Preventive Veterinary Medicine 119 (2015) 97–104

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

The feasibility of testing whether *Fasciola hepatica* is associated with increased risk of verocytotoxin producing *Escherichia coli* O157 from an existing study protocol

Graeme L. Hickey[a,*], Peter J. Diggle[a], Tom N. McNeilly[b], Sue C. Tongue[c,1], Margo E. Chase-Topping[d], Diana J.L. Williams[e,f]

[a] *Department of Epidemiology and Population Health, Institute of Infection and Global Heath, University of Liverpool, The Farr Institute@HeRC, Waterhouse Building, 1-5 Brownlow Street, Liverpool L69 3GL, UK*
[b] *Moredun Research Institute, Pentlands, Science Park, Bush Loan, Penicuik, Midlothian EH26 0PZ, UK*
[c] *SRUC (Scotland's Rural College), West Mains Road, Edinburgh, Scotland EH9 3JG, UK*
[d] *Centre for Immunity, Infection and Evolution, University of Edinburgh, Ashworth Laboratories, Charlotte Auerbach Road, Edinburgh, EH9 3FL, UK*
[e] *Department of Infection Biology, Institute of Infection and Global Heath, University of Liverpool, Liverpool L69 7ZJ, UK*
[f] *School of Veterinary Science, University of Liverpool, Leahurst Campus, Chester High Road, Neston CH64 7TE, UK*

ARTICLE INFO

Article history:
Received 8 October 2014
Received in revised form 19 January 2015
Accepted 22 February 2015

Keywords:
Power
Mixed effects model
Sample size
Co-infection
Fasciola hepatica
Verocytotoxigenic *Escherichia coli* O157

ABSTRACT

The parasite *Fasciola hepatica* is a major cause of economic loss to the agricultural community worldwide as a result of morbidity and mortality in livestock, including cattle. Cattle are the principle reservoir of verocytotoxigenic *Escherichia coli* O157 (VTEC O157), an important cause of disease in humans. To date there has been little empirical research on the interaction between *F. hepatica* and VTEC O157. It is hypothesised that *F. hepatica*, which is known to suppress type 1 immune responses and induce an anti-inflammatory or regulatory immune environment in the host, may promote colonisation of the bovine intestine with VTEC O157. Here we assess whether it is statistically feasible to augment a prospective study to quantify the prevalence of VTEC O157 in cattle in Great Britain with a pilot study to test this hypothesis. We simulate data under the framework of a mixed-effects logistic regression model in order to calculate the power to detect an association effect size (odds ratio) of 2. In order to reduce the resources required for such a study, we exploit the fact that the test results for VTEC O157 will be known in advance of testing for *F. hepatica* by restricting analysis to farms with a VTEC O157 sample prevalence of >0% and <100%.

From a total of 270 farms (mean 27 cows per farm) that will be tested for VTEC O157, power of 87% can be achieved, whereby testing of *F. hepatica* would only be necessary for an expected 50 farms, thus considerably reducing costs. Pre-study sample size calculations are an important part of any study design. The framework developed here is applicable to the study of other co-infections.

* Corresponding author at: Department of Epidemiology and Population Health, Institute of Infection and Global Heath, University of Liverpool, The Farr Institute@HeRC, Waterhouse Building, 1-5 Brownlow Street, Liverpool L69 3GL, UK. Tel.: +44 0151 794 6023/2000.
E-mail address: graeme.hickey@liverpool.ac.uk (G.L. Hickey).
[1] Current address: SRUC Epidemiology Research Unit, Drummondhill, Stratherrick Road, Inverness IV2 4JZ, UK.

http://dx.doi.org/10.1016/j.prevetmed.2015.02.022
0167-5877/

1. Introduction

Fasciolosis caused by *Fasciola hepatica*, more commonly referred to as liver fluke, is a major cause of economic loss to the agricultural community worldwide as a result of morbidity and mortality in livestock (Claridge et al., 2012). Verocytotoxigenic *Escherichia coli* (VTEC) O157 is a zoonotic bacteria of worldwide importance which, whilst largely asymptomatic in cattle, causes haemorrhagic colitis and potentially fatal haemolytic uraemic syndrome in humans as a result of systemic verocytotoxin activity (Chase-Topping et al., 2008). For both of these pathogens, cattle are a primary reservoir (Armstrong et al., 1996; McCann et al., 2010a,b).

To date there has been little empirical research on the interaction between *F. hepatica* and VTEC O157 in cattle. Recent evidence suggests a role for type 1 immune responses in control of VTEC O157 in cattle, with clearance of the bacteria from the bovine intestine associated with an up-regulation of T-helper type 1 associated transcripts within the rectal mucosa, the principle site of colonisation by this bacteria (Corbishley et al., 2014; Naylor et al., 2003). On the other hand, *F. hepatica* is known to suppress type 1 immune responses and induce an anti-inflammatory or regulatory immune environment in the host (Brady et al., 1999; Flynn et al., 2007). It is of interest to determine if infection with *F. hepatica* increases the risk that a cow will shed VTEC O157, as this will inform future control and risk management strategies aimed at reducing incidence of disease in humans.

