



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.

November 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 and Dr Simon Johnstone-Robertson for their supervision, feedback and advice during this project.
- Dr Haydar Demirhan for his advice regarding statistical tests.
- PhD candidate Daniel Longmuir, who helped me understand how to derive an important result.

## Abstract

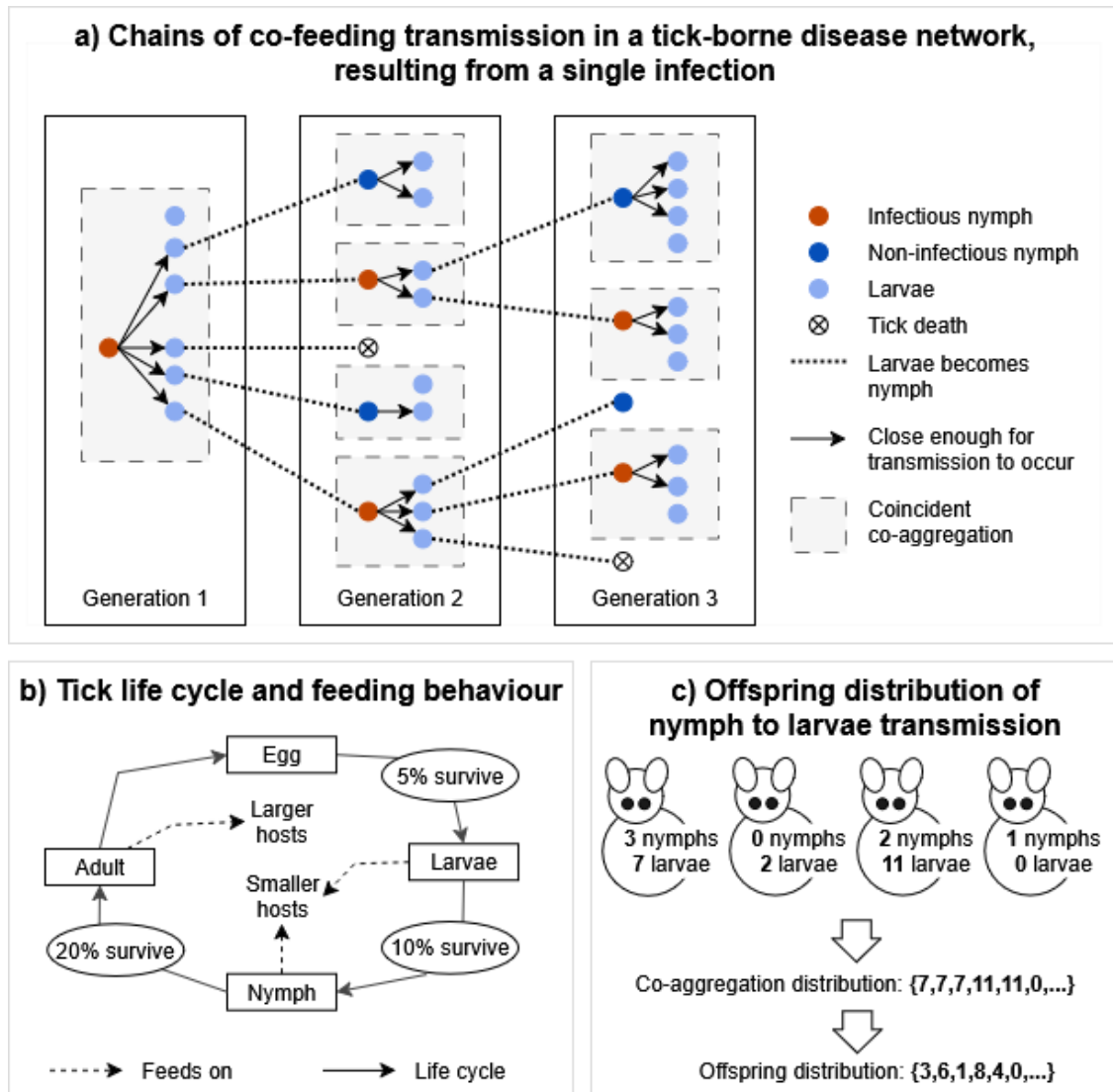
Importance of tick-borne diseases (one or two sentences).

Aim of research (one sentence).

Methods (two sentences).

Results and findings (four or more sentences).

## 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-aggregate 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 [1]. 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.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Tick biology . . . . .	4
1.2	Tick-borne diseases . . . . .	5
1.3	The definition and importance of $R_0$ . . . . .	7
1.4	Heterogeneity in disease networks . . . . .	8
1.5	Using the negative binomial distribution . . . . .	9
1.6	Thinking about TBD transmission using contact networks . . . . .	10
1.7	Galton-Watson Branching Processes . . . . .	11
<b>2</b>	<b>Research project</b>	<b>13</b>
2.1	Aims . . . . .	13
2.2	Methods . . . . .	13
2.3	Expected outcomes . . . . .	13
<b>3</b>	<b>Summaries of Kielder Forest data</b>	<b>15</b>
3.1	Evidence of co-aggregation . . . . .	15
3.2	Tick burden heterogeneity . . . . .	18
<b>4</b>	<b>Reproducing the results by Lloyd-Smith et al, 2005</b>	<b>22</b>
<b>5</b>	<b>Reparameterisation of the negative binomial distribution</b>	<b>24</b>
<b>6</b>	<b>The probability that a chain of TBD transmissions becomes extinct</b>	<b>25</b>
6.1	Finding a candidate offspring distribution . . . . .	25
6.2	Fitting a distribution to $X_\alpha$ . . . . .	26
6.3	Goodness of fit tests for co-aggregation distribution . . . . .	28
6.4	Fitted co-aggregation and offspring distributions . . . . .	31
6.5	Estimates for $\alpha$ and the effect on $R_0$ . . . . .	33
6.6	Estimates for the probability of extinction . . . . .	34
<b>7</b>	<b>Discussion</b>	<b>35</b>
<b>8</b>	<b>Conclusion</b>	<b>36</b>
<b>9</b>	<b>References</b>	<b>37</b>
<b>10</b>	<b>Appendix</b>	<b>40</b>

# 1 Introduction

Worldwide, tick-borne diseases (TBD) pose a significant health burden for humans and livestock. That is because ticks carry pathogens that can infect their vertebrate hosts when ticks take blood meals from their hosts. The pathogens that spread when ticks feed can cause diseases, and those diseases often have severe consequences for the individual vertebrate hosts that ticks feed on [2]. 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 and 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 [3].

When most ticks of a particular life stage feed on a small number of vertebrate hosts, then they are said to aggregate [4]. A necessary, but not sufficient, condition for co-feeding transmission to occur is the coincidence of larval and nymphal ticks aggregating on the same host [5], which 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 coincident co-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 ticks can be found on just 20% of vertebrate hosts [6].

FIND A NEW PLACE FOR THIS PARAGRAPH... IF IT'S NEEDED A TBD network is a web of transmissions: each edge is a transmission of a pathogen between individual ticks and their vertebrate hosts, and sometimes transmissions directly between individual ticks. TBD networks are complex; some networks of a single pathogen have multiple tick species and multiple host vertebrates [2]. Ticks also exhibit different behaviours depending on seasonal variation [7]. Adding to this complexity is multiple types of transmission, where systemic (tick-to-host) transmission is the most common and co-feeding transmission (tick-to-tick) is less-common [8].

## 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 [3]. There are three families of ticks described in research: *Ixodidae* (Hard Ticks), *Argasidae* (Soft Ticks) and *Nuttalliellidae*. *Ixodidae* has the most known genres and about 700 species. *Nuttalliellidae* has just one known species [9].

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 [10]. 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 [9]. Most *Argasidae* are nidicolous [11], while 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 [12]. 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 [9].

The biology of *Ixodidae* make them particularly competent vectors for many diseases. When they feed, they can stay attached for more than one week [13]. Their extended period of feeding means that the transmission of pathogens is more likely to occur [14]. They lack digestive enzymes, meaning that their midgut is relatively benign, which increases the probability of transstadial transmission. When they feed, they inject saliva into their hosts, which increases the chance of successful transmission [14].

Since nidicolous ticks such as *I. trianguliceps* are specialists, then usually they are not directly a threat to humans or livestock. However, some nidicolous ticks will bite any vertebrate that comes into contact with their habitat. Otherwise, many nidicolous tick species could maintain a pathogen among a population of reservoir hosts [15].

We can expect to find ticks of different life stages in different frequencies and can expect them to feed on different vertebrate hosts. Given that ticks die during and between their life stages, then larvae are most abundant, nymphs less common, and adults less common still [1]. A tick's life stage will 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 [16, 1]. In comparison, *I. trianguliceps* will feed almost exclusively on small mammal hosts, and is regarded as a specialist [17, 18].

Ticks vary in their lifespans and their yearly active periods. *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. *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.

## 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 of those have severe health implications for vertebrate hosts, including humans [2].

*I. ricinus* is the most important vector in central and western Europe, and is a vector for at least 20 pathogens, and at least eight pathogens that can affect humans, including Lyme disease and tick-borne encephalitis (TBE) [14]. 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 [19].

While the impact on human health is significant, so is the impact on the livestock industry. A

recent paper estimated the cumulative impact on livestock due to TBD in India’s dairy industry was \$787.63 million USD [20]. A 2006 article reported that economic losses in Tanzania’s cattle industry were estimated to be \$364 million USD [21].

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 [22]. *I. trianguliceps* has been reported across mainland Europe and in the British Isles [23]. 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 [18, 24].

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 [25]. Modelling suggests that under different climate change scenarios, the habitat of *I. ricinus* will probably spread northwards in Northern Europe [26]. 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 [27]. 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 [25].

There are three ways that a pathogen can spread through populations of ticks, as Harrison and Bennett explain in their 2012 article:

- Female ticks may transmit the pathogen to eggs via transovarial transmission (vertical transmission).
- Ticks may feed on, and infect, a host leading to a systemic infection; other ticks may then acquire an infection by ingesting the blood of the infected host (systemic transmission).
- Ticks may become infected by co-feeding, spatially or temporally, alongside infected ticks; this does not require the host to have a systemic infection but instead pathogens are passed from one tick to another as they feed together. This is referred to as co-feeding transmission [8].

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 must survive transstadially; transstadial transmission is the maintenance of the pathogen between life stages [10]. 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 disease systems where vertebrate hosts are incompetent vectors for maintaining a pathogen [8]. Several studies suggest that co-feeding transmission is required for TBE to persist in nature [28, 8], 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 [4, 29, 30]. Some studies show that co-feeding transmission may allow a pathogen to persist even without systemic transmission [29, 30]. 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 transmitting the pathogen 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*, commonly called a deer tick or sheep tick, the larva-to-nymph moulting success rate was reported at 91.9% under some laboratory conditions [31]. 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 [32]. For the species *I. trianguliceps*, mortality from larvae to nymph is density-dependent [33].

### 1.3 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 [34]. **It differs from the effective reproductive number  $R_t$** , which is used to quantify the efficacy of control measures in real time [35]. Taken together, research sometimes uses the term  $R$  to indicate the reproductive number at different stages of an outbreak [36].

A disadvantage 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:

$$I(t) \approx Ce^{rt} \tag{1}$$

where  $I$  is the total number of infections up until time  $t$ ,  $C$  is some constant, and  $r$  is the real time growth rate [34].

