

The distribution of the pathogenic nematode *Nematodirus battus* in lambs is zero-inflated

M. J. DENWOOD^{1*}, M. J. STEAR², L. MATTHEWS¹, S. W. J. REID¹, N. TOFT³
and G. T. INNOCENT¹

¹ *Comparative Epidemiology and Informatics, Institute for Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow, Bearsden Road, Glasgow G61 1QH, UK*

² *Veterinary Genes and Proteins Group, Institute for Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow, Bearsden Road, Glasgow G61 1QH, UK*

³ *Danish Meat Association, Vinkelvej 11, DK-8620 Kjellerup, Denmark*

(Received 4 February 2008; revised 6 April and 16 May 2008; accepted 16 May 2008; first published online 14 July 2008)

SUMMARY

Understanding the frequency distribution of parasites and parasite stages among hosts is essential for efficient experimental design and statistical analysis, and is also required for the development of sustainable methods of controlling infection. *Nematodirus battus* is one of the most important organisms that infect sheep but the distribution of parasites among hosts is unknown. An initial analysis indicated a high frequency of animals without *N. battus* and with zero egg counts, suggesting the possibility of a zero-inflated distribution. We developed a Bayesian analysis using Markov chain Monte Carlo methods to estimate the parameters of the zero-inflated negative binomial distribution. The analysis of 3000 simulated data sets indicated that this method out-performed the maximum likelihood procedure. Application of this technique to faecal egg counts from lambs in a commercial upland flock indicated that *N. battus* counts were indeed zero-inflated. Estimating the extent of zero-inflation is important for effective statistical analysis and for the accurate identification of genetically resistant animals.

Key words: *Nematodirus battus*, *Ovis aries*, distribution, Bayesian MCMC, faecal egg count, Zero-inflated negative binomial.

INTRODUCTION

Nematodirus battus is one of the most pathogenic organisms that infect sheep in cool temperate climates. Infected lambs develop acute enteritis with watery diarrhoea accompanied by inappetence and weight loss. Animals can die if the disease is not controlled by anthelmintic treatment (Armour and Coop, 1991). The life-history for *N. battus* is unusual because larvae develop to the infective third stage within the egg and these eggs only hatch when the temperature exceeds 10 °C following a cold spell. The simultaneous hatching of large numbers of infective larvae poses a severe disease risk. Historically, disease occurs in late spring in those years with suitable weather. However, more recently, warmer and more variable weather suggests that outbreaks could also occur in late autumn.

Despite its importance, *N. battus* has received surprisingly little attention. In particular, the distribution of parasites and egg output among hosts following natural infection is unknown, although the

distribution of *Nematodirus* spp. (predominantly *Nematodirus spathiger*) in 104 three-month-old Australian Merino lambs was similar to a negative binomial distribution (Barger, 1985). Mixed model analyses indicated that the heritability of faecal egg count following natural infection with *Nematodirus* spp. was similar to the heritability of egg counts due to other strongyle species in New Zealand Coopworth and Romney sheep (McEwan *et al.* 1995), higher for *Nematodirus* spp. than other strongyles in Scottish Blackface sheep (Bishop *et al.* 2004) and lower for *Nematodirus* spp. than other strongyles in selected lines of New Zealand Romney sheep (Morris *et al.* 2004). In all cases there were strong positive genetic correlations between *Nematodirus* spp. egg counts and those of other nematode species, suggesting that similar mechanisms underlay resistance to, or control of, all sheep nematodes, and that sheep could be simultaneously selected for resistance to *Nematodirus* spp. and other species.

However, an accurate knowledge of the distribution of egg counts and adult nematodes among hosts is essential for efficient experimental design, and analysis of the resulting data. Using an inappropriate distribution increases the risk of both false positive and false negative results (type I and type II errors). An inappropriate model of the frequency

* Corresponding author: Comparative Epidemiology and Informatics, Institute for Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow, Bearsden Road, Glasgow G61 1QH, UK. E-mail: m.denwood@vet.gla.ac.uk

distribution could lead to inaccurate estimates of variance components and inefficient selection schemes. In addition, the frequency distribution influences the impact of parasites upon the host population (Anderson and May, 1991) and influences the success of potential control measures such as selective breeding. Further, mathematical procedures that model the frequency distribution of egg counts can be used to improve the detection of anthelmintic resistance (Torgerson *et al.* 2005) and to determine the optimal time to treat grazing livestock (Morgan *et al.* 2005). Therefore an accurate description of the distribution among hosts of faecal egg counts is likely to improve the diagnosis and treatment of parasitic disease. The negative binomial distribution has been widely used to describe parasite distributions, but more recently zero-inflated models have also been suggested (Nødtvedt *et al.* 2002). These models are appropriate if the observed distribution results from a mixture of lambs that have not been exposed (zero distribution), and lambs that have been exposed (negative binomial distribution) to infection. An accurate and precise method to determine the frequency distribution should have widespread application among quantitative parasitologists.

The purposes of this study were to develop a suitable procedure to determine the frequency distribution of faecal egg counts following natural infection with *N. battus* and to determine the distribution of egg production among lambs.