Before a study with the sole purpose of establishing whether such an association exists is likely to be funded, pilot data that can be obtained within existing resource limitations and logistical constraints are required. A programme of work has been approved by the UK Food Standards Agency (FSA; Project FS101055), which includes a survey of VTEC O157 in beef cattle intended for the food chain in Scotland and in England and Wales. The sampling protocol for these studies has already been defined and is based on previously published methods (Gunn et al., 2007; Pearce et al., 2009). Therefore, the objective of this study is to establish whether it is feasible to augment the FSA VTEC O157 study with a pilot study to determine whether shedding of VTEC O157 is independent of *F. hepatica* infection in cattle. The statistical framework we have developed here is applicable to other studies of co-infection.

2. Materials and methods

2.1. Primary study protocol

For the purpose of quantifying the current prevalence of VTEC O157 across Great Britain, the sampling protocol developed for two previous surveys will be used (Gunn et al., 2007; Pearce et al., 2009). Briefly, a sufficient number of pat samples will be taken from each group of cattle to ensure 90% probability of detecting shedding of VTEC O157 if at least one shedding animal is present. The aim is to sample 110 Scottish and 160 English/Welsh farms; only farms with one or more store/finishing cattle will be included. Each farm will be visited on one occasion and visits will be spread over a 12-month period. Fresh faecal pat samples will be collected in accordance with the size of the group. In a previous Wellcome Trust funded International Partnership Research Award in Veterinary Epidemiology (IPRAVE), a cross-sectional survey of beef cattle was carried out in Scotland (Pearce et al., 2009), where between 1 and 113 samples were taken per group, with a mean of 27 and a median of 23 (Chase-Topping, unpublished results).

In the new study, samples from each farm will be sent to the Epidemiology Research Unit microbiological facilities, SAC Consulting Veterinary Services Disease Surveillance Centre, Inverness, within 48 h of collection. Samples of faeces from each pat will be tested for the presence of VTEC O157 using an immuno-magnetic separation technique as previously described in (Pearce et al., 2004). The required amount to fulfil the requirements of the FSA project will be removed and if sufficient sample is available, an aliquot will be removed and stored at −80 °C until the results of the VTEC O157 tests are known (approximately 1-week). Any further testing will be subject to the farmer having provided permission for their samples and data to be used for purposes other than those of the primary study.

2.2. Secondary study protocol

Following completion of VTEC O157 testing, stored samples will be made available for fluke testing. Samples will be transported under Containment Level 3 conditions to a separate laboratory (Moredun Research Institute, Edinburgh) to be tested for *F. hepatica* using a copro-antigen detection ELISA (Bio-X Diagnostics, Jemelle, Belgium).

A unique feature of this study is that the presence of the first pathogen (VTEC O157) will be known *prior* to any testing for the second pathogen (*F. hepatica*).

2.3. Statistical analysis

On observing the data, we will fit a mixed effects logistic regression model. In the absence of data on other explanatory variables, this model will be

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_j + \beta x_{ij} \quad (1)$$

where

- p_{ij} is the probability of cow i on farm j testing positive for VTEC O157.
- α_j is the intercept for farm j, such that the α_j are independently distributed according to a normal distribution with mean μ and standard deviation σ.
- $x_{ij} = 1$ if cow i on farm j tests positive for *F. hepatica*, and 0 otherwise.
- β is the natural log odds ratio (OR) for a positive *F. hepatica* test.

If data on additional explanatory variables were available, incorporating these into the statistical model would potentially reduce the heterogeneity currently captured by the random effects term, giving rise to a more powerful test

of the association with liver fluke. The results reported here can therefore be considered as being conservative.

We exploit the fact that each VTEC O157 test result will be known in advance of the samples being requested for liver fluke testing. If each sample were to be tested at the same time for both pathogens, then >7000 *F. hepatica* tests would need to be carried out. As VTEC O157 has a relatively low prevalence, many farms will have a sample prevalence of 0%. We expect for this study that these farms will contribute very little to the estimation of the model parameters; therefore we exclude them prior to fitting the regression model. By the same reasoning, some small farms with very small sample sizes might also have a sample prevalence of 100%, and these are also excluded prior to model fitting. Testing only pat samples from farms with a VTEC O157 sample prevalence between >0% and <100% leads to substantial cost-savings, which is the primary motivation here.

Power is the probability of a test to reject the null hypothesis when it is false. Conventionally, a test is considered reasonable if its power is >80%. The null hypothesis in this study is that $\beta = 0$ (*i.e.* odds ratio = 1). For the calculations reported here, we assume that a scientifically significant effect would be that a positive fluke test result doubles the odds of testing positive for VTEC O157 hence we set the alternative hypothesis to be $\beta = \log(2)$.

There is no closed-form solution for the power of the test, and we therefore use a simulation-based approach (Gelman and Hill, 2007) as follows.

1. Simulate a plausible synthetic dataset that adheres to any known constraints (detailed below) under the alternative hypothesis.
2. Fit the proposed regression model.
3. Test the null hypothesis at the 5% significance level.
4. Repeat 2500 times and calculate the power as the proportion of simulations where the null hypothesis was rejected.

