

Modelling anthelmintic resistance by extending eggCounts package to allow individual efficacy

Craig Wang^a, Paul R. Torgerson^b, Ray M. Kaplan^c, Melissa M. George^c, Reinhard Furrer^{a,d,*}

^a Department of Mathematics, University of Zurich, Winterthurerstrasse 190, Zurich, 8057, Switzerland

^b Section of Veterinary Epidemiology, Vetsuisse Faculty, University of Zurich, Winterthurerstrasse 260, Zurich, 8057, Switzerland

^c Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, GA, 30602, USA

^d Department of Computational Science, University of Zurich, Winterthurerstrasse 190, Zurich, 8057, Switzerland

ARTICLE INFO

Keywords:

Bayesian hierarchical model
Statistical analysis
Faecal egg count reduction test
Anthelmintic resistance
Simulation study

ABSTRACT

The same anthelmintic treatment can have variable efficacy on individual animals even if the parasite population is homogeneously susceptible. An extension of the R package eggCounts is proposed to take individual efficacy into account using a Bayesian hierarchical model. A simulation study is conducted to compare the performance of five different methods on estimating faecal egg count reduction and its uncertainty interval. Simulation results showed the individual efficacy model offered robust inference to two different data simulation procedures with low root mean squared error on the reduction estimate and appropriate uncertainty estimates. Different methods were used to evaluate the anthelmintic resistance in a dataset from USA with sheep and cattle faecal egg counts, where a strong anthelmintic resistance was detected. Open-source statistical tools were updated to include the proposed model.

1. Introduction

Helminth infections impose burden on human and livestock populations, and their control often relies on periodic mass administration of anthelmintics. Widespread anthelmintic resistance is now being reported (Rose et al., 2015). An important test to diagnose anthelmintic resistance is the faecal egg count reduction test (FECRT). Central to the FECRT is the estimation of the abundance of eggs, usually expressed as eggs per gram of faeces (epg) in helminth infected or potentially infected hosts. Egg counts are also used for monitoring in epidemiological studies and for developing alternative control methods that do not use anthelmintics. Thus, the accuracy of techniques to estimate eggs and hence the reduction will affect the results of such studies.

The FECRT was established in the World Association for the Advancement of Veterinary Parasitology (WAAVP) guideline (Coles et al., 1992) and is a straightforward group-based method to evaluate overall reduction of eggs. The FECRT is commonly used in practice and it has been shown to provide robust estimation of reductions in asymptotic settings (Levecke et al., 2012; Peña-Espinoza et al., 2016; Wang et al., 2017). However, it neglects some of the nowadays well-accepted issues, such as aggregation of egg counts between animals (Morgan et al., 2005; Dobson et al., 2012) and Poisson errors from sampling procedures (Torgerson et al., 2012). Improved formulations

have been proposed (Lyndal-Murphy et al., 2014; Levecke et al., 2015), but those methods do not completely address these issues. In addition, they cannot be used to obtain uncertainty estimates in some rare cases, for example, when 100% reduction is observed or when sample standard deviation of both before- and after-treatment counts are zero.

Recently, there has been an advocacy of Bayesian methods to evaluate FECRT. Bayesian methods can address those issues and provide uncertainty estimates when the FECRT cannot. At the moment, there are two open-source R packages that are commonly used for evaluating FECRT, namely eggCounts (Torgerson et al., 2014; Wang and Paul, 2018) and bayescount (Denwood et al., 2010). Peña-Espinoza et al. (2016) showed the paired model from eggCounts has poor coverage probability, as a result of narrow uncertainty intervals when analysing datasets with varying before- and after-treatment aggregation. Levecke et al. (2018) showed the model does not provide different inference when after-treatment counts are randomly reordered. Those undesirable results arise because of its model assumption, which assumes the same true efficacy (or egg count reduction) for every animal within a group. However, host related factors may effect efficacy. For example, it has been shown that the level of feed intake can effect the pharmacokinetics and hence efficacy in sheep (Ali and Hennessy, 1996). Diet has also been shown to effect the pharmacokinetics of benzimidazoles (e.g. Oukessou and Chkounda, 1997). Thus, even if all animals in a

* Corresponding author. Department of Mathematics, University of Zurich, Winterthurerstrasse 190, Zurich, 8057, Switzerland.

E-mail address: reinhard.furrer@math.uzh.ch (R. Furrer).

population are infected by equally susceptible helminth parasites, there may be variation in efficacy between individual animals due to variations in feed intake or other factors. Furthermore, the species composition of helminth fauna may not have the same proportions in every animal in a group and drug efficacy may vary between helminth species (Vidyashankar et al., 2012). The limitation of assuming the same true efficacy in humans was also discussed by Krücken et al. (2017) in a study on school children. In such a population, efficacy in individuals may have greater heterogeneity, as the children were from a number of different villages attending different schools, compared to a more homogeneous population of sheep grazing a single pasture.

Currently the only choice of Bayesian hierarchical model allowing for individual efficacy is by Geurden et al. (2015) and is implemented within the R package bayescount. However, the model recommends for replicated samples for individual animals (Denwood, 2015). The package also contains models that allow for varying before- and after-treatment aggregation, which can be a consequence of having individual efficacy. Peña-Espinoza et al. (2016) performed a simulation study comparing the paired and unpaired models in the eggCounts package with bayescount varying-aggregation model. However, the simulated data were generated under the assumption of the bayescount model, with a compound gamma-gamma-Poisson distribution and varying after-treatment aggregation. This violated the assumptions of the eggCounts model, which created a mismatch between the data and the model structure. Therefore, with such a violation, it is reasonable to expect eggCounts models to have inferior performance. The bayescount model provided better coverage probability, but no results regarding bias or mean squared error was provided in their simulation study. In this paper, we propose a new Bayesian hierarchical model that takes individual efficacy into account, while providing robust estimation of reduction in egg counts. We further provide a refined simulation study, with improvements reflected in two aspects. Firstly, we simulate datasets with two different procedures, one with our model assumptions and another with bayescount model assumptions. This encourages fair comparison between the methods. Secondly, we provide comparison of root mean squared error (RMSE) on reduction estimates in addition to coverage probabilities. This evaluates the methods in terms of both bias and variance.

The aim of the present study is to propose and to evaluate an improved Bayesian hierarchical model that addresses individual efficacy between animals. During this process, 1) we determine a suitable distribution for individual reduction of egg counts based on real dataset; 2) we provide a fair and thorough comparison of existing Bayesian and non-Bayesian methods against our proposed model; 3) we analyse a recent dataset to determine its anthelmintic resistance status; and 4) we offer updated open-source tools for evaluating FECR and general advice for using Bayesian hierarchical models.

