Dairy practitioners and our clients have unprecedented access to data through the advent of the internet. List serves like aabp-l provide bovine unprecedented access to data through the advent of the internet. The problem is how to keep up with all of the new information that is coming at us and most importantly, how to validate the information that we receive in order to use it in our daily practice of veterinary medicine. Our quest should be to manage information in a way that it can be useful to us for answering the daily questions that arise. Recent questions that have emerged in our practice include:

- What is the efficacy and risk of placing systemic antibiotics in the peritoneal cavity during routine displaced abomasum (DA) surgeries?
- What is the efficacy and risk of using penicillin in a mixture with a steroid administered in a subconjunctival injection for the treatment of pink eye in cattle?
- What is the diagnostic accuracy of the Pathoproof Mastitis PCR test in milk quality programs, especially when using samples not obtained via aseptic technique?

Veterinarians should strive to be "lifelong learners" and must resist sticking our head in the sand when it comes to harnessing access to new data. Utilizing the skills of Evidence-Based Medicine will help us successfully incorporate that data into the practice routine. The difficulty in this task is learning the skills of critically evaluating and efficiently incorporating new information into practice.

Diagnostic testing is a daily component of practice for dairy veterinarians. Diagnostic testing can be as simple as the daily palpation that some practitioners do hundreds of times a day to as complex as a genotypic test that looks to see if a particular gene is present in bacterium isolated from a diagnostic submission. This paper will concentrate on applying Evidence-Based Medicine techniques to the evaluation and application of diagnostic tests in veterinary practice.

WHAT IS EVIDENCE-BASED MEDICINE?

Evidence-Based Medicine was introduced into the human medical community when a former wartime prison doctor recognized that the medical community needed to substantiate medical decisions based on scientific evidence in the early 1970s. This process of forging new information and emerging technology into medical practice was formally named Evidence-Based Medicine by McMasters’ Medical School in Canada in the 1980s. The process has evolved into the publication of thousands of manuscripts about the topic in the human medical field. The topic of Evidence-Based Veterinary Medicine (EBVM) first appeared in the literature in the late 1990s.1

In order to apply EBVM in practice, practitioners must have the ability to2-3:

1. Turn a need for clinically relevant information into a question. This question can pertain to evaluating the performance of a diagnostic test for the animal or population in question, evaluating the efficacy of a preventative or therapeutic intervention, predicting the outcome of a disease, or predicting the cost or the risk associated with an intervention.
2. Perform a search for the best available information to answer the question with an emphasis on efficiency.
3. Critically appraise the information for its validity and usefulness to answer the question at hand.
4. Utilize the information to form clinical judgments and implement actions.
5. Perform an outcomes based evaluation of the success and the execution of the first four steps and seek information to improve the clinical outcomes and data evaluation process.

EBVM challenges us to place less weight on the subjective opinions that have developed from previous clinical experience, as these are often based on poor memory and/or record keeping and instead look to more rigid forms of evidence to guide us in the practice of veterinary medicine. The types of evidence that should be used include (listed from strongest to weakest evidence)2:

- Systematic reviews
- Meta-analysis
- Blinded random controlled trials
- Cohort studies
- Case control studies
- Case series
- Single case reports
- Editorials, opinions, comparative animal research
- In vitro studies

Finding a sufficient number of relevant information sources in the veterinary literature is more difficult due to the limited number of high quality references as compared with the human medical field.2-4,5 The veterinary research community is attempting to improve the quality of information sources specifically for the veterinary clinical community through the development of the REFLECT (Reporting Guidelines for Randomized Controlled Trials for Livestock and Food Safety) statement. The REFLECT statement resulted from several modifications and additions of the CONSORT(Consolidated Standards of Reporting Trials) statement to take into account the unique aspects of reporting livestock trials6,7 and is meant to be an evidence based minimum set of standards for trial reporting production, health, and food safety outcomes (www.reflect-statement.org).
APPLYING EVIDENCE-BASED MEDICINE TO DIAGNOSTIC TESTS

The presentation associated with this paper will summarize the results from several diagnostic laboratories that provide diagnostic service to geographic areas with higher populations of dairy cattle. The purpose of the presentation is to provide attendees with trends being seen from diagnostic submissions for common diseases associated with dairy cattle and to use that evidence in their practice. None of the actual data summaries will be presented in the remainder of this paper due to space and time constraints and the need to address the critical evaluation of diagnostic test results.

