Understanding Fecal Egg Counting

Paul Slusarewicz, PhD

1. Introduction

Fecal egg counting (FEC) methods underpin parasite control programs in pasture animals. Yet, many of those using these methods are unfamiliar with their basic operating principals and various strengths and limitations. A sound understanding of these principles has become particularly important in recent years due to the emergence of significant anthelmintic drug resistance (ADR) since FECs are instrumental in detecting it via FEC reduction (FECR) tests as well as meliorating it by way of informed deworming strategies. The purpose of this document is to provide you with an understanding of the performance parameters of these TEC testing methods, particularly when comparing results from the Parasight automated egg counting system with those from manual egg counting methods.

2. FEC Methods

Traditional FEC methods rely on separating ova from the feces before examination, and this is based on the egg's lower density compared to the bulk of the fecal content. This is achieved using a concentrated sugar and/or salt-based floatation medium (FM) with a density of approximately 1.2-1.25 g/mL. This is important because the ova represent only a tiny fraction of the feces, and their identification within the fecal background would be essentially impossible otherwise. In contrast, the Parasight system pre-treats the feces with a stain that causes the ova to fluoresce so that that they can be detected within the fecal background after a simple filtration step.

It should be noted that the lack of any regulatory environment of set methodological standards in this area has resulted in the proliferation of countless variants of each methodology (excepting the Parasight System). Thus, a McMaster test conducted in one laboratory may well produce consistently higher or lower results than a test at a different laboratory because of (often subtle) differences in the methods (see Section 3.1, Systematic Variation).

For the purposes of this document, FECs can be divided into the following three categories.

2.1. Centrifugal Methods

Centrifugal methods are generally referred to as Wisconsin- or Cornell-Wisconsin-type tests and utilize a centrifugal field to accelerate the floatation of eggs within a fecal sample. They generally analyze a relatively large amount of feces (typically 1 gram) and eggs are trapped on a cover slip that is placed on the top of the centrifuge tube and is in contact with the sample. In one version of the test, the coverslip is placed on the tube during centrifugation in order to use the centripetal force to help “stick” more eggs onto the glass. However, this method can be messy when coverslips detach during spinning, and also requires a more costly and bulky swing-out bucket centrifuge. In an alternative version, the sample is spun without a cover slip in a more commonly available fixed-angle centrifuge and then removed and topped-up with additional FM to form a meniscus onto which a coverslip is placed so that the ova can float upwards and attach under the influence of gravity. In some versions, samples are first resuspend in water and centrifuged, and then the pellet resuspended in FM prior to re-centrifugation and ova capture as described above.

Because these types of method are more time consuming, more technically demanding, less convenient and require more equipment (i.e. a centrifuge) they are not commonly used in large animal veterinary practice and are more frequently encountered in parasitology research laboratories.

2.2 McMaster Methods and the AAEP Guidelines

Developed in the McMaster Laboratory at the University of Sydney in the 1930s, this is by far the most commonly used test in large animal veterinary practices today. Here the sample is resuspended in FM, filtered to remove large particles, and then a portion placed into the two chambers of a specialized McMaster slide, which contains a cavity that is 1.5 mm high. The eggs are allowed to float to the surface within the space under gravity and can then be counted using a microscope whose focal depth-of-field is sufficiently shallow to defocus the fecal debris at the bottom of the slide and so allow identification of the sharply focused ova. The surface of each chamber is etched or painted with a 1 cm^2 grid consisting of six lanes that aid the analyst in keeping track of what has been counted. The traditional McMaster test uses 4 g of feces and 26 mL of FM to produce a slurry that contains 0.133 g of feces per mL. Since each grid contains 0.15 mL of liquid (1 cm x 1cm x 0.15 cm), each therefore also contains 0.02 g of feces (0.133 x 0.15 = 0.02), while the slide (i.e. both grids) contains 0.04 g.

