The art and mathematics of making a useful diagnosis

D. Smith, DVM, PhD, Dipl. ACVPM, Dipl. Epidemiology
Mississippi State University College of Veterinary Medicine Mississippi State, MS 39762

Abstract
Veterinarians make a diagnosis to summarize the state of health of an animal. A diagnosis explains the animal's clinical signs and helps to establish a prognosis and plan of action. Unfortunately, there is no foolproof way to make a diagnosis, regardless of experience, and, unfortunately, diagnostic errors are common. The steps to making an accurate diagnosis involve both art and basic mathematical skills. The art of the diagnosis is in conducting a complete history and physical examination, developing a comprehensive list of differential diagnoses, assigning reasonable likelihoods for those differentials, and recognizing and avoiding cognitive biases. Mathematics are used to combine test performance data and disease likelihood to understand if a diagnostic test is more likely to improve diagnostic accuracy or more likely to be misleading. The nature of diagnostic error is predictable. Tests for rare conditions have problems with poor positive predictive value, whereas tests for common conditions have problems of poor negative predictive value. Strategies for using multiple tests, such as testing in series or in parallel, may improve the likelihood of getting test results that reflect the animal's true condition.

Key words: diagnosis, predictive value, diagnostic test, screening test, cognitive bias

Introduction
Veterinarians make a diagnosis to summarize the state of health of an animal. A diagnosis explains the animal's clinical signs and helps to establish a prognosis and plan of action.[10] To make a diagnosis, the veterinarian collects information about the individual animal and, in some cases, the population from which the animal belongs. This fact-finding mission includes asking about clinical signs observed by the animal's caregiver, acquiring a pertinent history, and conducting a physical exam. At this point, the veterinarian may be consciously or unconsciously developing a list of possible diagnoses and considering which laboratory tests might aid in making the diagnosis. This process might be simplified among experienced clinicians who rely heavily on previous experience and pattern recognition; however, the process might be more complicated among novice clinicians or in uncommon situations.[7] The move toward the faster and simpler process of pattern recognition is an example of the clinician's thought process moving from the slower and more critical thinking of System 2 to the more rapid and efficient thinking of System 1, as experience is gained and the clinician becomes more comfortable using rules of thumb.[3] System 1 is especially efficient for making decisions rapidly in the moment -something clinicians often need to do. However, sometimes, before the crisis, System 2 can help one think through anticipated problems[3] – like making a diagnostic error.

There is no foolproof way to make a diagnosis, regardless of experience, and unfortunately, diagnostic errors are common.[1,2] Diagnostic error occurs when a diagnosis is 1) not made; 2) delayed; or 3) wrong.[1] Diagnostic errors occur because of 1) the lack of knowledge; 2) cognitive errors; or 3) poor understanding of the concepts surrounding diagnostic interpretation. The lack of knowledge may be either on the part of the veterinarian or the current state of medical knowledge. Cognitive errors result from failure to consider all possibilities or from either over- or under-estimating the likelihood of a disease process. In effect, cognitive errors are about being under- or over-imaginative. The art of diagnostics is about avoiding cognitive errors by realistically imagining the likelihood of an animal having a certain condition. The mathematics of diagnostics are about making an appropriate interpretation of a diagnostic test result by estimating the predictive value of a positive or negative test result.

A diagnostic test is any technique used to help distinguish between different states of a patient's health (e.g. normal and abnormal, or pregnant or not pregnant).[7] These may be laboratory-based tests or clinical diagnostic techniques, such as radiology, but also include some clinical examination procedures (e.g., rectal palpation to detect specific indicators of pregnancy). Veterinarians use diagnostic tests to detect, confirm, document or rule out a condition. The goal of the investigation may be to determine the cause of a condition, provide prognostic information, guide therapy, or document the effectiveness of a therapy (e.g., the clearance of an infectious agent). On a population basis, diagnostic tests are used to screen apparently healthy individuals for a condition, prevent infected individuals from entering into a herd, monitor the occurrence of a condition in a herd, or document the effectiveness of biosecurity of biocontainment programs.[7,9]

Cognitive bias
Cognitive biases are systematic errors in the clinical thought process.[6] These errors in cognitive reasoning may be subtle and there is no certain way to avoid them. For example, the cognitive error of failing to consider rare diseases is the opposite of the cognitive error of over-estimating the likelihood of a rare disease. In trying to overcome the one cognitive error, one risks making the other. The risk for making cognitive errors while making a diagnosis might be avoided by taking a moment after collecting history and physical examination data to consider:

- If this condition is not what I've tentatively diagnosed, what else could it be?
- What are the worst things this could be?
- What evidence is at odds with my tentative diagnosis?