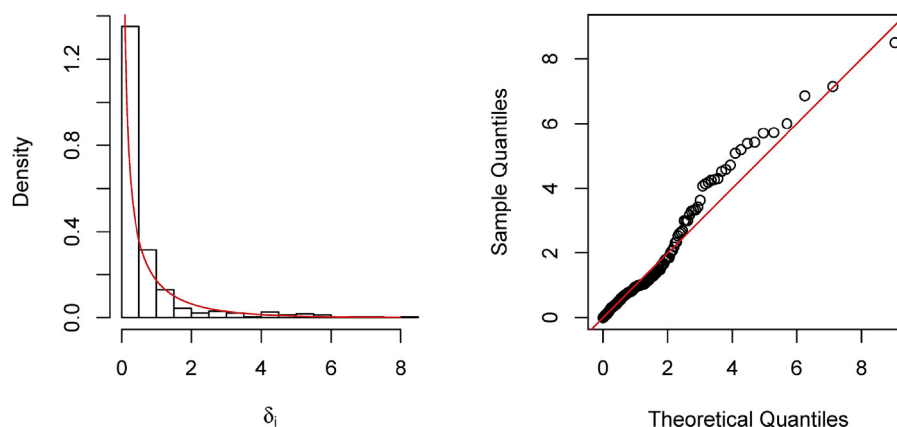


Fig. 1. Fitted gamma distribution. Left: histogram of individual reductions with fitted gamma distribution shown in red. Right: empirical quantiles with data points against fitted theoretical gamma quantiles, the $y = x$ line is shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2. Materials and methods

2.1. Data description

The before- and after-treatment faecal egg count data were collected from 287 sheep and 212 cattle, a total of 499 animals in 28 FECRTs (14 sheep FECRTs and 14 cattle FECRTs). All sheep tested were housed at the University of Georgia Sheep Unit, Athens, Georgia, USA. Sampling of sheep occurred between August 2011 and June 2015. Of the 14 FECRTs completed on sheep, 6 were completed with the high sensitivity 3 chamber modified-McMaster method (8 eggs per gram sensitivity) and 8 were completed with the 2 chamber modified-McMaster method (25 eggs per gram sensitivity). For modified-McMaster analysis, 4 g of faeces were placed into a cup with 26 mL (25 eggs per gram sensitivity) or 25 mL (8 egg per gram sensitivity) of sodium nitrate flotation solution (specific gravity = 1.25–1.30, FECA-MED, Vedco, Inc., St. Joseph, Missouri, USA). Homogenization of sample and sodium nitrate solution, slide preparation, and counting were completed as previously described (Noel et al., 2017). Cattle tests were performed on weaned stocker cattle on properties in Georgia and Mississippi, USA as part of a study addressing the use of composite faecal samples when performing a FECRT (George et al., 2017). Faecal egg counts were performed as described by George et al. (2017).

2.2. Proposed model

2.2.1. Determine the distribution of treatment efficacy

Using pooled data, we empirically select a suitable distribution to model treatment efficacy of individuals. It is common to firstly model the proportion of egg counts remaining after treatment, then subtract the proportion from 1 to obtain the reduction of egg counts. We denote this proportion as δ_i for the i th animal and compute it for each animal. 30 observations with zero before-treatment counts are excluded since these reductions are not well defined. In addition, 6 observations which had more than 15-fold increase in after-treatment counts were also excluded. Finally, a goodness-of-fit statistic was used to determine a suitable distribution based on the empirical cumulative distribution function. Several candidate distributions were considered, including gamma, inverse gamma, half-Cauchy, and half-normal distributions. Based on the computed proportions from 463 pairs of counts, the best fitted distribution according to Cramér-von Mises distance is the gamma distribution as shown in Fig. 1. A gamma distribution is then used to model the reduction in our proposed Bayesian hierarchical model.

2.2.2. Bayesian hierarchical model formulation

Suppose we have a group of animals from the same species with sample size n , and their true epg follow some unimodal distribution. For notational simplicity, the analytical sensitivity is assumed to be the same throughout the sampling procedures. A faecal egg count (FEC) is obtained twice from each animal with analytical sensitivity f , once before applying treatment and once some days after treatment. We denote the raw number of counts as Y_i^{*C} and Y_i^{*T} respectively, with $i = 1, 2, \dots, n$. Given the true epg of sampled faeces before treatment Y_i^C , Y_i^{*C} follows a binomial distribution with size Y_i^C and probability $1/f$. This captures the counting variability, where the number of eggs counted from diluted homogeneous samples is random. Then the true epg of sampled faeces Y_i^C follows a Poisson distribution with latent mean μ_i , this addresses Poisson error, which arises because of randomly distributed eggs within the faecal sample. Finally, the latent mean μ_i follows a gamma distribution with shape κ and rate κ/μ , which has mean μ and variance μ^2/κ . The gamma distribution captures FEC aggregation between animals. After treatment, we expect a change in the latent mean of each animal. We take a random effect model approach and assume their treatment efficacy is different. The raw number of counts Y_i^{*T} follow another binomial distribution with size Y_i^T and probability $1/f$, where the true epg of sampled faeces after treatment Y_i^T follows a Poisson distribution with latent mean $\delta_i\mu_i$. The new latent mean allows each animal to have its own treatment efficacy. The efficacies δ_i follows another gamma distribution as determined in section 2.2.1. with shape τ and rate τ/ν .

The proposed Bayesian hierarchical model is formulated as,

$$Y_i^{*C} \mid Y_i^C \sim \text{Bin}(Y_i^C, 1/f),$$

$$Y_i^C \mid \mu_i \sim \text{Pois}(\mu_i),$$

$$Y_i^{*T} \mid Y_i^T \sim \text{Bin}(Y_i^T, 1/f),$$

$$Y_i^T \mid \delta_i, \mu_i \sim \text{Pois}(\delta_i\mu_i),$$

$$\mu_i \mid \kappa, \mu \sim \text{Gamma}(\kappa, \kappa/\mu),$$

$$\delta_i \sim \text{Gamma}(\tau, \tau/\nu).$$

The median of reduction δ_i is computed based on the quantile function of a gamma distribution, which then is used as the reduction estimate. The model is implemented in R (R Core Team, 2018) using Stan modelling language (Carpenter et al., 2017), which is a modern computational tool for conducting Bayesian inference.

2.3. Simulation study

A simulation study is conducted to investigate the performance of the proposed Bayesian hierarchical model. The proposed model is compared against four other existing methods. Faecal egg counts are simulated under two different procedures, one that matches our model assumptions with individual efficacy and another that matches bayes-count model assumptions. Generally speaking, models that have the same assumptions as the simulation procedure have a natural advantage over models that assumes otherwise. Finally, the simulation parameters are also varied within each procedure to cover a broader range of possible scenarios.

