‘Super’ or just ‘above average’? Supershedders and the transmission of *Escherichia coli* O157:H7 among feedlot cattle.

Simon EF Spencer*, Thomas E Besser†, Rowland Cobbold‡ and Nigel P French§

Abstract

Supershedders have been suggested to be major drivers of transmission of *Escherichia coli* O157:H7 (*E. coli* O157:H7) among cattle in feedlot environments, despite our relatively limited knowledge of the processes that govern periods of high shedding within an individual animal. In this study we attempt a data-driven approach, estimating the key characteristics of high shedding behaviour, including effects on transmission to other animals, directly from a study of natural *E. coli* O157:H7 infection of cattle in a research feedlot, in order to develop an evidence-based definition of supershedding.

In contrast to the hypothesised role of supershedders, we found that high shedding individuals only modestly increased the risk of transmission: individuals shedding over 10^3 cfu/g faeces

*Dept of Statistics, University of Warwick, UK. Email: s.e.f.spencer@warwick.ac.uk
†Dept Veterinary Microbiology and Pathology, Washington State University, Pullman WA 99164
‡School of Veterinary Science, University of Queensland, Gatton QLD 4343
§EpiLab, Infectious Disease Research Centre, Institute of Veterinary Animal and Biomedical Sciences, Massey University, New Zealand

were estimated to pose a risk of transmission only 2.45 times greater than those shedding below that level. The data suggested that shedding above 10^3 cfu/g faeces was the most appropriate definition of supershedding behaviour and under this definition supershedding was surprisingly common, with an estimated prevalence of 31.3% in colonized individuals. We found no evidence that environmental contamination by faeces of shedding cattle contributed to transmission over timescales longer than 3 days and preliminary evidence that higher stocking density increased the risk of transmission.

Key Index words or phrases: Supershedder, *Escherichia coli* infections/veterinary, cattle diseases/epidemiology, Bayesian statistics

1 Introduction

Shiga toxin-producing *Escherichia coli* O157:H7 (*E. coli* O157:H7) continue to present a serious threat to public health in many countries around the world [1–3]. The pathogen is associated with severe haemorrhagic diarrhoea and serious sequelae which can result in loss of life [4, 5]. Cattle are considered the main reservoir host [6, 7] and, since the emergence of *E. coli* O157:H7 in 1982 [8] considerable efforts have been made to reduce the risk of zoonotic transmission via food and environmental pathways, prompting the development of interventions such as vaccination [9, 10]. Infection in cattle has major consequences for international trade in animal products, particularly meat from cattle, and this, combined with the direct public health costs, presents a major burden to the economies of many countries [11].

Mathematical models have been used to inform the design and impact of interventions [12–14], understand transmission dynamics in dairy [15, 16] and beef [17, 18] production systems, and determine the key processes that drive infection dynamics [16, 19]. To date, much of the modelling has been carried out by constructing and simulating from mechanistic models [16]

and, fitting these models to experimental infection data [20] or cross sectional data on prevalence and concentration of *E. coli* O157:H7 in cattle faeces [19].

A number of studies have focused on the role and importance of supershedders in transmission and the maintenance of infection in cattle [1, 19]. Although the precise definition varies, supershedders are usually defined as animals shedding relatively large concentrations of *E. coli* O157:H7 in their faeces, generally between 10^3 and 10^4 cfu/g, for an extended period [21, 22]. They are considered to be responsible for a disproportionate amount of transmission. For example, using data from Scotland [23], high shedders, defined as $\geq 3 \times 10^3$ cfu per gram of faeces were estimated to constitute 8% of the cattle but 99% of the bacteria shed. It has been suggested that these high shedders represent those animals that are colonized at the recto-anal junction (RAJ), whereas the subset of low shedders are not colonized, and in some way represent a distinct process, whereby non-colonized animals amplify *E. coli* O157:H7 elsewhere in the gut lumen and shed more briefly [1]. However, some important questions remain: is there robust evidence for the presence of a distinct, identifiable sub-population of cattle that are colonized and shed at high levels for a prolonged period and, if so, how important are these animals for ongoing transmission and maintenance of this pathogen in groups of animals?

In this study, we present a formal consideration of the relationship between the concentration of *E. coli* O157:H7 shed in faeces and transmission, using the results from a large-scale study of natural infection in feedlot cattle [22] and fitting Bayesian SIS models that consider in detail the relationship between shedding and R_0, and the characteristics of the tests used to detect shedding cattle. We derive the relationship between shedding and transmission and provide new insight into the role played by supershedders in the maintenance of infection in cattle populations. Finally we consider an appropriate definition of 'supershedding'.

2 Methods

2.1 Experimental data collection

The data originate from a study of naturally occurring infection in feedlot cattle, described elsewhere [22]. In brief, 20 pens of 8 steers were sampled approximately twice weekly over a 13 week grain feeding period. Animals were held in small research pens located on a working commercial feedlot with over 10,000 animals on the premises. The pens were non-adjacent, so that animals could only come into physical contact with their pen-mates. On each sampling date, samples consisted of a Recto-Anal Mucosal Swab (RAMS) sample [24] and a sample of freshly passed manure. Each sample was tested by PCR for the presence of *E. coli* O157:H7 and the shedding level measured by bacteriologic cultures as described [22]. The 12 pens in the North of the facility measured 6×17 m (102 m^2), whilst the 8 South pens measured 6×37 m (222 m^2).

2.2 Transmission model

A discrete time Susceptible-Infected-Susceptible (SIS) transmission model (Figure 1) was used to model the spread of infection through a pen. In this framework each individual is assumed to belong to one of two states for each day in the study: either susceptible or colonized (infected). Colonized individuals were assumed to shed sufficient quantities of bacteria to infect other individuals in the pen, possibly indirectly via the environment, and so being colonized was assumed to be the same as being infectious. Likewise we did not allow for any potential immunity, and so being uncolonized was assumed to be the same as being susceptible to infection. We denote the colonization status of individual i in pen p on day t by $X_{p,i,t}$ (for $p = 1, \ldots, 20$, $i = 1, \ldots, 8$, $t = 1, \ldots, 99$), where $X_{p,i,t} = 1$ if the individual was colonized and $X_{p,i,t} = 0$ otherwise.

Susceptible individuals were assumed to be at risk from colonization from both unmeasured sources outside of the pen (such as contaminated feed, wildlife etc.) and other colonized individuals within the pen. We assumed that the probability a susceptible individual avoids colonized from sources outside of the pen on any given day was $e^{-\alpha}$: the probability of observing no events in 1 time unit of a Poisson process with rate α. This parameterization therefore allows α to be interpreted as the colonization rate from outside of the pen.