MATERIALS AND METHODS

Experimental plan

The negative binomial distribution has a variety of parametrizations. The parametrization most commonly used by parasitologists characterizes the distribution by 2 parameters: the mean (m) and the shape (k), which is an inverse measure of aggregation. In this model, the variance is given by $m + m^2/k$ (McCullagh and Nelder, 1989). We used the ratio m/k (the scale parameter, θ) to represent dispersion, which is equivalent to the extra-Poisson variance to mean ratio suggested by Elliot (1977) as a measure of aggregation. θ is positively related to the degree of aggregation, so that as θ tends towards 0 the distribution converges on the Poisson. The variance in our model is therefore given by $m(1 + \theta)$, and the extra-Poisson variance by $m\theta$. The apparent prevalence of *Nematodirus* infections is always less than 1, therefore a zero-inflated negative binomial model was explored. All faecal egg counts greater than 0 are derived from the negative binomial distribution, but each of the zero counts can arise from either the negative binomial or the zero distributions, with a probability of belonging to the zero distribution equal to the zero-inflation parameter. This probability is

estimated by comparison of the observed number of zeros in the data to the expected number of zeros from a negative binomial distribution with given values for mean and dispersion. The zero-inflated negative binomial distribution therefore has 3 parameters; mean, scale and percentage zero-inflation. Parasitologists have usually estimated the mean and shape parameters by maximum likelihood (Bliss and Fisher, 1953). Other estimators also exist including method of moments, maximum quasi-likelihood and conditional likelihood (van de Ven, 1993). A simulation study suggested that maximum likelihood or conditional likelihood were the best-performing estimators (van de Ven, 1993); however, the desirable properties of the maximum likelihood estimator are most apparent at large sample sizes. Therefore we developed a method for Bayesian analysis with Markov chain Monte Carlo (MCMC) and compared this to a maximum likelihood method on simulated data. The best-performing procedure was then used to estimate the parameters in real data sets.

Animals

All sheep were straight-bred Scottish Blackface sheep from a commercial upland farm in Southwest Strathclyde. All husbandry procedures followed usual commercial practice and have been described previously (Stear *et al.* 1998). All lambs were kept on the same field after weaning at 3 or 4 months of age until the end of the study. To maintain health and productivity, all lambs were given anthelmintic (albendazole sulphoxide at the dose rate recommended by the manufacturer) every 28 days until 6 or 7 weeks before necropsy. Cohorts of 200 lambs were sampled monthly during the summer grazing season of 5 consecutive years to avoid bias due to unusually heavy or light infections in specific years. After removal of lambs due to death, missing records, lost tags or no rectal faecal sample, a mean of 169.4 (88–192) counts for 5 ($n=3$) or 6 ($n=2$) consecutive months between May and October were obtained. The lambs were the offspring of 39 sires and 496 dams.

Parasitological procedures

In addition to the monthly faecal egg counts performed on all 1000 animals, 530 animals were examined post-mortem at 6.5 months of age, in late October and early November. For the first 2 years, only female animals were sampled as all males were sent to slaughter. In the 3rd and 4th years, animals were selected on the basis of faecal egg counts, with individuals with less extreme egg counts being selected for post-mortem analysis. No animals from the 5th year of study were examined post-mortem. Standard procedures were used to estimate the concentration of nematode eggs per gram of faeces and to

enumerate fourth-stage, immature and adult fifth-stage nematodes in the abomasum and small intestine (Armour *et al.* 1966; Stear *et al.* 1998). Although it is possible to differentiate adult *N. battus* from other *Nematodirus* sp., this is not usually done. However, *N. battus* eggs were recorded separately.

Simulated data

In total, 1000 datasets were simulated for each sample size of 10, 100, and 1000 counts, using R (R Development Core Team, 2007). Each data set was generated using a zero-inflated negative binomial model with parameters randomly chosen from between 0 and 100 for mean count, 0 and 0.75 for the proportion of the population that is unexposed to infection (zero-inflation), and 0.01 and 100 for dispersion. The mean was sampled from a log uniform distribution to increase the number of datasets with lower mean counts. Dispersion was also sampled from a log uniform distribution to increase the number of datasets that were close to being Poisson distributed, since with unknown zero-inflation a small amount of dispersion with a low mean count is more difficult to accurately quantify than a large amount of dispersion with a high mean count. Zero-inflation was sampled from a uniform distribution. A total of 44 out of the 1000 simulated datasets for sample size 10 (but none of the sample size 100 or 1000) contained only zero counts and so were excluded from analysis. Treating the mean egg count and dispersion as independent could produce datasets that are unlikely to be observed in real faecal egg count data. However, since both models use the same distribution to fit data, it is unlikely that the conclusions would be any different if the two parameters were simulated with a dependency.

Statistical models

Each remaining dataset was analysed with MCMC methods using a zero-inflated negative binomial model in WinBUGS (Spiegelhalter, 2003). The model was run using 2 independent Markov chains, which effectively perform the simulation twice to ensure that the 2 sets of results have converged on the same distribution for each parameter. Each independent run of the simulation is expected to produce similar results when the chains have converged (Toft *et al.* 2007). The model description and additional details concerning the prior distributions and initial values can be found in Appendix A.

Convergence was assessed by calculating the Gelman-Rubin statistic, which compares the variance between chains to the variance within the chains (Gelman and Rubin, 1992). Given adequate convergence the 2 values should be similar, if they were equal the Gelman-Rubin statistic would equal 1. Convergence analysis was performed for all datasets

using the 'gelman.diag' function of the 'coda' package (Plummer *et al.* 2006) in R. Analyses were classified as converged if the point estimate of the Gelman-Rubin statistic for each of the 3 monitored variables was below 1.05.

The first 5000 iterations were discarded (burn in), before sampling from 10 000 iterations and saving the output to coda files. The MCMC model returned the error warning 'undefined real result' for a number of datasets after completing a varying number of iterations. This appeared to be associated with the Gibbs sampler attempting to sample from extreme values for dispersion. Datasets that achieved less than 1000 sampling iterations before this occurred (not including the 5000 iterations burn-in period) were rejected; the remainder were analysed for convergence along with successfully completed datasets. Outputs that failed this criterion were re-run with 100 000 and then, if necessary, 500 000 iterations in an attempt to improve convergence. The results for any datasets that failed to achieve convergence were removed. The outputs for the MCMC coda files were then analysed using R to determine the median value for the mean egg count and the 95% credible interval.