The simulation is setup as follows. Procedure one: for each animal before treatment, we firstly draw random samples of latent mean epg μ_i from a gamma distribution with shape κ and rate κ/μ . Then the true epg from observed sample y_i^C is drawn from a Poisson distribution with mean μ_i . The observed count y_i^C is simulated from a binomial distribution with probability $1/f$ and size y_i^C . Since this is a paired design, the latent mean epg μ_i remains the same for each animal after treatment. We assume each animal experiences various levels of efficacy, hence the true epg from observed sample after treatment y_i^T is now simulated from a Poisson distribution with mean $\delta_i\mu_i$, where δ_i is

sampled from a gamma distribution with shape τ and rate τ/ν . Finally, the observed count y_i^T is simulated from a binomial distribution with probability $1/f$ and size y_i^T . Procedure two: we draw again random samples of μ_i from a gamma distribution with shape κ and rate κ/μ . The true epg from observed sample y_i^C follows a gamma-Poisson (i.e. negative binomial) distribution mean μ_i and dispersion κ^C , the level of aggregation increases with decreasing κ^C . After treatment, y_i^T is simulated from another gamma-Poisson distribution with updated mean $\nu\mu_i$ and aggregation κ^T . The observed counts are again simulated from a binomial distribution. Such procedure allows a change of aggregation before and after treatment.

For each procedure, we simulate FEC datasets with eight combinations of parameters, each repeated 500 times. Specifically, we use latent mean epg for the group $\mu = \{150, 500\}$, between animal aggregation $\kappa = \{1, 2\}$ and mean reduction $\nu = \{0.03, 0.15\}$. For procedure one, τ is generated from a uniform distribution between 0.5 and 2.0 for each simulated dataset, representing large to medium variability in efficacy. This leads to *substantially* to *moderately* increased FEC aggregation after treatment. For example, when $\tau = 0.5$ and the before-treatment aggregation $\kappa = 1$, the observed after-treatment aggregation is reduced to 0.2. When $\tau = 2$, the observed after treatment aggregation is only reduced to 0.5. For procedure two, there is no variability in efficacy itself hence a single ν is used. We set $\kappa^C = \kappa^T = \kappa$ to represent cases of *no change* in FEC aggregation after treatment.

We assign a Beta(1,1) prior (equivalent to Uniform(0,1)) to the mean reduction ν . We adapt the prior of μ and κ from Wang et al. (2017). A weakly informative Gamma(1,0.001) is assigned to μ where 90% of the probability mass lies between 60 and 3000. A weakly informative Gamma(1,0.7) prior is assigned to κ , where 90% of probability mass lies between 0.1 and 4.3, corresponding to substantial and little aggregation. The shape parameter of reduction τ is often weakly identified by faecal egg count data due its small sample sizes in practice, hence we assign a moderately informative prior zero-truncated Normal(2,1), with 90% of probability mass lies between 0.36 and 3.6, corresponding to large and small variability in efficacy. This serves to regularize the model and enhance its stability without interfere with the result.

Next, we describe four additional existing methods to analyse the simulated FEC datasets for comparison.

2.3.1. Faecal egg count reduction test

According to the WAAVP guideline (Coles et al., 1992), the percentage reduction in FECs can be calculated by $1 - \bar{y}^T/\bar{y}^C$, where \bar{y}^C and \bar{y}^T are arithmetic means of before- and after-treatment counts. Its 95% uncertainty interval can be computed based quantiles of Student's t-distribution with $2n - 2$ degrees of freedom and asymptotic variance of their log ratio. This method is performed using *fecrtCI()* function in the *eggCounts* package.

2.3.2. Asymptotic variance method

The asymptotic variance method by Levecke et al. (2015) assumed a Poisson-negative binomial distribution of the observed FECs, and a single reduction within a population. The authors derived asymptotic expectation and variance of the reduction, and then computed its uncertainty interval based on quantiles of a gamma distribution. The shape and scale parameters of the gamma distribution are found via moment matching. The asymptotic expectation is $1 - \bar{y}^T/\bar{y}^C$, which is the same as the estimated reduction from FECRT. The asymptotic variance is

$$\left(\frac{\bar{y}^T}{\bar{y}^C}\right)^2 \left(\left(\frac{s^C}{\bar{y}^C}\right)^2 + \left(\frac{s^T}{\bar{y}^T}\right)^2 - 2\text{Cor}(y^C, y^T) \cdot \frac{s^C \cdot s^T}{\bar{y}^C \cdot \bar{y}^T} \right),$$

where s^C , s^T are the sample standard deviation of the before- and after-treatment counts respectively. $\text{Cor}(y^C, y^T)$ is the sample correlation between before- and after-treatment counts.

2.3.3. Model from bayescount package

The bayescount model assumes a compound gamma-gamma-Poisson distribution, which matches with the procedure two of our simulation study. The individual efficacy model implemented in bayescount recommends replicated samples for individual animals, which is not the spirit of our study, hence we select the model with fixed efficacy but allowing different aggregation. The level of aggregation is strongly associated with the amount of variability in efficacies, hence this does not put the bayescount model in an unfavourable situation. The function `fecrt.model()` is used with setting `paired.model`, `fix.controls` and `fix.efficacy` all set to TRUE, while `fix.variation` and `zero.inflation` set to FALSE. Default weakly informative priors were used for the mean parameters in the simulation study, while the default Beta(1,1) prior is used for the reduction. The same Gamma(1,0.7) prior from our individual efficacy model is used for the dispersion parameters to facilitate model comparison. Simulation results using default precision prior from bayescount is in the Supplementary Material.

2.3.4. Model from eggCounts package

The eggCounts paired model is a special case of our proposed model with a simplifying assumption. Instead of a gamma-distributed reduction δ_i , a single reduction is assumed for every animal. The function `fecr.stan()` is used to analyse the simulated FEC datasets with setting `paired` set to TRUE and `zero.inflation` set to FALSE.

The performance of our proposed individual efficacy model is compared against those four existing methods, in terms of the following criteria: 1) RMSE on the reduction estimates of FEC; 2) coverage probability of 95% confidence interval for the case of non-Bayesian methods and 95% credible interval for the case of Bayesian models, they are collectively referred to as 95% uncertainty intervals (UIs); and 3) the width of UIs.

3. Results

3.1. Simulation results

The performance criteria are evaluated based on the simulation results. Fig. 2 shows the width of 95% UIs and Fig. 3 shows their coverage probabilities. Table 1 shows the RMSE based on their FECR estimates. While generally one should not expect Bayesian credible intervals to have the same frequentist properties, namely having 95% coverage probability, we still compare them here since the estimation of uncertainty for FECR is important in the decision-making process of detecting anthelmintic resistance.

