Veterinary Parasitology 284 (2020) 109199

Contents lists available at ScienceDirect

Veterinary Parasitology

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

Research paper

Diagnostic performance of McMaster, Wisconsin, and automated egg counting techniques for enumeration of equine strongyle eggs in fecal samples

Jennifer L. Cain[a,*], Paul Slusarewicz[a,b], Matthew H. Rutledge[d], Morgan R. McVey[a], Kayla M. Wielgus[c], Haley M. Zynda[a], Libby M. Wehling[a], Jessica A. Scare[a], Ashley E. Steuer[a], Martin K. Nielsen[a]

[a] M.H. Gluck Equine Research Center, Department of Veterinary Science, University of Kentucky, Lexington, KY, USA
[b] MEP Equine Solutions, 3905 English Oak Circle, Lexington, KY, USA
[c] College of Veterinary Medicine, Lincoln Memorial University, Harrogate, TN, USA
[d] Department of Statistics, University of Kentucky, Lexington, KY, USA

ARTICLE INFO

Keywords:
Horse
Fecal egg count
McMaster
Automated
Wisconsin
Strongyle

ABSTRACT

Fecal egg counts are the cornerstone of equine parasite control programs. Previous work led to the development of an automated, image-analysis-based parasite egg counting system. The system has been further developed to include an automated reagent dispenser unit and a custom camera (CC) unit that generates higher resolution images, as well as a particle shape analysis (PSA) algorithm and machine learning (ML) algorithm. The first aim of this study was to conduct a comprehensive comparison of method precision between the original smartphone (SP) unit with the PSA algorithm, CC/PSA, CC/ML, and the traditional McMaster (MM) and Wisconsin (MW) manual techniques. Additionally, a Bayesian analysis was performed to estimate and compare sensitivity and specificity of all five methods. Feces were collected from horses, screened with triplicate Mini-FLOTAC counts, and placed into five categories: negative (no eggs seen), > 0 - ≤ 200 eggs per gram (EPG), > 200 - ≤ 500 EPG, > 500 - ≤ 1000 EPG, and > 1000 EPG. Ten replicates per horse were analyzed for each technique. Technical variability for samples > 200 EPG was significantly higher for MM than CC/PSA and CC/ML ($p < 0.0001$). Biological variability for samples > 0 was numerically highest for CC/PSA, but with samples > 200 EPG, MM had a significantly lower CV than MW ($p = 0.001$), MW had a significantly lower CV than CC/PSA ($p < 0.0001$), CC/ML had a significantly lower CV than both MW and SP/PSA ($p < 0.0001$, $p = 0.0003$), and CC/PSA had a significantly lower CV than CC/SP ($p = 0.0115$). Sensitivity was > 98 % for all five methods with no significant differences. Specificity, however, was significantly the highest for CC/PSA, followed numerically by SP/PSA, MM, CC/ML, and finally MW. Overall, the automated counting system is a promising new development in equine parasitology. Continued refinement to the counting algorithms will help improve precision and specificity, while additional research in areas such as egg loss, analyst variability at the counting step, and accuracy will help create a complete picture of its impact as a new fecal egg count method.

1. Introduction

Cyathostomins, commonly known as small strongyles, are found in horses worldwide with a prevalence of nearly 100 % and represent the main parasite of concern for modern equine anthelmintic treatment programs (Lyons et al., 1999; ESCCAP, 2019; Nielsen et al., 2019; Rendle et al., 2019). Cyathostomins can cause larval cyathostominosis when large numbers of L_4 larvae excyst from the mucosa of the large intestine, causing symptoms such as sudden onset diarrhea, weight loss, subcutaneous edema, fever, and colic, with a 50 % mortality rate (Love et al., 1999; Reid et al., 1995). Anthelmintic resistance is now rampant, and only one drug class out of three available for horses, the macrocyclic lactones, are still effective against cyathostomins, albeit with emerging resistance (Traversa et al., 2009; Peregrine et al., 2014; Bellaw et al., 2018).

Modern equine anthelmintic treatment programs utilize

* Corresponding author at: M.H. Gluck Equine Research Center, Department of Veterinary Science, University of Kentucky, 1400 Nicholasville Rd., Lexington, KY 40546-0099, USA.
E-mail address: jennifer.cain@uky.edu (J.L. Cain).

https://doi.org/10.1016/j.vetpar.2020.109199
Received 10 June 2020; Received in revised form 3 August 2020; Accepted 4 August 2020

surveillance-based methods, where only horses contributing the highest numbers of parasite eggs to pasture, and therefore increasing infection pressure, are treated with an anthelmintic drug (ESCCAP, 2019; Nielsen et al., 2019; Rendle et al., 2019). Fecal egg counts (FECs) are used in surveillance-based programs to identify high strongyle shedders, and in fecal egg count reduction tests (FECRT) determining treatment efficacy. The American Association of Equine Practitioners (AAEP) currently recommends two FEC methods for determining the number of strongyle parasite eggs in feces: a modified McMaster (MM) and modified Wisconsin technique (MW) (Nielsen et al., 2019). Both techniques require specialized equipment, including floatation solution with a specific gravity between 1.2 and 1.35, a microscope, and either a specialized counting slide or microscope slides (Cringoli et al., 2004; Nielsen et al., 2019). Additionally, the MW technique requires a centrifuge, which is not only inaccessible for people outside of a laboratory, but also greatly increases processing time (Ballweber et al., 2014). Reading the slides necessitates specialized training on identifying parasite eggs and discriminating them from fecal debris, in addition to training on general microscopy.

Despite the dire situation of anthelmintic resistance in cyathostomins, convincing horse owners to use FECs for their anthelmintic control programs has proven difficult. The number of owners and farms using FECs ranges from 10.6 %–30.6 % according to studies in the United States (Robert et al., 2015; Nielsen et al., 2018; Scare et al., 2018; Cain et al., 2019), and a multinational study including the United States, Germany, the Netherlands, Austria and Denmark found a similar trend (Becher et al., 2018). A more recent study in the United Kingdom showed a more optimistic trend, with 60.9 % of respondents using FECs (Tzelos et al., 2019). While education can help encourage people to use FECs more frequently (Cain et al., 2019), performing the counts can be time-consuming, results may be unreliable and variable depending on the operator, and recommended methods require equipment typically only found in scientific or veterinary laboratories (Vidyashankar et al., 2012).