All egg counts are traditionally reported in units of eggs per gram (EPG), and since the raw count produced from a McMaster slide is derived from than only 0.04 g of feces the result needs to be adjusted using a multiplication factor (MF) to obtain the count for 1g of the original sample. In the case of the McMaster test described above, the MF is 50x when counting only one grid (1/0.02 = 50) or 25x when counting both. In the case of Wisconsin-type tests where 1 g of feces is used, no multiplication factor is needed since 1 g was used in the test, while in versions where more than one gram was used, the MF will be less than one. It is often not appreciated that any changes to the system, such as the ratio of fecal material or the dimensions of the counting slide, need to be reflected in an appropriate modification of the MF. MEP has produced a simple Windows-based app to calculate MFs based on this information if you would like to check your numbers, and it is available for download at: https://www.dropbox.com/s/tlomflsuhqnqzec/MF%20Calculator.exe?dl=0.

Because McMaster methods analyze substantially less feces that Wisconsin-type tests, they tend to be less sensitive. However, because they are faster, simpler, and more convenient and do not require a centrifuge, they are by far the most common tests used in large animal clinical practice today. When developing the Parasight System, we therefore specifically chose to calibrate it to produce results comparable to those produced by the McMaster test used in the laboratory of renowned equine parasitologist Martin Nielsen at the Gluck Equine Research Center of the University of Kentucky, who is also the Chair of the American Association of Equine Practitioners' (AAEP) parasite control task force, which publishes the "AAEP Parasite Control Guidelines."

2.3 The Parasight System

The Parasight System differs from traditional methods because it is a filtration- and not floatation-based method. It further differs because it uses a computer to count eggs, and so eliminates many sources of human error. For example, it is not uncommon for us to hear vet techs claim that they can count their McMaster slides in a minute, and yet a recently published a paper[1] that shows that rushing manual eggs counts significantly reduces both their accuracy and precision, while the commonly used practice of cutting corners by only counting one grid has a similar effect on precision[1,6]. We are also in the process of writing a paper on a study that shows that the accuracy and performance of McMaster counts are also

affected by the level of analyst training. Such sources of variation do not apply to the Parasight System's automated counts. Furthermore, the Parasight system analyzes 0.5 g of feces, which, while being less than the Wisconsin, is over 10-times more than the McMaster. This additional difference lies at the heart of its superior performance over McMaster, as explained below.

Figure 1. Illustration of accuracy and precision in diagnostic tests.

3. Principles of Diagnostic Tests

Diagnostic tests are never perfect, and their results are subject to many sources of variability, both between different laboratories and even within a single laboratory. These errors are introduced by many factors, including differences in test protocols, the technique or training of the laboratory technician and factors inherent in the material being tested. FECs are especially susceptible to these three variables, and particularly to the latter, over which the analyst has almost no control.

Accuracy and precision are central concepts in diagnostic testing and respectively describe how close, on average, test results are likely to be to the actual value of the sample, and how reproducible results from a series of tests run from the same sample will be. These concepts are commonly illustrated as the clustering and positioning of a series of shots at a hypothetical target (see Figure 1). In this diagram, the bullseye centers represent the "true" value of the sample, while the black "shots" represent a series of results obtained for the same sample from four hypothetical diagnostic tests with different levels of accuracy and precision.

As shown in the diagram, low precision, even in an accurate test, leads to a higher probability of low accuracy for any single given test, since any single result is more likely to produce a readout that is distant from the true value. Highly accurate tests with low precision, therefore, require sufficient multiple replicate analyses in order average out the errors stemming from their poor precision and produce truly accurate results; unfortunately, this approach is seldom practical in a busy clinical laboratory. Increased precision, therefore, increases the confidence that any given result is truly representative of the sample being tested.