The main aim of this study was to quantify the relationship between the risk of colonization and shedding. We assume the probability that an individual in pen p which is susceptible on day t avoids colonization on day $t+1$ from individual j is given by $\exp\{-\beta f_{p,j,t}(\boldsymbol{X}_{p,j,1:t})\}$, where the function $f_{p,j,t}$ characterises the total force of infection from material shed between days 1 and t by individual j in pen p and $\boldsymbol{X}_{p,i,1:t}$ denotes the vector $(X_{p,i,1}, \ldots, X_{p,i,t})$. Therefore the probability that an individual susceptible on day t avoids colonization on day $t+1$ is given by

$$\mathrm{P}(X_{p,i,t+1} = 0 | X_{p,i,t} = 0) = \exp\left\{-\alpha - \beta \sum_{j=1}^{8} f_{p,j,t}(\boldsymbol{X}_{p,j,1:t})\right\}.$$

These assumptions yield a non-Markovian model that has the potential to allow infectious material to accumulate in the environment and cause colonizations later in the study. Note that we do not distinguish between different transmission pathways, so that all within-pen colonizations (such as direct contact between animals, or indirect transmission via the pen floor, the water trough etc.) are captured by the parameter β. We assumed that the rate of colonization between the study pens was negligible. The North pens and the South pens were of significantly different sizes and so we allowed for different transmission rates in these two environments, denoted by β_N and β_S respectively.

Once colonized, individuals were assumed to remain so for a number of days that followed a negative binomial distribution with shape parameter r and mean duration μ. More precisely,

for $r > 0$, $\mu > 1$ and $k = 1, 2, \ldots,$

$$\begin{aligned} &\mathrm{P}(X_{p,i,(t+1):(t+k)} = \mathbf{1}_k, X_{p,i,t+k+1} = 0 | X_{p,i,t} = 0, X_{p,i,t+1} = 1) \\ &= \frac{\Gamma(r+k-1)}{(k-1)!\Gamma(r)} \left(\frac{r}{r+\mu-1}\right)^r \left(\frac{\mu-1}{r+\mu-1}\right)^{k-1}, \end{aligned}$$

where $\mathbf{1}_k$ is a vector of k ones. This non-standard parameterization ensures that the minimum duration of colonization is one day, the mean duration is μ and the variance is $\mu - 1 + \frac{(\mu-1)^2}{r}$. Finally, we assumed that at the start of the study each animal was colonized independently with probability π.

2.3 Modelling the RAMS and faecal tests

Very little is known about the internal processes that govern RAJ colonization and the resulting shedding behaviour. Many existing studies have been cross-sectional and are therefore uninformative regarding the duration and intensity of shedding during a typical colonization episode. In this study we circumvented the need to model such processes internal to the animal by splitting the observed data (the RAMS counts and faecal counts) into two parts. First we used exponential smoothing to bridge the gaps between the tests and to combine the RAMS counts with the faecal counts. Where there were only negative tests we inserted the detection limit of the tests so that the output can be interpreted as the estimated shedding level over time conditional on the animal being colonized. Secondly, in an entirely separate inferential step, we fitted the SIS transmission model described in Section 2.2 to the binary outcomes of the RAMS and faecal tests (i.e. simply whether or not *E. coli* O157:H7 was detected).

Since each putative *E. coli* O157:H7 isolate was confirmed by multiple phenotypic and genetic markers [22], we assumed that both the RAMS and the faecal tests had 100% specificity. We estimated both test sensitivities separately, with parameters θ_R for the RAMS test and θ_F

for the faecal test. Note that these sensitivities represent the sensitivity of the whole sample acquisition and testing procedure and not simply the laboratory benchmark sensitivity, which could reasonably be assumed to be the same for both tests. We also assumed that the tests were conditionally independent given the colonization status of the animal. Under these assumptions if either the RAMS or the faecal test were positive then the animal must have been colonized. If neither of the tests were positive, the animal may still have been colonized, but neither of the tests was sufficiently sensitive to detect it.

2.4 Estimating shedding levels

Exponential smoothing was used to estimate the shedding level on each day for each animal, conditional on the animal being colonized. First we examined only the cases where both the RAMS and the faecal samples were positive. The mean for RAMS was 4.589 $\log_{10}$ cfu per gram and the mean for the faecal samples was 3.306, 1.283 lower. This difference is likely to be from the dilution of the bacteria from the recto-anal junction colonization site by faeces. Therefore for the purposes of combining these two counts into an overall shedding level we subtract 1.283 from each of the RAMS counts to bring them onto the same scale as the faecal counts.

The shedding level of individual i in pen p on day t was given by

$$S_{p,i,t} = \sum_{k \in T_{i,p}} w_k(t) o_k,$$

where $T_{i,p}$ is the index set for all of the RAMS and faecal tests for individual i in pen p; o_k is the kth observation and $w_k(t)$ is the weight of observation k.

When only one of the two tests was positive then the negative test was given a weight of zero. When neither test was positive, then since we wish to interpret the count as the shedding level given that the animal is positive, we included a single observation with the detection limit of

the RAMS sample, which equated to 0.64 $\log_{10}$ cfu per gram faeces. The weight function $w_k(t)$ was calculated using

$$w_k(t) \propto \exp(-\lambda|t - t_k|),$$

where t_k is the day sample o_k was taken. The weights were normalised to sum to 1. Finally, λ was chosen to be $\log(2)$ so that the influence of each observation halved with each additional day moved away from it.

2.5 Relating shedding to colonization

The principal aim of this study was to characterise the relationship between episodes of high shedding and the risk of colonization. We assumed that individuals that shed above a certain quantity contribute a relative increase in risk.

$$f_{p,j,t}(\boldsymbol{X}_{p,j,1:t}) = \begin{cases} \rho X_{p,j,t} & S_{p,j,t} \geq \tau, \\ X_{p,j,t} & S_{p,j,t} < \tau. \end{cases}$$

The threshold τ can be either chosen to match a prespecified definition of a supershedder or derived directly from the data. We considered the following predefined values for τ: 2, 2.717, 3, 3.333, 3.477 and 4 $\log_{10}$ cfu/g for the faecal test (see Table 2).

2.6 Extended model: environmental accumulation and decay

We extended the model in 2.5 so that infectious material shed earlier in the study can accumulate in the environment and retain some risk of infection. It was assumed that the risk reduced

by a fixed proportion δ each day, resulting in a geometric decay in risk.

$$f_{p,j,t}(\boldsymbol{X}_{p,j,1:t}) = \sum_{u=1}^{t} \delta^{t-u} \rho^{\mathbb{1}_{\{S_{p,j,u}>\tau\}}} X_{p,j,u}.$$

2.7 Model fitting

The models described in section 2.5 and 2.6 were fitted in a Bayesian framework using Markov chain Monte Carlo (MCMC). In all cases prior distributions were chosen to be relatively uninformative. Parameters representing probabilities (θ_R, θ_F and π) were given Beta$(1,1)$ priors and updated using Gibbs steps. Parameters on finite domains, such as δ and τ, were given uniform priors. The remaining parameters (see Table 1) with domains on $[0,\infty)$ were given Gamma$(1,1)$ priors and updated in blocks using multivariate normal Metropolis-Hastings proposals. Appropriate proposal covariance matrices were estimated from pilot runs. The hidden colonization status $X_{p,i,t}$ was updated using an adaptation of the method of O'Neill and Roberts [25], described in SI.