Diagnostic methods can be evaluated for the parameters of sensitivity, specificity, precision and accuracy. Sensitivity is the ability of a test to detect true positives, while specificity is the ability to detect true negatives. Precision measures a diagnostic method's ability to repeatedly produce the same quantitative result for a single sample, whereas accuracy reflects its ability to determine the true value of the relevant parameter in a sample. There are numerous variables that can affect precision in FECs, including biological ones, such as: uneven parasite egg distribution throughout a fecal sample; parasite species composition within the animal; and the number of internal parasites present, which may affect egg production (Denwood et al., 2012; Vidyashankar et al., 2012). Technical variables add additional variability to FEC analyses, and these include: egg distribution in, and subsampling of, the fecal suspension; egg loss in the sample preparation process; and variation in FEC performance due to user error, fatigue, skill level and/or subjectivity (Vidyashankar et al., 2012).

Automated systems that analyze images of parasite eggs are a promising development in the realm of parasitology, providing an opportunity to greatly reduce or eliminate operator variability, decrease processing time, and minimize the required training (Yang et al., 2001; Mes et al., 2007; Suzuki et al., 2013; Li et al., 2019). An automated system was recently developed that uses a machine to chemically process a fecal sample and produce fluorescently stained parasite eggs *via* attachment of a fluorophore to chitin present in the eggshell. The stained eggs can then be photographed by a smartphone (SP) incorporated into an imaging unit (IU). The SP then computationally analyzes the image with an algorithm based on particle shape analysis (PSA) to produce FECs (Slusarewicz et al., 2016). The original device utilized an 8-megapixel (MP) iPhone, and previous research on this version indicated that the system performed well with regards to precision and accuracy compared to the MM method (Scare et al., 2017). The imaging unit has evolved since these initial studies, firstly by incorporating an SP with a higher-resolution (18.1 M P) camera and, more recently, a custom-built camera (CC) with an 18.1 M P sensor.

In this study we evaluated the performance of both the more recent SP-based IU, also running a PSA-based algorithm, as well as the new CC unit running either PSA- or machine learning (ML)-based algorithms, and compared them with the two different manual FEC methods currently recommended by the AAEP (*i.e.,* MM and MW). The aims of this study were to: [1] analyze and compare the technical variability for MM, CC/PSA and CC/ML; [2] analyze and compare the biological variability for MM, MW, SP/PSA, CC/PSA, and CC/ML; and [3] determine and compare the sensitivity and specificity of all five methods.

2. Materials and methods

2.1. Horses

All fecal samples were collected from the University of Kentucky Parasitology Research Herds in accordance with IACUC protocol 2012−1046. One herd consists of mixed light breed horses with eighteen mares and one stallion ranging in age from six to seventeen years (Lyons et al., 1990). The other herd is made up of miniature horses with 20 mares and one stallion ranging in age from 5 to 23 years (Lyons et al., 2001). Three samples were collected from another research herd, made up of ten mares, two weeks post-anthelmintic treatment, to obtain negative samples. All animals were kept in fields exclusively used for horses, and samples were either collected from the ground in a clean stall immediately after a horse defecated, or rectally. Fecal samples were collected in a plastic bag, sealed, and stored in a refrigerator at 4 °C. A total of 35 fecal samples were collected between March and November 2018.

2.2. Cameras and algorithms

2.2.1. Sony experia camera and algorithm

This intermediate imaging unit prototype (SP/PSA) contained an Android-based Xperia Z3 smartphone (Sony Corporation, Tokyo, Japan) fitted with a macro lens and emission filter as well as a bank of blue LEDs to provide an excitative light source. The processed sample was placed into the unit and the phone was used to take the image and process the data using a custom application. The algorithm utilized a particle shape-recognition method that analyzed the parameters of particle aspect ratio, surface area and convexity to determine whether any given particle was an ovum and was based on the public domain OpenCV computer vision library (www.opencv.org). The automated system, including all camera variants, were provided by the manufacturer (MEP Equine Solutions, Lexington, Kentucky, USA).

2.2.2. CC camera and algorithms

The CC camera was custom-built and based on a UI-3591LE sensor (IDS Imaging Development Systems, Germany). The image taken by the unit was wirelessly transferred to a Galaxy S2 tablet (Samsung, Seoul, Republic of Korea) running Android, which was loaded with an application that subsequently performed the analysis. Two algorithms were used in conjunction with this unit: one that used the same PSA analysis as the SP unit with modified parameter values to account for differences in the sensor sizes of the two units (CC/PSA), and one that used a model developed from training a neural net with a training set of images of ova and non-ova particles using ML (CC/ML).

2.3. Sample screening

Fecal samples were screened in triplicate using Mini-FLOTAC (Noel et al., 2017) to determine strongyle egg level and subsequently separated into negative (no eggs seen), > 0 - ≤ 200 EPG, > 200 - ≤ 500 EPG, > 500 - ≤ 1000 EPG, and > 1000 EPG categories (Nielsen et al., 2019). These levels were used throughout the study.

