concerning from a public health policy perspective.

We can expect to find ticks of different life stages in differing frequencies. Ticks frequently die during and between their life stages, which implies larvae will be most abundant, nymphs less common, and adults less common still (Randolph, 1998). A tick’s life stage will also allow us to predict how it will feed. For *I. ricinus*, adults tend to feed on larger vertebrate hosts, whereas immature *I. ricinus* ticks will frequently feed on smaller vertebrate hosts (Herrmann & Gern, 2015; Randolph, 1998). In comparison, *I. trianguliceps* will feed almost exclusively on small mammal hosts, and is regarded as a specialist (K. J. Bown et al., 2003; K. Bown et al., 2008).

## 1.2 Tick-borne diseases

Ticks harbour pathogens (viruses, bacteria and protozoa) that spread to vertebrate hosts when they take blood meals from those hosts. Those pathogens can cause diseases and many have severe health implications for vertebrate hosts, including humans (Johnson, 2023c).

*I. ricinus* is the most important vector in central and western Europe; it is a vector for more than 20 pathogens (Sprong et al., 2018), and at least eight pathogens that can affect humans, including *Orthoflavivirus encephalitidis*, which is the causative agent of Tick-borne encephalitis (TBE). *I. trianguliceps* is a vector for at least 1 pathogen of human health concern. The pathogens for which these tick species are potentially the same vectors are summarised in Table 1.

Both species of tick are known to be vectors for *Babesia microti* (J. Gray et al., 2024; Rubel & Kahl, 2023). Pathogens transmitted by ticks are responsible for the majority of vector-borne disease in temperate North America, Europe and Asia, with Lyme Disease being the most prevalent tick-borne disease in the Northern Hemisphere (Rochlin & Toledo, 2020).



While the impact on human health is significant, so is the impact on the livestock industry. Research estimated that the estimated the global financial impact due to TBD was likely to be US\$22-30 billion per year (Lew-Tabor & Rodriguez Valle, 2016). A 2006 article reported that economic losses in Tanzania’s cattle industry were estimated to be \$364 million USD (Kivaria, 2006).

There are three ways that a pathogen can spread through populations of ticks (Harrison & Bennett, 2012):

- *Transovarial or vertical transmission:* Female ticks may transmit the pathogen to eggs.
- *Systemic transmission:* Ticks may feed on, and infect, a host leading to the host’s infection; other ticks may then acquire an infection by ingesting the infected host’s blood.
- *Co-feeding transmission:* Ticks may become infected by co-feeding alongside infected ticks; this does not require the host to have a systemic infection but instead pathogens are passed directly between ticks that feed in close proximity, and at the same time.

In addition, for a tick-borne pathogen to survive in nature, it is necessary for ticks to transmit the pathogens onwards. However, that is not sufficient for the pathogen’s survival. For a tick to become infected and then survive the moulting process to then pass on the disease, the pathogen first must survive transstadially; transstadial transmission is the tick’s maintenance of the pathogen between life stages (Johnson, 2023a). This is a feature of systemic, vertical and co-feeding transmission.

The importance of different transmission routes depends on the tick and host species. Co-feeding transmission is important in TBD systems where vertebrate hosts are incompetent vectors for maintaining a pathogen (Harrison & Bennett, 2012). Several studies suggest that co-feeding transmission is required for TBE to persist in nature (Harrison & Bennett, 2012; Hartemink et al., 2008), but it is less important for the persistence of Lyme disease. Several studies on TBD show that co-feeding transmission will increase  $R_0$ , which is defined later in this introduction (Johnstone-Robertson et al., 2020; Norman et al., 2004; Rosà et al., 2003). Some studies show that co-feeding transmission may allow a pathogen to persist even without systemic transmission (Norman et al., 2004; Rosà et al., 2003). The focus of this thesis is on such a scenario.

If an *Ixodidae* larva becomes infected during a blood meal (either via systemic or co-feeding transmission), then it becoming an infectious feeding nymph is not guaranteed. It must first survive the moulting process to become a nymph. It must then find another vertebrate host to feed upon. If the pathogen survives transstadially, then further transmission is possible. For the species *I. ricinus*, which is commonly called a deer tick or sheep tick, the larva-to-nymph moulting success rate was reported to be 91.9% under laboratory conditions (Hurry et al., 2021). Research from 2016 suggested that larvae-to-nymph moulting success of *I. ricinus* was dependent on environmental conditions, with evidence to suggest that snow cover would act as a temperature buffer during Germany’s winter (Dautel et al., 2016). For the species *I. trianguliceps*, mortality from larvae to nymph is density-dependent (Randolph, 1994). We note that, in trying to find the probability that an infected larva becomes an infectious feeding nymph, there is substantial variation.

### 1.3 Tick seasonal activity and distribution

Ticks vary in their lifespans and yearly host-seeking periods, which imposes a seasonal constraint on the potential for co-feeding transmission to be of importance; if nymphs and larvae feed at different times of the year on the same species of vertebrate hosts, then co-feeding transmission will not occur. *I. trianguliceps* has a lifespan of 2 to 5 years. All life stages seem to be active throughout the year but their activity peaks depending on where they live. In northern England, *I. trianguliceps* are most active in mid-autumn, while nymphal activity peaks in mid-summer, and infestations on rodents in the United Kingdom peak in summer and autumn (Pfäffle et al., 2017). *I. ricinus* is also active in northern England. Their life cycle typically lasts between 2 and 3 years. *I. ricinus* larvae activity frequently peaks in early summer while nymphal activity peaks in spring and autumn (Otranto et al., 2017). For both species there is at least some expected overlap in their activity.

The geographical ranges of *I. ricinus* and *I. trianguliceps* often overlap. *I. ricinus* has been reported in all parts of Europe, including Iceland, and throughout Northern Africa (Otranto et al., 2017). *I. trianguliceps* has been reported across mainland Europe and in the British Isles (Pfäffle et al., 2017). This thesis includes data from a single forest where *I. ricinus* and *I. trianguliceps* were collected from small mammals trapped during a longitudinal field study (K. J. Bown et al., 2011; K. Bown et al., 2008).

The habitat of *I. ricinus* is expanding in Europe, due to several factors. *I. ricinus* has increased its altitudinal range, based on studies in Bosnia and Herzegovina, the Czech Republic and Slovakia (Medlock et al., 2013). Modelling suggests that under different climate change scenarios, the habitat of *I. ricinus* will probably spread northwards in Northern Europe (Alkishe et al., 2017). Other modelling shows a similar northward spread of *I. ricinus*, although some southern regions of Europe are likely to become less suitable for *I. ricinus*, and some other species of ticks (Cunze et al., 2022). However, the spread of *I. ricinus* is not due to climate change alone; other drivers include the expansion of tick host populations, anthropogenic factors and land use changes (Medlock et al., 2013).

### 1.4 The definition and importance of $R_0$

The most important and widely-used concept in modelling transmissible diseases is the basic reproduction number  $R_0$ . This is a frequently-cited concept across epidemiology, and can be defined as the mean number of new infections per infection in an entirely susceptible population (Diekmann & Heesterbeek, 2000). It differs from the effective reproductive number  $R_t$ , which is used to quantify the efficacy of control measures in real time (Lim et al., 2020).

A consequence of using  $R_0$ , which is a notion of generational growth, is that it does not describe growth in real time. During the early stages of an outbreak in a susceptible population, we can understand growth as being exponential:

$$I(t) \approx Ce^{rt}$$

where  $I$  is the total number of infections up until time  $t$ ,  $C$  is some constant, and  $r$  is the real time growth rate (Diekmann & Heesterbeek, 2000).

$R_0$  is not a rate and is dimensionless, however, it has a relationship with the real-time growth rate,  $r$  (Diekmann & Heesterbeek, 2000):

$$\begin{aligned} R_0 < 1 &\iff r < 0 \\ R_0 > 1 &\iff r > 0. \end{aligned} \tag{1}$$

In other words, the outbreak will undergo exponential decay when  $R_0 < 1$ , but an outbreak can be expected to persist when  $R_0 > 1$ . An important nuance is that  $R_0 > 1$  is a necessary but insufficient condition for an outbreak to persist; due to stochasticity, the disease may still become extinct, even if  $R_0 > 1$ . This thesis will concentrate on  $R_0$ , rather than  $R_t$ , because we assume that populations of ticks are entirely susceptible when a single infected tick is introduced.

Determining  $R_0$  for wildlife disease systems often involves added complexities. Individual host species and individuals themselves have different susceptibilities, infectivity, and contacts. Given that an individual tick has different life stages, then these differences can exist within an individual throughout its life. Adding to the complexity is multiple transmission routes for many disease systems (Hartemink et al., 2008). Structured population models allow for modelling of these complexities (Diekmann & Heesterbeek, 2000). However, this thesis investigates the particular case where only co-feeding transmission is possible. Therefore,  $R_0$  in this thesis is simply the average number of new tick infections per infected tick. Finding  $R_0$  in previous research has involved using contact trace data (Lloyd-Smith et al., 2005), and this thesis introduces a method to simulate such data based on tick co-aggregation data.

## 1.5 Heterogeneity in disease networks

The extent to which individuals infected with transmissible diseases are able to infect their peers varies by multiple factors, including the specific pathogen, host population and environmental factors (Lloyd-Smith et al., 2005). In transmissible disease systems, heterogeneity in individual behaviour can mean that some infected individuals will not transmit the disease at all, while some might infect many of their peers.

An important notion to describe heterogeneity in disease systems is that of superspreading. Superspreading events are situations in which a few individuals are disproportionately responsible for infecting many others (Galvani & May, 2005), which is often the case with infections like gonorrhoea and HIV/AIDS. The individuals who transmit the disease in these situations are sometimes referred to as superspreaders. Important research in 2005 made the case that superspreading events are normal events in the spread of infectious diseases (Lloyd-Smith et al., 2005), and that work included a method to predict the frequency of superspreading events. Prior to 2005, there was already evidence of superspreading in disease networks. Data from the early-2000s SARS epidemic in Singapore showed that five individuals were identified as superspreaders after infecting another 103 individuals (CDC, 2003).

Superspreading is not just a feature of pathogens transmitted between humans; pathogens transmitted between other animals can exhibit the same behaviour, and there are multiple ways for superspreading to occur within those transmission networks. While it is true that some infected

individuals will come into contact with more of their peers, and are in a sense superspreaders, it is also true that some individuals have a higher infectiousness and are “super-shedders” (VanderWaal & Ezenwa, 2016). We can expect both of these types of heterogeneity (in contact and transmission) to be present in TBD systems.

For an individual tick to infect many others, a necessary but insufficient condition is co-incident co-aggregation (Ferreri et al., 2014), which we previously defined as the simultaneous aggregation of ticks from different life stages. Levels of aggregation are an indicator for, and pose a limit on, co-incident co-aggregation. To understand the heterogeneity in tick aggregation, the 80:20 rule is often cited. This is where 80% of cases are caused by 20% of infected individuals. Many vector-borne and sexually-transmitted diseases follow the 80:20 rule (Woolhouse et al., 1997). This heterogeneity is also present in TBD systems where there is evidence to suggest that tick aggregation data has strong heterogeneity (Brunner & Ostfeld, 2008; Devevey & Brisson, 2012; Ferreri et al., 2014; Shaw et al., 1998).

An interesting result in 2003 showed that when co-feeding ticks were considered, the transmission potential of the most-infested vertebrate hosts was even larger. Perkins et al. (2003) used Lorenz curves to quantify the heterogeneity in transmission potentials by ranking the most-tick infested vertebrate hosts. They found that 20% of vertebrate hosts were responsible for 74% of transmission potential, but that proportion jumped to 94% when only co-feeding ticks were considered (Perkins et al., 2003). This project takes inspiration from the work by Perkins et al. (2003) and will also use Lorenz curves to quantify transmission potential for co-feeding, but will use a different calculation.

The heterogeneity that is present in many transmission networks affects the choice of models we may choose to employ. Classically, models of disease outbreaks have assumed homogeneity in individual infectiousness and behaviour (Garske & Rhodes, 2008). These have typically been compartment models; the Susceptible-Infectious-Recovered systems of ordinary differential equations are common, although they can take many forms with more compartments defined. An assumption of those models is that the compartments are each large enough that the behaviour of individuals in those compartments is homogeneous. That is an appropriate assumption once an outbreak is established. However, at the start of an outbreak the number of infectious individuals must be small. This exposes a limitation: compartment models that lack a stochastic element are insufficient to describe the very early stages of an outbreak (Brauer, 2008a). Also in any population, the degree of infectiousness is distributed continuously, which further impedes our ability to divide a population into homogeneous subgroups (Lloyd-Smith et al., 2005). One approach is to use the SIR modelling framework to define stochastic differential equations (Allen, 2017), but there are other stochastic methods available. This thesis addresses the heterogeneity that occurs in TBD systems by using a branching process simulation, which is defined in 1.8.

## 1.6 Using the negative binomial distribution

The work of Lloyd-Smith et al. (2005) showed that the negative binomial (NB) distribution was a better fit than either the Poisson or geometric distributions to many contact tracing and surveillance datasets (Lloyd-Smith et al., 2005); this is an important result that this thesis will reproduce. While the NB binomial distribution is frequently presented in textbooks with an integer parameter, Lloyd-Smith et al. (2005) used a reparameterisation that allows for two continuous parameters:

sample mean  $m$  and dispersion parameter  $k$ , although this parameterisation was a known result (Rice, 2007). For a contact trace dataset, the sample mean is  $R_0$ .

The work by Lloyd-Smith et al. (2005) introduced the notion of an individual reproductive number  $v$ , which represents individuals' variation in disease history. They represented this as a gamma-distributed random variable and used that in a Poisson mixture to derive a NB distribution, as below:

$$\begin{array}{ll} \text{if} & v \sim \text{Gamma}\left(R_0, \frac{R_0}{k}\right) \\ \text{and} & Z \sim \text{Poisson}(v) \\ \text{then} & Z \sim \text{NB}(R_0, k). \end{array} \quad (2)$$

This is useful for several reasons. It allows us to query the effect of individual heterogeneity, and Lloyd-Smith et al. (2005) found systems with high individual variation had infrequent but explosive outbreaks. We can treat the fitted NB distribution as an offspring distribution, which is defined in 1.8. It allows for estimating the probability that a chain of transmission becomes extinct via numerical approximation.