Diagnostic tests range in complexity and cost, but the characteristics of critical appraisal will apply whether the test is a rectal palpation for pregnancy diagnosis or a multiplex polymerase chain reaction (PCR) test. Before spending much time about the complexities of critical appraisal, a veterinarian must consider the potential bias introduced into the diagnostic process. Ask the questions:

- What is the question I am trying to answer?
- Why am I performing diagnostic tests?
- What samples should I collect?
- Which animals should be tested?

Diagnostic testing should be done to help pare down a predetermined list of differentials, not to establish the differential list. Practitioners must understand that bias can be introduced into the situation through the selection of the wrong animals to sample, failing to sample the correct area of affected tissues or incomplete sampling, and failing to preserve tissues correctly. Unlike our swine and poultry counterparts, bovine veterinarians have a tendency to submit samples from the entirely wrong population of animals, the one that died after some prolonged (often unknown in exact duration) illness. Poultry and swine veterinarians will sacrifice acutely affected animals in order to establish an unbiased diagnosis. Likewise, submitting a sample from an animal housed in environment where infectious agents are ubiquitous (e.g., a fecal sample from a calf with diarrhea) will not provide a lot of long-term diagnostic value unless also concentrating on why the affected population is succumbing to the agent while other populations (animals from different cohorts, different farms, etc) are not being affected. Therefore, concentrating on management aspects of calf rearing would serve as a more prudent use of diagnostic dollars as compared to simply identifying the agent.

TERMINOLOGY OF DIAGNOSTIC TESTS

Practitioners utilize diagnostic tests in one of two ways, either as a screening test or a confirmatory diagnostic test. A screening test can be used to screen a herd for the presence of infected animals. An ideal screening test should identify as many diseased animals as possible (minimal false negatives). A diagnostic test could be used to confirm the presence of disease based on a practitioner’s physical exam findings or to confirm the presence of an infected animal as a follow up to a screening test. An ideal diagnostic test should correctly identify as many non-diseased animals as possible (minimal false positives).

In order to determine whether a particular test would be a good screening and/or diagnostic test, we must understand how accurately a test determines the status (diseased, infected, exposed or not affected) of an animal. There are several epidemiological terms used to describe how a diagnostic test performs in a population of infected, exposed, or diseased animals. Diagnostic test accuracy is determined by a comparison of the diagnostic test in question against a gold standard. The ideal gold standard test is one that can perfectly discriminate between diseased and normal. Of course, there are very few examples of a perfect test in the world. Therefore, descriptive terminology of new diagnostic tests is only as reliable as the perfectness of the gold standard.

- **Test sensitivity** is the ability of a test to identify a diseased animal in a population of diseased animals (based on the gold standard). If a new diagnostic test identifies 50 test positive (T+) out of 100 (D+) diseased animals (based on the gold standard), the sensitivity of the test is 50%.
- **Test specificity** is the ability of a test to identify a non-diseased animal in a population of non-diseased animals (based on the gold standard). If a new diagnostic test identifies 450 test negative (T-) out of 500 (D-) diseased animals (based on the gold standard), the specificity of the test is 90%.

When evaluating the performance of a diagnostic test, it is usually helpful to compare the results of the gold standard and the test in question in a 2 x 2 table. Table 1 is an example of sensitivity and specificity values for a commercially available Johnes ELISA.

<table>
<thead>
<tr>
<th>True status (D)</th>
<th>Test status (T)</th>
<th>Test status (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (T+)</td>
<td>127</td>
<td>5</td>
</tr>
<tr>
<td>Negative (T-)</td>
<td>120</td>
<td>712</td>
</tr>
<tr>
<td></td>
<td>247</td>
<td>717</td>
</tr>
<tr>
<td></td>
<td>964</td>
<td></td>
</tr>
</tbody>
</table>

The false-negative rate is the proportion of diseased animals that produce a negative result for the test in question. The equation to determine the false-negative rate for a test population is (1 – sensitivity). In the example from Table 1, the false-negative rate is 1 – 0.514 = 0.486 (48.6%).

Sensitivity and specificity by themselves tell us very little about how well a test performs. In life, there are very few tests that have 100% sensitivity and 100% specificity. When evaluating the performance of a test, sensitivity and specificity must be evaluated together as considering either separately does not provide the complete picture about the test. In addition, it is important to remember that if you desire to improve sensitivity or specificity, it will come at a cost of reducing the other.
Test results can be expressed in two forms, categorical or continuous. Categorical results produce a yes or no answer (positive or negative, pregnant or open). Continuous results produce a result along a quantitative scale at which a cut-point is utilized to determine if the outcome is abnormal. Sensitivity and specificity values are affected by where the arbitrary cut-point is established. While practitioners have very little to do with the establishment of the cut-point, it is extremely valuable for us to understand what shifting the cut-point up or down will have on the identification of diseased animals. Looking again at the example in Table 1, if it is the desire of a practitioner to identify as many potentially positive animals as possible (have a high sensitivity), the quantitative cut-point should be shifted downward in order to produce more positive animals. However, this will come at a cost of decreasing the specificity of the test. This will result in more false positives. However, if the desire of the practitioner is to have a test that minimizes the chance that an animal will be culled from a herd if she is MAP positive, then the current cut-point is sufficient as the false-positive rate is very low.