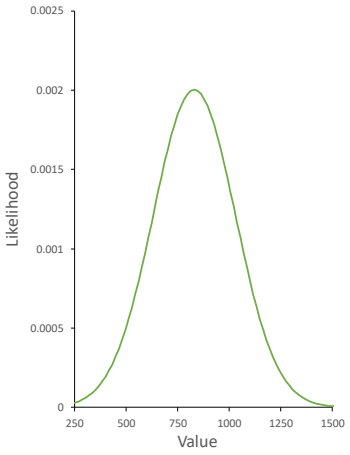

Figure 2. Precision represented as a bell curve.

Precision can also be illustrated by a bell curve (see Figure 2). The peak of the curve represents the true value of the sample being analyzed (in this case 830), and the height of any given point on the curve represents the probability of obtaining the result beneath it on any given test of the same sample. Less precise tests exhibit more flattened and broader curves, meaning that the likelihood of any given test producing a result close to the true value is reduced and the chances of results further away from the true value are increased.

FEC performance generally fits into one of the two categories depicted in the upper half of Figure 1 and, as a result, this can lead to apparent "discrepancies" when comparing test results for the same sample between not only two different methods, but even *the same* method, for the reasons discussed below.

3.1. Systematic Variation

Systematic discrepancies between results from two different egg counting techniques are those where counts from one method are consistently lower or higher than those from another; these generally arise from intrinsic differences between the methodologies. For example, although Wisconsin-type methods use substantially more fecal sample than others, they are known to be substantially less accurate and produce results that are almost half those of McMaster[2], perhaps due to losses during centrifugation. Thus, while comparisons of counts from the same samples conducted by both Wisconsin and McMaster do show a correlation, the Wisconsin counts are consistently lower. This is often a surprise to people not immersed in the field, since Wisconsin-type tests are commonly thought of as being "superior" because they analyze more fecal matter. In fact, in addition to their lower accuracy, Winsconsin-type tests also exhibit lower precision than both the McMaster and Parasight tests[5] (see Section 3.2, Sampling Variation, below).

Furthermore, results from the same samples run using different versions of the McMaster from different laboratories can also display such systematic variation. Such variants are often referred to as "modified McMaster" tests, where the modifications have been made to customize the test to a particular laboratory environment or to make it faster, easier or more convenient. Many such variant methods are available online (for example, see references 7-9), but It is our experience that the effects of these modifications are seldom, if ever, systematically evaluated (much less validated) to determine whether they affect test performance. Such modifications often lead to systematic result discrepancies compared to Parasight. The Parasight System has been calibrated to, and extensively validated against, the McMaster test used at the Gluck Equine Research Center and has consistently been shown to produce equivalent results in a number of papers published in peer-reviewed research journals[1,3-5]. Thus, any systematic discrepancies between Parasight System results and any given practice's version of the McMaster test are most likely due to the specific McMaster methodology being used, and the same discrepancies would be observed in a comparison with the McMaster test used at the Gluck.

We have been asked to investigate several such discrepancies over the past year, and in each case the root cause was identified as a key difference in the McMaster test being used compared to the Gluck method. In one case, the laboratory's McMaster was producing substantially higher counts than the Parasight System. Upon investigation, we could not find any obvious methodological problems until we manually counted a slide ourselves and then asked a tech to do so with the same slide. When the result came back substantially higher, we eventually discovered that the tech was counting all of the eggs that could find at all focal levels (that is, not those solely in the focal plane of the grid). Since *all* McMaster tests stipulate that only the upper plane be counted (because that is the very purpose of the test), this laboratory was not only overcounting relative to Parasight, but to *all* other properly conducted McMaster tests.

In another case of over-counting, we discovered that, while the laboratory was using the usual 4 g/ 26 mL McMaster ratio (producing a slurry of 0.133 g/mL) and 25x multiplication factor, their sample preparation vessel (which consisted of a plastic tube with two Sharpie marks on the side for the FM and fecal sample levels), was improperly calibrated. Thus, this device was measuring 6 g and 27.5 mL to produce a 0.218 g/mL slurry. Since the MF had not been adjusted downward, the result was an overestimation of EPG by the McMaster by a factor of 64%. When the MF was corrected to 15.25, the Parasight and McMaster results fell into agreement.

