Statistical and Biological Considerations in Evaluating Drug Efficacy in Equine Strongyle Parasites using Fecal Egg Count Data

Vidyashankar, A.N.[a]*, Hanlon, B.M.[b], and Kaplan, R.M.[c]

[a] Department of Statistics, George Mason University, Fairfax, VA, USA

[b] Department of Statistics, University of Wisconsin, Madison, WI, USA

[c] Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, Georgia, USA

*corresponding author

A.N. Vidyashankar, PhD

Department of Statistics, Volgenau School of Engineering

George Mason University

4400 University Drive, MSC 4A7

Fairfax, VA 22030

Email: avidyash@gmu.edu

Fax: (703) 993-1700

Abstract

Anthelmintic resistance (AR) is a serious problem for the control of equine gastrointestinal nematodes, particularly in the cyathostomins. The fecal egg count reduction test (FECRT) is the most common method for diagnosing AR and serves as the practical gold standard. However, accurate quantification of resistance and especially accurate diagnosis of emerging resistance to avermectin/milbemycin (A/M) drugs, is hampered by a lack of accepted standards for study design, data analysis, and data interpretation. In order to develop rational evidence-based standards for diagnosis of resistance, one must first take into account the numerous sources of variability, both biological and technical, that affect the measurement of fecal egg counts (FEC). Though usually ignored, these issues can greatly impact the observed efficacy. Thus, to accurately diagnose resistance on the basis of FECRT data, it is important to reduce levels of variability through improved study design, and then deal with inherent variability that cannot be removed, by performing thorough and proper statistical analysis. In this paper we discuss these issues in detail, and provide an explanation of the statistical models and methods that are most appropriate for analyzing these types of data. We also provide several examples using data from laboratory, field, and simulation experiments illustrating the benefits of these approaches.

Keywords: Equine parasites, anthelmintic resistance, fecal egg count reduction test, statistical models, bootstrap, Bayesian analysis

1. Introduction

Anthelmintic resistance is defined as a heritable genetic change in a population of parasites that enables a significantly greater proportion of individual parasites to survive drug treatments that previously were effective. This definition assumes the same species and developmental stage are being targeted at the same dose level as previously. At its biological core, anthelmintic resistance is a dynamic genetic process that occurs slowly over time as alleles conferring a resistant phenotype slowly increase in frequency with repeated cycles of drug selection. The rate at which resistance alleles increase in a given population of parasites to a given drug is affected by a complex interaction of many factors. These relate to the mode of inheritance and the number of genes involved, the parasite biology and epidemiology, the dynamics of the host-parasite relationship, the frequency and timing of treatments and the pharmacokinetics of the drugs (Churcher et. al., 2010).

Since resistance tends to evolve slowly over many years, the problem remains clinically unapparent until its later stages, when allele frequencies begin to reach high levels (Sangster, 1999). With parasites that have high pathogenic potential, resistance may manifest itself as a therapeutic failure with disease symptoms persisting after treatment. However, with parasites of relatively low pathogenic potential (e.g. cyathostomins of horses), detection of resistance is made only if specific testing is performed. Currently, there are no molecular assays that are useful for resistance diagnosis, and no *in vitro* bioassays for detecting resistance have been fully validated in equine parasites. Thus, direct *in vivo* measurements that evaluate the efficacy of drugs remain the only available approach in horses (Kaplan, 2002, 2009).

Before one can understand how to diagnose resistance, one must first understand how efficacy is determined and the factors that affect the observed efficacy. Unless efficacy is 100% in all animals, there will be variability in the measurement both within and between animals.

66 Consequently, every time a test for efficacy is performed the result will be different, and the
67 magnitude of the difference will depend on the amount of variability in the response to treatment.
68 Thus, the observed efficacy of a drug in any efficacy trial is not a fixed number, but instead is an
69 unknown value within a set of possible values. This set of values can be described using a
70 probability distribution whose parameters have both biological and statistical meaning.
71 The theoretical gold standard for assessing efficacy to anthelmintics is either counting the
72 total number of dead and live worms in each animal following treatment (critical test) or by
73 comparing the number of worms recovered from treated and untreated groups of animals
74 (controlled efficacy test). However, these types of trials are labor and resource intensive, which
75 limits the number of animals that can be tested. Additionally, data can be obtained only by
76 sacrificing the animals; hence, they are restricted to laboratory-based research and cannot be
77 used when studying resistance at the farm level. As an alternative, one can use surrogate
78 measurements (obtained from the live animal), such as the number of worm eggs measured in
79 feces before and after treatment. This procedure, referred to as the fecal egg count reduction test
80 (FECRT), is the most common means for determining the efficacy of the anthelmintics on horse
81 farms and serves as the practical gold standard (Kaplan, 2002).
82 When performing the FECRT on horse farms, fecal egg counts (FEC) typically are
83 compared in the same animals before treatment and at some established time-point after
84 treatment (typically 10-14 days). The most common approach to assess efficacy using FEC data
85 is to examine the arithmetic sample mean for percentage reduction. It has been suggested that
86 logarithmic or arcsine transformations be performed before calculating the mean (Dargatz et al.,
87 2000; Pook et al., 2002). Inferences regarding the presence or absence of resistance are then
88 made based on cutoff values for percent reduction in FEC; however, the cutoff values used vary

widely among published studies (Kaplan, 2002). In some studies the WAAVP standard cutoff for sheep was used (< 95% reduction, LCL < 90%) (Craven et al., 1998; Ihler, 1995) while others studies used a < 90% reduction in FEC (Reuber et al., 2000; Varady et al., 2000) as the cutoff for resistance. Still other investigators have used a more conservative cutoff of <80% for resistance, with results between 80% and 90% declared equivocal with resistance suspected (Kaplan et al., 2004; Repeta et al., 1993; Tarigo-Martinie et al., 2001; Woods et al., 1998). In all of these studies, no matter what cutoff value was used, it was the same for all drugs tested.

Such approaches are based on the observed efficacy of the drug rather than on the true efficacy of the drug at the time of treatment, which, as mentioned above, is always unknown. Furthermore, there are many sources of variability in FEC data that can impact the interpretation of results (Table 1). Consequently, making an accurate inference regarding resistance is complicated, even though the FECRT seems simple and straightforward.

In this paper we will review the major biological and statistical issues that must be accounted for in order to achieve the greatest possible accuracy in making resistance classifications when performing the FECRT. We will define efficacy and resistance and discuss the impact of variability on the interpretation of FECRT results. We also review various statistical models and methods that can be used to analyze FECRT data. These models help address the challenges of high variability, multi-modality, skewness, and zero-inflation inherent in FECRT data.

2. Resistance and Efficacy

111 Defining resistance solely on the basis of FEC reduction following treatment without an
112 independent means for verification, such as a calibrated and validated in vitro or molecular test,
113 is difficult. As explained above, resistance generally evolves slowly over time and when the
114 allele frequencies of relevant genes reach certain threshold levels a phenotype of reduced
115 efficacy appears relatively suddenly. Hence, diagnosing resistance in the early stages is a
116 challenging task.

117 Inferring the presence or absence of resistance depends on the measurement for efficacy
118 and the criteria used to evaluate the measurement. Thus, in order to develop a working definition
119 of resistance, we must first define efficacy and understand the factors affecting the measurement
120 of efficacy. Efficacy can be defined as a quantitative measure of the effectiveness of a drug
121 intended to produce a desired effect. With regard to anthelmintics, the expected or true efficacy
122 can be defined as the efficacy level of the drug when it was first introduced. Note that this value
123 is always less than 100%, is different for each drug, and also is likely to vary among the various
124 species within a host (the parasite infrapopulation) and among hosts, as described below.

125 In studies where the cyathostomins are of interest, it is common to use each horse as its
126 own control and collect pre-treatment and post-treatment fecal samples over a two-week period.
127 A biological justification for this study design is that cyathostomins undergo a long larval
128 development period yielding long life cycles. This causes the egg count levels to remain
129 relatively consistent over short periods (Gomez: and Georgi, 1991; Dopfer et al., 2004; Nielsen
130 et al., 2006). Also, due to complex host-parasite interactions, short-term changes in transmission
131 levels caused by weather and related factors are unlikely to greatly affect egg count levels over a
132 two-week period. A practical justification for this choice of the study design is that most horse
133 farms have few horses and it is difficult, if not impossible, to carry out controlled experiments

134 with multiple horses per treatment group. Also, since groups should be balanced by pre-
135 treatment FEC, the untreated control group would have to include some of the most highly
136 parasitized horses. It is difficult to convince horse owners to leave horses with high FEC
137 untreated, even if only for a few weeks.