The width of 95% UIs differ significantly between different methods. Our proposed individual efficacy model has moderate UI width. The eggCounts model has very narrow intervals across all the scenarios while the bayescount model has wide UIs under both simulation procedures, which are consistent with the simulation study in Peña-Espinoza et al. (2016). The asymptotic variance method and the FECRT sometimes have extremely wide intervals when the true reduction is 85%. There are 14 and 89 intervals respectively that are wider than 1 hence lie beyond the range of y-axis displayed in Fig. 2. An untruncated version of the figure is provided in Supplementary Material. Procedure two simulates the datasets with two separate gamma distributions, hence more variability in observed counts is expected compared to procedure one. The UIs have reasonable width when the true reduction is 97%. However, 69 out of 4000 (500×8) simulations resulted in undefined UIs with the asymptotic variance and the FECRT method. This occurs when all of the after-treatment counts are zero such that the sample variance is undefined.

Fig. 3 shows the coverage probabilities of those UIs, i.e. the percentage of datasets with the true reduction being within the estimated UI. As it can be expected based on the width of their UIs, the eggCounts model has low coverage probability while the bayescount model has high coverage probability. As a result of their differences in model

assumption, the eggCounts model provides insufficient coverage probabilities for both procedures. The bayescount model has almost 100% coverage when the datasets are simulated under procedure one, while it has reasonable coverage under procedure two. The asymptotic variance method also consistently provides insufficient coverage probability, but improved upon the eggCounts model. In contrast, the individual efficacy model has consistently reasonable coverage under both simulation procedures. Based on both Figs. 2 and 3, our individual efficacy model can achieve good coverage probability with much narrower UIs, while the bayescount model provides too wide intervals albeit having good coverages.

As shown in Tables 1 and 2, the level of RMSE are comparable across the methods, with bayescount model having higher RMSE when simulated under procedure two. Our proposed individual efficacy model has small RMSE for all the scenarios. When simulating using procedure one, the individual efficacy model has lowest RMSE. The model also has achieved low RMSE as well when simulating under procedure two, which uses the compound gamma-gamma-Poisson distribution.

3.2. Evaluating FECR for the USA dataset

The dataset consists of 28 flocks. Flocks here refer to both sheep flocks and cattle herds. Exploratory data analysis showed the flock-wise median reduction ranges from 100% to -322%, corresponding to complete elimination of egg counts in the sample and a 2.22-fold increase in egg counts after treatment. Within each flock, the efficacy between animals also varied dramatically. Since individual efficacy is not a part of original eggCounts model assumption, it is not considered in this evaluation. We also drop the asymptotic variance method in favour of the more widely used FECRT. In order to allow for increased egg counts in the FECR estimation, we place a Uniform(0,4) prior on ν in the individual efficacy model and the bayescount model. In practice, one should carefully check if the pre-assigned prior includes the possible outcome of the current dataset. Fig. 4 shows the estimated FECR along with UIs from four different methods. Their estimates are mostly similar, with the individual efficacy model having smaller UIs in most cases compared to the bayescount model.

Flock number 4, 5 and 28 result some discrepancies among the models. Flock 4 consists of 15 animals, two of already-heavily infected animals had a 4-fold and a 7-fold increase in egg counts respectively after treatments. Since FECRT does not explicitly model the pairwise relationship, the overall reduction is strongly influenced by those two animals. This results in a higher after-treatment mean egg counts in the flock hence a negative reduction estimate. In flock 5, 18 out of 19 animals experienced an increase in their after-treatment FECs, with a mean increase of 230%. There are 4 out of 19 animals that had less than 100% increase. The bayescount model was influenced by those observations hence the estimate is biased towards smaller increase. Since our individual efficacy model uses median reduction, it provides a reasonable estimate and it is similar to the result from FECRT. Flock 28 consists of 13 animals, where all of the after-treatment counts are zero, while there are only 3 non-zero before-treatment counts. Bayescount model provides an extremely wide UI. This observation is consistent with the simulation study, where bayescount model generally provides much wider UIs especially when the before-treatment mean count is low and reduction is large. Both the individual efficacy model and the FECRT provide reasonable estimates for flock 28, but only former is able to provide a valid UI.

4. Discussion

The efficacies of anthelmintic treatments can be different between animals due to the factors discussed previously. In this paper, we proposed an individual efficacy model within a Bayesian hierarchical framework to analyse FECR. The model extended upon the original

