

Modelling the epidemiology of Verocytotoxin-producing *Escherichia coli* serogroups in young calves

W.-C. LIU^{1*}, C. JENKINS², D. J. SHAW¹, L. MATTHEWS¹, M. C. PEARCE³,
J. C. LOW⁴, G. J. GUNN³, H. R. SMITH⁵, G. FRANKEL⁶ AND M. E. J. WOOLHOUSE¹

¹ Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian, Scotland, UK

² Microbiology, Royal Free Hospital, London, UK

³ Epidemiology Unit, Scottish Agricultural College Veterinary Science Division, Drummondhill, Inverness, Scotland, UK

⁴ Scottish Agriculture College, Penicuik, Midlothian, Scotland, UK

⁵ Laboratory of Enteric Pathogens, Specialist and Reference Microbiology Division, Health Protection Agency, London, UK

⁶ Centre for Molecular Microbiology and Infection, Department of Biological Sciences, Imperial College London, London, UK

(Accepted 10 December 2004)

SUMMARY

We investigate the epidemiology of 12 Verocytotoxin-producing *Escherichia coli* (VTEC) serogroups observed in a calf cohort on a Scottish beef farm. Fitting mathematical models to the observed time-course of infections reveals that there is significant calf-to-calf transmission of VTEC. Our models suggest that 40% of all detected infections are from calf-to-calf transmission and 60% from other sources. Variation in the rates at which infected animals recover from infection by different VTEC serogroups appears to be important. Two thirds of the observed VTEC serogroups are lost from infected calves within 1 day of infection, while the rest persist for more than 3 days. Our study has demonstrated that VTEC are transmissible between calves and are typically lost from infected animals in less than 1 week. We suggest that future field studies may wish to adopt a tighter sampling frame in order to detect all circulating VTEC serogroups in similar animal populations.

INTRODUCTION

Verocytotoxin-producing *Escherichia coli* (VTEC) are known to cause diarrhoea in young animals and humans [1, 2]. Human diarrhoeal disease can vary in severity from mild diarrhoea to haemorrhagic colitis (HC) and haemolytic uraemic syndrome (HUS) [3, 4]. VTEC infections in humans often occur via contact

with infected animals, or by consuming contaminated food and water [2, 5, 6], whereas the sources of infection in animal populations are less clear. The most common VTEC found in human patients in the United Kingdom is the serogroup O157 [2], while other serogroups such as O26, O103 and O111 have also been isolated [7]. Many VTEC serogroups, including those mentioned above, have also been detected in cattle, and it has been suggested that cattle are a potential reservoir of infection for humans [8, 9]. Thus, it is important and necessary to establish the sources of VTEC infections in animal populations

* Author for correspondence: W.-C. Liu, Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian, Scotland, EH25 9RG, UK.
(Email: W.Liu@ed.ac.uk)