138 Using the study design in which each horse serves as its own control, the observed
139 efficacy of the drug in a given horse on a given farm is defined as the relative change in the egg
140 count following treatment. That is,

$$eff = \frac{pre - post}{pre}$$

141 where pre represents the pre-treatment egg count from a horse on a specific farm and $post$
142 represents the post-treatment egg count from the same horse. The mean reduction in FEC can
143 then be calculated from the results of all horses tested that met the inclusion criteria. This mean
144 reduction then serves as the observed (or measured) efficacy for that farm. An observed efficacy
145 is often interpreted as a fixed value; however, due to variability from multiple sources (see Table
146 1) this value will change every time it is measured.

147 We consider experimental data that illustrate the variability in FEC measurements
148 (Figure 1). The data in this figure represent one hundred and ten separate FEC, on four different
149 horses. For each horse, five FEC were performed on each sample, which were collected
150 approximately every twelve hours over eleven days. These measurements are all pre-treatment.
151 Thus, from Figure 1 it is clear that the observed FEC values are better described by a probability
152 distribution, as opposed to a fixed number. It is critical to understand the difference between
153 genuine reduction in efficacy and changes in efficacy caused due to variability. This leads to
154 describing the sources of variability.
155

3. Sources of Variability in Fecal Egg Counts

There are many sources of animal-related and farm-related variability in FEC data that can impact the interpretation of results concerning efficacy. This issue is amplified especially when performing multiple farm studies. Some of the important sources of variability and their consequences are presented in Table 1. Quality fecal samples, appropriate storage conditions, and sound experimental practices will typically lead to reduction in technical variability. However, this alone will not eliminate all sources of inherent variability in the number of eggs counted in the fecal samples. No matter how well the FEC are performed on a technical level, factors such as egg loss during the procedure and non-uniform distribution of eggs in the fecal solutions cannot be eliminated. To address this point, we compared manual stirring with constant mixing (using a stir bar) of the fecal solution (prior to removing the sample aliquot) and found that there was no significant difference in the FEC ($p = 0.2489$). These results suggest that non-uniform distribution of eggs in fecal solutions cannot be overcome by thorough mixing. A study evaluating the Cornell-Wisconsin centrifugal floatation method for egg counting demonstrated that only 60-69% of eggs are recovered from feces (Egwang and Slocombe, 1982). Experiments in our laboratory using the modified-Wisconsin centrifugal method also yielded similar results. Egg loss during the egg counting procedure will lead to egg count measurements that are lower than the actual levels of eggs per gram (EPG) in the feces. When performing FEC for routine clinical examinations this issue is of minor importance, as long as the technique is performed in a consistent manner each time. However, in the context of a FECRT, such egg loss has the effect of yielding increasing numbers of zeros in the post treatment measurement, making the finding of 100% efficacy increasingly likely when EPG are very low. A practical consequence is that efficacy of drugs is often overestimated.

If variability in the data is too high, it is frequently impossible to statistically identify a small reduction in efficacy. When attempting to diagnose resistance, particularly in the early stages, such small changes are critical. As described above, there are several sources that cause variability. One approach to reducing variability is to increase the number of horses used for evaluation. While this is perhaps the best approach, for practical reasons it is often not possible to increase the number of animals involved in the study. In such cases, it is important to be able to reduce variability from other sources.

Barring drug treatment soon before the egg count is performed, FECs of any given mature horse are fairly consistent over time (Nielsen et al., 2006), but the observed FEC does fluctuate up and down for a variety of biological as well as technical (procedural) reasons. The impact of this fluctuation is further exacerbated by the fact that egg counts typically are performed only once for each animal. It is a basic statistical principle that replicating a measurement and averaging will decrease variability. Indeed, increasing the number of egg counts $k-$ fold and averaging will decrease the variability due to egg counts by a factor of k. This fact is considered when measuring practically all biological parameters (e.g. when performing biochemical assays, immunological assays, in vitro drug resistance assays or quantitative PCR), even though these measurements likely have far less variability than FEC data. Replicated measurements are not routinely performed with FEC in livestock; however, in human parasitology, the use of replicated measurements for FEC and the collection of double samples is a relatively common practice (Glinz et al., 2010; Knopp et al., 2008; Knopp et al., 2011). Therefore, the recommendation that FEC be done in triplicate when performing a FECRT as part of a scientific investigation should be considered as bringing this measurement into the same realm as other biological measurements, not as something novel and burdensome. Practical

202 considerations and cost relative to benefit will have to be taken into account to determine the
203 number of egg counts for a particular study, or when performing a single farm clinical
204 investigation.

205 Furthermore, since post-treatment FEC values are usually small, using a method with a
206 lower detection limit of < 5 EPG, and repeating egg counts to improve the accuracy of the count
207 also have very important implications when studying efficacy. Of course, if post-treatment FEC
208 are high all the discussions are moot. Since egg loss seems to be a consistent feature of FEC
209 methods (approximately 30% of eggs are lost), with a single egg count the chances of getting
210 zero eggs will be enhanced if FEC are very low. The difference between 100% and less than
211 100% FECR can be important---particularly with ivermectin and moxidectin. Seeing or not
212 seeing 1 or 2 eggs can make this difference. By repeating FEC three times, a zero result is more
213 likely to be a true zero.

214 To reiterate the above issue concerning repeated measurements for FEC, we performed a
215 biological experiment and a simulation experiment to demonstrate the gain in precision. In the
216 biological experiment, approximately 200 grams (g) of feces were collected from a horse, mixed
217 well, and then six 10 g subsamples were used for the analysis. Ten separate FEC were then
218 performed on each of the 10 g subsamples, for a total of 60 FEC. To see what would happen to
219 the variability of the measured FEC if we used the results of more than one measurement, we
220 combined two or three consecutive FEC measurements and then calculated the mean. The results
221 from this experiment are displayed in Figure 2. The lines in Figure 2 represent 60 individual
222 counts, 30 counts obtained by averaging any two consecutive FEC, and 20 counts obtained by
223 averaging any three consecutive FEC. The results from the simulation experiments, which are
224 based on 5000 simulated data sets from a negative binomial distribution, are displayed in Table

2. The results in Table 2a are based on mean =25 and variance =125, while those in Table 2b are based on mean 250 and variance 5000. The choices of the parameters were based on the results of experiments performed in our laboratory. The simulation evidence displays clear gains in precision when averaging over multiple FEC compared to using a single FEC. For example, consider Table 2b, where the true mean is 250. If FEC are repeated three times, then the observed mean egg count will approximate the true mean almost 80% of the times as compared to about 50% of the time for a single count.

4. Scientific Questions and Statistical hypotheses

Scientific questions concerning efficacy and resistance can be formulated in terms of testing statistical hypotheses and confidence intervals. As a specific example, suppose a drug is known to be efficacious with an average efficacy of 99% in non-resistant parasites and one wants to investigate if there is reduction in efficacy. As explained above, the observed efficacy can be any number as suggested by its distribution. If the observed efficacy is greater than 99%, it is suggestive of no onset of reduction. However, if the observed efficacy is less than 99%, it is important to understand the tradeoff between the reduction in efficacy and variability. If the efficacy is much smaller than what can be attributed to variability, then it is reasonable to conclude a reduction in efficacy.

To describe the tradeoff between efficacy and variability in the context of a hypothesis testing problem, Type I error plays an important role. The probability of Type I error of a statistical test describes the probability of incorrectly rejecting the null-hypothesis and concluding a reduction in efficacy. This probability is pre-set by the scientist. As one would expect, the smaller the probability, the less frequently one would reject the null hypothesis. It is

common practice to set the probability of Type I error at 5%. The conclusion based on a statistical test of efficacy can be referred to as statistical identification of reduction in efficacy.

Confidence intervals can also be used to evaluate a decrease in efficacy. The lower-limit of a 95% confidence interval can be used to identify a reduction in efficacy as follows: using data one constructs a confidence interval for the efficacy parameter. If the lower-limit is less than the pre-defined lower-confidence threshold, then one can surmise that there is statistical evidence of reduction in efficacy. One could also identify a reduction in efficacy based on the upper limit of the confidence interval. However, this will not work in highly efficacious drugs, like ivermectin, since in most cases the upper-limit will be one (100%).