3 Results

3.1 Raw data

E. coli O157:H7 was isolated in all 20 study pens, with an average prevalence of 13.5% in North pens and 11.2% in South pens. The RAMS tests (473 positives) were found to be more sensitive than the faecal samples (283 positives). Figure 2 (left plots) shows the timecourses of positive tests for pens 7 and 8, which contain 4 of the 5 supershedders as defined in Cobbold *et al.* [22]. Plots for the remaining pens are given in SI. Figure 2 (right plots) show the level of shedding for the two individuals highlighted in the left plots. To ensure comparability between RAMS

and faecal levels, the RAMS levels have been adjusted to the same scale as the faecal tests in order to account for the dilution of the bacteria with faeces.

Using their definition of a supershedder (an individual with mean shedding level greater than 4 $\log_{10}$ cfu per positive RAMS and at least 4 consecutive positive RAMS), Cobbold *et al.* [22] identified 5 supershedding animals contained in 3 pens. In this study we reconsidered the concept of identifying individual animals as supershedders and instead focus on supershedding behaviour, which we define to be periods of time in which shedding exceeds a certain threshold. We considered both a list of predefined thresholds for supershedding (see Table 2) and a threshold estimated from the data.

3.2 Smoothed shedding levels

The first stage of our analysis was to estimate the shedding level of each animal for each day in the study, given that it was colonized, via exponential smoothing. The smoothed shedding levels for two animals are given in Figure 2. When both the RAMS and faecal tests were negative, we assumed that the animal was shedding at the detection limit of the RAMS given that it was colonized. Plots of the smoothed shedding levels for every animal are given in SI.

3.3 Natural history parameters

The primary focus for this study was in developing an appropriate model to relate the risk of colonization to the shedding level. Nonetheless fitting an explicit transmission model allowed us to estimate several biologically interesting parameters relating to the natural history of *E. coli* O157:H7 colonization and the study procedures. These parameters appear in all of the models and since their inferred values were broadly unaltered by the choice of supershedding threshold (τ), we summarise the results here from the model described in 2.5, in which the

value of τ was inferred from the data. Posterior summaries and parameter interpretations are given in Table 1.

We estimated the posterior median colonization period to be 9.6 days, with a posterior median shape parameter of 1.5. We also estimated that 9.8% of the animals were colonized at the start of the study. The posterior median test sensitivity of the RAMS was estimated to be 78%, and the faecal test 47%.

We inferred the posterior probability of colonization for every animal in every day of the study. Animal 6 in pen 8 is given as a typical example in Figure 3(a), with the RAMS and faecal tests superimposed at 1 if positive and at 0 if negative. Full details of these posterior distributions are given in the supplement.

3.4 Colonization parameters

The posterior median rate of external transmission was 0.009 which implies each pen receives one external infectious contact on average every 14.5 days. In the North pens the posterior median of the within-pen colonization rate was 0.011, which implies that a single colonized individual within the pen exerts approximately the same infectious pressure as all of the external sources of infection. In the South pens the within-pen colonization rate was roughly half that of the North pens, recalling that the South pens were approximately twice the area of the North pens. The posterior median of the ratio β_N/β_S was 2.41 (95% CI 1.31–5.70).

Figure 3(c) shows the posterior distribution of the supershedding threshold τ. Recall that individuals shedding at levels above τ pose a relative risk ρ times that of individuals shedding below τ. The posterior distribution of ρ is given in Figure 3(d). The joint posterior distribution of ρ and τ (given in SI) reveals that when $\tau > 4$ the shape of the posterior follows the shape of the prior distribution and so the data were only weakly informative about the risk posed by

individuals shedding above this level.

Clearly the supershedding threshold τ and the relative risk ρ are highly dependent parameters, and so to better illustrate their relationship we have calculated the posterior expected reproduction function (Figure 3(b)). The reproduction function is an extension of the basic reproduction number R_0 stratified as a function of the level of shedding. We define the reproduction function to be the expected number of secondary colonizations that would result from a single colonized individual shedding at a fixed level throughout their colonization period, in an otherwise uncolonized pen. This quantity encapsulates all of the epidemic parameters except for the external transmission rate, and has the property that where the reproduction function is greater than one the level of transmission within the pen is sufficient to result in a sustained outbreak without any external contributions.

3.5 Fixed supershedding threshold

We wished to explore further the effect that different definitions of supershedding would have on the risk posed by supershedders. We refitted the model with supershedding threshold fixed at a range of predefined values.

When τ was chosen to be 10^3 cfu/g faeces we found that the posterior median relative risk was 2.449 for supershedding cattle. Results for other choices of τ are given in Table 2. The 95% credible interval for the relative risk did not include 1 under this definition of supershedding, showing significant evidence of an increase in risk. The remaining definitions did not show significant evidence for an increase.

3.6 Extended model: environmental accumulation and decay

Finally we consider an extended model described in 2.6 in which faecal material accumulates in the environment once it has been shed, before decaying away. This allows individuals to become colonized from animals that were shedding earlier in the study but are not shedding contemporaneously. Figure 4 displays the posterior distribution for the half-life of risk attributed to faecal material shed into the environment. The posterior median half-life for the risk of colonization was 0.45 days (95% CI: 0.15–1.68). The majority of the posterior distribution lies shorter than 2 days, which is less than the shortest interval between tests. Therefore there is no evidence from these data that faecal material shed into the environment posed a risk of colonization to other animals in the pen for substantial amounts of time after it had been shed. Posterior distributions for the remaining parameters (given in SI) are remarkably similar to the ones for the model without environmental accumulation.

4 Discussion

The three key findings from this study are that (i) supershedding individuals appear to pose approximately double the risk of transmission resulting in colonization, compared to low shedding individuals; (ii) a data-driven estimate for the definition of a supershedding threshold was 3 $\log_{10}$ cfu/g faeces and (iii) there was no evidence of environmental transmission occurring over timescales longer than 2 days.

Finding (i) contrasts sharply with the prevailing view expressed in the literature on supershedders of *E. coli* O157:H7, that since high shedding animals are expelling several orders of magnitude more bacteria, then their risk of transmission must be raised by a similar amount [19]. Our finding appeared to be robust to whether the threshold for supershedding was estimated from the data (Figure 3(d)) or whether it was predetermined (Table 2). Even low dose ex-

posures have been shown to lead to colonizations in challenge studies [26, 27], and this might explain why the risk of colonization was not strongly related to the number of bacteria shed. It has been suggested that interventions targeting high shedding individuals would be disproportionally effective [19, 22] but our results suggest that such interventions would not reduce within-herd transmission by as much as anticipated.

