Zero-inflated hierarchical models for faecal egg counts to assess anthelmintic efficacy

Craig Wang

Department of Mathematics, University of Zurich, Zurich, Switzerland.

Paul R. Torgerson

Section of Veterinary Epidemiology, Vetsuisse Faculty, University of Zurich, Switzerland

Johan Höglund

Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Section for Parasitology, Uppsala, Sweden

Reinhard Furrer

Department of Mathematics and Department of Computer Science, University of Zurich, Zurich, Switzerland.

Abstract

The prevalence of anthelmintic resistance has increased in recent years, as a result of the extensive use of anthelmintic drugs to reduce the infection of parasitic worms in livestock. In order to detect the resistance, the number of parasite eggs in animal faeces is counted. Typically a subsample of the diluted faeces is examined, and the mean egg counts from both untreated and treated animals are compared. However, the conventional method ignores the variabilities introduced by the counting process and by different infection levels across animals. In addition, there can be extra zero counts, which arise as a result of the unexposed animals in an infected population or animals . In this paper, we propose the zero-inflated Bayesian hierarchical models to estimate the reduction in faecal egg counts. The simulation study compares the Bayesian models with the conventional faecal egg count reduction test and other methods such as bootstrap and quasi-Poisson regression. The results show the Bayesian models are more robust and they perform well in terms of both the bias and the coverage. We further illustrate the advantages of our proposed model using a case study

about the anthelmintic resistance in Swedish sheep flocks.

Keywords: Bayesian hierarchical model, faecal egg count reduction test, anthelmintic resistance, zero-inflated models, statistical analysis

1. Introduction

Gastrointestinal nematodes are parasitic worms that survive in livestock hosts, such as sheep, cattle and horses. The infection is common in the livestock populations in some regions (Waruiru et al., 2001; Mortensen et al., 2003; Pfukenyi et al., 2007; Tariq, 2014; Zanzani et al., 2014). Such infection can lead to numerous problems including reduction in skeletal growth, live-weight gain and milk yield (Houtert and Sykes, 1996), which can impose great economic burden on ruminant production (Perry and Randolph, 1999). The regular administration of anthelmintic treatments is a widely used method to control the infection. It aims not to eliminate the infection, but to reduce the infection intensity and prevent transmission (Levecke et al., 2012a). However, anthelmintic resistant nematodes appeared in different regions across the globe since late 1950s (Kaplan, 2004). The extensive use of anthelmintic treatments has led to an increasing problem of anthelmintic resistance. Once a resistance is detected, alternative treatments are needed in order to avoid any further production losses. Accurate and reliable methods to assess the treatment efficacy are thus essential to effectively control and monitor the infection.

The widely used faecal egg count reduction test (FECRT) was established in the early 1990s (Coles et al., 1992). It is a straightforward method to calculate the reduction in faecal egg counts (FECs), by comparing the mean pre-treatment and post-treatment FECs. For sheep and goats, if both the percentage reduction in mean FECs is less than 95% and the corresponding lower confidence limit is less than 90%, then the anthelmintic resistance is declared to be present. A standard method to obtain the FECs, the modified McMaster counting technique, is detailed in the guideline of the World Association for the Advancement of Veterinary Parasitology (WAAVP) (Coles et al., 1992). New WAAVP guidelines

are not yet developed, but Levecke et al. (2017) have made recommendations to improve and standardize the FECRT.

Although the FECRT and the McMaster technique were widely used in practice, some limitations have been pointed out in recent years. First of all, the McMaster counting technique introduces substantial variability in the results which is not accounted for in the FECRT (Torgerson et al., 2012). As a consequence of this, the estimated efficacy were found to be quite variable particularly for the samples with low pre-treatment FECs and efficacy in the range between 90%–95% (Miller et al., 2006). The use of refined techniques with a high analytical sensitivity such as FLOTAC (Giuseppe et al., 2010) and Cornell-Wisconsin (Egwang and Slocombe, 1982) can reduce but not eliminate the variability (Torgerson et al., 2012; Levecke et al., 2012b). Secondly, the distribution of egg counts is typically aggregated or overdispersed within the host population (Grenfell et al., 1995). Levecke et al. (2012a) evaluated the FECRT under different scenarios, highlighted that test results should be interpreted with caution when the sample size is small and the aggregation level is high. There were several attempts to propose more elaborate statistical models in the past years. Torgerson et al. (2005) assumed a negative binomial distribution for the counts, and used parametric bootstrap to calculate the confidence interval (CI) of the FECs reduction. More recently, methods have emerged that formulate the problem in a Bayesian framework. Denwood et al. (2010) considered a Poisson-gamma distribution for the counts, with the post-treatment mean linked to the pre-treatment mean via a scale factor. The inference is then done using Markov chain Monte Carlo (MCMC). Dobson et al. (2012) proposed a novel way to determine the confidence limits of the FECs reduction using Jeffrey intervals, which is derived from Bayesian procedures using a non-informative prior, however it requires high counts and high analytical sensitivity. Paul et al. (2014) proposed a hierarchical model that uses binomial distribution to capture the counting variability, and a Poisson-gamma distribution to model the overdispersion. The posterior median for the reduction and its 95% highest posterior density (HPD) interval is used for its point and

interval estimate respectively. An easy-to-use web interface was implemented and made available online (Torgerson et al., 2014). However the models themselves were not published and well-documented. Levecke et al. (2015) proposed another Bayesian model with a slightly different formulation. It used a Poisson distribution to capture the variability in the counting process and a negative binomial distribution to capture the overdispersion. The Bayesian models do not only provide credible intervals on the reduction, but also generate posterior distributions for each of the model parameters, hence offering a probabilistic view on the efficacy rather than a yes or no answer. To the best of our knowledge, a common assumption made by those recent Bayesian models is that all animals in an infected population are exposed. However, Denwood et al. (2008) showed the underlying distribution of the nematodes FECs can be zero-inflated negative binomial (ZINB). The zero-inflation component can arise as a result of the unexposed livestock in an infected population. Models with zero inflation have already been used in the context of disease mapping (Vounatsou et al., 2009; Soares Magalhães et al., 2011).

In this paper, we propose zero-inflated Bayesian hierarchical models to estimate the reduction in FECs. We build on the models in (Paul et al., 2014) and explicitly formulate the model structures. The models account for the extra variabilities that arise from both the sampling process and the between-animal variations. In addition, the models allow for extra zero counts by introducing the zero-inflation components. Overall, the models are more flexible and are suitable for a wide range of scenarios. The rest of this paper is organized as follows. Section 2 briefly reviews the conventional FECRT and efforts made to modify it. Section 3 introduces the zero-inflated Bayesian hierarchical models. Section 4 conducts a simulation study, where the bias and coverage of the estimated FECs reduction are compared across different methods. In Section 5, a case study is used to illustrate the proposed methods for estimating the reduction in FECs, where anthelmintic resistance was investigated in Swedish sheep flocks. Finally, Section 6 concludes with a discussion.