The methods described above are examples of frequentist methods for evaluating reduction in efficacy. Alternatively, one can also use Bayesian methods. In Box 1, we describe an example to show how proper analysis using frequentist and Bayesian methods yield similar conclusions. We emphasize here that both methods of analysis, if performed correctly, should yield similar conclusions.

5. Statistical Models

In this section we present statistical models for FEC data. Here we distinguish between single farm data and multiple farm data. We begin by considering the single farm case.

Let n denote the number of horses in a given farm. Let $(X_1 \ Y_1)$, ... $(X_n \ Y_n)$ denote the pre-treatment and post-treatment egg counts for the horses. For instance, $(X_i \ Y_i)$ denotes the pre-treatment and post-treatment egg counts for the i^{th} horse. Then, using the definition of

efficacy described in Section 2, the observed efficacy of the i^{th} horse, denoted by eff_i is given by

$$eff_i = \frac{X_i - Y_i}{X_i}$$

5.1 Distribution of Pre-treatment Egg Counts

The distribution of pre-treatment egg counts X, which we often refer to as the *pre-treatment distribution*, is unknown. We denote this unknown distribution by . Based on multiple studies and large data sets, parasitologists have suggested the use of the negative binomial distribution for G (Hunter and Quenouille, 1952; Crofton, 1971; Grenfell et al., 1995; Wilson and Grenfell, 1997). There are other distributions that can be used for modeling the pre-treatment egg counts. For instance, zero-inflation is an important consideration in modeling FEC data. Therefore, zero-inflated negative binomial models can also be used for modeling pre-treatment data (Walker et al., 2009; Hilbe, 2011).

The choice of which pre-treatment distribution to use in the analysis is complicated by the small sample size n and over dispersion. Over dispersion in a horse farm is typically caused due to most horses having low egg counts while few have high egg counts. Statistical distributions that account for over dispersion have the property that the variance in egg counts is larger than the mean egg count. Even though negative binomial distribution is a reasonable assumption for the pre-treatment distribution when considering large population of animals, when the sample size is small, as is common on horse farms, this assumption may be difficult to verify. A typical example is described in Table 3. This variability between the pre-treatment counts of different horses leads to complications in modeling the pre-treatment distribution. Typical analysis of these data (e.g. as performed in standard statistical software like R and SAS)

will depend on the choice of pre-treatment distribution. However, we advocate using a bootstrap methodology proposed in (Vidyashankar et al., 2007) that does not make any assumptions about the specific form of the pre-treatment distribution.

5.2 Post-treatment Egg Counts

To describe the distribution of the post-treatment egg counts, we assume that the drug is efficacious at level p %. Hence, conditioned on the pre-treatment counts, we model the post-treatment counts as binomial, specifically

$$Y_i \mid X_i \sim Bin(X_i, 1-p).$$

Combining this assumption with the pre-treatment distribution we have the following model

$$\begin{aligned} Y_i \mid X_i &\sim Bin(X_i, 1-p), \\ X_i &\sim G\,. \end{aligned} \qquad (1)$$

We postulate that the data on a single farm contains $n$samples from the model (1). We will refer to this as the fixed p model. Table 4 presents such samples at three simulated egg counts for horses in Table 3 with $p = .95$.

5.3 A Random Effect Model

As mentioned above, the efficacy of a drug in a given horse ultimately depends on the interaction of the parasite with the drug and this is greatly affected by the pharmacokinetics and pharmacodynamics. The following horse features interact to influence the pharmacokinetics and pharmacodynamics of the drug: gastrointestinal motility, size, level of body fat, general immune state, general physiological state, and diet. These factors and others are likely responsible for the observation that anthelmintics are less effective in young (yearling) horses as compared to adult

horses (Herd and Gabel, 1990). Furthermore, there are more than fifty different species of cyathostomins and the relative sensitivity of any one of them to the drug is unknown. Also, which of these fifty species is populating the gut of the horse is also unknown. Thus, the assumption that the efficacy p is the same for all horses is biologically infeasible. For this reason, we model the variation statistically by assuming that the efficacy levels of the drug on each horse are random draws from some distribution H. This yields the following statistical model for the egg count data from a single farm

$$\begin{aligned} Y_i \mid X_i &\sim Bin(X_i, 1-p_i) \\ X_i &\sim G \qquad (2) \\ p_i &\sim H \end{aligned}$$

In the statistical literature, model (2) is referred to as a random effect model, a hierarchical model, or a multilevel model. These models induce correlations in FEC from horses within the same farm.

5.4 Examples

We now present specific examples illustrating various statistical models. If the efficacies p_i follow a beta distribution $Beta(\alpha, \beta)$ then the resulting model is a *beta-binomial* model. The parameters α and β determine the shape of the distribution; the ratio $\frac{\alpha}{\alpha+\beta}$ represents the expected efficacy of the drug. Figure 3 provides densities of the beta distribution for various values of the parameters α and β. Table 4 contains post-treatment counts from this model for horses from Table 3. The results in Table 4 demonstrate that the post FEC values for fixed p case are more similar across simulations than those for the random p case. Now, understanding that each simulation represents the particular instance of performing FEC, it is clear from Table 4 and

Figure 3 that a random effect model with flexible random effect distribution is ideal for modeling the post-treatment egg counts.

Another common model is the *logit-normal* model. Normal distribution cannot be used to model the distribution of the efficacies p_i because probabilities are restricted to the interval (0,1). A standard solution is to use a logit transform, $\log\text{it}(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$ which lies in the interval $(-\infty, \infty)$. In the logit-normal model, it is assumed that $\log\text{it}(p_i)$ follows a normal distribution.

With random effect models, each horse has a different efficacy, therefore it does not make sense to talk about a fixed value $p$at the farm level. An important concept is *model based efficacy*. Model based efficacy is defined as the average theoretical efficacy in the population of horses in that farm, as described by the statistical model. It is denoted $E(p)$ which represents the expected value or mean of the efficacy distribution. As explained above, for the beta-binomial model,

$$E(p) = \frac{\alpha}{\alpha + \beta}.$$

Such a closed form solution does not exist for the logit-normal model. See (Vidyashankar et. al., 2007) and (McCulloch et al., 2008) for more detailed descriptions of these models and related concepts.

5.5 Multiple Farm Data

To study the prevalence of resistance, one has to deal with multiple farms. In this case, it is necessary to take into account between farm variability along with other sources for each farm. We now describe the statistical model. Let (X_{ij}, Y_{ij})denote the pre-treatment and post-treatment egg counts of the i^{th} horse on the j^{th} farm. We assume that for each farm j, X_{ij} are random

draws from G_j where G_j is the pre-treatment distribution for the j^{th} farm. The post-treatment egg counts are modeled as before, but here the efficacy levels change between horses within a farm and between farms; that is,

$$Y_{ij} | X_{ij}, p_{ij} \sim Bin\,(X_{ij}, 1 - p_{ij})$$
$$X_{ij} \sim G_j \qquad (3)$$
$$p_{ij} \sim H_j.$$

Importance of between farm variability and its impact on the interpretation of results is described in Box 2.

6. Data Analysis Methods

Several statistical methods are available to evaluate reduction in efficacy using FECRT data. The methods depend on the assumptions made on the distributions of the egg counts (pre and post) and the efficacies. Statistical methods for analyzing FEC data can be classified into frequentist methods and Bayesian methods. These methods include both parametric and non-parametric methods. Additionally, both likelihood and bootstrap based methods are commonly used in the frequentist setting. More details are provided in Boxes 3, 4, and 5.

The most common approach to infer reduction in efficacy is based on the sample (arithmetic) mean for percentage reduction (this is a frequentist method); some studies have advocated the use of logarithmic or arcsine transformations before calculating the arithmetic mean (Dargatz et al., 2000; Pook et al., 2002) so that the resulting sample means have an approximate normal distribution. However, (Vidyashankar et. al., 2007) demonstrate that even after these transformations, the resulting inference concerning reduction in efficacy can be incorrect, especially when the sample size is small and heterogeneities are present. This situation

375 is frequently encountered when the number of horses within a farm is small or the farm contains
376 several small groups of horses, where the horses are homogeneous within a group but highly
377 variable between groups. Non-parametric bootstrap methods are useful since they do not require
378 data transformations and are also easy to implement. Furthermore, they do not require detailed
379 modeling assumptions concerning egg count distributions and efficacies.