Our results do suggest a potential intervention to reduce transmission within pens. The South pens were 2.18 times the area of the North pens, and the transmission rate 2.41 times lower in the South pens. This might suggest reducing animal density as a means to reduce within-pen transmission in the feedlot environment. It was notable that in the South pens the reproduction functions were generally below one, implying that infection from outside the pen was necessary for the infection to persist in the pen. The animal densities in these study pens approximate those of high density (North pens) and low density (South pens) commercial feedlot pens, so the range of stocking density here is directly relevant to those in use in the industry [28]. Furthermore, the use of increased stocking density during summer months as a management tool to reduce feedlot dust emissions [29, 30] roughly correlates with the higher average *E. coli* O157:H7 faecal prevalence in feedlot cattle during that period [31]. Our study was not designed to investigate pen area as a risk factor for transmission, however our data is supported by a previous study in which it was observed that cattle density within feedlot pens correlated with the prevalence of faecal *E. coli* O157:H7 shedding [32]. On the other hand, Renter *et al.* did not detect an effect of confinement on the prevalence of faecal shedding of *E. coli* O157:H7 by cattle when comparing feedlot and pastured cattle, after controlling for the younger age of the feedlot cattle [33]. It may be that there are different drivers of transmission in pasture environments, such as infection from wild animals and increased contact between cattle around water sources.

Cobbold *et al.* [22] defined a supershedder to be an individual with a mean RAMS greater than 4 $\log_{10}$ cfu/swab with at least 4 consecutive RAMS tests. Although this definition characterises

individuals that are colonized at the recto-anal junction (RAJ) with high levels of bacteria and for a sustained period, there are two important drawbacks with this as a general definition of supershedding. Firstly, it is difficult to generalize to cross-sectional studies or longitudinal studies with a different sampling frequency because of the requirement for 4 consecutive positive tests. Secondly, because the average RAMS level is taken over all positive tests in the study it will regress towards the mean shedding level as greater numbers of positives appear. Consequently this definition actually favours individuals with fewer positives. There is clear evidence of this effect in Figure 2, where 3 of the 4 supershedders identified by Cobbold *et al.* (labelled on the left hand side with "SS") have close to the minimum number of positive RAMS tests – just 4 consecutive ones. The remaining supershedder identified in the study also had 4 positive RAMS tests (not shown).

These considerations motivate the need for a more generalisable definition of supershedding behaviour. There remains little evidence from longitudinal studies on how long supershedding behaviour persists in individual animals. It is certainly not well established that individuals identified to be shedding at high levels will again shed at high levels after a period free from colonization. Cattle infected by *E. coli* O157:H7 respond with specific mucosal and systemic immune responses, and these responses may contribute to clearance of infection [34, 35]. However, cattle have been shown to be susceptible to recolonization shortly after clearance in both observational and experimental studies, albeit often accompanied by faecal shedding reduced in level and duration [36, 37]. Therefore the process through which individual animals clear colonization remains undefined. One hypothesis might be that supershedding is the indirect result of interactions between different components of the RAJ microbiome that permits temporary or persistent dominance by *E. coli* O157:H7. Likewise, the process of decolonization may be the result of changes in the competing bacterial flora at the RAJ, perhaps acting in concert with RAJ mucosal immune responses.

These considerations lead us to suggest moving away from the idea that individual animals

can be categorized as supershedders, but instead that supershedding behaviour is a transient property that can appear and disappear over time. We propose that supershedding behaviour be defined by achieving a single shedding level above a certain threshold. Ideally the shedding level would be measured using the RAMS test as we found it to be more sensitive than the faecal test and it is presumably more specific to RAJ colonization. However since faecal sampling is more widespread and far easier to undertake, we have described our results in terms of this test. All that remains is to decide on an appropriate threshold.

A variety of different thresholds for supershedding of *E. coli* O157:H7 in cattle have been used in the literature, including 10^4 cfu/RAMS [22], 10^4 cfu/g faeces [38–40], 10^3 cfu/g faeces [41] and 3×10^3 cfu/g faeces [12]. In this study we have converted these into cfu/g faeces, based on an average swab weight of 0.242g and a faecal dilution factor of 5.21% ($10^{-1.283}$) when comparing RAMS and faecal tests. These conversions are summarised in Table 2. By fitting a model in which the supershedding threshold τ could be estimated from the data, this study has been the first to provide a data-driven definition of supershedding. The posterior median supershedding threshold was 3.2 $\log_{10}$ cfu/g faeces however the posterior mode was closer to 3. This suggests that 10^3 cfu/g faeces might be the most appropriate threshold for supershedding. This definition also produced the highest relative risk posed by supershedders (Table 2). However note that there is considerable uncertainty in the posterior distribution of the supershedding threshold (Figure 3(c)), and so this finding will be worth re-evaluating as more evidence becomes available.

The raw prevalence, based on samples in which at least one of the two tests was positive, was 12.6%. The posterior expected prevalence, which takes into account colonized animals that failed to test positive was 14.6%. Using 10^3 cfu/g faeces as the threshold for supershedding behaviour we find a prevalence of supershedding behaviour of 4.6% and hence the prevalence of supershedding behaviour during colonization was 31.3%. The abundance of supershedding behaviour adds further weight to the idea that supershedding may not be particularly exceptional or influential to transmission and is therefore somewhat misnamed.

In Section 3.6 an extended model was fitted which had the ability to capture the accumulation and removal of bacteria in the environment as the study progressed. In particular, the risk of infection from faeces continued beyond the day on which it was shed. This model estimated that the half-life for the risk of colonization from faeces (Figure 4) was 0.45 days. This is much smaller than existing survival estimates in experimental settings, such as survival in soils [42,43] and on farmyard material surfaces [44]. We found no significant risk of colonization beyond 2 or 3 days after shedding, which coincides with the shortest time interval between tests in our study. Consequently there was not the temporal resolution in our dataset to determine whether the risk of infection was spread across a few days or concentrated around the time that shedding occurs. However the data appear to contain no evidence of longer-term risk of transmission. For reasons of parsimony it therefore appears most sensible to use a Markovian transmission model, in which the risk of transmission occurs only on the day of shedding, because such models require fewer parameters.

If we link together the findings that the risk of transmission is not strongly linked to the level of shedding, but it is related to pen area, then this could support the hypothesis that ingestion of faecal material from the pen floor is the prevailing risk-factor for within pen transmission. In particular the density of recent faecal pats on the pen floor (irrespective of the concentration of bacteria within the faeces) may be worth further investigation as a driver of transmission within the feedlot environment.

The two diagnostic tests produced very different posterior median sensitivity estimates: 0.78 (95% CI 0.73–0.82) for the RAMS and 0.46 (95% CI 0.42–0.51) for the faecal test. Note that these sensitivities encompass the whole sampling procedure (sample capture, transportation and lab work); and not just the laboratory benchmark sensitivity, which could reasonably be assumed to be the same for both tests. The faecal test sensitivity is likely to be lower due to the dilution of the bacteria with faeces, which coincides with most of the existing literature [24,45,46], but not all [47–49]. It is unsurprising that there is some variation between

studies as the sensitivity is likely to depend strongly on the exact sampling protocol and study design.