2. Faecal egg count reduction test

The FECRT was suggested in the WAAVP guideline for estimating the reduction in FECs and its corresponding CI (Coles et al., 1992). In order to reduce the counting variability, using groups of at least 10-15 animals was suggested. In addition, the mean pre-treatment FECs should be at least 150 epg, otherwise the FECRT can give unreliable results.

Suppose a group of n_T animals received anthelmintic treatment and a group of n_C animals serves as control. The percentage reduction in FECs can be calculated as

$$\text{Percentage reduction } = 100 \times \left(1 - \frac{\bar{x}_T}{\bar{x}_C}\right), \tag{1}$$

where $\bar{x}_T$ and $\bar{x}_C$ denote the mean counts of the treatment and the control group. Assuming independence, the estimated asymptotic variance of the log ratio is given by

$$\widehat{\text{Var}}\left(\log \frac{\bar{X}_T}{\bar{X}_C}\right) = \frac{s_T^2}{n_T \bar{x}_T^2} + \frac{s_C^2}{n_C \bar{x}_C^2}. \tag{2}$$

where $\bar{X}_T$ and $\bar{X}_C$ denote the means of random samples, s_T^2 and s_C^2 denote the sample variance of the treatment and the control group counts. The variance can be used to construct an approximate 95% CI of the log ratio using the 97.5% and the 2.5% quantile of a Student's t-distribution with $n_T + n_C - 2$ degrees of freedom. The CI for the log-ratio can be then transformed back to obtain the 95% CI for the estimated reduction. The WAAVP guideline (Coles et al., 1992) states that for sheep and goats, the resistance is present if (i) the percentage reduction in FECs is less than 95% and (ii) the corresponding lower 95% confidence limit is less than 90%. If only one of these two criteria is met, then resistance is suspected. Different thresholds have been suggested for other livestock.

Over the past years, modified versions of the FECRT have been proposed in the literature. Wood et al. (1995) suggested to use the geometric mean in the FECRT instead of arithmetic mean. Davison and Hinkley (1997) suggested the

95% CI can also be calculated using nonparametric bootstrap. In the unpaired design, there is one group of animals that receives the treatment and another group is chosen to act as the control group. McKenna (1990) suggested that instead of taking samples from two groups of animals, the pre-treatment counts from the treatment group can be used as the baseline, hence eliminated the need of a distinct control group. We refer to this as the paired design. In this case, the FECRT becomes inappropriate since it does not take the paired structure into account in calculating the variance.

3. Bayesian hierarchical models

There are two designs that can be used for detecting anthelmintic resistance in a livestock population. For each design, we propose a zero-inflated Bayesian hierarchical model to estimate the reduction in FECs.

3.1. The unpaired design

Suppose we have two groups of animals from the same population, a control group with size n_C and a treatment group with size n_T. A faecal sample from each animal is collected and counted with an analytical sensitivity f_i, where i is the index of each animal in the corresponding group. We assume the counts belong to the same species, more specifically the counts follow a unimodal distribution. For notational simplicity, we assume every sample has the same analytical sensitivity, hence the index in f_i is dropped for the rest of the paper. The faecal sample is thoroughly mixed after dilution, hence we assume the eggs are homogeneously distributed within each sample. A proportion of the diluted sample $p = 1/f$ is then counted. Denote the raw number of eggs in the diluted sample of the ith control animal as Y_i^{*C}, with $i = 1, 2, ..., n_C$. Given the true number of eggs per gram of faeces Y_i^C, the raw count Y_i^{*C} follows a binomial distribution with size Y_i^C and probability p. This captures both the dilution and the McMaster counting variability. Then the true epg Y_i^C follows a zero-inflated Poisson (ZIP) distribution with mean μ_i^C and zero-inflation parameter

ϕ, it implies Y_i^C is 0 with probability ϕ, and follows the Poisson distribution with mean μ_i^C with probability $(1-\phi)$. The zero-inflation component captures the excess number of zero counts that could come from unexposed animals, while the Poisson component captures the animals with zero counts that are below the detection limit. Finally the mean μ_i^C is gamma-distributed with shape κ and rate κ/μ. It has mean μ and variance μ^2/κ, the gamma distribution captures the overdispersion of the egg counts. This yields the following model for the control group animals,

$$\begin{aligned} Y_i^{*C}|Y_i^C &\sim \text{Bin}(Y_i^C, p), \\ Y_i^C|\mu_i^C, \phi &\sim \text{ZIP}(\mu_i^C, \phi), \\ \mu_i^C|\kappa, \mu &\sim \text{Gamma}(\kappa, \kappa/\mu). \end{aligned} \tag{3}$$

For the treatment group, the number of eggs in faecal samples is likely to decrease after some days receiving the treatment, hence we introduce a reduction factor $(1-\delta)$ where δ represents the proportion of eggs remaining. The treatment may significantly reduce the infection level but it is very unlikely to completely eliminate the infection, hence the zero-inflation component remains the same. In addition, we assume the reduction in FECs occurs at individual level, such that the parameters μ and κ also stay the same. This yields the following model for the treatment group,

$$\begin{aligned} Y_i^{*T}|Y_i^T &\sim \text{Bin}(Y_i^T, p), \\ Y_i^T|\mu_i^T, \phi &\sim \text{ZIP}(\delta\mu_i^T, \phi), \\ \mu_i^T|\kappa, \mu &\sim \text{Gamma}(\kappa, \kappa/\mu). \end{aligned} \tag{4}$$

where the superscript T denotes the parameters for the treatment group. The priors for the flock parameters μ, κ and ϕ need to be specified in advance. If previous knowledge about the distribution of those parameters is available, they can be taken into account in the model as priors. Otherwise, diffuse priors should be used.

3.2. The paired design

In the paired design, there is only one group of animals of size n. A faecal sample from each animal is counted twice, once before the treatment and once some days after the treatment. The baseline counts of each animal is used as the corresponding control. The model for the paired design is

$$\begin{aligned} Y_i^{*C}|Y_i^C &\sim \text{Bin}(Y_i^C, p), \\ Y_i^C|\mu_i^C, \phi &\sim \text{ZIPois}(\mu_i^C, \phi), \\ Y_i^{*T}|Y_i^T &\sim \text{Bin}(Y_i^T, p), \\ Y_i^T|\mu_i^C, \phi &\sim \text{ZIPois}(\delta\mu_i^C, \phi), \\ \mu_i^C|\kappa, \mu &\sim \text{Gamma}(\kappa, \kappa/\mu). \end{aligned} \tag{5}$$

The only difference in the model comparing with the unpaired design is that, the pre-treatment epg Y_i^C and post-treatment epg Y_i^T are now based on the same Poisson mean μ_i^C to indicate that they belong to the same animal. The priors for the flock parameters should be specified in a similar way as for the unpaired design.

