Veterinary Parasitology 296 (2021) 109509

Contents lists available at ScienceDirect

Veterinary Parasitology

journal homepage: www.elsevier.com/locate/vetpar

Review article

What makes a good fecal egg count technique?

Martin K. Nielsen

M.H. Gluck Equine Research Center, Department of Veterinary Science, University of Kentucky, Lexington, KY, USA

ARTICLE INFO

Keywords:
Validation
Diagnosis
Fecal egg count
Performance
Design

ABSTRACT

The first parasite fecal egg counting techniques were described over 100 years ago, and fecal egg counting remains essential in parasitology research as well as in clinical practice today. Several novel techniques have been introduced and validated in recent years, but this work has also highlighted several current issues in this research field. There is a lack of consensus on which diagnostic parameters to evaluate and how to properly design studies doing so. Furthermore, there is a confusing and sometimes incorrect use of terminology describing performance of fecal egg counting techniques, and it would be helpful to address these. This manuscript reviews qualitative and quantitative diagnostic performance parameters, discusses their relevance for fecal egg counting techniques, and highlights some of the challenges with determining them. Qualitative parameters such as diagnostic sensitivity and specificity may be considered classic diagnostic performance metrics, but they generally only have implications at low egg count levels. The detection limit of a given technique is often referred to as the "analytical sensitivity", but this is misleading as the detection limit is a theoretically derived number, whereas analytical sensitivity is determined experimentally. Thus, the detection limit is not a diagnostic performance parameter and does not inform on the diagnostic sensitivity of a technique. Quantitative performance parameters such as accuracy and precision are highly relevant for describing the performance of fecal egg counting techniques, and precision is arguably the more important of the two. An absolute determination of accuracy can only be achieved by use of samples spiked with known quantities of parasite ova, but spiking does not necessarily mimic the true distribution of eggs within a sample, and accuracy estimates are difficult to reproduce between laboratories. Instead, analysis of samples from naturally infected animals can be used to achieve a relative ranking of techniques according to egg count magnitude. Precision can be estimated in a number of different approaches, but it is important to ensure a relevant representation of egg count levels in the study sample set, as low egg counts tend to associate with lower precision estimates. Coefficients of variation generally provide meaningful measures of precision that are independent of the multiplication factor of the techniques evaluated. Taken together, there is a need for clear guidelines for studies validating fecal egg counting techniques in veterinary parasitology with emphasis on what should be evaluated, how studies could be designed, and how to appropriately analyze the data. Furthermore, there is a clear need for better consensus regarding use of terminology describing the diagnostic performance of fecal egg count techniques.

1. Introduction

The principle of fecal egg counting remains a cornerstone in veterinary parasitology. Over 100 years ago, Bass (1909) described using flotation to recover and count parasite ova in fecal samples, and multiple techniques exist today. The classic flotation techniques can be divided into two main types; 1) counting chamber or 2) test tube and cover slip. A classic technique using an egg counting chamber is the McMaster method, which was first described in the 1930s (Gordon and Whitlock, 1939), and is widely regarded as an industry standard today. Multiple modifications have been developed and several counting chamber based techniques such as the Moredun method (Jackson, 1974), FECPAK (Presland et al., 2005), FLOTAC (Cringoli et al., 2010), and Mini-FLOTAC (Cringoli et al., 2017) can be regarded as modifications of the McMaster principle. The test tube and cover slip approach to fecal egg counting was first described in 1928, where Lane outlined a technique based on centrifugation of fecal matter suspended in flotation medium contained in centrifuge tubes with glass cover slips on top (Lane, 1928). This technique was later adapted for counting *Haemonchus contortus* eggs in sheep feces (Stoll, 1930), and it is commonly referred to as the Stoll technique today. The widely used Wisconsin (Cox and Todd, 1962) and Cornell-Wisconsin (Egwang and Slocombe, 1982) methods

E-mail address: martin.nielsen@uky.edu.

https://doi.org/10.1016/j.vetpar.2021.109509
Received 31 March 2021; Received in revised form 19 June 2021; Accepted 23 June 2021
Available online 27 June 2021

are modifications of the Stoll technique.

Today, fecal egg counts are used for a variety of different purposes; 1) evaluation of anthelmintic treatment efficacy by the fecal egg count reduction test (FECRT), 2) identification of animals in need of anthelmintic treatment as part of a targeted anthelmintic treatment scheme (Kenyon et al., 2009; Kenyon and Jackson, 2012; Kaplan, 2013; Nielsen et al., 2014), 3) identification of different ova types present in the feces to guide treatment decisions (Nielsen et al., 2019), and 4) identification of animals with low trichostrongylid egg counts to be used as target phenotypes in sheep breeding programs in some parts of the world (Bisset et al., 1996; Gowane et al., 2020). The ever-increasing levels of anthelmintic resistance reported in parasites of livestock (Kaplan, 2004; Howell et al., 2008; Gasbarre, 2014; Ploeger and Everts, 2018) and horses (Samson-Himmelstjerna, 2012; Matthews, 2014; Peregrine et al., 2014; Nielsen et al., 2020) have led to an increased emphasis on the need for surveillance-based parasite control regimens to reduce the reliance on routine administration of anthelmintic products and thereby slow further development of resistance (Kenyon et al., 2009; Kaplan and Nielsen, 2010; Kenyon and Jackson, 2012). Furthermore, it is of utmost importance to routinely evaluate anthelmintic treatment efficacy by means of the FECRT. For the interpretation of the FECRT, it is important to reliably distinguish a true egg count reduction from random data variation, and it has been shown that the choice of egg counting technique has a major influence on the outcome of the results (Levecke et al., 2011, 2012). Recent work has illustrated the importance of considering the magnitude of raw egg counts (prior to converting to eggs per gram of feces) in the FECRT analysis (Dobson et al., 2012; Levecke et al., 2018). The rationale is that the statistical power is driven by the number of eggs counted, and not the fecal egg count expressed as the number of eggs per gram (EPG). Depending on the technique used, an egg count of 200 EPG could represent as much as 200 eggs counted under the microscope or, for some commonly used methods, it could be a matter of only four eggs being counted. The number of eggs counted pre-treatment determines whether a reduced anthelmintic efficacy can reliably be detected with statistical significance. This “eggs counted” principle is now fully incorporated into the new guidelines for conducting FECRT studies and is part of a guideline document currently under evaluation by the World Association for the Advancement of Veterinary Parasitology (WAAVP) (Kaplan, R.M., Levecke, B., Denwood, M.J., Torgerson, P.R., Dobson, R. J., Thamsborg, S.M., Nielsen, M.K., Gilleard, J.S., Vercruysse, J. World Association for the Advancement of Veterinary Parasitology (WAAVP) guidelines for diagnosing anthelmintic resistance using the faecal egg count reduction test in ruminants, horses and swine.). This approach, obviously, emphasizes the importance of the choice of egg counting technique and this requires due consideration of relevant performance parameters. The FECRT remains the only field test capable of evaluating anthelmintic efficacy in all host species and of all anthelmintic drug classes, and it will continue to be an important research tool for years to come. Taken together, there is increasing emphasis on the use of fecal egg counts as parasite monitoring tools, and this, in turn, increases the importance of egg count test performance.