In another laboratory, we discovered that a Whitlock, rather than a McMaster, slide was being used, and yet the 25x MF had not been modified. Whitlock slides contain 0.5 mL of slurry under each grid, and so each slide contains 1 mL. As explained above (Section 2.2., McMaster Methods), a true McMaster slide contains 0.3 mL under both grids, and it is from this that the MF is partially derived. Thus, this laboratory should have been using an MF of 7.5x, not 25x, and so systematically overestimating all counts by 3.33-fold.

In a final example, one laboratory that also reported "low" Parasight counts and was using a commercial "modified" McMaster that eliminated the filtration step to save time and improve convenience. Upon investigation at the Gluck Equine Research Center, we discovered that this same test also produced substantially higher counts than the Gluck version because it was not subject to the egg loss that usually occurs during filtration. Furthermore, the lack of filtration produced extremely murky slides that were difficult to read, and which likely reduced test precision (although this was not formally investigated). When we contacted the manufacturer and asked whether they had any validation data for their test we received, perhaps unsurprisingly, no response.

Taken together, these examples will hopefully illustrate that if you are observing a systematic discrepancy between your in-house McMaster and Parasight, it is possible that it may due to such a hidden discrepancy between your McMaster method and that used at the Gluck Research Center.

3.2. Sampling Variation

Sampling variation is perhaps the most difficult aspect of egg counting to understand, and the most commonly misunderstood, for someone who has not been fully immersed in the field of parasitology. In order to address this, we need to first discuss what we mean by "sampling".

Let us consider an intact fecal sample that may weigh thousands of grams. Obviously, it is impractical to process the entire sample, count every single egg, and then divide by its weight to get an accurate EPG value. As a result, we need to take a smaller sample (usually between 1 and 10 grams) that we hope will represent the whole stool. Unfortunately, ova are not necessarily distributed evenly within the feces, and so there will some chance variation in the number of eggs in the fecal subsample, depending on whether it contains egg-rich or egg-poor regions. We refer to this kind of variation as "biological variability".

Now let us consider the slurry that is produced from this material. Again, it is impractical to count all of the eggs in the slurry, since it consists of many tens of mLs of liquid. Instead, we take another subsample, and accept that it may not be completely representative of the actual concentration of eggs in the slurry. Unlike chemical analyses, where subsamples containing countless billions of molecules are statistically a good representation of the sample itself, this is not the case in fecal samples, where each subsample contains only a few to a few thousand ova. One simply cannot expect the same kind of precision as one might obtain from, for example, a blood test for glucose because a few thousand eggs are not as evenly dispersed in the sample as are a few billion or more molecules. This effect is particularly problematic at low egg counts since the likelihood of uneven particle distribution increase with decreasing particle number.

Next, we add our sample to the McMaster grid, two-thirds of whose area lies outside the grid to be counted. It is entirely random as to whether an egg will come to rest in or outside the grid, with some grids being over- or under-represented with eggs by sheer chance. We refer to these two latter sources of variation (i.e. slurry subsampling and grid distribution) as "technical variability", with analyst-to-analyst comprising the third such source.

Now that we have defined our terms, let us examine what effects these phenomena may have on actual egg counts. While many people understand that one should not necessarily expect "the same" egg count from the repeated counts of the same samples due to the variation inherent in the various stages of sampling and subsampling, very many don't have any kind of feel of the magnitude of this variation (not surprisingly, since practitioners don't have time to run multiple counts from the same sample and so have not experienced it for themselves). In contrast, since we are in the egg-counting business, we have extensive experience of this kind of variability and what to expect from it.

Figure 3 depicts paired McMaster counts from 90 separate strongyle-positive stool samples, where each count pair was conducted on a separate subsample of the same stool. The paired data have been plotted against each other in order to visualize the agreements between the paired counts. The first thing to notice is that, while there is a reasonably strong positive correlation, the data are extremely scattered, indicating significant quantitative disagreement within the sets of paired counts.