The NB distribution is also important in modelling TBD systems. It has become the de facto choice for tick aggregation data, and for parasite aggregation data more generally. Shaw et al. (1998) showed that the NB distribution was an appropriate fit for 90% of aggregation datasets that they looked at (Shaw et al., 1998). Many articles concerning tick aggregation data have used the NB distribution (K. J. Bown et al., 2003; Brunner & Ostfeld, 2008; Harrison & Bennett, 2012). Other work suggests that when tick aggregation data has a heavy right tail, which indicates extreme heterogeneity in vectors on hosts, then the power law distribution is a better fit for some purposes (Bisanzio et al., 2010; Ferreri et al., 2014).

## 1.7 Thinking about tick-borne transmission using contact networks

Contact networks are a frequently-used tool in the analysis of how pathogens spread. Contact networks are graphs where vertices represent individuals and edges between them represent contacts. Such contacts are necessary for transmission i.e. the edges and nodes of a transmission network are contained within a contact network.

Network thinking can help overcome some of the problems associated with deterministic compartment models because network models allow for representing the heterogeneity that exists in individual behaviour, where superspreaders are represented as nodes with many outgoing edges. Network approaches can lead to useful outcomes, as some simulations suggest that stochastic methods lead to more accurate estimates for  $R_0$  when early outbreak data is used (Brauer, 2008b).

Bisanzio et al. (2010) modelled the potential spread of TBE and Lyme disease. Their analysis focused on co-aggregation data for *I. ricinus*. They made a bipartite contact network to represent the potential for transmission between vectors and hosts. Their analysis, which used bipartite networks to represent contacts between vectors and hosts, showed that extreme aggregation on hosts would greatly impact the epidemic threshold (Bisanzio et al., 2010). Their work showed

some evidence that a power law distribution is a better fit than the NB distribution for the right tail of aggregation on hosts. The work of Ferreri et al. (2016) followed the work of Bisanzio et al. (2010). Their model represented aggregation as star graphs, with vertebrates at the centre and ticks as peripheral nodes. Their analysis focused on co-feeding transmission, but the model did not differentiate between the different life stages of ticks (Ferreri et al., 2016).

The work of Johnstone-Robertson et al. (2020) characterised a tick-host contact network as a directed, acyclic, bipartite graph. There are nodes for ticks and vertebrates. The number of edges leading away from a vertebrate host node, denoted by  $k_{out}$ , represent the number of larvae that co-aggregate, and  $k_{in}$  represent the number of nymphs that co-aggregate. However, there is only one node type per tick. This implies that the tick nodes each represent a tick in its larval and then nymphal life stages. Johnstone-Robertson et al. (2020) then superimposed a transmission network that featured co-feeding transmission (edges between ticks) and systemic transmission (edges from tick-to-host, host-to-tick) (Johnstone-Robertson et al., 2020). This work is the main inspiration for the method, which is presented later in this thesis, of simulating outbreaks using co-aggregation data. An important result in Johnstone-Robertson et al. (2020) is that for co-feeding transmission:

$$R_0 \propto \frac{\langle k_{in}k_{out} \rangle}{\langle k_{in} \rangle} \quad (3)$$

where  $k_{in}$  is the number of nymphs and  $k_{out}$  is the number of larvae that are found on an individual vertebrate host. This result is also used to quantify the heterogeneity that exists in co-aggregation data.

## 1.8 Galton-Watson Branching Processes

A Galton-Watson Branching Process (GWBP) is a useful device for analysing how a disease spreads, but its original use was concerned with the extinction of surnames in Victorian Britain (Athreya & Ney, 1972). A surname that becomes extinct because each father has no sons to carry on the family name, and a disease outbreak that becomes extinct because infected individuals infect no one else, can be analysed with the same technique. The key to understand both situations is that each individual has the chance to produce offspring in a stochastic process.

A single-type GWBP branching process is such that all offspring, in all generations, belong to the same category; for a TBD, that could mean a specific species, particular life stage or infected by a particular pathogen. GWBP are also discrete-time Markov chains where the number of offspring in generation  $n$  depends on the probability of extinction in generation  $n - 1$ , but not on previous generations (Allen, 2019).

A requirement for using any branching process simulation is to define an offspring distribution. This is a discrete random variable that represents the number of new infections per infection (or sons per father, or new cells per cell, and so on). Let  $q_k$  be the offspring distribution, such that  $\sum_{k=0}^{\infty} q_k = 1$  (Diekmann & Heesterbeek, 2000), where  $k$  is the number of offspring. Note that if  $q_k$  is known, then we can determine  $R_0$  as:

$$R_0 = \sum_{k=1}^{\infty} kq_k \quad (4)$$

For a single-type GWBP, let  $z$  be the probability of extinction, and introduce  $g(z)$  as a probability generating function:

$$g(z) = \sum_{k=0}^{\infty} q_k z^k, |z| < 1 \quad (5)$$

Now, let  $z_n$  be the probability that the chain is extinct at or before  $n$  generations. Then substitute  $z_n = g(z_{n-1})$ .

The intuition for the  $z^k$  term is if there are  $k$  individuals then the probability  $z$  must be considered  $k$  times, since we require all individuals at that point to have no offspring.  $z$  is applied to individuals in the branching process in the sense that for extinction to occur, every infected individual must have 0 offspring.

$$z_n = g(z_{n-1}) = \sum_{k=0}^{\infty} q_k (z_{n-1})^k \quad (6)$$

Note this has the Markov property, since the probability of extinction for generation  $n$  only relies on generation  $n - 1$ .

This indicates  $z_n$  increases monotonically. But since all  $|z_n| \leq 1$ , then this series converges, which leads to the simplification  $z = g(z)$ . Thanks to that simplification, we find that the function is equal to its independent variable, and that means we can find an approximation for it via fixed point iteration. While that result is useful, we can avoid the calculation when extinction is guaranteed: since  $R_0$  is known (4), then we can use  $R_0 \leq 1 \implies z_{\infty} = 1$  (Diekmann & Heesterbeek, 2000).

The single-type branching process defined above is useful when offspring are the same type as their parent (Allen, 2019). However, many disease systems can be modelled using a multi-type branching process, which are capable of representing more complex systems that have more than just one pathogen, multiple species of host, and ticks in multiple different life-stages.

Multi-type GWBP simulations have been used to model TBD previously, which have effectively addressed the small number of infected individuals that limits the use of deterministic compartment models. In one article on TBD, authors compared predictions made using a deterministic compartment model and a multi-type GWBP to represent systemic transmission between hosts and vectors. They found the deterministic model would predict a stable endemic equilibrium if  $R_0 > 1$ , whereas the stochastic models would sometimes still lead to disease extinction even if  $R_0 > 1$  (Maliyoni et al., 2017).

In this thesis, we model transmission as infectious nymphs beget new infectious nymphs during co-feeding. Furthermore, this project assumes that the tick species are not competent vectors for the same pathogens and that the vertebrate hosts have different behaviours. We also note different larval and nymphal activity for both tick species varies by year. So, we use a single-type GWBP for each combination of tick species, host species and year.

Co-aggregation data (count data) Kielder Forest					
		<i>I. ricinus</i>		<i>I. trianguliceps</i>	
		Larvae	Nymphs	Larvae	Nymphs
2004	common shrew	451	14	1302	137
	field vole	150	13	432	4
2005	common shrew	314	0	512	79
	field vole	1458	24	764	78

**Table 2:** Overall counts of nymphs and larvae, with several vertebrate host species removed due to low counts of ticks.

## 2 Summaries of Kielder Forest data

This project makes use of data provided by Dr Kevin Bown, who has used the same data to author several papers (K. J. Bown et al., 2009, 2011; K. Bown et al., 2008). The data was collected in surveys in the Kielder Forest in the United Kingdom in 2004 and 2005. The provided dataset contains the numbers of larvae and nymphs found on each vertebrate, where vertebrates were included in the collection once only.

The data includes only the tick species *I. trianguliceps* and *I. ricinus*. The host species were mostly ‘field voles’, (*Microtus agrestis*), and ‘common shrews’ (*Sorex araneus*). The researchers also encountered small numbers of other small mammals. The common shrew and field voles vastly outnumber other vertebrate hosts in the Kielder Forest data. The use of that data in this project focuses on the common shrew and field vole observations.

### 2.1 Evidence of co-aggregation

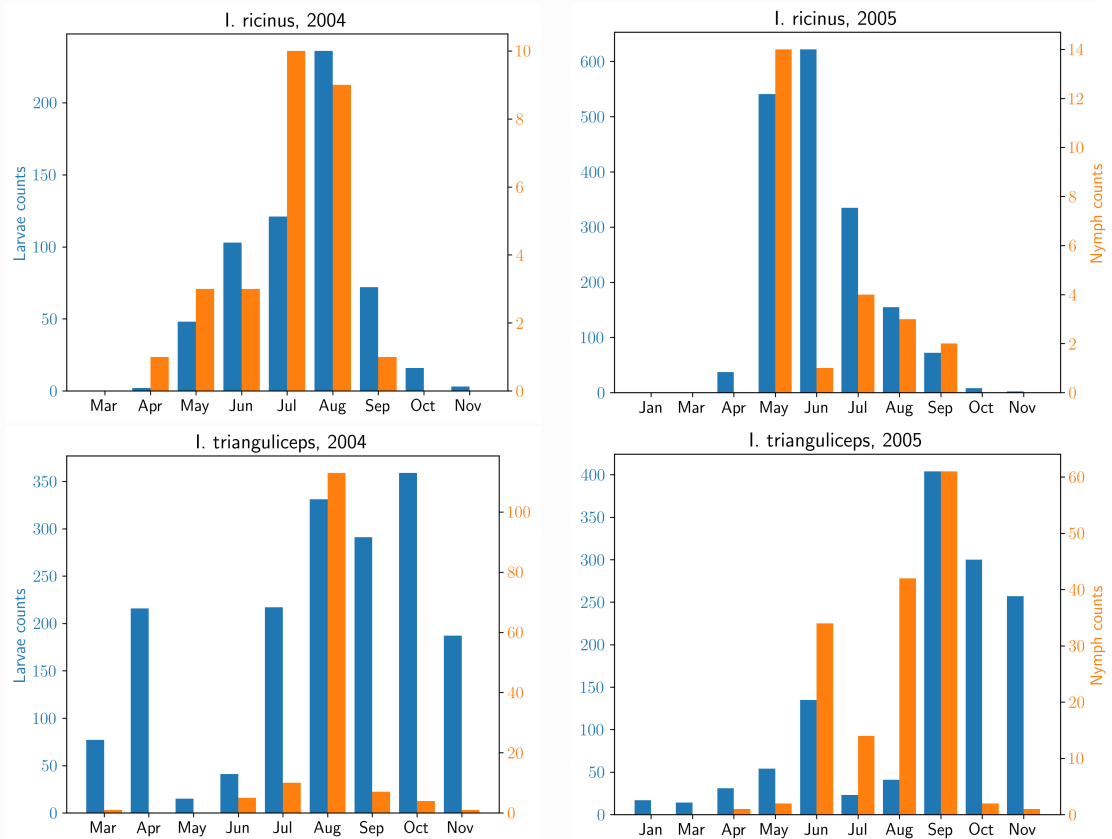
Of interest is whether larvae and nymphs of a particular tick species have the same seasonal variation in activity. If the larvae and nymphs are active at different times, then the opportunity for co-feeding transmission to occur is limited. Seasonal activity for nymphs and larvae, for *I. ricinus* and *I. trianguliceps*, collected at the Kielder Forest is presented in Figure 2. Overall counts of larval and nymphal burdens for both species of tick, and for the common shrew and field vole host species, are presented in Table 2.

This project makes the assumption that co-feeding is the only viable route of transmission, and that co-feeding transmission occurs from an infectious nymph to larvae that are in close proximity. Furthermore, if we assume *I. ricinus* and *I. trianguliceps* carry pathogens for which the other tick species are not susceptible, then we can refine the data to include only the larvae that co-aggregate with nymphs of the same species, as in Table 3. This means that larvae found on vertebrates without a nymphal burden of the same species were excluded.

Due to low tick burden counts, we exclude several combinations of year, host and tick species, but will investigate the following combinations using the Kielder Forest data:

- 2004: *I. ricinus* found on common shrews.
- 2004: *I. trianguliceps* found on common shrews.





**Figure 2:** Seasonal nymphal and larval activity for *I. trianguliceps* and *I. ricinus* for 2004 and 2005. The data here shows the combined frequency of tick counts found on field vole and common shrew individuals.

Co-aggregation data (count data) Kielder Forest					
		<i>I. ricinus</i>		<i>I. trianguliceps</i>	
		Larvae	Nymphs	Larvae	Nymphs
2004	common shrew	48	14	369	137
	field vole	7	13	1	4
2005	common shrew	0	0	179	79
	field vole	190	24	9	78

**Table 3:** Counts of nymphs and larvae, but with larvae removed if they do not co-feed with nymphs of the same species.

- 2005: *I. trianguliceps* found on common shrews.
- 2005: *I. ricinus* found on field voles.
- 2005: *I. trianguliceps* found on field voles.

There are several sensible reasons to separate these data into these 5 separate categories, rather than analysing all data together:

- *I. ricinus* is non-nidicolous while *I. trianguliceps* is nidicolous, indicating different behaviour (Nicholson et al., 2019).
- The pathogens transmitted by each tick species are frequently not the same pathogens.
- An adult common shrew is smaller than the adult field vole (Mammal Society, 2024a, 2024b), which implies the probability of nymphs and larvae co-feeding on a field vole, but in close enough proximity for co-feeding transmission to occur, is lower than for ticks that co-feed on common shrews.
- The two species of vertebrate host might have different rates of attracting ticks.
- Yearly seasonal variation shown in Figure 2 suggests that the different tick species have peak activities at different times of the year.

Central to this project is the co-incident co-aggregation of juvenile ticks on vertebrate hosts. To check the data reflected this, we used the Spearman Correlation Coefficient to test the existence of a monotonic relationship; generally, the vertebrate hosts that have more nymphs should also have more larvae. We used the SciPy package in Python to perform the test. We found weak evidence of correlation for some host and tick combinations. The results are presented in Table 4.

Spearman (ranked) Correlation Analysis Kielder Forest				
	<i>I. ricinus</i>		<i>I. trianguliceps</i>	
	statistic	p-value	statistic	p-value
common shrew	0.13401	~0	0.24473	~0
field vole	0.22757	~0	-0.03027	0.10164
common shrew and field vole	0.19450	~0	0.12317	~0

**Table 4:** The ranked correlations between nymphs and larvae, obtained by analysing the Kielder Forest data provided by K. Bown et al. (2008).

Assuming confidence of 0.95, then the p-value  $> 0.05$  of *I. trianguliceps* on field voles means we cannot reject the null hypothesis that there does not exist a monotonic relationship between the counts of larvae and nymphs for that combination of tick and host species. This pair of tick and vertebrate host species also has the lowest overall number of larvae, indicating low levels of parasitism. It is interesting that later in this project, we show co-feeding transmission between *I. trianguliceps* that co-feed on *field voles* will not allow for a pathogen's survival.