For this particular study, before performing step 2 we will exclude farms with 0% or 100% VTEC O157 sample prevalence from the synthetic dataset, as described above. In what follows, we explore the implications of randomly sampling $M < 270$ farms, which, if conferring adequate power, would reduce the necessary resource liability and allow for 'drop outs' due to inadequate faecal sample amounts or farmers refusing permission for secondary reuse of the samples, which we assume could be treated as being non-informative (Little and Rubin, 1987). We would note from experience, however, that farmers are generally receptive to *F. hepatica* testing, as it is non-controversial; therefore refusal for secondary reuse in this context would be unlikely.

Sensitivity analyses are performed to evaluate the dependence of statistical power on: (1) the choice of the alternative hypothesis and (2) the initial prevalence estimates. We also illustrate our proposed approach by simulating one additional dataset and fitting the described model.

All analyses are run in R version 3.1.0 (R Foundation for Statistical Computing, Vienna, Austria). Optimisation and numerical quadrature were performed using base functions 'uniroot' and 'integrate' respectively. Mixed effects models are fitted with the lme4 (version 1.1-6) R package (Bolker et al., 2009). Systems of non-linear equations were solved using the rootSolve (version 1.6.5) R package (Soetaert and Herman, 2009). Graphs were produced using the ggplot2 (version 1.0.0) R package (Wickham, 2009). All logarithms are to the natural base.

3. Data simulation

As the data have not yet been observed, we must simulate a plausible synthetic dataset, adhering to constraints identified from other available data. To account for the uncertainty in the data, we simulate multiple datasets. There are four components to this simulation: (1) the farm-level sample sizes; (2) the *F. hepatica* infection; (3) the farm-level random-effects; and (4) the VTEC O157 infection.

3.1. Farms

Let N_j be a random variable denoting the sample size for farm j. To match the proposed study design, the mean and median sample size should be approximately 27 and 23 cows per farm respectively, with a range of 1–113. We simulate the number of cows that will be tested in each of $M = 270$ farms from a Beta-Binomial model – a generalisation of the Binomial model that allows for the probability of the event to be random, and modelled using a Beta distribution, rather than fixed – thus allowing for over-dispersion to be incorporated. To achieve this, for each farm j we first randomly draw q_j from a Beta distribution with shape parameters $a = 1.32$ and $b = 4.35$ (see Appendix for explanation). We then randomly draw values of $N_j - 1$ from a Binomial distribution with size $113 - 1 = 112$ and event probability q_j. The Binomial model permits random draws of 0 (*i.e.* 'no events'), whereas the simulation model ensures that every farm has at least one cow tested.

3.2. F. hepatica infection

Based on existing data, we want the approximate marginal prevalence of *F. hepatica* among individual cows (ignoring clustering effects) to be 20%, and the farm-level prevalence (the proportion of farms with ≥ 1 cow testing positive for *F. hepatica*) to be 80% (McCann et al., 2010a,b; Salimi-Bejestani et al., 2005) (http://www.eblex.org.uk/). To achieve this, within each farm we infect cows with *F. hepatica*, *in silico*, with a within-farm probability r_j sampled from a Beta distribution with shape parameters $u = 0.99$ and $v = 3.97$ (see Appendix for explanation).

3.3. Farm effects and VTEC O157 infection

We expect VTEC O157 to be clustered within farms, thus driving heterogeneity. We simulate farm-level random effects α_j on the logit scale from a normal distribution with mean μ and standard deviation σ. Based on existing data, we want the approximate marginal prevalence of VTEC O157 among individual cows to be 4%, and the

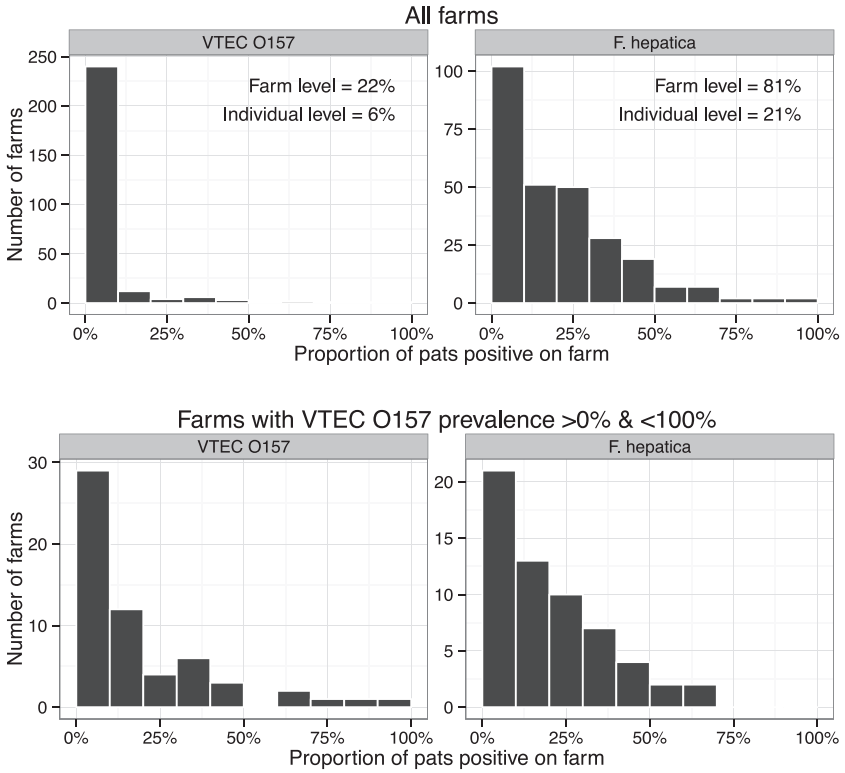

Fig. 1. Farm-level prevalence distribution for VTEC O157 and *F. hepatica* for a single simulated synthetic dataset of 270 farms. Bottom row shows data after excluding farms with either 0% or 100% VTEC O157 sample prevalence.