380 A key idea behind the bootstrap methods for hypothesis testing is that one uses random
381 number generators to produce multiple new data sets, which are similar to field data assuming
382 the postulated value for efficacy. For example, if the null-hypothesis states that the true efficacy
383 is 90%, we use the observed pre-treatment egg count to simulate new data with 90% efficacy.
384 Efficacies are then calculated for each of these newly simulated data sets. It should be noted that
385 even though the average of efficacies from these new data sets will be close to 90% due to
386 variability, values smaller than 90% and larger than 90% are likely to occur. Using these
387 simulated efficacies, we obtain what is called the bootstrap distribution of efficacy. Exact
388 algorithms for calculating this distribution are available in the literature (Vidyashankar et. al.,
389 2007)

390 The logic behind the bootstrap hypothesis test is then to calculate the probability that the
391 efficacy obtained from the field data is at least as large as that obtained from the bootstrap
392 distribution. If this probability is larger than 0.05, we conclude that the field data does not exhibit
393 a reduction in efficacy, while if it is less than 0.05 we conclude that the data exhibits a reduction
394 in efficacy.

395 In Bayesian methods, we assume the Binomial model for the post-treatment egg count
396 and use a prior distribution to take into account variability in the efficacies. Then one uses the
397 mean of the posterior distribution---i.e., the distribution of efficacies given the data---to identify

reduction in efficacy. For a more detailed description of Bayesian methods see Basanez et al. (2004) and Hanlon et al. (2009). We illustrate the ideas using two examples.

We consider a data set with six horses with observed efficacy values given in Table 5. This single farm data is part of a larger study of Danish horse farms, where the horses are treated with pyrantel (Nielsen et al., 2010). The bootstrap and Bayesian analysis yield similar results for this data set. Using Method 1 of (Vidyashankar et. al., 2007), a 95% bootstrap confidence interval for the farm efficacy is (0.9584, 0.9759); a Bayesian analysis using an improper $Beta(0,0)$ prior yields a 95% credible region (0.9612, 0.9737). Because pyrantel has a nominal efficacy of 90% and because these intervals' lower limits are above 0.9, both the Bayesian and bootstrap analysis suggest that there is no evidence of resistance. We emphasize that this example is performed to illustrate these methods in a relatively simple situation. One needs to be careful in making resistance classifications based on a single farm analysis, especially when the sample size is only n=6. As discussed in Nielsen et al. (2010) a more careful analysis would take into account variability between farms (also see Vidyashankar, et. al., 2007) and the covariates of different horses, such as age, gender, pre-treatment infection level. For illustrative purposes we also display the bootstrap distribution of efficacy values and the Bayesian posterior distribution for the efficacy in Figure 4. We provide an analysis comparing bootstrap methods and Bayesian methods for multiple farms in Box 1.

7. Conclusion

The FECRT is the practical gold standard to detect resistance on horse farms. However, there are numerous sources of variability that complicate interpretation of data. Thus, proper statistical analysis is necessary to accurately diagnose resistance. We have reviewed many of the

biological and statistical issues and challenges that must be addressed. To summarize, the following key issues should be taken into account when performing FECRT and subsequent analysis of data. It is important to have accurate diagnosis of resistance when the prevalence is rare. Hence, it is important to distinguish true reduction in efficacy from variability. The key to achieving this is to use methods that reduce the amount of inherent variability. One of the ways this can be achieved is by increasing the number of horses tested and increasing the number of egg counts performed on each sample. Finally, to analyze the data, one should carefully model the data and use appropriate statistical methods, understanding both the assumptions and limitations of these methods.

Conflict of interest statement:

The authors have no financial or personal relationship with other people or organisations that could inappropriately influence or bias the paper entitled "Statistical and biological considerations in evaluating drug efficacy in equine strongyle parasites using fecal egg count data."

References

Basanez, M.-G., Marshall, C., Carabin, H., Gyorkos, T., Joseph, L., 2004. Bayesian statistics for parasitologists. Trends Parasitol. 20, 85-91.

Churcher, T.S., Kaplan, R.M., Ardelli, B.F., Schwenkenbecher, J.M., Basáñez, M.G., Lammie, P.L., 2010. Mass treatment of parasitic disease: implications for the development and spread of anthelmintic resistance, In: Webber, J.T. (Ed.) Antimicrobial Resistance - Beyond the Breakpoint, Issues in Infectious Diseases. Karger, Basel, pp. 120-137.

445 Craven, J., Bjorn, H., Henriksen, S.A., Nansen, P., Larsen, M., Lendal, S., 1998. Survey of
446 anthelmintic resistance on Danish horse farms, using 5 different methods of calculating
447 faecal egg count reduction. Equine Vet. J. 30, 289-293.

448 Crofton, H.D., 1971. A quantitative approach to parasitism. Parasitol. 62, 179-193.

449 Dargatz, D.A., Traub-Dargatz, J.L., Sangster, N.C., 2000. Antimicrobic and anthelmintic
450 resistance. Vet. Clin. North Am. Equine Pract. 16, 515-536.

451 Dopfer, D., Kerssens, C.M., Meijer, Y.G.M., Boersema, J.H., Eysker, M., 2004. Shedding
452 consistency of strongyle-type eggs in Dutch boarding horses. Vet. Parasitol. 124, 249-
453 258.

454 Efron, B., Tibshirani, R., 1993. An introduction to bootstrap, Vol 57. Chapman & Hall/CRC.

455 Egwang, T.G., Slocombe, J.O.D., 1982. Evaluation of the Cornell-Wisconsin centrifugal
456 flotation technique for recovering trichostrongylid eggs from bovine feces. Can. J. Comp.
457 Med. 46, 133-137.

458 Glinz, D., Silue, K.D., Knopp, S., Lohourignon, L.K., Yao, K.P., Steinmann, P., Rinaldi, L.,
459 Cringoli, G., N'Goran, E.K., Utzinger, J., 2010. Comparing diagnostic accuracy of Kato-
460 Katz, Koga agar plate, ether-concentration, and FLOTAC for Schistosoma mansoni and
461 soil-transmitted helminths. PLoS Negl. Trop. Dis. 4, e754.

462 Gomez, H., Georgi, J., 1991. Equine helminth infections: control by selective chemotherapy
463 [published erratum appears in Equine Vet. J. 1991 Jul;23(4):252] [see comments]. Equine
464 Vet. J. 23, 198-200.

465 Grenfell, B.T., Wilson, K., Isham, V.S., Boyd, H.E.G., Dietz, K., 1995. Modelling patterns of
466 parasite aggregation in natural populations: Trichostrongylid nematode-ruminant
467 interactions as a case study. Parasitol. 111, S135-S151.

468 Hanlon, B., Vidyashankar, A.N., Petersen, S., Kaplan, R.M., Nielsen, M.K., 2009. A Bayesian
469 approach for evaluating drug efficacy using fecal egg count data. In: 24th International
470 Workshop on Statistical Modelling, Ithaca, NY USA.

471 Herd, R.P., Gabel, A.A., 1990. Reduced efficacy of anthelmintics in young compared with adult
472 horses. Equine Vet. J. 22, 164-169.

473 Hilbe, J.M., 2011. Nagative binomial regression, 2nd Edition. Cambridge University Press, New
474 York.

475 Hunter, G.C., Quenouille, M.H., 1952. A statistical examination of the worm egg count sampling
476 technique for sheep. J. Helminthol. 26, 157-170.

Ihler, C.F., 1995. A field survey on anthelmintic resistance in equine small strongyles in Norway. Acta Vet. Scand. 36, 135-143.

Kaplan, R.M., 2002. Anthelmintic resistance in nematodes of horses. Vet. Res. 33, 491-507.

Kaplan, R.M., 2004. Drug resistance in nematodes of veterinary importance: a status report. Trends Parasitol. 20, 477-481.

Kaplan, R.M. 2009. In vitro diagnosis of anthelmintic resistance in equine parasites: do we really have any useful assays? In World Association for the Advancement of Veterinary Parasitology, 22nd International Conference, Boeckh, A., ed. (Calgary, Canada, Fort Dodge Animal Health), pp. 1-5.

Kaplan, R.M., Klei, T.R., Lyons, E.T., Lester, G.D., French, D.D., Tolliver, S.C., Courtney, C.H., Vidyanshankar, A.N., Zhao, Y., 2004. Prevalence of anthelmintic resistant cyathostomes on horse farms. J. Am. Vet. Med. Assoc. 225, 903-910.