The hierarchical model given in Eq. (5) without zero-inflation in Y_i^C and Y_i^T was proposed in (Paul et al., 2014), however the authors used the posterior median as the point estimate for the reduction, and the 95% HPD credible interval as the interval estimate. The model was implemented in the "eggCounts" package version $\leq$ 0.4-1 (Wang and Paul, 2016) in R along with the hierarchical model for the unpaired design without zero-inflation. In addition, the authors used $(1 - \bar{Y}_i^C/\bar{Y}_i^T)$ as the posterior samples for the reduction in the unpaired model rather than using $(1 - \delta)$ directly. Typically, the posterior mode is used in conjunction with the HPD interval. In the simulation study, we show that using the posterior mode of the reduction parameter as the estimate gives a smaller bias compared to using the posterior median.

4. Simulation study

In order to investigate the performance of the proposed Bayesian models, we conduct a simulation study to estimate the FECs reduction. We first simulate the FECs data under different scenarios, then use our proposed models and other methods to estimate the reduction. The bias and the coverage of the 95% CIs or credible intervals are compared across different methods.

4.1. Simulation setup

FECs for both unpaired and paired designs are simulated. For each design, we consider 16 different scenarios that vary in terms of the baseline mean count μ (150 epg or 500 epg), the dispersion κ (1 or 2), the reduction $(1-\delta)$ (90% or 95%) and the zero-inflation ϕ (0 or 30%). Sample size is chosen to be 15 for all scenarios, and the analytical sensitivity is 50. For each scenario in each design, 1000 dataset are simulated. The pre-treatment FECs are simulated as follows: we firstly draw the mean epg μ_i^C from a gamma distribution with shape κ and rate κ/μ. Then the true number of eggs y_i^C are drawn from a ZIP distribution with mean μ_i^C and zero-inflation ϕ. Finally, the observed counts are drawn from another Poisson distribution with mean y_i^C/f where f is the analytical sensitivity. The post-treatment FECs are simulated in a similar way but with different parameters. Note the process of simulating the data does not exactly match our proposed model. In addition, the simulation parameters are chosen such that the FECRT is suitable to use under the guideline of WAAVP. This encourages a fair comparison across the different methods. If we simulate exactly as our model specifications, we expect the results will be even more favorable.

We compare several different methods for estimating the mean FECs reduction and its confidence interval. For the unpaired design, we consider the FECRT with the approximate CI (FECRT); the hierarchical model in Eq. (3)–(4) without zero-inflation and using posterior median as the point estimate, as implemented in (Wang and Paul, 2016) (PoGa(median)) and the same model

but using posterior mode as the point estimate (PoGa(mode)); our proposed zero-inflated hierarchical model for the unpaired design (ZIPoGa); and finally parametric bootstrap, assuming zero-inflated negative binomial distributions and using 1999 bootstrap samples (pBoot).

The FECRT does not distinguish between paired and unpaired designs, hence it is applicable to both. The zero-inflated negative binomial regression does not perform well when the sample size is small, and it sometimes does not produce sensible results (Denwood et al., 2008). Hence for the paired design, in addition to the FECRT, we consider a quasi-Poisson regression, excluding zero pre-treatment counts and using log pre-treatment counts as the offset term (qPois); the proposed hierarchical model in (Paul et al., 2014) using posterior median as the point estimate (PoGa(median)) and the same model but using posterior mode as the point estimate (PoGa(mode)); and finally our proposed zero-inflated hierarchical model for the paired design (ZIPoGa).

The Bayesian models are implemented in the "eggCounts" package version 1.1-1 (Wang and Paul, 2016) using the modelling language Stan (Carpenter, 2015), Stan uses an effective MCMC sampling technique and is available through the "stan" package (Guo et al., 2015) in R (R Core Team, 2015). The prior for the reduction follows a Beta$(1, 1)$ distribution, which assigns uniform density between 0 and 1. For the parameters μ and κ, we use Gamma$(1, 0.001)$ and Gamma$(1, 0.7)$ prior respectively. For each Bayesian model, 12,000 MCMC samples are generated with 2,000 samples for burn-in without thinning. The posterior mode is used as the estimate for the reduction parameter in our proposed models, and the 95% HPD interval of the posterior samples was obtained. All the simulations are conducted in R version 3.2.1.

4.2. Simulation results

Fig. 1 and Fig. 2 show the bias and the coverage probability of 95% CIs or 95% HPD interval for the FECs reduction, in the case of unpaired designs. The PoGa(median) model slightly underestimate the reduction in most cases, however it is improved by using the posterior mode as the point estimate as shown

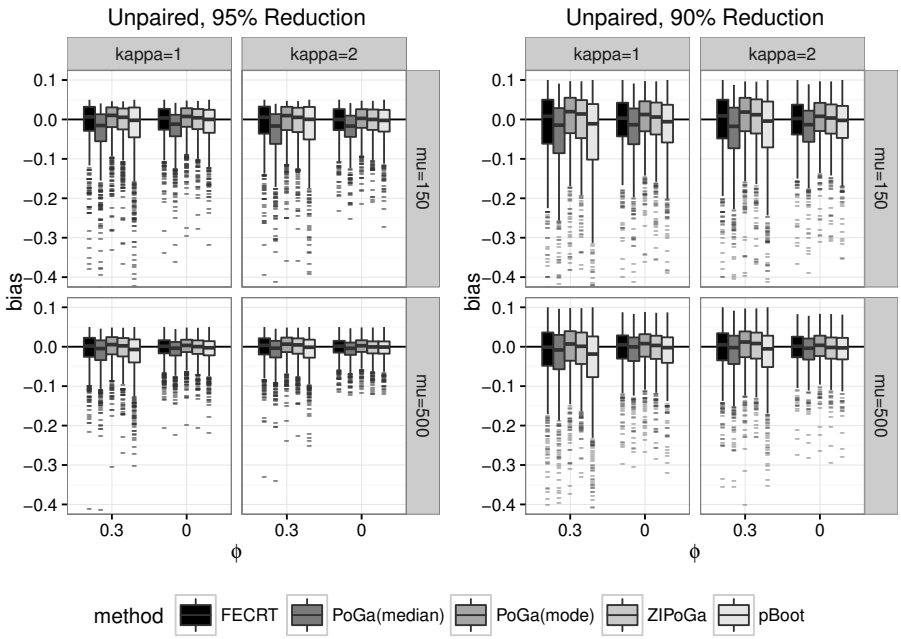

Fig. 1: Boxplots of the estimated FECs reduction in the paired design, using FECRT with approximated CI (FECRT); the hierarchical model without zero-inflation (Paul et al., 2014) and using the posterior median as the point estimate (PoGa(median)); the hierarchical model without zero-inflation (Paul et al., 2014) and using the posterior mode as the point estimate (PoGa(mode)); our proposed zero-inflated hierarchical model for the paired design (ZIPoGa); and quasi-Poisson regression (qPois). The horizontal line indicates zero bias.