Maximum likelihood analysis used the 'zicounts' package (Mwalili) in R. The maximum likelihood analysis did not produce results for some datasets. In addition, for some datasets, the standard error of one or more parameters was returned as impossible to calculate or infinite. Since a maximum likelihood estimate with an unknown confidence interval is of little use, these results were also removed.

R was used to compare the maximum likelihood and Bayesian MCMC analyses. To assess the accuracy of the MCMC median and maximum likelihood estimates, the mean bias was calculated using the simulated (true population) value for each parameter. The error of the estimate was calculated in each case by subtracting the log of the simulation parameter from the log of the estimate. The mean of these estimate errors was taken as the mean bias. Because subtracting the log of the simulation parameter from the log of the estimate is equivalent to taking the log of the estimate divided by the simulation parameter, the resultant bias is relative to the simulation parameter.

The mean value presented is the mean of the negative binomial distribution describing the infected animals, multiplied by 1 – the proportion of zero-inflated animals. It is similar to the arithmetic mean of all the animals, but the two will not be identical since the distributional mean takes into account variation due to sampling, and the arithmetic mean is merely the mean of a single sample. For the Bayesian MCMC analyses the mean value was estimated from the median of the mean value over the different iterations. The dispersion parameter was calculated as the scale parameter (equal to the mean divided by k).

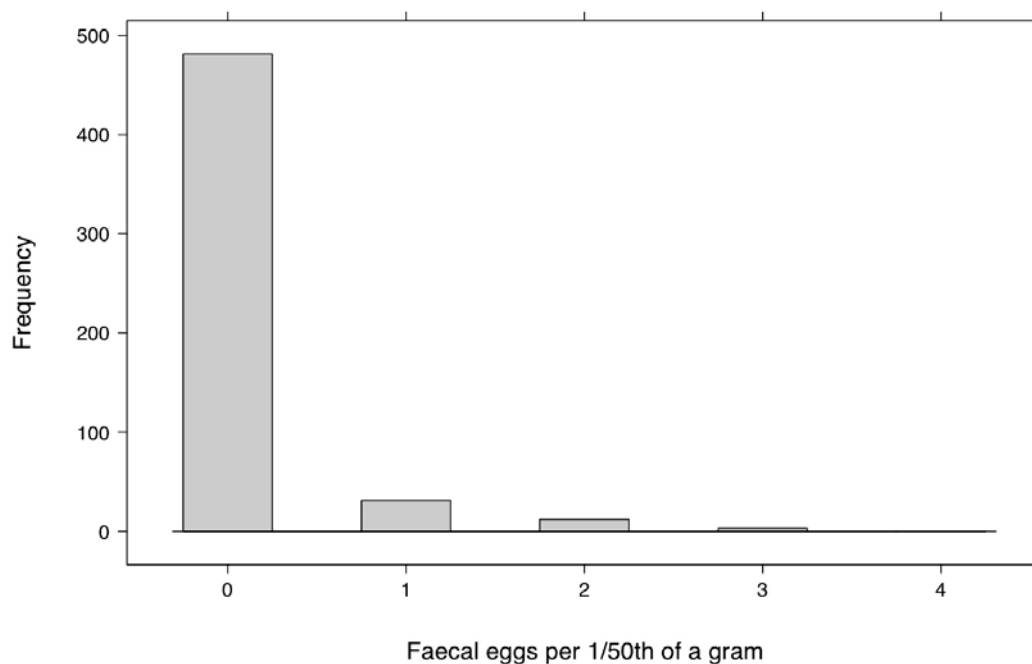


Fig. 1. The distribution of *Nematodirus battus* eggs in the faeces of 530 Scottish Blackface lambs at post-mortem examination over a 4-year period.

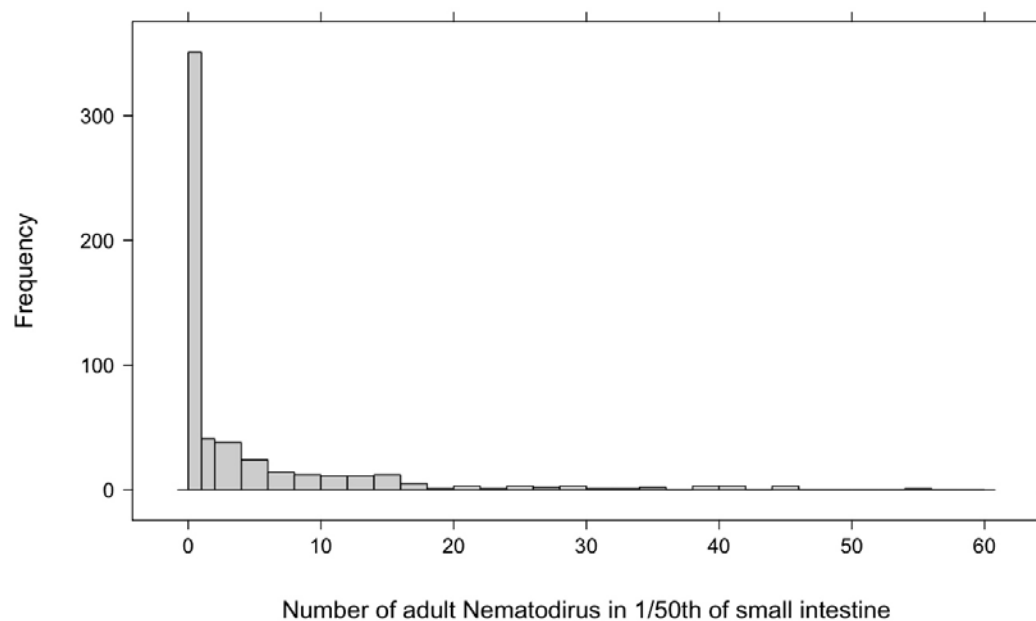


Fig. 2. The distribution of adult *Nematodirus* sp. at post-mortem examination of 530 Scottish Blackface lambs over a 4-year period.