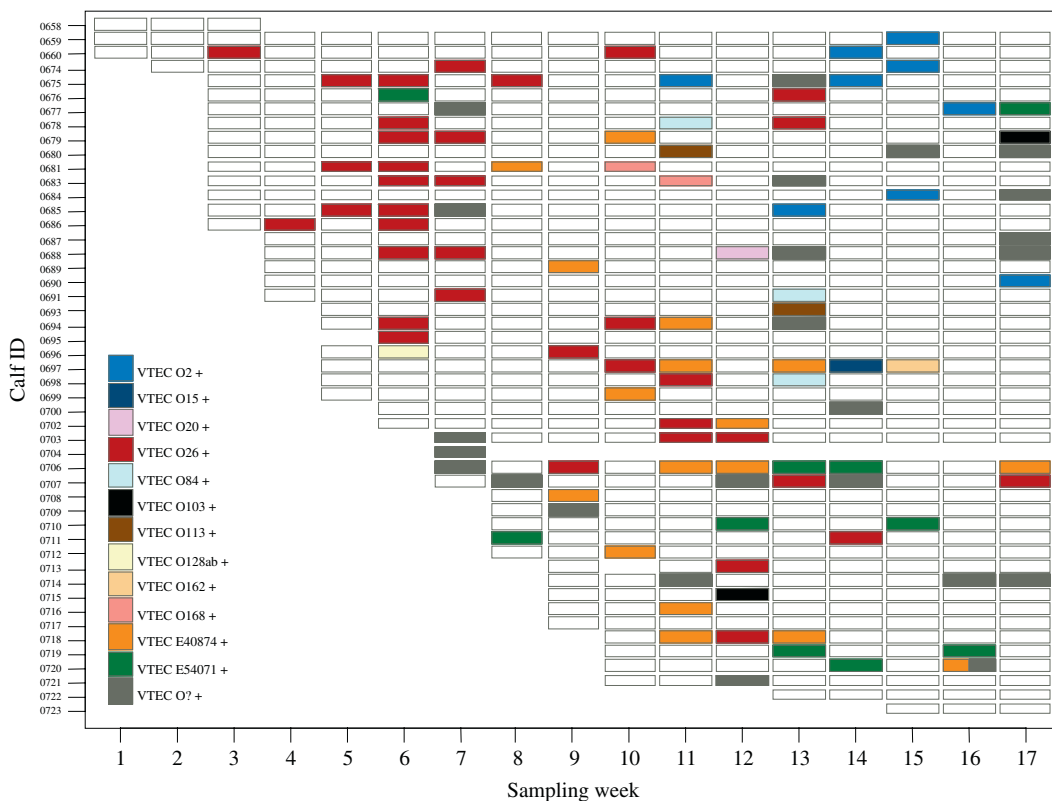


Fig. 1. VTEC serogroups isolated from 49 calves plotted by sampling weeks during the course of study. Each box represents a sample taken from a particular calf (identified by the labels on the vertical axis) on a particular week (identified by the week number on the horizontal axis). Empty boxes represent samples with no VTEC serogroups detected. Coloured boxes represent samples with VTEC serogroups detected.

and the epidemiological characteristics of VTEC serogroups.

Recently, a field study was carried out on a Scottish beef farm to monitor the presence of VTEC in calves from birth over a period of 5 months [10]. The methodology and the results of this study are summarized in [10]. In brief, faecal samples were taken from 49 calves once a week from the time of birth, and isolation of VTEC was carried out by a PCR/DNA probe hybridization method [2, 10]. In that particular study, 12 VTEC serogroups were isolated from 86 calf samples (out of a total 570 collected). Faecal samples were also taken from dams at the beginning and the end of the study, and 40 samples (out of a total 86 collected) were VTEC-positive. However, no calves excreted the same VTEC serogroup as that isolated from the dams at birth. Among those detected as VTEC, serogroup O26 was most frequently observed throughout the calf population (Fig. 1).

In this paper, we construct stochastic models to describe the epidemiology of VTEC serogroups found in the calf cohort study described above. We assume

infections can be acquired either by transmission between calves or contact with other contaminated sources. We aim to quantify the relative importance of these different sources of infection. Specifically, are all infections observed in the calf cohort a result of random acquisition of VTEC from the environment, or is there calf-to-calf transmission of infection? During our study period, different serogroups were present at very different prevalences, with O26 being the serogroup most frequently isolated. We are interested in whether these observed differences are due to different epidemiologies of the VTEC serogroups, or can be explained as a consequence of chance events in the infection and recovery processes. Within the model framework, we ask whether the dynamics of different VTEC serogroups observed in this calf cohort are governed by the same epidemiological processes and same parameter values.

This paper is organized as follows: first, we describe the data used in our models and the methods used for parameter estimation. The basic building blocks of our stochastic models are then introduced. Models

are then constructed sequentially, starting with the simplest model with the least number of parameters. We then present the results of our models and discuss their implications. This paper ends with a short summary of our findings and future directions.

METHODS

Data description

The VTEC serogroups isolated from individual calves by sampling week are summarized in Figure 1. Twelve VTEC serogroups were observed: 85 samples with single identified serogroups, 23 samples with unknown identity and one sample with multiple serogroups (i.e. O720 on week 16 in Fig. 1). For each serogroup, we define three quantities to characterize the data. The first is the *number of positive infections*, which is the number of samples that tested positive for a given serogroup. For instance, this quantity for VTEC serogroup O2 is nine positive samples. The second is the *number of animals ever infected with a serogroup*: for VTEC serogroup O2 this is eight animals. The third is *infection weeks*, which is defined as the number of weeks for which at least one sample tested positive for a given serogroup: for VTEC serogroup O2, this is 6 weeks. Figure 2 summarizes the means, variances and distributions of these three quantities for the 12 serogroups. Two thirds of observed VTEC serogroups have low numbers of positive infections, animals ever infected, and infection weeks. Serogroup O26 is the most frequently observed serogroup with the highest numbers of positive infections, animals ever infected and infection weeks.

Maximum-likelihood (ML) method of parameter estimation

We use ML methods to estimate parameter values and their confidence intervals [11, 12]. Following these methods, the best parameter set is the one for which a given model is most likely to produce the observed data. Here, the observed data are the numbers of positive infections, the numbers of animals ever infected and the infection weeks for each of the 12 observed serogroups. To calculate the probability of observing the data for a given model and parameter set, we simulate the model N times (here $N = 1\,000\,000$) to construct a probability space. We then count the number of simulations for which we reproduce the observed data. Once we have the probability of observing the data, the negative log likelihood value