## 2.2 Tick burden heterogeneity

Some vertebrate hosts attract many more ticks than others, and measuring that heterogeneity has been addressed or estimated in many studies (Brunner & Ostfeld, 2008; Devevey & Brisson, 2012; Perkins et al., 2003; Woolhouse et al., 1997). Of particular interest is research from Perkins et al. (2003), which quantified inequality of hosts' co-feeding transmission potentials by plotting Lorenz curves, and then calculating a Gini coefficient to describe the overall inequality. This project will also use Lorenz curves and Gini coefficients to quantify the heterogeneity of tick burdens for particular populations of vertebrate hosts. While this thesis takes inspiration from the work of Perkins et al. (2003), we use a different calculation inspired by the work of Johnstone-Robertson et al. (2020).

A limitation of using the formula presented by Perkins et al. (2003) is that burdens of ticks in different life stages are grouped together in the term  $v_i^2$ . However, this project makes the simplifying assumption that co-feeding transmission is from infected nymph to nearby larvae only. Given that simplifying assumption, the Perkins calculation is unsuitable since it will not reflect that co-feeding transmission will not occur when either the larval or nymphal burdens for a particular vertebrate host are 0. Therefore, this project seeks another method to measure the inequalities in transmission potentials that vertebrate hosts have, due to heterogeneity in co-aggregation observed on those hosts. We use a modified version of the Johnstone-Robertson formula (3) since it addresses the above-mentioned complaint.

In Johnstone-Robertson et al. (2020), an important nuance is that  $k_{in}$  and  $k_{out}$  are respectively the nymphal and larval burdens of a host during its lifetime; to apply the calculation of  $R_0$  to co-feeding transmission, a term  $\nu_{ln}$  is needed to consider temporal and spatial requirements. But, if we use tick burden data collected in field surveys (as with the Kielder Forest data) then we can guarantee that the temporal requirement for co-feeding transmission to occur - that there must be co-incident co-aggregation - is satisfied.

So, let:

$b_n$  be the nymphal burden observed on an individual vertebrate host, which is analogous to  $k_{in}$ , except that  $b_n$  only applies at the time of data collection.

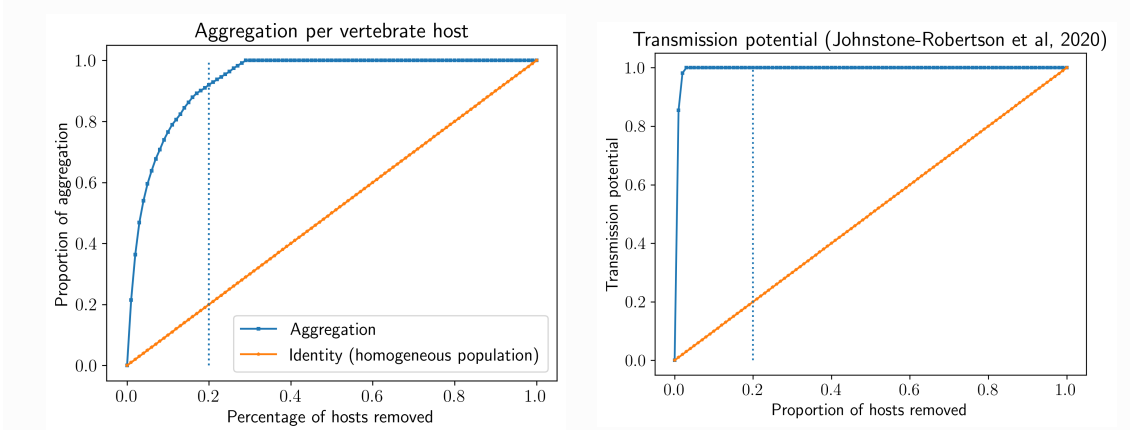
$b_l$  be the larval burden observed, which is analogous to  $k_{out}$ .

$m$  be the number of vertebrate hosts.

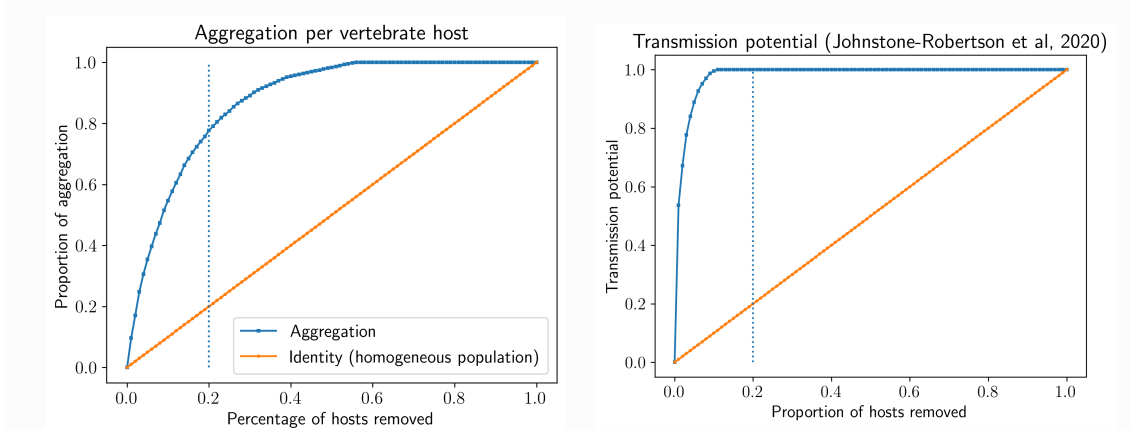
The modified version of (3) is now:

$$\begin{aligned} R_0 \propto \frac{\langle b_n b_l \rangle}{\langle b_l \rangle} &= \frac{(1/m) \sum_{a=1}^m (b_n b_l)_a}{(1/m) \sum_{a=1}^m (b_n)_a} \\ &= \frac{\sum_{a=1}^m (b_n b_l)_a}{\sum_{a=1}^m (b_l)_a} \\ &= \frac{(b_n b_l)_1}{\sum_{a=1}^m (b_l)_a} + \frac{(b_n b_l)_2}{\sum_{a=1}^m (b_l)_a} + \dots + \frac{(b_n b_l)_m}{\sum_{a=1}^m (b_l)_a} \end{aligned} \quad (7)$$

The goal is to obtain a cumulative estimate of vertebrate hosts' contributions to total transmission potentials, but without first having to find  $R_0$ , as did the example set by Perkins et al.



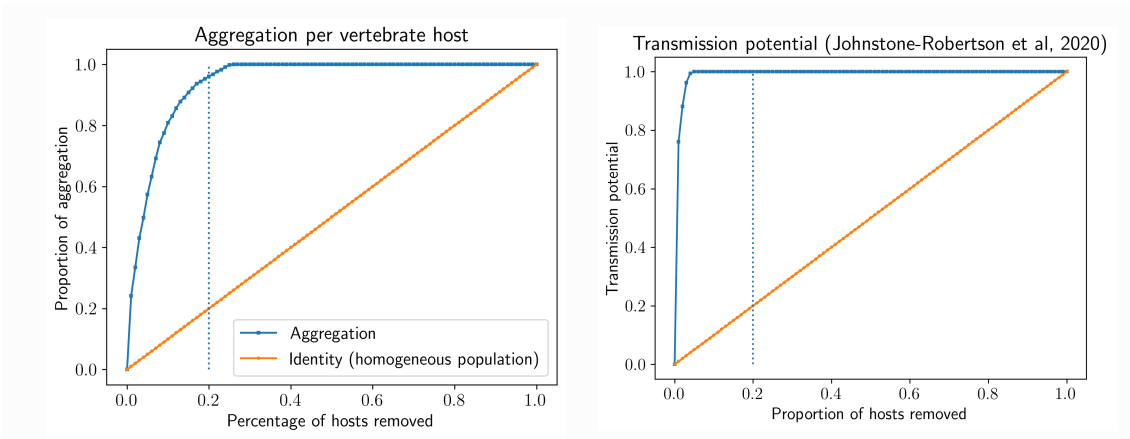
**Figure 3:** Aggregation (left) and cumulative transmission potential (right) of *common shrews* that were found with *I. ricinus* in 2004 in the Kielder Forest data.



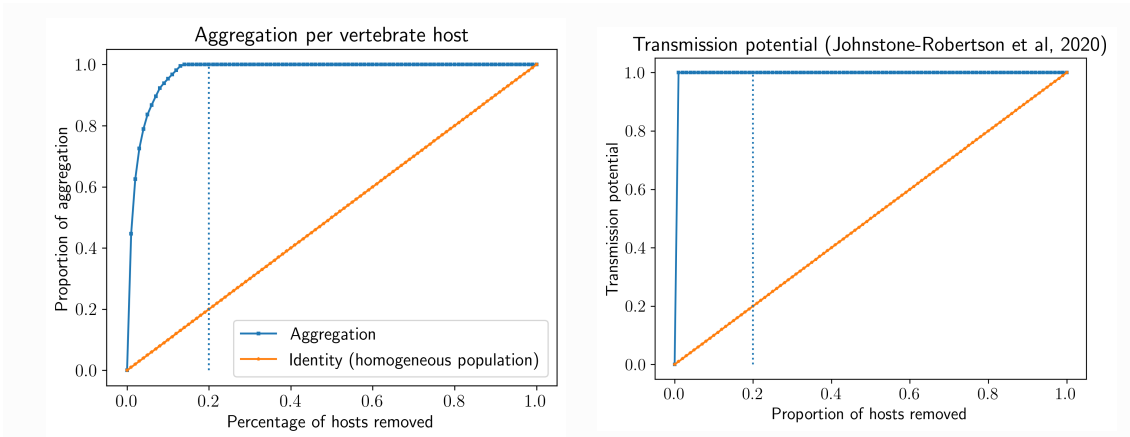
**Figure 4:** Aggregation (left) and cumulative transmission potential (right) of *common shrews* that were found with *I. trianguliceps* in 2004 in the Kielder Forest data.

(2003). We can order the vertebrates by the product  $b_n b_l$  in descending order, with  $a = 1$  denoting the vertebrate with the largest relative transmission potential. Then, we remove groups of vertebrates that have the most transmission potentials, following the Perkins et al. (2003) example. The cumulative effect of removing subsequently larger whole-number percentages of vertebrate hosts is shown in Figures 3, 4, 5, 6, 7. In those figures, the cumulative totals for tick aggregation (total numbers of nymph and larvae) are plotted next to the cumulative transmission potential calculated using (7). Each curve is found by removing subsequently larger proportions of the vertebrate host populations, starting with 1%, then 2%, and so on. The shape of each Lorenz curve is useful as a visual aide for understanding the inequality in transmission potentials. We also use those to calculate Gini coefficients, which are reported in Table 5.

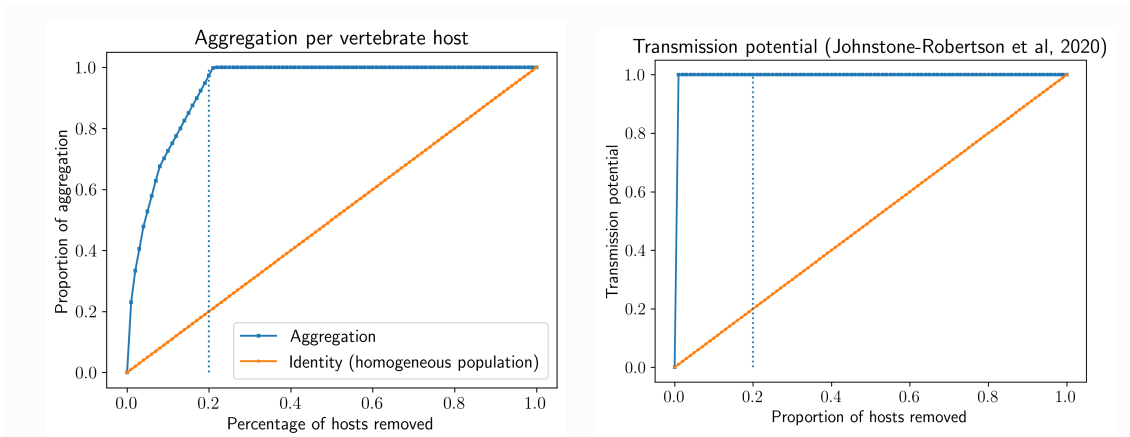
To find the Gini coefficient for each subset of data, we calculate the area between each Lorenz curve and the identity function. Since the identity function represents an idealised homogeneous population, where every vertebrate host has the same magnitude of tick immature burdens, then the Gini coefficient summarises the total amount of inequality present in the Kielder Forest data.



**Figure 5:** Aggregation (left) and cumulative transmission potential (right) of *common shrews* that were found with *I. trianguliceps* in 2005 in the Kielder Forest data.



**Figure 6:** Aggregation (left) and cumulative transmission potential (right) of *field voles* that were found with *I. ricinus* in 2005 in the Kielder Forest data.

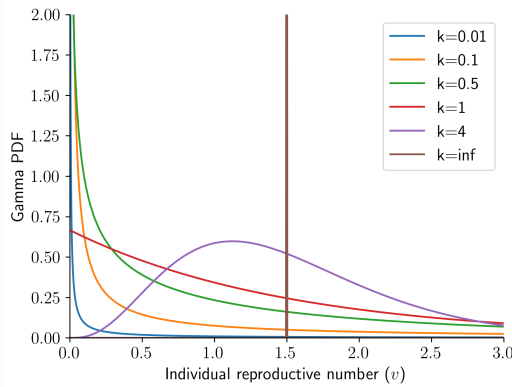


**Figure 7:** Aggregation (left) and cumulative transmission potential (right) of *field voles* that were found with *I. trianguliceps* in 2005 in the Kielder Forest data.

As expected, aggregation counts follow the 80:20 rule somewhat, which is commonly reported for tick aggregation and parasite aggregation more generally (Brunner & Ostfeld, 2008; Ferreri et al., 2014; Woolhouse et al., 1997). Also as expected, co-feeding transmission potential is even more heterogeneous, which matches the finding by Perkins et al. (2003). However, the heterogeneity for co-feeding transmission potentials in the Kielder Forest data is greater than the finding by Perkins et al. (2003): Figures 6 and 7 show that just 1% of vertebrate hosts were responsible for all transmission potential of field voles in 2005, for *I. ricinus* and *I. trianguliceps*.