$R_0$  is not a rate and is dimensionless, however, it has a relationship with the real-time growth rate,  $r$  [34]:

$$\begin{aligned} R_0 < 1 &\iff r < 0 \\ R_0 > 1 &\iff r > 0 \end{aligned} \tag{2}$$

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 involve 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 [28]. Structured population models allow for modelling these complexities [34]. 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 infections per infection. Finding  $R_0$  in previous research has involved using contact trace data [37], and this thesis introduces a method



to simulate such data based on tick aggregation data.

## 1.4 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 disease, host population and environmental factors [37]. In transmissible disease systems, heterogeneity in individual behaviour can mean that some of the infected 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 [38], which is often the case with infections like gonorrhoea and HIV/AIDS. The individuals who transmit the disease in these situations are superspreaders. Important research in 2005 made the case that superspreading events are normal events in the spread of infectious diseases [37], and their 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 [39].

To address the heterogeneity in the ... TALK ABOUT THE INDIVIDUAL REPRODUCTIVE NUMBER.

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 "superheaders" [40]. This implies we can expect many reasons to find TBD transmission networks that feature heterogeneity.

For an individual tick to infect many others, a necessary but insufficient condition is 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 [6]. This heterogeneity is also present in TBD systems; there is significant evidence to suggest that aggregation data follows the 80:20 rule [5, 41, 4]. When larvae and nymphs aggregate on the same vertebrate hosts, they are said to co-aggregate.

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. Transmission potential was calculated as:

$$R_0 \propto \sum_i^m v_i^2 h_i \quad (3)$$

where  $m$  is the number of vertebrates,  $v_i$  is the proportion of all ticks found on vertebrate  $i$ , and  $h_i$  is the proportion of vertebrates. With this, Perkins et al used Lorenz curves to quantify the heterogeneity in transmission potential. They found that 20% of vertebrate hosts were responsible for 74% of transmission potential, but that approximation jumps to 94% when only co-feeding ticks were considered [42].

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 [43]. 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. This 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 [44]. One approach is to use the SIR modelling framework to define stochastic differential equations [45], but there are other stochastic methods available. Also in any population, the degree infectiousness is distributed continuously, which further impedes our ability to divide a population into homogeneous subgroups [37]. This thesis addresses the heterogeneity that occurs in TBD systems by using a branching process simulation, which is defined in a later section.

## 1.5 Using the negative binomial distribution

The work of Lloyd-Smith and others in 2005 showed that the negative binomial distribution was a better fit than either the Poisson or geometric distributions to many contact trace and surveillance datasets [37]; this is an important result that this thesis will reproduce. While the negative binomial distribution is frequently presented in textbooks with an integer parameter, Lloyd-Smith et al used a reparameterisation that allows for two continuous parameters: sample mean  $m$  and dispersion parameter, although that was a known result  $k$  [46]. For a contact trace dataset, the sample mean is  $R_0$ . Lloyd-Smith et al showed that  $R_0$  can be low and a pathogen may still spread if the over-dispersion parameter  $k$  is low enough, since a low  $k$  indicates heterogeneity in the data, meaning there exist superspreading events that allow a transmissible pathogen to spread.

The work by Lloyd-Smith et al showed 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 negative binomial distribution, as below:

$$\begin{aligned} v &\sim \text{Gamma}\left(R_0, \frac{R_0}{k}\right) \\ Z &\sim \text{Poisson}(v) \\ Z &\sim \text{negbinom}(R_0, k) \end{aligned} \tag{4}$$

This is useful for several reasons. It allows us to query the effect of individual heterogeneity, and Lloyd-Smith et al found systems with high individual variation had infrequent but explosive outbreaks. This is a useful result because it allows for estimating the probability that a chain of transmission becomes extinct via numerical approximation, which is explored later in this project.

The negative binomial 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. In

1998, Shaw et al showed that the NB distribution was an appropriate fit for 90% of aggregation datasets that they looked at [47]. Many articles concerning tick aggregation data have used the negative binomial distribution [17, 8, 41]. 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 [5, 48].

## 1.6 Thinking about TBD 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. Those contacts may also facilitate transmission; the edges and nodes of a transmission network is contained within a contact network GET CITATION FOR THIS. In the case of a vector-borne disease, a bipartite graph is useful to represent contacts between vectors and hosts.

Network thinking can help overcome some of the problems associated with deterministic compartment models; network models allow for representing the heterogeneity that exists in individual behaviour. Superspreaders can feature in these networks as nodes This can lead to more useful outcomes, as some simulations suggest that stochastic methods lead to more accurate estimates for  $R_0$  when early outbreak data is used [49].

In 2010, Bisanzio et al 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 vector and host, showed that extreme aggregation on hosts would greatly impact the epidemic threshold [48]. In 2016, the work of Ferreri et al followed the work of Bisanzio et al. 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 tick.

In 2020, the work of Johnstone-Robertson et al characterised a tick-host contact network as a directed, acyclic, bipartite graph. The ticks represent one node, but the edges leading away from the vertebrate host node  $k_{out}$  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 per tick. This implies that the tick nodes represent a tick in its larval and then nymphal life stages. Johnstone-Robertson et al then superimposed a transmission network; co-feeding (edges between ticks) and systemic (edges from tick-to-host, host-to-tick) [4]. This work is the main inspiration for the method, which is presented later in this thesis, of simulating outbreaks by using co-aggregation data. An important result in the paper is that for co-feeding transmission,

$$R_0 \propto \frac{\langle k_{in}k_{out} \rangle}{\langle k_{in} \rangle} \quad (5)$$

where  $k_{in}$  is the number of nymphs and  $k_{out}$  is the number of larvae that are found on an individual vertebrate host.

## 1.7 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 families in Victorian Britain [50]. Whether a family becomes extinct because each father has no sons to carry on the family name, or a disease outbreak becomes extinct because each infected person infects no one else, are both addressed with the same technique.

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; the number of offspring in generation  $n$  depends on the probability of extinction in generation  $n - 1$ , but not on previous generations [51].

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, etc). Let  $q_k$  be the offspring distribution, such that  $\sum_{k=0}^{\infty} q_k = 1$  [34], where  $k$  is the number of offspring. Below is the technique applied to a single-type branching process.

Let  $z$  be the probability of extinction, and note that  $g(z)$  is a probability generating function.  $z$  is applied to individuals in the branching process in the sense that for extinction to occur, every individual must have 0 offspring.

$$g(z) = \sum_{k=0}^{\infty} q_k z^k, |z| < 1 \quad (6)$$

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.

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

$$z_n = g(z_{n-1}) = \sum_{k=0}^{\infty} q_k (z_{n-1})^k \quad (7)$$

and 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. That leads to the simplification  $z = g(z)$ . Thanks to the 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, then we can use  $R_0 < 1 \implies z_{\infty} = 1$  [34].

The single-type branching process defined above is useful when offspring are the same type as their parent [51]. 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  $R_0 > 1$  [52].

This project assumes that ticks are not competent vectors for the same pathogens, and that different vertebrate hosts have different behaviours, but infected nymphs beget infected nymphs during co-feeding transmission, so therefore a single-type branching process is appropriate.

## 2 Research project

### 2.1 Aims

- Implement a technique, which we believe is novel, to generate an offspring distribution by using tick burden data, which field ecologists commonly-collect. An offspring distribution is required to use a GWBP simulation.
- Publish and document the use of a new method for use by TBD researchers, who may wish to calculate the probability that a TBD goes extinct, where co-feeding is the only viable transmission route.
- Find estimates for parameters to be used as a threshold; passing the threshold would be necessary, but not sufficient, for a tick-borne pathogen to persist in nature via co-feeding transmission alone. Parameters below the threshold would necessarily cause extinction.

### 2.2 Methods

1. Analyse and summarise provided tick co-aggregation datasets.
2. Reproduce results from an important paper that simulates extinctions of directly-transmitted disease chains. Then, apply these extinction simulations to simulated TBD networks, where co-feeding transmission is a form of direct transmission.
3. Implement a parameterisation of the Negative Binomial distribution that uses that simulated offspring distribution's sample mean as a shape parameter. This is currently not supported in Python's SciPy. This will be one candidate Offspring Distribution.
4. Use the co-aggregation data to derive offspring distributions, which is required for the GWBP analysis, for each species of tick to determine the probability that a chain of transmissions becomes extinct.
5. Fit the reparameterised negative binomial, geometric and Poisson distributions to each subset of data, as candidates for offspring distributions. Use a goodness of fit method to check which distribution best-fits each subset of data.
6. Once we have a candidate offspring distribution for each tick species, use the theory of branching processes to find the probability of pathogen extinction based on given starting parameters.
7. Ensure all code is documented, easy to read and usable for future researchers. Ensure code is published online for others to use.

### 2.3 Expected outcomes

- The negative binomial distribution will be a better fit for simulated offspring data, when compared with Poisson and geometric distributions.

- Some combinations of parameters will allow for  $R_0 > 1$ , indicating that co-feeding transmission might be sufficient to maintain some combinations of: pathogen, tick, and vertebrate host species, in different environmental conditions.
- A single-type GWBP may be too simple, and there will be a cause to recommend a multi-type GWBP simulation in future research. We can expect this to feature in the discussion section of this thesis.

### 3 Summaries of Kielder Forest data

This project makes use of data provided by **INSERT** Kevin Bown. The **INSERT** collected that data in the Kielder Forest in the United Kingdom between April and August in 2004 and 2005. The provided dataset presents the numbers of larvae and nymphs found on each vertebrate. Vertebrates were included in the collection once only.

The analysis by Bown et al originally **INSERT DETAILS ABOUT WHY THEY COLLECTED THE DATA, HOW, ETC**.

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.

#### 3.1 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 1.

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

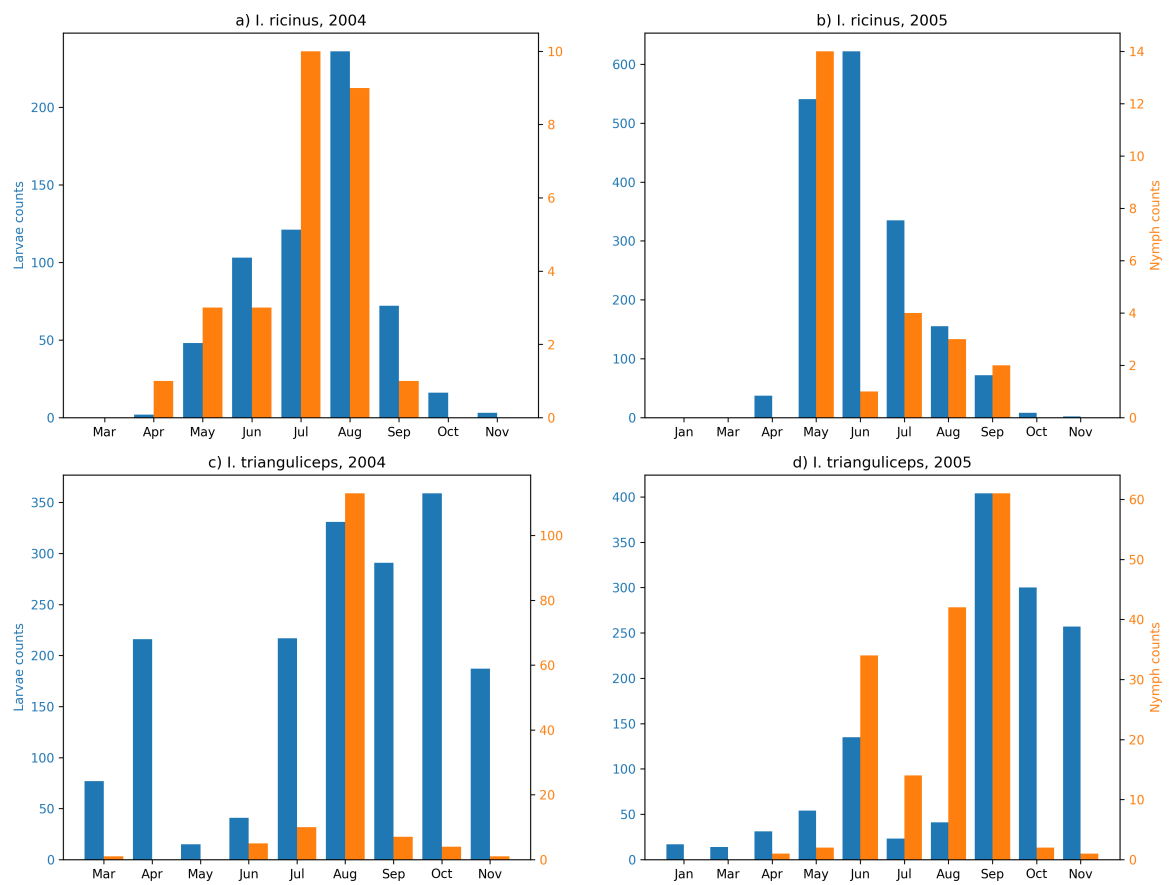
**Table 1:** Overall counts of nymphs and larvae, with several vertebrate host species removed due to low counts of ticks.