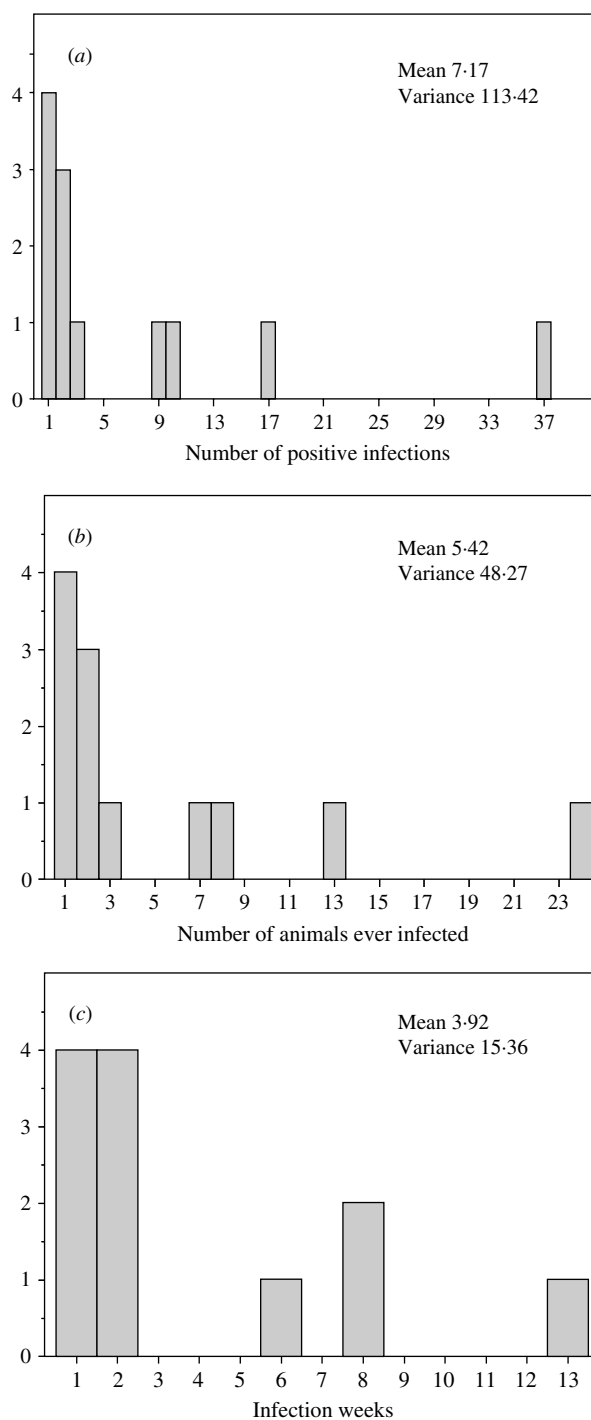


Fig. 2. Distributions of (a) the number of positive infections, (b) the number of animals ever infected and (c) the infection weeks for VTEC serogroups observed in the data.

for a model with a particular parameter set can be easily calculated [11]: a negative log likelihood value is simply the product of -1 and the logarithm of the probability of observing the data.

Trial simulations are conducted to ascertain biologically relevant areas of parameter space. We then

systematically search through the appropriate parameter space and calculate the probability of observing the data and the associated negative log likelihood value for each parameter set. For each parameter, the search range is chosen such that it incorporates the lower and upper 95% confidence intervals of the best parameter value. The best parameter set is defined as the one with the highest probability of observing the data (i.e. the smallest negative log likelihood value) [11].

The basics

For each VTEC serogroup, we construct a stochastic Susceptible–Infected–Susceptible (SIS) process confined by the cohort structure of the calf population. During this farm study, individual calves were born into the cohort at different times. Therefore, we introduce calves into our models according to their actual dates of birth. Two calves died before the completion of the study (i.e. 0658 and 0704 in Fig. 1), and were, therefore, removed from the model on their dates of death. During this study we also kept a record of the exact dates when samples were taken from calves. Therefore, during the simulation of our models, we take samples from individual calves according to the actual sampling dates.

A simple stochastic SIS process is assumed to describe the infections of calves by each VTEC serogroup. Susceptible calves (S) can acquire infection via two different routes. First is by direct or indirect contact with infected calves (I) at a rate βSI , where β is the internal transmission coefficient of a VTEC serogroup. Indirect contact here includes cases where a susceptible calf becomes infected after an encounter with an infectious source that was originally contaminated by another infected calf. Second, calves can become infected by contact with other infectious sources on a farm at a rate θS , where θ is the external transmission coefficient. The external route of transmission includes all other infectious sources not considered by the internal route. This includes infectious dams and all contaminated objects as long as infection and contamination were not caused by infectious calves; otherwise they are considered as indirect transmission from infectious calves. We assume an initial condition with no infected calves and, therefore, the very first infection in the cohort must be from external sources. Infected calves (I) recover at a rate γI , where γ is the recovery parameter. The infection dynamics for a given VTEC serogroup are illustrated in Figure 3,

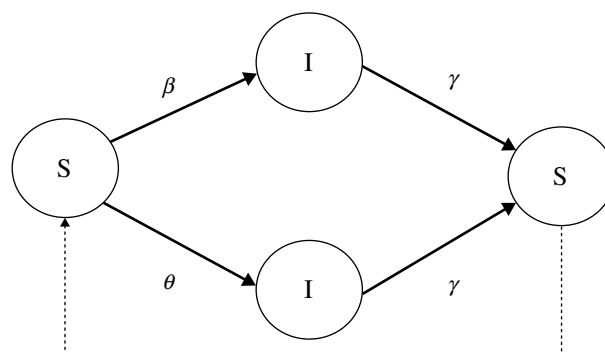


Fig. 3. A simple SIS process: a susceptible calf (S) acquires infection either by the internal transmission route (β) or the external transmission route (θ). An infected calf (I) can regain susceptible status again after recovery (γ).