Additional terms that become useful in determining the accuracy of a test are the positive and negative predictive values. In a comparison of a new diagnostic test versus a gold standard, the true disease status of an animal is known based on the outcomes of the gold standard. However, in real life a practitioner does not know the actual disease status when evaluating a herd. The positive and negative predictive values provide an estimate of the proportion of animals that are classified correctly. The positive predictive value (PPV) for a test is the proportion of the animals with a positive test that actually have disease. The following equation is used to determine the PPV:

$$PPV = \frac{\text{True positive animals}}{\text{true positives} + \text{false positives}}$$

Looking at the example from Table 1, the test in question produced 127 true positives and 5 false positives.

$$PPV = \frac{127}{(127 + 5)} = 96.2\%$$

The negative predictive value (NPV) for a test is the proportion of the animals with a negative test that are truly not affected by the disease. The following equation is used to determine the NPV:

$$NPV = \frac{\text{True negative animals}}{\text{true negatives} + \text{false negatives}}$$

Again looking at the example from Table 1, the test in question produced 712 true negatives and 120 false negatives.

$$NPV = \frac{712}{(712 + 120)} = 85.6\%$$

Based on these results, approximately 4% of the animals that tested positive are truly negative while approximately 15% the animals that tested negative are truly positive.

The apparent prevalence is the predicted prevalence of the herd based on the new diagnostic test. From the example from Table 1, the apparent prevalence of the herd in question is 13.7% (132 true positives/964 animals in the test population). The true prevalence is the determined by taking the number of diseased identified by the gold standard test in the test population. In the example herd in Table 1, the true prevalence of the herd is 25.6% (247 true positive animals/964 animals in the test population).

As can be seen from the above definitions, positive and negative predictive values, as well, as apparent prevalence are based on the specificity and sensitivity of the test in question. However, they are also affected by the true prevalence in a herd. This example is demonstrated in Tables 2 and 3 below comparing two hypothetical 1000-cow herds, one with a true prevalence of 30% (Table 2) and the second with a low prevalence, 5% (Table 3). In both scenarios, the test sensitivity and specificity are identical.

In the scenario in Table 2, the apparent prevalence is 15.9% compared to a true prevalence of 30%. In the low prevalence herd, the apparent prevalence is closer to the true prevalence (3.3% vs. 5%, respectively). In comparing the positive and negative predictive values, the higher prevalence produces a pool of positive animals that are nearly all truly positive while 17% of the test negative animals are false negatives. In the lower prevalence herd, while the test more closely predicted the true prevalence and the true negative status of the test negative animals, 21% of the animals that tested positive are actually not diseased. This demonstrates the importance of truly understanding what the cut-point is established. While practitioners have very little to do with the establishment of the cut-point, it is extremely valuable for us to understand what shifting the cut-point up or down will have on the identification of diseased animals.

<table>
<thead>
<tr>
<th>Test status</th>
<th>True status</th>
<th>Predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>154</td>
<td>97%</td>
</tr>
<tr>
<td>Negative</td>
<td>146</td>
<td>83%</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>1000</td>
</tr>
</tbody>
</table>

Table 2. Positive and Negative Predictive Values for a Hypothetical 1000-Cow Herd with a 30% True Prevalence

<table>
<thead>
<tr>
<th>Test status</th>
<th>True status</th>
<th>Predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>26</td>
<td>7</td>
</tr>
<tr>
<td>Negative</td>
<td>24</td>
<td>943</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>950</td>
</tr>
</tbody>
</table>

Table 3. Positive and Negative Predictive Values for a Hypothetical 1000-Cow Herd with a 5% True Prevalence
point that a test is more reliable in a herd with a higher prevalence. In order to see how prevalence affects PPV and NPV, see Figure 1.

The Likelihood Ratio is the likelihood that a given result would be expected in a diseased animal versus the likelihood that the same result would be expected in a non-diseased animal. The positive Likelihood Ratio ($LR^+$) is the likelihood that a positive result will be found in a diseased animal versus a non-diseased animal. The equation to determine the $LR^+ = \text{Sensitivity} / (1 – \text{Specificity})$. From the scenario from Table 1, (sensitivity = 51.4% and specificity = 99.3%) the $LR^+ = 0.514/(1 – 0.993) = 73.4:1$.