farm-level prevalence (the proportion of farms with ≥1 cow testing positive for VTEC O157) to be 19% (Pearce et al., 2009). We determine that this is achieved by selecting $\mu = -7.09$ and $\sigma = 3.52$ when $\beta = \log(2)$ (see Appendix for explanation). We then infect each cow i on farm j, *in silico*, with probability p_{ij}, as defined by Eq. (1).

4. Results

4.1. Power analysis

For each of $M = 90$, 135, 180, 225 and 270 farms, we simulated 2500 fake datasets. Fig. 1 summarises a single simulation of 270 farms. The farm-level and individual-level infection prevalence figures are not identical to those specified because of random sample variation. We observe that a large number of farms have 0% infection for VTEC O157, and are therefore excluded prior to model fitting. The simulated datasets satisfied the desired attributes on average, as shown in Fig. 2. Of the 2500 datasets with 270 farms, the mean (standard deviation) of the farm-level and individual-level sample prevalence for VTEC O157 was 18.9% (2.4%) and 4.0% (1.0%) respectively. Similarly, for *F. hepatica* they were 79.9% (2.4%) and 20.0% (1.3%) respectively. The mean (standard deviation) of the average number of cows per farm across datasets was 27.1 (1.1) for the mean, and 23.4 (1.5) for the median.

The power curve is shown in Fig. 3. It shows that from a synthetic dataset of 270 farms included in the FSA survey, only 50 farms on average, equating to an average of 1645 pat samples, would have a sample VTEC O157 prevalence of >0% or <100% and thus require testing for *F. hepatica*. This would yield power of 87% to detect an odds ratio of 2, hence there is potential to test fewer farms. Repeating the exercise with 225 farms, we find that we expect to apply fluke testing to 42 farms, equating to approximately 269 fewer pat sample tests, whilst yielding power of 82%.

4.2. Sensitivity analyses

We simulated 2500 synthetic datasets under alternative hypothesis effect sizes of $\beta = \log(1.2)$, $\log(1.4)$, …, $\log(2.0)$, …, $\log(3.0)$, in each case with the maximum number of farms available for testing ($M = 270$). In each case we recalculated the random-effects distribution parameters to match the marginal VTEC O157 farm-level prevalence. Power ranged from 13.0% (for OR = 1.2) to 99.6% (for OR = 3.0) (Fig. 4). The power to detect an OR of 1.8 would be 75%.

We considered all combinations of the underlying prevalence values: farm-level VTEC O157 (10% and 25%); farm-level *F. hepatica* (70% and 80%); and individual-level *F. hepatica* (15% and 25%). The individual-level VTEC O157 prevalence was left unchanged as the theory underlying the transmission dynamics suggest this value should be stable. A reduction in the farm-level VTEC O157 was commensurate with a reduction in statistical power, ranging between 54% and 67% for the different *F. hepatica* prevalence values