This project makes the assumption that co-feeding is the only viable route of transmission, and that co-feeding transmission is passed from 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 is 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 2. This means that larvae found on vertebrates without a nymphal burden of the same species were excluded.

Based on low tick burden counts, we exclude several combinations but will investigate these 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 2:** Counts of nymphs and larvae on vertebrates, but larvae were found to co-aggregate with one or more 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 REF.
- The pathogens transmitted by each tick species are frequently not the same pathogens REF.
- An adult common shrew is larger than the adult field vole REF, which implies the chances of nymphs and larvae co-aggregating in close enough proximity
- Yearly seasonal variation shown in figure 2 suggests that these ticks 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 presented in Table 3.

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 3:** The ranked correlations between nymphs and larvae, obtained by analysing the Kielder Forest data provided by Bown et al.

Assuming confidence of 0.95, then the pvalue  $> 0.05$  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 co-aggregation. Not finding evidence of correlation should not prevent us from doing the analysis.

### 3.2 Tick burden heterogeneity

Some vertebrate hosts attract many more ticks than others, and measuring that heterogeneity has been addressed or estimated in many studies [**<empty citation>**]. Of particular interest is research from Perkins et al in 2003 [42], which quantified heterogeneity in transmission potential for co-feeding transmission by plotting Lorenz curves, and then calculating a Gini coefficient to describe the overall heterogeneity. This project will also use Lorenz curves and Gini coefficients to quantify heterogeneity. However, this paper will compare the Perkins approach to finding Lorenz curves with another approach that uses a modified version of the formula provided by Johnstone-Robertson et al (5).

A limitation of using the Perkins formula (3) 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 (see the section "Finding a candidate offspring distribution" for more on this reasoning). Given that simplifying assumption, the Perkins is unsuitable since it will not reflect that co-feeding transmission is not possible on a vertebrate host when either the larval or nymphal burdens 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 (5) for this.

In the Johnstone-Robertson paper, 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 data of vertebrates at a particular time 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  $k_{out}$ .

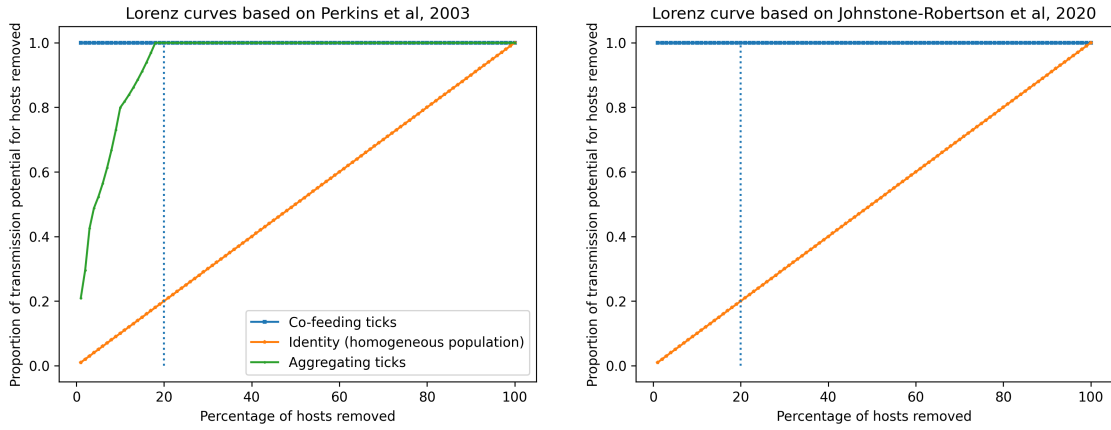
$m$

be the number of vertebrate hosts.

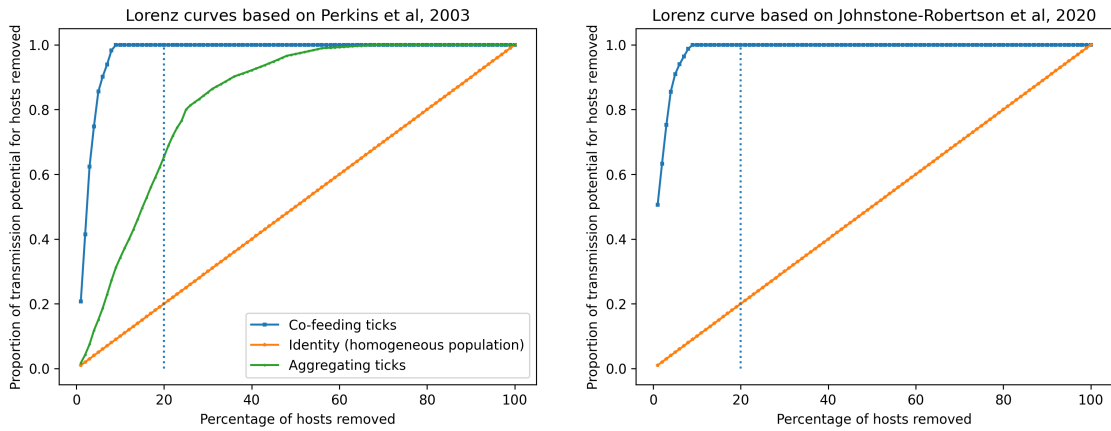
The modified version of (5) is now:

$$\begin{aligned}
R_0 \propto \frac{\langle b_n b_l \rangle}{\langle b_l \rangle} &= \frac{(1/m) \sum_{a=1}^m (b_n b_l)_a}{(1/m) \sum_{a=1}^m (b_n)_a} \\
&= \frac{\sum_{a=1}^m (b_n b_l)_a}{\sum_{a=1}^m (b_n)_a} \\
&= \frac{(b_n b_l)_1}{\sum_{a=1}^m (b_n)_a} + \frac{(b_n b_l)_2}{\sum_{a=1}^m (b_n)_a} + \dots + \frac{(b_n b_l)_m}{\sum_{a=1}^m (b_n)_a}
\end{aligned} \tag{8}$$

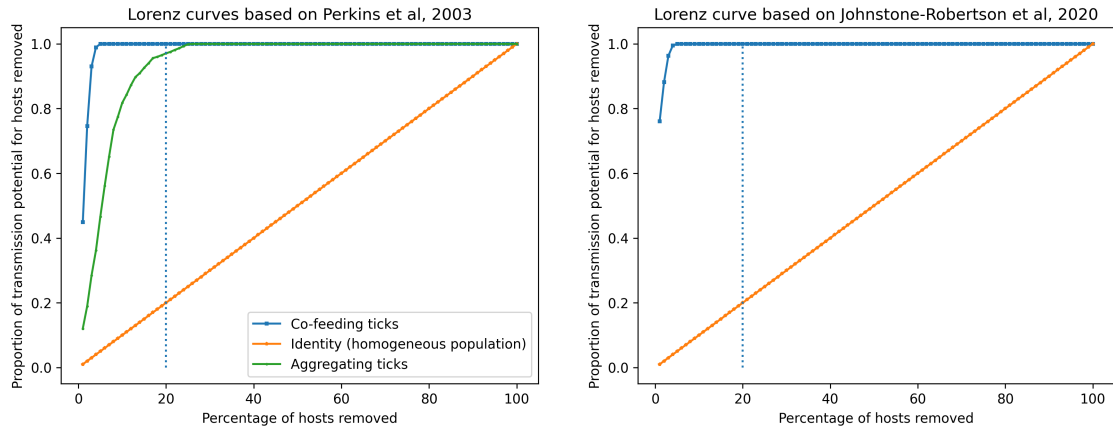
Then, to obtain a result that is similar to that of Perkins et al, we can order the vertebrates by the product  $b_n b_l$  in descending order, with  $a = 1$  denoting the vertebrate with the largest relative transmission potential. Then, we remove the vertebrates that have the most transmission potentials. The cumulative effect of removing subsequently larger whole number percentages of vertebrate hosts is show in figures . In those figures, to determine the relative effect of removing the vertebrates that have the largest transmission potential, we plot the Perkins formula next to the modified version of the Johnstone-Robertson formula. The shape of the Lorenz curve is useful as a visual aide for understanding the inequality in transmission potentials.



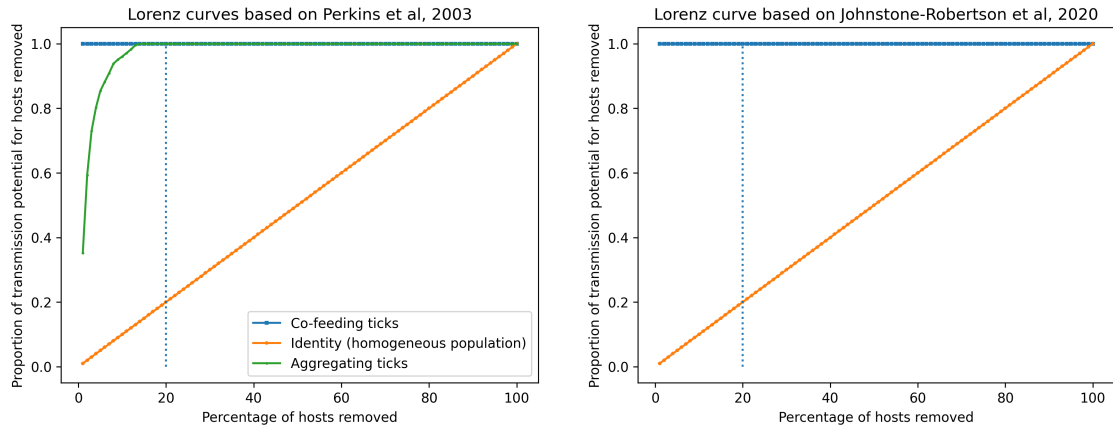
**Figure 3:** Cumulative transmission potential of *common shrews* that were found with *I. ricinus* in 2004 in the Kielder Forest data.