Tick species	Host species	Year	Aggregation	Transmission potential
<i>I. trianguliceps</i>	<i>common shrew</i>	2004	0.747	0.963
<i>I. trianguliceps</i>	<i>common shrew</i>	2005	0.885	0.984
<i>I. trianguliceps</i>	<i>field vole</i>	2005	0.869	0.993
<i>I. ricinus</i>	<i>common shrew</i>	2004	0.871	0.989
<i>I. ricinus</i>	<i>field vole</i>	2005	0.95	0.993

**Table 5:** Gini coefficients for each of the five subsets of data, for total counts of immature tick aggregation (left) and for co-feeding transmission potentials (right), both derived from larval burdens collected in the Kielder Forest.



**Figure 8:** The individual reproductive number  $v \sim \text{Gamma}(R_0, \frac{R_0}{k})$ . This chart shows the effect of fixing  $R_0 = 1.5$  and allowing  $k$  to vary. Note that for  $k = \infty$ , we used  $k = 10^7$ .

### 3 Reproducing the results by Lloyd-Smith et al, 2005

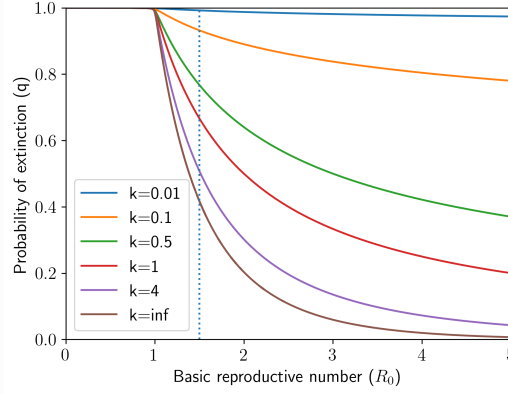
Before we follow the example set by Lloyd-Smith et al. (2005), it will be useful to re-implement parts of their work. Their work included using a stochastic simulation of pathogen transmissions, and numerical approximation to predict the probability that a chain of transmissions becomes extinct, for some pathogen and population where  $R_0$  and  $k$  could be determined. They also used many interesting visualisations, which we make use of later in this project.

Figures 8, 9 and 10 are reproduced versions of the charts shared by Lloyd-Smith et al. (2005). These illustrate the effect of the individual reproductive number  $v$ . A link to the code to generate these charts is shared in the appendix. To produce each chart, when  $k = \infty$ , we used  $10^b$  where  $b$  is the largest integer such that  $10^b$  did not produce numerical errors.

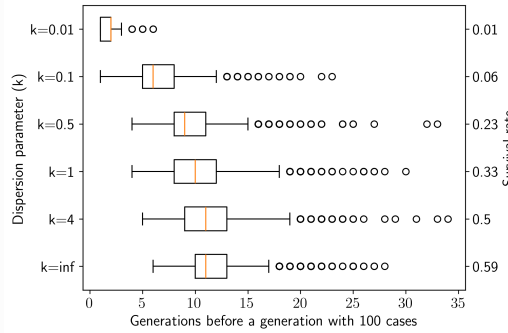
Figure 10 highlights the effect of superspreaders. We first use a stochastic simulation that randomly draws the individual reproductive number  $v \sim \text{Gamma}(R_0, \frac{R_0}{k})$ . Numbers of offspring are then randomly drawn from the Poisson Mixture:  $Z \sim \text{Poisson}(v)$ . That means there are two sources of randomness in the simulations presented in Figure 10. We record the number of generations required before a generation with 100 offspring is found. We also record the rate of outbreaks that survive 10,000 generations. A low value for  $k$  implies a heterogeneous population where few individuals infect many others. Inversely,  $k = \infty$  implies a homogeneous population where each individual infects (close to)  $R_0$  others. For low values of  $k$ , the related low survival rate (along the second vertical axis) combined with the smaller average and smaller spread in the number of generations before a generation with 100 offspring is found, indicates that when  $k$  is small an outbreak either explodes in numbers quickly or not at all.

By comparing Figures 9 and 10, we can see that when  $R_0 = 1.5$ , then the extinction probability  $q$  is roughly the complement of the survival rate. That is not surprising, but it is interesting and encouraging to see.

Lloyd-Smith et al. (2005) noted that if  $v \sim \text{Gamma}$  and if there is a Poisson mixture  $Z \sim \text{Poisson}(v)$ , then that Poisson mixture simplifies to a NB distribution. If  $Z$  is also the offspring distribution of a GWBP, then we can use the probability generating function of that NB distribution to approximate the probability that a chain of transmissions becomes extinct. By first



**Figure 9:** If  $Z \sim \text{Poisson}(v)$ , then the resulting Poisson Mixture is a NB distribution, which is a candidate offspring distribution. The probability that a chain of transmissions becomes extinct for a combination of  $R_0$  and  $k$ , after the introduction of a single infected individual, is found by fixed point iteration over the NB distribution's probability generating function. Each simulation generates a random value for  $v \sim \text{Gamma}(R_0, \frac{R_0}{k})$ . Note that for  $k = \infty$ , we used  $k = 10^{12}$ . The vertical dotted line is  $R_0 = 1.5$  that we use in Figure 10.



**Figure 10:** Lloyd-Smith et al. (2005) used a simulation, with the first generation to reach 100 offspring used as an arbitrary benchmark. Boxes represent the interquartile ranges for how many generations were required before a single generation contained 100 offspring, given different values of  $k$  and a fixed  $R_0 = 1.5$ . Note that for  $k = \infty$ , we used  $k = 10^7$ . The values along the right vertical axis are the proportion of outbreaks that survived 10,000 generations.



reparameterising the NB distribution, we can express the probability generating function in terms of  $R_0$  and  $k$ . The derivation for that particular probability generating function is presented in (11). By using the more general result of convergence presented in (6), we can approximate the probability of extinction by fixed point iteration, which is the process that Figure 9 shows for different combinations of  $R_0$  and  $k$ .

In this project we use count data of larvae per nymph. Before we find the offspring distribution's parameters  $R_0$  and  $k$ , for use in a branching process analysis, we introduce a method to derive a transmission network from the larval count data. For our purposes,  $v$  is already represented by the larval count data. We also test whether the NB distribution is a good fit to the data.

## 4 Reparameterisation of the negative binomial (NB) distribution

While the work by Lloyd-Smith et al. (2005), presented above, uses the NB distribution, it remains to be seen that it is the best fit for our tick larval count data. To test it, we first need to derive the same parameterisation that Lloyd-Smith et al. (2005) used. Then, it will be a candidate for the offspring distribution. Other candidate discrete distributions are the Poisson and geometric distributions; later in this project, we fit all distributions to the same data and then compare goodness-of-fit for each.

Commonly, the NB distribution is presented as:

$$Z \sim NB(x|p, k) = \binom{x+k-1}{x} p^k q^x, x \in \mathbb{N}. \quad (8)$$

But, a more convenient parameterisation uses the sample mean  $m$  and dispersion parameter  $k$  as parameters (Rice, 2007). This is possible with  $p = \frac{k}{m+k}$ :

$$Z \sim NB(x|m, k) = \frac{\Gamma(x+k)}{x!\Gamma(k)} \left(1 + \frac{m}{k}\right)^{-k} \left(\frac{m}{m+k}\right)^x, x \in \mathbb{N}. \quad (9)$$

Before fitting the distribution to some data the sample mean  $m$  can already be found from the data itself. Finding the dispersion parameter  $k$  is achievable with maximum likelihood estimation (MLE). Since the parameters  $m$  and  $k$  have asymptotically orthogonal limits (Lloyd-Smith et al., 2005), then we can find  $k$  via MLE without doing so for  $m$ , which is useful since  $m$  is known.

Lloyd-Smith et al. (2005) fit the reparameterised NB distribution to contact trace data, where the sample mean for each dataset is  $m = R_0$ . An important nuance is that to find  $R_0$  from counts of larvae per nymph, since we do not have contact trace data, another step is required to account for stochasticity in transmission. In this thesis  $m$  is the sample mean of larvae per nymph counts and  $m \neq R_0$ , but we derive an expression for finding  $R_0$  in a later section.

In this project, we tested the standard negative binomial distribution, available in Python's Scipy package, against the reparameterisation (9) by fitting both distributions to the same data. Results indicate a similar fit. A link to the code is available in the appendix.

We next derive the probability generating function for the negative binomial distribution (8), and then perform a change of variables, as in (9). Note the definition of a probability generating function given in (5) is employed below.

$$\begin{aligned}
g(z) &= \sum_{x=0}^{\infty} \binom{x+k-1}{x} p^k q^x z^x \\
&= p^k \sum_{x=0}^{\infty} \binom{x+k-1}{x} (qz)^x \\
&= p^k (1 - qz)^{-k}
\end{aligned} \tag{10}$$

$$\begin{aligned}
&= \left( \frac{k}{m+k} \right)^k \left( 1 - \frac{m}{m+k} z \right)^{-k} \\
&= \left( 1 + \frac{m}{k} (1 - z) \right)^{-k}
\end{aligned} \tag{11}$$

Above, (10) is a result of the negative binomial theorem. The equation on line (11) is the final form of the probability generating function, for use with numerical approximation later in this project.

## 5 Determining the offspring distribution's parameters

Given the scenario where a single infected tick is introduced to a population of vertebrate hosts for the first time, there is no guarantee that the tick will infect any other tick. Even if it did, those infected ticks might not infect any others for a number of reasons. There is no reason to think that all ticks will share the same individual reproductive number  $v$ . Deterministic compartment models are not appropriate for estimating the probability that a chain of transmissions in a TBD network becomes extinct in the first stages of an outbreak due to their underlying assumption of homogeneous populations in each compartment. Since the infected compartment in such a model must be small at the start of an outbreak, then that compartment can not be said to be homogeneous in any meaningful way.

To calculate the probability that a chain of TBD transmissions becomes extinct, we can use a single-type GWBP analysis similar to the example set by Lloyd-Smith et al. (2005). To do that, however, we first require a candidate for the offspring distribution.

### 5.1 A process to find parameters $R_0$ and $k$

Lloyd-Smith et al. (2005) estimated  $R_0$  and then found an estimate for  $k$  by using contact-tracing network data. In our case, we do not have contact-trace data. We instead have empirical data for the co-aggregation of larvae and nymphs on individual vertebrate hosts and co-aggregation does not mean that each pair of larva and nymph was a transmission. Instead, we can use a stochastic process to simulate a transmission network based on the empirical data in the sense that each pair of co-feeding larva and nymph is a potential transmission.

Let us note some characteristics of pathogen transmission where *Ixodidae* tick species are the vectors:

- Since an *Ixodidae* larva will take one blood meal before it moults into a nymph, then the only way for an that larva to be infectious when it takes its blood meal would be if it was already infected via transovarial transmission.
- Given that ticks die during and between their life stages, then larvae will be relatively common, nymphs less common, and adults less common still (Randolph, 1998).
- Adult ticks tend to feed on larger vertebrates, whereas immature ticks (larvae and nymphs) frequently feed on small vertebrate hosts (Herrmann & Gern, 2015; Randolph, 1998).
- Since immature ticks prefer to feed on smaller vertebrate hosts, then they are more likely to feed in close proximity.
- Nymph-to-nymph co-feeding transmission is rare (Voordouw, 2014).

So, if a TBD was to spread through a population of ticks via co-feeding transmission, in the absence of transovarial and systemic transmission, then the large majority of infections would pass from nymphs to larvae. Each larva would need to moult into a nymph and then successfully find another host before it could then pass the pathogen onto larvae that it co-aggregates with. During moulting the pathogen would need to survive transstadially as each infected larva becomes an

infectious nymph. The ticks that co-aggregate would need to feed close enough for transmission to occur.

Let us now introduce the following parameters:

$c$  : the contact rate, or the probability that a larva and infectious nymph co-feed close enough for transmission to occur.

$v$  : the probability of transmission, since individual variation means not every infectious nymph has the same capability to transmit the pathogen.

$\sigma$  : the probability that a larvae survives to become a nymph (moulting success).

$\tau$  : the probability that transstadial transmission occurs.

Then for each larvae that co-feeds with an infectious nymph, the probability that it becomes an infectious nymph is found by:

$$\alpha = cv\sigma\tau. \quad (12)$$

Hence,  $\alpha$  will depend on the species of tick, the pathogen being transmitted, the vertebrate host ticks feed on, seasonal variation and perhaps other factors.

The network thinking by Johnstone-Robertson et al. (2020) is important for this project. Their superimposition of tick-to-tick transmission onto the tick-host contact network is the inspiration for the following idea: we can use count data of larvae per nymph to generate a nymph-larvae contact network for each nymph. The definitions below are placed into context in Figure 11, which serves as a visual aide in understanding how we intend to find estimates for  $R_0$  and  $k$ .

We introduce the following variables:

$X$  : a vector of the out degree for each nymph's larval contact network.

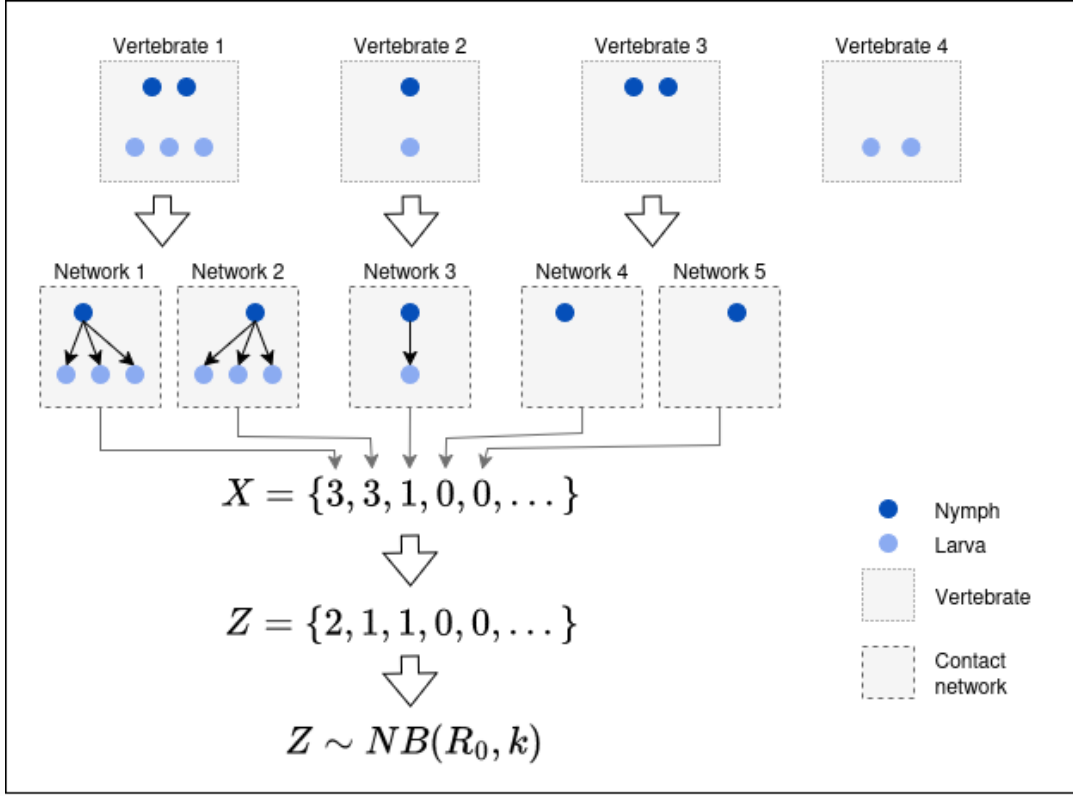
$n$  : the count of nymphs.