in PoGa(mode). All the other methods have small biases. Both the FECRT and the parametric bootstrap method have inaccurate coverage probabilities when the pre-treatment mean count is low. As expected, the FECRT has accurate coverage when the pre-treatment mean is high, since the asymptotic variance improves. The PoGa(median) model provides low coverage probability when the pre-treatment mean count is high, and it is improved by using $(1 - \delta)$ as the posterior samples for the reduction directly. In contrast, our proposed zero-inflation models offers good coverage while maintaining small bias. Note the Bayesian credible intervals do not have a long-run property like the CIs where 95 percent of the 95% CIs should cover the true parameter value (Spiegelhalter et al., 2004), but the coverage probability for the Bayesian methods can still be

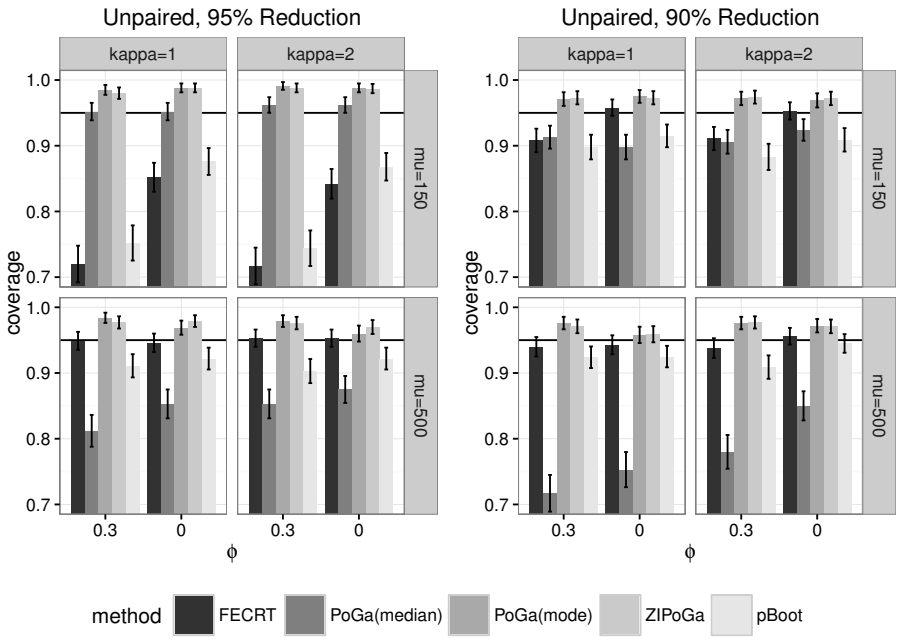

Fig. 2: Barplots of the coverage probability of the 95% CIs, or HPD credible intervals for the FECs reduction in the unpaired design. The error bars are calculated based on the 95% binomial confidence interval. The horizontal line indicates 95% coverage. The methods are the same as described in Fig. 1.

used as a rule of thumb to assess the models.

Fig. 3 and Fig. 4 show the bias and the coverage probability for the paired designs. The biases are small for all the methods except the PoGa(median) model. It is improved again by using the posterior mode as the estimate. In term of the coverage, the FECRT method tends to have wide confidence intervals since they do not take the paired structure into account, resulting almost 100% coverage when the pre-treatment mean is high. The Bayesian models provide slight over-coverage in all the scenarios.

Overall, the zero-inflated Bayesian models are robust methods. They consistently provide small bias and accurate coverage in the simulated scenarios. In the following case study, we further illustrate the advantages of the zero-inflated hierarchical models.

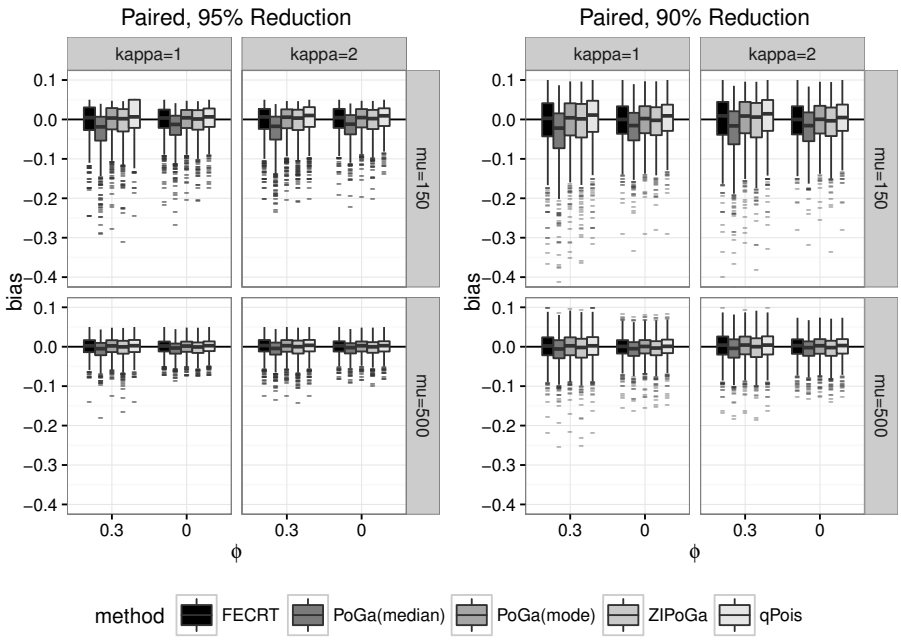

Fig. 3: Boxplots of the estimated FECs reduction in the paired design, using FECRT with approximated CI (FECRT); the hierarchical model without zero-inflation (Paul et al., 2014) and using the posterior median as the point estimate (PoGa(median)); the hierarchical model without zero-inflation (Paul et al., 2014) and using the posterior mode as the point estimate (PoGa(mode)); our proposed zero-inflated hierarchical model for the paired design (ZIPoGa); and quasi-Poisson regression (qPois). The horizontal line indicates zero bias.

5. Case study: anthelmintic resistance in Swedish sheep flocks

In order to illustrate our proposed model, we re-analyze the data in a study where the prevalence of anthelmintic resistance in parasitic nematodes in Swedish sheep flocks was investigated (Höglund et al., 2009). The FEC data was collected and analyzed using both the FECRT and molecular testing methods. In the study, a total of 45 farms were randomly selected throughout Sweden, each with a minimum of 20 ewes. During the grazing season of 2006 and 2007, two flocks of approximately 15 lambs were selected from each farm, each flock was treated with either a benzimidazole (BZ), albendazole (Valbazen®, Pfizer) or a macrocyclic lactone, ivermectin (Ivomec®, Merial). In this paper, we only consider the efficacy of BZ, which was received by 45 out of all 90