Knopp, S., Mgeni, A.F., Khamis, I.S., Steinmann, P., Stothard, J.R., Rollinson, D., Marti, H., Utzinger, J.r., 2008. Diagnosis of soil-transmitted helminths in the era of preventive chemotherapy: effect of multiple stool sampling and use of different diagnostic techniques. PLoS Negl. Trop. Dis. 2, e331.

Knopp, S., Speich, B., Hattendorf, J., Rinaldi, L., Mohammed, K.A., Khamis, I.S., Mohammed, A.S., Albonico, M., Rollinson, D., Marti, H., Cringoli, G., Utzinger, J.r., 2011. Diagnostic accuracy of Kato-Katz and FLOTAC for assessing anthelmintic drug efficacy. PLoS. Negl. Trop. Dis. 5, e1036.

McCulloch, C.E., Searle, S.R., Neuhaus, J.M., 2008. Generalized, linear, and mixed models, 2nd Edition. John Wiley & Sons Inc., Hoboken, NJ.

Nielsen, M.K., Haaning, N., Olsen, S.N., 2006. Strongyle egg shedding consistency in horses on farms using selective therapy in Denmark. Vet. Parasitol. 135, 333-335.

Nielsen, M.K., Vidyashankar, A.N., Hanlon, B., Petersen, S.L., Kaplan, R.K. 2010. Increased accuracy for detecting anthelmintic resistance in livestock using statistical modeling of observational data. In ICOPA XII (Melbourne Australia, August 15-20, 2010).

Pook, J.F., Power, M.L., Sangster, N.C., Hodgson, J.L., Hodgson, D.R., 2002. Evaluation of tests for anthelmintic resistance in cyathostomes. Vet. Parasitol. 106, 331-343.

Repeta, D.L., Birmbaum, N., Courtney, C.H., 1993. Anthelmintic resistance on pleasure horse farms in north central Florida. Equine Pract. 15, 8-12.

Reuber, K., Beelitz, P., Gothe, R., 2000. Anthelmintic resistance of small strongyles of horses in Upper Bavaria. Tierarztl. Umsch. 55, 216-222.

510 Sangster, N.C., 1999. Pharmacology of anthelmintic resistance in cyathostomes: will it occur
511 with the avermectin/milbemycins? Vet. Parasitol. 85, 189-204.

512 Tarigo-Martinie, J.L., Wyatt, A.R., Kaplan, R.M., 2001. Prevalence and clinical implications of
513 anthelmintic resistance in cyathostomes of horses. J. Am. Vet. Med. Assoc. 218, 1957-
514 1960.

515 Varady, M., Konigova, A., Corba, J., 2000. Benzimidazole resistance in equine cyathostomes in
516 Slovakia. Vet. Parasitol. 94, 67-74.

517 Vidyashankar, A.N., Kaplan, R.M., Chan, S., 2007. Statistical approach to measure the efficacy
518 of anthelmintic treatment on horse farms. Parasitol. 134, 2027-2039.

519 Walker, M., Hall, A., Anderson, R.M., Basanez, M.G., 2009. Density-dependent effects on the
520 weight of female *Ascaris lumbricoides* infections of humans and its impact on patterns of
521 egg production. Parasites & Vectors 2, 1-18.

522 Wilson, K., Grenfell, B.T., 1997. Generalized linear modelling for parasitologists. Parasitol.
523 Today 13, 33-38.

524 Woods, T.F., Lane, T.J., Zeng, Q.Y., Courtney, C.H., 1998. Anthelmintic resistance on horse
525 farms in north central Florida. Equine Pract. 20, 14-17.