An obvious weakness with the models discussed in this paper is the assumption that the sensitivity of the RAMS and faecal tests does not depend on the shedding level of the animal. *A priori* we would expect that high shedding individuals would be easier to detect than individuals shedding at levels close to the detection limit of the tests and this should be addressed in future modelling work. A further weakness is that the model does not account for any uncertainty in the shedding levels, which were estimated via exponential smoothing in a separate inferential step. Ideally the shedding level would be estimated concurrently with the colonization status, but to do so would require the model to describe the way that shedding levels evolve through time within an individual animal. At present there is not enough available data or understanding of the biological processes involved to parametrise such a model.

External validity. The data used in this study come from research pens of 8 cattle which, due to their small size, may not entirely reflect the dynamics of transmission within the much larger pens used in commercial feedlots, although pen densities are likely to be comparable. However, the research pens were located outside on a fully operational commercial feedlot and, except for the sampling procedures, the treatment was the same as for the full-scale commercial setting. The costs of performing such intensive sampling on commercial scale pens would be prohibitive and so having sufficient data to reconstruct the transmission dynamics must necessarily come at a cost to external validity. A recent review [50] highlighted only one longitudinal study with larger pen sizes [38] with an average pen size of 31.9, however the sampling frequency was once per month, which would make reconstructing the transmission process between tests extremely challenging from these data. Our observations regarding the effects of supershedder thresholds and animal density on transmission rates are plausible but should be applied tentatively to other cattle production systems or to other host-agent systems for which supershedders are proposed to play a role.

To our knowledge this study has been the first to employ a data-driven approach to characterise supershedding behaviour. Although our findings were based on data from only one study on a much smaller scale than a commercial feedlot, our modelling approach could readily be applied to more comprehensive longitudinal datasets as they emerge. In contrast to previous efforts to model the effects of supershedders, we report here little support for dramatically higher effects of these animals in driving transmission. In addition, our data give little indication that individual cattle persistently exhibit supershedder behaviour but rather, are consistent with periodic or episodic supershedding among many or most colonized cattle.

Author contributions

S.E.F.S. designed the model, carried out the statistical analyses and drafted the manuscript; T.E.B., R.C. and N.P.F. provided input on contextual interpretation and practical impacts of findings and drafted the manuscript. R.C. provided the data. All authors gave final approval for publication.

Funding statement

This work was funded by The Beef Checkoff, with support from National Institutes of Health Public Health Service grants U54-AI-57141, P20-RR16454, and P20-RR15587 and from the US Department of Agriculture National Institute of Food and Agriculture grant 2010-04487.

References

[1] Chase-Topping M, Gally D, Low C, Matthews L, Woolhouse M. 2008 Super-shedding and the link between human infection and livestock carriage of *Escherichia coli* O157. *Nat. Rev. Microbiol.* **6**, 904–912.

[2] Karmali MA, Gannon V, Sargeant JM. 2010 Verocytotoxin-producing *Escherichia coli* (VTEC). *Vet. Microbiol.* **140**, 360–370.

[3] Mead PS, Griffin PM. 1998 *Escherichia coli* O157:H7. *The Lancet* **352**, 1207–1212.

[4] Karch H, Tarr PI, Bielaszewska M. 2005 Enterohaemorrhagic *Escherichia coli* in human medicine. *Int. J. Med. Microbiol.* **295**, 405–418.

[5] Karmali MA. 2004 Infection by Shiga toxin-producing *Escherichia coli*. *Mol. Biotechnol.* **26**, 117–122.

[6] Ferens WA, Hovde CJ. 2011 *Escherichia coli* O157:H7: animal reservoir and sources of human infection. *Foodborne Pathog. Dis.* **8**, 465–487.

[7] Wells J, Shipman L, Greene K, Sowers E, Green J, Cameron D, Downes F, Martin M, Griffin P, Ostroff S. 1991 Isolation of *Escherichia coli* serotype O157:H7 and other Shiga-like-toxin-producing *E. coli* from dairy cattle. *J. Clin. Microbiol.* **29**, 985–989.

[8] Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, *et al.* 1983 Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* **308**, 681–685.

[9] Snedeker K, Campbell M, Sargeant J. 2012 A systematic review of vaccinations to reduce the shedding of *Escherichia coli* O157 in the faeces of domestic ruminants. *Zoonoses Public Hlth* **59**, 126–138.

[10] Varela N, Dick P, Wilson J. 2013 Assessing the existing information on the efficacy of bovine vaccination against *Escherichia coli* O157:H7 – a systematic review and meta-analysis. *Zoonoses Public Hlth* **60**, 253–268.

[11] Scharff RL. 2012 Economic burden from health losses due to foodborne illness in the United States. *J. Food Prot.* **75**, 123–131.

[12] Matthews L, Reeve R, Gally DL, Low JC, Woolhouse ME, McAteer SP, Locking ME, Chase-Topping ME, Haydon DT, Allison LJ, *et al.* 2013 Predicting the public health benefit of vaccinating cattle against *Escherichia coli* O157. *Proc. Natl. Acad. Sci. U.S.A* **110**, 16265–16270.

[13] Vosough Ahmadi B, Frankena K, Turner J, Velthuis AG, Hogeveen H, Huirne RB. 2007 Effectiveness of simulated interventions in reducing the estimated prevalence of *E. coli* O157:H7 in lactating cows in dairy herds. *Vet. Res.* **38**, 755–771.

[14] Vogstad A, Moxley R, Erickson G, Klopfenstein T, Smith D. 2014 Stochastic simulation model comparing distributions of STEC O157 faecal shedding prevalence between cattle vaccinated with type III secreted protein vaccines and non-vaccinated cattle. *Zoonoses Public Hlth* **61**, 283–289.

[15] Turner J, Bowers RG, Begon M, Robinson SE, French NP. 2006 A semi-stochastic model of the transmission of *Escherichia coli* O157 in a typical UK dairy herd: Dynamics, sensitivity analysis and intervention/prevention strategies. *J. Theor. Biol.* **241**, 806–822.

[16] Turner J, Bowers R, Clancy D, Behnke M, Christley R. 2008 A network model of *E. coli* O157 transmission within a typical UK dairy herd: The effect of heterogeneity and clustering on the prevalence of infection. *J. Theor. Biol.* **254**, 45–54.

[17] Liu WC, Jenkins C, Shaw D, Matthews L, Pearce M, Low J, Gunn G, Smith H, Frankel G, Woolhouse M. 2005 Modelling the epidemiology of Verocytotoxin-producing *Escherichia coli* serogroups in young calves. *Epidemiol. Infect.* **133**, 449–458.

[18] Ayscue P, Lanzas C, Ivanek R, Gröhn Y. 2009 Modeling on-farm *Escherichia coli* O157:H7 population dynamics. *Foodborne Pathog. Dis.* **6**, 461–470.