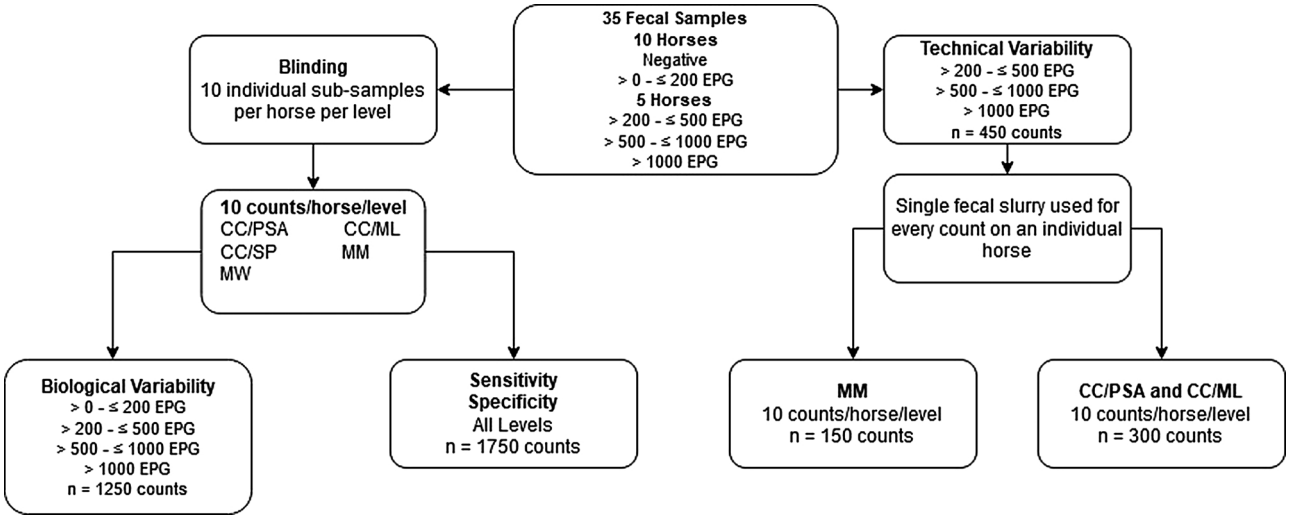

Fig. 1. Flow chart illustrating methodology for data collection. Total biological variability counts n = 1250 and total technical variability counts n = 450. MM = McMaster, MW = Wisconsin, CC = custom camera, SP = smartphone, PSA = particle size analysis algorithm, ML = machine learning algorithm.

2.4. Technical variability

Fecal samples from five mares in each of the following categories were used for this portion of the study: > 200 - ≤ 500 EPG, > 500 - ≤ 1000 EPG, and > 1000 EPG. Glucose-salt floatation solution (specific gravity 1.25) was used and prepared as follows for the duration of the study (Nielsen et al., 2019): 375 g of glucose monohydrate and 250 g of sodium chloride were combined with distilled water to a final volume of 1 L. Two fecal subsamples from each horse were prepared using 6.0 g of feces and 54 mL of floatation medium to produce a dilution factor of 10 × . The subsamples were placed into a separate sample preparation tool (SPT) of the automated system and homogenized (Figure S1). The resulting slurries were then passed through the filters built into the SPT cap, pooled, and used to perform ten MM counts by filling both chambers of a two-chambered McMaster slide (Chalex, Park City, Utah, USA) and counting within the grids, using a multiplication factor (MF) of 33.3. Ten automated counts were performed by pipetting 4 mL of the same pooled fecal suspension onto filter chambers, processed with the automated preparation system, and imaged and counted using the CC camera/ tablet system. Images produced from the CC camera were analyzed using both the PSA and ML algorithms using an MF of 2.5 (Fig. 1). Note that, while the automated system does not require a floatation medium (the manufacturer suggests using tap water), but it was utilized in this study in order to conduct both manual MM and automated counts from the same slurry.

2.5. Biological variability

2.5.1. Blinding

Four trained analysts performing FECs were blinded and were unaware of the screening counts, performed by two different trained analysts, or which horse's FEC they were determining. Plastic bags were filled with 10 g of feces and each was labeled with collection date, a random number, and which counting method was to be performed on that sample. Samples from five horses were used per category for the > 200 - ≤ 500 EPG, > 500 - ≤ 1000 EPG and > 1000 EPG levels, while samples from ten horses were used for the > 0 - ≤ 200 EPG category. Ten subsamples per horse per method were collected and used for this study, as summarized in Fig. 1.

2.5.2. Egg counting methods

FECs were performed using MM (MF 25) and MW (MF 1) methods outlined in the AAEP guidelines (Nielsen et al., 2019) using glucose-salt floatation medium. The passive floatation option, where cover slips are placed on top of 15 mL tubes after centrifugation to allow eggs to float to the top, was used for the MW counts instead of the final centrifugation step as outlined in the AAEP guidelines (Nielsen et al., 2019). Counts for the automated system (MF 2.5) were performed by placing 54 mL of floatation medium and 6.0 g of feces in the SPT, homogenizing, and dispensing 4 mL into the egg chamber following manufacturer's instructions. The same egg chamber was used to perform counts with the SP/PSA, CC/PSA and CC/ML camera systems.

2.6. Statistical analyses

2.6.1. Statistical analysis for precision

The coefficient of variation (CV) was calculated by dividing the standard deviation by the mean count of all the replicates of the same sample and then multiplying by 100 to obtain a percentage. The R^2 values of linear correlations were used for comparisons between methods and *p*-values were calculated using a Pearson correlation. A repeated measures ANOVA analysis with a Tukey post-hoc test was performed to compare CV values between methods. The plot of CV scores and residuals were investigated for approximate normality and homogenous variance, and the assumptions of the repeated measures ANOVA were adequately satisfied. All statistics were performed in R (R Core Team, 2019). The 0.05 α level was used to determine statistical significance.

2.6.2. Bayesian analysis

Sensitivity and specificity were estimated using a modified version of previously described Bayesian methodology (Joseph et al., 1995; Martinez et al., 2008) to account for the fact that each test was conducted multiple times on each horse. Additional counts on negative (no eggs seen) samples were performed ten times for ten different horses and included in the analysis. Posterior means were used to estimate the sensitivity and specificity of each test. Priors for sensitivity and specificity were uninformative (*i.e.*, uniformly distributed) on (0,1) and the prior for prevalence was a Beta (24.3, 9.7) which was based on the results of the mini-FLOTAC screening tests. Analysis was performed in R (R Core Team, 2019) and 95 % credible intervals, obtained from the 2.5 % and 97.5 % quantiles of the posterior distribution, for differences in sensitivity and specificity between methods were used to make pairwise comparisons between methods.

