Accepted Manuscript

Title: Novel Insights in the faecal egg count reduction test for monitoring drug efficacy against gastrointestinal nematodes of veterinary importance

Authors: B. Levecke, R.J. Dobson, N. Speybroeck, J. Vercruysse, J. Charlier

PII: S0304-4017(12)00144-6
DOI: doi:10.1016/j.vetpar.2012.03.020
Reference: VETPAR 6294

To appear in: Veterinary Parasitology

Received date: 22-12-2011
Revised date: 9-3-2012
Accepted date: 13-3-2012

Please cite this article as: Levecke, B., Dobson, R.J., Speybroeck, N., Vercruysse, J., Charlier, J., Novel Insights in the faecal egg count reduction test for monitoring drug efficacy against gastrointestinal nematodes of veterinary importance, Veterinary Parasitology (2010), doi:10.1016/j.vetpar.2012.03.020

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Novel Insights in the faecal egg count reduction test for monitoring drug efficacy against gastrointestinal nematodes of veterinary importance

Levecke B.¹, Dobson R.J.², Speybroeck N.³, Vercruysse J.¹, Charlier J.¹

¹ Department of Virology, Parasitology and Immunology, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium
² Division of Health Sciences, School of Veterinary and Biomedical Sciences, Murdoch University, Murdoch, Western Australia
³ Institute of Health and Society, Université Catholique de Louvain, Louvain, Belgium

*Corresponding author: Bruno Levecke, Department of Virology, Parasitology and Immunology, Ghent University, Faculty of Veterinary Medicine, Salisburylaan 133, 9820 Merelbeke, Belgium, Tel. +32 9 264 74 04, Fax. +32 9 264 74 96, e-mail: Bruno.Levecke@UGent.be

Keywords: faecal egg count reduction test; anthelmintic efficacy; monitoring programmes; anthelmintic resistance; Monte Carlo simulation; classification trees
Abstract

The faecal egg count reduction test (FECRT) is the method of choice to monitor anthelmintic efficacy against gastro-intestinal nematodes in livestock. Guidelines on how to conduct a FECRT are made available by the World Association for the Advancement of Veterinary Parasitology (WAAVP). Since the publication of these guidelines in the early 1990s, some limitations have been noted, including (i) the ignorance of host-parasite interactions that depend on animal and parasite species, (ii) their feasibility under field conditions, (iii) appropriateness of study design, and (iv) the low analytic sensitivity of the recommended faecal egg count (FEC) method. Therefore, the objective of the present study was to empirically assess the impact of the level of excretion and aggregation of FEC, sample size and detection limit of the FEC method on the sensitivity and specificity of the FECRT to detect reduced efficacy (<90% or <95%) and to develop recommendations for surveys on anthelmintic resistance. A simulation study was performed in which the FECRT (based on the arithmetic mean of grouped FEC of the same animals before and after drug administration) was conducted under varying conditions of mean FEC, aggregation of FEC (inversely correlated with k), sample size, detection limit and „true” drug efficacies. Classification trees were built to explore the impact of the above factors on the sensitivity and specificity of detecting a truly reduced efficacy. For a reduced-efficacy threshold of 90%, most combinations resulted in a reliable detection of reduced and normal efficacy. For the reduced-efficacy threshold of 95% however, unreliable FECRT results were found when sample sizes <15 were combined with highly aggregated FEC (k = 0.25) and detection limits ≥5 EPG or when combined with detection limits ≥15 EPG. Overall, an increase in sample size and mean preDA FEC, and a decrease in detection limit improved the diagnostic accuracy. FECRT remained inconclusive under any evaluated condition for drug efficacies ranging from 87.5% to 92.5% for a reduced-efficacy-threshold of 90% and from 92.5% to 97.5% for a threshold of
95%. The results highlight that (i) the interpretation of this FECRT is affected by a complex interplay of factors, including the level of excretion and aggregation of FEC and (ii) the diagnostic value of FECRT to detect small reductions in efficacy is limited. This study, therefore, provides a framework allowing researchers to adapt their study design according to a wide range of field conditions, while ensuring a good diagnostic performance of the FECRT.
1. Introduction