and can be expressed as a set of differential equations that describe the rates of changes in the numbers of susceptible and infected calves:

$$\frac{dS}{dt} = -\beta SI - \theta S + \gamma I, \quad (1)$$

$$\frac{dI}{dt} = \beta SI + \theta S - \gamma I. \quad (2)$$

To simulate stochastically our simple SIS process based on equations (1) and (2) for a finite population, we construct a sequence of events by adopting the methodology in [13]. Imagine we are at the end of the i th event, the methodology involves the determination of waiting time s , until the occurrence of the $i+1$ th event and what happens in the $i+1$ th event. Here, one of the following can happen in the $i+1$ th event: the infection of one susceptible calf via the internal route with a probability of $\beta S_i I_i / (\beta S_i I_i + \theta S_i + \gamma I_i)$; the infection of one susceptible calf via the external route with a probability of $\theta I_i / (\beta S_i I_i + \theta S_i + \gamma I_i)$; or the recovery of an infected calf with a probability of $\gamma I_i / (\beta S_i I_i + \theta S_i + \gamma I_i)$. In those expressions, S_i and I_i are the numbers of susceptible and infected calves at the end of i th event. If an infection occurs, then the number of susceptible calves decreases by 1, and the number of infected calves increases by 1; if a recovery event occurs, then the number of infected calves decreases by 1, and the number of susceptible calves increases by 1. We assume the waiting time s , from the i th event to the $i+1$ th event is an exponentially distributed random variable [13]. To incorporate the calf cohort structure into the SIS process, we simply adjust the numbers of calves throughout according to the observed occurrence of births and deaths.

In addition to the cohort structure and SIS process described above, we sample from the process

according to the actual sampling dates used during the farm study. In our models, faecal samples are taken from calves once a week during a period of 17 weeks; infections, therefore, will not be detected if they occur between two sampling dates. As such, it is possible that some serogroups may not be detected within the sampling frame even though infections do occur in the calf cohort. We estimate how many serogroups may have been missed by this particular sampling frame.

Homogeneous model

The homogeneous model (Hom) assumes all VTEC serogroups have the same epidemiological parameters. The model also assumes no interactions between different serogroups, and allows for multiple VTEC infections in a sample. Since 12 VTEC serogroups were observed in the data, we require 12 independent SIS processes in the model. We estimate the best parameter set and determine its confidence intervals using the ML method described above [11, 12]. For β , the internal transmission coefficient, we search in a range from 0 to 0.02/day per calf with an increment of 0.001/day per calf. For θ , the external transmission coefficient, we search in a range from 0.00025 to 0.005/day with an increment of 0.00025/day. The range searched for γ , the recovery parameter, is from 0.05 to 1/day with an increment of 0.05/day.

Heterogeneous model

All assumptions are the same as the homogenous model except that heterogeneity is introduced in one of the three epidemiological parameters. There are three heterogeneous models: we refer to models with heterogeneity in the internal transmission coefficient (β), external transmission coefficient (θ) and recovery parameter (γ) as models $\text{Het}(\beta)$, $\text{Het}(\theta)$ and $\text{Het}(\gamma)$ respectively. For simplicity, we choose a binary division for each epidemiological parameter: for model $\text{Het}(\beta)$ there are two parameters for internal transmission coefficient, β_1 and β_2 ; for model $\text{Het}(\theta)$ there are two external transmission coefficients, θ_1 and θ_2 ; and similarly for model $\text{Het}(\gamma)$ there are two recovery parameters, γ_1 and γ_2 . For each heterogeneous model, we define another new parameter m , which is the number of VTEC serogroups that have epidemiological parameters with subscript 1. Like the homogeneous model, we use the ML method [11, 12] to estimate the best parameter sets and determine their confidence intervals. For β , the internal transmission coefficient,

we search in a range from 0 to 0.02/day per calf with an increment of 0.002/day per calf. For θ , the external transmission coefficient, we search in a range from 0.0005 to 0.00165/day with an increment of 0.0005/day. The range searched for γ , the recovery parameter, is from 0.1 to 12/day with the following increments: for $\gamma = 0.1$ to 3/day, the increment is 0.1/day; for $\gamma = 4$ to 12/day, the increment is 1/day. A smaller increment for lower values of γ is needed because the best value lies within the lower range (see the Results section) and we wish to determine a more accurate estimate. We also search the best parameter sets with different values of m , starting from $m = 0$ to $m = 12$.