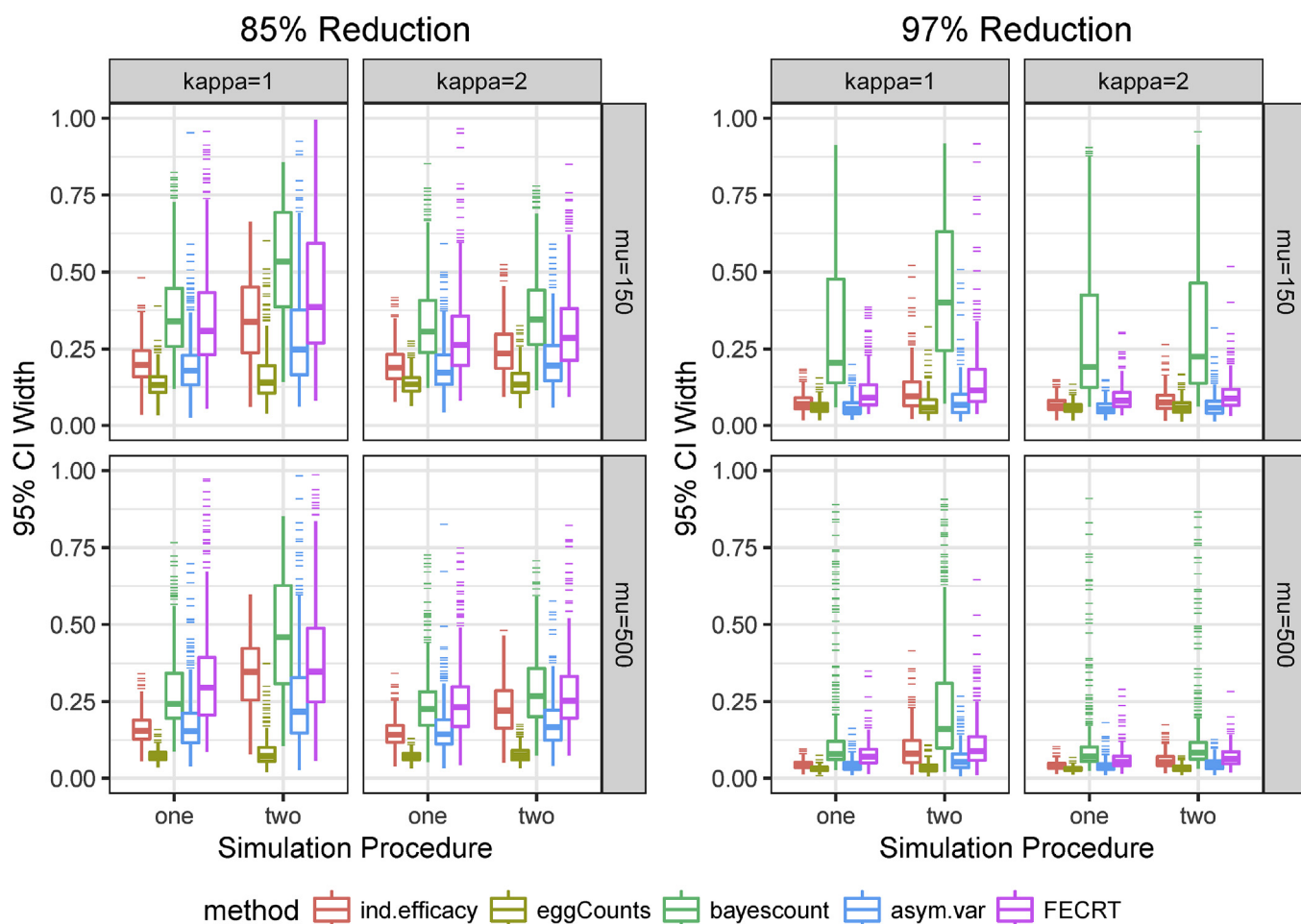


Fig. 2. Uncertainty interval widths of simulations. Boxplots of 95% uncertainty interval width, grouped by true reduction percentage, simulation procedure, before-treatment mean and dispersion. Data outside the y-axis range are not shown. An untruncated version of the figure is available in Supplementary Material.

eggCounts paired model to allow varying individual efficacy and offers robust estimation of the reduction in egg counts.

The original eggCounts model assumes every animal in the same group has the same reduction, consequently there is a perfect correlation between the latent before- and after-treatment egg. Although this does not automatically suggest that observed counts are also perfectly correlated due to Poisson errors in the sampling process, it does indicate a strong correlation is expected. The correlation approaches 1 asymptotically as the observed FECs increases. In our pooled sample of FECs, we observe a highly significant correlation of 0.417 ($p < 0.001$) of the observed counts, nevertheless it is much lower than 1, which indicates that the individual efficacy model we propose is a reasonable approach to model this correlation and will give more robust results than the assumption of equal efficacy across all individuals.

The original eggCounts model provided narrow UIs on the reduction estimate in all scenarios. The small uncertainty can be explained by its model assumption, where a single reduction parameter is assumed for every animal within the same group. Hence the variation is forcibly captured by the dispersion parameter such that the uncertainty for reduction remains relatively small. The bayescount model has the flexibility of allowing varying before- and after-treatment aggregation. However, in our simulation study with two different simulation procedures, the model with varying aggregation consistently provided wide UIs and slightly higher RMSE, which can lead to undesirable results of having many inconclusive decisions. One possible reason is that the model has more hierarchical layers. With limited data available, there is hardly any information on latent parameters of the model. In addition, it uses mean reduction as the reduction estimate, this exposes

it to be easily influenced by extreme observations. For the bayescount model, we have also experimented with *fix.efficacy* set to FALSE while *fix.variation* set to TRUE, and it did not lead to significantly different results, suggesting the different efficacy between animals can be effectively addressed by either allowing for varying aggregation or varying efficacy itself. The FECRT is based on straightforward computation of the after- and before-treatment mean counts ratio, hence it neglects features of the data generating process, such as individual efficacy, between-animal aggregation and Poisson errors. The simulation showed relatively good performance for the FECRT, but the method also generated undefined estimate and wider than 100% UIs. The asymptotic variance method improved upon the FECRT and captures some of the data generating process. However, it suffers the same weakness as the FECRT which can generate undefined reduction estimates. Meanwhile, our newly proposed individual efficacy model does not suffer the constraint of having the same efficacy for every animal, which avoids the potential mismatch between model assumption and data, hence leads to improvement in coverage probability.

When simulating under procedure two, separate gamma-Poisson distributions are assumed for before- and after-treatment mean. This leads a difference in the observed reduction, which is “heavy-tailed” as a result of over-dispersion. When an extreme reduction is generated in a dataset, the mean estimate of FECR can be strongly influenced by this single observation. Although the bayescount method can effectively model the distribution in procedure two, it tends to generate biased estimate of FECR hence having higher RMSE. However, for the original eggCounts and the asymptotic variance/FECRT method, the pairwise before- and after-treatment relationship is not explicitly modelled,