The periodic administration of anthelmintics is currently the most widely used method to control gastrointestinal nematode infections in animals. Rather than aiming to achieve elimination, these control programmes focus on reducing infection intensity and transmission to prevent production losses (Shaw et al., 1998; Charlier et al., 2009). The four major anthelmintic families are the benzimidazoles, imidazothiazoles/tetrahydroxypyrimidines, macrocyclic lactones and the amino-acetonitrile derivates. The intensive use of anthelmintics has led to an increasing problem of anthelmintic resistance and the need to monitor and detect changes in drug efficacy (Wolstenholme et al., 2004; Kaplan, 2004). Now, the faecal egg count reduction test (FECRT) is the method of choice to monitor anthelmintic efficacy and guidelines on how to conduct a FECRT were described by the World Association for the Advancement of Veterinary Parasitology (WAAVP) (Coles et al., 1992). These guidelines provide recommendations on the experimental set up (randomized control trial), sample size (≥10 or ≥15 animals per treatment group, each excreting at least 150 eggs per gram of faeces (EPG)), the faecal egg count (FEC) method (McMaster egg counting method with detection limit of 10 to 50 EPG, (MAFF, 1986)), statistical analysis (FECRT based on the arithmetic mean of grouped FEC after drug administration) and criteria defining reduced efficacy (FECRT <90% or FECRT <95% and lower limit of 95% confidence interval <90%), and this for a variety of animal species, including small ruminants (sheep and goats), cattle, horses and pigs. The WAAVP guidelines do not consider the variation in fecundity and aggregation between (e.g. Ascaris suum and Trichuris suis infections in pigs) and within parasite species (e.g. the influence of immunity development on egg shedding). Moreover, the recommendations on the study design and the sample size are stringent, and can be difficult to meet under field conditions. McKenna (2006) and Dobson et al. (2011) found that sensitivity was reduced when using the randomized controlled study design advocated in the WAAVP
guidelines; the latter also found it to be a poor study design when parasites were highly aggregated. Hence, FECRT based on the arithmetic mean of grouped FEC of the same animals before and after drug administration (Kochapakdee et al., 1995) is likely to provide more reliable results. Finally, the high detection limit of the recommended FEC method may thwart the precision of FECRT results (Levecke et al., 2011a). Alternative FEC methods are FLOTAC (Cringoli et al., 2010) and FECPAK (www.fecpak.com), allowing for the detection of 1 and 5 EPG, respectively.

The objective of the present study was to empirically assess the impact of sample size, detection limit of the FEC method, and level of excretion and aggregation of FEC on the interpretation of the FECRT in the absence of a control group. To this end, data were generated using a statistical simulation and subsequently analysed using tree-based models. From the results, we provide recommendations for future study designs to monitor drug efficacy against gastrointestinal nematodes of veterinary importance.

2. Methods

The study consisted of two consecutive methodological procedures. First, data were generated using a simulation in which the „true” drug efficacy (TDE) was evaluated by the FECRT under varying conditions of level of excretion and aggregation of FEC across the host population, sample size and detection limit of the FEC methods. Subsequently, the obtained data were analyzed using tree-based models, to determine their impact on the interpretation of FECRT and to assess critical values in terms of specificity to detect normal efficacy and sensitivity to detect reduced efficacy.
2.1. Data generation

Data were generated by Monte Carlo simulation as previously described by Levecke et al. (2011b) and are illustrated in Supplemental Figure 1. First, the distribution of parasites within the host population before administration of drugs was defined by a negative binomial distribution. This distribution is determined by two parameters: the mean level of egg excretion across animals (mean pre-drug administration (preDA) FEC) and the level of aggregation of FEC across animals (k). Low values of k indicate that only few animals are excreting the majority of the eggs, high values indicate that egg counts are more equally distributed across the host population. From this pre-defined distribution, a number of individual animals were randomly drawn representing the sample size. The preDA FEC observed, however, will be different from the ‘true’ preDA FEC due to the variation (i.e. stochasticity) introduced by sampling eggs associated with the FEC method. This component of variation was simulated using the Poisson distribution defined by the expected number of eggs counted (=‘true’ preDA FEC/detection limit). In order to simulate a TDE, the ‘true’ preDA FECs of these animals were multiplied by 1-TDE. The observed FEC after the administration of the drug (postDA FEC) was generated as described above for the preDA FEC. Subsequently, the FECRT was calculated as described in the formula below (Kochapakdee et al., 1995), and is based on the arithmetic mean of pre- and postDA FEC of the same animals. Finally, the entire process was iterated 500 times, to obtain 500 estimates of FECRT given a pre-defined parasite population, sample size, detection limit and TDE.

\[
\text{FECRT} \, (\%) = 100\% \times \frac{\text{arithmetic mean (preDA FEC)} - \text{arithmetic mean (postDA FEC)}}{\text{arithmetic mean (preDA FEC)}}
\]