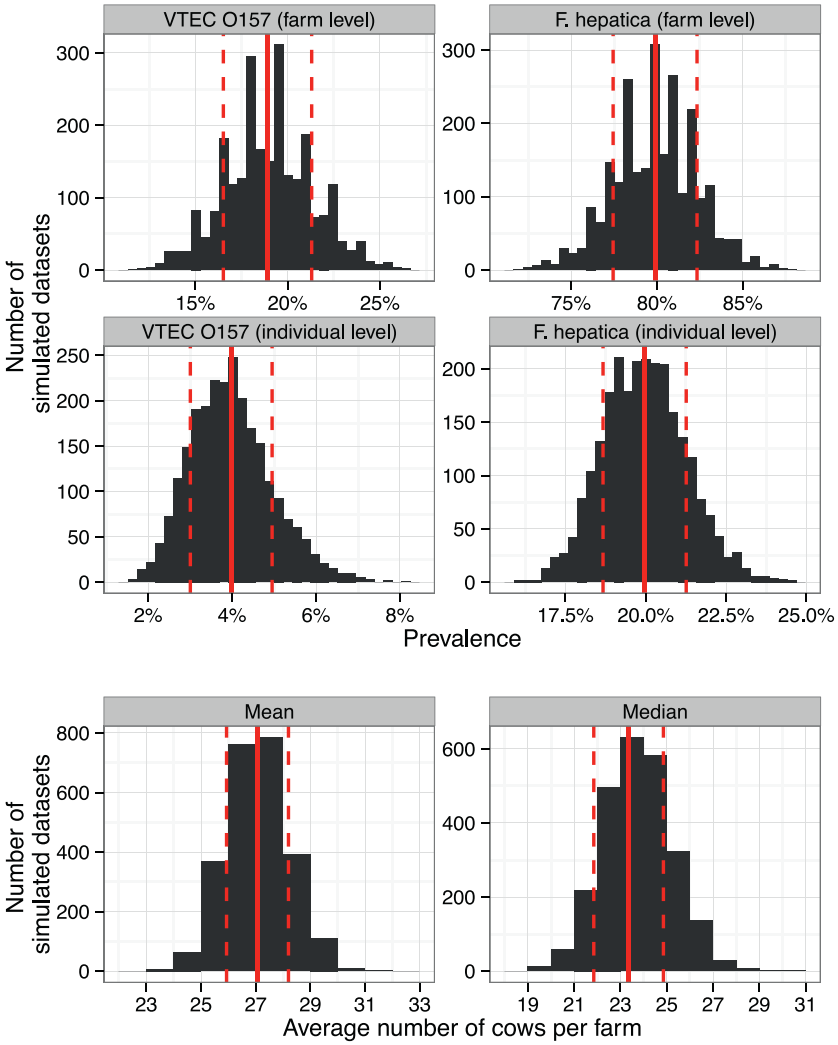

Fig. 2. Distributions of summary statistics calculated for each of 2500 synthetic datasets. Top four panels show the farm-level and individual-level sample prevalence for VTEC O157 and *F. hepatica*. The bottom two panels show the distribution of the sample mean and sample median for the number of cows per farm. Solid red lines denote the mean of the distribution with dashed red lines denoting ±1 standard deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Sensitivity analysis results showing statistical power for a combination of different underlying prevalence values. The analysis is based on first performing VTEC O157 testing on all $M = 270$ farms.

VTEC O157 (farm-level)	*F. hepatica* (farm-level)	*F. hepatica* (individual-level)	Power
25%	70%	15%	88.4%
25%	70%	25%	88.8%
25%	85%	15%	93.8%
25%	85%	25%	95.3%
10%	70%	15%	54.0%
10%	70%	25%	55.3%
10%	85%	15%	60.8%
10%	85%	25%	67.3%

considered (Table 1). Changes in the individual-level prevalence for *F. hepatica* had little effect on the power, except when the farm-level prevalence was higher and the VTEC O157 farm-level prevalence low. Increases in the farm-level prevalence for *F. hepatica* were also commensurate with increases in statistical power (Table 1).

4.3. Example

We generated a single synthetic dataset, summarised in Fig. 1, representing a total of 7596 cows from 270 farms, with between 1 and 99 cows sampled from each farm (mean = 28.0, median = 22.5). The VTEC O157 prevalence

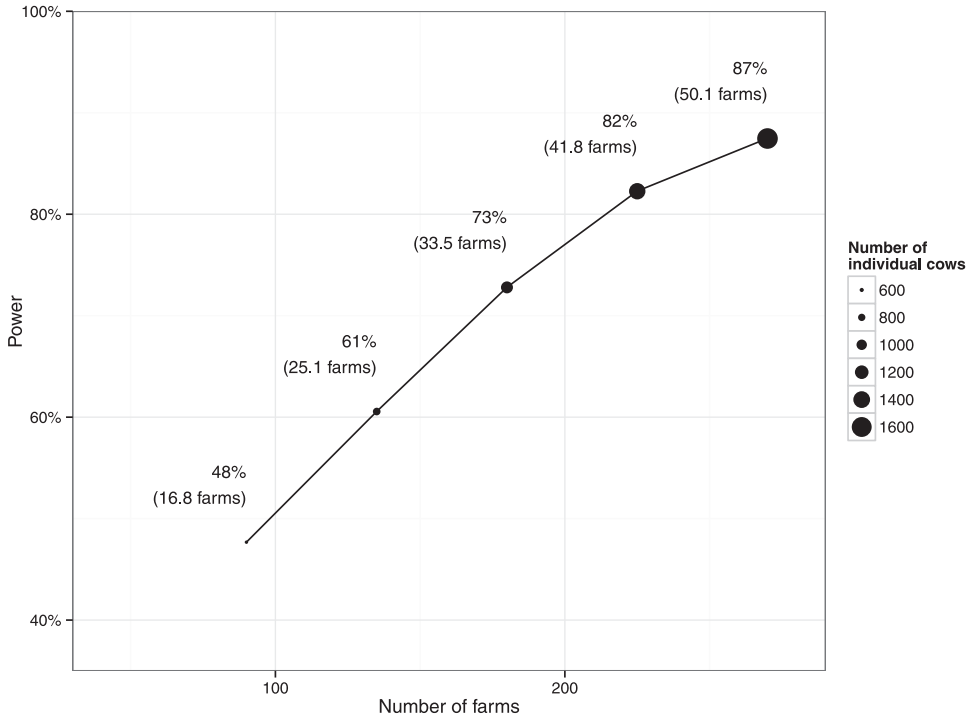

Fig. 3. Power curve to detect an odds ratio of 2 (equivalently $\beta = \log(2)$) for a positive *F. hepatica* test for varying number of farms available for testing. The horizontal axis denotes the total number of farms undergoing VTEC O157 testing, with the actual number of farms undergoing *F. hepatica* testing shown in parentheses.