Despite the century-old principle behind fecal egg counting, a remarkable number of new techniques has been introduced in recent years. The widely used FLOTAC and Mini-FLOTAC techniques were both introduced in the past 15 years (Cringoli et al., 2010, 2017), and more recently, several novel digital imaging techniques have been developed for computerized parasite egg counting. While the principle of image analysis for fecal counting was already demonstrated in the 1990s (Sommer, 1996; Mes et al., 2007), techniques using this technology have generally been introduced within the past five years. The FECPAK^{G2} system was the first image-based system to be made widely available (Rashid et al., 2018; Tyson et al., 2020). However, eggs in the images are still counted manually, so this technique is not fully automated at this stage. Several other recently developed image-based systems, including Parasight, Telenostic, VETSCAN IMAGYST, and Kubic FLOTAC, are based on digital analysis of images taken and computerized counting of parasite ova present (Cain et al., 2020; Elghryani et al., 2020; Nagamori et al., 2020; Cringoli et al., 2021). Given the remarkable influx of these novel egg counting systems in recent years, there are reasons to expect that additional similar systems will be developed in the future, and these existing systems are likely to be refined and updated as well. Consequently, there will be an increasing need for validating and evaluating the performance of these systems and their modifications in the years to come.

It is clear from the above that there is an increasing need for evaluating, assessing and validating the performance of fecal egg counting techniques in veterinary parasitology. However, it is equally clear from the many recently published papers describing the performance of these techniques, that there is a pronounced lack of consensus of how this should be done, which diagnostic performance metrics that should be determined, and how they should be interpreted. Furthermore, there is a confusing use of terminology describing some of these performance characteristics for fecal egg counting techniques and there appears to be some misconceptions in need of being addresses and corrected.

The aim with this article is to 1) summarize performance metrics for fecal egg counting techniques, describe their appropriate interpretation, and outline challenges with determining them, 2) address common misconceptions related to some of these performance metrics, and 3) make a case for the need for developing uniform guidelines for validating fecal egg counting techniques in veterinary parasitology.

2. Qualitative diagnostic performance parameters

Diagnostic sensitivity and specificity are commonly used and readily recognized qualitative parameters describing the performance of diagnostic tests. These can be expanded to include positive and negative predictive values and likelihood ratios, which are all calculated from the same basic 2 × 2 table principle. Receiver operator characteristic (ROC) curve analysis is another commonly used methodology for evaluating qualitative diagnostic performance by determining the area under the curve, which can be done for the diagnostic sensitivity as well as specificity (Nielsen et al., 2010; Daş et al., 2017).

Despite often made claims, a direct connection between diagnostic sensitivity and quantitative parameters such as accuracy and precision does not exist, and a more sensitive technique cannot be expected to generate higher counts. As outlined in Section 2.4, the diagnostic sensitivity will primarily determine whether a given technique is capable of detecting true egg counts below 50 EPG (Noel et al., 2017; Scare et al., 2017), whereas neither diagnostic sensitivity nor detection limit of the technique is predictive of quantitative performance (Cain et al., 2020). However, as will be discussed in this manuscript, accuracy and precision can both affect diagnostic sensitivity at low egg count levels (Fig. 1).

A few different approaches can be taken to determine diagnostic sensitivity, but they each have limitations. Most approaches require a defined gold standard, against which these qualitative performance parameters can be calculated, but alternative options exist for scenarios without a designated gold standard. In the following, each of these study design approaches is outlined and discussed.

2.1. True gold standard evaluation

The true gold standard is presence or absence of adult fecund helminths in the intestinal tracts of the animals included in the study population. In other words, this generally requires a terminal study or access to abattoir material, although this information could theoretically be obtained by confining animals post anthelmintic therapy and recovering excreted parasites from the feces. This latter approach has been used for recovering ascarid parasites in human parasitology (Walker et al., 2009), but may be less feasible for smaller sized parasites, which could be more difficult to recover from the feces following deworming. A deworm and collect approach has been applied in equine

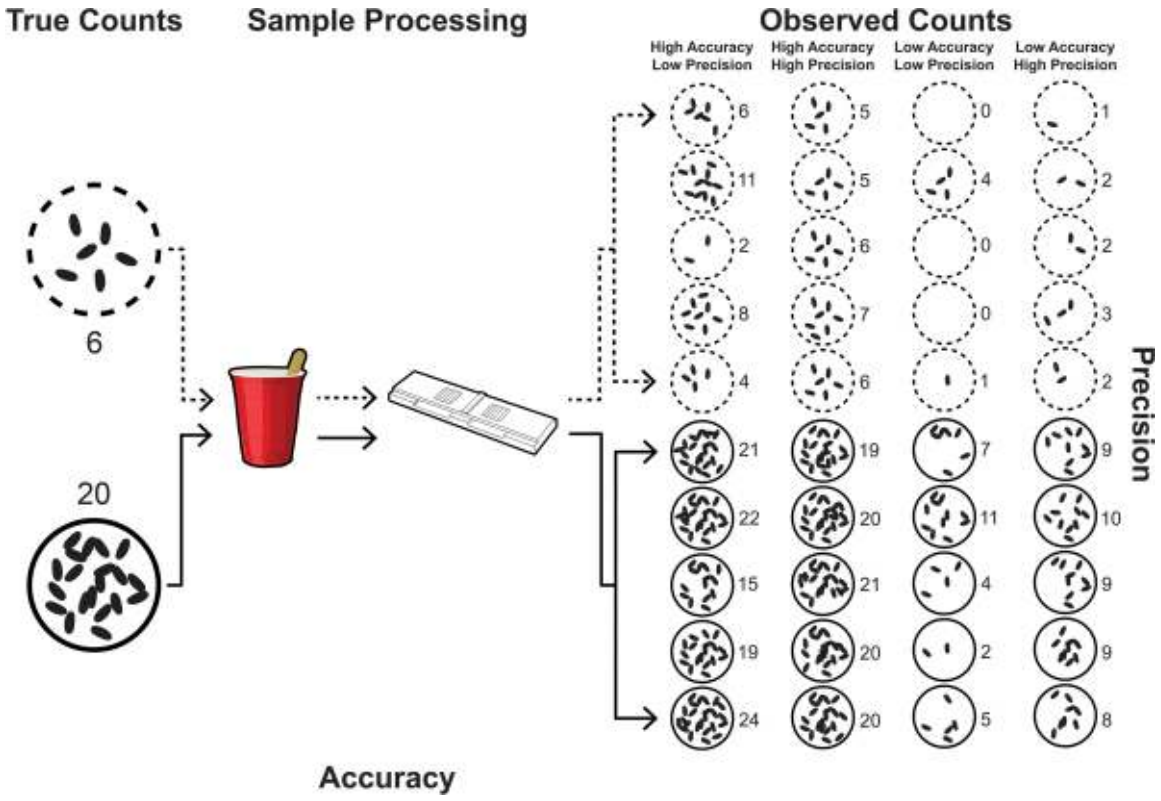

Fig. 1. Illustration of accuracy and precision in the context of fecal egg count techniques. The figure describes scenarios for two different fecal samples; one with an egg count of 6 eggs and one with a count of 20 eggs. Egg counting accuracy is defined by the egg loss occurring during sample processing and flotation, while precision can be described as the variation between repeated counts on the same sample. The figure exemplifies counts determined with four hypothetical techniques with different levels of accuracy and precision. In the low count example, some of the counts are negative for the low accuracy/low precision technique, illustrating the potential relationship between accuracy, precision and diagnostic sensitivity at low egg count levels. Figure courtesy Jamie K. Norris.