RESULTS

Homogeneous model

For the homogeneous model, the best parameter value for the internal transmission coefficient (β) is 0.004/day per calf (95% CI 0.002–0.009). This suggests that on average, a calf encounters a given infected calf and becomes infected once every 250 days (1/0.004). For the external transmission coefficient (θ) the best fit value is 0.001/day (95% CI 0.0005–0.00325). This implies that a given calf on average acquires a VTEC serogroup once every 1000 days (1/0.001). For the recovery parameter (γ) the best fit value is 0.2/day (95% CI 0.15–0.5) and this suggests that the average duration of infectiousness is ~ 5 days (1/0.2).

To check how well the homogeneous model describes the data, we can simulate the model 10000 times with the best parameter values. For each realization, the means and variances of the number of positive infections, number of animals ever infected and infection weeks are calculated. We then plot distributions of those model means and variances and ask where the means and variances of the observed data are placed within the model distributions. All observed data means and variances are all placed within the 95% confidence region in the model distributions except the variance for infection weeks (Fig. 4a); the homogeneous model thus fails to explain the variation in infection weeks.

Heterogeneous models

According to the likelihood ratio test [11], for the heterogeneous models (with five parameters) to represent a significant improvement over the homogeneous model (with three parameters) at the 0.05

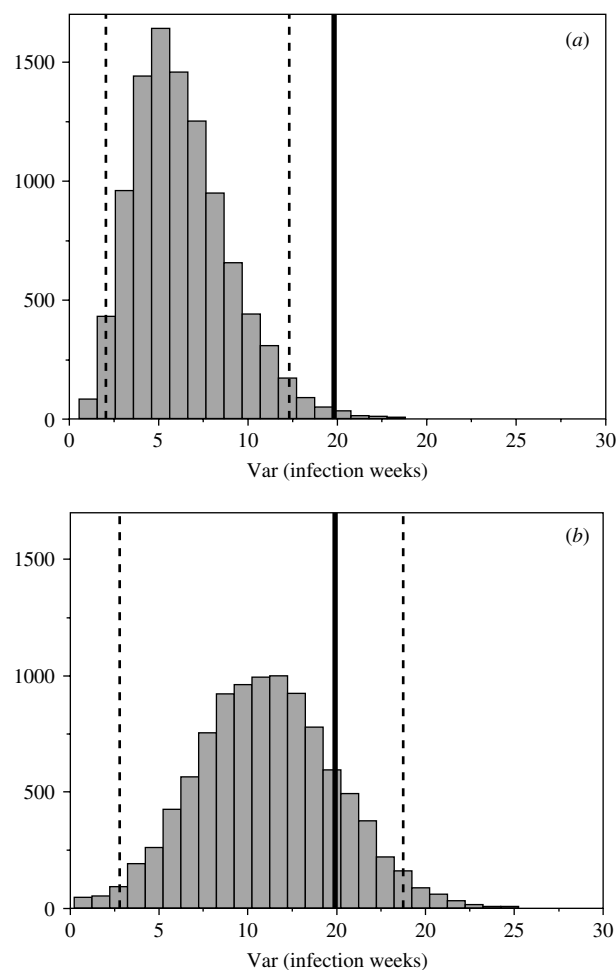


Fig. 4. (a) Distribution of model variance for infection weeks for the homogeneous model. The bold line indicates where the data variance is placed within the model distribution. The dashed lines are the 95% confidence limits. (b) Distribution of model variance for infection weeks for model Het(γ). The bold line indicates where the data variance is placed within the model distribution. The dashed lines are the 95% confidence limits.

level, a reduction of 3 or more in the negative log likelihood value is required. The negative log likelihood values given in Table 1 show that only the model with heterogeneity in the recovery rate [i.e. Het(γ)] is significantly better than the homogeneous model, although other heterogeneous models are also improvements on the homogeneous model.

For model Het(γ), the best parameter value for the internal transmission coefficient (β) is 0.004/day per calf (95% CI 0.002–0.012), and the best fit value for the external transmission coefficient parameter (θ) is 0.005/day (95% CI 0.002–0.0125). For the two recovery parameters γ_1 and γ_2 , the best fit values are 0.3/day (95% CI 0.2–0.7) and 1.9/day (95% CI

Table 1. Negative log likelihood values for all models

Model	Negative log likelihood values	Difference to the homogeneous model
Hom	24.49	0
Het(β)	23.64	0.85
Het(θ)	21.76	2.73
Het(γ)	19.87	4.62

1.10–8.0) respectively. Note that the best estimates for the internal transmission coefficient (β) for both the homogeneous model and model Het(γ) are the same. For the external transmission coefficient (θ) and the recovery parameters (γ_1 or γ_2), the estimates for model Het(γ) are much higher than the homogeneous model. In other words, increases in the external transmission coefficient (θ) are compensated by faster recovery rates (γ_1 or γ_2). Model Het(γ) suggests that four serogroups have a lower recovery rate γ_1 and eight serogroups have a much higher recovery rate γ_2 . Furthermore, model Het(γ) is the model that best explains the observed variance in infection weeks (Fig. 4b).