$Z$  : a vector for how many larvae become infectious nymphs.

The vector  $Z$  offspring data, to which we will fit an offspring distribution. Note that  $X$  and  $Z$  have the same number of elements  $n$  (count of nymphs) and  $X_i \geq Z_i$  for all  $i \dots n$ . This section has talked about using the NB distribution for the offspring distribution. Several discrete distributions could be the best choice: geometric, Poisson and NB distributions could all fit. The NB distribution is often a good fit to parasite aggregation data, and, Lloyd-Smith et al. (2005) showed this was the best fit for several disease transmission networks; we can expect that the NB distribution will be the best fit. Later, we test the NB distribution's goodness-of-fit against the Poisson and geometric distributions.

## 5.2 Avoiding simulation of the offspring distribution to find $R_0$

Conceptually, to find an offspring distribution we could i) fit a discrete distribution to  $X$  ii) draw random values from  $X$ , iii) for each value drawn, apply a Bernoulli trial of  $\alpha$  to find  $Z$ , iv) fit



**Figure 11:** A toy example of how we find the parameters for the offspring distribution. Each nymph has its own contact network with larvae. The out degree for each nymph is the number of larvae that it co-feeds with. A stochastic step is applied to the vector  $X$  to determine the number of larvae that become infectious nymphs, which is a vector  $Z$ .  $R_0$  and  $k$  are then found by fitting the reparameterised NB distribution to  $Z$  as the offspring distribution.

a distribution to  $Z$  to determine  $R_0$  and  $k$ . This section shows that  $k$  can be found without the stochastic step to generate  $Z$ .

Here we place existing variables into greater context:

$m = \frac{1}{n} \sum_i^n L_i$  : is the mean of  $X$ .

$X \sim \mathbf{NB}(m, k)$  : is the co-aggregation distribution, fit to  $X$ . This is the reparameterised NB distribution (9).  $X_i$  is the number of larvae that are found to co-feed with a nymph of index  $i$ .

$I_{i,j} \sim \mathbf{Bern}(p = \alpha)$  : is a Bernoulli trial for each nymph of index  $i$  and for each larvae that it co-aggregates with, indexed by  $j \in 1 : X_i$ . This is a step where an infectious nymph generates a new infectious nymph in a branching process.

$R_0$  : is the basic reproductive number. This is also the mean value of the offspring data  $Z$ .

$Z \sim \mathbf{NB}(R_0, k)$  : is the offspring distribution. This can be achieved by simulating offspring data  $Z$  using  $X, \alpha$  and then fitting the distribution to it.

To get the expected number of new infectious nymphs per current infectious nymph, dependent on the number of larvae that each infectious nymph co-feeds with, the number of new infections is a series of Bernoulli trials, where  $X_i$  is the number of trials. This is otherwise known as a binomial experiment:

$$I_i = \sum_j^{X_i} I_{i,j}$$

$$I_i \sim \text{Binom}(X_i, \alpha) \quad (13)$$

The expected value of the binomial distribution in (13) is  $\mu_i = X_i \alpha$ , which is the expected number of new infections for the nymph of index  $i$ .

Then to determine  $R_0$ , which is the expected number of new infections in an offspring distribution, this is:

$$R_0 = \frac{1}{n} \sum_i^n \mu_i$$

$$R_0 = \frac{1}{n} \sum_i^n L_i \alpha$$

$$R_0 = \alpha \frac{1}{n} \sum_i^n L_i$$

$$R_0 = \alpha m. \quad (14)$$

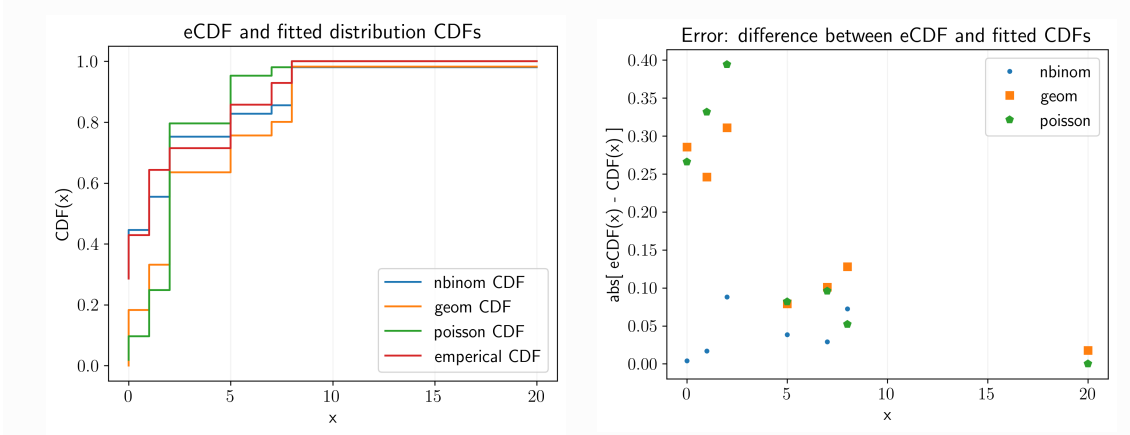
Since the reparameterised NB distribution has parameters with asymptotically orthogonal limits, then  $k$  will not vary between the distributions fitted to  $X$  and  $Z$ . This implies that we can obtain the parameters of  $Z$ ,  $R_0$  and  $k$ , directly from the distribution fitted to  $X$ .

### 5.3 Goodness of fit tests for distributions fit to co-aggregation data $X$

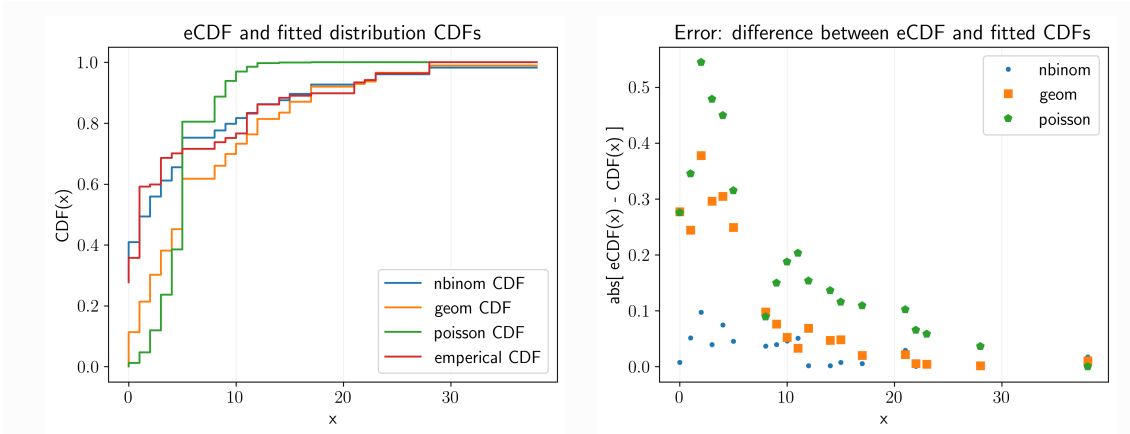
Before using the reparameterised NB distribution (9) for determining the probability of extinction with a single-type GWBP, we can check which distribution fits the co-aggregation data  $X$  best. Candidate distributions are the negative binomial, Poisson and geometric distributions. We compared the fitted distributions visually, and then by using the Akaike Information Criterion (AIC) (Lloyd-Smith et al., 2005).

For the visual analysis of each subset of Kielder Forest data, we calculated the error of each fitted distribution by finding the difference between its cumulative density function (CDF) and the data's empirical cumulative density function (eCDF) for the independent variable  $x$  of each function. These are presented as Figures: 12, 13, 14, 15 and 16.

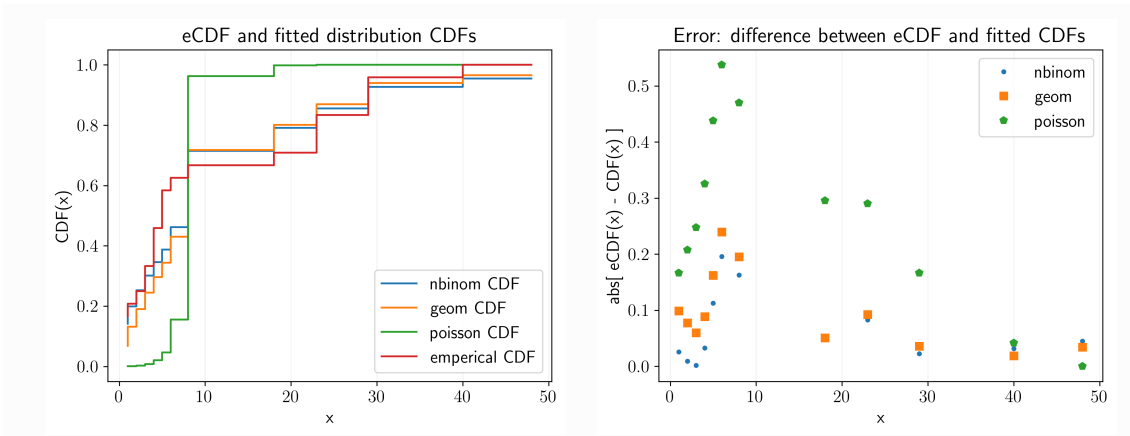
To calculate goodness of fit, for each distribution and for each subset of data, we used  $AIC_c$ , which is  $AIC$  corrected for small datasets.  $AIC_c$  rewards the best-fitting models while penalising models by their number of parameters. In the sense that the NB, Poisson and geometric distribu-



**Figure 12:** Discrete probability distributions fitted to the co-aggregation data  $X$  for *I. ricinus* found on *common shrews* in 2004.

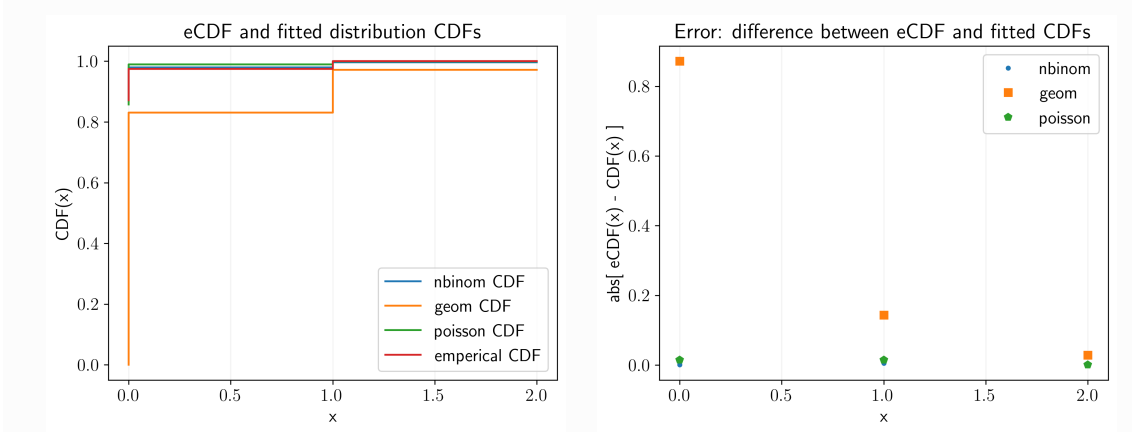


**Figure 13:** Discrete probability distributions fitted to the co-aggregation data  $X$  for *I. trianguliceps* found on *common shrews* in 2004.

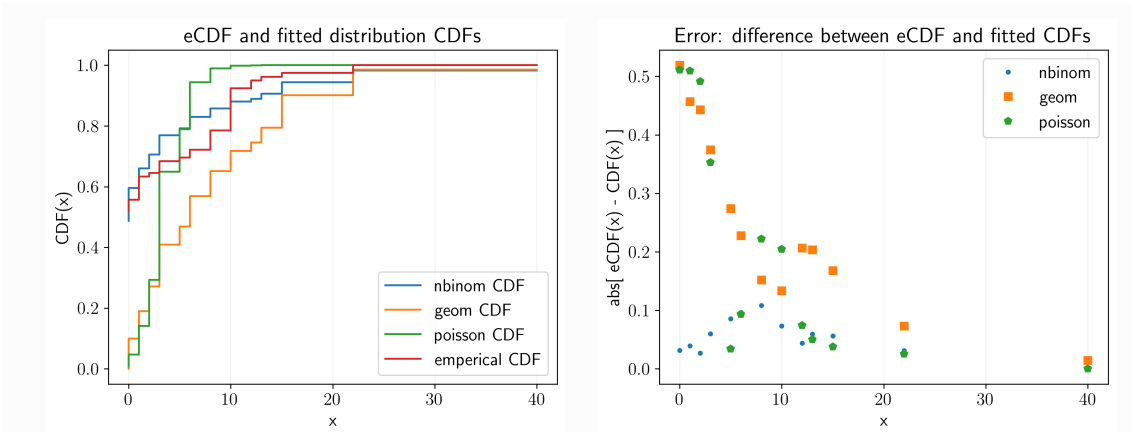


**Figure 14:** Discrete probability distributions fitted to the co-aggregation data  $X$  for *I. ricinus* found on *field voles* in 2005.





**Figure 15:** Discrete probability distributions fitted to the co-aggregation data  $X$  for *I. trianguliceps* found on *field voles* in 2005.



**Figure 16:** Discrete probability distributions fitted to the co-aggregation data  $X$  for *I. trianguliceps* found on *common shrews* in 2005.

tions are models, the NB distribution will be penalised for its extra parameter, while the Poisson and geometric distributions will each be rewarded for having just one parameter.