RESULTS

Fig. 1 shows the distribution of *N. battus* eggs in the faeces of these lambs at necropsy. The majority of animals have zero egg counts indicating a relatively low prevalence of *N. battus* infection in early November. Of the 530 animals tested, only 51 (9.6%) had egg counts above zero.

Fig. 2 shows the combined distribution of *Nematodirus* sp. (predominantly *N. battus*, *N. filicollis*,

N. spathiger) in 6.5-month-old lambs from a commercial farm. The data were collected from 110, 100, 170 and 150 lambs that were necropsied in 1992, 1993, 1994 and 1995 respectively. *Nematodirus* spp. were detected in 34.9% (185/530) of the lambs necropsied.

The comparison of Bayesian MCMC methods and maximum likelihood showed that the percentage of simulated data sets that were analysed successfully was higher for MCMC than ML (Table 1). This

Table 1. Comparison of maximum likelihood (ML) and Bayesian Monte Carlo Markov chain methods (MCMC) for zero-inflated negative binomial simulated data sets

	Sample size 10 MCMC	Sample size 10 ML	Sample size 100 MCMC	Sample size 100 ML	Sample size 1000 MCMC	Sample size 1000 ML
Analysis did not return usable results	17%	42%	4%	25%	2%	22%
True value was outside the 95% CI for mean count	5%	15%	5%	9%	6%	8%
True value was outside the 95% CI for dispersion	4%	26%	5%	15%	6%	10%
True value was outside the 95% CI for zero-inflation	3%	8%	5%	12%	6%	9%

difference was especially marked with a sample size of 10 where 42% of the datasets analysed by maximum likelihood did not produce meaningful estimates, compared to only 17% of the data sets analysed by Bayesian MCMC methods. Table 1 also shows that there was a larger proportion of maximum likelihood analyses where the 95% confidence interval did not contain the true (simulated) value. This was especially marked at the smallest sample size, where the maximum likelihood analysis returned the true value for the mean in only 85% of cases and dispersion in 74% of cases, although the true value for zero-inflation was identified in 92% of cases. The MCMC model returned the true value 94–97% of the time in all cases, compared to the value of 95% that would be expected from a 95% credible interval.

Fig. 3 compares the mean bias in the estimates for mean egg count, the dispersion parameter and zero-inflation obtained by maximum likelihood and Bayesian MCMC analyses of the simulated data sets. Maximum likelihood estimates were, on average, more biased than the Bayesian MCMC estimates for all except the estimates of zero-inflation in the data sets of size 1000.

The Bayesian MCMC method was then used to analyse faecal egg counts from Scottish Blackface lambs on a commercial farm in Strathclyde. Table 2 shows that all the data sets from the naturally infected animals showed evidence of zero-inflation, with median estimates ranged from 2.5 to 86%, and lower 95% credible intervals ranged from 0.1 to 82.5% (data not shown). The mean number of counted eggs in the lambs ranged from 0.02 in October 1996 to 9.8 in June 1995. As each egg counted represented 50 eggs per gram, mean egg output ranged from a low of 1 egg per gram to a high of 490 eggs per gram. For comparison the arithmetic mean egg production ranged from 1.5 to 440 eggs per gram (data not shown). The degree of dispersion also showed considerable temporal variation and appeared to be higher at increased levels of infection, while at low intensities of infection, the estimates of dispersion were close to zero (Fig. 4). Spearman's rank correlation coefficient between the estimates of

the mean and the dispersion parameter was 0.63 ($P < 0.001$). The May 1994 counts had a relatively high mean egg count (5.3) and a very high estimate for the scale parameter (43). Even after excluding this sample, the rank correlation was still highly significant at 0.59 ($P < 0.002$). There was less evidence for a relationship between the sample mean and the estimate of zero-inflation (Spearman's rank correlation = -0.34 ; $P = 0.085$).

DISCUSSION

The number of lambs with demonstrable *N. battus* eggs in their faeces, or with *Nematodirus* sp. in their intestines was always less than 100%. This differs from *Teladorsagia circumcincta* (Stear *et al.* 1998) but is typical of most parasitic infections, and supports the possibility that the distribution of *N. battus* contains a zero-inflation component. We developed a Bayesian MCMC method to estimate the parameters of the zero-inflated negative binomial distribution. This method performed better than maximum likelihood in simulated data sets of size 10 to 1000. Analysis of faecal egg counts from lambs on a commercial farm indicated that most *Nematodirus* distributions were zero-inflated and the median estimate of zero-inflation ranged from 2.5 to 86%. There was also considerable variation over time in mean egg counts from 1 to just below 500 eggs per gram. The frequency distributions at times of low egg output had a variance close to the mean, but showed evidence of greater dispersion when egg output was higher. A repeating yearly pattern was evident, with relatively high estimates for mean count and dispersion seen at the start of the grazing season in May and June which gradually reduced over the summer (data not shown).