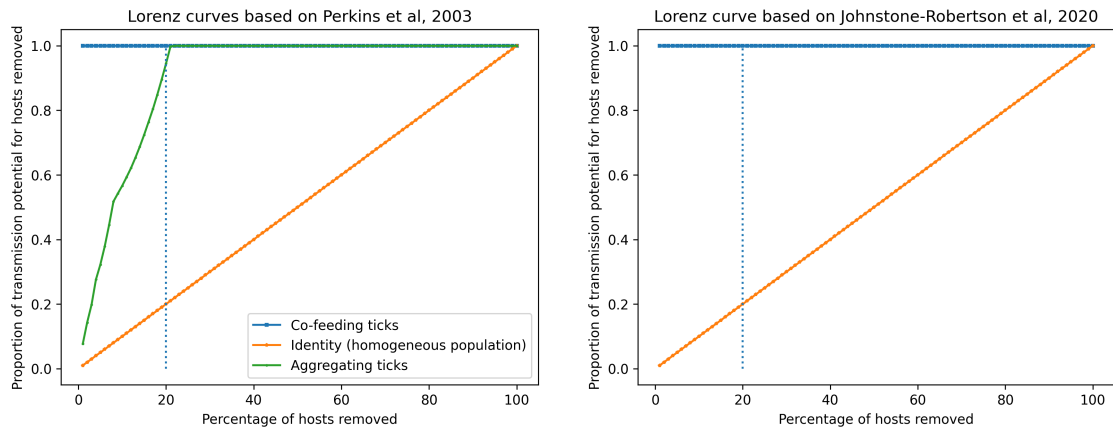
**Figure 4:** Cumulative transmission potential of *common shrews* that were found with *I. trianguliceps* in 2004 in the Kielder Forest data.



**Figure 5:** Cumulative transmission potential of *common shrews* that were found with *I. trianguliceps* in 2005 in the Kielder Forest data.



**Figure 6:** Cumulative transmission potential of *field voles* that were found with *I. ricinus* in 2005 in the Kielder Forest data.



**Figure 7:** Cumulative transmission potential of *field voles* that were found with *I. trianguliceps* in 2005 in the Kielder Forest data.

To calculate the GINI coefficient for each subset of data, we calculate the area between the Lorenz curve and the identity function. Since the identity function represents an idealised homo-

Tick species	Host species	Year	Perkins et al, 2003		Johnstone -Robertson et al 2020
			Aggregation	Co-aggregation	Co-aggregation
<i>I. trianguliceps</i>	<i>Common shrew</i>	2004	0.651	0.942	0.956
<i>I. trianguliceps</i>	<i>Common shrew</i>	2005	0.872	0.968	0.975
<i>I. trianguliceps</i>	<i>Field vole</i>	2005	0.809	0.98	0.98
<i>I. ricinus</i>	<i>Common shrew</i>	2004	0.879	0.98	0.98
<i>I. ricinus</i>	<i>Field vole</i>	2005	0.945	0.98	0.98

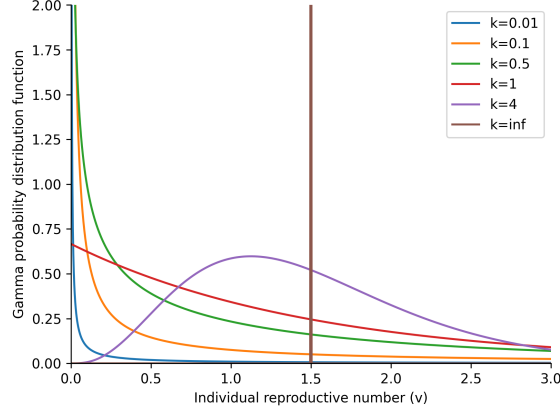
**Table 4:** The GINI coefficient for each of the five subsets of data, for which we have sufficient data, indicates heterogeneity among vertebrates for their co-feeding transmission potential, compared to the just aggregating tick heterogeneity, based on larval burdens collected in the Kielder Forest.

geneous population, the GINI coefficient summarises the total amount of inequality in transmission potentials. For all subsets of data analysed, the GINI coefficient obtained using the modified version of the Johnston-Robertson formula (5) indicates greater heterogeneity than the Perkins formula.

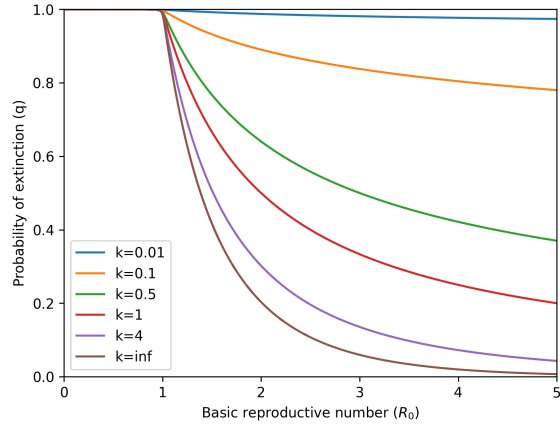
## 4 Reproducing the results by Lloyd-Smith et al, 2005

Before we follow the example set by Lloyd-Smith and others, it will be useful to re-implement parts of their work. In their 2005 paper, Lloyd-Smith and others using a branching process to analyse disease transmissions [37].

Figures 8, 9, 10 are reproduced versions of the charts shared by Lloyd-Smith et al. These illustrate the effect of the individual reproductive number  $v$ . A link to the code to generate these charts is shared in the appendix.



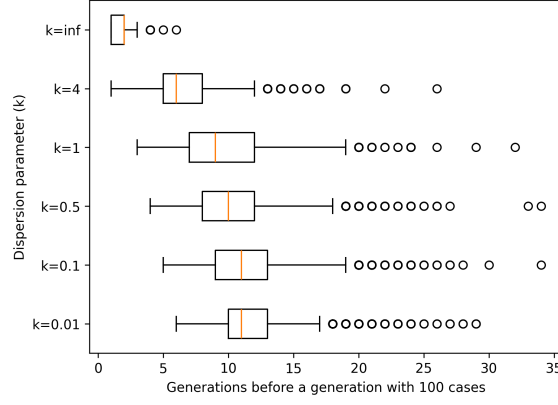
**Figure 8:** The individual reproductive number  $v \sim \text{Gamma}(R_0, \frac{R_0}{k})$ . This chart shows the effect of fixing  $R_0 = 1.5$  and allowing  $k$  to vary.



**Figure 9:** If  $Z \sim \text{Poisson}(v)$ , then the resulting Poisson Mixture is a negative binomial distribution, which is the 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 negative binomial distribution's probability generating function. Each simulation generates a random value for  $v \sim \text{Gamma}(R_0, \frac{R_0}{k})$ , and these are used to generate different

FIX THIS PARAGRAPH Lloyd-Smith et al also showed that if  $v \sim \text{Gamma}(R_0, \frac{R_0}{k})$ , then the resulting offspring distribution is the Negative Binomial distribution.

And, if  $v \sim \text{Gamma}(R_0, \frac{R_0}{k})$  then (2) becomes:



**Figure 10:** Lloyd-Smith et al 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  $R_0 = 1.5$ .

$$g(s) = \left(1 + \frac{R_0}{k}(1 - z)\right)^{-1} \quad (9)$$

The derivation of (9) is presented in (15).

Then, if we wish to calculate the probability that a chain of transmissions becomes extinct at or before generation  $m$ , then we can use (3) as a recurrence relation:

$$z_{m+1} = g(z_m) = \left(1 + \frac{R_0}{k}(1 - z_m)\right)^{-k} \quad (10)$$

$$z = g(z) = \left(1 + \frac{R_0}{k}(1 - z)\right)^{-k} \quad (11)$$

The equation (10) simplifies to (11), because all  $z_m \leq 1$  and  $z_{m+1} \geq z_m$ , meaning the series monotonically increases to a finite value. In other words, it converges.

Then  $z = g(z)$ , which is a function that is equal to its independent variable, which is known as a fixed point. We can approximate the value, via Fixed Point Iteration, to some acceptable measure of error. Then this result we will use later to estimate the probability that a chain of co-feeding transmissions becomes extinct. The code used to recreate the work by Lloyd-Smith and others is available in the appendix at the end of this project.



## 5 Reparameterisation of the negative binomial distribution

While the work by Lloyd-Smith et al, presented above, uses the negative binomial distribution, it remains to be seen that it is the best fit for our tick co-aggregation data. To test it, we first need to derive the same parameterisation that Lloyd-Smith et al 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 negative binomial distribution is presented as:

$$Z \sim NB(x|p, k) = \binom{x+k-1}{x} p^k q^x, x \in \mathbb{N} \quad (12)$$

But, a more convenient parameterisation uses the sample mean  $m$  and dispersion parameter  $k$  as parameters [46]. This is possible with  $p = \frac{k}{m+k}$ :

$$Z \sim NB(x|m, k) = \frac{\Gamma(x+k)}{x!\Gamma(k)} \left(1 + \frac{m}{k}\right)^{-k} \left(\frac{m}{m+k}\right)^x, x \in \mathbb{N} \quad (13)$$

Finding  $R_0$  by analysing contact-trace data also determines  $m$  for this distribution, by setting  $m = R_0$ . Finding the dispersion parameter  $k$  is achievable with maximum likelihood estimation (MLE). Since the parameters have asymptotically orthogonal limits [37], then we can find  $k$  via MLE without doing so for  $m$ , since  $m = R_0$  is known.

In this project, we tested the standard negative binomial distribution, available in Python Scipy, against the reparameterisation (13) by fitting both distributions to the same data. Results indicate a similar fit. A link to the code is available in the appendix.

Below, we derive the probability generating function for the reparameterisation, using (12). Note the definition of a probability generating function given in (6).

$$\begin{aligned} g(z) &= \sum_{x=0}^{\infty} \binom{x+k-1}{x} p^k q^x z^x \\ &= p^k \sum_{x=0}^{\infty} \binom{x+k-1}{x} (qz)^x \\ &= p^k (1 - qz)^{-k} \end{aligned} \quad (14)$$

$$\begin{aligned} &= \left(\frac{k}{m+k}\right)^k \left(1 - \frac{m}{m+k} z\right)^{-k} \\ &= \left(\frac{m}{k}\right) (1 - z)^{-k} \end{aligned} \quad (15)$$

Above, (14) is a result of the negative binomial theorem. The equation on line (15) is the final form of the probability generating function, for use with numerical approximation later in this project.

## 6 The probability that a chain of TBD transmissions becomes extinct

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 systems of equations, and their underlying assumption of homogeneity, 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.

To calculate the probability that a chain of TBD transmissions becomes extinct, we can use a GWBP analysis similar to the example set by Lloyd-Smith and others used in their 2005 paper [37]. To do that, however, we first require a candidate for the offspring distribution.

### 6.1 Finding a candidate offspring distribution

Lloyd-Smith and others estimated  $R_0$  and then found an estimate for  $k$  by using contact-tracing network data, AND COUNTING ZEROS. In our case, we have empirical data for the co-aggregation of larvae and nymphs, but this does not mean that each pair of larva and nymph was a transmission. We could use a stochastic process simulate a transmission network based on the empirical data.

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

- Since a larva will take one blood meal before it moults into a nymph, then the only way for an *Ixodidae* 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 [1].
- Adult ticks tend to feed on larger vertebrates, whereas immature ticks (larvae and nymphs) frequently feed on small vertebrate hosts [16, 1].
- 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 [53].

So, if a TBD was to spread through a population of ticks via co-feeding only, in the absence of transovarial and systemic transmission, then the 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 the transmission to occur.