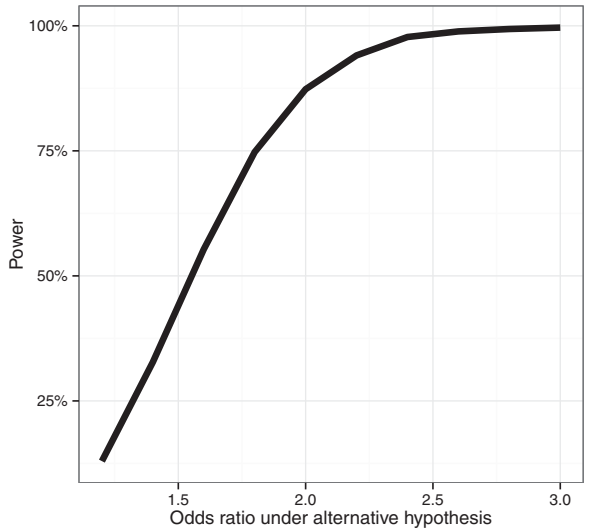

Fig. 4. Power curve as a function of the odds ratio (OR) for detection under the alternative hypothesis. The analysis is based on first performing VTEC O157 testing on all $M = 270$ farms.

was 21.9% at the farm-level and 5.5% at the individual-level. Similarly, *F. hepatica* prevalence was 81.5% at the farm-level and 21.3% at the individual-level. A total of 59 farms had a VTEC O157 prevalence >0% and <100%. Fitting the above model to these data (Eq. (1)), we estimate the odds ratio for a positive *F. hepatica* test result to be 2.36 (95% CI: 1.66-3.37; $P < 0.001$), in which case we would reject the null hypothesis of no association of co-infection.

5. Discussion

The design of an experiment should always consider the number of test samples required. A dedicated stand-alone study to examine whether there is an association between the co-infection of *F. hepatica* and VTEC O571 is probably not feasible in the absence of any pilot data, as the costs would be prohibitive. A unique element of the study here is that the fluke testing is performed *post hoc* to the primary FSA funded study, which incorporates the testing of VTEC O157. The consequence of this is that the desired power can be achieved whilst reducing the number of samples that require testing by approximately 80%, thus reducing the costs.

Trimming the data by excluding farms with either 0% or 100% sample prevalence for VTEC O157 confers a substantial reduction in cost but could, in principle, invalidate the subsequent analysis. For this study we confirmed by simulation that the analysis gives approximately the correct significance level (between 4.2% and 5.2% of datasets simulated under the null hypothesis that $\beta = 0$ rejected this hypothesis at the nominal 5% level). Also, bias in estimation of β was small; with 270 farms available for testing and 2500 simulated synthetic datasets, trimming resulted in a negligible bias of -0.032, whereas estimation using the complete dataset had smaller bias (-0.010). In general, the accuracy of nominal significance levels, and bias in estimation of β, when using our approach will be application

dependent. We recommend that users of the approach verify the properties of the study by simulation using scenarios tailored to each application.

There are several other limitations to this study. First, the exact sample sizes of each farm being tested by the FSA were unknown during the analysis. By replacing these with a suitable sampling distribution determined from existing data, a decision on the study protocol was possible. However, farm sizes might have changed over the past decade. An increase in average farm size will be commensurate with an increase in statistical power, but also resource demands, as it would be expected that there will be fewer farms with zero VTEC O157 prevalence excluded from the analysis. Second, we have assumed that the prevalence estimates of VTEC O157, which are based on data >10 years old, have remained constant at both the individual- and farm-level, and that the distribution in Scotland is similar in England and Wales. We showed that the statistical power would be sensitive to a lower farm-level prevalence value for VTEC O157, but are otherwise quite robust. As we randomly simulated datasets, we note that natural sampling variation in the prevalence values was accounted for. Thirdly, we ignored farm-level explanatory variables and spatiotemporal effects. However, as noted, the additional adjustment variables would potentially reduce the standard error of the effect size estimate; hence the current setup should be considered as conservative. Finally, we have discounted the issue of the test sensitivities and specificities, effectively assuming perfect tests (100% sensitivity and 100% specificity for both tests). In practice, this assumption might not hold and the power calculations would require further complexity, similar to that performed for prevalence estimation using imperfect tests (Diggle, 2011).

Simple formulae are not available for sample size calculations beyond the most basic of statistical tests. The incorporation of random-effects into the model immediately introduces complexity. Simulation-based sample size calculations can be tailored to any hypothesised data-generating mechanism and analysis protocol where either pilot data or prevalence estimates are obtainable, and are therefore applicable to any prospective experimental design.

6. Conclusions

We have developed a feasible experimental design to test whether infection with *F. hepatica* and VTEC O157 in cattle are independent, which exploits an existing comprehensive study and achieves statistical power of 87%. This statistical approach might be suitable for other co-infection studies.

Acknowledgements

VTEC O157 data used in this study was collected as part of the International Partnership Research Award in Veterinary Epidemiology (IPRAVE), *Epidemiology and Evolution of Enterobacteriaceae Infections in Humans and Domestic Animals* funded by the Wellcome Trust.