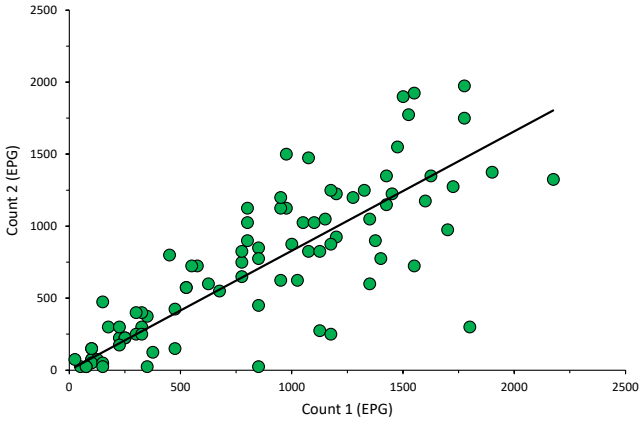

Figure 3. Comparison of duplicate McMaster counts of the same samples.

The second thing is that, while the straight line denotes the least-squares best fit to the data, *very few of the data points lie on the line.* This demonstrates that, rather than expecting the results from two counts of the same sample to agree (or at least agree very closely, one should instead *expect the results to be quite different*, and in some cases *very* different. While in in approximately two-thirds of cases the differences are below 50%, in some they can be as high as 600%. This is illustrated in Figure 4, where the same data are replotted as the fold difference of the second count of each sample relative to the first. In this graph, test pairs that were in complete agreement lie on the x-axis and represent only 2 out of the 90 test pairs (i.e., 2.2%), while only twenty pairs (22.2%) were within 10% of each other. About one third of all the paired counts differed by 20-50%, and over a quarter did so by over 50%, while 11% differed by as much as 300% or more. The full frequency distribution of these data is presented in Figure 5 and provides a good snapshot of the amount of variation one should expect with a properly conducted McMaster test. It should therefore be noted that even a hypothetically perfect test (i.e., one that produced the “true” result each time) would exhibit apparent discrepancies when compared to McMaster because of the inherent variability of the latter.

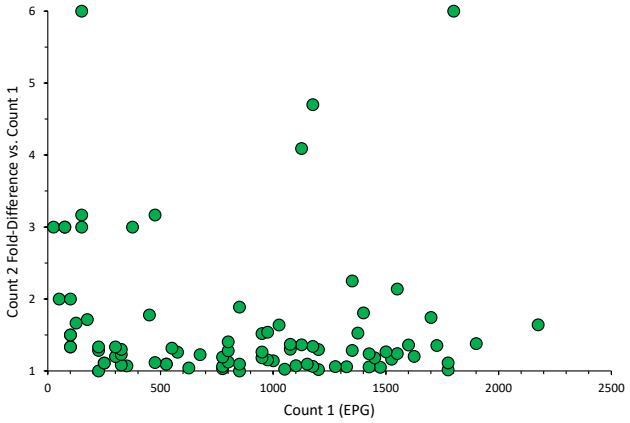

Figure 4. Fold-difference of second McMaster count relative to first for each paired count.

It should also be noted that these results illustrate the variability between replicate counts using the *exact same test.* When comparing two different tests, such as Parasight and a McMaster test that is a

variant of the one used at the Gluck Center, this variability could also be layered atop the systematic variability discussed above (Section 3.1, Systematic Variation).

These data should also demonstrate the kind of rigor required to fully characterize an analytical assay, and this is the kind of rigor that vendors of other FEC systems have never subjected their products to. In addition, practices simply do not have the time to properly characterize their own in-house FEC system, and so many may be unaware that if are performing poorly (see the examples given in Section 3.1, Systematic Variation). It should also be clear that comparing methods by examining single repeated counts of a handful of samples is an entirely inadequate approach towards accurately evaluating system performance. This is why we have done the work for you in order to produce a system that performs to the levels of a research parasitologist and yet can be used by anyone with almost no training.