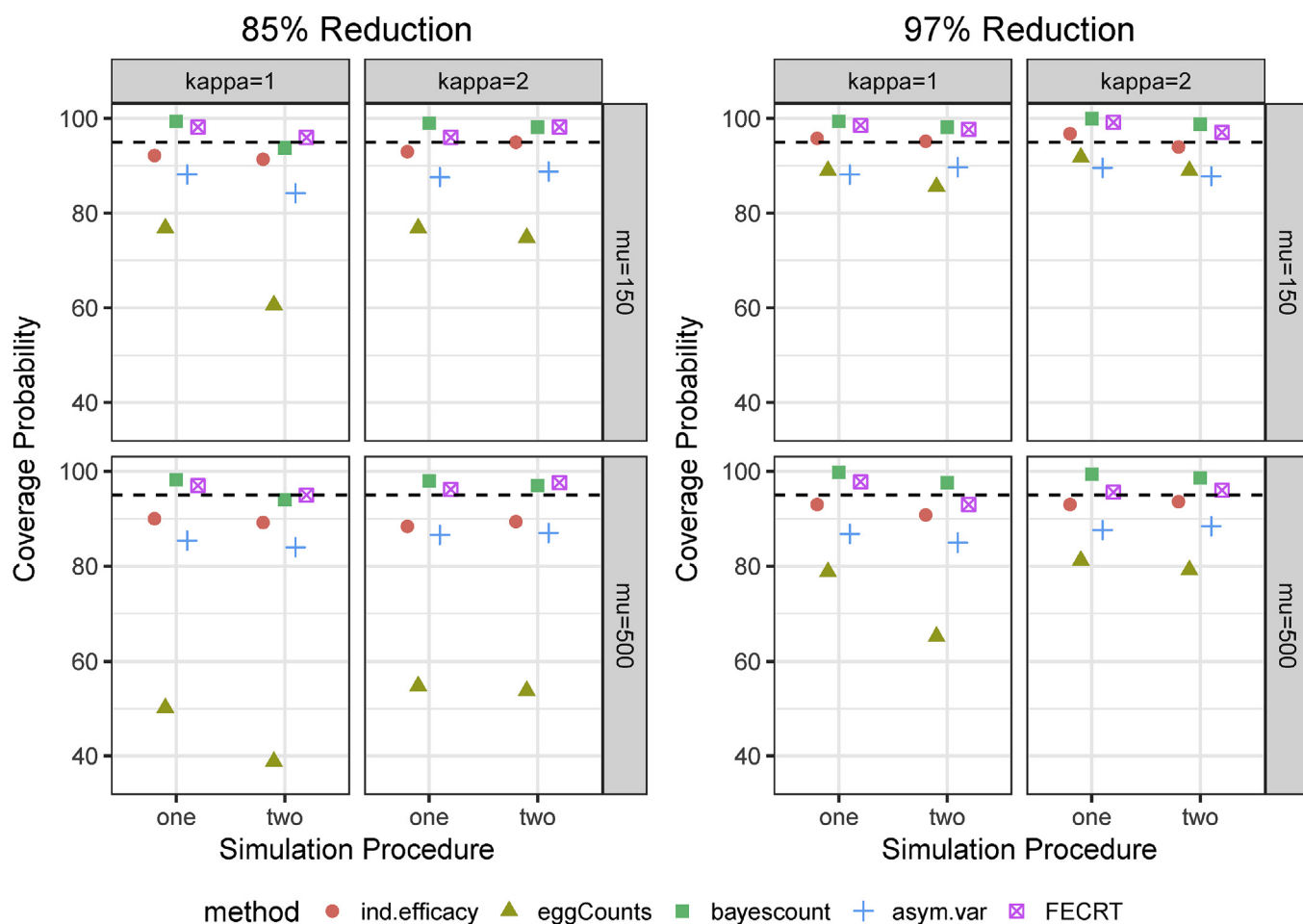


Fig. 3. Coverage probability of simulations. Coverage probabilities of 95% credible or confidence intervals, grouped by true reduction percentage, simulation procedure, before-treatment mean and dispersion parameter.

hence mitigates this problem. Our individual efficacy model uses median reduction as the reduction estimate, providing robust inference while take the pairwise relationship into account. The simulation study showed superior performance of the individual efficacy model compared to other methods, it has low RMSE and reasonable UIs across all the scenarios under both simulation procedures. Another advantage of Bayesian models is that they allow a probabilistic view on the anthelmintic resistance status. Rather than providing an estimated reduction and 95% UI as in the asymptotic variance/FECRT methods, the Bayesian models provide entire distributions of the estimated reduction, which allows for the computation of anthelmintic resistance probability given an arbitrary threshold.

When evaluating the present dataset, the FECR estimates were similar across the individual efficacy model, bayescount model and the FECRT. This particular dataset contains flocks with increased after-

treatment FECs, such that the commonly used Beta(1,1) prior for the reduction is deemed inappropriate. Therefore, after exploring the dataset, we selected a Uniform(0,4) prior to allow for up to 3-fold increased after-treatment mean egg. In practice, researchers should apply Bayesian models with care. First of all, every Bayesian model comes with assumptions and constraints from the model structure itself and from the priors. For example, Wang et al. (2017) proposed zero-inflation models that deals with over-represented zero counts; the egg-Counts paired model assumes the same reduction for every animal. As shown in our simulation study, the paired model provides overly-narrow UIs when the assumption is violated, which can lead to misclassification in the end. In terms of prior constraints, researchers should check if the priors at least cover the possible parameter values observed from the dataset. For example, the default prior for modelling reduction is Beta(1,1). This does not allow for increased mean egg after

Table 1

Root mean squared error of the faecal egg count reduction estimates from five different methods, based on 500 simulations for each scenario under procedure one.

Procedure One			Individual efficacy	eggCounts	Bayescount	asym.variance/FECRT
85% reduction	$\kappa = 1$	$\mu = 150$	0.0519	0.0602	0.0643	0.0602
		$\mu = 500$	0.0457	0.0565	0.0522	0.0565
	$\kappa = 2$	$\mu = 150$	0.0515	0.0573	0.0660	0.0579
		$\mu = 500$	0.0464	0.0486	0.0483	0.0487
97% reduction	$\kappa = 1$	$\mu = 150$	0.0171	0.0197	0.0309	0.0197
		$\mu = 500$	0.0113	0.0123	0.0211	0.0122
	$\kappa = 2$	$\mu = 150$	0.0149	0.0167	0.0297	0.0168
		$\mu = 500$	0.0110	0.0131	0.0171	0.0128

Table 2

Root mean squared error of the faecal egg count reduction estimates from five different methods, based on 500 simulations for each scenario under procedure two.

Procedure Two			Individual efficacy	eggCounts	Bayescount	asym.variance/FECRT
85% reduction	$\kappa = 1$	$\mu = 150$	0.1011	0.1020	0.1488	0.1025
		$\mu = 500$	0.1053	0.0945	0.1360	0.0942
	$\kappa = 2$	$\mu = 150$	0.0607	0.0601	0.0846	0.0606
		$\mu = 500$	0.0689	0.0562	0.0737	0.0565
97% reduction	$\kappa = 1$	$\mu = 150$	0.0297	0.0268	0.0593	0.0269
		$\mu = 500$	0.0258	0.0214	0.0450	0.0214
	$\kappa = 2$	$\mu = 150$	0.0203	0.0204	0.0381	0.0204
		$\mu = 500$	0.0148	0.0139	0.0217	0.0139

treatment. In evaluating the present dataset, an alternative strategy is to use Beta(1,1) prior only for those flocks with reduced after-treatment counts, and use Uniform(0,4) prior only for increased after-treatment counts. We expect the flocks with Beta(1,1) prior to have a slightly narrower UIs compared to our analysis. Arguably this may not be of clinical relevance because an increase in FECs after treatment clearly indicates lack of anthelmintic efficacy. However, the use of this individual efficacy model would not be confined to anthelmintic efficacy testing but could also be used for experiments where alternatives to anthelmintics are used for parasite control and hence allowing for an increase in FECs with the Uniform(0,4) prior adds additional flexibility. Secondly, researchers should be aware of the relevant contribution of information from priors and data. This can be checked by plotting posterior parameter distributions overlapped with corresponding prior distributions (See section 4 in Wang and Furrer, 2018). While our default priors are weakly informative (except for τ which is moderately informative) for most faecal egg count datasets, the relative contribution of information also depends on the data available (Gelman et al., 2017). Researchers should ensure the priors used are sensible for their

data. Another potential issue with Bayesian models implemented using Markov chain Monte Carlo (MCMC) sampling is that there can be convergence problems. This can occur when there is limited information available, including low counts and small sample sizes. The issues can be detected by checking the effective sample size of MCMC chains (Gelman et al., 2014) and potential scale reduction factors (Brooks and Gelman, 1998). The updated eggCounts package version 2.0 which implemented the individual efficacy model also has built-in warning messages when convergence problems occur. We agree with Peña-Espinoza et al. (2016), that unexperienced users should seek statistical advice when problems occur, otherwise unreliable results could be obtained. When there are convergence problems, a possible solution is to further constrain the priors. In some cases, Monte Carlo simulation can be used to obtain some information about the weakly identified latent parameter from the data. This leads to empirical Bayes approach (Carlin and Louis, 1997).