Following this stage of self-reflection, the differential diagnosis list might include conditions that may not have initially been considered and may point to the need to collect additional information, including from the use of diagnostic tests.[6]

Diagnostic test interpretation
It is not always helpful to run a diagnostic test. In fact, diagnostic tests can make matters worse by leading the clinician down a path of misdiagnosis and mis-directed therapy. Before

choosing to use a test, the clinician should consider whether:

- The condition is likely or unlikely
- There are diagnostic errors associated with the test
- The test will help to distinguish conditions
- There are any adverse health risks associated with running the test
- The results of the test would alter the treatment plan

The first two points, the likelihood of the condition and the diagnostic errors associated with the test, are the components of estimating the predictive value of a test – whether the test results are likely to represent the true health status of the animal. The other points are also practical considerations for whether the test results will be clinically useful. For example, recovering a pathogen from the tissue of a sick animal means something different than finding serological evidence of exposure to the same agent that occurred at some point during the animal's life. Also, the likelihood of finding an agent among animals with and without the condition should be considered before deciding that the agent is the cause of the condition. For example, culturing generic Escherichia coli from the feces of a calf with diarrhea is not convincing evidence that the organism is responsible for the diarrhea, because it is equally likely to be recovered from the feces of any normal calf. Sometimes the stress or risk of injury from collecting the sample or restraining the animal can be more detrimental than the value of the test. Finally, a diagnostic test may have little value if the results are unlikely to change the course of therapy.

The performance of diagnostic tests is evaluated by parameters of sensitivity and specificity.[4,5] Sensitivity (SENS) is the conditional probability of testing positive, *given that the individual truly has the condition*.

$$SENS = P(T+| D+)$$

Specificity (SPEC) is the conditional probability of testing negative, *given that the individual truly does not have the condition.*

$$SPEC = P(T- |D-)$$

By themselves, SENS and SPEC provide no directly useful information to the clinician because the parameter only apply to the condition of knowing the status of the animal's condition. However, these test performance statistics are useful for test interpretation when they are used in combination with an accurate pre-test estimate of the likelihood of the animal having the condition. Accurately estimating the pre-test probability of an animal having a condition is an art that improves with experience and the ability to reasonably overcome cognitive biases. Pre-test probability (PD) can be thought of as the proportion of animals having the condition from an imaginary population of animals all with the same clinical presentation and history as the animal under evaluation. We estimate PD from what we know from the literature, other epidemiologic knowledge, the history and physical examination results of the animal, and previous experience.

The astute clinician is interested in knowing how likely a test result is to reflect the true condition of the animal. This is called the post-test probability or the predictive value of a test.[4,5] If the result is not likely to represent the condition, then perhaps the test should be avoided or, at least, some results should be viewed skeptically (Figure 1). Positive predictive value (PPV) is the conditional probability of an animal having a condition, *given that the test was positive.*

$$PPV= P(D+ |T+)$$

Negative predictive value (NPV) is the conditional probability of an animal not having the condition, *given that the test was negative.*

$$NPV= P(D- |T-)$$

Both PPV and NPV are functions of the two test performance statistics (SENS and SPEC) and PD.[4]

$$PPV = \frac{SENS \times P_D}{(SENS \times PD) + (1–SPEC) \times (1–P_D)}$$

The formula for PPV is not as complicated as it looks. It is the probability for a true positive divided by the sum of probabilities for a true positive and a false positive.

$$NPV = \frac{SENC \times (1–P_D)}{SPEC \times (1 –PD) + (1–SENS) \times P_D}$$

Similarly, the formula for NPV is the probability for a true negative divided by the sum of probabilities for a true negative and a false negative.

For a given individual, the test result either represents the true condition of the animal, or it does not. We can think of PPV and NPV as the proportion of that imaginary population of animals with the exact set of history and physical examination findings that would have truly had the condition (in the case of PPV), or not had the condition (in the case of NPV), given they had tested positive or negative, respectively. Therefore, PPV and NPV tell us how likely the test result is to be believable regarding the condition of the animal. This is not trivial. Any non-perfect test -and they are all non-perfect -could have extremely high or extremely low predictive value depending on the PD. Understanding this helps the clinician know whether to trust a positive or negative test result, or whether the test is more likely to be misleading than be helpful in a given situation. The nature of the diagnostic error is predictable. In circumstances with low PD, such as when apparently healthy animals are screened for a disease, then PPV is low. In this situation, positive test results should be suspect because of the number of false positive test results is high relative to the number of true positives. On the other hand, in circumstances when PD is high, such as when testing beef cattle for pregnancy in a herd with a good reproduction program, then NPV is low. In this situation, negative test results should be suspect because of the number of false negatives is high relative to the number of true negatives.

Improving test performance with multiple test strategies

A combination of tests may be used together to improve either test sensitivity at the cost of specificity, or to improve specificity at the cost of sensitivity. These strategies may help in situations where a single test has poor positive or negative predictive value.