Each distributions' fit to co-aggregation data were calculated as:

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

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1}. \quad (15)$$

For the Kielder Forest data, these  $AIC_c$  scores are presented in Table 6. A point of difference is that scaled their  $AIC_c$  by subtracting the smallest value in each dataset, but the smallest value in all tick burden datasets used in this project is always 0.

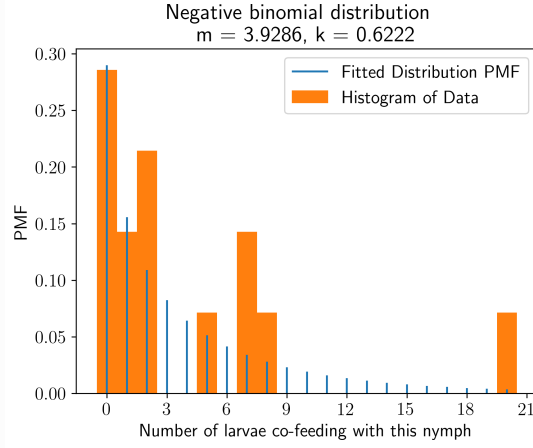
Tick species	Host species	Year	AICc		
			Negative binomial	Geometric	Poisson
<i>I. trianguliceps</i>	<i>common shrew</i>	2004	773.634	$\infty$	1811.540
<i>I. trianguliceps</i>	<i>common shrew</i>	2005	370.802	$\infty$	939.5
<i>I. trianguliceps</i>	<i>field vole</i>	2005	74.768	$\infty$	73.748
<i>I. ricinus</i>	<i>common shrew</i>	2004	73.725	$\infty$	115.538
<i>I. ricinus</i>	<i>field vole</i>	2005	182.205	176.696	471.78

**Table 6:** AICc (corrected AIC), fitted to subsets of the Kielder Forest data. The NB distribution is the best fit for 3 subsets of data, while it is nearly the best fit in 2 other cases.

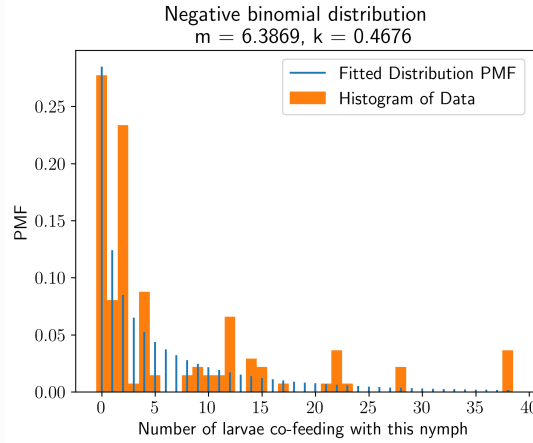
## 5.4 Results of fitting the negative binomial (NB) distribution to co-aggregation data $X$

With the NB distribution confirmed as the best (or close to the best) fit for the co-aggregation data,  $X$ , we can now fit the reparameterised NB distribution to subsets of the Kielder Forest data, to obtain the sample mean  $m$  and dispersion parameter  $k$ .

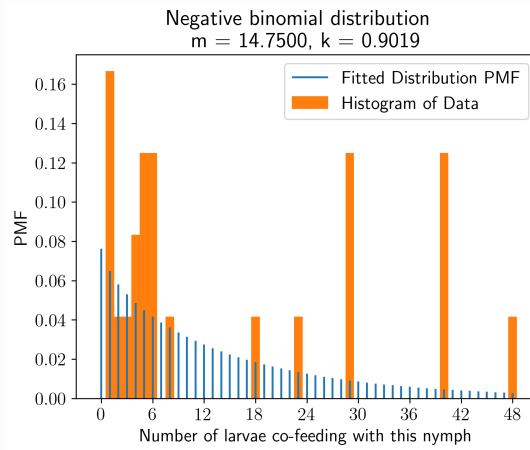
Figures 17, 18, 19, 20 and 21, show the results of MLE applied to the reparameterised negative binomial distribution for each of the 5 subsets of Kielder Forest data.



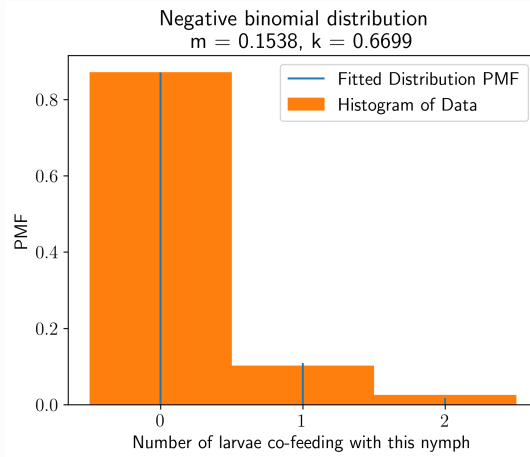
**Figure 17:** For counts of *I.ricinus* found on common shrews in 2004.



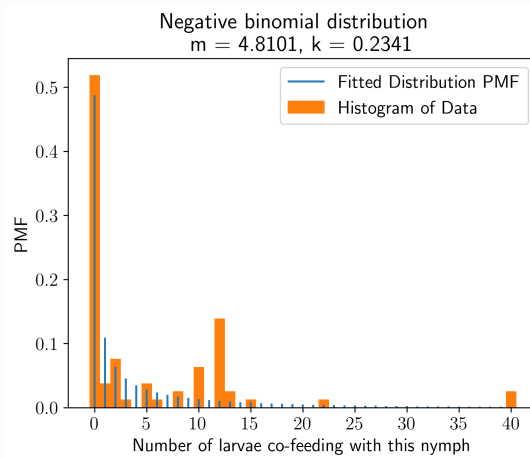
**Figure 18:** For counts of *I.trianguliceps* found on common shrews in 2004.



**Figure 19:** For counts of *I. ricinus* found on field voles in 2005.



**Figure 20:** For counts of *I. trianguliceps* found on field voles in 2005.



**Figure 21:** For counts of *I. trianguliceps* found on common shrews in 2005.

## 5.5 Estimates for $\alpha$ and its effect on $R_0$

Now that we have estimates for the sample mean of the co-aggregation data  $X$ , obtained from subsets of the Kielder Forest data, we could obtain a particular estimate for  $R_0$  based on some  $\alpha$  that we calculate (12). Researchers who wish to approximate the probability that a new outbreak becomes extinct would need to provide estimates about their particular tick species, host species, pathogen and environmental conditions.

However, we can understand the effect on  $R_0$  without specifying the value of  $\alpha$  by utilising the simple linear equation (14). Since a chain of transmission will certainly become extinct if  $R_0 \leq 1$ , then we are interested in the case where  $R_0 > 1$ . In other words, what is the minimum value of  $\alpha$  such that a chain of transmissions might persist? This is found with:

$$R_0 = m\alpha > 1 \implies \alpha > \frac{1}{m}.$$

Since  $m$  is known for each subset of Kielder Forest data, then letting  $\alpha$  vary allows us to see the effect on  $R_0$ . The minimum values of  $\alpha$  that would allow for  $R_0 > 1$ , for each subset of the Kielder Forest data, are shown in Table 7. Note that since  $0 \leq \alpha \leq 1$  then this imposes bounds on the value that  $R_0$  can possibly take. Also,  $m < 1 \implies R_0 < 1$ , hence the N/A result in Table 7.

Year	Tick species	Host species	$m$	$\alpha$ threshold
2004	<i>I. ricinus</i>	<i>common shrew</i>	3.929	0.255
2004	<i>I. trianguliceps</i>	<i>common shrew</i>	6.387	0.157
2005	<i>I. ricinus</i>	<i>field vole</i>	14.75	0.068
2005	<i>I. trianguliceps</i>	<i>common shrew</i>	4.81	0.208
2005	<i>I. trianguliceps</i>	<i>field vole</i>	0.154	N/A

**Table 7:** This table presents the threshold that  $\alpha$  must surpass that would allow  $R_0 > 1$ .

## 6 Estimates for the probability that a chain of transmissions becomes extinct

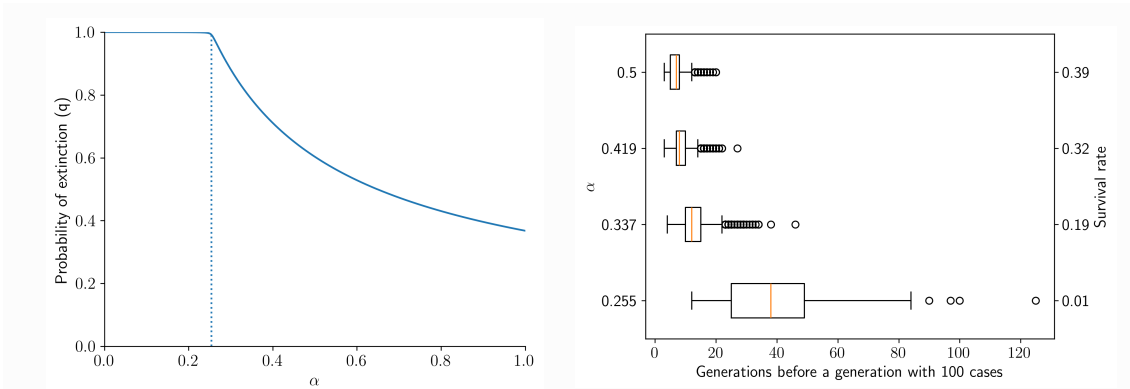
Finally we can use estimates of  $k$  and  $R_0$ , obtained for each subset of Kielder Forest data, to approximate the probability that a chain of transmissions becomes extinct after the introduction of a single infectious tick. By using the negative binomial distribution's probability generating function (11), reparameterised to use parameters  $R_0$  and  $k$ , we can use fixed point iteration to approximate the probability of eventual extinction. However, since we do not know the actual value of  $R_0$ , we instead graphically show the range of extinction probabilities based on the possible values of  $\alpha$ .

We take inspiration from Figures 9 and 10, which are reproduced to follow the example set by Lloyd-Smith et al. (2005), to generate Figures 22, 23, 24 and 25. We use the same methodology to generate both sets of figures, however, where  $R_0$  and  $k$  vary in Figure 9, we will instead let  $\alpha$  vary for each subset of Kielder Forest data while we fix the value of  $k$ , which is known. And where  $R_0 = 1.5$  for different values of  $k$  in Figure 10, we will instead fix  $k$  for each subset of Kielder Forest data while allowing  $\alpha$  to vary in those simulations. In each simulation (right-hand side) plot of Figures 22, 23, 24 and 25, we chose four equally-spaced values of  $\alpha$  between the minimum that would allow  $R_0 > 1$  and  $\alpha = 0.5$ . While values of  $\alpha > 0.5$  are possible, there may be unknown factors that reduce the probability of a co-feeding larva becoming an infectious nymph. So, lower values of  $\alpha$  are assumed to be more realistic.

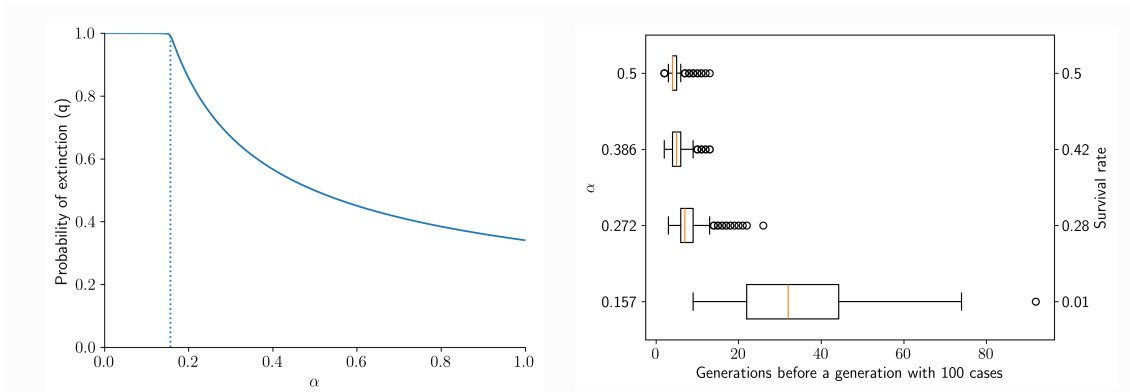
In Figures 22, 23, 24 and 25 the minimum value of  $\alpha$  such that  $R_0 > 1$  are displayed as broken vertical lines. Extinction with  $q = 1$  is guaranteed below those threshold values. The result of simulation (right-hand side in each figure) shows that for lower values of  $\alpha$ , on average, there are more generations required before an outbreak reaches the benchmark defined by Lloyd-Smith et al. (2005). We also see lower rates of outbreaks surviving for lower values of  $\alpha$ , which is expected.

The results of these simulations show varied outcomes. In Figure 24 we can see a rapid decline of the extinction probability  $q$  for relatively low values of  $\alpha$ , and relatively few generations before a generation with 100 offspring is found. In stark contrast, Figure 25 shows only a modest decline in  $q$  as  $\alpha$  increases, while many more generations on average are required before a generation with 100 offspring is found. The unsurprising result is that in some combinations of tick and host species, co-feeding in the absence of other forms of transmission is viable to maintain a pathogen in nature if a pathogen can have a sufficiently high  $\alpha$  (previously defined as the probability that a co-feeding larva becomes an infectious nymph). However, this is only part of the story.