studies characterizing equine cyathostomin species communities (Kuzmina et al., 2005, 2011), but this has not been used for validating diagnostic techniques at this stage, and would only be appropriate if the anthelmintic used is fully efficacious. There are obvious limitations to the necropsy-based approach, as it can be challenging to compile enough data to allow reliable estimates of the desired diagnostic performance parameters. One option is to gather data from animals subjected to euthanasia due to other studies, but that often requires collecting data across years and decades, which can be problematic because of changing personnel, procedures, and protocols. There is a limited number of such published gold standard-based studies validating fecal egg count methods (Proudman and Edwards, 1992; Bøgh et al., 1994; Nilsson et al., 1995; Nielsen et al., 2010; Skotarek et al., 2010; Daş et al., 2017; Byrne et al., 2018) including one carried out with horses (Nielsen et al., 2010), and this low number of publications likely reflects the many challenges with taking this approach. The equine study was a retrospective study conducted with archived data accumulated from anthelmintic efficacy trials conducted over five decades (Nielsen et al., 2010). Here, we encountered another limitation of the necropsy approach; virtually all horses in the study were both worm count and fecal egg count positive for strongylid type parasites. As estimation of these qualitative performance parameters requires representation of both test positives and negatives as well as true positives and negatives in the data set, this can be difficult to achieve for very commonly occurring parasites.

Though considered the ultimate gold standard, it is important to recognize that necropsy techniques have clear limitations as well. In horses for example, counts of cyathostomin larval and adult stages are estimated from representative subsamples due to the large volume of gastrointestinal content (Chapman et al., 2003), and this could affect the recovery of species present in low quantities. Larger sized parasites such as *Parascaris* spp., *Anoplocephala perfoliata*, adult *Oxyuris equi*, and *Strongylus* spp., on the other hand, can be recovered without subsampling the ingesta, and infection with these parasites could, therefore, be expected to more reliably detected, when present.

2.2. Simulated gold standard studies

In the absence of necropsy-based validation studies, there are several alternative study design approaches to be taken; 1) usage of samples spiked with known quantities of parasite ova, 2) using a composite reference standard, and 3) using latent class modelling. Instead of evaluating whether a given technique is capable of diagnosing presence of adult parasites in the intestinal tract of the animals, these approaches all have a very different aim: They evaluate the ability to detect parasite ova present in the feces. This is a very different gold standard definition, as it has been documented that egg count negative animals can still harbor worms (Nielsen et al., 2010; Byrne et al., 2018). In the following, the different study design approaches will be outlined with a brief discussion of pros and cons for each.

The spiking approach has been used in numerous recent studies (Noel et al., 2017; Bosco et al., 2018; Paras et al., 2018; Nápravníková et al., 2019; Daş et al., 2020). While the principle seems straight-forward, it comes with several challenges. The source of ova and the method employed for retrieving them are obviously critical. Often, ova are isolated by means of washing and sieving (Bosco et al., 2018; Paras et al., 2018; Daş et al., 2020), during which they could sustain damage potentially affecting their subsequent recovery rates in flotation-based systems. Furthermore, flotation of ova is sometimes employed to assist recovery as well (Noel et al., 2017; Nápravníková et al., 2019; Daş et al., 2020), and this could further introduce a bias, as better floating ova are more likely to be retrieved. However, these

potential sources of bias have yet to be demonstrated and quantified in research studies, and they remain hypothetical at this point. Adding to this are concerns whether the artificially introduced ova can be fully distributed and integrated within the fecal matter. See further discussion of issues with spiking studies in Section 3.2. Taken together, it is clear that spiking remains a simulation of naturally infected samples, and results have to be interpreted with caution.

Using a composite reference standard (Alonzo and Pepe, 1999) is a well-known approach for validating egg counting techniques in human parasitology (Habtamu et al., 2011; Moser et al., 2018). Here, the idea is to employ several different diagnostic tests in a serial manner, and then regard the sample as a true positive, if ova are detected with at least one of these techniques. However, this has been demonstrated to result in biased performance estimates, if classification errors made by the tests included in the composite reference standard and the technique under evaluation are correlated (Dendukuri et al., 2018). Since fecal egg counting techniques are largely based on the same fundamental principles, there is strong reason to expect classification errors to be largely correlated, and using a composite reference standard is therefore likely to lead to biased estimates.

A better alternative for estimating qualitative diagnostic performance parameters in the absence of a true gold standard is to make use of latent class modelling (Reitsma et al., 2009; Dendukuri et al., 2018). This analysis can take into account prior knowledge about target organism prevalence in the study population and adjust for the performance of each technique included in the study and the possible conditional dependence between them. As an example of this approach, a recent study employed Bayesian modelling to estimate diagnostic sensitivity and specificity of egg counting techniques for detection of equine strongylid ova (Cain et al., 2020), and this appears to be a viable approach for future studies.

It should be emphasized that all of these approaches discussed here are based on positive egg counts either detected in samples from naturally infected animals or artificially created through spiking of ova-free samples. It is well-known that estimation of qualitative diagnostic performance parameters are highly affected by the prevalence of the target organism in the study population (Leeflang et al., 2013; Dendukuri et al., 2018), so the study outcome can be largely affected by the choice of study sample set. This also means that study results cannot necessarily be replicated between laboratories and that a given sensitivity and specificity are not fixed parameters for any given techniques. Furthermore, the sensitivity of egg counting techniques will largely depend on the magnitude of the counts among the positive samples included. If counts are relatively high, most egg counting techniques would detect them as positive (Noel et al., 2017, Scare et al., 2017), whereas some techniques would struggle if a large proportion of low egg count samples are included in the study. Taken together, it is clear that the outcome of a validation study is largely dependent on prevalence and egg count magnitude in the sample set included in the study, and it is important that this is appropriately accounted for in the study design.

2.3. Detection limit vs. diagnostic sensitivity

Some of the most commonly encountered misconceptions surrounding parasite fecal egg counting stems from misleading and confusing terminology often used about the detection limits of various techniques.

The detection limit of a parasite egg counting technique is synonymous with its multiplication (or conversion) factor. It cannot be emphasized enough that this is a theoretically derived number that does not inform about test performance. For a flotation-based technique, the detection limit is determined by three factors: 1. The mass of feces weighed and processed, 2. The volume of flotation medium, in which the feces are suspended, and 3. The volume of this suspension examined under the microscope. For a simple McMaster technique, this calculation will look as follows: 4 g of feces suspended in 56 mL of flotation medium, yielding 60 mL of suspension. From this, 0.3 mL is then examined within the McMaster chambers under the microscope. Based on this, the detection limit can be calculated as $(60/4)/0.3 = 50$ EPG. This number is based on two assumptions that are inherently wrong: 1) It assumes that no eggs are lost during the filtration, mixing and flotation, and 2) it assumes a perfect homogenization and distribution of eggs within the flotation medium. As will be covered in Section 3.2 Accuracy, there is always an egg loss involved with the procedure, regardless of the technique employed, so the detection limit does not equate the lowest detectable true egg count in a sample. Similar issues apply to the assumption that the subsample examined under the microscope is representative of the distribution of ova within the entire suspension, as it is ignoring subsample variability. See Section 3.3 Precision for further discussion.