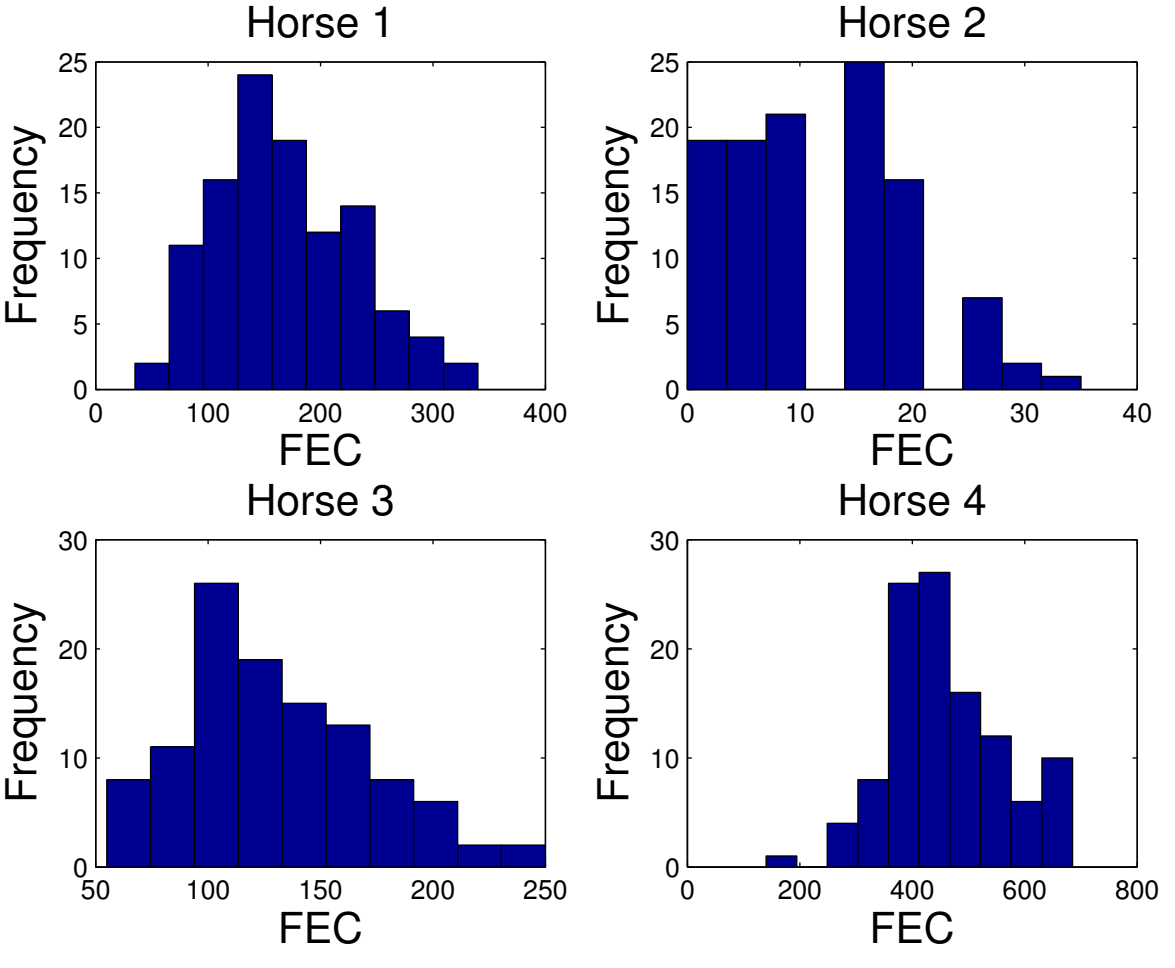
526

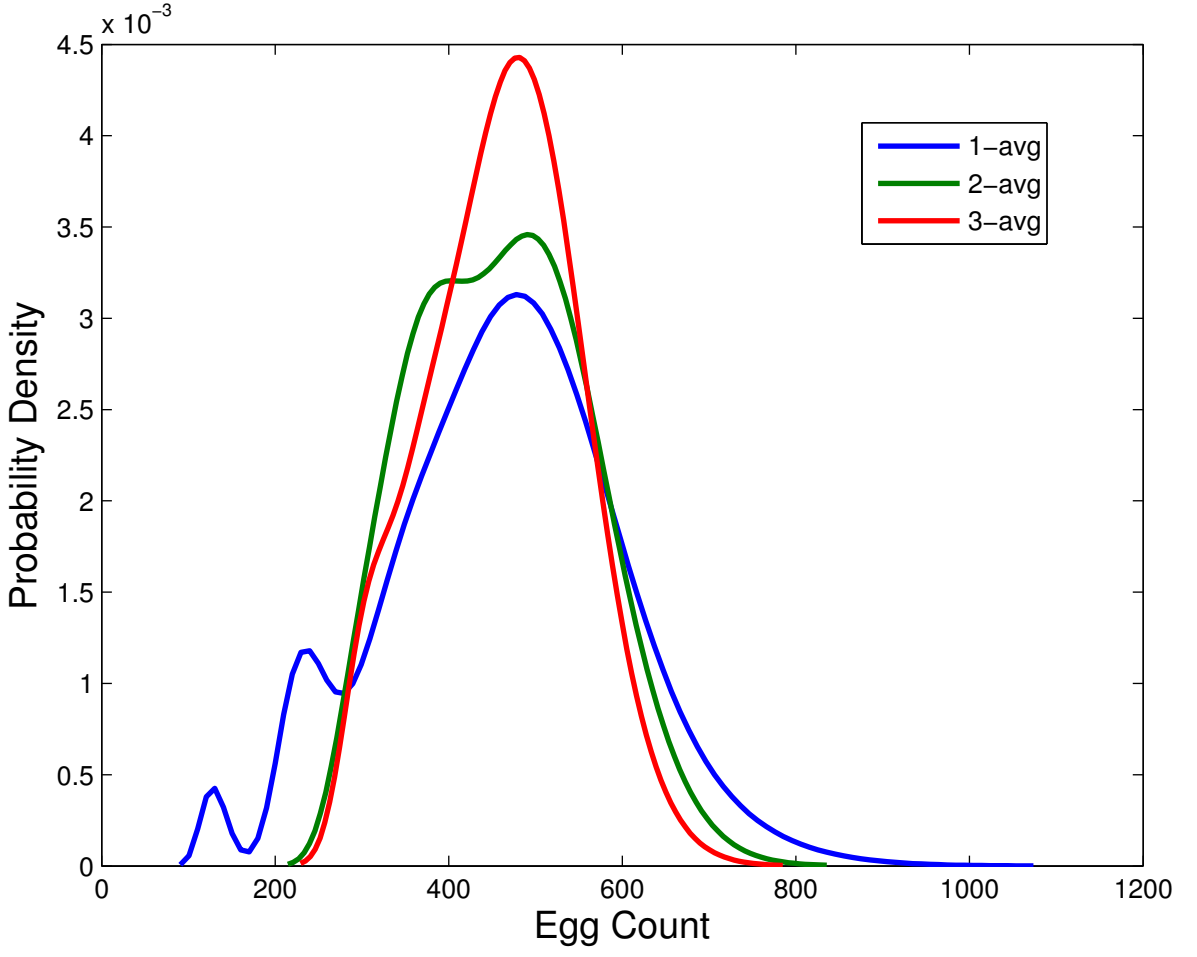
Figure 1: Variability in the egg count distributions of horses (experimental data). The data in this figure represent one hundred and ten separate fecal egg counts FEC on the same horse, for four different horses. For each horse, five FEC were performed on each sample, which were collected approximately every twelve hours over eleven days. Note the different magnitudes on the x-axis.

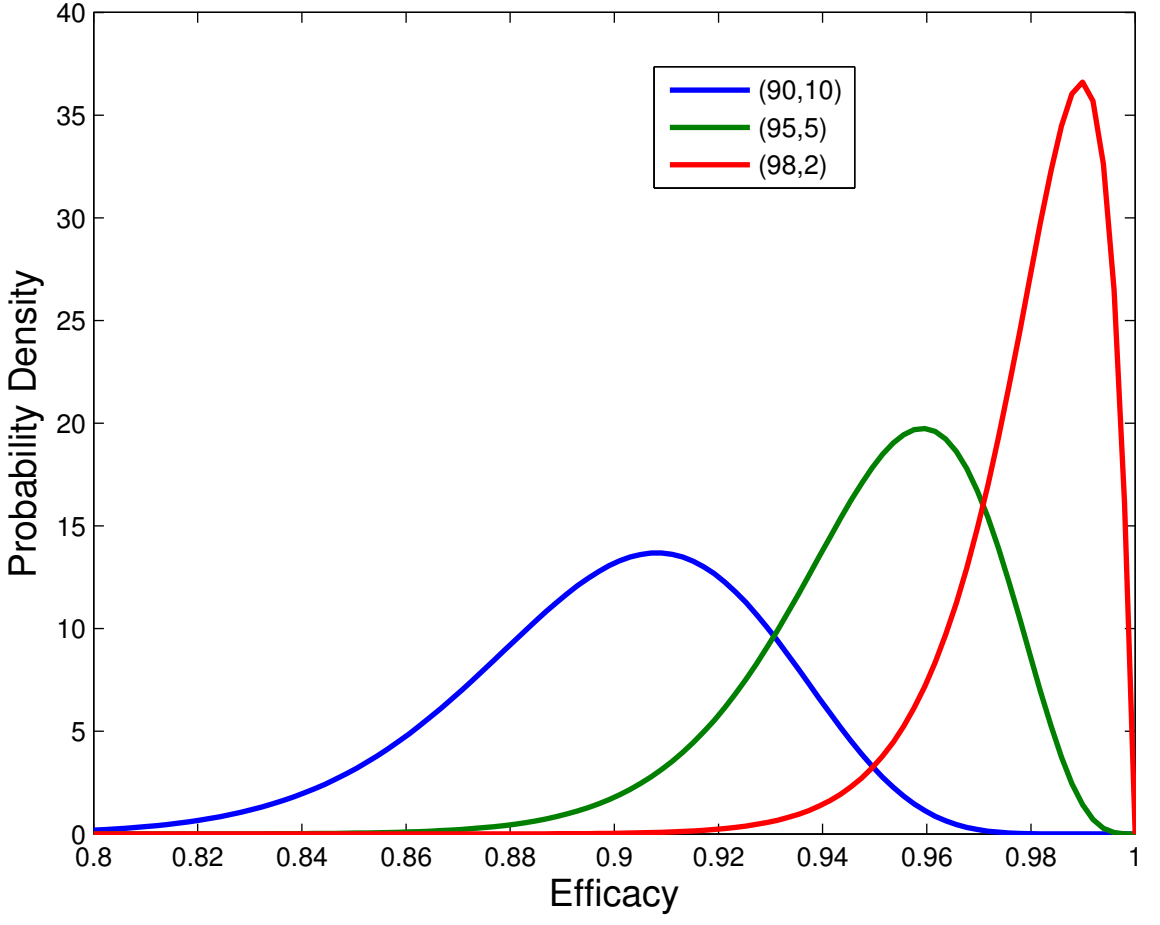
Figure 2: Improved precision with multiple egg counts (experimental data). Kernel density plots showing the effect of performing multiple fecal egg counts (FEC) on the precision of the resulting value. Ten individual FEC were performed on 6 separate 10-gram subsamples of a single fecal sample from a horse, for a total of 60 FEC performed in total. The blue line (1-avg) represents the kernel density plot of the 60 individual FEC. The green line (2-avg) represents the average of every two consecutive FEC; for a total of 30 data points. The red line (3-avg) represents the average of every three consecutive FEC; for a total of 20 data points. Note that the precision of the measurement improves by increasing the number of FEC and averaging the results.

Figure 3: Probability Density for the Beta Distribution. This figure provides densities of the beta distribution for various values of the parameters α and β.

Figure 4: Bootstrap distribution and Bayesian posterior distribution of the efficacy values (experimental data). The figure gives the Bayesian posterior distribution and a kernel density estimate of the bootstrap distribution (based on B = 2000) bootstrap samples based on the dataset in Table 5.