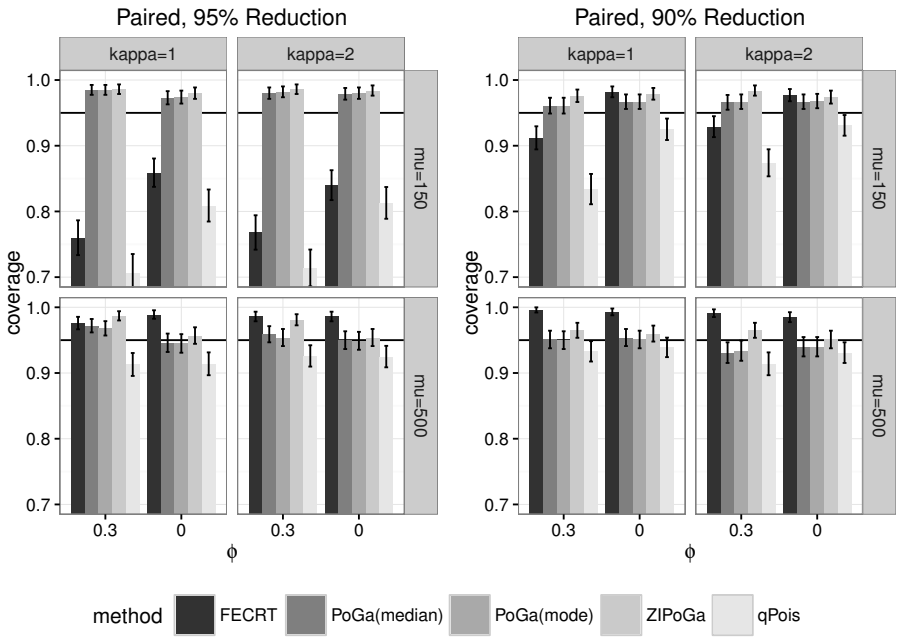

Fig. 4: Barplots of the coverage probability of the 95% CIs, or HPD credible intervals in the case of Bayesian models, for the FECs reduction in the paired design. The error bars are calculated based on the 95% binomial confidence interval. The horizontal line indicates 95% coverage. The methods are the same as described in Fig. 3.

flocks selected. However the model is applicable for other treatments as well as other livestock. Each lamb was sampled before treatment using the modified McMaster technique with an analytical sensitivity of 50. 39 out of 45 flocks with mean of at least 50 epg was re-sampled using the same setting 7-10 days after treatment, with flock sizes varying between 10 to 17 animals. In addition to the McMaster counting technique, the BZ-resistance of parasites was tested using a pyrosequencing assay. Larval cultures indicated that Teladorsagia and Trichostrongylid nematode infection were predominant.

There are 39 flocks consisting of 575 animals in total, all of them were treated with BZ. The post-treatment FECs are missing in 28 animals, hence they are excluded from the analysis. In addition, one animal had a pre-treatment epg of 30, which is not possible with a correction factor of 50. In this case, the author clarified that 3 eggs were observed outside the grid area on the McMaster slide,

hence a correction factor of 10 was applied. However according to WAAVP guideline, eggs outside the grid should not be counted, hence this particular observation was set to zero. Using FECRT, we first calculate the reduction in FECs and its approximate 95% CI. The decision rule for sheep and goats suggested in the WAAVP guidelines is used for deciding anthelmintic resistance. In 35 flocks, all of the post-treatment counts were zero which resulted 100% reduction in each flock. The CI for those flocks cannot be computed using the FECRT. Out of the remaining 4 flocks, the parasite in 2 flocks (flock 33 and 39) are anthelmintic resistant according to the FECRT. The results based on the molecular testing suggested 5 out of 39 flocks (flock 24, 33, 36, 37 and 39) have anthelmintic resistance present using the codon 200 TAC allele frequency of $\geq 95\%$ as the indicator. In the end, the authors concluded that the prevalence of anthelmintic resistance in the Swedish sheep population is relatively low, however it is more widespread than the FECRT indicated. The paper pointed out the urgent need to develop alternative diagnostic procedures. The quasi-Poisson regression gave similar results, failing to detect the remaining resistance.

In the following, we re-analyze the FECs data from the Swedish sheep study using our proposed model. The worm burden differs depending on the animals and the type of parasites eggs that is being counted, hence the choice of hyperparameters for the prior should be based on similar studies. According to another study of the distribution of trichostrongylid eggs in the sheep flocks (Morgan et al., 2005), the mean pre-treatment FECs ranged from 43 to 1915, and the estimated dispersion parameter based on negative binomial regressions ranged from 0.18 (95% CI: 0.10 to 0.32) to 2.3 (95% CI: 0.2 to 4.2). Hence we assign a weakly informative prior Gamma(1, 0.001) to μ, where 90% of the probability mass lies between 59 and 2996, and assign a Gamma(1, 0.7) prior for κ, where 90% of the probability mass lies between 0.1 and 4.3 with a prior median of 1. We assume the overall level of infection does not increase after treatment is applied, hence the reduction should always be between 0 to 100%. A non-informative prior Beta(1, 1) is assigned to the parameter δ, such that all the values between 0 and 1 are equally likely a priori. Finally for the zero-inflation

Flock	FECRT	quasi-Poisson	PoGa(mode)	ZIPoGa
24	99.0 (96.3, 99.8)	99.0 (97.2, 99.7)	99.0 (98.5, 99.4)	97.8 (95.8, 98.9)
33	**82.2 (65.4, 90.8)**	**82.2 (68.6, 90.0)**	**81.3 (77.4, 85.9)**	**76.8 (70.6, 81.8)**
36	97.5 (90.6, 99.4)	97.5 (93.2, 99.1)	97.6 (93.1, 99.2)	97.4 (93.1, 99.2)
37	100.0 (–, –)	100.0 (100.0, 100.0)	*99.3 (89.5, 100.0)*	*98.8 (49.3, 100.0)*
39	**92.3 (62.9, 98.4)**	**93.9 (90.1, 96.3)**	**92.6 (89.0, 94.8)**	**93.1 (89.7, 95.6)**

Table 1: Analysis results for the five BZ treated flocks which the molecular testing indicated anthelmintic resistance are present. Results are shown for the estimated percentage reduction in FECs using the FECRT, quasi-Poisson regression, the PoGa and the ZIPoGa hierarchical models. The 95% CI are shown for the first two methods, while the 95% HPD intervals are shown for the hierarchical model. The text is in **bold** if a resistance is detected, and is in *italic* if a resistance is suspected.

parameter ϕ, we assign a non-informative Beta(1, 1) prior.