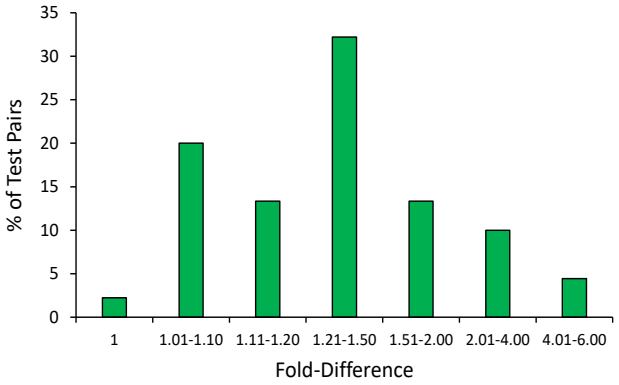

Figure 5. Frequency distribution of fold-differences between paired McMaster counts.

We have also conducted studies to formally quantify the degree of expected technical variation of both the McMaster and Parasight tests[1]. To do this, we used a statistical metric of precision known as the coefficient of variance (CoV), which is defined as the standard deviation of a population (in this case a group of replicate test results) expressed as a percentage of its mean. This metric gives an estimate of how close to the mean one would expect two-thirds of the repeated counts to cluster – in other words, it is a measurement of the broadness of the bell curve shown in Figure 2. For example, if we conducted 100 repeated counts on a sample using a test with a CoV of 20% and found a mean of 1000, we would expect, on average, that approximately 66 of those tests would return counts between 800 and 1200 EPG. Similarly, if the CoV of the test was 10%, then two-thirds of the counts should cluster between 900 and 1100 EPG.

As you can see, results from Parasight produce a substantially narrower and taller curve than McMaster. The practical result of this difference is that Parasight is 2.5-times more likely to produce the "true" result from a single test of any given sample than the McMaster test. Conversely, the McMaster test has the potential to produce results that are substantially further away from the "true" value than the corresponding Parasite result. Thus, while the accuracies of the two tests are the same (by design) when averaged over a large number of repeated measurements, the likelihood of a single result being accurate is substantially higher when using Parasight. To give an idea of the potential impact of these differences, consider Figure 6, which shows the experimentally derived probability distributions of the two tests for a sample averaging approximately 470 EPG. The probability of any given McMaster test giving a count below 200 EPG in such samples (and therefore potentially leading to a decision not to treat when treatment should be considered) is 5.5%. In

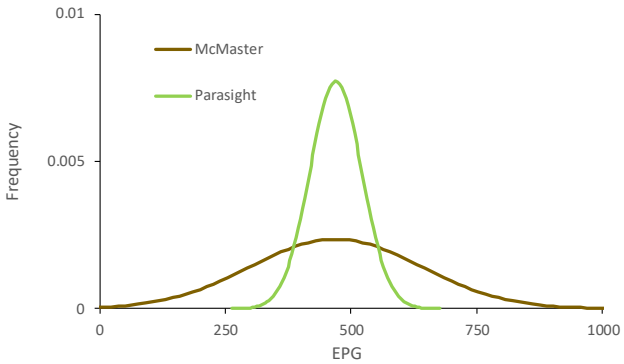

Figure 6. Technical precision comparison of McMaster with

contrast, and because of its superior precision, the possibility of such a misdiagnosis using the Parasite system is less than 0.003%, which is at least 2000 times less likely than when using McMaster.

Part of the reason behind this elevated precision is the larger amount of fecal sample analyzed by Parasight. This means that Parasight can reduce the inherent sampling errors we discussed earlier, and so increase the reliability of any given single test result. It should also be noted that McMaster is a highly operator-dependent procedure, and that technicians are prone to fatigue when counting multiple samples at a microscope or can be placed under time-pressure to conduct many counts in a short period. In contrast, since the process is highly automated, Parasight is essentially operator independent. These differences were not accounted for in the above studies, where the McMaster test was performed by highly experienced analysts over a long period of time; thus, "real-world" differences may well be even greater than those presented here.