Judged by the number of unsuccessful analyses, as well as the failure of the 95% confidence interval to contain the true value and by the mean bias, the Bayesian MCMC method performed better than the ML procedure in the simulated data sets. Although maximum likelihood methods have mainly desirable properties (Edwards, 1992), they are only asymptotically efficient and it is clear that for negative

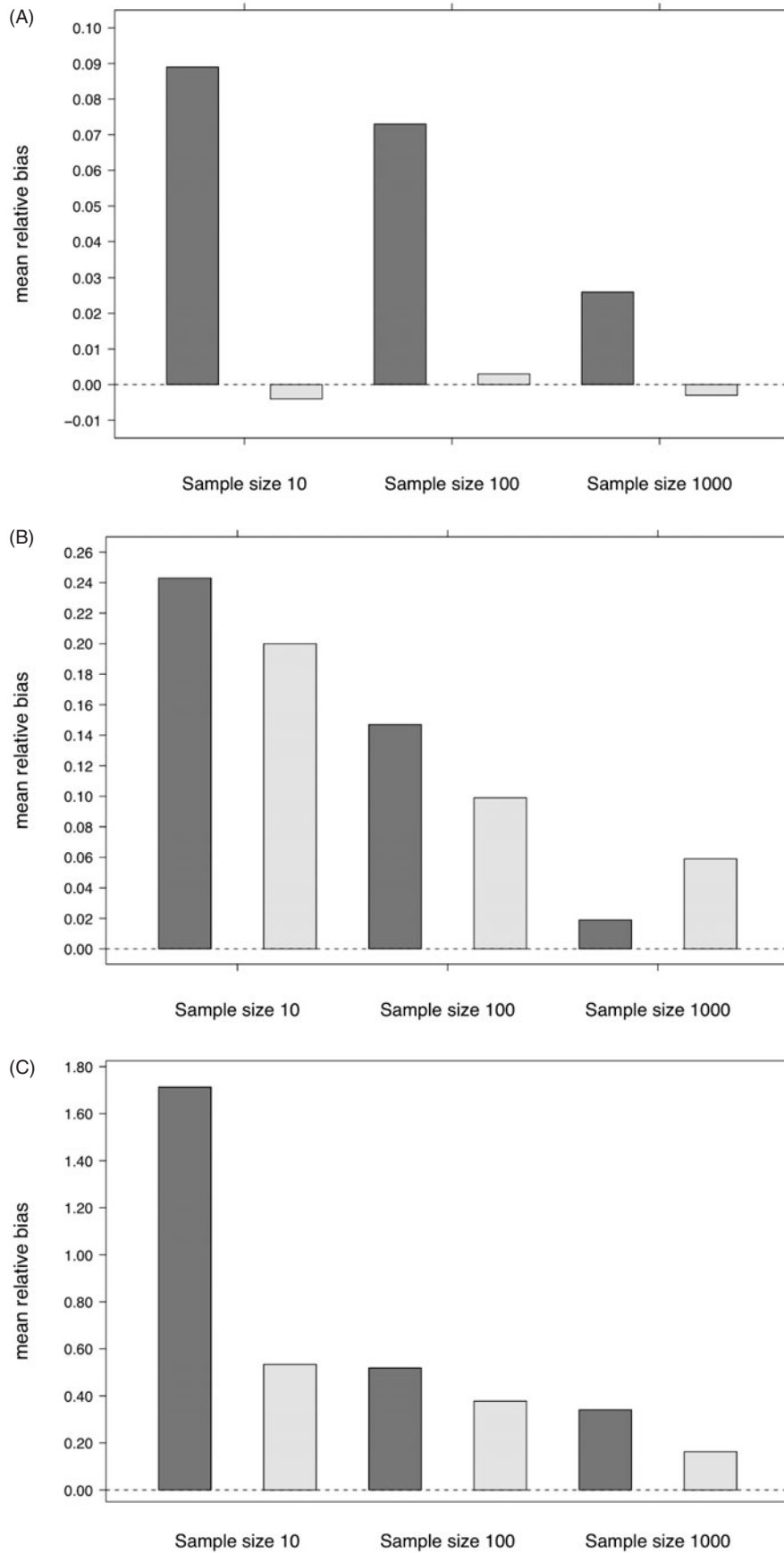


Fig. 3. (A) Mean bias for the estimate of mean count at each sample size. Maximum likelihood shown in dark grey, MCMC median estimate in light grey. (B) Mean bias for the estimate of zero-inflation at each sample size. Maximum likelihood shown in dark grey, MCMC median estimate in light grey. (C) Mean bias for the estimate of the dispersion parameter at each sample size. Maximum likelihood shown in dark grey, MCMC median estimate in light grey.

Table 2. Parameters of the zero-inflated negative binomial distribution in *Nematodirus battus* egg counts

(The mean value in Table 2 is the mean value estimated in the proportion of infected animals multiplied by 1 – the proportion of zero-inflated animals. It is not the arithmetic mean of all the animals. The mean value was the median of the different runs. The dispersion parameter is the mean divided by k.)