The negative Likelihood Ratio ($LR^-$) is the likelihood that a negative result will be found in a non-diseased animal versus a diseased animal. The equation to determine the $LR^- = (1 – \text{Sensitivity})/\text{Specificity}$. From the scenario from Table 1, (sensitivity = 51.4% and specificity = 99.3%) the $LR^- = (1 – 0.514)/0.993 = 0.49:1$.

![Positive and Negative Predictive Values](image)

**Figure 1.** Interaction between PPV and NPV as prevalence changes.

### CRITICAL APPRAISAL OF DIAGNOSTIC TESTS

As a practicing veterinarian, why do I care how well a test performs as we just send the sample to the state diagnostic lab and they run the test of their choosing? In many parts of the country that may be fine. However, some of the diagnostic test manufacturers are talking directly to your clients and more small reference labs are popping up in our communities offering diagnostic services either directly to your clients or to your clients through a local association. In some of these labs, the diagnostic test is completed and supported by a tech that was trained by the company that marketed the test with little understanding of what the outcomes actually mean or what the outcomes should be. With that said, many of these labs do not know what it a positive or negative outcome really means in a population of animals.

Critical appraisal of manuscripts evaluating diagnostic tests can be much harder to complete in comparison to evaluating the quality of a research manuscript from a randomized clinical trial, even for seasoned epidemiologists. To help epidemiologists and practicing veterinarians evaluate the quality of a diagnostic test assessment, there are several guidelines/checklists available that take us in a step by step evaluation through the manuscript. Probably the best available guideline/checklist for veterinarians assessing diagnostic test manuscripts is QUADAS (Available at: http://www.biomedcentral.com/1471-2288/3/25). "The QUADAS tool (Figure 2) was developed to provide an evidence based quality assessment tool of diagnostic accuracy studies." It is a 14-item questionnaire that evaluates the report for patient spectrum, reference standard, disease progression bias, verification bias, review bias, clinical review bias, incorporation bias, test execution, study withdrawals, and indeterminate results. When utilizing this tool to evaluate studies of a new test, nine of the fourteen items concentrate on identifying potential bias introduced by the study authors as bias will limit the validity of the study more than issues of variability and reporting.

The 14 items in the excerpt below (Figure 2) are taken verbatim from the original QUADAS description to provide the reader with as much information as possible in reviewing diagnostic articles. For more information about using the QUADAS tool, readers are referred to Whiting or O’Connor.

### SUMMARY

Most veterinarians already practice forms of Evidence-Based Veterinary Medicine on a regular basis. The issue that most of us get into overtime is that we get comfortable with what we’ve done in the past and become complacent about completing objective outcomes based assessments. For individuals who have limited access to multiple journals electronically, getting access to new unbiased evidence to apply to evidence based medicine is difficult and cumbersome. As communication technology continues to improve, access to multiple electronic evidence sources will become easier. Likewise, as diagnostic technology continues to advance, in areas such as molecular diagnostics for example, numerous new tests are being introduced based on technology that many veterinarians do not fully comprehend. These advances in technology will require veterinarians to search for the evidence to support the use of these newer, more expensive tests in practice. Unfortunately, this type of well reported, objective data has not been as readily available in the veterinary literature as compared to human medical literature. As reporting guidelines such as the REFLECT statement are adopted by manuscript authors, this type of literature should improve and become more abundant. Now the question is: Are you ready to resist the urge to stick your head in the sand and strive to be a lifelong learner by applying Evidence-Based Veterinary Medicine techniques to your daily practice?
Figure 2. The QUADAS Tool

- Item 1. Was the spectrum of patients representative of the patients who will receive the test in practice?
- Item 2. Were selection criteria clearly described?
- Item 3. Is the reference standard likely to correctly classify the target condition?
- Item 4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
- Item 5. Did the whole sample or a random selection of the sample, receive verification using a reference standard?
- Item 6. Did patients receive the same reference standard regardless of the index test result?
- Item 7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
- Item 8. Was the execution of the index test described in sufficient detail to permit replication of the test?
- Item 9. Was the execution of the reference standard described in sufficient detail to permit its replication?
- Item 10. Were the index test results interpreted without knowledge of the results of the reference standard?
- Item 11. Were the reference standard results interpreted without knowledge of the results of the index test?
- Item 12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
- Item 13. Were uninterpretable/ intermediate test results reported?
- Item 14. Were withdrawals from the study explained?

QUADAS publisher’s note:\(^8\): “This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article’s original URL.”

REFERENCES