The parasite/host population parameter values chosen for mean preDA FEC (50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 and 1,000) and k (0.1, 0.25, 0.5, 1.0) were used in the simulations.
0.75, 1, 1.5 and 2) were based on previously conducted studies in which gastrointestinal
nematodes were quantified in goats (Hoste et al., 2002), sheep (Morgan et al., 2005), cattle
(El-Abdellati et al., 2010), horses, pigs and camelids (Laboratory of Parasitology, Faculty of
Veterinary Medicine, Ghent University (Belgium), unpublished data). The values for the
sample size were 6, 10, 15, 20, 25, 50, 75, 100, 125, 150, 175 and 200, covering a large range
of herd sizes. The values for detection limit represented those of four currently used FEC
methods: FLOTAC (detection limit = 1 and 2 EPG) (Cringoli et al., 2010), FECPAK
(detection limit = 5 and 10 EPG) (www.fecpak.com) and McMaster (detection limit = 10, 15,
25, 33.3 and 50 EPG) (MAFF, 1986). The TDE was set on 50, 60, 70, 80, 82.5, 85, 87.5, 90,
92.5, 95, 97.5 and 99%, resulting in 120,960 combinations (15 (mean preDA FEC) x 7 (k) x
(sample size) x 8 (detection limit) x 12 (TDE)) that were each iterated 500 times.

2.2. Analysis of data using tree-based models

The impact of the various factors on the sensitivity and specificity of the FECRT was
evaluated. Every TDE that was less than 90% or 95% was considered as a truly reduced
efficacy and as truly efficacious if different.

A combination of evaluated factors (500 iterations) was considered to be “sensitive”
when a FECRT could be calculated (observed mean preDA FEC >0) and a truly reduced
efficacy (TDE <90% or <95%) was correctly detected in at least 95% of the iterations or
“insensitive” (i.e. false negative) otherwise. A combination of evaluated factors was
considered to be “specific” when a FECRT could be calculated (observed mean preDA FEC
>0) and TDE ≥90 or ≥95% was correctly detected in at least 95% of the iterations or
“aspecific” (false positive) otherwise.

Subsequently, tree-based models (classification trees) were built (Speybroek et al., 2004)
to determine critical values for the parasite-host population (mean preDA FEC and k), the
sample size, and the detection limit affecting the interpretation of the FECR. To this end, the
together with the packages „rpart“ (Speybroeck, 2011) and „randomforest“ from the statistical software R was
used (version 2.10.0, 2009, The R Foundation for Statistical Computing) with both sensitivity
and specificity as a binary outcome variable and mean preDA FEC, aggregation (k), sample
size, detection limit and TDE as predictor variables.

3. Results

3.1. Detection of a normal or reduced efficacy for a threshold of 90%

The different combinations of sample size, detection limit, mean preDA FEC and k, and
their respective TDE limits for which the FECRT cannot reliably provide a correct diagnosis
(sensitivity and specificity <85%) are summarized in Figure 1.

For small sample sizes (6 and 10 animals), detection limit was the most important factor:
the reliability of diagnosing reduced and normal efficacy increased when the detection limit
decreased. With a detection limit ≥15 EPG reduced and normal efficacy could only reliably be
detected when the TDE ≤60% and ≥95%, respectively, whereas a detection limit of 1 and 2
EPG allowed to accurately detect reduced and normal efficacy up to a TDE of ≤87.5% and
≥95.0%, respectively.

For moderate (15 and 25 animals) sample sizes, FECRT results were also affected by
mean preDA FEC: FECRT results were less discriminatory when mean preDA FEC
decreased. However, the impact of mean preDA FEC on the detection of reduced and normal
efficacy decreased when detection limit decreased. For detection limits ≥15 EPG, the TDE
limits where ≤60% and ≥95% when mean preDA FEC were low (50 and 100 EPG), whereas
detection of normal and reduced efficacy became more reliable when mean preDA FEC were
≥150 EPG (≤82.5% and ≥95.0%). For detection limits of 1 and 2 EPG, the TDE limits to
accurately detect reduced and normal drug efficacy were ≤87.5% and ≥95.0%, regardless of
the mean preDA FEC.

For large sample sizes (≥ 50 animals), the FECRT was able to detect reduced and normal
efficacy when TDE were ≤87.5% and ≥92.5%, except for combinations where detection limits
≥10 EPG were combined with a mean preDA FEC of 50 and 100 EPG, for these combinations
the TDE limits were ≤82.5% and ≥92.5%.

For the 90% threshold aggregation of the FEC (k) had no effect on specificity or
sensitivity.