In the scientific literature, the detection limit is often referred to as the "sensitivity"', "analytical sensitivity" or "technical sensitivity" (Levecke et al., 2012, 2018; Ballweber et al., 2014; Castro et al., 2017; George et al., 2021). However, these terms all imply aspects of diagnostic test performance, making them highly misleading. As demonstrated in a recent study, a technique with a lower detection limit does not necessarily perform with a higher diagnostic sensitivity (Cain et al., 2020). Some veterinary parasitologists seem to prefer the term analytical sensitivity (Levecke et al., 2012, 2018; Ballweber et al., 2014; Rinaldi et al., 2019), but even that is misleading. The proper definition of analytical sensitivity is the lowest detectable concentration of the target substance or organism in a sample, and its identification requires an appropriately designed study (Saah and Hoover, 1997). Since the detection limit (multiplication/conversion factor) of a given egg counting technique is not identified in a study, but is a theoretically derived number, it is not appropriate to refer to it as analytical sensitivity.

Taken together, "detection limit" is more appropriate than any term including the word "sensitivity". However, given the points presented above, it could be argued that even "detection limit" can be construed as implying aspects of test performance, so perhaps "multiplication" or "conversion" factor would be the most appropriate. Furthermore, it should be clear from this discussion that it is misleading to refer to techniques with low detection limits/multiplication factors as "more sensitive", unless diagnostic sensitivity has been appropriately documented.

2.4. The value of qualitative diagnostic performance

While diagnostic sensitivity and specificity are regarded standard diagnostic performance parameters, it should be recognized that fecal egg counts are, by definition, quantitative in nature. Thus, quantitative aspects of test performance are, arguably, more important. Adding to this, it should be kept in mind that the qualitative parameters only really matter at very low egg count levels. For example, studies completed in our laboratory have demonstrated how the different techniques evaluated demonstrated different diagnostic sensitivities at egg count levels at 50 EPG or below, whereas all evaluated techniques returned positive counts from samples above this level (Noel et al., 2017; Scare et al., 2017). See Fig. 1 for a schematic illustration of potential issues with diagnostic performance issues at low egg count levels. The question then becomes how important it is to reliably detect these lower counts. This, of course, depends on the purpose with the testing, but if the aim is to identify higher shedders for targeted anthelmintic treatment, the diagnostic sensitivity will be irrelevant, and quantitative aspects will be more important. In small animal practice, on the other hand, fecal samples are often examined qualitatively with little or no emphasis on the count (Dryden et al., 2005), and the diagnostic sensitivity is arguably important in this setting. Furthermore, it can be argued that detection of low egg count levels is important for the FECRT, as low egg counts in the post treatment samples could indicate reduced anthelmintic efficacy. Finally, the aforementioned "eggs counted" approach now taken for

ensuring adequate statistical power in FECRT studies (Dobson et al., 2012; Levecke et al., 2018), does emphasize the value of using techniques with lower detection limits, as they tend to count more eggs.

3. Quantitative diagnostic performance parameters

Correlation between egg counts determined with various techniques is often included in validation studies, but the two main quantitative diagnostic performance parameters to consider for quantitative techniques are accuracy and precision. These are two distinctly different measures, and the terms should not be used interchangeably. Fig. 1 illustrates the roles of accuracy and precision in fecal egg counting. Accuracy is determined by the proportion of ova lost during sample processing and flotation, while precision is a measure of the variation between repeated counts on the same sample. In the following, these quantitative measures are briefly outlined with a discussion of their relevance and choice of methodology to determine them.

3.1. Correlation

Many studies have included linear correlation analyses between the technique(s) being evaluated and commonly used techniques (Noel et al., 2017; Rashid et al., 2018; Nagamori et al., 2020; Elghryani et al., 2020; Cringoli et al., 2021). Demonstrating agreement with established methods is a generally meaningful way of providing proof of concept for a novel technique, but a few aspects should be kept in mind when interpreting such data. First of all, the linear correlation coefficient would be affected by precision of both techniques being correlated (see Section 3.3), and the correlation analysis does not allow determining if one or both techniques are major contributors to the data variability. If the expected level of precision is not known for the technique being compared against, it cannot be discerned if the variability observed is due to the technique under evaluation or the technique included for comparison purposes. The other aspect worth emphasizing is that demonstrating a linear correlation between egg counts determined with two techniques does not constitute a diagnostic validation, but merely a proof of concept.

If the study is using an intestinal worm count gold standard, linear correlations between worm and egg counts can be evaluated. However, although results may vary depending on the parasite type, host species, and age groups evaluated, such linear relationships should not be a general expectation (Roberts and Swan, 1981; Coadwell and Ward, 1982; Rehbein et al., 1997; Cabaret et al., 1998; Nielsen et al., 2010).

3.2. Accuracy

Diagnostic accuracy is a measure of how close a test measures to the true value of a given sample (Fig. 1). In other words, in this context, it is a measure of how close a determined egg count is to the true count of the sample. Some authors have used the term “bias” in this context, which is the inverse of accuracy (Levecke et al., 2012; Went et al., 2018), and therefore essentially describes the same aspects of quantitative performance.

The challenge with determining accuracy is that the true egg count is never truly known. Samples can be spiked with known quantities of ova, but as discussed in Section 2.2, there are several potential issues with this approach. In fact, these issues have been illustrated in several published studies. A study conducted in our laboratory determined egg recovery rates and compared the Mini-FLOTAC with the McMaster technique. On a set of spiked samples, the Mini-FLOTAC returned counts significantly closer to the true spiked value than the McMaster, but when comparing egg counts determined from naturally infected samples, there was no difference in the magnitude of counts determined with the two techniques (Noel et al., 2017). If the Mini-FLOTAC technique truly performed with a significantly higher accuracy, this should have been evident from the naturally infected samples as well. The possible issues with the spiked samples have also been illustrated by the wide range of accuracy estimates reported from different laboratories for the same technique, animal host, and ova type. For example, published accuracy estimates for determination of equine strongylid egg counts with the simple McMaster technique range from 20–25% (Noel et al., 2017; Scare et al., 2017) to over 90 % (Bosco et al., 2018; Nápravníková et al., 2019), and for the Mini-FLOTAC technique estimates have been reported ranging from 42 to 98 % (Noel et al., 2017; Scare et al., 2017; Bosco et al., 2018; Nápravníková et al., 2019; Amadesi et al., 2020). The studies referenced here strongly suggest that accuracy estimates are generally not reproducible between laboratories, while results can be more consistent for studies performed by the same group. As an example, our laboratory produced accuracy estimates that were largely repeatable for McMaster and Mini-FLOTAC strongylid egg counts in two different studies (Noel et al., 2017; Scare et al., 2017). The apparent lack of consistency between studies probably reflects differences in study design, source of ova, methodology used for recovering these, methodology used for spiking the samples, and flotation medium used, to mention some of the most obvious sources of error. Thus, it is clear that study designs for accuracy studies need standardization and clear guidelines. One recent study demonstrated that similar accuracy estimates can be obtained from different laboratories, if study protocols are followed stringently (Amadesi et al., 2020), illustrating that it is possible to develop and implement meaningful study protocols. Thus, protocols for standardized ring testing between laboratories should be developed and implemented, ensuring that diagnostic performance meets expected standards between laboratories.