The proposed individual efficacy model was extended based on the predecessor eggCounts paired model, it is shown to have superior performance in both the simulation study and when analysing the

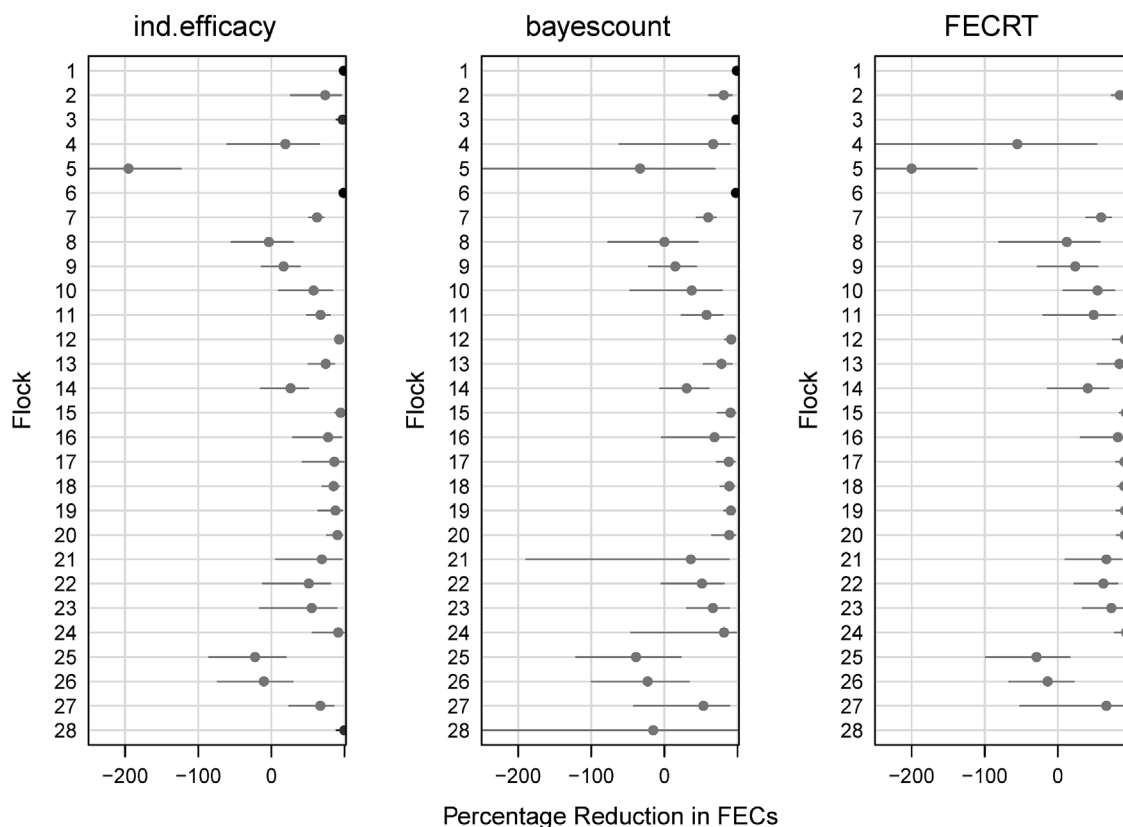


Fig. 4. Analysis results of USA data. Analysis results of faecal egg count reduction from the individual efficacy model, bayescount model and FECRT. Grey colored uncertainty intervals and reduction estimates indicate the flock is classified as anthelmintic resistant according to the WAAVP guideline, while black colored indicates the flock is classified as non-resistant, and there were no flocks having suspected resistance.

present dataset. The eggCounts package version 2.0 (Wang and Paul, 2018) and the online analysis tool (Torgerson et al., 2014; Furrer et al., 2016) have been updated to include the newly proposed individual efficacy model, and a vignette to guide using the eggCounts package was made available (Wang and Furrer, 2018). As we learn more about the different generating process of egg counts, the model used for FECR evaluation should be continuously iterated to ensure model assumptions closely matches the actual data generating process.

Conflicts of interest

The authors declare that they have no conflicts of interests.

Acknowledgements

This work was supported by the Swiss National Science Foundation [grant number: 175529]. We gratefully thank Matthew Denwood and the editorial team for valuable comments on an earlier version of this manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ijpdr.2018.07.003>.