3.2. Detection of a normal or reduced efficacy for a threshold of 95%

In analogy with a threshold of 90%, the different combinations of sample size, detection
limit, mean preDA FEC and k, and their respective TDE limits are summarized in Figure 2. In
comparison with a threshold of 90%, four important differences can be noted. First, the
aggregation of the FEC did affect the FECRT results. Moreover, for small sample sizes (6 and
10 animals) combined with highly aggregated FEC (k = 0.1) and detection limits of 5, 10 and
15 EPG or combined with detection limits ≥15 EPG and mean preDA FEC of 50 and 100
EPG did not allow for a reliable detection of a reduced and normal efficacy. Finally, the
variation in TDE limits across the remaining combinations was small. The least
discriminatory FECRT results were observed for sample sizes (6 and 10 animals) (TDE
limits: ≤85.0% and ≥97.5%). The most discriminatory combinations (TDE limits: ≤92.5% and
≥97.5%) were found for: (1) sample sizes of 15 and 25 animals combined with detection
limits of 1, 2.5 and 10 EPG and mean preDA FEC ≥250 EPG, (2) sample sizes ≥50 combined
with detection limit of 1, 2.5 and 10 EPG EPG, and (3) sample sizes ≥50 combined with a
detection limit ≥15 EPG and preDA FEC ≥150 EPG. In the remaining combinations the TDE limits were ≤87.5% and ≥97.5%.

4. Discussion

The present study provides novel insights into three aspects of the FECRT in absence of a control group. The first important finding is that a correct interpretation of the FECRT is not always possible. This can be due to factors inherent to the study design, but also to the level of excretion and aggregation of eggs within the host population. For a threshold of 95%, we must accept poor diagnostic performance when small sample sizes (6 and 10 animals) are combined with highly aggregated FEC (k = 0.1) and detection limits of 1, 2, 5, 10 and 15 EPG or when these sample sizes are combined with high detection limits (≥15 EPG) and low mean preDA FEC (50 and 100 EPG). As a consequence, previous studies that assessed anthelmintic efficacy performed under these conditions may have resulted in inconclusive results and should be interpreted with caution (e.g. El-Abdellati et al., 2010). Secondly, the results highlight that their will always be an interval of TDE for which the FECRT remains inconclusive, even in optimal conditions of sample size, detection limit, mean preDA FEC and aggregation (cfr. Figure 1 and 2). These intervals range from 87.5% to 92.5% and from 92.5% to 97.5% for a 90% and 95% threshold, respectively. Finally, the results indicate that recommendations on the sample size can be less stringent in cases where FEC methods that are more sensitive than McMaster are applied. We, therefore, believe that Figures 1 and 2 can provide researchers with the flexibility to adapt their study design according to a wide range of field conditions.

Based on a comparison of FECRT results with control slaughter data to detect anthelmintic resistance both McKenna (2006) and Dobson et al. (2011) found specificity to be 100%, that is, susceptible nematode isolates were correctly diagnosed as susceptible. They
found sensitivity to be around 90-95%, that is, most resistant nematode isolates were correctly
diagnosed as resistant. The results of this study were consistent with their findings as the
range of true drug efficacies that provide doubtful results were generally greater below the
threshold than above it (cfr. Figure 1 and 2). In practice other factors that may lead to a false
declaration of reduced efficacy, such as faulty equipment, incorrect dosage, missed animals,
sub-standard product, miss-labeled samples etc. were not considered in this exercise. In the
laboratory and on farm these are errors that can be difficult to eliminate completely and will
tend to reduce FECRT specificity. By comparison with preDA FEC, detection level and
sample size aggregation of FEC (k) played a minor role in FECRT precision when the
threshold was 95% and had no effect on precision when the threshold was 90%. This was
because the simulated study design (pre- and post-treatment FEC of the same animals)
minimizes the negative impact of small k on FECRT precision. The advantage of this design
over FECRT involving separate treated and control groups was demonstrated by Dobson et al.
(2011) for k set at 0.5 and 2.