[19] Matthews L, Low J, Gally D, Pearce M, Mellor D, Heesterbeek J, Chase-Topping M, Naylor S, Shaw D, Reid S, *et al.* 2006 Heterogeneous shedding of *Escherichia coli* O157 in cattle and its implications for control. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 547–552.

[20] Tildesley MJ, Gally DL, McNeilly TN, Low JC, Mahajan A, Savill NJ. 2011 Insights into mucosal innate responses to *Escherichia coli* O157:H7 colonization of cattle by mathematical modelling of excretion dynamics. *J. R. Soc. Interface* **9**, 518–527.

[21] Omisakin F, MacRae M, Ogden I, Strachan N. 2003 Concentration and prevalence of *Escherichia coli* O157 in cattle feces at slaughter. *Appl. Environ. Microbiol.* **69**, 2444–2447.

[22] Cobbold RN, Hancock DD, Rice DH, Berg J, Stilborn R, Hovde CJ, Besser TE. 2007 Rectoanal junction colonization of feedlot cattle by *Escherichia coli* O157:H7 and its association with supershedders and excretion dynamics. *Appl. Environ. Microbiol.* **73**, 1563–1568.

[23] Chase-Topping ME, McKendrick IJ, Pearce MC, MacDonald P, Matthews L, Halliday J, Allison L, Fenlon D, Low JC, Gunn G, *et al.* 2007 Risk factors for the presence of high-level shedders of *Escherichia coli* O157 on Scottish farms. *J. Clin. Microbiol.* **45**, 1594–1603.

[24] Rice DH, Sheng HQ, Wynia SA, Hovde CJ. 2003 Rectoanal mucosal swab culture is more sensitive than fecal culture and distinguishes *Escherichia coli* O157:H7-colonized cattle and those transiently shedding the same organism. *J. Clin. Microbiol.* **41**, 4924–4929.

[25] O'Neill PD, Roberts GO. 1999 Bayesian inference for partially observed stochastic epidemics. *J. R. Stat. Soc. A* **162**, 121–129.

[26] Besser T, Richards B, Rice D, Hancock D. 2001 *Escherichia coli* O157:H7 infection of calves: infectious dose and direct contact transmission. *Epidemiol. Infect.* **127**, 555–560.

[27] Cornick N, Helgerson A. 2004 Transmission and infectious dose of *Escherichia coli* O157:H7 in swine. *Appl. Environ. Microbiol.* **70**, 5331–5335.

[28] Animal Health Australia. 2013. Cattle standards and guidelines – beef feedlots. discussion paper. http://www.animalwelfarestandards.net.au/cattle/consultative-process.

[29] Auvermann B, Romanillos A. 2000 Manipulating stocking density to manage fugitive dust emission from cattle feedyards. In Vigil M, Pieper L, editors, *Proc. of the Innovative Technologies for Planning Animal Feeding Operations*, pages 61–70. Akron, Colo.: USDA-ARS Central Great Plains Research Station.

[30] Sweeten JM. 1998 Cattle feedlot manure and wastewater management practices. In Hatfield J, Stewart B, editors, *Animal waste utilization: Effective use of manure as a soil resource*, pages 125–155. Ann Arbor Press Chelsea, MI.

[31] Hancock D, Besser T, Lejeune J, Davis M, Rice D. 2001 The control of VTEC in the animal reservoir. *Int. J. Food Microbiol.* **66**, 71–78.

[32] Vidovic S, Korber DR. 2006 Prevalence of *Escherichia coli* O157 in Saskatchewan cattle: characterization of isolates by using random amplified polymorphic DNA PCR, antibiotic resistance profiles, and pathogenicity determinants. *Appl. Environ. Microbiol.* **72**, 4347–4355.

[33] Renter DG, Sargeant JM, Hungerford LL. 2004 Distribution of *Escherichia coli* O157:H7 within and among cattle operations in pasture-based agricultural areas. *Am. J. Vet. Res.* **65**, 1367–1376.

[34] Nart P, Naylor SW, Huntley JF, McKendrick IJ, Gally DL, Low JC. 2008 Responses of cattle to gastrointestinal colonization by *Escherichia coli* O157:H7. *Infect. Immun.* **76**, 5366–5372.

[35] Bretschneider G, Berberov E, Moxley RA. 2007 Isotype-specific antibody responses against *Escherichia coli* O157:H7 locus of enterocyte effacement proteins in adult beef cattle following experimental infection. *Vet. Immunol. Immunopathol.* **118**, 229–238.

[36] Sanderson M, Besser T, Gay J, Gay C, Hancock D. 1999 Fecal *Escherichia coli* O157:H7 shedding patterns of orally inoculated calves. *Vet. Microbiol.* **69**, 199–205.

[37] Besser TE, Hancock DD, Pritchett LC, McRae EM, Rice DH, Tarr PI. 1997 Duration of detection of fecal excretion of *Escherichia coli* O157:H7 in cattle. *J. Infect. Dis.* **175**, 726–729.

[38] Arthur TM, Keen JE, Bosilevac JM, Brichta-Harhay DM, Kalchayanand N, Shackelford SD, Wheeler TL, Nou X, Koohmaraie M. 2009 Longitudinal study of *Escherichia coli* O157:H7 in a beef cattle feedlot and role of high-level shedders in hide contamination. *Appl. Environ. Microbiol.* **75**, 6515–6523.

[39] Hallewell J, Niu Y, Munns K, McAllister T, Johnson R, Ackermann H, Thomas J, Stanford K. 2014 Populations of endemic bacteriophage differ between cattle shedding high and low numbers of *Escherichia coli* O157:H7 in feces. *Appl. Environ. Microbiol.* **80**, 3819–3825.

[40] Cernicchiaro N, Pearl DL, McEwen SA, Zerby HN, Fluharty FL, Loerch SC, Kauffman MD, Bard JL, LeJeune JT. 2010 A randomized controlled trial to assess the impact of dietary energy sources, feed supplements, and the presence of super-shedders on the detection of *Escherichia coli* O157:H7 in feedlot cattle using different diagnostic procedures. *Foodborne Pathog. Dis.* **7**, 1071–1081.

[41] Low JC, McKendrick IJ, McKechnie C, Fenlon D, Naylor SW, Currie C, Smith DG, Allison L, Gally DL. 2005 Rectal carriage of enterohemorrhagic *Escherichia coli* O157 in slaughtered cattle. *Appl. Environ. Microbiol.* **71**, 93–97.

[42] Berry ED, Wells JE. 2012 Soil solarization reduces *Escherichia coli* O157:H7 and total *Escherichia coli* on cattle feedlot pen surfaces. *J. Food Prot.* **75**, 7–13.

[43] Ma J, Ibekwe AM, Yi X, Wang H, Yamazaki A, Crowley DE, Yang CH. 2011 Persistence of *Escherichia coli* O157:H7 and its mutants in soils. *PloS One* **6**, e23191.