Estimating the number of VTEC serogroups present

In our model simulations, some infected calves will not be recorded as positive for VTEC serogroups simply because we sample only on specific days. Thus, for some model simulations fewer than 12 VTEC serogroups will be detected. We can simulate models with different numbers of VTEC serogroups (i.e. different numbers of SIS processes) and calculate the average number of serogroups observed given the sampling frame. Our results suggest that in order to observe an average of 12 serogroups, we have to include an average of 14 serogroups in the model. Furthermore, if more than 17 serogroups are included in the model, then in fewer than 5% of model simulations will 12 VTEC serogroups be detected (Fig. 5). Thus, in order to observe 12 serogroups in the data, we estimate that 12–17 serogroups are circulating in the calf cohort.

Sources of infection

With 14 serogroups (as this gives an average of 12 observed serogroups), we simulate 10 000 times the homogeneous model and the model with heterogeneity in recovery rate [i.e. model Het(γ)] using the ML parameter estimates. For each realization, we

Table 2. Numbers of samples out of a total of 86 involving single, double, triple and quadruple consecutive infections

	Data	Model
Out of 86 positive samples		
Single infection	68 (80%)	65.6 (76.3%, 63–93)
Double consecutive infection	18 (20.9%)	16.3 (18.9%, 5–31)
Triple consecutive infection	0 (0%)	3.41 (3.96%, 0–12)
Quadruple consecutive infection	0 (0%)	0.006 (0.007%, 0–6)

The values within parentheses are the percentages of samples and their 95% confidence intervals.

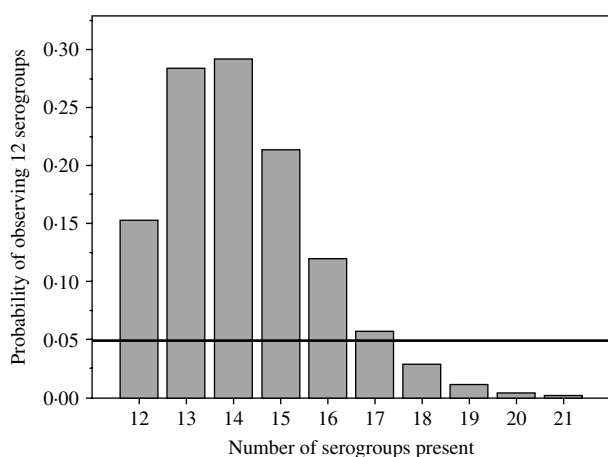


Fig. 5. Probability profile of observing 12 VTEC serogroups when different numbers of VTEC serogroups are incorporated into models. The bold line represents the cut-off at the 0.05 level.

determine the source of infection for each infected sample. For the homogeneous model, our result suggests 70% (95% CI 48–83) of all infections arise through calf-to-calf transmission (both direct and indirect), whereas this figure decreases to 40% (95% CI 19–55) for model $\text{Het}(\gamma)$. Thus, according to the homogeneous model, only 30% of infections are from external sources. However, this figure increases to 60% for model $\text{Het}(\gamma)$. This rise in the proportion of infections from external sources is expected since the value for the external transmission coefficient (θ) for model $\text{Het}(\gamma)$ is five times larger than that for the homogeneous model.

We can also track each infected sample and determine the frequency of single and consecutive infections. For consecutive infections, we then determine whether the infection following the initial infection is a consequence of persistence or re-infection (i.e. a calf recovering and then acquiring infection again the following week). The number of observed single and

consecutive infections in the data (there were 86 positive samples, among which 68 were single infections and 18 were involved in double consecutive infections) is predicted well by model $\text{Het}(\gamma)$ (Table 2; $\chi^2 = 3.69$, D.F. = 3, $P = 0.298$). Whilst the 95% confidence intervals do include zero samples with triple consecutive infections (Table 2), this is only true for 31% of the model simulations. No quadruple consecutive infections were observed in the data (Table 2) nor in 85% of the model simulations. Only 15% of the model simulations showed the occurrence of quadruple consecutive infections. Furthermore, the model predicts that among those double consecutive infections observed in the data, 88% of them are the result of persistence of VTEC serogroups within the calves, whereas only 12% are the result of re-infections.

Estimating the number of samples with multiple infections

We have assumed independent SIS processes for all VTEC serogroups in all models; therefore, more than one serogroup may be present in any one calf faecal sample. To estimate the expected number of samples with multiple infections, we can run simulations with 14 serogroups (because this gives an average of 12 observed serogroups) 10 000 times, and for each realization we determined the number of serogroups per sample. Of 570 samples collected, the data shows 484 samples with no infections, 85 samples with a single serogroup and only one sample with two serogroups (Table 3). Model $\text{Het}(\gamma)$ predicts the observed data reasonably well (Table 3; $\chi^2 = 5.22$, D.F. = 3, $P = 0.156$). On average, the model predicts 484 samples with no infections, 79.2 samples with a single serogroup and 6.56 samples with double infections. The 95% confidence intervals do include only one sample with double infections (Table 3); however, only 12% of the model simulations have one sample with double

Table 3. Numbers of samples infected with 0, 1, 2 and 3 VTEC serogroups

Out of 570 samples		
Number of serogroups	Data	Model
0	484 (84.9%)	484 (84.9%, 75–94)
1	85 (14.9%)	79.2 (13.9%, 6–22)
2	1 (0.175%)	6.56 (1.15%, 0–3)
3	0 (0%)	0.0969 (0.02%, 0–0)

The values within parentheses are the percentages of samples and their 95% confidence intervals.

infections. Samples with three or more serogroups were not observed in the data (Table 3) or in 75% of the model simulations. Only 25% of the model simulations have samples infected with three serogroups.