Table 1
Coefficient of variation (CV) and 95 % confidence intervals (CI) results for technical variability.

			Shedding Category CV		
	CV	95 % CI	> 200 - ≤ 500	> 500 - ≤ 1000	> 1000
MM	27.4_a	(20.1 – 26.2)	36.0	25.9	20.4
CC/PSA	11.7_b	(9.3 – 12.9)	12.9	11.4	10.7
CC/ML	10.6_b	(8.3 – 12.0)	11.4	11.0	9.3

All values shown in percentage (%) and different letters indicate a significant difference.
MM = McMaster, CC = custom camera, PSA = particle size analysis algorithm, ML = machine learning algorithm.

3. Results

3.1. Technical variability study

A total of 450 counts were performed for this portion of the study using only the CC imaging unit. The CV and 95 % confidence interval values for technical variability are shown in Table 1. Overall, MM was had a significantly higher CV compared to both CC/PSA and CC/ML ($p < 0.0001$). There was no significant difference in CV between CC/PSA and CC/ML. Counts for all methods were significantly correlated ($p < 0.0001$; Table S1; Fig. 2).

3.2. Biological variability study

A total of 1250 counts were performed for this portion of the study. The CV and 95 % confidence interval values are shown in Table 2. There were no significant differences in CV between the five methods, however MW had the numerically lowest CV of 48.5 % and CC/PSA had the numerically highest CV of 84.8 %. When only the categories of > 200 - ≤ 500 EPG, > 500 - ≤ 1000 EPG, and > 1000 EPG were included in the analysis, there were some notable differences (Table 2). MM had a significantly lower CV than MW ($p = 0.001$), MW had a significantly lower CV than CC/PSA ($p < 0.0001$), CC/ML had a significantly lower CV than both MW and SP/PSA ($p < 0.0001$, $p = 0.0003$), and CC/PSA had a significantly lower CV than CC/SP ($p = 0.0115$). Counts for all methods were significantly correlated ($p < 0.0001$; Table S1), and as can be seen in Fig. 3, MW tended to produce lower counts and SP/PSA tended to produce higher counts compared to the other methods.

3.3. Sensitivity and specificity

A total of 1750 counts were used for Bayesian analysis of sensitivity and specificity, and results are summarized in Fig. 4. All samples were egg count positive at levels > 200 EPG. Sensitivity was similar between all five methods with mean sensitivity ranging from 98.0 % (SP/PSA) to 99.4 % (MW, MM). Mean specificity ranged from 51.0 % (MW) to 90.6 % (CC/PSA). All credible intervals for pairwise comparisons between Sp_CC/PSA - Sp_m (where m is any other method) were greater than 0, indicating that this method had the highest specificity. A summary of credible intervals can be found in Table S2.

4. Discussion

This study demonstrated that the automated egg count approach significantly improves precision over manual methods for samples with egg counts are above 200 EPG. The precision values obtained here are also in good agreement with those of a previous study that compared the precision and accuracy of the CC/ML to MM, when manual counting time for the analyst was restricted (Slusarewicz et al., 2019). Evaluating precision at the lower egg count levels is complicated by the fact that some methods (*i.e.,* the McMaster technique) will return a large proportion of 0 EPG results when fecal egg content is low, which may falsely increase the observed precision. All evaluated methods had similar sensitivity, whereas specificity levels varied by technique.

This is the first study evaluating precision by separating it into biological and technical variability. Previous research has found that a large proportion of biological variability occurs due to egg aggregation within the feces and variation between fecal piles (Denwood et al., 2012; Carstensen et al., 2013). Using a single, homogenous fecal slurry facilitates an analysis of repeatability for a given method, which we dubbed “technical variability”, in the absence of additional variables contributed by separate subsampling a non-homogeneous matrix. As expected, and likely due to biological factors such as uneven distribution of parasite eggs within a sample, precision was lower in the biological variability portion of the study (Denwood et al., 2012; Vidyashankar et al., 2012). Precision generally increased with increasing FEC levels for both biological and technical variability,