Given the many potential biases associated with the spiking approach, an alternative to evaluating accuracy is to use naturally infected samples and simply compare egg count magnitudes determined by different techniques. While this alleviates many of the concerns outlined in Section 2.2, it does not allow an estimation of accuracy as a percentage of the true egg count. However, the techniques evaluated can be ranked according to their relative accuracy, which can still be useful. This approach is justified by the observation in numerous spiking studies that egg count accuracy never exceeds 100 %, which means that a degree of egg loss should always be expected (Noel et al., 2017; Scare et al., 2017; Bosco et al., 2018; Bortoluzzi et al., 2018; Nápravníková et al., 2019). Thus, a pragmatic interpretation is to simply conclude that the technique yielding the highest egg counts is the most accurate, and this has been the approach in several of our recent studies (Went et al., 2018; Slusarewicz et al., 2019; Cain et al., 2020).

It should be noted that although it may seem counter-intuitive, accuracy has limited implications for the performance of fecal egg counting techniques. First of all, accuracy of different egg counting techniques has been shown to be roughly constant across a wide range of egg count levels (Noel et al., 2017; Scare et al., 2017; Bosco et al., 2018; Nápravníková et al., 2019). Therefore, the FECRT calculation will not be affected by egg counting accuracy (Levecke et al., 2012), because this is a proportion calculated between the pre- and post-treatment egg counts, which are determined with the same accuracy. Another main use of fecal egg counts is for classifying animals into egg shedding categories (Nielsen et al., 2019). Here, the magnitude of counts determined with a given technique can obviously affect classification, but thresholds could always be adjusted to the accuracy of the technique employed, if necessary, so accuracy is not a major issue here either. Finally, since strongylid and ascarid fecal egg counts do not correlate with adult worm burdens (Nielsen et al., 2010), egg count magnitude has limited diagnostic implications, so there is little added value of using more accurate techniques. Using a more precise technique has the potential to improve the correlation coefficient, while accuracy will only affect the slope of the linear correlation, if found.

3.3. Precision

As mentioned above, precision is a measure of variation between

repeated measures on the same sample (Fig. 1), and is synonymous with repeatability. It is well documented that parasite fecal egg counts are highly variable (Denwood et al., 2012; Carstensen et al., 2013), so it is well justified to document and consider the precision for any given technique.

Broadly spoken, sources of egg count variation can be divided into biological and technical sources. Biological sources include the distribution of ova within the fecal matter and variation within and between samples taken from the same animal. Despite often made claims of the opposite, strongylid egg count variation between defined time points within days or between days has been shown to be either negligible or non-detectable (Carstensen et al., 2013). However, one recognized biological source of variation is a density-dependent fecundity, with female worms apparently suppressing each other's egg production through unrecognized mechanisms (Kotze and Kopp, 2008; Walker et al., 2009). Technical sources include loss of ova during filtration and flotation, mixing and suspending samples, the flotation capacity of ova present, and the training and experience of the analyst reading the sample (Cain et al., 2020, 2021). It should be clear that modifications made to improve egg counting technique precision will primarily address the technical sources of variability.

In general, estimating precision is much more straight-forward than accuracy, as it is generally just a matter of determining repeated counts on the same samples. However, sources of egg count variability can be viewed in a hierarchical manner, with variation existing between different fecal samples from the same animal, different time points of sample collection from these animals within and between days, different subsamples taken from the same fecal samples, and for repeated counts on the same subsamples (Denwood et al., 2012; Carstensen et al., 2013). Thus, a decision needs to be made as to which of these hierarchical levels of variation to address for precision estimation. A strict approach would be to only evaluate the precision of repeated counts on the same subsample and fecal slurry, as this arguably is where most techniques will differ in precision. However, this ignores other sources of variation and does not allow evaluation of the effects of homogenization methods, which have been shown to affect quantitative egg count performance (Went et al., 2018). Furthermore, precision estimates will differ widely depending on which hierarchical levels of variation (sample, subsample, slurry) that are considered in the study design (Cain et al., 2020). Thus, a study evaluating coefficients of variation for repeated counts on the same slurries might report significant differences between techniques, but this may or may not translate to different performance of counts determined on different subsamples. For precision estimates to be meaningful, it is clear that the study design needs due consideration of relevant sources of variability, and better standardization is desired on these aspects.

There are numerous ways to quantify and compare data variability for different techniques. Standard deviation and confidence intervals can be useful, but they will be largely augmented by the multiplication factor (Noel et al., 2017). Instead, the coefficient of variation is independent of the multiplication factor and expresses precision as a percentage, and is often used in studies evaluating fecal egg counting techniques (Bosco et al., 2018; Nápravníková et al., 2019; Daş et al., 2020). The estimation of precision tends to depend on the fecal egg count level with lower precision demonstrated at lower counts (Nápravníková et al., 2019; Daş et al., 2020; Cain et al., 2020, 2021), so it is important to ensure a representation of relevant egg count levels in the samples included. In our recent work, we have adopted a screening procedure to identify a set number of samples for different predetermined egg count level categories, to ensure that the technique is evaluated at a relevant range of counts. We typically define negative, low, moderate, and high egg count categories as determined by a technique not being evaluated in the study, and then allocated equal numbers of samples to each category (Went et al., 2018; Cain et al., 2020, 2021).

In recent years, attention has been drawn to the Poisson distribution describing observed variation of fecal egg counts (Torgerson et al., 2012). Given the sources of variation described above, this Poisson process can be viewed as describing the technical components of egg count variability. As pointed out by Torgerson et al. (2012), scenarios where the data do not follow a Poisson distribution are suggesting excess variation due to other possible sources of variation, such as those associated with processing, mixing, or counting. Adding to this is the additional variability within and between subsamples mentioned above (Denwood et al., 2012; Carstensen et al., 2013) as well as all the other sources of variability mentioned in the previous paragraphs, so it is clear that the Poisson distribution concept for fecal egg counting is theoretical and does not account for these additional sources of variability that also exist.

As a reviewer and editor, I often encounter attempts to evaluate egg counting precision (repeatability) by analyzing for statistical differences of mean counts between replicates. This is not an appropriate approach, as absence of a statistical difference between replicate counts does not suggest good precision. Rather, poor precision will increase data variability, which, in turn, decreases the likelihood of detecting a statistically significant difference. Furthermore, a significant difference between replicates can only be expected if there is either an upwards or downwards trend between them. For precision estimation, where repeated counts are determined on the same samples, such trends should not be expected and this type of analysis is, therefore, not meaningful.

As opposed to accuracy, precision has strong implications for fecal egg counting. For the FECRT, the challenge remains to delineate between random variation and true reduction between pre- and post-treatment counts, and increasing technique precision will help address this. Similarly, classification of animals according to their egg shedding status is complicated by a high level of egg counting variability. For these reasons, precision is of utmost importance in veterinary medicine, and far more relevant than any of the other diagnostic parameters covered in this manuscript. Given this, it is remarkable that several of the novel automated image-analysis based techniques have been published without precision estimates (Lu et al., 2018; Li et al., 2019; Sukas et al., 2019; Elghryani et al., 2020; Inácio et al., 2020; Nagamori et al., 2020; Cringoli et al., 2021). This could reflect the proof-of-concept nature of many of these publications, and precision data may well be provided in subsequent publications. But the fact that precision was not prioritized in these initial publications could also imply that it was not deemed important by study authors. However, this seems counterintuitive given that a main aim with these novel systems is to reduce or eliminate operator error, so it seems highly relevant to document the precision gained with these. Clearly, the veterinary parasitology community needs to have a discussion of which target parameters to evaluate in studies validating fecal egg counting techniques.