We apply the zero-inflated Bayesian model for the paired design separately to each flock. In order to diagnose the potential non-convergence, 4 MCMC chains were requested. Each has 12,000 MCMC samples, 2,000 samples for burn-in and without thinning. There was no evidence of non-convergence with potential scale reduction factors (Brooks and Gelman, 1998) approximately equal to 1. The sensitivity analysis showed similar results with wide uniform priors on the mean and dispersion, here we only present the main results. Table 1 shows the results for the five flocks which the molecular data indicated anthelmintic resistance. The approximate CI cannot be computed for flock 37 using the FECRT, since all the post-treatment FECs are zero. Because the standard FECRT method does not take the paired structure into account, the approximate CI is wider in general compares to the quasi-Poisson regression and the Bayesian models. The Bayesian models are able to obtain an interval estimate even when the reduction is 100%. The posterior mode estimate for the Bayesian model without zero-inflation is similar to the FECRT, however the zero-inflated Bayesian model gave slightly different estimates. In particular, the posterior mode for the reduction in flock 33 is 76.8% using our proposed model, compare to 82.2% and 81.3% in the FECRT and PoGa(mode). Indeed, the mean reduction calculated using Eq. (1) is 82.2%, however this completely ignores the paired structure. The actual mean pairwise reduction for flock 33 is 73.1%, hence our proposed ZIPoGa model provide a more sensible result in this case. For flock 37, the Bayesian models

classify it as suspected resistance due to its lower confidence limit. Since no parasite eggs were detected in 7 out of 13 sheep before treatment, the uncertainty in the treatment efficacy is high. Hence the interval estimate is much wider, which is only captured by the zero-inflation model. The other classification results stay the same.

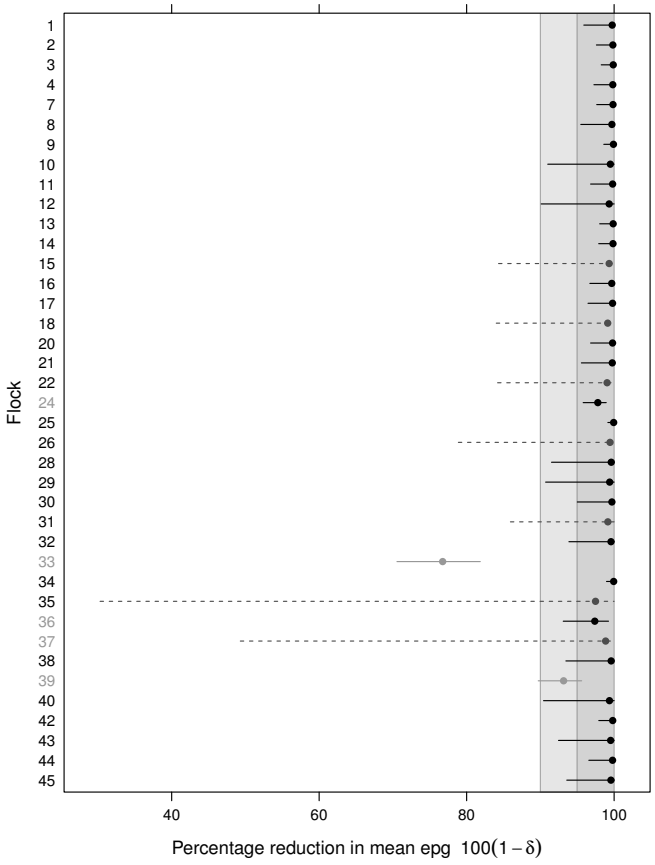

Fig. 5: Estimated reduction in mean FECs and its 95% HPD interval for the 39 flocks that were sampled both before and after treated with BZ. Using the WAAVP guideline for the decision of anthelmintic resistance, the intervals in solid black lines belong to flocks with no anthelmintic resistance, intervals in dashed lines belong to flocks with suspected resistance and intervals in solid gray lines belong to flocks with resistance. The flock numbers that were flagged as resistant using molecular data are colored in grey.

Fig. 5 shows the estimated reductions and its 95% HPD intervals for all 39 flocks considered in the case study. There are several flocks that are flagged as suspected resistance even though there were no eggs present in the post-treatment FECs. For example, flock 35 has 15 sheep, all of which had zero post-treatment FECs. However, 10 out of 15 sheep had zero pre-treatment counts, those could be the unexposed individuals that should not contribute to the estimation of treatment efficacy. This is captured by the zero-inflated model, hence the HPD credible interval for this flock is wide.

6. Discussion

In this paper we propose zero-inflated Bayesian hierarchical models for estimating the reduction in FECs. The models capture the additional sources of variability in the data, and allow for extra zero counts that are frequently observed in practice due to unexposed animals. The simulation results suggest that the zero-inflated Bayesian hierarchical models are robust methods to estimate the reduction, in both unpaired and paired designs. They consistently provide small bias and good coverage in all the simulated scenarios even though we did not simulate exactly according to our model specification. The case study further illustrated the advantages of our proposed model, which can accurately model the paired structure and provide an interval estimate where the conventional method cannot. The extra uncertainty in reduction introduced by the zero counts was only reflected in the proposed zero-inflation model.

An advantage of the Bayesian approach is that it does not only provide the reduction estimate and the credible interval, but also it offers density distributions of the model parameters. Denwood et al. (2010) pointed out that Bayesian methods allow for probabilistic classification on the efficacy, in terms of the probability that a true reduction is below a given threshold. According to the WAAVP guidelines, there are three possible decision outcomes on resistance status, namely “yes”, “suspected” and “no”. Such trichotomy outcome should be interpreted with caution, especially at the decision boundaries. We illustrate

the probabilistic view using flock number 37 and 39. Fig. 6 shows the posterior marginal density of the reduction parameter $(1-\delta)$ from the proposed model. Coles et al. (2006) stated that a reduction greater than 95% is considered as beneficial, hence we use this as the threshold. The shaded area in each case corresponds to the probability that the reduction in mean FECs is less than 95%, i.e. the probability that anthelmintic resistance is present using a 95% reduction as the threshold. Based on the posterior marginal distribution, the probability that the resistance is present in flock 37 is 0.42, indicating moderate evidence for resistance. For flock 39, the probability is 0.94 which suggests a very strong evidence that the resistance is present.

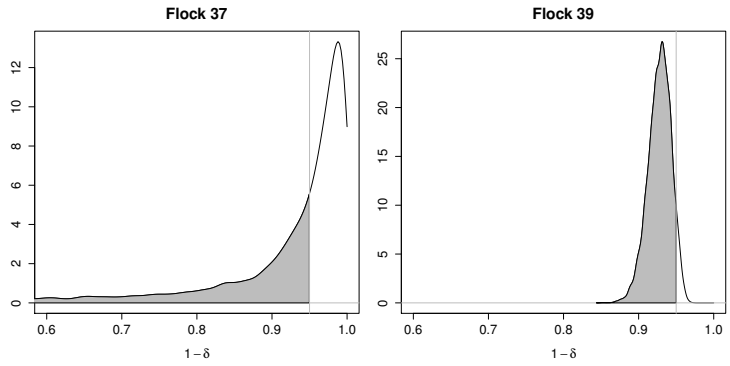

Fig. 6: The marginal posterior density for the reduction (1-δ) for flock 37 and 39. The shaded area represents the density mass for reduction less than 95%.