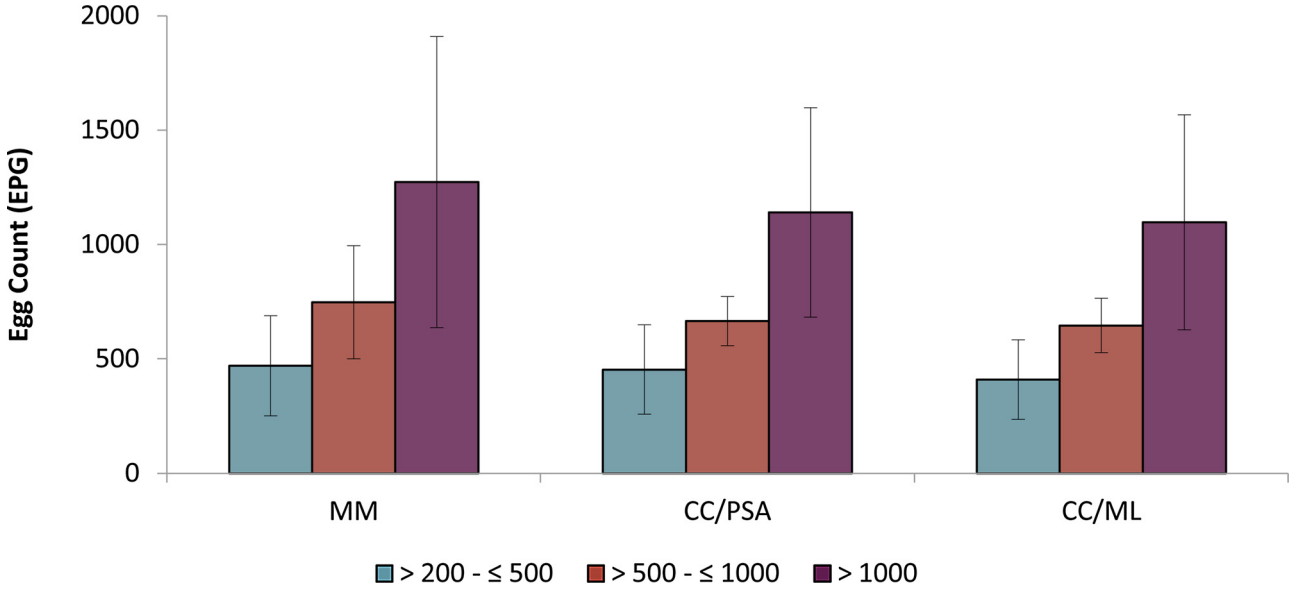

Fig. 2. Mean egg count in eggs per gram (EPG) for technical variability separated by fecal egg count level. Error bars represent standard deviation. MM = McMaster, CC = custom camera, PSA = particle size analysis algorithm, ML = machine learning algorithm.

Table 2
Coefficient of variation (CV) results for biological variability for all levels combined, all levels > 200 eggs per gram (EPG), and CV for individual EPG levels.

	All Levels		Levels > 200		Individual Level CV			
	CV	95 % CI	CV	95 % CI	> 0 - ≤ 200	> 200– ≤ 500	> 500– ≤ 1000	> 1000
MM	58.8	(34.8–82.8)	$23.9_{b,c}$	(19.8 – 28.0)	111.1	26.9	26.2	18.6
MW	48.5	(39.7 – 57.3)	$36.3_{a,b}$	(30.5 – 42.2)	66.7	38.6	32.9	37.5
CC/PSA	84.8	(38.4 – 131.2)	21.0_{c}	(16.6 – 25.9)	180.2	23.2	20.2	20.4
CC/ML	71.2	(38.5 – 104.0)	20.0_{c}	(14.8 – 25.1)	148.0	21.8	20.8	17.2
SP/PSA	66.0	(31.9 – 100.1)	$31.3_{a,b}$	(23.6 – 38.9)	118.1	32.9	27.2	33.9

All values shown in percentage (%) and different letters indicate a significant difference.
MM = McMaster, MW = Wisconsin, CC = custom camera, SP = smartphone, PSA = particle size analysis algorithm, ML = machine learning algorithm.

presumably because high egg densities in the slurries reduced sampling errors (Levecke et al., 2012). The most drastic differences were observed in the biological variability section because it included the lowest FEC category, which had a CV two-to-nine times higher than the next lowest category. This demonstrates that low FEC levels are more affected by the sources of variability contributing to precision, but also emphasizes the importance of systematically ensuring a representation of different FEC levels when estimating precision for a given method (Levecke et al., 2012).

While two automated methods (SP/PSA and CC/PSA) exhibited a lower false positive rate (higher specificity) than MM, the third (CC/ML) did not. Furthermore, MW produced the highest false-positive rate. It is possible that non-ovum chitin-containing objects in the automated counts, such as insect fragments and fungal spores, were the cause of some of the false positive counts, however, this does not explain why false positives were also generated by manual MW counts. The fact that both of the manual methods produced false positives might appear surprising at a first glance, since one might only expect such results when analysts misidentify extraneous particles as ova for reasons such as inadequate training, fatigue, time pressure or lack of motivation. However, this was unlikely to be the case in this study, where such factors were absent. Instead, these observations could be explained by the nature of the Bayesian analysis, which was performed in the absence of a defined gold standard. Thus, test results were classified as false positives based on comparison between all evaluated techniques. Consequently, a method detecting a higher proportion of positives than others in the analysis may have a proportion of these results classified as false-positives, even though they may have been correctly identified. Additionally, the theoretical detection limits of MW, CC/ML and MM are 1, 2.5 and 25 EPG respectively, which also happens to be the order of their respective false positive rates (MW > CC/ML > MM). This again suggests that specificity estimation in this study could indeed