Year	Month	Number of lambs	Mean faecal egg count	Dispersion	Zero Inflation (%)
1992	May	118	1.1	2.5	26
1992	June	164	2.0	0.4	2.5
1992	July	152	0.5	0.01	46
1992	August	183	0.3	0.01	34
1992	September	187	0.2	0	27
1993	May	148	0.7	2.6	18
1993	June	183	2.8	2.0	5.6
1993	July	192	0.8	0.03	36
1993	August	176	0.3	0.03	65
1993	September	191	0.03	0	42
1994	May	88	5.3	43	26
1994	June	164	2.8	4.3	17
1994	July	171	3.6	3.3	30
1994	August	178	0.2	0.1	40
1994	September	186	0.06	0.9	9
1994	October	190	0.08	0.6	68
1995	May	108	4.3	0.05	86
1995	June	171	9.8	5.0	7
1995	July	188	1.3	0.2	43
1995	August	185	0.3	0.8	38
1995	September	187	0.05	0	48
1995	October	171	0.1	0.1	71
1996	June	165	5.8	2.5	6.5
1996	July	180	0.9	1.4	10
1996	August	176	0.2	0.01	57
1996	September	182	0.3	0.7	77
1996	October	190	0.02	0.37	3.1

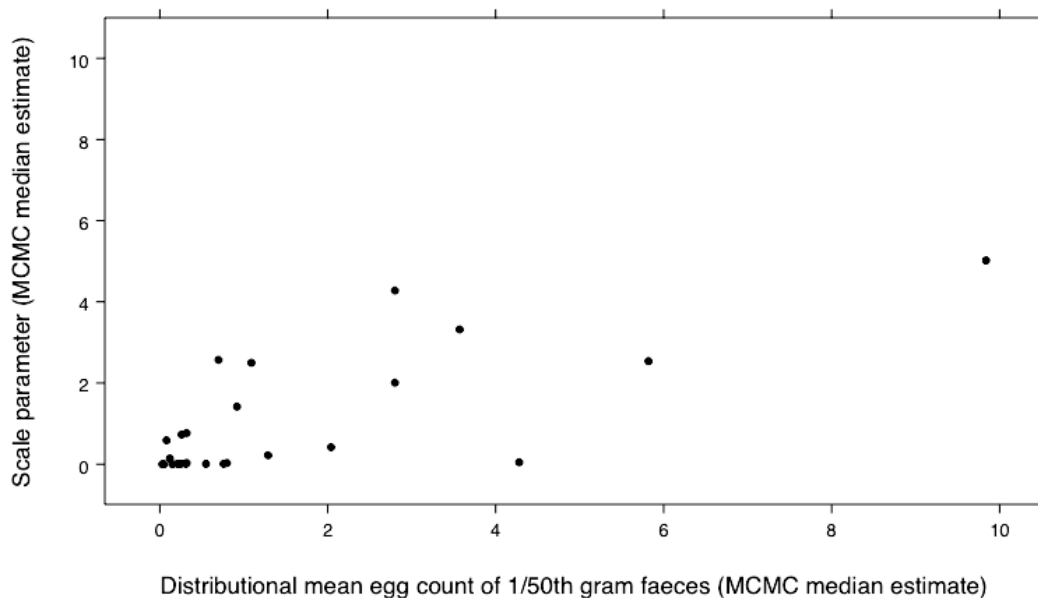


Fig. 4. The relationship between the mean *Nematodirus battus* egg count and the scale parameter for 28 populations of Scottish Blackface lambs (outlier point for May 1994 omitted).

binomial distributions, the approach to the asymptote is slow, especially for the dispersion parameter (van de Ven, 1993; van de Ven and Weber, 1999). MCMC methods have the advantage of not assuming

normality, which is one reason why the technique appears to have out-performed the ML method here. The Bayesian MCMC method is computationally intensive, but the time and effort required is

considerably less than the effort required to collect the parasitological data. However, even with the Bayesian MCMC method, obtaining accurate and precise estimates for the parameters of the zero-inflated negative binomial distribution requires reasonably large and homogeneous data sets. Our analyses of simulated data sets suggests that sample sizes in the order of 100 animals or even larger are desirable. These results suggest that the likelihoods, especially for the zero-inflation and dispersion parameters, were not normally distributed, even at sample sizes of 1000.

Analysis of *N. battus* egg counts indicated that zero-inflated models were more appropriate to describe the distribution of *N. battus* in lamb faeces than models that do not take into account zero-inflation. The large proportion of zero counts in Figs 1 and 2 indicate the possibility that some zero-inflation may be present in the data, and the consistent identification of a zero-inflation component using MCMC methods confirms this in the *N. battus* egg counts, although the relatively low mean egg count increased the uncertainty in zero-inflation estimates. Low egg counts are typical of *N. battus*, especially in mid-summer, but our results may have been influenced by the fact that the lambs were kept in byre (close to the farmer's house) and given anthelmintic every 28 days. The lower 95% confidence interval for zero-inflation was greater than 5% in 9 of the 28 data sets, while the upper 95% confidence interval was greater than 5% in all cases. It is convenient to think of a lower 95% confidence interval of less than 5% as the possibility that the data may not contain an appreciable zero-inflated component, and an upper 95% confidence interval of greater than 5% as the possibility that an appreciable zero-inflated component may be present. On the basis of this, the model retained the possibility that zero-inflation was absent in 10 of the 28 (36%) datasets, while retaining the possibility that zero-inflation was present in 100% of the datasets. This can be compared with an additional study using 100 datasets of 100 egg counts simulated from a negative binomial distribution with similar parameters to those used for the simulation study in this paper, where the lower 95% credible interval for zero-inflation was below 5% in 91 out of 97 (94%) of the successfully converged analyses, and the upper 95% credible interval was above 5% in only 53 (55%) of the analyses (data not shown). The use of zero-inflated models is therefore valid where the extent of zero-inflation (if any) is unknown.

The observed distribution of *N. battus* eggs in faeces can therefore be considered as a mixture of 3 processes: presence or absence of exposure to infection, an underlying distribution of egg production among infected animals (Hunter and Quenouille, 1952), and a Poisson process reflecting variation in the counting process. Zero counts may arise in an individual because it has not been exposed or because

it has been exposed but is producing no eggs, or very few eggs. Very low egg output may not be detected. While it is impossible to distinguish these based on faecal egg counts, it is important to consider both possibilities to avoid misclassification of a nematode-resistant animal. Analyses that ignore zero-inflation may therefore give misleading results, with a tendency to over-estimate the dispersion parameter. To our knowledge, only one group has so far taken zero-inflation into account in the analysis of egg counts, although the method used was not Bayesian MCMC (Nødtvedt *et al.* 2002).