Another advantage of the Bayesian hierarchical models is its flexibility in model formulation. In this paper we have assumed the reduction in FECs is the same for every animal, as one can expect the efficacy of anthelmintic treatment across animals are similar within a resistant community. However each animal can experience different efficacy due to different metabolism or drug availability (Cabaret and Berrag, 2004), one can adjust the model to introduce animal-specific reductions. Sufficient data are required to ensure the convergence of the model. In the case study, if researchers are interested in assessing the anthelmintic resistance in the Swedish sheep population in general, a hierarchical meta-analysis model over all the flocks can be formulated. The corresponding

model parameters from each flock would follow the same distributions, for example, the parameter μ from each flock together follows a normal distribution with some population mean. This can be particular useful if one wishes to consider some national treatment schemes applied to the entire sheep population.

The proposed Bayesian models are implemented using efficient MCMC algorithm in the "eggCounts" package (Wang and Paul, 2016) in R. A website application that features all the basic functionalities of the package is available at `http://t.uzh.ch/D1` (Furrer et al., 2016), it has a easy-to-use interface and is designed for practitioners who do not have sufficient R knowledge.

Currently, the models assume the counts belong to the same species of parasites. However if there is a mixture of parasite species with different infection level, one expects a multi-modal distribution from the counts. Additionally if there is a different reduction for each species of the mixture, then the reduction parameter also follows a multi-modal distribution. Instead of a gamma distribution in Eq. (3)–(5), a mixture of Gamma distribution with an additional weight parameter for each component of the mixture could be used. Different possibilities of reduction from each species need to be carefully considered in the presence of mixture.

With the proposed models in mind, one can also design more efficient sampling process in order to obtain the estimated FEC reduction with sufficient statistical power. The sample size and the analytical sensitivity are the two important factors involved in a study design. The CIs are expected to be narrower for larger sample size and higher analytical sensitivity. The minimum sample size required for a reliable estimation of the reduction and the influence of analytical sensitivity can be further investigated for the zero-inflated Bayesian hierarchical models.

7. Acknowledgements

CW and RF acknowledge support of the Swiss National Science Foundation SNSF-144973. The authors thank Dr. Michaela Paul for initiating the eggCounts

package and drafting an earlier version of the paper based on the previous models. The authors thank two reviewers for helpful comments that led to improvements of the paper.

References

Brooks, S. P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comp. Graph. Stat. 7 (4), 434–455.

Cabaret, J., Berrag, B., 2004. Faecal egg count reduction test for assessing anthelmintic efficacy: average versus individually based estimations. Vet. Parasitol. 121 (1 - 2), 105–113.

Carpenter, B., 2015. Stan: A Probabilistic Programming Language. J Stat. Softw. In press.

Coles, G., Bauer, C., Borgsteede, F., Geerts, S., Klei, T., Taylor, M., Waller, P., 1992. World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) methods for the detection of anthelmintic resistance in nematodes of veterinary importance. Vet. Parasitol. 44 (1 - 2), 35–44.

Coles, G., Jackson, F., Pomroy, W., Prichard, R., von Samson-Himmelstjerna, G., Silvestre, A., Taylor, M., Vercruysse, J., 2006. The detection of anthelmintic resistance in nematodes of veterinary importance. Vet. Parasitol. 136 (3-4), 167–185.

Davison, A., Hinkley, D., 1997. Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Denwood, M., Reid, S., Love, S., Nielsen, M., Matthews, L., McKendrick, I., Innocent, G., 2010. Comparison of three alternative methods for analysis of equine faecal egg count reduction test data. Prev. Vet. Med. 93 (4), 316–323.

Denwood, M. J., Stear, M. J., Matthews, L., Reid, S. W. J., Toft, N., Innocent, G. T., 2008. The distribution of the pathogenic nematode Nematodirus battus in lambs is zero-inflated. Parasitol. 135, 1225–1235.

Dobson, R., Hosking, B., Jacobson, C., Cotter, J., Besier, R., Stein, P., Reid, S., 2012. Preserving new anthelmintics: a simple method for estimating faecal egg count reduction test (FECRT) confidence limits when efficacy and/or nematode aggregation is high. Vet. Parasitol. 186 (1-2), 79–92.

Egwang, T. G., Slocombe, J. O., 1982. Evaluation of the Cornell-Wisconsin centrifugal flotation technique for recovering trichostrongylid eggs from bovine feces. Can. J. Comp. Med. 46 (2), 133–137.

Furrer, R., Flury, R., Wang, C., 2016. Modelling faecal egg counts with shiny.
URL http://t.uzh.ch/D1

Giuseppe, C., Laura, R., Paola, M. M., Jurg, U., 2010. FLOTAC: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. Nat. Protoc. 5 (3), 503–515.

Grenfell, B. T., Wilson, K., Isham, V. S., Boyd, H. E. G., Dietz, K., 1995. Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study. Parasitol. 111, 135–151.

Guo, J., Gabry, J., Goodrich, B., 2015. rstan: R interface to Stan. R package version 2.8.2.
URL http://CRAN.R-project.org/package=rstan

Höglund, J., Gustafsson, K., Ljungström, B.-L., Engström, A., Donnan, A., Skuce, P., 2009. Anthelmintic resistance in Swedish sheep flocks based on a comparison of the results from the faecal egg count reduction test and resistant allele frequencies of the β-tubulin gene. Vet. Parasitol. 161 (1-2), 60–68.

Houtert, M. F. V., Sykes, A. R., 1996. Implications of nutrition for the ability of ruminants to withstand gastrointestinal nematode infections. Int. J. Parasitol. 26 (11), 1151–1167.

494 Kaplan, R. M., 2004. Drug resistance in nematodes of veterinary importance: a
495 status report. Trends Parasitol. 20 (10), 477–481.

496 Levecke, B., Anderson, R. M., Berkvens, D., Charlier, J., Devleesschauwer, B.,
497 Speybroeck, N., Vercruysse, J., Aelst, S. V., 2015. Chapter Five - Mathemat-
498 ical Inference on Helminth Egg Counts in Stool and Its Applications in Mass
499 Drug Administration Programmes to Control Soil-Transmitted Helminthiasis
500 in Public Health. In: Mathematical Models for Neglected Tropical Diseases:
501 Essential Tools for Control and Elimination, Part A. Vol. 87 of Adv. Parasitol.
502 Academic Press, pp. 193–247.

503 Levecke, B., Dobson, R., Speybroeck, N., Vercruysse, J., Charlier, J., 2012a.
504 Novel insights in the faecal egg count reduction test for monitoring drug effi-
505 cacy against gastrointestinal nematodes of veterinary importance. Vet. Para-
506 sitol. 188 (3–4), 391–396.

507 Levecke, B., Kaplan, R. M., Thamsborg, S. M., Torgerson, P. R., Vercruysse, J.,
508 Dobson, R. J., 2017. How to improve the standardization and the diagnostic
509 performance of the fecal egg count reduction test? Vet. Parasitol. In press.

510 Levecke, B., Rinaldi, L., Charlier, J., Maurelli, M., Bosco, A., Vercruysse,
511 J., Cringoli, G., 2012b. The bias, accuracy and precision of faecal egg
512 count reduction test results in cattle using McMaster, Cornell-Wisconsin and
513 FLOTAC egg counting methods. Vet. Parasitol. 188 (1–2), 194–199.

