Precisely Parasight by Paul Slusarewicz, PhD, ARCS

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. The McMaster fecal egg count test is particularly susceptible to these three variables, resulting in its generally accepted poor performance.

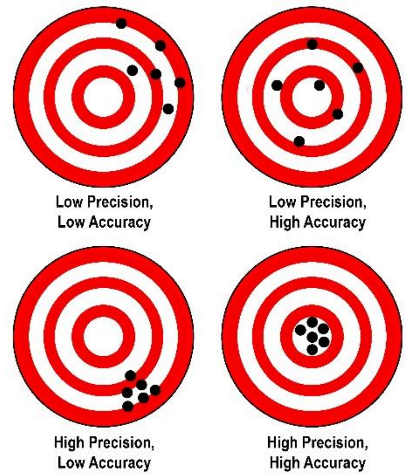

Figure 1. Illustration of accuracy and precision in diagnostic tests.

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 to average out the errors from their poor precision in order to produce truly accurate results; this is seldom practical in a busy clinical laboratory. Increased precision, therefore, increases the confidence that any given result is representative of the sample being tested.

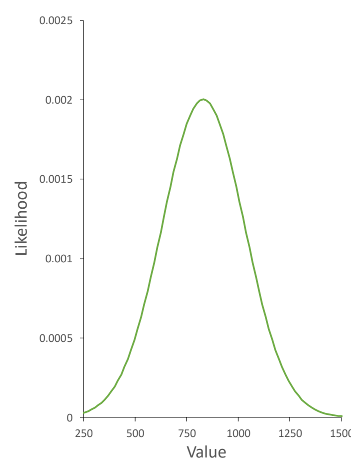

Figure 2. Precision represented as a bell curve.

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.

The Parasight test has been optimized for both accuracy and precision. Its accuracy was calibrated to produce, on average, the same egg counts as the McMaster test described in the parasitology guidelines issued by the American Association of Equine Practitioners (AAEP). As a result, veterinarians can continue using the same familiar metrics (such as treatment thresholds) for their parasite control programs. Since the McMaster test is notoriously imprecise, calibration was achieved by running hundreds of tests on dozens of samples in order to average out these errors. Since this process involved counting multiple samples multiple times using both Parasight and McMaster, it also allowed us to statistically evaluate the relative precision of both tests.

Example bell curves corresponding to the McMaster and Parasight tests are shown in Figure 3. As you can see, results from Parasight produce a significantly narrower and taller curve than McMaster, indicating a much higher degree of precision. The practical result of this difference is that Parasight is 2.5 more likely to produce the "true" result from a single test of any given sample as the McMaster test for samples containing more than 500 eggs per gram (EPG). Furthermore, the McMaster test has the potential to produce results that are substantially further away from the "true" values than corresponding Parasite results. Thus, while the accuracies of the two tests are similar (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. Furthermore, in samples containing less than 500 EPG, which encompasses the 200 EPG level that is commonly used as the threshold for making treatment decisions, Parasite was 3.5-times more precise as McMaster (as depicted in Figure 3). To give and idea of the potential impact of these differences, consider Figure 3, which shows the experimentally derived probability distributions of the two tests around samples 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 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. We are in the process of preparing a manuscript describing these results for submission to a peer-reviewed scientific journal.

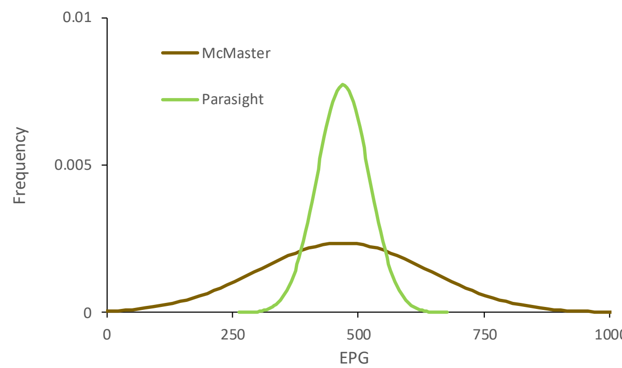

Figure 3. Precision comparison of McMaster with Parasight.

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 sampling errors inherent in subsampling and so increase the reliability of the final 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. 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 a highly experienced analyst over a long period of time. Furthermore, numerous "modified" McMaster tests are currently being marketed and used that have never been validated with respect to their performance, and these too could represent additional sources of "real world" variation.

Surprisingly, the effects of such methodological modifications and operator-dependent effects have never been assessed quantitatively, even though the McMaster test is almost a century old, but it does seem likely that the performance superiority of Parasight would be even higher relative to McMaster in real-world situations. We are in the process of performing such assessments; for example, another parameter that can introduce variability to a test is counting speed. This is of particular importance in busy laboratories, where sample throughput needs to be maximized, and it has been our experience that it is not uncommon to encounter laboratory technicians who count McMaster slides in under a minute, although the effect of “rushing” counts has never been assessed. Since the Parasight system is fully automated and takes 2.5 minutes to run a test, it is independent of such considerations. We have just completed a study, which we are also preparing for publication, that shows that counting slides for 1 minute reduces the counts (i.e. the accuracy) of McMaster tests by 50% relative to counts performed with no time constraints (taking approximate 4 minutes per slide). Furthermore 1-minute counting also decreases precision by one-third, thereby further increasing the performance improvements of Parasight over McMaster.

Another approach to increasing throughput is to count only one of the two grids on a McMaster slide. While this does decrease the sensitivity of the test, its effect on accuracy and precision has never been assessed. We also examined this in our study and found that while accuracy (over a large number of averaged counts) was unaffected, precision decrease by one third, as it had had when counting both grids in a minute. Once again, Parasight is unaffected by such time-constraint considerations.

A final advantage of Parasight is its simplicity of operation, which eliminates other human factors such as analyst weariness and skill level. In fact, one of our customers’ 10-year-old daughter is currently conducting the fecals for her practice using the Parasight system. We are also currently conducting a study to compare McMaster and Parasight counts using analysts both before and after training to assess the effect of skill level on test results and are again expecting to find Parasight to be superior to McMaster in this respect.

In summary, our testing and validation of the Parasight system have shown that even the least-trained analyst is now capable of producing equine fecal egg counts that surpass the results obtained from the McMaster test detailed in the AAEP guidelines when performed by even experienced parasitologists.