[44] Williams A, Avery L, Killham K, Jones D. 2005 Persistence of *Escherichia coli* O157 on farm surfaces under different environmental conditions. *J. Appl. Microbiol.* **98**, 1075–1083.

[45] Davis MA, Rice DH, Sheng H, Hancock DD, Besser TE, Cobbold R, Hovde CJ. 2006 Comparison of cultures from rectoanal-junction mucosal swabs and feces for detection of *Escherichia coli* O157 in dairy heifers. *Appl. Environ. Microbiol.* **72**, 3766–3770.

[46] Greenquist M, Drouillard J, Sargeant J, Depenbusch B, Shi X, Lechtenberg K, Nagaraja T. 2005 Comparison of rectoanal mucosal swab cultures and fecal cultures for determining prevalence of *Escherichia coli* O157:H7 in feedlot cattle. *Appl. Environ. Microbiol.* **71**, 6431–6433.

[47] Cernicchiaro N, Pearl DL, McEwen SA, LeJeune JT. 2011 Assessment of diagnostic tools for identifying cattle shedding and super-shedding *Escherichia coli* O157:H7 in a longitudinal study of naturally infected feedlot steers in Ohio. *Foodborne Pathog. Dis.* **8**, 239–248.

[48] Khaitsa M, Bauer M, Gibbs P, Lardy G, Doetkott D, Kegode R. 2005 Comparison of two sampling methods for *Escherichia coli* O157:H7 detection in feedlot cattle. *J. Food Prot.* **68**, 1724–1728.

[49] Williams K, Ward M, Dhungyel O, Van Breda L. 2014 Relative sensitivity of *Escherichia coli* O157 detection from bovine feces and rectoanal mucosal swabs. *J. Food Prot.* **77**, 972–976.

[50] Munns KD, Selinger LB, Stanford K, Guan L, Callaway TR, McAllister TA. 2015 Perspectives on super-shedding of *Escherichia coli* O157:H7 by cattle. *Foodborne Pathog. Dis.* **12**, 89–103.

Parameter	Posterior median		Interpretation
α	0.009	(0.006, 0.012)	External colonization rate (days^{-1})
β_N	0.011	(0.006, 0.016)	Within-pen colonization rate, North (days^{-1})
β_S	0.004	(0.002, 0.009)	Within-pen colonization rate, South (days^{-1})
τ	3.199	(2.319, 5.170)	Supershedding threshold ($\log_{10}$ cfu/g faeces)
ρ	1.910	(0.111, 4.093)	Supershedding relative risk
μ	9.578	(8.170, 11.04)	Mean colonization duration (days)
κ	1.524	(0.910, 2.518)	Shape parameter of colonization period
π	0.098	(0.056, 0.154)	Probability colonized initially
θ_R	0.776	(0.733, 0.817)	Sensitivity of RAMS
θ_F	0.464	(0.423, 0.507)	Sensitivity of faecal test

Table 1: Interpretation and posterior medians of the model parameters, for the model in which the supershedding threshold τ is inferred from the data (see Section 2.5). Values in parenthesis indicate 95% credible intervals.

Supershedding definition	Supershedding threshold ($\log_{10}$ cfu/g faeces)	Posterior median relative risk	
10^2 cfu/g faeces	2	1.209	(0.527, 2.653)
10^4 cfu/g RAMS	2.717	1.843	(0.852, 4.034)
10^3 cfu/g faeces [41]	3	2.449	(1.232, 4.618)
10^4 cfu/swab RAMS [22]	3.333	1.930	(0.766, 3.742)
3×10^3 cfu/g faeces [12]	3.477	2.221	(0.793, 4.654)
10^4 cfu/g faeces [38–40]	4	1.622	(0.142, 4.286)

Table 2: Posterior median relative risk for selected supershedding thresholds converted to $\log_{10}$ cfu/g faeces. Values in parenthesis indicate 95% credible intervals.

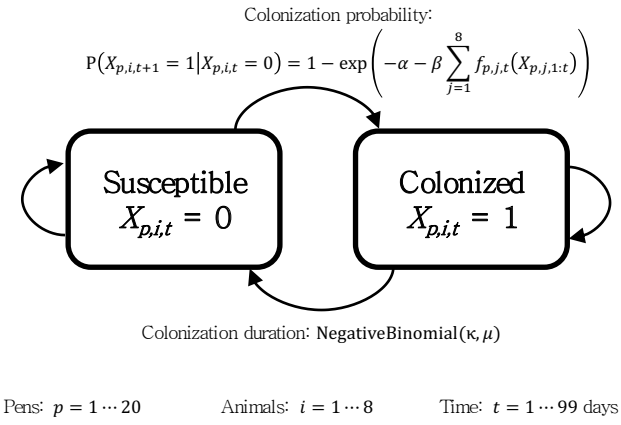

Figure 1: Illustration of the transmission model.