Many parasite species (with the notable exception of *T. circumcincta*) show frequent faecal egg counts of zero, and our results suggest that there would be value in examining these infections for the possibility of zero-inflation. Where there is sufficient information in the data to be confident that zero-inflation is not present, our Bayesian MCMC models return an estimate of approximately zero for the zero-inflation parameter, and the estimates for mean and the dispersion parameter are very similar to the simple negative binomial model. In the presence of zero-inflation, a zero-inflated model returns more accurate estimates for both mean and the dispersion parameter. Unlike the lognormal-Poisson, negative binomial, or continuous distributions such as the Weibull (Gaba *et al.* 2005), zero-inflated models allow a bimodal distribution to be described. This has implications for many aspects of quantitative parasitology, including the identification of lambs that are genetically resistant to nematode infection. An appropriate analysis that takes into account a zero-inflated distribution, such as that applied by Nødtvedt *et al.* (2002) using commercially available software, should be used to analyse zero-inflated data.

The mean faecal egg counts ranged from 1 to 490 eggs per gram. These data are consistent with the view that *N. battus* is present most of the time at a low level but reaches dangerous levels only occasionally. The only deaths due to *Nematodirus* infections on this farm occurred in May and June 1995 during the period of peak intensity. There was also a significant positive relationship between the mean and the scale parameter. At low levels of infection the scale parameter was close to zero, so there was little evidence for overdispersion. As overdispersion tends to zero, the negative binomial converges to the Poisson distribution. As the Poisson distribution describes the uncertainty due to the counting of eggs in the faeces, this implies that at low levels of *N. battus* infection, host variation in grazing behaviour or in protective immune responses do not noticeably contribute to the observed variation in egg output among lambs. Protective immune responses against nematode infections are slow to develop in sheep and cattle and our results support Dineen (1963) who argued that effective immunity requires prolonged and moderately heavy exposure. It is also possible to measure

Table 3. The prior distributions and initial values used for each parameter with the Bayesian MCMC model in WinBUGS

	Prior distribution	Initial value chain 1	Initial value chain 2
Mean.correction	Uniform(0.001,1000)	0.5	2
Log.beta	Uniform(-7,7)	-2.3	2.3
Probability	Uniform(0,1)	0.05	0.95

dispersion as the ratio of the extra-Poisson standard deviation to the mean, which may affect the relationship between the mean and dispersion.

The negative binomial is not the only distribution that can be used to describe the frequency distribution of parasites among hosts. Other distributions may be more appropriate given the biological assumptions, such as the lognormal Poisson distribution, or give a better empirical fit, such as the Weibull (Gaba *et al.* 2005). Now that we can estimate the parameters of the negative binomial distribution more accurately and precisely, and can account for zero-inflation, we can more meaningfully compare the fits of different distributions.

Conclusions

In conclusion, we have developed a Bayesian Markov chain Monte Carlo method to estimate the parameters of the zero-inflated negative binomial distribution. The analysis of 3000 simulated data sets indicated that this method out-performed the maximum likelihood procedure. Analysis of egg counts from lambs in a commercial upland flock indicated that *N. battus* counts were zero-inflated. As egg

output increased, the estimates for the dispersion parameter increased, but at low intensities of egg production, there was little evidence for over-dispersion.

This research was funded by DEFRA: VTRI projects 0101 and 0102.

APPENDIX A: MCMC MODEL DESCRIPTION

A brief description of the MCMC model used is as follows. Each count is defined as originating from a Poisson distribution with a mean value that is unique to each count. These mean values are then multiplied by a variable of either 0 or 1 for each count, which is generated from a Bernoulli trial with a probability value that represents the prevalence of infection within the group. The adjusted mean values are then defined as part of a gamma distribution with a mean and variance that describes the non-zero group. The mean and ratio of mean to variance of the gamma distribution represent mean count and dispersion respectively, and the prevalence is taken as one minus the probability that a count is truly zero.

```

Model {
  for(i in 1: Sample.Size){
    Count[i] ~ dpois(lambda[i])
    lambda[i] <- true.count[i] * corrected.lambda[i]
    corrected.lambda[i] <- gamma[i] * mean.correction
    true.count[i] ~ dbern(Probability)
    gamma[i] ~ dgamma(alpha, beta)
  }
  alpha <- beta * input.mean
  beta <- exp(log.beta)
  # Prior distributions
  mean.correction ~ dunif(0.001,1000)
  log.beta ~ dunif(-7, 7)
  Probability ~ dunif(0,1)
}
Calculated in R from coda files after WinBUGS analysis:
mean.count <- mean.correction * Input.Mean
dispersion <- mean.correction/beta
zero.inflation <- 1 - Probability

```

In this syntax, 'alpha' is defined as 'beta' multiplied by 'Input.Mean', so that the mean of the gamma distribution is defined as the additional data-driven parameter 'Input.Mean', which is the arithmetic mean of the data as calculated prior to Bayesian analysis to the nearest integer with a minimum value of 2. Mean count is therefore equivalent to 'input.mean' multiplied by 'mean.correction', and dispersion as the inverse of 'beta' multiplied by 'mean.correction'.