Then, let us introduce these terms:

- $c$  be the contact rate, or the probability that a larva and infectious nymph co-aggregate close enough for transmission to occur.
- $v$  be the probability of transmission, since individual variation means not every infectious nymph has the same capability to transmit the disease.
- $\sigma$  be the probability that a larvae survives to become a nymph (moulting success).
- $\tau$  be the probability that transstadial transmission occurs.

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

$$\alpha = cv\sigma\tau \tag{16}$$

$\alpha$  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, in their 2020 paper [4], is important for this project. Their superimposition of tick-to-tick transmission onto the tick-host contact network is the inspiration for this idea: we can use co-aggregation data to generate a nymph-larvae contact network for each nymph.

Then let:

$X$

be a vector of the out degree for each nymph's larval contact network.

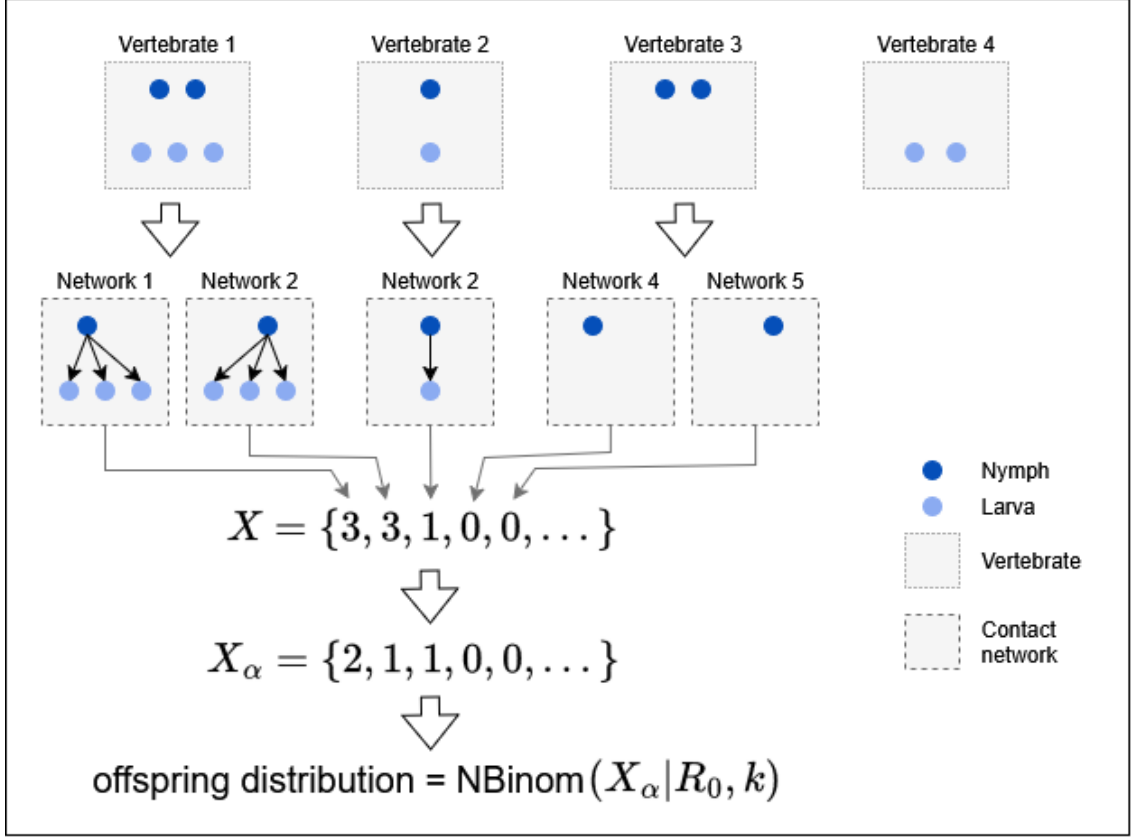
$X_\alpha$

be a vector for how many larvae become infectious nymphs. This is offspring data to which we will fit a distribution, as the offspring distribution. Note that  $X$  and  $X_\alpha$  have the same number of elements.

Conceptually, we could fit a probability distribution to  $X$  and then draw random values from that, then apply the stochastic process using  $\alpha$  to find  $X_\alpha$ , to then fit a probability distribution to it as the offspring distribution. That would be everything that we need to use the GWBP approach to numerically determine the probability that a chain of transmissions becomes extinct. However, we can avoid the stochastic simulation, as shown later in this thesis.

## 6.2 Fitting a distribution to $X_\alpha$

Before proceeding, we can find a probability distribution that is a good fit to the co-aggregation data  $X$ , and then the offspring data  $X_\alpha$ . Several discrete distributions could be the best choice: geometric, Poisson and negative binomial distributions could all fit. The negative binomial distribution is often a good fit to parasite aggregation data, and, Lloyd-Smith and others showed this was the best fit for several disease transmission networks; we can expect that the negative binomial distribution will be the best fit. The following assumes that the reparameterised negative binomial distribution is an appropriate fit to  $X$ . Later, we test its goodness-of-fit against the Poisson and geometric distributions.



**Figure 11:** A toy example of how we find 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 coaggregates with. A stochastic step is applied to the vector  $X$  to determine the number of larvae that become infectious nymphs.  $R_0$  and  $k$  are then found by fitting the reparameterised negative binomial distribution to  $X_\alpha$ .

Let:

$n$  be the number of nymphs in the analysis, also the length of  $X$

$k$  be the dispersion parameter.

$$m = \frac{1}{n} \sum_i^n L_i$$

be the mean of  $X$ .

$$L_i \sim \mathbf{NB}(X|m, k)$$

be the co-aggregation distribution, fit to  $X$ . This is the reparameterised negative binomial distribution (13).  $L_i$  is the number of larvae that co-aggregate with a nymph of index  $i$ .

$$I_{i,j} \sim \mathbf{Bern}(p = \alpha)$$

be a Bernoulli trial, or, a coin flip for each nymph of index  $i$  and for each larvae that it co-aggregates with, indexed by  $j \in 1 : L_i$ . This is a step where an infectious nymph generates a new infectious nymph, which is an offspring in the GWBP.

$$R_0$$

be the basic reproductive number. This is also the mean value of the offspring data  $X_\alpha$

$$Z \sim \mathbf{NB}(X_\alpha | m = R_0, k)$$

be the offspring distribution. This is achieved by simulating offspring data  $X_\alpha$  and then fitting the distribution to it. This will have mean  $R_0$ .

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

$$\begin{aligned} I_i &= \sum_j^{L_i} I_{i,j} \\ I_i &\sim \text{Binom}(L_i, \alpha) \end{aligned} \tag{17}$$

The expected value of the binomial distribution in (17) is  $\mu_i = L_i \alpha$ , 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:

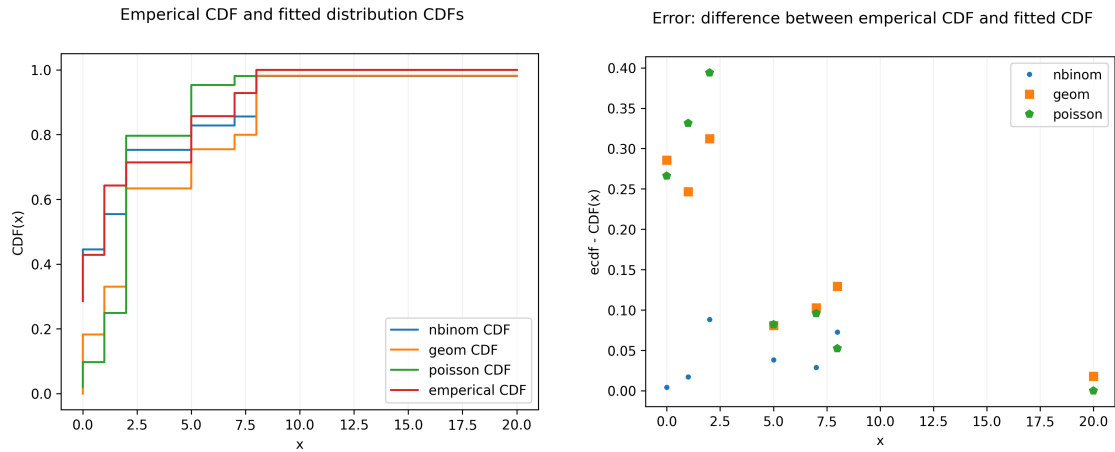
$$\begin{aligned} R_0 &= \frac{1}{n} \sum_i^n \mu_i \\ R_0 &= \frac{1}{n} \sum_i^n L_i \alpha \\ R_0 &= \alpha \frac{1}{n} \sum_i^n L_i \\ R_0 &= \alpha m \end{aligned} \tag{18}$$

In summary, by fitting the co-aggregation distribution to  $X$ , we can find the mean parameter for the  $X_\alpha$  by scaling the  $X$  mean value by  $\alpha$ . This allows us to avoid using computation to determine which larvae become infectious nymphs.

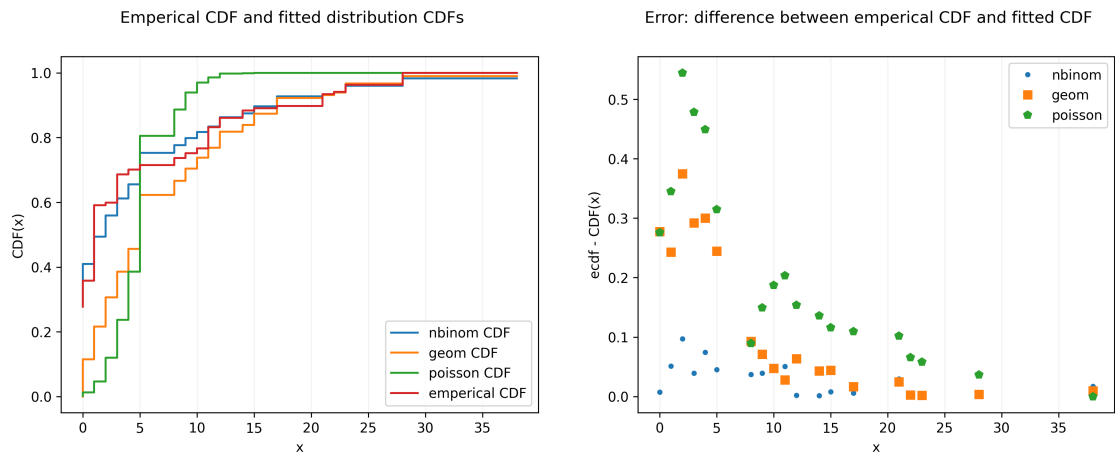
### 6.3 Goodness of fit tests for co-aggregation distribution

Before using the reparameterised negative binomial distribution (13) for determining the probability of extinction with a 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 by using the Akaike Information Criterion (AIC) [37].