Using 2 or more tests concurrently and considering animals that test positive to any of the tests as having the condition is called testing in parallel. Parallel testing increases test sensitivity and results in fewer-false negative results, but more false positive results because of lowered specificity. Parallel testing is used when missing a condition is more costly than having false-positive test results. Parallel testing is useful for ruling out a condition, for example when the NPV of a single test is low. The animal "proves" it does not have the condition by testing negative to multiple tests. "sNout" is the mnemonic used to remind us that negative results from very sensitive test strategies help to rule OUT a condition.[8]

For two independent tests:

$$SENSP = 1- [(1-SENS_A) \times (1-SENS_B)]$$

$$SPEC_P = SPEC_A \times SPEC_B$$

Using 2 or more tests sequentially based on positive test results on the previous test is called serial testing. Typically, the subsequent tests are chosen because they are more specific. Animals that test positive on all tests are considered to have the condition. Serial testing increases test specificity and results in fewer false positive results, but more false-negative results because of lowered sensitivity. Serial testing is used when the consequences of a false positive test result are more costly than the consequences of missing a condition. For example, when a positive test result might lead to euthanasia of the animal, or where a single positive test result might classify an entire herd as having a condition. Serial testing is useful for ruling in a condition, for example, when the PPV of a single test is low. This is the typical situation when screening apparently healthy individuals for rare conditions. The animal "proves" it has the condition by repeatedly testing positive to increasingly specific tests. "sPin" is the mnemonic used to remind us that positive test results from very specific test strategies help to rule IN a condition.[8]

For two independent tests:

$$SENS_S = SENS_A \text{ x } SENS_B$$

$$SPECS = 1 - [(1-SPEC_A) \text{ x } (1-SPEC_B)]$$

Conclusion

Veterinarians are in the business of making medical diagnoses. However, making an accurate and useful diagnosis involves both art and mathematics. Diagnostic prowess requires the veterinarian to be skilled in the art of physical examination, collecting a good history, and being aware of their cognitive biases. Basic mathematical skills applied to calculating the predictive value of a diagnostic test help the veterinarian gain insight into when a test will aid or hinder an accurate diagnosis.

Acknowledgement

A contribution of the Beef Cattle Population Health and Reproduction Program at Mississippi State University. Supported by the Mikell and Mary Cheek Hall Davis Endowment for Beef Cattle Health and Reproduction. This work is supported by the Agriculture and Food Research Initiative Competitive Grants Program grant no. 2018-69003-28706 from the USDA National Institute of Food and Agriculture. Any opinions, finding, conclusions, or recommendations expressed in this paper are those of the author and do not necessarily reflect the view of the U.S. Department of Agriculture.

References

1. Diagnostic Errors: Technical Series on Safer Primary Care. Geneva: World Health Organization, 2016.

2. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf* 2013;22 Suppl 2:ii21-ii27.

3. Kahneman D. *Thinking fast and slow*. New York: Farrar, Straus and Giroux, 2011.

4. Martin SW. The evaluation of tests. *Can J Comp Med* 1977;41:19-25.

5. Martin SW. Estimating Disease Prevalence and the Interpretation of Screening Test Results. *Prev Vet Med* 1984;2:463-472.

6. McGee D. Cognitive errors in clinical decision making. *Merck Manual*, 2018.

7. Radostits OM, Mayhew I, Houston D. *Veterinary Clinical Examination and Diagnosis*. 1st ed: Saunders Ltd., 2000.

8. Sackett D, Straus S, Richardson W, et al. *Evidence-based medicine. How to practice and teach EBM*. New York: Churchill Livingstone, 2000.

9. Smith DR. Epidemiologic tools for biosecurity and biocontainment. *Vet Clin North Am Food Anim Pract* 2002;18:157-175.

10. Smith DR. I'm not positive that's a positive. Become indispensable to your clients by knowing how to choose, interpret and incorporate diagnostic testing in bovine practice. Conference of the Am Assoc Bov Pract 2018;126-133.

Figure 1:
The predictive value of positive and negative test results over the full range of pre-test probabilities with different test performances. Graph A: Sensitivity = 1, Specificity = 0. This is a non-informative test. All tests are called positive. Graph B: Sensitivity = 0, Specificity = 1. This is a non-informative test. All tests are called negative. Graph C: Sensitivity = 0.5, Specificity = 0.5. This test is non-informative. This is like flipping a coin to determine the test result. Graph D. Sensitivity = 0.9, Specificity = 0.9. This test is informative. This test provides predictive value that exceeds what was known before the test was conducted (pre-test probability). Note that even non-informative tests have high predictive value in some circumstances and even informative tests have poor predictive value in some circumstances. Regardless of test performance, positive test results are suspect when the condition is rare and negative test results are suspect when the condition is common.
PD = Pre-test probability of the disease. This is the probability of the animal having the condition estimated before the test results were known.
PPV= positive predictive value. This is the probability an animal with a positive result has the condition.
NPV = negative predictive value. This is the probability that an animal with a negative result does not have the condition.