3.3 Resolution

Another aspect overlooked when using some modified McMaster test is the granularity of data they produce, which is related to the MF and therefore to the ratio of feces to FM used when producing the fecal slurry. MF for these modified tests commonly ranges between 25 and 100, and sometimes even 200. As a consequence, each egg counted increases the final count by the same degree as the MF. Thus, consider a test with an MF of 50[7]. In this case, counting 3 versus 4 eggs will produce counts of 150 vs 200 EPG, while with a test with an MF of 100[9], counting 1 versus 2 eggs will yield results of 100 and 200 EPG, respectively. Because Parasight counts very many more eggs, it's MF is 3.5. Thus small differences in the number of eggs counted due to the variability discussed above result in smaller differences in the final egg count.

4. What Does This All Mean?

The preceding discussion has hopefully given you a clearer understanding of the kind of considerable variation one should expect when conducting FECs. The fact of the matter is that such variation is unavoidable due to the nature of the material being analyzed, i.e., low levels of anylate (in this case, ova) in relatively large volumes. Thus, rather than expecting counts of the same sample to yield the same value, one should in fact expect them not to, and be surprised when they occasionally, by happy chance, do. But what does that mean in reality?

Perhaps the greatest customer concern lies with respect to treatment decisions in the case where a comparison of McMaster and Parasight results yield numbers that respectively lie above and below the treatment threshold (commonly 200 EPG). However, such a discrepancy in fact represents a false dilemma because it also applies to the McMaster (or any other FEC) method itself. For example, what if we were to conduct two McMaster counts on the same sample and get a result of 150 EPG with one and 250 with the other (a difference that is easily within the range one could, and in fact, should, expect)? Would we treat or not treat? The fact of the matter is that practitioners seldom, if ever, face this quandary because they (understandably) never conduct multiple counts. If the single count that was conducted gave a result of 150 EPG they would not have treated, while if it had instead been 250 EPG they would have, irrespective of what the "actual" count of the sample was. Asking the same question with respect comparing a single Parasight and single McMaster count is tantamount to the same thing. Since we can never know the "true count" (because analyzing the whole stool is impractical) we are forced to be pragmatic and need to understand that 200 EPG is not a magical clinical life-or-death threshold, but rather an approximate level that has been empirically determined to produce a good balance between worm control and maintaining

refugia in sufficient numbers to ameliorate the development of ADR. In reality, some horses under 200 EPG will inevitably be treated and some above 200 EPG will not, regardless of what egg counting method is used. For example, let us take the McMaster comparison data shown in Figures 3 and four and consider just the horses in the 0-500 EPG group. This region from Figure 3 is enlarged in Figure 7. As you can see, 9 of the samples lie in the green region, where both McMaster counts were below 200 EPG and so there would be agreement to "not treat." However, there are 6 samples that lie in the red zones, where one test result was above 200 and one below, resulting in an apparent contradiction. In this case, 20% of the under-500 EPG test group could potentially have been "misdiagnosed" by the McMaster test. Furthermore, these data were produced using the AAEP method, which has an MF of 25. Had they been generated with a method with a higher MF, and therefore lower resolution (see Section 3.3, Resolution, above), it is likely that substantially more of the data points would have fallen in the red zones.

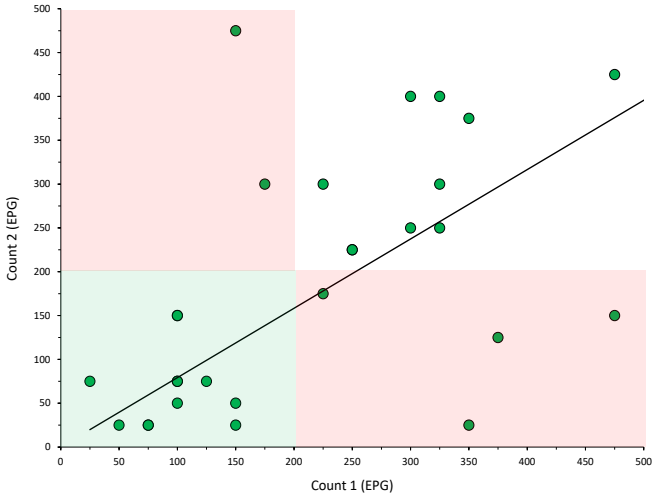

Figure 7. McMaster variation around the treat/not-treat level.