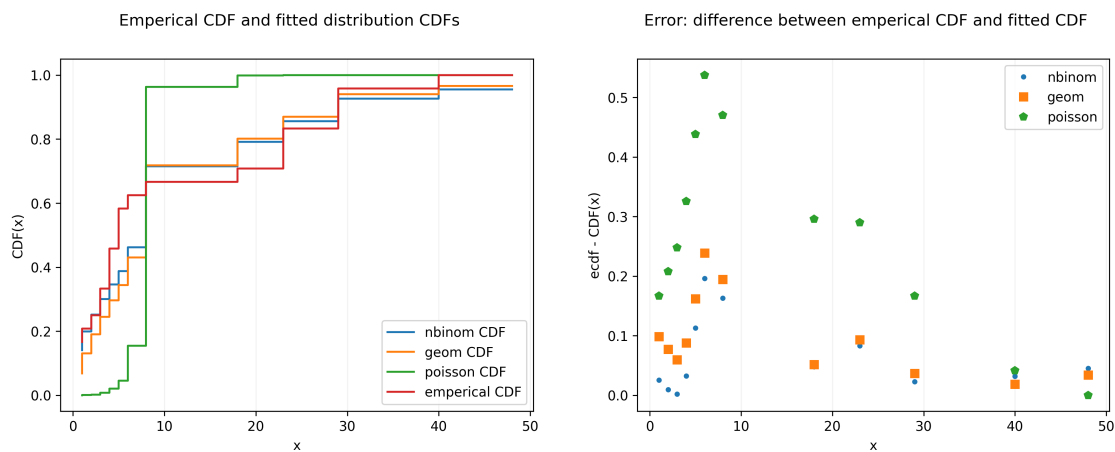
For the visual analysis, we calculated the error of each fitted distribution by finding the difference between its cumulative density function and the data's empirical cumulative density function. We plotted this error for each subset of the Kielder Forest data. These are presented in the results section, in figures: ....



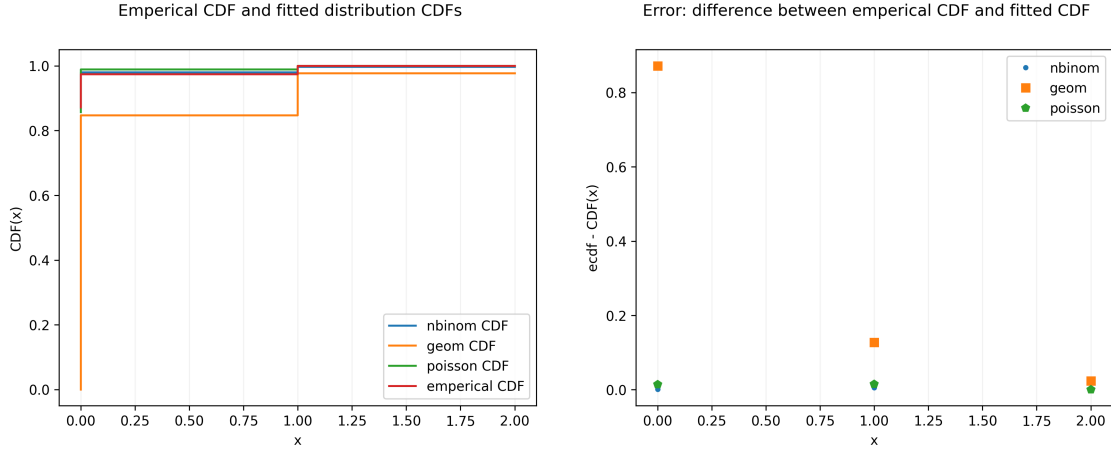
**Figure 12:** Discrete probability distributions fitted to the co-aggregation data  $X$  for *I. ricinus* found on *common shrews* in 2004.



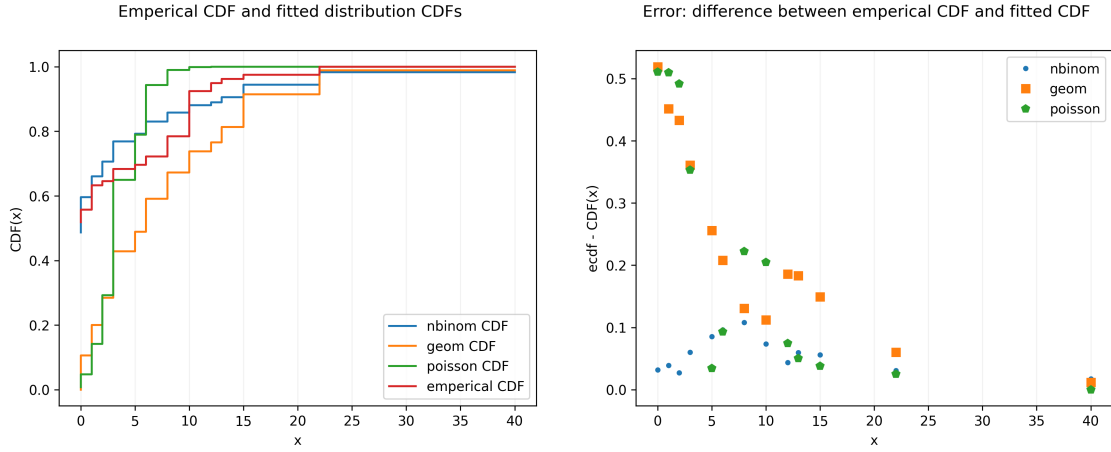
**Figure 13:** Discrete probability distributions fitted to the co-aggregation data  $X$  for *I. trianguliceps* found on *common shrews* in 2004.



**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.

To calculate goodness of fit, for each distribution and for each subset of data, we used AIC as did the Lloyd-Smith et al paper. The AIC method rewards the best-fitting models while penalising models by their number of parameters. In the sense that the negative binomial, Poisson and geometric distributions are models, the negative binomial's fit will be offset by its extra parameter, while the Poisson and geometric distributions each have just one parameter. REF DETAILS ABOUT PARAMS.

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

$$\begin{aligned}
 AIC &= 2K - 2\ln(L(\hat{\theta}|X)) \\
 AIC_c &= AIC + \frac{2K(K+1)}{n-K-1}
 \end{aligned}
 \tag{19}$$

For the Kielder Forest data, these  $AIC_c$  scores are presented in table 5. A point of difference is that Lloyd-Smith and others 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.

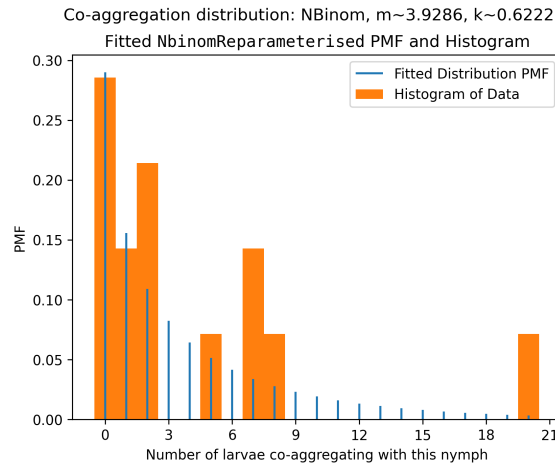
Tick species	Distribution	$AIC_c$
<i>I. ricinus</i> <i>I. trianguliceps</i>	geometric	$\infty$
	Poisson	4955
	negative binomial	2013
<i>I. ricinus</i>	geometric	$\infty$
	Poisson	806
	negative binomial	332
<i>I. trianguliceps</i>	geometric	$\infty$
	Poisson	3544
	negative binomial	1318

**Table 5:** Each subset of Kielder Forest data includes the common shrew and field vole species of vertebrates, and both years of data.

## 6.4 Fitted co-aggregation and offspring distributions

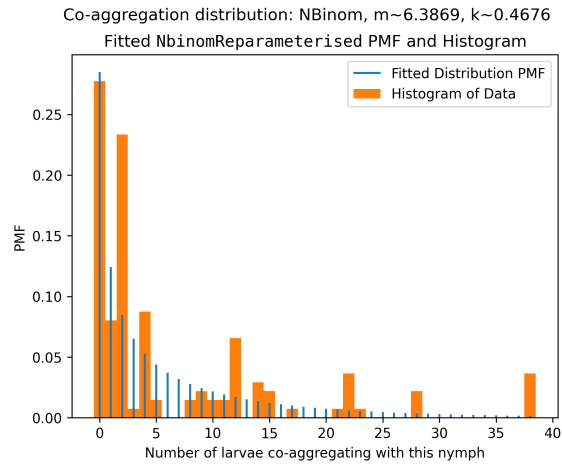
With the negative binomial distribution confirmed as the best-fit for the co-aggregation data,  $X$ , we can now fit the reparameterised negative binomial distribution to subsets of the Kielder Forest data, to obtain the sample mean  $m$  and dispersion parameter  $k$ .

Figures:NUMBERS show the results of MLE applied to the reparameterised negative binomial distribution for each of the 5 subsets of Kielder Forest data we identified in SECTION REF.

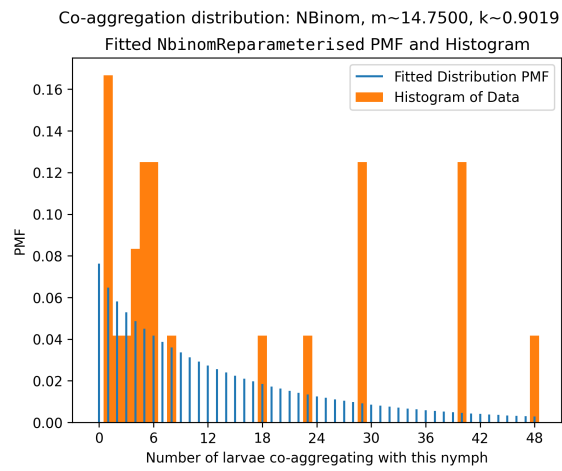


**Figure 17:** For counts of *I. ricinus* found on Common Shrews in 2004

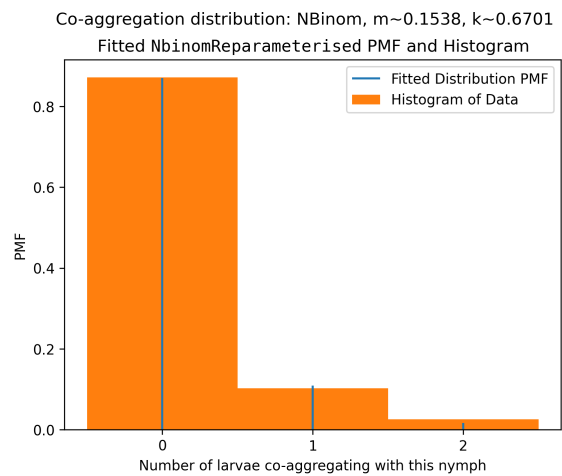




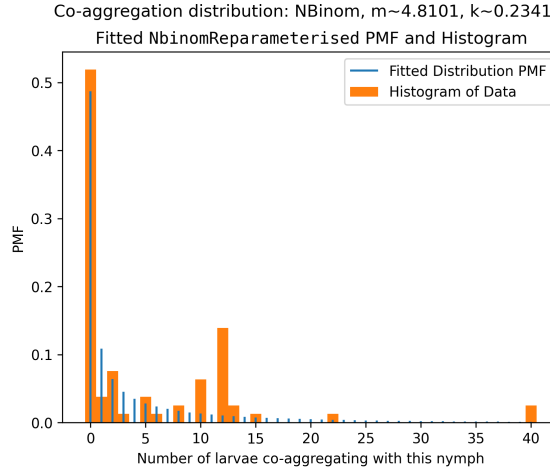
**Figure 18:** For counts of *I.trianguiceps* found on Common Shrews in 2004



**Figure 19:** For counts of *I.ricinus* found on Field Voles in 2005



**Figure 20:** For counts of *I.trianguiceps* found on Field Voles in 2005



**Figure 21:** For counts of *I. trianguliceps* found on Common Shrews in 2005

## 6.5 Estimates for $\alpha$ and the effect on $R_0$

Now that we have estimates for the sample mean of the co-aggregation data, obtained from subsets of the Kielder Forest data, we can obtain a particular estimate for  $R_0$ , based on some  $\alpha$  that we nominate based on (16). These estimates will be illustrative only; researchers who wish to calculate the probability that a new outbreak becomes extinct would do well to provide better estimates about their particular tick species, host species, pathogen and environmental conditions.