References

- Ali, D.N., Hennessy, D.R., 1996. The effect of level of feed intake on the pharmacokinetic disposition and efficacy of ivermectin in sheep. *J. Vet. Pharmacol. Therapeut.* 19, 89–94. <https://doi.org/10.1111/j.1365-2885.1996.tb00018.x>.
- Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph Stat.* 7, 434–455.
- Carlin, B.P., Louis, T. a., 1997. Bayes and empirical Bayes methods for data analysis. *Stat. Comput.* 7, 153–154. <https://doi.org/10.1023/A:1018577817064>.
- Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: a probabilistic programming language. *J. Stat. Software* 76, 1–32. <https://doi.org/10.18637/jss.v076.i01>.
- Coles, G.C., Bauer, C., Borgsteede, F.H.M., Geerts, S., Klei, T.R., Taylor, M.A., Waller, P.J., 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, vol 44. pp. 35–44. [https://doi.org/10.1016/0304-4017\(92\)90141-U](https://doi.org/10.1016/0304-4017(92)90141-U).
- Denwood, M., 2015. Bayescount: power calculations and bayesian analysis of count distributions and FECRT data using MCMC. <http://cran.r-project.org/package=Bayescount>.
- Denwood, M.J., Reid, S.W.J., Love, S., Nielsen, M.K., Matthews, L., McKendrick, I.J., Innocent, G.T., 2010. Comparison of Three Alternative Methods for Analysis of Equine Faecal Egg Count Reduction Test Data, vol 93. pp. 316–323. <https://doi.org/10.1016/j.prevetmed.2009.11.009>.
- Dobson, R.J., Hosking, B.C., Jacobson, C.L., Cotter, J.L., Besier, R.B., Stein, P.A., Reid, S.A., 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, 79–92. <https://doi.org/10.1016/j.vetpar.2011.11.049>.
- Furrer, R., Flury, R., Wang, C., 2016. Modelling faecal egg counts with shiny. <http://t.uzh.ch/D1>.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2014. *Bayesian Data Analysis*. CRC Press.
- Gelman, A., Simpson, D., Betancourt, M., 2017. The prior can often only be understood in the context of the likelihood. *Entropy* 19. <https://doi.org/10.3390/e19100555>.
- George, M.M., Paras, K.L., Howell, S.B., Kaplan, R.M., 2017. Utilization of composite fecal samples for detection of anthelmintic resistance in gastrointestinal nematodes of cattle. *Vet. Parasitol.* 240, 24–29. <https://doi.org/10.1016/j.vetpar.2017.04.024>.
- Geurden, T., Chartier, C., Fanke, J., di Regalbono, A.F., Traversa, D., von Samson-Himmelstjerna, G., Demeler, J., Vanimisetti, H.B., Bartram, D.J., Denwood, M.J., 2015. Anthelmintic resistance to ivermectin and moxidectin in gastrointestinal nematodes of cattle in Europe. *Int. J. Parasitol. Drugs Drug Resist.* 5, 163–171. <https://doi.org/10.1016/j.ijpdr.2015.08.001>.
- Krücken, J., Fraundorfer, K., Mugisha, J.C., Ramünke, S., Sift, K.C., Geus, D., Habarugira, F., Ndoli, J., Sendegeya, A., Mukampungu, C., Bayingana, C., Aebischer, T., Demeler, J., Gahutu, J.B., Mockenhaupt, F.P., von Samson-Himmelstjerna, G., 2017. Reduced efficacy of albendazole against *Ascaris lumbricoides* in Rwandan schoolchildren. *Int. J. Parasitol. Drugs Drug Resist.* 7, 262–271. <https://doi.org/10.1016/j.ijpdr.2017.06.001>.
- Levecke, B., Anderson, R.M., Berkvens, D., Charlier, J., Devleeschauwer, B., Speybroeck, N., Vercruysse, J., Aelst, S. Van, 2015. Chapter five - mathematical inference on helminth egg counts in stool and its applications in mass drug administration programmes to control soil-transmitted helminthiasis in public health. In: *Mathematical Models for Neglected Tropical Diseases: Essential Tools for Control and Elimination*, Part A. Academic Press, pp. 193–247. <https://doi.org/10.1016/bs.ap.2015.01.001>.
- Levecke, B., Dobson, R.J., Speybroeck, N., Vercruysse, J., Charlier, J., 2012. Novel insights in the faecal egg count reduction test for monitoring drug efficacy against gastrointestinal nematodes of veterinary importance. *Vet. Parasitol.* 188, 391–396. <https://doi.org/10.1016/j.vetpar.2012.03.020>.
- Levecke, B., Kaplan, R.M., Thamsborg, S.M., Torgerson, P.R., Vercruysse, J., Dobson, R.J., 2018. How to improve the standardization and the diagnostic performance of the faecal egg count reduction test? *Vet. Parasitol.* 253, 71–78. <https://doi.org/10.1016/j.vetpar.2018.02.004>.
- Lyndal-Murphy, M., Swain, A.J., Pepper, P.M., 2014. Methods to determine resistance to anthelmintics when continuing larval development occurs. *Vet. Parasitol.* 199, 191–200. <https://doi.org/10.1016/j.vetpar.2013.11.002>.
- Morgan, E.R., Cavill, L., Curry, G.E., Wood, R.M., Mitchell, E.S.E., 2005. Effects of aggregation and sample size on composite faecal egg counts in sheep. *Vet. Parasitol.* 131, 79–87. <https://doi.org/10.1016/j.vetpar.2005.04.021>.
- Noel, M.L., Scare, J.A., Bellaw, J.L., Nielsen, M.K., 2017. Accuracy and precision of mini-FLOTAC and McMaster techniques for determining equine strongyle egg counts. *J. Equine Vet. Sci.* 48, 182–187. e1. <https://doi.org/10.1016/j.jevs.2016.09.006>.
- Oukessou, M., Chkounda, S., 1997. Effect of diet variations on the kinetic disposition of oxfendazole in sheep. *Int. J. Parasitol.* 27, 1347–1351. [https://doi.org/10.1016/S0020-7519\(97\)00107-0](https://doi.org/10.1016/S0020-7519(97)00107-0).
- Peña-Espinoza, M., Thamsborg, S.M., Denwood, M.J., Drag, M., Hansen, T.V., Jensen, V.F., Enemark, H.L., 2016. Efficacy of ivermectin against gastrointestinal nematodes of cattle in Denmark evaluated by different methods for analysis of faecal egg count reduction. *Int. J. Parasitol. Drugs Drug Resist.* 6, 241–250. <https://doi.org/10.1016/j.ijpdr.2016.10.004>.
- R Core Team, 2018. *R: a Language and Environment for Statistical Computing*.
- Rose, H., Rinaldi, L., Bosco, A., Mavrot, F., De Waal, T., Skuce, P., Charlier, J., Torgerson, P.R., Hertzberg, H., Hendrickx, G., Vercruysse, J., Morgan, E.R., 2015. Widespread anthelmintic resistance in European farmed ruminants: a systematic review. *Vet. Rec.* <https://doi.org/10.1136/vr.102982>.
- 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, 299–303. <https://doi.org/10.1016/j.ijpara.2014.01.005>.
- 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, 397–401. <https://doi.org/10.1016/j.vetpar.2012.03.043>.
- Vidyashankar, A.N., Hanlon, B.M., Kaplan, R.M., 2012. Statistical and biological considerations in evaluating drug efficacy in equine strongyle parasites using fecal egg count data. *Vet. Parasitol.* 185, 45–56. <https://doi.org/10.1016/j.vetpar.2011.10.011>.
- Wang, C., Furrer, R., 2018. eggCounts: a Bayesian Hierarchical Toolkit to Model Faecal Egg Count Reductions. *ArXiv:1804.11224 [stat.CO]*.
- Wang, C., Paul, M., 2018. eggCounts: hierarchical modelling of faecal egg counts. R package version 2.0. <http://cran.r-project.org/package=eggCounts>.
- Wang, C., Torgerson, P.R., Höglund, J., Furrer, R., 2017. Zero-inflated hierarchical models for faecal egg counts to assess anthelmintic efficacy. *Vet. Parasitol.* 235. <https://doi.org/10.1016/j.vetpar.2016.12.007>.