3.4. Operator error

With the exception of the recently developed techniques mentioned above, it is important to recognize that parasite fecal egg counts are determined manually by analysts/operators. Thus, any study evaluating the performance of manual techniques does not just quantify the performance of the technique, but rather the performance of the technique in the hands of the given study personnel. Thus, a validation study should not be viewed as a validation of a given technique, but rather as determination of the performance of the technique as employed in the laboratory, in which the study was conducted. Despite this, very few published studies have evaluated the impact of operator proficiency or training on test performance (Slusarewicz et al., 2019; Cain et al., 2021). Given the recent progress with reducing the room for operator error in the counting process, there is substantial reason to consider these sources of variability in validation studies going forward.

4. Conclusions

This review has illustrated several pressing needs in veterinary parasitology research. It is abundantly clear that our field lacks consensus with regards to use of terminology describing diagnostic performance of fecal egg counting techniques. Furthermore, we need better agreement on which diagnostic parameters to evaluate in validation studies, and how to most appropriately conduct such studies. This review has outlined that the inclusion of samples with low egg count levels in the study material will likely negatively affect both precision and sensitivity estimates, and it is important that ethically and scientifically justified decisions are made regarding the composition of the sample material used in a validation study. Given the strong emphasis on parasite fecal egg counting in veterinary parasitology now and in the future, it is highly warranted to address these aspects. The WAAVP has taken leadership in developing guidelines for other essential aspects of veterinary parasitology research, and it seems obvious for this association to take the initiative to develop guidelines for validating the techniques that make up the foundation for so much of our research today. Furthermore, parasite egg count monitoring is widely recommended for veterinary practitioners in the field, further emphasizing the need for uniform standards for diagnostic performance and quality control.

CRediT authorship contribution statement

As the sole author, **M.K. Nielsen:** did all the work on this manuscript.

Declaration of Competing Interest

The author holds stock in MEP Equine Solutions, LLC, which is commercializing an automated image-based parasite egg counting technology.

Acknowledgement

Jamie K. Norris is warmly acknowledged for putting together Fig. 1.

References

Alonzo, T.A., Pepe, M.S., 1999. Using a combination of reference tests to assess the accuracy of a new diagnostic test. Stat. Med. 18, 2987–3003.
Amadesi, A., Bosco, A., Rinaldi, L., Cringoli, G., Claerebout, E., Maurelli, M.P., 2020. Cattle gastrointestinal nematode egg-spiked faecal samples: high recovery rates using the Mini-FLOTAC technique. Parasit. Vectors 13, 230.
Ballweber, L.R., Beugnet, F., Marchiondo, A.A., Payne, P.A., 2014. American Association of Veterinary Parasitologists' review of veterinary fecal flotation methods and factors influencing their accuracy and use—Is there really one best technique? Vet. Parasitol. 204, 73–80.
Bass, C.C., 1909. Mild *Uncinaria* infections. Arch. Intern Med. 3, 446–450.
Bisset, S.A., Vlassoff, A., Douch, P.G.C., Jonas, W.E., West, C.J., Green, R.S., 1996. Nematode burdens and immunological responses following natural challenge in Romney lambs selectively bred for low or high faecal worm egg count. Vet. Parasitol. 61, 249–263.
Bøgh, H.O., Eriksen, L., Lawson, L.G., Lind, P., 1994. Evaluation of an enzyme-linked-immunosorbent-assay and a histamine-release test system for the detection of pigs naturally infected with *Ascaris suum*. Prev. Vet. Med. 21, 201–214.
Bortoluzzi, C., Paras, K.L., Applegate, T.J., Verocai, G.G., 2018. Comparison between McMaster and Mini-FLOTAC methods for the enumeration of *Eimeria maxima* oocysts in poultry excreta. Vet. Parasitol. 254, 21–25.
Bosco, A., Maurelli, M.P., Ianniello, D., Morgoglione, M.E., Amadesi, A., Coles, G.C., Cringoli, G., Rinaldi, L., 2018. The recovery of added nematode eggs from horse and sheep faeces by three methods. BMC Vet. Res. 14, 7.
Byrne, R.L., Fogarty, U., Mooney, A., Marples, N.M., Holland, C.V., 2018. A comparison of helminth infections as assessed through coprological analysis and adult worm burdens in a wild host. Int. J. Parasitol. Parasites Wildl. 7, 439–444.
Cabaret, J., Gasnier, N., Jacquiet, P., 1998. Faecal egg counts are representative of digestive-tract strongyle worm burdens in sheep and goats. Parasite 5, 137–142.
Cain, J.L., Slusarewicz, P., Rutledge, M.H., McVey, M.R., Wielgus, K.M., Zynda, H.M., Wehling, L.M., Scare, J.A., Steuer, A.E., Nielsen, M.K., 2020. Diagnostic performance of McMaster, Wisconsin, and automated egg counting techniques for enumeration of equine strongyle eggs in fecal samples. Vet. Parasitol. 284, 109199.
Cain, J.L., Peters, K.T., Suri, P., Roher, A., Rutledge, M.H., Nielsen, M.K., 2021. The effect of analyst training on fecal egg counting variability. Parasitol. Res. 120, 1363–1370.
Carstensen, H., Larsen, L., Ritz, C., Nielsen, M.K., 2013. Daily variability of strongyle fecal egg counts in horses. J. Equine Vet. Sci. 33, 161–164.