There are several other methods of specifying a similar model in WinBUGS, such as allowing 'alpha' and 'beta' to vary independently, and by defining the mean of the gamma distribution as one. Each of these 3 specifications achieves generally similar results, although the specification used gives more reliable results when information in the data is poor and/or at extremes of values for dispersion (data not shown). An exhaustive justification of the specification is neither warranted nor practical in this paper; however, more information is available from the corresponding author.

The prior distributions used are minimally informative, although dispersion and mean adjustment are restricted to sensible values in order to reduce the number of times the simulation returned an error warning when sampling from a gamma distribution. The prior distribution used for 'log.beta' is log uniform in order to avoid biasing towards larger values of 'beta' of >0.5 (small amounts of dispersion) when there is little information in the data. 'beta' is defined as the exponent of 'log.beta' in the model.

The prior distributions and initial values used for 'mean.correction', 'log.beta', and 'Probability' or one minus zero-inflation are given in Table 3.

REFERENCES

- Anderson, R. M. and May, R. M.** (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford Science Publications, Oxford, UK.
- Armour, J. and Coop, R. L.** (1991). Pathogenesis and control of gastrointestinal helminthiasis. In *Diseases of Sheep* (ed. Martin, W. B. and Aitken, I. D.), pp. 122–130. Blackwell Scientific Publications, Oxford, UK.
- Armour, J., Jarrett, W. F. H. and Jennings, F. W.** (1966). Experimental *Ostertagia circumcincta* infections in sheep: development and pathogenesis of a single infection. *American Journal of Veterinary Research* **27**, 1267–1278.
- Barger, I. A.** (1985). The statistical distribution of trichostrongylid nematodes in grazing lambs. *International Journal for Parasitology* **15**, 645–649.
- Bishop, S. C., Jackson, F., Coop, R. L. and Stear, M. J.** (2004). Genetic parameters for resistance to nematode infections in Texel lambs and their utility in breeding programmes. *Animal Science* **78**, 185–194.
- Bliss, C. I. and Fisher, R. A.** (1953). Fitting the negative binomial distribution to biological data. *Biometrics* **9**, 176–200.
- Dineen, J. K.** (1963). Immunological aspects of parasitism. *Nature, London* **197**, 268–269.
- Edwards, A. W. F.** (1992). *Likelihood Expanded Edition*. The Johns Hopkins University Press, Baltimore, MD, USA.
- Elliott, J. M.** (1977). *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates*, 2nd Edn. Freshwater Biological Association, Ambleside, UK.
- Gaba, S., Ginot, V. and Cabaret, J.** (2005). Modelling macroparasite aggregation using a nematode-sheep system: the Weibull distribution as an alternative to the negative binomial distribution? *Parasitology* **131**, 393–401.
- Gelman, A. and Rubin, D. B.** (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.
- Hunter, G. C. and Quenouille, M. H.** (1952). A statistical examination of the worm egg count sampling technique for sheep. *Journal of Helminthology* **26**, 157–170.
- McCullagh, P. and Nelder, J. A.** (1989). *Generalized Linear Models*. Chapman and Hall, London.
- McEwan, J. C., Dodds, K. G., Greer, G. J., Bain, W. E., Duncan, S. J., Wheeler, R., Knowler, K. J., Reid, P. J., Green, R. S. and Douch, P. G. C.** (1995). Genetic estimates for parasite resistance traits in sheep and their correlations with production traits. *New Zealand Journal of Zoology* **22**, 177.
- Morgan, E. R., Cavill, L., Curry, G. E., Wood, R. M. and Mitchell, E. S. E.** (2005). Effects of aggregation and sample size on composite faecal egg counts in sheep. *Veterinary Parasitology* **128**, 79–87.
- Morris, C. A., Bisset, S. A., Vlassoff, A., West, C. J. and Wheeler, M.** (2004). Genetic parameters for *Nematodirus* spp. egg counts in Romney lambs in New Zealand. *Animal Science* **79**, 33–39.
- Nødtvedt, A., Dohoo, I., Sanchez, J., Conboy, G., DesCôteaux, L., Keefe, G., Leslie, K. and Campbell, J.** (2002). The use of negative binomial modelling in a longitudinal study of gastrointestinal parasite burdens in Canadian dairy cows. *Canadian Journal of Veterinary Science* **66**, 249–257.
- Plummer, M., Best, N., Cowles, K. and Vines, K.** (2006). *coda: Output Analysis and Diagnostics for MCMC*. R package version 0.10–7.
- R Development Core Team** (2007). *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Spiegelhalter, D. J., Thomas, A. and Best, N. G.** (2003). *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit.
- Stear, M. J., Bairden, K., Bishop, S. C., Gettinby, G., McKellar, Q. A., Park, M., Strain, S. A. J. and Wallace, D. S.** (1998). The processes influencing the distribution of parasitic nematodes among naturally infected lambs. *Parasitology* **117**, 165–171.
- Toft, N., Innocent, G. T., Gettinby, G. and Reid, S. W. J.** (2007). Assessing the convergence of Markov chain Monte Carlo methods: an example from

- evaluation of diagnostic tests in absence of a gold standard. *Preventive Veterinary Medicine* **79**, 244–256.
- Torgerson, P. R., Schnyder, M. and Hertzberg, H.** (2005). Detection of anthelmintic resistance: a comparison of mathematical techniques. *Veterinary Parasitology* **128**, 291–298.
- van de Ven, R.** (1993). Estimating the shape parameter of the negative binomial distribution. *Journal of Statistical Computation and Simulation* **46**, 111–123.
- van de Ven, R. and Weber, N.** (1999). On the asymptotic behaviour of estimators of the shape parameter of the negative binomial distribution. ISI 99 52nd Session. *Bulletin of the International Statistical Institute*, pp. 419–420. Edita Ltd, Helsinki, Finland.