Parameter	Estimate	Reason
$c$	0.4	Ticks tend to feed on or near ears GET REF. Most vertebrates have two ears, but some vertebrates will attach away from an ear.
$v$	0.95	Arbitrarily set to be not perfect.
$\sigma$	0.5	Estimates range between 0.1 and 0.9.

**Table 6:**  $\alpha$  estimates for a generic combination of tick species, host species, pathogen and environmental conditions.

Based on the estimates provided in 6 for the analysis presented in the results section,  $\alpha = cv\sigma = 0.19$ . While this estimate is based on some naive assumptions, it is useful for illustration purposes and will be useful for running simulations later. REMOVE THIS IF YOU DON'T GET TO RUN THE SIMULATIONS.

However, we can understand the effect on  $R_0$  even if we do not have a particular

We can also understand the relationship between  $\alpha$  and  $R_0$  with the simple linear equation (18). Since a chain of transmission will certainly become extinct if  $R_0 < 1$ , then we are interested in the case where  $R_0 > 1$ . Other words, what is the minimum value of  $\alpha$  such that a chain of transmissions might persist? This is found with:

I HAVE FORGOTTEN, WHAT HAPPENS WHEN  $R_0 = 1$

$$\begin{aligned}
R_0 > 1 &\implies m\alpha < 1 \\
\alpha &< \frac{1}{m}
\end{aligned} \tag{20}$$

Since  $m$  is known for each subset of Kielder Forest data, then letting  $\alpha$  vary allows us to see the effect on  $R_0$ ; in FIGURE REF

FIGURE OF STRAIGHT LINES INTERSECTING THE ORIGIN, WITH THE GRADIENT OF EACH EQUAL TO SAMPLE MEAN, ALPHA AS INDEPENDENT VARIABLE.

Simply put, the effect of  $\alpha$  is a line that intersects  $R_0 = 0$  at  $\alpha = 0$ , where the gradient is the sample mean of the co-aggregation data,  $X$ .

## 6.6 Estimates for the probability of extinction

Finally, we can use estimates of  $R_0$  and  $k$ , obtained for each subset of Kielder Forest data, to determine the probability that a chain of transmissions that starts with a single infectious tick eventually becomes extinct. We do this using our estimate of  $R_0$  for each subset of Kielder Forest data (tick species, host species, and year) based on the naive estimate of  $\alpha$ . We then do this as we let  $R_0$  vary, since we can not be sure about the value of  $\alpha$ .

By using the negative binomial distribution's probability generating function REF EQUATION, reparameterised to use  $R_0$  and  $k$  for the offspring distribution  $X_\alpha$ , we can use Fixed Point Iteration to approximate the probability of eventual extinction.

TABLE FOR THE PROBABILITIES OF EXTINCTION BASED ON THE NAIVE ESTIMATE OF ALPHA.

We take inspiration from the charts by Lloyd-Smith et al [37], presented in REF FIGURES, and plot the probability of extinction and the simulation where the first generation to reach 100 offspring as an arbitrary benchmark. However, where Lloyd-Smith et al fixed  $R_0$  and let  $k$  vary, we will instead fix  $k$  for each subset of the Kielder Forest data and let  $R_0$  vary, to reflect our uncertainty about the value of  $\alpha$ .

## 7 Discussion

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 *Ixodidae* (Hard Ticks), which feed only once per life stage [**<empty citation>**].

In this project, we have made some assumptions, which vary in the reason for their introduction; some are simplifying assumptions, 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 and alternative assumptions we could have used. Finally, we talk about opportunities for further research based on this project.

The first and most obvious assumption made 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 [**<empty citation>**]. Therefore, it is possible that a ticks carrying a pathogen might move into an area where a naive population of vertebrate hosts exists, and that those vertebrate hosts would be incompetent hosts to maintain a pathogen. Therefore, the role of systemic transmission in that scenario would be less important. The role of trans-ovarial transmission is assumed to be minor, since adult ticks are the least-common life stage to be found in nature [**<empty citation>**]. Assuming that co-feeding transmission is the only viable route of transmission also allows us to use a single-type GWBP, which is effective for understanding the risk of extinction at the start of an outbreak for a directly-transmissible pathogen [**<empty citation>**]. Assuming that co-feeding transmission is the only viable form of transmission will be an insufficient technique for estimating the probability of extinction for a tick-borne disease system where we know systemic transmission is common, but in this scenario, we can interpret the probability of extinction for co-feeding transmission as an upper bound on the overall risk of extinction for all forms of transmission.

The assumption that co-feeding transmission is from infectious nymph to a nearby larvae only is a simplifying assumption. In reality, it is possible that an infected nymph will feed close to other nymphs and that those other nymphs could become infected. This is particularly possible for nidicolous ticks, which tend to WHAT [**<empty citation>**]. If 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. However, adults tend to feed on larger hosts and immature ticks tend to feed on smaller hosts [**<empty citation>**], so the that particular chain of transmissions would likely end there.

We also assume that at most one infectious nymph will co-feed with larvae on each host. This is reasonable for non-nidicolous tick species, which tend to drop off their vertebrate hosts and moult in a wider range of locations [**<empty citation>**]. Nidicolous ticks, however, tend to WHAT [**<empty citation>**]. Therefore, for nidicolous ticks, having two infectious nymphs co-feed on the same vertebrate might increase the amount of pathogens available for infecting the co-feeding larvae. However, if we would continue to use a single-type GWBP, then the co-feeding larvae would need to be divided amongst the infected nymphs.

The division of the Kielder Forest data has implications for how we can interpret the estimates of  $R_0$ ,  $k$  and the probability that a chain of transmissions becomes extinct. Firstly, the tick species under study, *I. ricinus* and *I. trianguliceps*, are assumed to not be vectors for the same pathogens. Therefore, finding one species of larvae co-feeding with the other species of nymph would mean

that transmission could not occur. More problematic is dividing the data by host species as well. We did this to reflect that the probability of transmission for a single species of tick is different depending on which vertebrate host they feed on. The Field Vole is smaller than the Common Shrew, so the probability that ticks co-feed close enough for transmission to occur would depend on the host species on which they feed. However, partitioning the data by host species means the co-aggregation distribution misses information; for example an infectious nymph could infect a larvae would co-feeding on a Common Shrew, and then, that larvae could survive to become a infectious nymph and co-feed with larvae on a Field Vole. Potentially, this could be addressed with a multi-type GWBP. That was outside the scope of this project, but would be interesting future research.

Some work is required by anyone who wishes to use the results of this research. Firstly, 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. One can then fit the reparameterised negative binomial distribution to obtain the dispersion parameter. The real work would be determining a viable estimate of  $\alpha$ , which is the probability that a co-feeding nymph gets infected, survives the moulting process and maintains the pathogen transstadially. This would likely depend on the tick species, pathogen, host species and environmental condition in which the transmission takes place. In this project, we were not able to determine accurate estimates for  $\alpha$ , SO WHAT DID WE DO. In using the method presented in this project to determine  $R_0$  based on  $\alpha$ , one may add terms to  $\alpha$  since it is understood simply as the overall probability that a larva co-feeding with an infectious nymph becomes itself an infectious nymph. There might be factors in determining  $\alpha$  that we did not discuss in this project.

## 8 Conclusion

## 9 References

- [1] S.E. Randolph. ‘Ticks are not Insects: Consequences of Contrasting Vector Biology for Transmission Potential’. In: *Parasitology Today* 14.5 (May 1998), pp. 186–192. ISSN: 0169-4758. DOI: 10.1016/s0169-4758(98)01224-1.
- [2] N. Johnson. ‘An introduction to tick-borne disease’. In: *Ticks*. Elsevier, 2023, pp. 65–74. ISBN: 9780323911481. DOI: 10.1016/b978-0-323-91148-1.00004-6.
- [3] N. Johnson. ‘A brief introduction to ticks’. In: *Ticks*. Elsevier, 2023, pp. 1–7. ISBN: 9780323911481. DOI: 10.1016/b978-0-323-91148-1.00001-0.
- [4] S.P. Johnstone-Robertson, M.A. Diuk-Wasser and S.A. Davis. ‘Incorporating tick feeding behaviour into  $\text{jmm1:math xmlns:mml=}$ ”<http://www.w3.org/1998/Math/MathML>”  $\text{display=}$ ”inline”  $\text{id=}$ ”d1e939”  $\text{altimg=}$ ”si151.svg”  $\text{}$   $\text{R0}$  for tick-borne pathogens’. In: *Theoretical Population Biology* 131 (Feb. 2020), pp. 25–37. ISSN: 0040-5809. DOI: 10.1016/j.tpb.2019.10.004.
- [5] L. Ferreri et al. ‘Pattern of Tick Aggregation on Mice: Larger Than Expected Distribution Tail Enhances the Spread of Tick-Borne Pathogens’. In: *PLoS Computational Biology* 10.11 (Nov. 2014). Ed. by James Lloyd-Smith, e1003931. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003931.
- [6] M.E.J. Woolhouse et al. ‘Heterogeneities in the transmission of infectious agents: Implications for the design of control programs’. In: *Proceedings of the National Academy of Sciences* 94.1 (Jan. 1997), pp. 338–342. ISSN: 1091-6490. DOI: 10.1073/pnas.94.1.338.
- [7] N. Johnson. ‘Climate change and control of ticks and tick-borne diseases’. In: *Ticks*. Elsevier, 2023, pp. 171–175. ISBN: 9780323911481. DOI: 10.1016/b978-0-323-91148-1.00007-1.
- [8] A. Harrison and N.C. Bennett. ‘The importance of the aggregation of ticks on small mammal hosts for the establishment and persistence of tick-borne pathogens: an investigation using the  $\text{R0}$  model’. In: *Parasitology* 139.12 (July 2012), pp. 1605–1613. ISSN: 1469-8161. DOI: 10.1017/s0031182012000893.
- [9] W.L. Nicholson et al. ‘Ticks (Ixodida)’. In: *Medical and Veterinary Entomology*. Elsevier, 2019, pp. 603–672. ISBN: 9780128140437. DOI: 10.1016/b978-0-12-814043-7.00027-3.
- [10] N. Johnson. ‘Blood feeding as a life choice and the multiple functions of tick saliva’. In: *Ticks*. Elsevier, 2023, pp. 45–63. ISBN: 9780323911481. DOI: 10.1016/b978-0-323-91148-1.00011-3.
- [11] L. Vial. ‘Biological and ecological characteristics of soft ticks (Ixodida: Argasidae) and their impact for predicting tick and associated disease distribution’. In: *Parasite* 16.3 (Sept. 2009), pp. 191–202. ISSN: 1776-1042. DOI: 10.1051/parasite/2009163191.
- [12] P. Parola and D. Raoult. ‘Ticks and Tickborne Bacterial Diseases in Humans: An Emerging Infectious Threat’. In: *Clinical Infectious Diseases* 32.6 (Mar. 2001), pp. 897–928. ISSN: 1537-6591. DOI: 10.1086/319347.
- [13] N. Johnson. ‘Tick classification and diversity’. In: *Ticks*. Elsevier, 2023, pp. 9–23. ISBN: 9780323911481. DOI: 10.1016/b978-0-323-91148-1.00013-7.
- [14] J. Gray, O. Kahl and A. Zintl. ‘Pathogens transmitted by *Ixodes ricinus*’. In: *Ticks and Tick-borne Diseases* 15.6 (Nov. 2024), p. 102402. ISSN: 1877-959X. DOI: 10.1016/j.ttbdis.2024.102402.