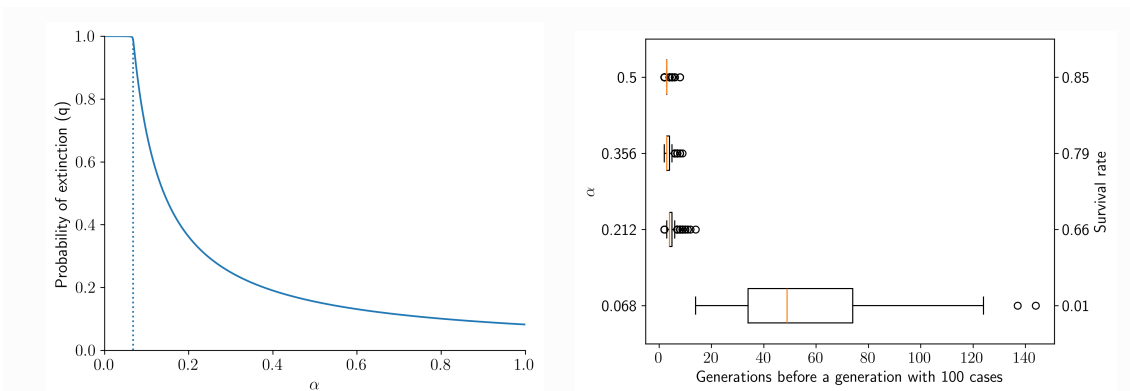
To see the effect of  $k$  in each subset of Kielder Forest data, we can instead assume that each subset has an offspring distribution with  $R_0 = 1.5$ . Figure 26 is a re-implementation of Figure 10, using exactly the same methodology, but instead of using arbitrarily-chosen values of  $k$  we will use the values of  $k$  found by fitting the NB distribution to each subset of Kielder Forest data. Figure 27 is another iteration of the same simulation that also uses the fitted values of  $k$ , but instead shows the total and cumulative populations of outbreaks that became extinct. Taken together, Figures 26 and 27 reveal that if these particular combinations of tick and host species involved a pathogen, or pathogens, that could infect larvae at a rate of  $R_0 = 1.5$  per infectious nymph, then those combinations that have lower  $k$  will on average reach the arbitrary benchmark in fewer generations. Outbreaks that become extinct will have roughly the same cumulative infected



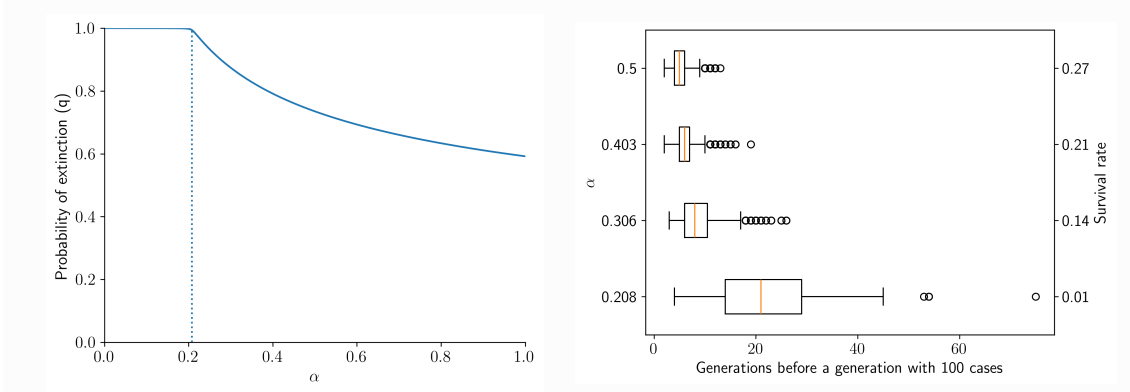
**Figure 22:** (Left) probability of extinction depending on  $\alpha$  and (right) the number of generations to reach a generation with 100 offspring depending on  $\alpha$ , and outbreaks' survival rates after 10,000 generations, for *I. ricinus* found on *common shrews* in 2004, and where  $m = 3.939, k = 0.622$



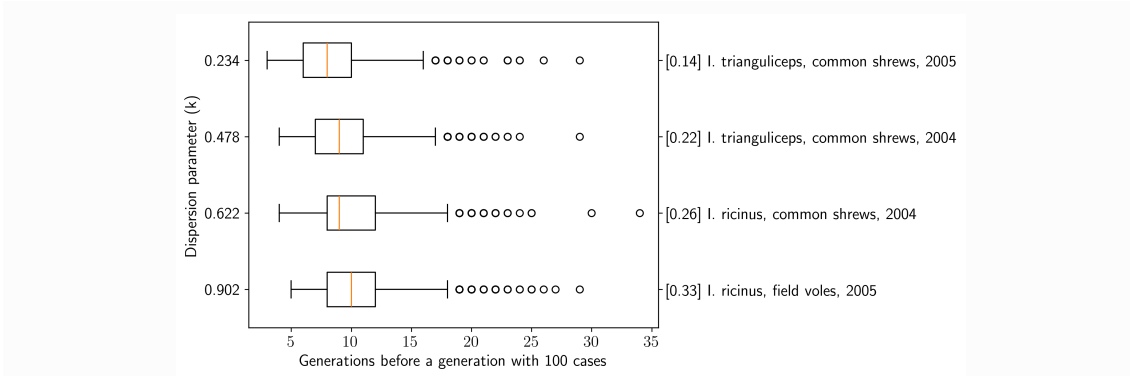
**Figure 23:** (Left) probability of extinction depending on  $\alpha$  and (right) the number of generations to reach a generation with 100 offspring depending on  $\alpha$ , and outbreaks' survival rates after 10,000 generations, for *I. trianguliceps* found on *common shrews* in 2004, and where  $m = 6.387, k = 0.478$ .



**Figure 24:** (Left) probability of extinction depending on  $\alpha$  and (right) the number of generations to reach a generation with 100 offspring depending on  $\alpha$ , and outbreaks' survival rates after 10,000 generations, for *I. ricinus* found on *field voles* in 2005, and where  $m = 14.75, k = 0.9019$ .



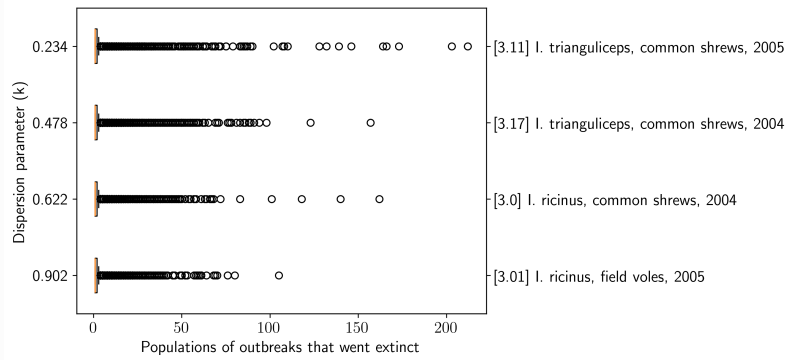
**Figure 25:** (Left) probability of extinction depending on  $\alpha$  and (right) the number of generations to reach a generation with 100 offspring depending on  $\alpha$ , and outbreaks' survival rates after 10,000 generations, for *I. trianguliceps* found on *common shrews* in 2005, and where  $m = 4.81$ ,  $k = 0.234$ .



**Figure 26:** The number of generations to reach a generation with 100 offspring depending on  $k$ , where  $R_0 = 1.5$ , for all combinations of Kielder Forest data that have  $m \geq 1$ . The values in square brackets are outbreaks' survival rates after 10,000 generations.

populations on average, but the outbreaks that have a lower  $k$  will include more extreme values. This shows the influence of superspreading: some nymphs co-feed with a disproportionately high number of larvae, leading to explosive growth that still sometimes ends with eventual extinction.





**Figure 27:** The cumulative numbers of infected individuals in extinct outbreaks, where  $R_0 = 1.5$ , for all combinations of Kielder Forest data that have  $m \geq 1$ . The values in square brackets are means of outbreak cumulative population sizes. These results show that for low values of  $k$ , outbreaks frequently reach extreme sizes but still become extinct.

## 7 Discussion

In this project we have made assumptions that vary in the reason for their introduction; some are simplifying assumptions that make the task more tractable, while some are appropriate due to the biological realism we wish to emulate. This discussion will focus on the assumptions made and the consequences of using those. We will also consider alternative assumptions we could have used, and potential ideas for future research based on modified assumptions.

The first and most obvious assumption in this project was to make co-feeding transmission the only viable route of transmission. This assumption is of interest since various factors are causing ticks to spread (Alkishe et al., 2017; Cunze et al., 2022; Medlock et al., 2013) and therefore, it is possible that a tick carrying some pathogen might move into an area where a naive population of vertebrate hosts are incompetent to maintain that pathogen. In that case, the role of systemic transmission would be limited. Assuming that co-feeding is the only viable route of transmission also allows us to use a single-type GWBP, which is effective for understanding the probability of extinction at the start of an outbreak for a directly-transmissible pathogen (Lloyd-Smith et al., 2005), but only when every new offspring is the same type and with the same offspring distribution for the entire analysis (Allen, 2019). Only considering co-feeding transmission would of course be inappropriate for estimating the probability of extinction for many TBD systems, where we know systemic transmission is important. However even in those cases, it would still be interesting to find that some pathogens could exist with some combinations of tick and host species by co-feeding transmission alone.

The assumption that co-feeding transmission is from infectious nymph to a co-feeding larvae only is a simplifying assumption; in reality it is possible that an infected nymph will feed close to, and infect, other nymphs or even adult ticks. The Kielder Forest data, for example, has many recorded observations of captured vertebrates that each have multiple nymphs of the same species. But if we suppose that an infectious nymph did infect another nymph, then the newly-infected nymph would have to survive moulting before it could feed on another vertebrate host as an adult, which would most-likely be a large vertebrate host (Randolph, 1998). Given that the more-numerous immature ticks frequently feed on smaller vertebrate hosts, then that particular adult tick would be unlikely to co-feed with other ticks in close proximity and that particular chain of transmissions would most likely end there.

We also make the simplifying assumption that more than one larvae that become infected in the same chain of transmissions will then co-feed as infectious nymphs on different vertebrate hosts. This is reasonable for non-nidicolous tick species, which tend to drop off their vertebrate hosts and moult in a wider range of locations (Nicholson et al., 2019). Nidicolous ticks, however, tend to stay close to where their vertebrate hosts live. An alternative assumption would have been to allow for multiple larvae to become infected and then to co-feed with other infectious nymphs on a single vertebrate host, as we might expect with nidicolous ticks. The effect of this might be to increase the overall probability  $\alpha$  (12) that co-feeding larvae become infected, due to more infectious nymphs increasing the overall load of pathogens. Alternatively, this complication could also be handled by dividing the co-feeding larvae amongst the infectious nymphs before applying  $\alpha$ , in the sense that a co-feeding larva can only become infected once.

The division of the Kielder Forest data into subsets that are each a combination of one tick

species, one host species and year of capture has implications for how we can interpret the estimates of  $R_0$ ,  $k$  and the probability that a chain of transmissions becomes extinct. The data is segregated by tick species, because the tick species under study, *I. ricinus* and *I. trianguliceps*, are known to be vectors for only one pathogen in common, which is *Babesia microti* (J. Gray et al., 2024; Rubel & Kahl, 2023). Therefore for most pathogens, a larva that co-feeds with an infectious nymph of a different species will not become an infectious nymph for that pathogen. Even if the two tick species under study were vectors for the same pathogen, then  $\alpha$  (the probability of a co-feeding larva becoming an infectious nymph) for each species would be different due to different moulting success rates and different magnitudes of vector competence. However if we wished to study *Babesia microti* in particular then separating the data by tick species is inappropriate. We could instead use a multi-type GWBP to represent that a nymph of either *I. ricinus* or *I. trianguliceps* could infect a larva of the other species. Using a multi-type GWBP was outside the scope of this project.

This project also divided the Kielder Forest data into subsets by host species. This reflects that the probability of transmission for a single species of tick is different depending on which vertebrate host they feed on. The probability that ticks co-feed close enough for transmission to occur depends on the host species on which they feed; since the field vole is larger than the common shrew, then we can expect different rates of transmission success during co-feeding for the two different species of host. However, partitioning the data by host species means the co-aggregation distribution leads to a loss of information; for example an infectious nymph could infect a larvae of the same species while co-feeding on a common shrew, and then, that larvae could survive to become a infectious nymph and co-feed with larvae of the same species on a field vole. Once again, a multi-type GWBP seems like an appropriate tool. This would be interesting future research.

Some work is required by anyone who wishes to use the results of this research; calculating the probability of extinction in a TBD transmission network requires multiple steps. Firstly, once field data is obtained, it should be simple to count the number of larvae that co-feed with each nymph, from which one can obtain the sample mean for the co-aggregation data. Then, one can fit the reparameterised NB distribution to obtain the dispersion parameter  $k$ . The real work would be determining a viable estimate of  $\alpha$  (12), which is needed to determine  $R_0$ . This would likely depend on the tick species, pathogen, host species and environmental conditions in which the transmission takes place. In this project, we were not able to determine accurate estimates for  $\alpha$ , so instead we analysed what effect  $\alpha$  has on  $R_0$  during co-feeding transmission. If one was to obtain plausible and reliable estimates for  $\alpha$ , then  $R_0$  due to co-feeding transmission could easily be obtained by using (14).

## 8 Conclusion

This project has investigated a scenario where co-feeding transmission is the only viable route of transmission for a tick-borne pathogen. We specifically looked at data from the Kielder Forest in England, that concerns *Ixodidae* (hard-bodied) ticks. The project aim was to use a single-type GWBP to find an approximate probability that a chain of transmissions becomes extinct, which is of particular interest at the start of an outbreak. We sought to implement what we believe is a novel technique to find an offspring distribution, which the single-type GWBP analysis requires.

In the presentation of data summaries, we implemented a novel technique to understand the heterogeneity in vertebrates' co-feeding transmission potentials. While aggregation was found to be somewhat consistent with the frequently-cited 80:20 rule, co-feeding transmission potential for the Kielder Forest data has much greater heterogeneity.

The analysis presented here indicates that for Kielder Forest, and for some combinations of tick and host species, there may be cases where co-feeding transmission alone is viable to maintain a pathogen in nature. However it would require further work to determine the probability that a co-feeding larva becomes an infectious nymph, which we have called  $\alpha$ , for a particular combination of tick, host and pathogen. The discussion reviews the many assumptions made and indicates that using a multi-type GWBP would be a good opportunity for future research.