Castro, L.L.Dd., Abrahão, C.L.H., Buzatti, A., Molento, M.B., Bastianetto, E., Rodrigues, D.S., Lopes, L.B., Silva, M.X., de Freitas, M.G., Conde, M.H., de Almeida Borges, F., 2017. Comparison of McMaster and Mini-FLOTAC fecal egg counting techniques in cattle and horses. Vet. Parasitol. Reg. Stud. Rep. 10, 132–135.
Chapman, M.R., Kearney, M.T., Klei, T.R., 2003. Equine cyathostome populations: accuracy of species composition estimations. Vet. Parasitol. 116, 15–21.
Coadwell, W.J., Ward, P.F.V., 1982. The use of faecal egg counts for estimating worm burdens in sheep infected with *Haemonchus contortus*. Parasitology 85, 251–256.
Cox, D.D., Todd, A.C., 1962. Survey of gastrointestinal parasitism in Wisconsin dairy cattle. J. Am. Vet. Med. Assoc. 141, 706–709.
Cringoli, G., Rinaldi, L., Maurelli, M.P., Utzinger, J., 2010. FLOTAC: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. Nat. Protoc. 5, 503–515.
Cringoli, G., Maurelli, M.P., Levecke, B., Bosco, A., Vercruysse, J., Utzinger, J., Rinaldi, L., 2017. The Mini-FLOTAC technique for the diagnosis of helminth and protozoan infections in human and animals. Nat. Protoc. 12, 1723–1732.
Cringoli, G., Amadesi, A., Maurelli, M.P., Celano, B., Piantadosi, G., Bosco, A., Ciuca, L., Cesarelli, M., Bifulco, P., Montresor, A., Rinaldi, L., 2021. The Kubic FLOTAC Microscope (KFM): a new compact digital microscope for helminth egg counts. Parasitology 48, 427–434.
Daş, G., Hennies, M., Sohnrey, B., Rahimian, S., Wongrak, K., Stehr, M., Gauly, M., 2017. A comprehensive evaluation of an ELISA for the diagnosis of the two most common ascarids in chickens using plasma or egg yolks. Parasit. Vectors 10, 187.
Daş, G., Klauser, S., Stehr, M., Tuchscherer, A., Cornelia, C., Metges, C.C., 2020. Accuracy and precision of McMaster and Mini-FLOTAC egg counting techniques using egg-spiked faeces of chickens and two different flotation fluids. Vet. Parasitol. 283, 109158.
Dendukuri, N., Schiller, I., de Groot, J., Libman, M., Moons, K., Reitsma, J., van Smeden, M., 2018. Concerns about composite reference standards in diagnostic research. Brit. Med. J. 360, j5779.
Denwood, M.J., Love, S., Innocent, G.T., Matthews, L., McKendrick, I.J., Hillary, N., Smith, A., Reid, S.W.J., 2012. Quantifying the sources of variability in equine faecal egg counts: implications for improving the utility of the method. Vet. Parasitol. 188, 120–126.
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.
Dryden, M.W., Payne, P.A., Ridley, R., Smith, V., 2005. Comparison of common fecal flotation techniques for the recovery of parasite eggs and oocysts. Vet. Ther. 6, 15–28.
Egwang, T.G., Slocombe, J.O., 1982. Evaluation of the Cornell-Wisconsin centrifugal flotation technique for recovering trichostrongylid eggs from bovine feces. Can. J. Comp. Med. 46, 133–137.
Elghryani, N., Crispell, J., Ebrahimi, R., Krivoruchko, M., Lobaskin, V., McOwan, T., O'Connor, W., Power, E., Voisin, B., Scholz, D., de Waal, T., 2020. Preliminary evaluation of a novel, fully automated, Telenostic device for rapid field-diagnosis of cattle parasites. Parasitology 24, 1–5.
Gasbarre, L.C., 2014. Anthelmintic resistance in cattle nematodes in the US. Vet. Parasitol. 204, 3–11.
George, M.M., Vatta, A.F., Howell, S.B., Storey, B.E., McCoy, C.J., Wolstenholme, A.J., Redman, E.M., Gilleard, J.S., Kaplan, R.M., 2021. Evaluation of changes in drug susceptibility and population genetic structure in *Haemonchus contortus* following worm replacement as a means to reverse the impact of multiple-anthelmintic resistance on a sheep farm. Int. J. Parasitol. Drug. Drug Resist. 15, 134–143.
Gordon, H.M., Whitlock, H.V., 1939. A new technique for counting nematode eggs in sheep faeces. J. Counc. Sci. Ind. Res. 12, 52.
Gowane, G.R., Swarnkar, C.P., Misra, S.S., Kumar, R., Kumar, A., Singh, D., Prince, L.L.L., 2020. Selecting sheep for *Haemonchus contortus* resistance and susceptibility: flock dynamics and genetic architecture. Res. Vet. Sci. 132, 116–126.
Habtamu, K., Degarege, A., Ye-Ebiyo, Y., Erko, B., 2011. Comparison of the Kato-Katz and FLOTAC techniques for the diagnosis of soil-transmitted helminth infections. Parasitol. Int. 60, 398–402.
Howell, S.B., Burke, J.M., Miller, J.E., Terrill, T.H., Valencia, E., Williams, M.J., Williamson, L.H., Zajac, A.M., Kaplan, R.M., 2008. Prevalence of anthelmintic resistance on sheep and goat farms in the southeastern United States. J. Am. Vet. Med. Assoc. 233, 1913–1919.
Inácio, S.V., Gomes, J.F., Falcão, A.X., Suzuki, C.T.N., Nagata, W.B., Loiola, S.H.N., dos Santos, B.M., Soares, F.A., Rosa, S.L., Baptista, C.B., Alves, G.B., Bresciani, K.D.S., 2020. Automated diagnosis of canine gastrointestinal parasites using image analysis. Pathogens 9, 139.
Jackson, F., 1974. New technique for obtaining nematode ova from sheep faeces. Lab. Pract. 23, 65–66.
Kaplan, R.M., 2004. Drug resistance in nematodes of veterinary importance: a status report. Trends Parasitol. 20, 477–481.
Kaplan, R.M., 2013. Recommendations for control of gastrointestinal nematode parasites in small ruminants: these ain't your father's parasites. Bovine Pract. 47, 97–109.
Kaplan, R.M., Nielsen, M.K., 2010. An evidence-based approach to equine parasite control: it ain't the 60s anymore. Equine Vet. Educ. 22, 306–316.
Kenyon, F., Jackson, F., 2012. Targeted flock/herd and individual ruminant treatment approaches. Vet. Parasitol. 186, 10–17.
Kenyon, F., Greer, A.W., Coles, G.C., Cringoli, G., Papadopoulos, E., Cabaret, J., Berrag, B., Varady, M., Van Wyk, J.A., Thomas, E., Vercruysse, J., Jackson, F., 2009. The role of targeted selective treatments in the development of refugia-based