Horse 1
Frequency
FEC
0
5
10
15
20
25
0
100
200
300
400
Horse 2
Frequency
FEC
0
5
10
15
20
25
0
10
20
30
40
Horse 3
Frequency
FEC
0
10
20
30
50
100
150
200
250
Horse 4
Frequency
FEC
0
10
20
30
0
200
400
600
800


x 10^-3
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Probability Density
0
200
400
600
800
1000
1200
Egg Count
1–avg
2–avg
3–avg


(90,10)
(95,5)
(98,2)
Probability Density
Efficacy
0
5
10
15
20
25
30
35
40
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1

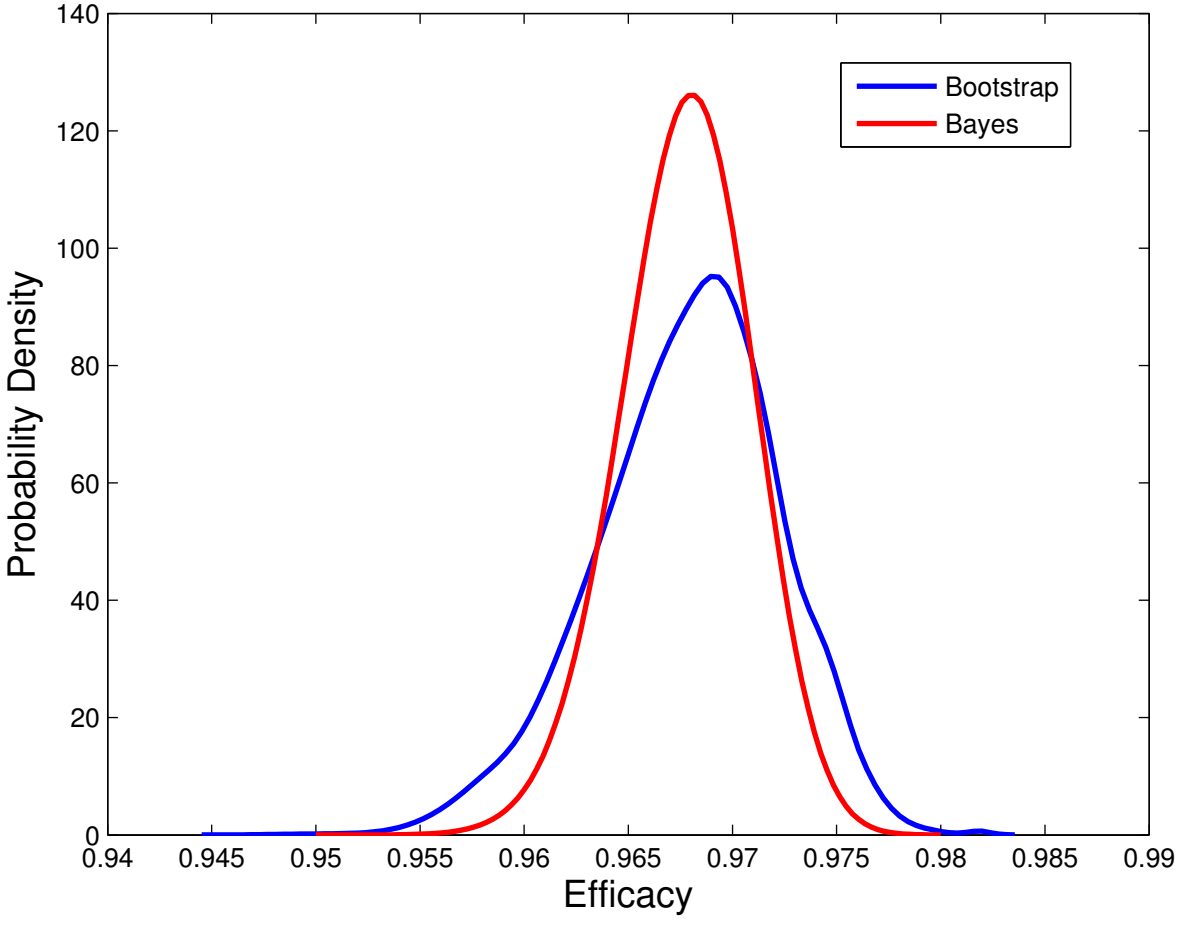

Bootstrap
Bayes
Probability Density
Efficacy
0
20
40
60
80
100
120
140
0.94
0.945
0.95
0.955
0.96
0.965
0.97
0.975
0.98
0.985
0.99

Table 1: Factors that impact the outcome of fecal egg count reduction tests in horses.

Factor	Outcome
non-Gaussian overdispersed distribution of parasites in host animals	causes large differences in pre-treatment values between animals on the same farm
low and zero pre-treatment FEC are common	reduces the numbers of animals available to test
differences in parasite infection intensities between farms	causes large differences in pre-treatment values between farms
inherent variability in parasite egg numbers within the fecal output of an animal	results in the collection of non-uniform samples
non-uniform distribution of eggs in solutions used for FEC analysis	causes variability in FEC
technical variation in performance of FEC	causes variability in FEC
overall health and body condition of animals	impacts drug pharmacokinetics and pharmacodynamics
differences in age, breed, and sex of animals both on and between farms	non-specific increase in variability
differences in management practices, diet type and nutritional programs between farms	non-specific increase in variability; variation in levels and moisture content of feces produced
spatial differences due to location of farms	non-specific increase in variability
it is difficult to ensure that full dose of the paste dewormer is swallowed	Some horses may not consume the full dose of anthelmintic administered
Different cyathostomin species have different predilection sites in the host (cecum, dorsal colon, ventral colon)	May impact drug-parasite interaction leading to unequal exposure to the drug among the species of cyathostomins present leading to species-specific variations in efficacy
multiple cyathostomin species present and relative proportions vary from farm to farm	species-specific efficacy will vary – leads to increased farm to farm variability
multiple cyathostomin species present and relative proportions vary from horse to horse on a given farm	species-specific efficacy will vary – leads to increased animal to animal variability
Density dependent effects of worm numbers on egg production, and multiple species present	Numbers of worms and relative proportion of the different species present in the host may impact the number of eggs produced per worm increasing variability
temporal differences resulting from non-uniform sampling times	non-specific increase in variability
Small numbers of horses available to test	increases the impact of all other sources of variability and increases the affect that any single horse with an ‘outlier response’ has on the resistance classification for the farm

Table 2: Precision with multiple egg counts (simulated data). The following displays evidence from a simulation experiment to demonstrate the gain in precision when using multiple egg counts. Each table presents results from a study based on 5000 simulations. The data are simulated from a negative binomial distribution with: a) mean = 25, variance = 125 and b) mean = 250, variance = 5000. The table gives the frequency for the number of times the averaged egg count falls into different intervals.

mean = 25	avg-1	avg-2	avg-3	avg-4	avg-5
[0, 10)	0.05	0.0112	0.0024	0.0008	0.0002
[10,20)	0.29580	0.2590	0.21960	0.18560	0.15480
[20, 30)	0.34540	0.4720	0.55340	0.61980	0.67220
[30, ∞)	0.30880	0.25780	0.22460	0.19380	0.17280

mean = 250	avg-1	avg-2	avg-3	avg-4	avg-5
[0, 100)	0.0038	0	0	0	0
[100,200)	0.24380	0.15420	0.10180	0.0690	0.04920
[200, 300)	0.52260	0.6840	0.7820	0.84340	0.88340
[300, ∞)	0.22980	0.16180	0.11620	0.08760	0.06740

Table 3: Fecal egg count data (experimental data). This table displays fecal egg counts (pre and post) and observed efficacy for a United States farm (Georgia) containing 13 horses that were treated with one of the following: ivermectin (Tx ID =1), fenbendazole (Tx ID =2), oxibendazole (Tx ID =3), or pyrantel (Tx ID =4) (Kaplan et al., 2004).

Horse	Tx ID	Pre count	Post count	Efficacy
1	1	3250	0	1.0000
2	1	450	0	1.0000
3	1	445	5	0.98876
4	1	300	0	1.0000
5	1	260	0	1.0000
6	2	1420	705	0.50352
7	2	505	50	0.90099
8	2	45	30	0.33333
9	3	1230	315	0.74390
10	3	325	40	0.87692
11	3	325	280	0.13846
12	3	100	0	1.0000
13	4	80	25	0.68750

Table 4: Post-treatment egg count variability simulated using a fixed p model and a beta-binomial model, pi ~ Beta(95, 5). Pre-treatment data is from 13 horses on a single farm (see Table 3). Data represents three simulated post-treatment egg counts produced using the two different models assuming a 95% efficacy (p = .95).