FS101055. *E. coli* O157 super-shedding in cattle and mitigation of human risk is funded by the U.K. FSA website at http://www.food.gov.uk/science/research/foodborneillness/ecoliresearch/fs101055/.

Appendix A.

Calculation of Beta distribution shape parameters for farm sample sizes

We assume q_j is a Beta distributed variable with shape parameters a and b. Since each farm must have at least one cow sampled, we randomly sample $N_j - 1$ from a Binomial distribution with size $113 - 1 = 112$ and event probability q_j. By fixing the mean of the distribution, which equals $a/(a + b)$, to be $(27 - 1)/112$, we obtain the condition $86a = 26b$. Furthermore, if we fix the approximate median, as given by Kerman (2011), to be $(23 - 1)/112$, we obtain the condition $90a = 22b + 68/3$. Solving this simultaneous pair of equations admits the solution $a = 1.32$ and $b = 4.35$.

Calculation of Beta distribution shape parameters for *F. hepatica* infection

Let X_{ij} denote a random variable that equals 1 if cow i on farm j tests positive for *F. hepatica*, and 0 otherwise. Also, let $X_j = X_{1j} + X_{2j} + \cdots + X_{n_j j}$, where n_j is the known total number cows tested on farm j. It is believed that the individual-level and farm-level *F. hepatica* prevalence are approximately 20% and 80% respectively, which is equivalent to $P[X_{ij}] = 0.20$ and $P[X_j > 0] = 0.80$. By the law of total expectation we have, for sufficiently large n_j:

$$E_r[E_X[X_{ij}|r_j]] = 0.20 \tag{2}$$

$$E_r[P[X_j > 0|r_j]] = 0.80 \tag{3}$$

where the outer expectation is taken with respect to a Beta distributed random variable with shape parameters u and v. For Eq. (2), the inner expectation reduces to the mean of a Bernoulli random variable, namely r_j, so that the outer expectation reduces to the expectation of a Beta distributed random variable, namely $u/(u + v)$. Hence, we can re-write the equation as $v = 4u$. For Eq. (2), we can rewrite the inner probability as $1 - P[X_j = 0]$, which is given by $1 - (1 - r_j)^{n_j}$. Taking the expectation of this term with respect to the distribution of r_j, we obtain:

$$1 - \frac{1}{B(u, v)} \int_0^1 r^{u-1}(1 - r)^{v+n_j-1} dr = 0.80$$

which reduces to $5B(u, v + n_j) = B(u, v)$, where $B(x, y)$ is the Beta function. Combining this with the condition from Eq. (2), we obtain the simultaneous condition that $5B(u, 4u + n_j) = B(u, 4u)$, which, conditional on observed n_j, can be solved for u, and subsequently v can be retrieved.

The solution to this system of equations depends on n_j. In general, there are no solutions for this equation for $n_j < 8$, which is a problem here; however such cases are permissible in this study. In practice, however, we overcome this by noting that the constraints only need to hold true *on average*. Therefore we replace n_j with the random variable N_j (as described earlier) in the equation above, and take the

expectation to obtain:

$$\sum_{x=0}^{112}\binom{112}{x}\frac{B(x+a,112-x+b)}{B(a,b)}B(u,4u+x+1)$$
$$=0.2B(u,4u)$$

where $a=1.32$ and $b=4.35$ are the shape parameters calculated for sampling from N_j. Hence, the solution, determined by numerical optimisation, is found to be $u=0.99$ and $v=3.97$.

Calculation of random-effects distribution parameters for VTEC O157

Let Y_{ij} denote a random variable that equals 1 if cow i on farm j tests positive for VTEC O157, and 0 otherwise. Also, let $Y_j = Y_{1j} + Y_{2j} + \cdots + Y_{n_j j}$, where n_j is the total number cows tested on farm j. Current data suggests that the marginal individual-level and farm-level VTEC O157 prevalence are 4% and 19% respectively, which is equivalent to $P[Y_{ij}]=0.04$ and $P[Y_j>0]=0.19$. However, each Y_{ij} depends on farm-level random effect, the co-infection effect size and test result for *F. hepatica* on the same cow. We exploit knowledge about the sampling distribution for the *F. hepatica* infection, and using Eq. (1), we have, conditional on the farm-level effects:

$$E_X[E_Y[Y_{ij}|\alpha_j,\beta,X_{ij}]]=0.8\left(\frac{e^{\alpha_j}}{1+e^{\alpha_j}}\right)+0.2\left(\frac{e^{\alpha_j+\beta}}{1+e^{\alpha_j+\beta}}\right)$$
$$E_X[E_Y[P[Y_j>0|\alpha_j,\beta,X_{ij}]]]=1-\left[0.8\left(\frac{1}{1+e^{\alpha_j}}\right)+0.2\left(\frac{1}{1+e^{\alpha_j+\beta}}\right)\right]^{n_j} \quad (4)$$