Anthelmintic drug efficacy in vivo is monitored by the FECRT, whereas the
evaluation/registration of efficacy of novel anthelmintic compounds for veterinary application
requires the assessment of reduction in gastrointestinal nematode counts (Vercruysse et al.,
2001; 2002). In analogy with the FECRT, there is a large variation in the number and the
aggregation of gastrointestinal nematodes across animals. Current guidelines recommend
counting at least 2% of the gastrointestinal nematode population collected (detection limit =
50 gastrointestinal nematodes), but, as for the FECRT guidelines, the level of understanding
of the effects of factors inherent to study design, parasite and/or host species to support these
guidelines is poor. Therefore, a similar approach as in this study might also be used to deduct
implications for study designs assessing drug efficacy based on gastrointestinal nematode
counts.
Although this simulation provides novel insights in the FECRT, we should also underscore the limitations of the present study. First, this analysis only improves the selection of a study design. For the reliability of a conclusion towards presence of reduced efficacy on an individual case, 95% confidence intervals are still to be calculated (Coles et al., 1992, Dobson et al., 2011, Vidyashankar et al., 2012). Second, we did not consider additional variation caused by differences in properties of the FEC method beyond the detection limit (Cringoli et al., 2004; Pereclienè et al., 2007; Cringoli et al., 2010). Extrapolation of the performance of FECR across different FEC methods should be done with care as recent studies show that FEC methods with the same detection limit, but different methodologies, do not guarantee an equal level of performance of FECR (Levecke et al., under review). Finally, this simulation is restricted to parasite populations defined by a mean baseline FEC between 50 and 1,000 EPG, k between 0.01 and 2, and sample sizes between 6 and 200, and did only explore a limited number of values between these intervals.

In conclusion, this study assessed the impact of various factors on the interpretation of probably the most reliable FECRT design. The results highlight that the interpretation of the FECRT is affected by a complex interplay of various factors, including the mean and level of aggregation of FEC. This study provides an empirical-based framework allowing researchers to adapt their study design according to a wide range of field conditions, while ensuring a good diagnostic performance of the FECRT.

Acknowledgements

BL is funded by the Fund for Scientific Research-Flanders (Belgium) (F.W.O.-Vlaanderen). This study received funding from the E.U. FP7 GLOWORM project (Grant agreement no 288975CP-TP-KBBE.2011.1.3-04).
References


Figure captions

Figure 1. The detection of reduced (true drug efficacy <90%) and normal drug efficacy (true drug efficacy ≥90%). The white zone indicates which drug efficacies can be accurately determined by FECRT as having reduced efficacy, whereas the black zone indicates which drug efficacies can be accurately determined as having a normal efficacy. The grey zone indicates drug efficacies that cannot be accurately determined as having either reduced or normal efficacy.

Figure 2. The detection of reduced (true drug efficacy <95%) and normal drug efficacy (true drug efficacy ≥95%). The white zone indicates which drug efficacies can be accurately determined by FECRT as having reduced efficacy, whereas the black zone indicates which drug efficacies can be accurately determined as having a normal efficacy. The grey zone indicates drug efficacies that cannot be accurately determined as having either reduced or normal efficacy.

Supplemental Figure 1. Schematic presentation of the data generation based on Monte Carlo simulation.
<table>
<thead>
<tr>
<th>Sample size</th>
<th>Detection limit (EPG)</th>
<th>Mean pre DA FEC (EPG)</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-10</td>
<td>1-2</td>
<td>50-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td>5-10</td>
<td></td>
<td>50-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td>15-50</td>
<td></td>
<td>50-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td>15-25</td>
<td>1-2</td>
<td>50-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>50-400</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>450-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>50-100</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150-400</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>450-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td>15-50</td>
<td>50-100</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td>50-200</td>
<td>1-5</td>
<td>50-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td>10-50</td>
<td></td>
<td>50-100</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150-1,000</td>
<td>0.1-2</td>
</tr>
</tbody>
</table>

50%  | 60%  | 70%  | 80%  | 90%  | 100% |

- Reduced
- Doubtful
- Normal

Figure 1
<table>
<thead>
<tr>
<th>Sample size</th>
<th>Detection limit (EPG)</th>
<th>Mean pre DA FEC (EPG)</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-10</td>
<td>1-2</td>
<td>50-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td>5-10</td>
<td>50-1,000</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50-1,000</td>
<td>0.25-2</td>
</tr>
<tr>
<td>15-50</td>
<td>50-100</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>150-1,000</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.25-2</td>
</tr>
<tr>
<td>15-25</td>
<td>1-10</td>
<td>50-200</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>250-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td>15-50</td>
<td>50-1,000</td>
<td>0.1-2</td>
<td></td>
</tr>
<tr>
<td>50-200</td>
<td>1-10</td>
<td>50-1,000</td>
<td>0.1-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150-1,000</td>
<td>0.1</td>
</tr>
</tbody>
</table>

![Bar chart showing 'True' drug efficacy with reduced, doubtful, and normal categories.](chart.png)