This, however, is not the problem that it may appear to be *en face*. The point of performing well-conducted egg counts is to identify high shedding animals and treat them to reduce the infection pressure from larvae or ova in the environment while maintaining sufficient non-resistant refugia to dilute drug-resistance genes in the helminth population. In this manner, worm burdens in the equine population are sup

pressed to the point where they do not cause disease while keeping anthelmintic use to a minimum to prevent the further expansion of ADR. The small amount of over- or under-treatment that results from egg count variability will, in the end, not affect this goal because it will average out and with respect to maintaining the egg-reduction/refugia balance. Thus, counts are essential to provide the data that needed to reliably inform such a parasite control strategies, and do not necessarily have to identify with 100% specificity each horse above and below the 200 EPG level (which is an unattainable goal). Nevertheless, based on the extensive data produced during the validation of the Parasight System, we are confident that it comes far closer to reaching that lofty goal than does the McMaster or any other FEC testing methodology in use by veterinarians today.

5. Conclusion

Although ubiquitous in clinical practice, FECs are the subject of much misunderstanding with respect to their actual performance parameters. In reality, the results produced by all FEC methods are inherently variable, and while this may be recognized by most veterinarians and laboratory technicians, most would be surprised as to the actual magnitude of that variation, because such information usually resides in the domain of research parasitologists. In this document, we have presented data to illustrate the variability inherent to the McMaster test as well as experimental results to show why the Parasight System, while not eliminating it entirely, goes a long way toward providing more consistent and reliable results.

6. References

1. Slusarewicz, M., Slusarewicz, P., and M.K. Nielsen. 2019. The effect of counting duration on quantitative fecal egg count test performance. *Vet. Parasitol. X:* **doi** 10.1016/j.vpoa.2019.100020.
2. Levecke, B., L. Rinaldi, J. Charlier, M.P. Maurelli, A. Bosco, J. Vercruysse, and G. Cringoli. 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.
3. Slusarewicz, P., S. Pagano, C. Mills, G. Popa, K.M. Chow, M. Mendenhall, D.W. Rodgers, and M.K. Nielsen. 2016. Automated parasite faecal egg counting using fluorescence labelling, smartphone image capture and computational image analysis. *Int. J. Parasitol.* **46:**485-493.
4. Scare, J.A., P. Slusarewicz, M.L. Noel, K.M. Wielgus, and M.K. Nielsen. 2017. Evaluation of accuracy and precision of a smartphone based automated parasite egg counting system in comparison to the McMaster and Mini-FLOTAC methods. *Vet. Parasitol.* **247:**85-92.
5. 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., and M.K. Nielsen. 2020. Diagnostic performance of McMaster, Wisconsin, and automated egg counting techniques for enumeration of equine strongyle eggs in fecal samples. *Vet. Parasitol.* In Press.
6. Kochanowski, M., J. Dabrowska, J. Karamon, T. Cencek, and Z. Osinski. 2013. Analysis of the accuracy and precision of the McMaster method in detection of the eggs of Toxocara and Trichuris species (Nematoda) in dog faeces. *Folia Parasitol. (Praha).* **60:**264-272.
7. https://www.rvc.ac.uk/Review/Parasitology/EggCount/Purpose.htm
8. https://content.ces.ncsu.edu/modified-mcmasters-fecal-egg-counting-technique
9. http://www.vetlab.com/McMaster%20technique%20for%20insert.pdf