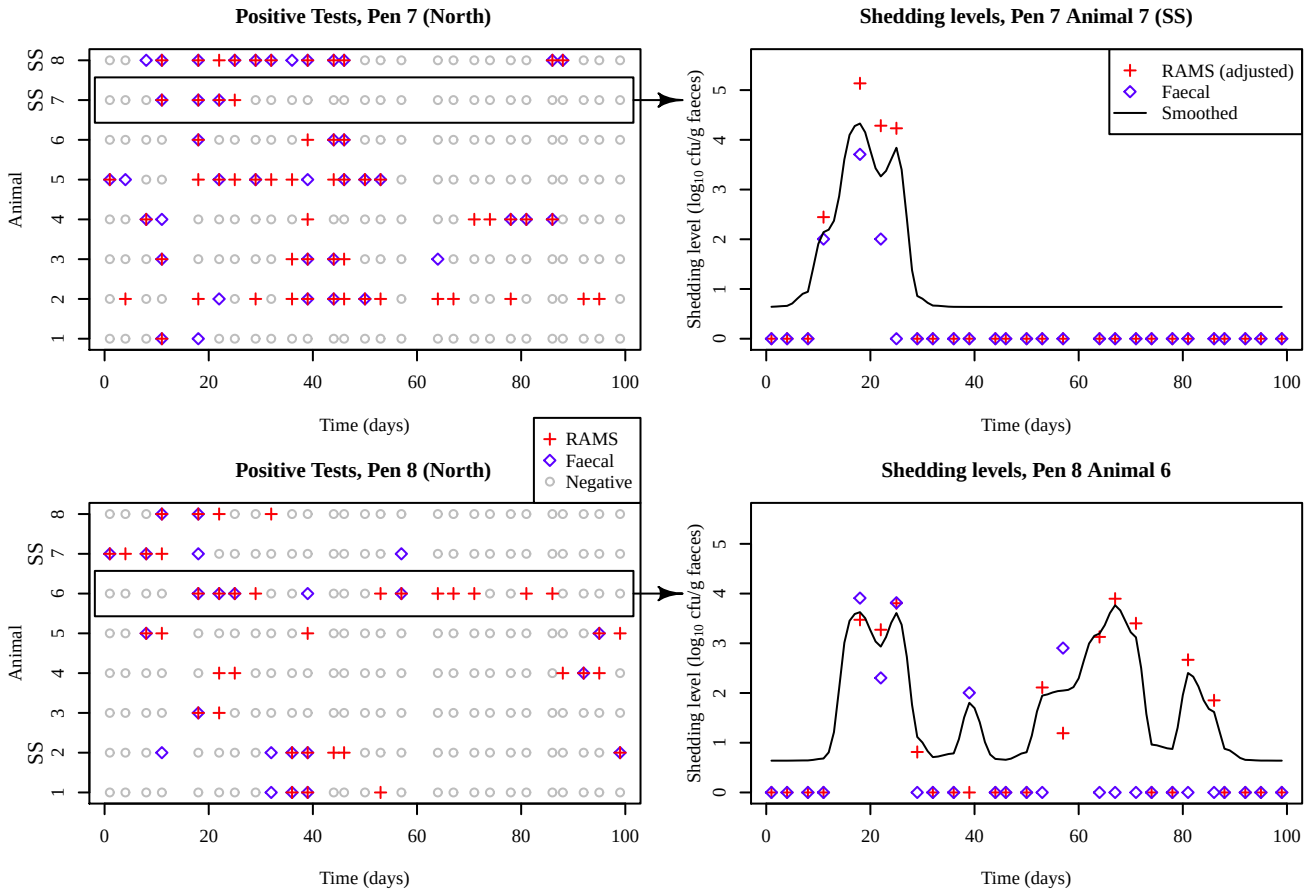

Figure 2: *E. coli* O157:H7 positive tests against time (left plots) for pens 7 and 8 with measured shedding levels and exponentially smoothed shedding levels (right plots) for two animals. The letters "SS" indicate that the animal was termed a supershedder under the definition of [22]. Complete plots given in SI.

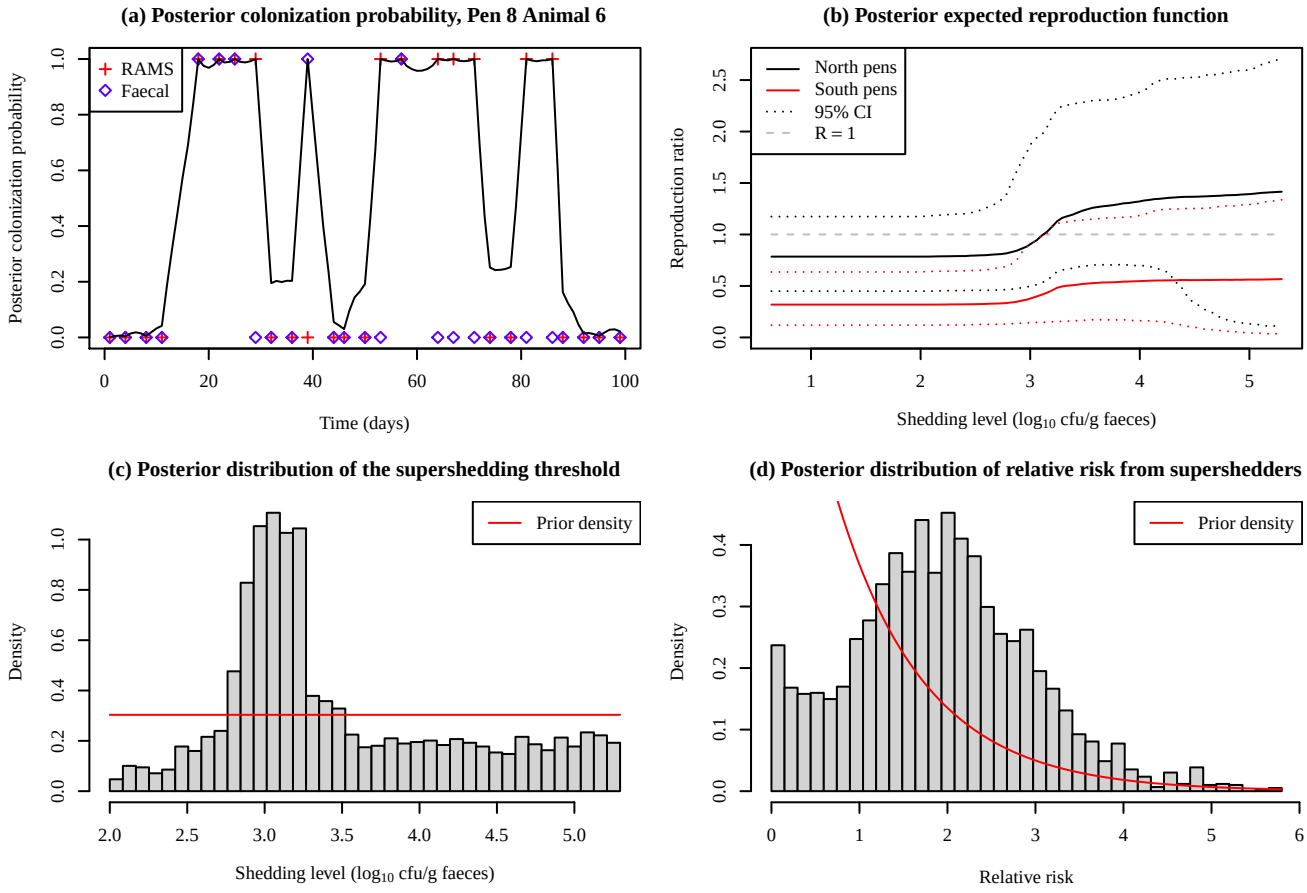

Figure 3: (a) Posterior colonization probability with positive and negative tests marked at 1 and 0 respectively. (b) Posterior expected reproduction function against shedding level with 95% credible intervals as dotted lines. (c) Posterior and prior distributions of the supershedding threshold τ. (d) Posterior and prior distributions for the relative risk ρ posed by individuals shedding above the supershedding threshold τ.

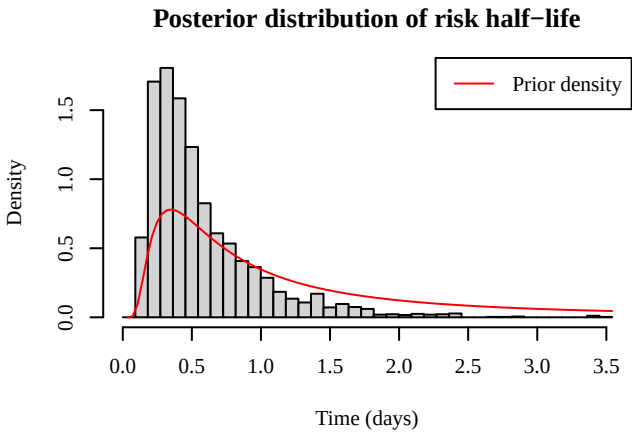

Figure 4: Posterior and prior distributions for the half-life of the risk from faecal material shed into the environment.