DISCUSSION

A previous analysis of this calf cohort [10] has shown that VTEC serogroups isolated from calves were different to those isolated from their dam at birth; it has, therefore, been suggested that calves might acquire their VTEC infections from environmental sources [10]. Here, we have assumed that calves can acquire infections either from other infected calves (directly or indirectly) or from other sources (this also includes an infected dam). Our estimates show the transmission coefficient to be significantly greater than zero. Therefore, the model results suggest there is calf-to-calf VTEC transmission within the calf cohort. Moreover, our results reveal that a substantial proportion of infections arises via this route [70% of all infections are from calf-to-calf transmission for the homogeneous model and 40% for model $\text{Het}(\gamma)$].

Heterogeneity can be introduced into the homogeneous model in many different ways: one can have heterogeneity in one, two or all epidemiological parameters in various combinations; or one can go beyond the binary division and have 3, 4, 5 or more different values for each epidemiological parameter. To determine fully the differences between different VTEC serogroups in terms of epidemiological characteristics will require a complete exploration of all possible heterogeneous models (and ultimately estimation of parameter values for each serogroup). Therefore, the conclusion we draw from our simpler heterogeneous models is conservative. We have found that there is, at least, a difference between those observed

VTEC serogroups. For this particular dataset, model $\text{Het}(\gamma)$ suggests that this difference is in the rate at which VTEC serogroups are lost from calves. In particular, we estimate four serogroups to have low recovery rates and eight to have much higher recovery rates. A higher recovery rate can reduce the total number of infections observed and as a result reduce the number of animals ever infected and infection weeks. This division of two classes of VTEC serogroups is observed in the data where eight serogroups have small numbers of total infections and animals ever infected, as well as very short infection weeks (Fig. 2). Recovery rate also has an impact on how often we see consecutive infections. The data shows that there are nine occurrences of double consecutive infections (i.e. 18 samples involved in double consecutive infections), and our model suggest that eight of these occurrences (88%) are the result of persistence of VTEC serogroups within calves rather than re-infection.

Our results also suggest that calves are typically infected for less than a week, with estimates for recovery periods ranging from 5 days for the homogeneous model, to as little as <1 day for the heterogeneous model. It is possible, therefore, that the weekly sampling adopted in practice failed to detect some additional circulating serogroups. Our results suggest that in order to observe an average of 12 serogroups, there may be on average 14 serogroups circulating in the calf cohort. A tighter sampling frame would be required to observe all the serogroups circulating in this particular calf cohort. Imperfect sensitivity of the PCR/DNA probe hybridization method might also result in missing serogroups [2]. To test this, we incorporated a sensitivity parameter into our homogeneous model. We define sensitivity as the probability of a sample being tested positive if it is infected with a serogroup. We found that the parameter values for the homogeneous model are robust across a wide range of sensitivities (from 0.2 to 1), and that the homogeneous model with sensitivity is not statistically a significantly better model. Therefore, we concluded that sensitivity is not an important factor in explaining the data.

Although the 95% confidence intervals for the number of samples infected with two serogroups do include one sample (Table 3), 88% of the model simulations show that there are more than one sample with double infections (on average, 6.56 samples are expected, see Table 3). Thus, the model clearly overestimates the number of samples with double infections; or there should be more than one sample with

double infections in the data. There are several potential explanations to account for this. First, the assumption of independent infections for all serogroups might be inappropriate in our models. Experiments in microbiology are thus required to investigate the possibility of interactions between different VTEC serogroups. Second, the detection/isolation method for VTEC serogroup may not be sufficiently sensitive to detect more than one serogroup per sample. For the method to detect a particular VTEC serogroup, the number of organisms in each faecal sample must be above a detection threshold [10, 14]. Furthermore, Pearce et al. [15] used immunomagnetic separation (IMS) to detect VTEC in the same calf cohort and detected serogroup O145, which was not isolated by the PCR/DNA probe hybridization method used by Shaw et al. [10]. Therefore, it is possible that the number of serogroups observed per sample (or even across the whole calf cohort) might be limited by the detection method used. Finally, it might simply be due to VTEC serogroups in the material taken from each faecal sample for detection not being fully representative of those in individual calves: this is a potential worry in our study since only a small amount (i.e. 1 g) was taken from each faecal sample for VTEC detection.