		Fixed *p* Model			Beta-Binomial Model		
horse	pre	post (sim 1)	post (sim 2)	post (sim 3)	post (sim 1)	post (sim 2)	post (sim 3)
1	3250	163	167	163	168	210	93
2	450	28	26	26	17	21	21
3	445	25	21	21	11	17	25
4	300	20	17	14	5	9	14
5	260	16	9	16	37	16	6
6	1420	73	65	78	117	88	42
7	505	20	20	27	18	7	19
8	45	4	4	3	1	1	3
9	1230	70	60	59	52	86	28
10	325	12	19	14	4	32	17
11	325	18	9	22	16	13	16
12	100	2	1	2	12	4	8
13	80	4	1	1	2	1	5

Table 5: Example Data used for the Bootstrap and Bayesian analysis (experimental data). This table displays fecal egg counts (pre and post) and observed efficacy for a Danish farm containing 6 horses that were treated with pyrantel. In this study only horses with a fecal egg count of 200 eggs per gram or higher were treated (Nielsen et al., 2010)

Horse	Pre count	Post count	Efficacy
1	200	0	1.0000
2	840	40	0.95238
3	300	20	0.93333
4	1120	40	0.96429
5	340	0	1.0000
6	300	0	1.0000

Box 1: Data Analysis.

We revisit the data analysis presented in Vidyashankar et al. (2007), which analyzes an experimental data set that was collected as part of a study on anthelmintic efficacy across various farms in the southeastern United States (Kaplan et al., 2004). Horses on each farm were randomly assigned to one of four anthelmintic treatments, ivermectin, fenbendazole, oxibendazole, or pyrantel. Vidyashankar et al. (2007) present a non-parametric bootstrap analysis. Here we present both a bootstrap analysis and a Bayesian analysis to illustrate the similarity of conclusions from the two methods.

For this illustration, we focus on the farms in the state of Louisiana and those horses treated with ivermectin. Nine farms were included in the Louisiana study. On all of the Louisiana farms except for two, ASHU and EHS, ivermectin demonstrated an observed efficacy of 100% in every horse treated. It is important in this context to consider the variation across the farms so this analysis was done using data from all nine farms in Louisiana. The bootstrap and Bayesian analysis yield very similar results for these data, a 95% bootstrap confidence interval is (.9981, .9996) and a 95% Bayesian credible interval is (.9991, .9994). Because ivermectin has a nominal efficacy of 99% and because these intervals' lower limits are above .99, both the Bayesian and bootstrap analysis suggest that there is no overall reduction in efficacy. This demonstrates that both the bootstrap analysis and Bayesian analysis, when properly performed, will often yield similar conclusions.

Box 2: Example of how variability can impact interpretation of results in multiple-farm studies if appropriate statistical analysis is not performed.

In Kaplan et al. (2004) a fecal egg count reduction test (FECRT) study was performed in the southern United States to determine the prevalence of anthelmintic resistance on horse farms. The study involved 786 horses on 44 different farms; 214 of the horses were treated with ivermectin. Focusing on the 214 horses treated with ivermectin, 205 demonstrated 100% reduction in fecal egg counts (FEC), 6 horses had a reduction of 95-99.9%, and 3 horses had a reduction of < 90%. Interestingly, 2 of the 3 horses with FEC reductions < 90% were on the same farm (out of 13 horses tested). The mean FEC reduction across all farms was > 99.9%, but on this one farm the arithmetic mean reduction was 96.7% with a bootstrap 95% confidence interval (96.2, 97.1). Thus, resistance was suspected. However, when the 95% bootstrap confidence interval was recalculated taking into account inter-farm variability the interval increased to (99.2, 99.9), suggesting that there was no resistance present. To further investigate whether we were truly dealing with a case of ivermectin resistance on this farm (note that resistance to ivermectin in cyathostomins had not yet been reported when this study was performed in 2002), we performed a second FECRT. Fortunately, this was a large farm, so we divided horses into 2 groups, each with 20 horses. One group received a label dose of ivermectin (200 μg/kg), and one group received a half dose of ivermectin (100 μg/kg). Even at 100 μg/kg, the efficacy of ivermectin against cyathostomins is known to be > 99%, thus if resistance was present, the reduced dose would be a more sensitive means to detect it. This method of diagnosis for ivermectin resistance is commonly performed in Australian sheep. The second FECRT on this farm yielded results consistent with ivermectin susceptibility. Percent reduction in FEC was > 99.9% in both groups of horses. These findings indicate that the reduced efficacy observed on the first test was due to variability and not to reduced effectiveness of the drug. These findings also validated the results of the initial bootstrap analysis that took inter-farm variability in account. This example illustrates the need to account for inter-farm variability when performing multiple-farm studies. In our experience, approximately 5% of horse farms will yield reduced efficacy that is not associated with ivermectin resistance. Thus, not accounting for inter-farm variability will lead to an overestimation of resistance prevalence. This issue is most important when diagnosing the first cases of resistance, as is the current situation with ivermectin and moxidectin.

Box 3: Statistical Methods and Software: Likelihood based methods.
Parametric Likelihood based methods

Description: In this approach we take the model to be as described in (2) and base our inference on the maximum likelihood procedure.
Software: Statistical software packages such as SAS and R can be used to fit this model. In SAS, this is achieved using PROC GENMOD while in R it is achieved using lmer. These methods also allow for flexible modeling of various distributions for egg counts and appropriate regression models.
Advantages: If the assumed statistical models are correct one can obtain improved power in detecting resistance.
Disadvantages: The methods may not be “robust" to parametric assumptions and presence of outliers in the data. Additionally, these methods are based on conditioning on the observed pre-treatment egg counts and hence variability in pre-treatment counts is not accounted for in the standard analysis outputs.

Other Likelihood based Methods

Description: These methods use variations of model (2) that allow one to empirically determine appropriate correlations induced by random effects. Inference is based on modifications to the likelihood that enable the ease of maximization.
Software: Statistical programming languages such as SAS and R can be used to fit this model. In SAS, this is achieved using PROC GENMOD, PROC GLIMMIX, and PROC NLMIXED while in R it is achieved using lmer.
Advantages: If the parametric model and empirically chosen correlation matrix are correct, then one can get improved power.
Disadvantages: Apart from the disadvantages stated before the fitted models may not be stable.

Box 4: Statistical Methods and Software: Frequentist Bootstrap.

Parametric Bootstrap

Description: In this approach we model the pre-treatment and post-treatment data using parametric models. Then we estimate the parameters of the distribution (using MLE, for example) and construct confidence intervals based on the parametric bootstrap (Efron and Tibshirani, 1993). The basic idea of the parametric bootstrap is to simulate data from the assumed parametric distributions using the estimated parameters. We outline the basic steps below:

1. Fit a parametric model to the pre-treatment data; for example, Poisson, negative binomial, zero-inflated Poisson, or zero-inflated Poisson. Call the parameter θ and the estimated parameter $\hat{\theta}$. For example, in the negative binomial model θ is two dimensional.
2. Fit a binomial model for the post-treatment data using the given pre-treatment data. This produces an estimate $\hat{p}$.
3. Simulate the pre-treatment data from the assumed pre-treatment distribution using $\hat{\theta}$.
4. Simulate the post-treatment data from the simulated pre-treatment data and the binomial distribution with the estimated $\hat{p}$.
5. Repeat Steps 3 and 4 B times (often B = 2000) and proceed as in Vidyashankar et al. (2007).

Software: Software is not readily available but easily implementable using standard software like SAS and R
Advantages: This method can increase the accuracy of inference in small samples if the assumed parametric model is correct. This method also provides an easy approach to take into account variability in pre-treatment egg counts.
Disadvantages: Validity of inference is critically dependent on the assumed parametric model.

Non-parametric Bootstrap

Description: Details, including a step-by-step procedure, are given in Vidyashankar et al. (2007).
Software: This code typically needs to be written by a statistician. This can be accomplished in either R or SAS.
Advantages: This method allows provides an easy non-parametric approach to take into account pre-treatment variability. More detailed descriptions are provided in Vidyashankar et al. (2007)
Disadvantages: Since the method is non-parametric, it could lead to loss of power in small samples. More detailed descriptions are provided in Vidyashankar et al. (2007).

Box 5: Statistical Methods and Software: Bayesian Methods.

Bayesian Parametric Methods

Description: In this approach, the scientist specifies a prior distribution for the parameters and bases inference on the posterior distributions given the data.
Software: The methods can be implemented using SAS, R, and BUGS (or the linux version JAGS). Proc Bgenmod in SAS allows a variety of flexible models for both the pre- and post-treatment egg count distributions.
Advantages: These methods allow one to take into account pre-treatment FEC variability. In complicated models Markov chain Monte Carlo (MCMC) algorithms work efficiently and are more stable than the likelihood based methods.
Disadvantages: Convergence of MCMC algorithms and choice of right prior are usually complicated and require care.