approaches to the control of gastrointestinal nematodes of small ruminants. Vet. Parasitol. 164, 3–11.
Kotze, A.C., Kopp, S.R., 2008. The potential impact of density dependent fecundity on the use of the faecal egg count reduction test for detecting drug resistance in human hookworms. PLoS Negl. Trop. Dis. 2, e297.
Kuzmina, T.A., Kharchenko, V.A., Starovir, G.M., Dvojnos, G.M., 2005. Analysis of the strongylid nematodes (Nematoda: Strongylidae) community after deworming of brood horses in Ukraine. Vet. Parasitol. 131, 283–290.
Kuzmina, T.A., Kornas, S., Basiaga, M., Kharchenko, V.A., Vyniarska, A., 2011. Biodiversity of strongylids (Nematoda: Strongylidae) communities in domestic horses from Poland and Ukraine. Helminthologia 48, 77–84.
Lane, C., 1928. The mass diagnosis of hookworm infection. Am. J. Hyg. (May), 1–148.
Leeflang, M.M.G., Rutjes, A.W.S., Reitsma, J.B., Hooft, L., Bossuyt, P.M.M., 2013. Variation of a test's sensitivity and specificity with disease prevalence. Can. Med. Assoc. J. 185, E537–E544.
Levecke, B., Rinaldi, L., Charlier, J., Maurelli, M.P., Morgoglione, M.E., Vercruysse, J., Cringoli, G., 2011. Monitoring drug efficacy against gastrointestinal nematodes when faecal egg counts are low: do the analytic sensitivity and the formula matter? Parasitol. Res. 109, 953–957.
Levecke, B., Rinaldi, L., Charlier, J., Maurelli, M.P., Bosco, A., Vercruysse, J., Cringoli, G., 2012. The bias, accuracy and precision of faecal egg count reduction test results in cattle using McMaster, Cornell-Wisconsin and FLOTAC egg counting methods. Vet. Parasitol. 188, 194–199.
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 fecal egg count reduction test? Vet. Parasitol. 15, 71–78.
Li, Y., Zheng, R., Wu, Y., Chu, K., Xu, Q., Sun, M., Smith, Z.J., 2019. A low-cost, automated parasite diagnostic system via a portable, robotic microscope and deep learning. J. Biophotonics 12, e201800410.
Lu, Q., Liu, G., Xiao, C., Hu, C., Zhang, S., Xu, R.X., Chu, K., Xu, Q., Smith, Z.J., 2018. A modular, open-source, slide-scanning microscope for diagnostic applications in resource-constrained settings. PLoS One 13 (3), 15.
Matthews, J.B., 2014. Anthelmintic resistance in equine nematodes. Int. J. Parasitol. Drugs Drug. Resist. 4, 310–315.
Mes, T.H.M., Eysker, M., Ploeger, H.W., 2007. A simple, robust and semi-automated parasite egg isolation protocol. Nat. Prot. 2, 486–489.
Moser, W., Bärenbold, O., Mirams, G.J., Cools, P., Vlaminck, J., Ali, S.M., Ame, S.M., Hattendorf, J., Vounatsou, P., Levecke, B., Keiser, J., 2018. Diagnostic comparison between FECPAK^{G2} and the Kato-Katz method for analyzing soil-transmitted helminth eggs in stool. PLoS Negl. Trop. Dis. 12, e0006562.
Nagamori, Y., Hall Sedlak, R., DeRosa, A., Pullins, A., Cree, T., Loenser, M., Larson, B.S., Smith, R.B., Goldstein, R., 2020. Evaluation of the VETSCAN IMAGYST: an in-clinic canine and feline fecal parasite detection system integrated with a deep learning algorithm. Parasit. Vectors 13, 346.
Nápravníková, J., Petrtyl, M., Stupka, R., Vadlejch, J., 2019. Reliability of three common fecal egg counting techniques for detecting strongylid and ascarid infections in horses. Vet. Parasitol. 272, 53–57.
Nielsen, M.K., Baptiste, K.E., Tolliver, S.C., Collins, S.S., Lyons, E.T., 2010. Analysis of multiyear studies in horses in Kentucky to ascertain whether counts of eggs and larvae per gram of feces are reliable indicators of numbers of strongyles and ascarids present. Vet. Parasitol. 174, 77–84.
Nielsen, M.K., Pfister, K., Samson-Himmelstjerna, G.v., 2014. Selective therapy in equine parasite control - application and limitations. Vet. Parasitol. 202, 95–103.
Nielsen, M.K., Mittel, L., Grice, A., Erskine, M., Graves, E., Vaala, W., Tully, R.C., French, D.D., Bowman, R., Kaplan, R.M., 2019. AAEP Parasite Control Guidelines. https://aaep.org/sites/default/files/Guidelines/AAEPParasiteControlGuidelines_0.pdf.
Nielsen, M.K., Banahan, M., Kaplan, R.M., 2020. Importation of macrocyclic lactone resistant cyathostomins on a US Thoroughbred farm. Int. J. Parasitol. Drugs Drug Resist. 14, 99–104.
Nilsson, O., Ljungström, B.L., Höglund, J., Lundquist, H., Uggla, A., 1995. *Anoplocephala perfoliata* in horses in Sweden: prevalence, infection levels and intestinal lesions. Acta Vet. Scand. 36, 319–328.
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.
Paras, K.L., George, M.M., Vidyashankar, A.N., Kaplan, R.M., 2018. Comparison of fecal egg counting methods in four livestock species. Vet. Parasitol. 257, 21–27.
Peregrine, A.S., Molento, M.B., Kaplan, R.M., Nielsen, M.K., 2014. Anthelmintic resistance in important parasites of horses: does it really matter? Vet. Parasitol. 201, 1–8.
Ploeger, H.W., Everts, R.R., 2018. Alarming levels of anthelmintic resistance against gastrointestinal nematodes in sheep in the Netherlands. Vet. Parasitol. 262, 11–15.
Presland, S.L., Morgan, E.R., Coles, G.C., 2005. Counting nematode eggs in equine faecal samples. Vet. Rec. 156, 208–210.
Proudman, C.J., Edwards, G.B., 1992. Validation of a centrifugation/flotation technique for the diagnosis of equine cestodiasis. Vet. Rec. 131, 71–72.
Rashid, M.H., Stevenson, M.A., Waenga, S., Mirams, G., Campbell, A.J.D., Vaughan, J.L., Jabbar, A., 2018. Comparison of McMaster and FECPAK^{G2} methods for counting nematode eggs in the faeces of alpacas. Parasit. Vectors 2, 278.
Rehbein, S., Lindner, T., Kollmannsberger, M., Visser, M., Winter, R., 1997. Untersuchungen zum Helminthenbefall von Schlachtschafen in Oberbayern 2. Mitt.: Beziehungen zwischen der Eiausscheidung und der Wurmbürde [German]. Berl. Munch. Tierarztl. Wochenschr. 110, 148–152.
Reitsma, J.B., Rutjes, A.W.S., Khan, K.S., Coomarasamy, A., Bossuyt, P.M., 2009. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. J. Clin. Epidemiol. 62, 797–806.
Rinaldi, L., Amadesi, A., Dufourd, E., Bosco, A., Gadanho, M., Lehebel, A., Maurelli, M.P., Chauvin, A., Charlier, J., Cringoli, G., Ravinet, N., Chartier, C., 2019. Rapid assessment of faecal egg count and faecal egg count reduction through composite sampling in cattle. Parasit. Vectors 12, 353.
Roberts, J.L., Swan, R.A., 1981. Quantitative studies of ovine haemonchosis. I. Relationship between faecal egg counts and total worm counts. Vet. Parasitol. 8, 165–171.
Saah, A.J., Hoover, D.R., 1997. "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. Ann. Intern. Med. 126, 91–94.
Samson-Himmelstjerna, G.v., 2012. Anthelmintic resistance in equine parasites – detection, potential clinical relevance and implications for control. Vet. Parasitol. 185, 2–8.
Scare, J.A., Slusarewicz, P., Noel, M.L., Wielgus, K.M., Nielsen, M.K., 2017. Evaluation of accuracy and precision of a smartphone based automated parasite egg counting system in comparison to the McMaster and Mini-FLOTAC method. Vet. Parasitol. 247, 85–92.
Skotarek, S.L., Colwell, D.D., Goater, C.P., 2010. Evaluation of diagnostic techniques for *Anoplocephala perfoliata* in horses from Alberta, Canada. Vet. Parasitol. 172, 249–255.
Slusarewicz, M., Slusarewicz, P., Nielsen, M.K., 2019. The effect of counting duration on quantitative fecal egg count test performance. Vet. Parasitol. X 2, 100020.
Sommer, C., 1996. Digital image analysis and identification of eggs from bovine parasitic nematodes. J. Helminthol. 70, 143–171.
Stoll, N.R., 1930. On methods of counting nematode ova in sheep dung. Parasitology 22, 116–136.
Sukas, S., Van Dorst, B., Kryj, A., Lagatie, O., De Malsche, W., Stuyver, L.J., 2019. Development of a Lab-on-a-Disk platform with digital imaging for identification and counting of parasite eggs in human and animal stool. Micromachines 10, 852.
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.
Tyson, F., Dalesman, S., Brophy, P.M., Morphew, R.M., 2020. Novel equine faecal egg diagnostics: validation of the FECPAKG2. Animals 10, 1254.
Walker, M., Hall, A., Anderson, R.M., Basáñez, M.-G., 2009. Density-dependent effects on the weight of female *Ascaris lumbricoides* infections of humans and its impact on patterns of egg production. Parasit. Vectors 2, 11.
Went, H.A., Scare, J.A., Steuer, A.E., Nielsen, M.K., 2018. Effects of homogenizing methods on accuracy and precision of equine strongylid egg counts. Vet. Parasitol. 261, 91–95.