CONCLUSION

In this paper, we have modelled the epidemiology of 12 VTEC serogroups observed in a calf cohort on a Scottish beef farm. Our models suggest there is calf-to-calf transmission of VTEC serogroups within the calf cohort (the transmission coefficient of VTEC within the calf population is significantly greater than zero), but a substantial proportion of transmission is from other sources (60% of all infections). Our findings suggest that some VTEC serogroups have a very short duration of infection of <1 day, while there is evidence that some VTEC serogroups have a longer duration of infection (3–5 days). We also demonstrated there is no good evidence that internal and external transmission rates vary among different VTEC serogroups. Our model predicts that of all consecutive infections that were observed in the data, the majority (88%) are from the persistence of VTEC serogroups within infected calves. Our findings also suggest that some VTEC serogroups are transient in this calf cohort with calves recovering within a day of infection, and therefore a tighter sampling frame may be required to detect all circulating serogroups in future

studies. The use of complementary detection methods in future studies could also help to ensure that the number of serogroups observed is not limited by the use of a particular isolation method. We are aware that the data used were from a particular farm, which may not be a truly representative of all the farms in the United Kingdom. We plan to apply this methodology to other datasets and test the robustness of our conclusions.

ACKNOWLEDGEMENTS

This study was funded by the Wellcome Trust International Partnership Research Award in Veterinary Epidemiology (IPRAVE) project. We thank the Scottish Agricultural College, Inverness, and in particular Alastair Smith and Hazel Knight, for the processing of calf faecal samples during the study. L.M. is grateful to the Wellcome Trust for a Mathematical Biology Research Training Fellowship.

REFERENCES

1. **Sherwood D, Snodgrass DR, O'Brien AD.** Shiga-like toxin production from *Escherichia coli* associated with calf diarrhoea. *Vet Rec* 1985; **116**: 217–218.
2. **Willshaw GA, Cheasty T, Smith HR, O'Brien SJ, Adak GK.** Verocytotoxin-producing *Escherichia coli* (VTEC) O157 and other VTEC from human infections in England and Wales: 1995 to 1998. *J Med Microbiol* 2001; **50**: 135–142.
3. **Ackman D, Marks S, Mack P, Caldwell M, Root T, Birkhead G.** Swimming-associated haemorrhagic colitis due to *Escherichia coli* O157:H7 infection: Evidence of prolonged contamination of a fresh water lake. *Epidemiol Infect* 1997; **119**: 1–8.
4. **Karmali MA, Petric M, Lim C, Fleming PC, Arbus GS, Lior H.** The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*. *J Infect Dis* 1985; **151**: 775–782.
5. **Milne LM, Plom A, Strudley I, et al.** *Escherichia coli* O157 incident associated with a farm open to members of the public. *Commun Dis Public Health* 1999; **2**: 22–26.
6. **Trevena WB, Hooper RS, Wray C, Willshaw GA, Cheasty T, Dominique G.** Vero cytotoxin-producing *Escherichia coli* O157 associated with companion animals. *Vet Rec* 1996; **138**: 400.
7. **Willshaw GA, Scotland SM, Smith HR, Rowe B.** Properties of verocytotoxin-producing *Escherichia coli* of human origin of O-serogroups other than O157. *J Infect Dis* 1992; **166**: 797–802.
8. **Beutin L, Geier D, Steinruck H, Zimmermann S, Scheutz F.** Prevalence and some properties of verotoxin (shiga-like toxin)-producing *Escherichia coli* in 7 different

- species of healthy domestic animals. *J Clin Microbiol* 1993; **31**: 2483–2488.
9. **Kobayashi H, Shimada J, Nakazawa M, et al.** Prevalence and characteristics of Shiga toxin-producing *Escherichia coli* from healthy cattle in Japan. *Appl Environ Microbiol* 2001; **67**: 484–489.
 10. **Shaw DJ, Jenkins C, Pearce MC, et al.** Analysis of characteristics of shedding patterns on verocytotoxin-producing *Escherichia coli* strains in a cohort of calves and their dams on a Scottish beef farm. *Appl Environ Microbiol* 2004 (in press).
 11. **Hilborn R, Mangel M.** The confrontation: likelihood and maximum likelihood. In: *The ecological detective: confronting models with data*. Princeton, NJ: Princeton University Press, 1997: 131–179.
 12. **Hudson DJ.** Interval estimation from the likelihood function. *J Roy Stat Soc B* 1971; **33**: 256–262.
 13. **Renshaw E.** Epidemic processes. In: *Modelling biological populations in space and time*. Cambridge University Press, 1991: 324–359.
 14. **Willshaw GA, Smith HR, Roberts D, Thirlwell J, Cheasty T, Rowe B.** Examination of raw beef products for the presence of verocytotoxin producing *Escherichia coli*, particularly those of serogroup O157. *J Appl Bacteriol* 1993; **75**: 420–426.
 15. **Pearce MC, Jenkins C, Vali L, et al.** Temporal shedding patterns and virulence factors of *Escherichia coli* serogroups O26, O103, O111, O145 and O157 in a cohort of beef calves and their dams. *Appl Environ Microbiol* 2004; **70**: 1708–1716.