514 McKenna, P., 1990. The detection of anthelmintic resistance by the faecal egg
515 count reduction test: an examination of some of the factors affecting perfor-
516 mance and interpretation. N.Z. Vet. J. 38 (4), 142–147.

517 Miller, C., Waghorn, T., Leathwick, D., Gilmour, M., 2006. How repeatable is
518 a faecal egg count reduction test? N.Z. Vet. J. 54 (6), 323–328.

519 Morgan, E., Cavill, L., Curry, G., Wood, R., Mitchell, E., 2005. Effects of
520 aggregation and sample size on composite faecal egg counts in sheep. Vet.
521 Parasitol. 131 (1-2), 79–87.

Mortensen, L. L., Williamson, L. H., Terrill, T. H., Kircher, R. A., Larsen, M., Kaplan, R. M., 2003. Evaluation of prevalence and clinical implications of anthelmintic resistance in gastrointestinal nematodes in goats. J. Am. Vet. Med. Assoc. 223 (4), 495–500.

Paul, M., Torgerson, P. R., Höglund, J., Furrer, R., 2014. Hierarchical modelling of faecal egg counts to assess anthelmintic efficacy. ArXiv e-prints.

Perry, B., Randolph, T., 1999. Improving the assessment of the economic impact of parasitic diseases and of their control in production animals. Vet. Parasitol. 84 (3-4), 145–168.

Pfukenyi, D., Willingham, A., Mukaratirwa, S., Monrad, J., 2007. Epidemiological studies of parasitic gastrointestinal nematodes, cestodes and coccidia infections in cattle in the highveld and lowveld communal grazing areas of Zimbabwe. Onderstepoort J. Vet. Res. 74 (2).

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.R-project.org/

Soares Magalhães, R. J., Biritwum, N.-K., Gyapong, J. O., Brooker, S., Zhang, Y., Blair, L., Fenwick, A., Clements, A., 2011. Mapping helminth co-infection and co-intensity: geostatistical prediction in Ghana. PLoS Negl. Trop. Dis. 5 (6), 1–13.

Spiegelhalter, D. J., Abrams, K. R., Myles, J. P., 2004. An Overview of the Bayesian Approach. John Wiley & Sons, Ltd, pp. 64–67.

Tariq, K. A., 2014. A review of the epidemiology and control of gastrointestinal nematode infections of small ruminants. Proc. Nat. Acad. Sci., India B: Biol. Sci. 85 (2), 693–703.

Torgerson, P., Schnyder, M., Hertzberg, H., 2005. Detection of anthelmintic resistance: a comparison of mathematical techniques. Vet. Parasitol. 128 (3-4), 291–298.

Torgerson, P. R., Paul, M., Furrer, R., 2014. Evaluating faecal egg count reduction using a specifically designed package 'eggCounts' in R and a user friendly web interface. Int. J. Parasitol. 44 (5), 299–303.

Torgerson, P. R., Paul, M., Lewis, F. I., 2012. The contribution of simple random sampling to observed variations in faecal egg counts. Vet. Parasitol. 188 (3-4), 397–401.

Vounatsou, P., Raso, G., Tanner, M., N'Goran, E. K., Utzinger, J., 2009. Bayesian geostatistical modelling for mapping schistosomiasis transmission. Parasitol. 136, 1695–1705.

Wang, C., Paul, M., 2016. eggCounts: Hierarchical Modelling of Faecal Egg Counts. R package version 1.1-1.
URL http://CRAN.R-project.org/package=eggCounts

Waruiru, R., Thamsborg, S., Nansen, P., Kyvsgaard, N., Bogh, H., Munyua, W., Gathuma, J., 2001. The epidemiology of gastrointestinal nematodes of dairy cattle in central Kenya. Trop. Anim. Health Prod. 33 (3), 173–187.

Wood, I., Amaral, N., Bairden, K., Duncan, J., Kassai, T., Jr., J. M., Pankavich, J., Reinecke, R., Slocombe, O., Taylor, S., Vercruysse, J., 1995. World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) second edition of guidelines for evaluating the efficacy of anthelmintics in ruminants (bovine, ovine, caprine). Vet. Parasitol. 58 (3), 181–213.

Zanzani, S. A., Gazzonis, A. L., Di Cerbo, A., Varady, M., Manfredi, M. T., 2014. Gastrointestinal nematodes of dairy goats, anthelmintic resistance and practices of parasite control in northern Italy. BMC Vet. Res. 10 (1), 1–10.

Figure captions

Fig. 1 Boxplots of the estimated FECs reduction in the unpaired design, using FECRT with approximated CI (FECRT); the hierarchical model without zero-inflation (Wang and Paul, 2016) and using the posterior median of $(1 - \bar{Y}_i^C/\bar{Y}_i^T)$ as the point estimate (PoGa(median)); the hierarchical model without zero-inflation (Wang and Paul, 2016) and using the posterior mode of $(1 - \delta)$ as the point estimate (PoGa(mode)); our proposed zero-inflated hierarchical model for the unpaired design (ZIPoGa); and parametric bootstrap (pBoot). The horizontal line indicates zero bias.

Fig. 2 Barplots of the coverage probability of the 95% CIs, or HPD credible intervals for the FECs reduction in the unpaired design. The error bars are calculated based on the 95% binomial confidence interval. The horizontal line indicates 95% coverage. The methods are the same as described in Fig. 1.

Fig. 3 Boxplots of the estimated FECs reduction in the paired design, using FECRT with approximated CI (FECRT); the hierarchical model without zero-inflation (Paul et al., 2014) and using the posterior median as the point estimate (PoGa(median)); the hierarchical model without zero-inflation (Paul et al., 2014) and using the posterior mode as the point estimate (PoGa(mode)); our proposed zero-inflated hierarchical model for the paired design (ZIPoGa); and quasi-Poisson regression (qPois). The horizontal line indicates zero bias.

Fig. 4 Barplots of the coverage probability of the 95% CIs, or HPD credible intervals in the case of Bayesian models, for the FECs reduction in the paired design. The error bars are calculated based on the 95% binomial confidence interval. The horizontal line indicates 95% coverage. The methods are the same as described in Fig. 3.

Fig. 5 Estimated reduction in mean FECs and its 95% HPD interval for the 39 flocks that were sampled both before and after treated with BZ. Using the WAAVP guideline for the decision of anthelmintic resistance, the intervals in solid black lines belong to flocks with no anthelmintic resistance, intervals in dashed lines belong to flocks with suspected resistance and intervals in solid

603 gray lines belong to flocks with resistance. The flock numbers that were flagged
604 as resistant using molecular data are colored in grey.

605 Fig. 6 The marginal posterior density for the reduction (1-δ) for flock 37 and
606 39. The shaded area represents the density mass for reduction less than 95%.