We next average each of these over the distribution of $\alpha\sim N(\mu,\sigma^2)$. Additionally, as Eq. (4) also depends on n_j, we again average over its sampling distribution to yield:

$$1-\sum_{x=0}^{112}\binom{112}{x}\frac{B(x+a,112-x+b)}{B(a,b)}\int_{-\infty}^{\infty}\phi(u|\mu,\sigma^2)$$
$$\times\left[0.8\left(\frac{1}{1+e^{\alpha_j}}\right)+0.2\left(\frac{1}{1+e^{\alpha_j+\beta}}\right)\right]^{x+1}du$$

where $\phi(u|\mu,\sigma^2)$ is the probability density function for a normally distributed random variable with mean μ and standard deviation σ, and the 0.8 and 0.2 multipliers denote $P[X_{ij}=0]$ and $P[X_{ij}=1]$ respectively. Equating these equations to 0.04 and 0.19 respectively, and solving by numerical integration yields $(\mu,\sigma)=(-7.09,3.52)$.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.prevetmed.2015.02.022.

References

Armstrong, G.L., Hollingsworth, J., Morris, J.G., 1996. Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. Epidemiol. Rev. 18, 29–51.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. 24, 127–135.

Brady, M.T., O'Neill, S.M., Dalton, J.P., Mills, K.H., 1999. *Fasciola hepatica* suppresses a protective Th1 response against *Bordetella pertussis*. Infect. Immun. 67, 5372–5378.

Chase-Topping, M.E., 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.

Claridge, J., Diggle, P., McCann, C.M., Mulcahy, G., Flynn, R., McNair, J., Strain, S., Welsh, M., Baylis, M., Williams, D.J.L., 2012. *Fasciola hepatica* is associated with the failure to detect bovine tuberculosis in dairy cattle. Nat. Commun. 3, 853.

Corbishley, A., Ahmad, N.I., Hughes, K., Hutchings, M.R., McAteer, S.P., Connelley, T.K., Brown, H., Gally, D.L., McNeilly, T.N., 2014. Strain dependent cellular immune responses in cattle following *Escherichia coli* O157:H7 colonisation. Infect. Immun. 82, 5117–5131.

Diggle, P.J., 2011. Estimating prevalence using an imperfect test. Epidemiol. Res. Int., 2011.

Flynn, R.J., Mannion, C., Golden, O., Hacariz, O., Mulcahy, G., 2007. Experimental *Fasciola hepatica* infection alters responses to tests used for diagnosis of bovine tuberculosis. Infect. Immun. 75, 1373–1381.

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York.

Gunn, G.J., McKendrick, I.J., Ternent, H.E., Thomson-Carter, F., Foster, G., Synge, B.A., 2007. An investigation of factors associated with the prevalence of verocytotoxin producing *Escherichia coli* O157 shedding in Scottish beef cattle. Vet. J. 174, 554–564.

Kerman, J., 2011. A Closed-form Approximation for the Median of the Beta Distribution. arxiv:1111.0433.

Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis With Missing Data, Wiley Series in Probability and Mathematical Statistics. Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

McCann, C.M., Baylis, M., Williams, D.J.L., 2010a. The prevalence and spatial distribution of *Fasciola hepatica* infected dairy herds in England and Wales. Vet. Rec. 166, 612–617.

McCann, C.M., Baylis, M., Williams, D.J.L., 2010b. The development of linear regression models using environmental variables to explain the spatial distribution of *Fasciola hepatica* infection in dairy herds in England and Wales. Int. J. Parasitol. 40, 1021–1028.

Naylor, S.W., Low, J.C., Besser, T.E., Mahajan, A., Gunn, G.J., Pearce, M.C., McKendrick, I.J., Smith, D.G.E., Gally, D.L., 2003. Lymphoid follicle-dense mucosa at the terminal rectum is the principal site of colonization of enterohemorrhagic *Escherichia coli* O157:H7 in the bovine host. Infect. Immun. 71, 1505–1512.

Pearce, M.C., Chase-Topping, M.E., McKendrick, I.J., Mellor, D.J., Locking, M.E., Allison, L., Ternent, H.E., Matthews, L., Knight, H.I., Smith, A.W., Synge, B.A., Reilly, W., Low, J.C., Reid, S.W.J., Gunn, G.J., Woolhouse, M.E.J., 2009. Temporal and spatial patterns of bovine *Escherichia coli* O157 prevalence and comparison of temporal changes in the patterns of phage types associated with bovine shedding and human *E. coli* O157 cases in Scotland between 1998–2000 and 2002–2004. BMC Microbiol. 9, 276.

Pearce, M.C., Fenlon, D., Low, J.C., Smith, A.W., Knight, H.I., Evans, J., Foster, G., Synge, B.A., Gunn, G.J., 2004. Distribution of *Escherichia coli* O157 in bovine fecal pats and its impact on estimates of the prevalence of fecal shedding. Appl. Environ. Microbiol. 70, 5737–5743.

Salimi-Bejestani, M.R., Daniel, R.G., Felstead, S.M., Cripps, P.J., Mahmoody, H., Williams, D.J.L., 2005. Prevalence of *Fasciola hepatica* in dairy herds in England and Wales measured with an ELISA applied to bulk-tank milk. Vet. Rec. 156, 729–731.

Soetaert, K., Herman, P.M.J., 2009. A Practical Guide to Ecological Modelling. Using R as a Simulation Platform. Springer.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer, New York.