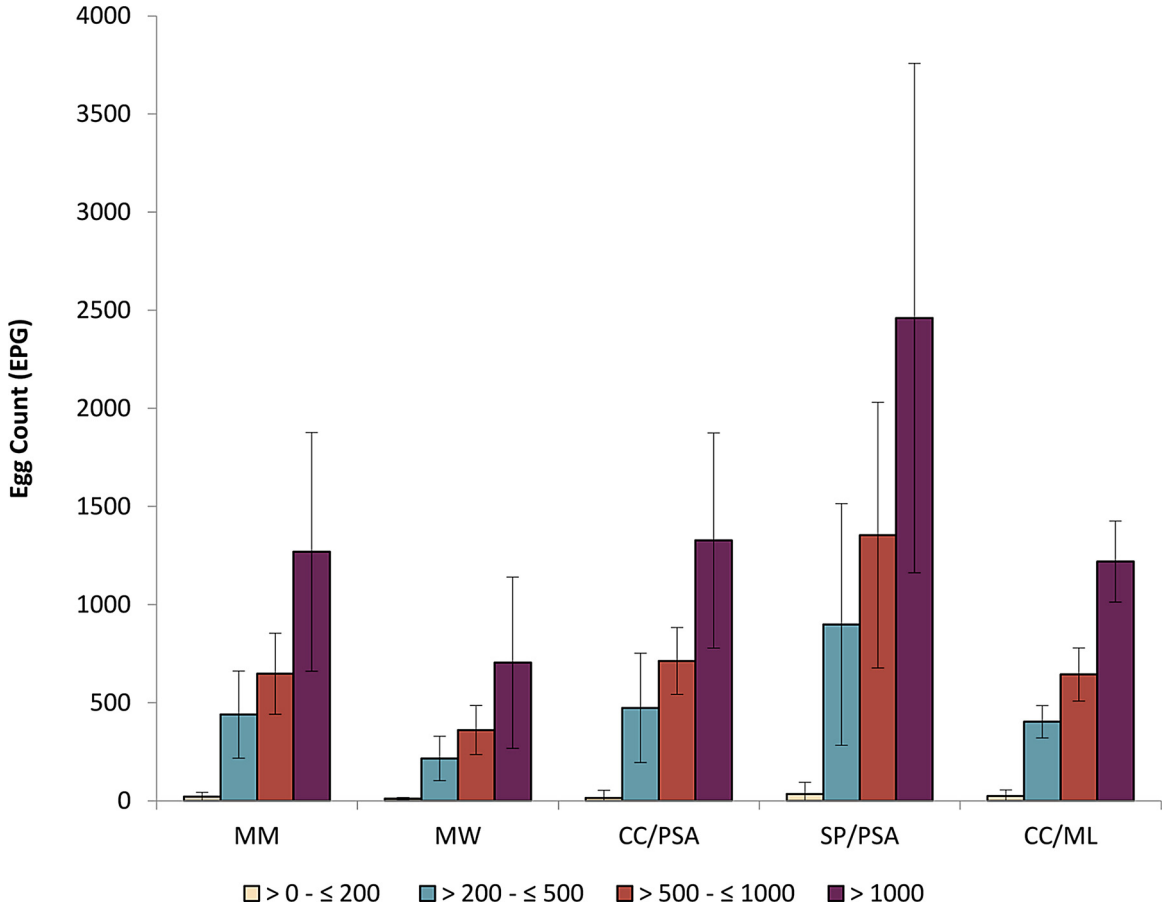

Fig. 3. Mean egg count in eggs per gram (EPG) for biological variability separated by fecal egg count level. Error bars represent standard deviation. MM = McMaster, MW = Wisconsin, CC = custom camera, SP = smartphone, PSA = particle size analysis algorithm, ML = machine learning algorithm.

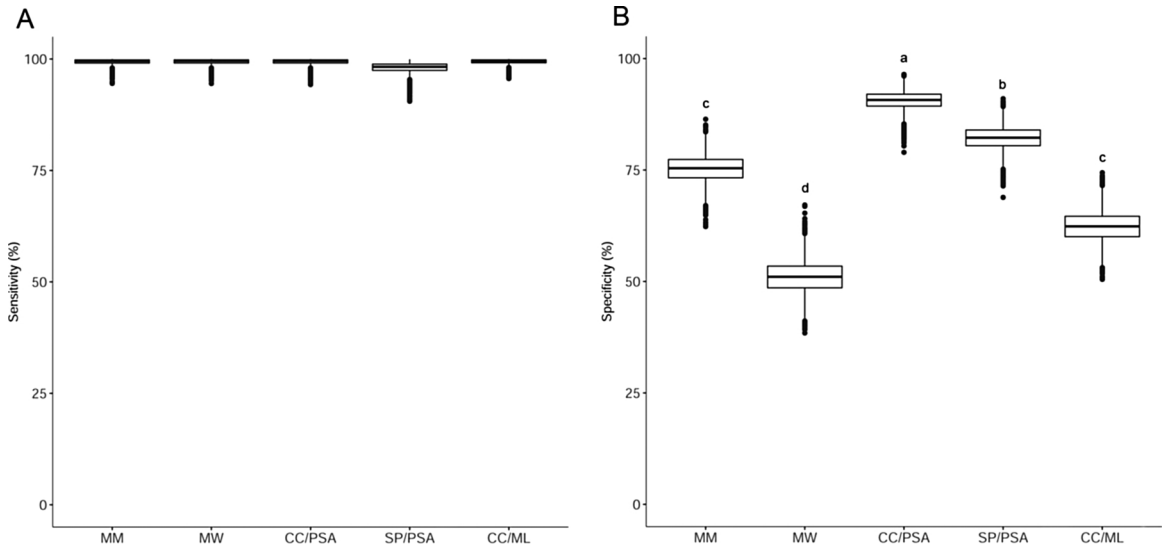

Fig. 4. Bayesian estimates of sensitivity (A) and specificity (B) for all five methods. Different letters indicate a significant difference. MM = McMaster, MW = Wisconsin, CC = custom camera, SP = smartphone, PSA = particle size analysis algorithm, ML = machine learning algorithm.

have been influenced by differential detection of low levels of ova in some tests of samples *i.e.*, that some of the false positive results were, in fact, mis-classed true positives. It was also interesting to note that the two PSA-based automated methods produced fewer false positives than the ML-based system, even though all three systems shared the same 2.5 EPG theoretical detection limit. This suggests that the more sophisticated, machine learning-based algorithm, is better at detecting parasite eggs than the PSA-based algorithm, and that the latter's actual detection limit is higher than its theoretical limit.

The results of this study showed that sensitivity for all methods was similar and high, with a range of only 1.4 % between the highest and lowest sensitivities and with no significant differences. It is interesting to note, however, that sensitivity estimates were not proportional to the vastly different theoretical detection limits (also known as the multiplication factor, MF). This deserves attention, because researchers, parasitologists, and clinicians often assume that all parameters of a test scale linearly with this limit, which is derived arithmetically from the amount of fecal material subjected to analysis in the test. This reasoning leads to the assumption that if a test has a lower MF, it must be more sensitive. Furthermore, it is also assumed that as long as the correct MF is used, the counts from one kind of test must be equivalent to those derived from an entirely different method. The results reported herein clearly demonstrate that this is not necessarily the case. In addition to the non-scalability of the MF with respect to sensitivity, the results presented here with respect to MM and MW clearly show that the latter produces lower counts than the former and that the MFs cannot therefore scale arithmetically with respect to accuracy, an observation that is supported by other studies (Levecke et al., 2012; Bosco et al., 2018). Previous research using spiked samples and an older iteration of the automated method indicated that MM, which was the technique with the highest MF, performed with the lowest sensitivity (Scare et al., 2017). However, a study evaluating performance of different egg counting techniques for counting cattle trichostrongylid eggs, found no difference in diagnostic sensitivity despite MFs ranging from 1 to 10 (Levecke et al., 2012), which is in agreement with our findings.