## 9 References

- Alkishe, A. A., Peterson, A. T., & Samy, A. M. (2017). Climate change influences on the potential geographic distribution of the disease vector tick *Ixodes ricinus* (R. M. Wooten, Ed.). *PLOS ONE*, 12(12), e0189092. <https://doi.org/10.1371/journal.pone.0189092>
- Allen, L. J. (2017). A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2), 128–142. <https://doi.org/10.1016/j.idm.2017.03.001>
- Allen, L. J. (2019, December). Branching processes. In *Encyclopedia of theoretical ecology* (pp. 112–119). University of California Press. <https://doi.org/10.1525/9780520951785-022>
- Athreya, K. B., & Ney, P. E. (1972). The galton-watson process. In *Branching processes* (pp. 1–65). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-65371-1\\_1](https://doi.org/10.1007/978-3-642-65371-1_1)
- Bisanzio, D., Bertolotti, L., Tomassone, L., Amore, G., Ragagli, C., Mannelli, A., Giacobini, M., & Provero, P. (2010). Modeling the spread of vector-borne diseases on bipartite networks (C. V. Munayco, Ed.). *PLoS ONE*, 5(11), e13796. <https://doi.org/10.1371/journal.pone.0013796>
- Bown, K. J., Begon, M., Bennett, M., Woldehiwet, Z., & Ogden, N. H. (2003). Seasonal dynamics of anaplasma phagocytophilum in a rodent-tick (*Ixodes trianguliceps*) system, united kingdom. *Emerging Infectious Diseases*, 9(1), 63–70. <https://doi.org/10.3201/eid0901.020169>
- Bown, K. J., Lambin, X., Ogden, N. H., Begon, M., Telford, G., Woldehiwet, Z., & Birtles, R. J. (2009). Delineating anaplasma phagocytophilum ecotypes in coexisting, discrete enzootic cycles. *Emerging infectious diseases*, 15(12), 1948. <https://doi.org/10.3201/eid1512.090178>
- Bown, K. J., Lambin, X., Telford, G., Heyder-Bruckner, D., Ogden, N. H., & Birtles, R. J. (2011). The common shrew (*Sorex araneus*): A neglected host of tick-borne infections? *Vector-Borne and Zoonotic Diseases*, 11(7), 947–953. <https://doi.org/10.1089/vbz.2010.0185>
- Bown, K., Lambin, X., Telford, G., Ogden, N., Telfer, S., Woldehiwet, Z., & Birtles, R. (2008). Relative importance of *Ixodes ricinus* and *Ixodes trianguliceps* as vectors for *Anaplasma phagocytophilum* and *Babesia microti* in field vole (*Microtus agrestis*) populations. *Applied and Environmental Microbiology*, 74(23), 7118–7125. <https://doi.org/10.1128/aem.00625-08>
- Brauer, F. (2008a). Compartmental models in epidemiology. In *Mathematical epidemiology* (pp. 19–79). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-78911-6\\_2](https://doi.org/10.1007/978-3-540-78911-6_2)
- Brauer, F. (2008b). An introduction to networks in epidemic modeling. In *Mathematical epidemiology* (pp. 133–146). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-78911-6\\_4](https://doi.org/10.1007/978-3-540-78911-6_4)
- Brunner, J. L., & Ostfeld, R. S. (2008). Multiple causes of variable tick burdens on small-mammal hosts. *Ecology*, 89(8), 2259–2272. <https://doi.org/10.1890/07-0665.1>
- CDC. (2003, May). Severe acute respiratory syndrome — singapore, 2003 [Accessed: 2025-05-20]. <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5218a1.htm>
- Cunze, S., Glock, G., Kochmann, J., & Klimpel, S. (2022). Ticks on the move—climate change-induced range shifts of three tick species in europe: Current and future habitat suitability for *Ixodes ricinus* in comparison with *Dermacentor reticulatus* and *Dermacentor marginatus*. *Parasitology Research*, 121(8), 2241–2252. <https://doi.org/10.1007/s00436-022-07556-x>
- Dautel, H., Kämmer, D., & Kahl, O. (2016, November). 24. how an extreme weather spell in winter can influence vector tick abundance and tick-borne disease incidence. In *Ecology*

- and prevention of lyme borreliosis (pp. 335–349). Brill — Wageningen Academic. [https://doi.org/10.3920/978-90-8686-838-4\\_24](https://doi.org/10.3920/978-90-8686-838-4_24)
- Devevey, G., & Brisson, D. (2012). The effect of spatial heterogeneity on the aggregation of ticks on white-footed mice. *Parasitology*, 139(7), 915–925. <https://doi.org/10.1017/s003118201200008x>
- Diekmann, O., & Heesterbeek, J. (2000, January). *Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation*. Wiley.
- Ferreri, L., Bajardi, P., & Giacobini, M. (2016). Non-systemic transmission of tick-borne diseases: A network approach. *Communications in Nonlinear Science and Numerical Simulation*, 39, 149–155. <https://doi.org/https://doi.org/10.1016/j.cnsns.2016.02.034>
- Ferreri, L., Giacobini, M., Bajardi, P., Bertolotti, L., Bolzoni, L., Tagliapietra, V., Rizzoli, A., & Rosa, R. (2014). Pattern of tick aggregation on mice: Larger than expected distribution tail enhances the spread of tick-borne pathogens (J. Lloyd-Smith, Ed.). *PLoS Computational Biology*, 10(11), e1003931. <https://doi.org/10.1371/journal.pcbi.1003931>
- Galvani, A. P., & May, R. M. (2005). Dimensions of superspreading. *Nature*, 438(7066), 293–295. <https://doi.org/10.1038/438293a>
- Garske, T., & Rhodes, C. (2008). The effect of superspreading on epidemic outbreak size distributions. *Journal of Theoretical Biology*, 253(2), 228–237. <https://doi.org/10.1016/j.jtbi.2008.02.038>
- Gray, J., Kahl, O., & Zintl, A. (2024). Pathogens transmitted by ixodes ricinus. *Ticks and Tick-borne Diseases*, 15(6), 102402. <https://doi.org/10.1016/j.ttbdis.2024.102402>
- Gray, J. S., Estrada-Peña, A., & Vial, L. (2014). Ecology of nidicolous ticks. In *Biology of ticks volume 2* (pp. 39–60). OUP USA. <https://books.google.com.au/books?id=gck4AAAAQBAJ>
- Harrison, A., & Bennett, N. C. (2012). The importance of the aggregation of ticks on small mammal hosts for the establishment and persistence of tick-borne pathogens: An investigation using the  $r_0$  model. *Parasitology*, 139(12), 1605–1613. <https://doi.org/10.1017/s0031182012000893>
- Hartemink, N., Randolph, S., Davis, S., & Heesterbeek, J. (2008). The basic reproduction number for complex disease systems: Defining  $r_0$  for tick-borne infections. *The American Naturalist*, 171(6), 743–754. <https://doi.org/10.1086/587530>
- Herrmann, C., & Gern, L. (2015). Search for blood or water is influenced by borrelia burgdorferi in ixodes ricinus. *Parasites & Vectors*, 8(1), 6. <https://doi.org/10.1186/s13071-014-0526-2>
- Hurry, G., Maluenda, E., Sarr, A., Belli, A., Hamilton, P., Duron, O., Plantard, O., & Voordouw, M. (2021). Infection with borrelia afzelii reduces moulting time of ixodes ricinus ticks. <https://doi.org/10.21203/rs.3.rs-139147/v1>
- Johnson, N. (2023a). Blood feeding as a life choice and the multiple functions of tick saliva. In *Ticks* (pp. 45–63). Elsevier. <https://doi.org/10.1016/b978-0-323-91148-1.00011-3>
- Johnson, N. (2023b). A brief introduction to ticks. In *Ticks* (pp. 1–7). Elsevier. <https://doi.org/10.1016/b978-0-323-91148-1.00001-0>
- Johnson, N. (2023c). An introduction to tick-borne disease. In *Ticks* (pp. 65–74). Elsevier. <https://doi.org/10.1016/b978-0-323-91148-1.00004-6>
- Johnson, N. (2023d). Tick classification and diversity. In *Ticks* (pp. 9–23). Elsevier. <https://doi.org/10.1016/b978-0-323-91148-1.00013-7>
- Johnstone-Robertson, S. P., Diuk-Wasser, M. A., & Davis, S. A. (2020). Incorporating tick feeding behaviour into  $R_0$  for tick-borne pathogens. *Theoretical Population Biology*, 131, 25–37. <https://doi.org/10.1016/j.tpb.2019.10.004>

- Kivaria, F. (2006). Estimated direct economic costs associated with tick-borne diseases on cattle in tanzania. *Tropical Animal Health and Production*, 38(4), 291–299. <https://doi.org/10.1007/s11250-006-4181-2>
- Lew-Tabor, A., & Rodriguez Valle, M. (2016). A review of reverse vaccinology approaches for the development of vaccines against ticks and tick borne diseases. *Ticks and Tick-borne Diseases*, 7(4), 573–585. <https://doi.org/10.1016/j.ttbdis.2015.12.012>
- Lim, J.-S., Cho, S.-I., Ryu, S., & Pak, S.-I. (2020). Interpretation of the basic and effective reproduction number. *Journal of Preventive Medicine and Public Health*, 53(6), 405–408. <https://doi.org/10.3961/jpmph.20.288>
- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066), 355–359. <https://doi.org/10.1038/nature04153>
- Maliyoni, M., Chirove, F., Gaff, H. D., & Govinder, K. S. (2017). A stochastic tick-borne disease model: Exploring the probability of pathogen persistence. *Bulletin of Mathematical Biology*, 79(9), 1999–2021. <https://doi.org/10.1007/s11538-017-0317-y>
- Mammal Society. (2024a). Common shrew [Accessed: 2025-10-20]. <https://mammal.org.uk/british-mammals/common-shrew>
- Mammal Society. (2024b). Field vole [Accessed: 2025-10-20]. <https://mammal.org.uk/british-mammals/field-vole>
- Medlock, J. M., Hansford, K. M., Bormane, A., Derdakova, M., Estrada-Peña, A., George, J.-C., Golovljova, I., Jaenson, T. G., Jensen, J.-K., Jensen, P. M., et al. (2013). Driving forces for changes in geographical distribution of ixodes ricinus ticks in europe. *Parasites & Vectors*, 6(1). <https://doi.org/10.1186/1756-3305-6-1>
- Nicholson, W., Sonenshine, D., Noden, B., & Brown, R. (2019). Ticks (ixodida). In *Medical and veterinary entomology* (pp. 603–672). Elsevier. <https://doi.org/10.1016/b978-0-12-814043-7.00027-3>
- Norman, R., Ross, D., Karen Laurenson, M., & Hudson, P. J. (2004). The role of non-viraemic transmission on the persistence and dynamics of a tick borne virus? louping ill in red grouse (*lagopus lagopus scoticus*) and mountain hares (*lepus timidus*). *Journal of Mathematical Biology*, 48(2), 119–134. <https://doi.org/10.1007/s00285-002-0183-5>
- Otranto, D., Dantas-Torres, F., & Santos-Silva, M. (2017). *Ixodes ricinus* (linnaeus, 1758) (figs. 6769). In *Ticks of europe and north africa* (pp. 189–195). Springer International Publishing. [https://doi.org/10.1007/978-3-319-63760-0\\_38](https://doi.org/10.1007/978-3-319-63760-0_38)
- Parola, P., & Raoult, D. (2001). Ticks and tickborne bacterial diseases in humans: An emerging infectious threat. *Clinical Infectious Diseases*, 32(6), 897–928. <https://doi.org/10.1086/319347>
- Perkins, S. E., Cattadori, I. M., Tagliapietra, V., Rizzoli, A. P., & Hudson, P. J. (2003). Empirical evidence for key hosts in persistence of a tick-borne disease. *International Journal for Parasitology*, 33(9), 909–917. [https://doi.org/10.1016/s0020-7519\(03\)00128-0](https://doi.org/10.1016/s0020-7519(03)00128-0)
- Pfäffle, M., Petney, T., & Madder, M. (2017). *Ixodes trianguliceps* birula, 1895 (figs. 57–59). In *Ticks of europe and north africa* (pp. 167–171). Springer International Publishing. [https://doi.org/10.1007/978-3-319-63760-0\\_34](https://doi.org/10.1007/978-3-319-63760-0_34)
- Randolph, S. (1994). Density-dependent acquired resistance to ticks in natural hosts, independent of concurrent infection with babesia microti. *Parasitology*, 108(4), 413–419. <https://doi.org/10.1017/s003118200007596x>

- Randolph, S. (1998). Ticks are not insects: Consequences of contrasting vector biology for transmission potential. *Parasitology Today*, 14(5), 186–192. [https://doi.org/10.1016/s0169-4758\(98\)01224-1](https://doi.org/10.1016/s0169-4758(98)01224-1)
- Rice, J. A. (2007). *Mathematical statistics and data analysis* (3. ed., internat. student ed.) [teilweise auch mit CD-ROM, ISBN der CD-ROM: 0-495-10986-X]. Thomson Brooks/Cole.
- Rochlin, I., & Toledo, A. (2020). Emerging tick-borne pathogens of public health importance: A mini-review. *Journal of Medical Microbiology*, 69(6), 781–791. <https://doi.org/10.1099/jmm.0.001206>
- Rosà, R., Pugliese, A., Norman, R., & Hudson, P. J. (2003). Thresholds for disease persistence in models for tick-borne infections including non-viraemic transmission, extended feeding and tick aggregation. *Journal of Theoretical Biology*, 224(3), 359–376. [https://doi.org/10.1016/s0022-5193\(03\)00173-5](https://doi.org/10.1016/s0022-5193(03)00173-5)
- Rubel, F., & Kahl, O. (2023). The eurasian shrew and vole tick ixodes trianguliceps: Geographical distribution, climate preference, and pathogens detected. *Experimental and Applied Acarology*, 90(1–2), 47–65. <https://doi.org/10.1007/s10493-023-00797-0>
- Shaw, D., Grenfell, B., & Dobson, A. (1998). Patterns of macroparasite aggregation in wildlife host populations. *Parasitology*, 117(6), 597–610. <https://doi.org/10.1017/s0031182098003448>
- Sprong, H., Azagi, T., Hoornstra, D., Nijhof, A. M., Knorr, S., Baarsma, M. E., & Hovius, J. W. (2018). Control of lyme borreliosis and other ixodes ricinus-borne diseases. *Parasites & Vectors*, 11(1). <https://doi.org/10.1186/s13071-018-2744-5>
- VanderWaal, K. L., & Ezenwa, V. O. (2016). Heterogeneity in pathogen transmission: Mechanisms and methodology (D. Hawley, Ed.). *Functional Ecology*, 30(10), 1606–1622. <https://doi.org/10.1111/1365-2435.12645>
- Vial, L. (2009). Biological and ecological characteristics of soft ticks (ixodida: Argasidae) and their impact for predicting tick and associated disease distribution. *Parasite*, 16(3), 191–202. <https://doi.org/10.1051/parasite/2009163191>
- Voordouw, M. J. (2014). Co-feeding transmission in lyme disease pathogens. *Parasitology*, 142(2), 290–302. <https://doi.org/10.1017/s0031182014001486>
- Woolhouse, M. E., Dye, C., Etard, J.-F., Smith, T., Charlwood, J., Garnett, G., Hagan, P., Hii, J. x., Ndhlovu, P., Quinell, R., et al. (1997). Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academy of Sciences*, 94(1), 338–342. <https://doi.org/10.1073/pnas.94.1.338>



## 10 Appendix

The author completed all analysis using Python and Jupyter notebooks, and compiled this document using LaTeX (TeXstudio). All analysis, code, and LaTeX source files are available at [https://github.com/JasonThomasData/honours\\_project](https://github.com/JasonThomasData/honours_project).

Figures 1, 11 were made using <https://app.diagrams.net/>.

Tables were formatted with <https://www.tablesgenerator.com/#>.