- [15] J.S. Gray, A. Estrada-Peña and Vial L. ‘Ecology of Nidicolous Ticks’. In: *Biology of Ticks Volume 2*. Biology of Ticks Volume 2 v. 2. OUP USA, 2014, pp. 39–60. ISBN: 9780199744060. URL: <https://books.google.com.au/books?id=gck4AAAAQBAJ>.
- [16] C. Herrmann and L. Gern. ‘Search for blood or water is influenced by *Borrelia burgdorferi* in *Ixodes ricinus*’. In: *Parasites & Vectors* 8.1 (2015), p. 6. ISSN: 1756-3305. DOI: 10.1186/s13071-014-0526-2.
- [17] K.J. Bown et al. ‘Seasonal Dynamics of Anaplasma phagocytophilum in a Rodent-Tick (*Ixodes trianguliceps*) System, United Kingdom’. In: *Emerging Infectious Diseases* 9.1 (Jan. 2003), pp. 63–70. ISSN: 1080-6059. DOI: 10.3201/eid0901.020169.
- [18] K.J. Bown et al. ‘Relative Importance of *Ixodes ricinus* and *Ixodes trianguliceps* as Vectors for *Anaplasma phagocytophilum* and *Babesia microti* in Field Vole (*Microtus agrestis*) Populations’. In: *Applied and Environmental Microbiology* 74.23 (Dec. 2008), pp. 7118–7125. ISSN: 1098-5336. DOI: 10.1128/aem.00625-08.
- [19] I. Rochlin and A. Toledo. ‘Emerging tick-borne pathogens of public health importance: a mini-review’. In: *Journal of Medical Microbiology* 69.6 (June 2020), pp. 781–791. ISSN: 1473-5644. DOI: 10.1099/jmm.0.001206.
- [20] K. Singh et al. ‘Economic impact of predominant ticks and tick-borne diseases on Indian dairy production systems’. In: *Experimental Parasitology* 243 (Dec. 2022), p. 108408. ISSN: 0014-4894. DOI: 10.1016/j.exppara.2022.108408.
- [21] F.M. Kivaria. ‘Estimated direct economic costs associated with tick-borne diseases on cattle in Tanzania’. In: *Tropical Animal Health and Production* 38.4 (May 2006), pp. 291–299. ISSN: 1573-7438. DOI: 10.1007/s11250-006-4181-2.
- [22] D. Otranto, F. Dantas-Torres and M.M. Santos-Silva. ‘*Ixodes ricinus* (Linnaeus, 1758) (Figs. 6769)’. In: *Ticks of Europe and North Africa*. Springer International Publishing, 2017, pp. 189–195. ISBN: 9783319637600. DOI: 10.1007/978-3-319-63760-0\_38.
- [23] M.P. Pfäffle, T.N. Petney and M. Madder. ‘*Ixodes trianguliceps* Birula, 1895 (Figs. 57–59)’. In: *Ticks of Europe and North Africa*. Springer International Publishing, 2017, pp. 167–171. ISBN: 9783319637600. DOI: 10.1007/978-3-319-63760-0\_34.
- [24] K.J. Bown et al. ‘The Common Shrew (*Sorex araneus*): A Neglected Host of Tick-Borne Infections?’ In: *Vector-Borne and Zoonotic Diseases* 11.7 (July 2011), pp. 947–953. ISSN: 1557-7759. DOI: 10.1089/vbz.2010.0185.
- [25] J.M. Medlock et al. ‘Driving forces for changes in geographical distribution of *Ixodes ricinus* ticks in Europe’. In: *Parasites & Vectors* 6.1 (Jan. 2013). ISSN: 1756-3305. DOI: 10.1186/1756-3305-6-1.
- [26] A.A. Alkishe, A. Townsend Peterson and A.M. Samy. ‘Climate change influences on the potential geographic distribution of the disease vector tick *Ixodes ricinus*’. In: *PLOS ONE* 12.12 (Dec. 2017). Ed. by R. Mark Wooten, e0189092. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0189092.
- [27] S. Cunze et al. ‘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*’. In: *Parasitology Research* 121.8 (June 2022), pp. 2241–2252. ISSN: 1432-1955. DOI: 10.1007/s00436-022-07556-x.

- [28] N.A. Hartemink et al. ‘The Basic Reproduction Number for Complex Disease Systems: Defining  $R_0$  for Tick-Borne Infections’. In: *The American Naturalist* 171.6 (June 2008), pp. 743–754. ISSN: 1537-5323. DOI: 10.1086/587530.
- [29] R. Rosà et al. ‘Thresholds for disease persistence in models for tick-borne infections including non-viraemic transmission, extended feeding and tick aggregation’. In: *Journal of Theoretical Biology* 224.3 (Oct. 2003), pp. 359–376. ISSN: 0022-5193. DOI: 10.1016/s0022-5193(03)00173-5.
- [30] R. Norman et al. ‘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*)’. In: *Journal of Mathematical Biology* 48.2 (Feb. 2004), pp. 119–134. ISSN: 1432-1416. DOI: 10.1007/s00285-002-0183-5.
- [31] G. Hurry et al. ‘Infection with *Borrelia afzelii* reduces moulting time of *Ixodes ricinus* ticks’. In: (Jan. 2021). DOI: 10.21203/rs.3.rs-139147/v1.
- [32] H. Dautel, D. Kämmer and O. Kahl. ‘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*. Brill — Wageningen Academic, Nov. 2016, pp. 335–349. ISBN: 9789086868384. DOI: 10.3920/978-90-8686-838-4\_24.
- [33] S.E. Randolph. ‘Density-dependent acquired resistance to ticks in natural hosts, independent of concurrent infection with *Babesia microti*’. In: *Parasitology* 108.4 (May 1994), pp. 413–419. ISSN: 1469-8161. DOI: 10.1017/s003118200007596x.
- [34] O. Diekmann and J.A.P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley, Jan. 2000.
- [35] J. Lim et al. ‘Interpretation of the Basic and Effective Reproduction Number’. In: *Journal of Preventive Medicine and Public Health* 53.6 (Nov. 2020), pp. 405–408. ISSN: 2233-4521. DOI: 10.3961/jpmph.20.288.
- [36] N.P. Jewell and J.A. Lewnard. ‘On the use of the Reproduction Number for SARS-COV-2: Estimation, Misinterpretations and Relationships with other Ecological Measures’. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 185.Supplement<sub>1</sub> (June 2022), S16–S27. ISSN: 1467-985X. DOI: 10.1111/rssa.12860.
- [37] J.O. Lloyd-Smith et al. ‘Superspreading and the effect of individual variation on disease emergence’. In: *Nature* 438.7066 (Nov. 2005), pp. 355–359. ISSN: 1476-4687. DOI: 10.1038/nature04153.
- [38] A.P. Galvani and R.M. May. ‘Dimensions of superspreading’. In: *Nature* 438.7066 (Nov. 2005), pp. 293–295. ISSN: 1476-4687. DOI: 10.1038/438293a.
- [39] *Severe Acute Respiratory Syndrome — Singapore, 2003*. <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5218a1.htm>. Accessed: 2025-05-20. May 2003.
- [40] K.L. VanderWaal and V.O. Ezenwa. ‘Heterogeneity in pathogen transmission: mechanisms and methodology’. In: *Functional Ecology* 30.10 (Mar. 2016). Ed. by Dana Hawley, pp. 1606–1622. ISSN: 1365-2435. DOI: 10.1111/1365-2435.12645.
- [41] J.L. Brunner and R.S. Ostfeld. ‘MULTIPLE CAUSES OF VARIABLE TICK BURDENS ON SMALL-MAMMAL HOSTS’. In: *Ecology* 89.8 (Aug. 2008), pp. 2259–2272. ISSN: 0012-9658. DOI: 10.1890/07-0665.1.



- [42] S.E. Perkins et al. ‘Empirical evidence for key hosts in persistence of a tick-borne disease’. In: *International Journal for Parasitology* 33.9 (Aug. 2003), pp. 909–917. ISSN: 0020-7519. DOI: 10.1016/s0020-7519(03)00128-0.
- [43] T. Garske and C.J. Rhodes. ‘The effect of superspreading on epidemic outbreak size distributions’. In: *Journal of Theoretical Biology* 253.2 (July 2008), pp. 228–237. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2008.02.038.
- [44] F. Brauer. ‘Compartmental Models in Epidemiology’. In: *Mathematical Epidemiology*. Springer Berlin Heidelberg, 2008, pp. 19–79. ISBN: 9783540789116. DOI: 10.1007/978-3-540-78911-6\_2.
- [45] L.J.S. Allen. ‘A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis’. In: *Infectious Disease Modelling* 2.2 (May 2017), pp. 128–142. ISSN: 2468-0427. DOI: 10.1016/j.idm.2017.03.001.
- [46] J.A. Rice. *Mathematical statistics and data analysis*. 3. ed., internat. student ed. Duxbury advanced series. teilweise auch mit CD-ROM, ISBN der CD-ROM: 0-495-10986-X. Belmont, Calif.: Thomson Brooks/Cole, 2007. 60363 pp. ISBN: 9780495118688.
- [47] D.J. Shaw, B.T. Grenfell and A.P. Dobson. ‘Patterns of macroparasite aggregation in wildlife host populations’. In: *Parasitology* 117.6 (Dec. 1998), pp. 597–610. ISSN: 1469-8161. DOI: 10.1017/s0031182098003448.
- [48] D. Bisanzio et al. ‘Modeling the Spread of Vector-Borne Diseases on Bipartite Networks’. In: *PLoS ONE* 5.11 (Nov. 2010). Ed. by Cesar V. Munayco, e13796. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0013796.
- [49] F. Brauer. ‘An Introduction to Networks in Epidemic Modeling’. In: *Mathematical Epidemiology*. Springer Berlin Heidelberg, 2008, pp. 133–146. ISBN: 9783540789116. DOI: 10.1007/978-3-540-78911-6\_4.
- [50] K.B. Athreya and P.E. Ney. ‘The Galton-Watson Process’. In: *Branching Processes*. Springer Berlin Heidelberg, 1972, pp. 1–65. ISBN: 9783642653711. DOI: 10.1007/978-3-642-65371-1\_1.
- [51] L.J.S. Allen. ‘Branching Processes’. In: *Encyclopedia of Theoretical Ecology*. University of California Press, Dec. 2019, pp. 112–119. DOI: 10.1525/9780520951785-022.
- [52] M. Maliyoni et al. ‘A Stochastic Tick-Borne Disease Model: Exploring the Probability of Pathogen Persistence’. In: *Bulletin of Mathematical Biology* 79.9 (July 2017), pp. 1999–2021. ISSN: 1522-9602. DOI: 10.1007/s11538-017-0317-y.
- [53] M.J. Voordouw. ‘Co-feeding transmission in Lyme disease pathogens’. In: *Parasitology* 142.2 (Oct. 2014), pp. 290–302. ISSN: 1469-8161. DOI: 10.1017/s0031182014001486.

## 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).