It is important to bear in mind that the MF is a purely theoretical value assuming homogenous egg distribution in the fecal suspension, zero egg loss in sample preparation, no additional losses during floatation, 100 % adhesion of ova to a coverslip (for MW-type tests), and that subsamples are representative of the entire fecal sample. These assumptions, however, are largely incorrect, which is underscored by the observation here that the CV for MW was significantly inferior to both MM and all three the automated methods in samples of > 200EPG, presumably due to factors such as those described above. Previous research in horses has shown that there is a high amount of variability between FECs from different subsamples (Carstensen et al., 2013). Additionally, FECs are prepared in floatation solution that causes parasite eggs to move upwards rapidly, so analysts must be cognizant of this fact and ensure homogenization of the suspension occurs immediately before subsampling for slide preparation. There are numerous opportunities for egg loss to occur during sample preparation, such as eggs becoming stuck to containers and filtration media and suspension loss between containers (Egwang and Slocombe, 1982). As a result, parasitologists should not use MFs as sole descriptors of the absolute and relative performances of different tests, because the available evidence suggests that the test operating parameters can vary wildly and can only be determined experimentally.

An alternative study design approach to the one we took here would be to use spiked, rather than naturally infected, samples. While spiking a known number of eggs within an egg-free fecal ball may potentially represent a gold standard allowing for an evaluation of true accuracy, this method is not necessarily a better alternative to the naturally infected samples used herein. Spiked samples act as a simulation of samples obtained from infected animals and cannot completely mimic either the natural distribution of eggs throughout fecal matter or their incorporation into (and subsequent release from) the fecal matrix. Recent equine studies using spiked samples reported vastly different accuracy levels for the same methods, EPG levels, and parasite egg type, illustrating that this method may not represent a better alternative study design (Noel et al., 2017; Scare et al., 2017; Bosco et al., 2018; Napravnikova et al., 2019).

Ultimately, FECs are used practically in parasite control programs, where horses above a particular EPG level are treated with anthelmintics (ESCCAP, 2019; Nielsen et al., 2019; Rendle et al., 2019). The most recent automated methods, CC/PSA and CC/ML, are significantly and 2–3 times more precise than traditional methods at > 200 EPG levels with respect to technical variability. As a result, the improved technical precision was most likely mainly a function of in the variability of stochastic sampling from the fecal slurry, since 4 mL was used for each automated count and only 0.3 mL (the volume beneath two McMaster grids) was used for MM. This improved precision was largely negated by the introduction of biological variability, and while the automated counts remained significantly more precise than MW, they were only slightly so in comparison to MM, and these small differences were not significant. Thus, the large degree of biological variation was sufficient to mask the technical superiority of the automated system. It

should be noted that the manual counts were conducted under optimal conditions (*i.e.*, in a parasitology laboratory at a large research university), and many of the potential sources of technical variability mentioned previously (*e.g.*, analyst training, fatigue, time constraints, motivation *etc.*) were likely or be absent or relatively small. We have already shown that reducing counting time by either time restriction or counting only one grid of a McMaster slide further decrease MM precision by 2-fold each relative to CC/ML (Slusarewicz et al., 2019). We are also currently assessing the effect of analyst training on the performance of a range of egg counting methods. In the future, it would be interesting to determine whether these additional real-world technical variables that can affect manual counts would be sufficient to produce significant differences between methods even in the presence of biological variability.

In summary, automated FEC methods that largely remove operator involvement, decrease time for sample processing, and require little training are a promising development in equine parasitology. This study highlights new advancements while also indicating areas for future improvement to these methods. Additional research in areas such as egg loss, analyst variability at the counting step, and accuracy of ML is necessary to fully gauge the impact that automated FEC systems will have on equine parasitology.

CRediT authorship contribution statement

Jennifer L. Cain: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Project administration, Writing - original draft, Writing - review & editing. **Paul Slusarewicz:** Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing - review & editing. **Matthew H. Rutledge:** Formal analysis, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Morgan R. McVey:** Investigation. **Kayla M. Wielgus:** Investigation. **Haley M. Zynda:** Investigation. **Libby M. Wehling:** Investigation. **Jessica A. Scare:** Conceptualization, Methodology. **Ashley E. Steuer:** Conceptualization, Methodology. **Martin K. Nielsen:** Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

Dr. Martin Nielsen and Dr. Paul Slusarewicz both hold stock in MEP Equine Solutions, a company that is manufacturing an automated parasite egg counting technique. Dr. Slusarewicz is an employee of said company. None of the other authors have any conflicts to declare.

Acknowledgements

The authors would like to thank Chris Mills for technical support, and Eric Roemmele for assistance with the statistical analysis.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.vetpar.2020.109199.

References

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.

Becher, A.M., van Doorn, D.C., Pfister, K., Kaplan, R.M., Reist, M., Nielsen, M.K., 2018. Equine parasite control and the role of national legislation – a multinational questionnaire survey. Vet. Parasitol. 259, 6–12.

Bellaw, J.L., Krebs, K., Reinemeyer, C.R., Norris, J.K., Scare, J.A., Pagano, S., Nielsen, M.K., 2018. Anthelmintic therapy of equine cyathostomin nematodes – larvicidal efficacy, egg reappearance period, and drug resistance. Int. J. Parasitol. 48, 97–105.

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.

Cain, J.L., Foulk, D., Jedrzejewski, E., Stofanak, H., Nielsen, M.K., 2019. The importance of anthelmintic efficacy monitoring: results of an outreach effort. Parasitol. Res. 118 (10), 2877–2883.

Carstensen, H., Larsen, L., Ritz, C., Nielsen, M.K., 2013. Daily variability of strongyle fecal egg counts in horses. J. Equine Vet. Sci. 33 (3), 161–164.

Cringoli, G., Rinaldi, L., Veneziano, V., Capelli, G., Scala, A., 2004. The influence of flotation solution, sample dilution and the choice of McMaster slide area (volume) on the reliability of the McMaster technique in estimating the faecal egg counts of gastrointestinal strongyles and *Dicrocoelium dendriticum* in sheep. Vet. Parasitol. 123, 121–131.

Denwood, M.J., Love, S., Innocent, G.T., Matthews, L., McKendrick, I.J., Hillary, N., Smith, A., Reid, S.W., 2012. Quantifying the sources of variability in equine faecal egg counts: implications for improving the utility of the method. Vet. Parasitol. 188, 120–126.

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

ESCCAP, 2019. A Guide to the Treatment and Control of Equine Gastrointestinal Parasite Infections. Online at. 2nd ed. European Scientific Counsel Compantion Animal Parasites. https://www.esccap.org/uploads/docs/rtjqmu6t_0796_ESCCAP_Guideline_GL8_v7_1p.pdf.

Joseph, L., Gyorkos, T.W., Coupal, L., 1995. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am. J. Epidemiol. 141 (3), 263–272.

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. Par. 188 (1-2), 194–199.

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 (9), e201800410.

Love, S., Murphy, D., Mellor, D., 1999. Pathogenicity of cyathostome infection. Vet. Parasitol. 85, 113–122.

Lyons, E.T., Drudge, J.H., Tolliver, S.C., 1990. Prevalence of some internal parasites found (1971–1989) in horses born on a farm in Central Kentucky. J. Equine Vet. Sci. 10, 99–107.

Lyons, E.T., Tolliver, S.C., Drudge, J.H., 1999. Historical perspective of cyathostomes: prevalence, treatment and control programs. Vet. Parasitol. 85, 97–112.

Lyons, E.T., Tolliver, S.C., Drudge, J.H., Collins, S.S., Swerczek, T.W., 2001. Continuance of studies on Population S benzimidazole-resistant small strongyles in a Shetland pony herd in Kentucky: effect of pyrantel pamoate (1992–1999). Vet. Parasitol. 94, 247–256.

Martinez, E.Z., Louzada-Neto, F., Derchain, S.F., Achcar, J.A., Gontijo, R.C., Sarian, L.O., Syrjänen, K.J., 2008. Bayesian estimation of performance measures of cervical cancer screening tests in the presence of covariates and absence of a gold standard. Cancer Inform. 6, 33–46.

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.

Napravnikova, 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., Branan, M.A., Wiedenheft, A.M., Diganantonio, R., Garber, L.P., Kopral, C.A., Phillippi-Taylor, A.M., Traub-Dargatz, J.L., 2018. Parasite control strategies used by equine owners in the United States: a national survey. Vet. Parasitol. 250, 45–51.

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. American Association of Equine Practitioners Online at www.aaep.org on 10/14/2019.

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, 1–6.

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.

R Core Team, 2019. R: a Language and Environment for Statistical Computing. Online at. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reid, S.W., Mair, T.S., Hillyer, M.H., Love, S., 1995. Epidemiological risk factors associated with a diagnosis of clinical cyathostomiasis in the horse. Equine Vet. J. 27, 127–130.

Rendle, D., Austin, C., Bowen, M., Cameron, I., Furtado, T., Hodgkinson, J., McGorum, B., Matthews, Jacqueline, 2019. Equine de-worming: a consensus on current best practice. UK-Vet Equine 3 (1), 1–14.

Robert, M., Hu, W., Nielsen, M.K., Stowe, C.J., 2015. Attitudes towards implementation of surveillance-based parasite control on Kentucky Thoroughbred farms – Current strategies, awareness and willingness-to-pay. Equine Vet. J. 47, 674–700.

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 MiniFLOTAC methods. Vet. Parasitol. 247, 85–92.

Scare, J.A., Steuer, A.E., Gravette, H.S., Kálmán, Cs., Ramires, L., Dias de Castro, L.L., Norris, J.K., Miller, F., Camargo, F., Lawyer, A., De Pedro, P., Jolly, B., Nielsen, M.K.,

2018. Management practices associated with strongylid parasite prevalence on horse farms in rural counties of Kentucky. Vet. Parasitol. Reg. Stud. Rep. 14, 25–31.
Slusarewicz, P., Pagano, S., Mills, C., Popa, G., Chow, K.M., Mendenhall, M., Rodgers, D.W., Nielsen, M.K., 2016. Automated parasite faecal egg counting using fluorescence labelling, smartphone image capture and computational image analysis. Int. J. Parasitol. 46, 485–493.
Slusarewicz, M., Slusarewicz, P., Nielsen, M.K., 2019. The effect of counting duration on quantitative fecal egg count test performance. Vet. Par.: X 2, 100020.
Suzuki, C.T.N., Gomes, J.F., Falcão, A.X., Papa, J.P., Hoshino-Shimizu, S., 2013. Automatic segmentation and classification of human intestinal parasites from microscopy images. IEEE Trans. Biomed. Eng. 60, 803–812.
Traversa, D., von Samson-Himmelstjerna, G., Demeler, J., Milillo, P., Schürmann, S., Barnes, H., Otranto, D., Perrucci, S., Regalbono, A.F., Beraldo, P., Boeckh, A., Cobb, R., 2009. Anthelmintic resistance in cyathostomin populations from horse yards in Italy, United Kingdom and Germany. Parasit. Vectors 2, 1–7.
Tzelos, T., Morgan, E.R., Easton, S., Hodgkinson, J.E., Matthews, J.B., 2019. A survey of the level of horse owner uptake of evidence-based anthelmintic treatment protocols for equine helminth control in the UK. Vet. Parasitol. 274.
Vidyashankar, A.N., Hanlon, B.M., Kaplan, R.M., 2012. Statistical and biological considerations in evaluating drug efficacy in equine strongyle parasites using fecal egg count data. Vet. Parasitol. 185, 45–56.
Yang, Y.S., Park, D.K., Kim, H.C., 2001. Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. IEEE Trans. Biomed. Eng. 48, 718–730.