# **Module 5: Use of Diagnostic Tests**

Contributors: Scott Wells, Amy Kinsley, Sandra Godden

In this module, you will explore the following:

1. What factors affect predictive values? Sensitivity, Specificity, and Prevalence
2. Interpreting test results in a real-world situation without knowledge of true disease status (No gold standard test available and unknown true prevalence)
3. How can I increase my confidence in these test results? How to improve predictive values?
4. How do I decide when to apply a test? - Pre- and post-test probability of disease

## **Factors affecting the Predictive Values of Test Results**Predictive values are affected by three factors:

1. **Test sensitivity**: As we increase test sensitivity, we reduce the number of false negatives and improve predictive value of a negative test. In other words, if we get a negative test result, we are more inclined to trust it, given that a highly sensitive test yields few false negatives.
2. **Test specificity**: As we increase test specificity, we reduce the number of false positives and improve the predictive value of a positive test. In other words, if we get a positive test result, we are more inclined to trust it, given that a highly specific test yields few false positives.
3. **True prevalence of the disease in the population being tested**: As a disease becomes more common (e.g., higher prevalence), the predictive value of a positive test increases and the predictive value of a negative test decreases. The reverse is true if the disease is rare (e.g., low prevalence): If the prevalence is low, the predictive value of a positive test decreases and the predictive value of a negative test increases. This is observed over time in disease eradication programs: As the prevalence of the disease decreases over time, the predictive value of a positive test will fall, and the predictive value of a negative test will increase.

Test sensitivity and specificity generally remain constant for a diagnostic test (are not influenced by prevalence). However, the prevalence of disease can vary over a wide range in different populations. This prevalence can have a very large effect on the predictive values of your test results.

### **Case 5.1. Effect of prevalence on predictive values in individual animal medicine**

After graduation, you decide not to live in Florida. Instead, you live and practice, quite contentedly, in Minnesota. You are doing Heartworm testing in late May, using the SNAP PF Heartworm antigen test kit, on client dogs in your practice. Your first dog tested yields a positive test result. Your second dog tested yields a negative test result.

Now, fill in the 2x2 table below to recalculate the predictive value of both a positive and negative test when using this test on Minnesota dogs.

What do you need to know to estimate the predictive values of positive or negative tests?

* Test sensitivity = 67% (unchanged from Florida study in Case 3.2)
* Test specificity = 98% (unchanged from Florida study)
* Prevalence = \_1%\_ (theoretical, much lower than for Florida, right?)
 Assume 1,000 fictitious dogs in your client population to fill in the table below.

Steps to fill in the table:

1. Construct a fictitious population of 1,000 dogs with 10 (1%) being truly infected (a+c). Thus 990 must be truly uninfected (b+d).
2. Calculate 67% (sensitivity) of 10 and fill in the true positive cell a (7 true positives).
3. Calculate 98% (specificity) of 990 and fill in the true negative cell ‘d’ (970 true negatives).
4. Fill in remaining cells (calculate the differences).



***Predictive Value of a Positive Test (PV+ve)***

The proportion of test-positive animals that truly are diseased.
or the probability, given a positive test result, that the animal actually has the disease.

= a / (a + b) = \_\_\_\_\_

***Predictive Value of a Negative Test (PV-ve)***

The proportion of test negative animals that truly are not diseased.
or the probability, given a negative test result, that the animal does not have the disease.

 = d / (c + d) = \_\_\_\_\_

Before you initiate treatment, what can you do to increase your confidence in a Positive test result from a MN dog? One answer is to use multiple tests: Repeat the blood test using a different test kit (blood test or radiography) with a higher specificity.

Answers:

|  | True Disease Positive | True Disease Negative | Total |
| --- | --- | --- | --- |
| Test Positive | a 7 | b 20 | a + b 27 |
| Test Negative | c 3 | d 970 | c + d 973 |
| Total | a + c 10 | b + d 990 | n 1000 |

***Predictive Value of a Positive Test (PV+ve)***

The proportion of test-positive animals that truly are diseased.
or the probability, given a positive test result, that the animal actually has the disease.

= a / (a + b) = 7/27 = 25.9%

***Predictive Value of a Negative Test (PV-ve)***

The proportion of test negative animals that truly are not diseased.
or the probability, given a negative test result, that the animal does not have the disease.

 = d / (c + d) = 970/973 = 99.7%

|  | Florida Clinic (from previous session) | Minnesota Clinic  |
| --- | --- | --- |
| True Prevalence  | 59% | 1% |
| PV +ve | 98% | 26% |
| PV -ve | 67% | 99.7% |

### **Case 5.2: Example of how prevalence affects predictive values in an eradication program.**

You are working with a producer on a test and cull program to eradicate *Neospora caninum* infection from his dairy herd (an intracellular protozoal parasite that causes mid-to-late gestation abortion in 5-20% of infected animals). The program involves an annual serological screening (ELISA) of the entire herd to detect animals with a positive antibody titer. Antibody-positive animals are presumed infected and will be culled from the herd as soon as is economically feasible. The **sensitivity** of this ELISA test is known to be **89%** and the **specificity** is **97%**. The within-herd **true prevalence of Neospora caninum is 35%.** The true prevalence is also called the **pre-test probability of disease**.

Calculate the predictive value of a positive test and the predictive value of a negative test.

The PV+ve = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. This means that there is a \_\_\_% probability that an animal testing positive will truly be infected.



The PV-ve = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. This means that an animal testing negative has a \_\_\_% probability of truly being disease-free.

Are you pretty confident either keeping, or culling, animals based on a negative or positive test result, respectively?

**Answers:**

The PV+ve = \_94% (312/331). This means that there is a 94% probability that an animal testing positive will truly be infected.

The PV-ve = \_94% (631/669). This means that an animal testing negative has a 94% probability of truly being disease-free.

Are you pretty confident either keeping, or culling, animals based on a negative or positive test result, respectively? Yes, in this situation, you are pretty certain (94% probability) of the infection status of both test-positive and test-negative cattle.

Three years later the producer has been able to cull several infected animals, reducing the true herd prevalence to **5%**.

Recalculate the predictive values for this test.



The PV+ve = \_\_\_\_\_\_\_\_\_. The PV-ve = \_\_\_\_\_\_\_\_\_.

What happened to the predictive value of the test when the prevalence of the disease decreased to 5%?

Are you still very confident in either keeping, or culling, animals based on a negative or positive test result, respectively?

What will be the consequences of lower positive predictive values in an eradication program?

If this is a real eradication program, what do you do now to improve your predictive values?

**Answers:**

The PV+ve = 44.5/73 = 70%.

The PV-ve = 921.5/927 = 99.4%.

What happened to the predictive value of the test when the prevalence of the disease decreased?

* PPV decreased from 94% to 70%. NPV increased from 94% to 99%.

Are you still very confident in either keeping, or culling, animals based on a negative or positive test result, respectively?

* Extremely confident in negative test results, but less confident in positive test results (30% of animals that test-positive are false positive).

What will be the consequences of lower positive predictive values in an eradication program?

* This is likely to result in culling of test-positive uninfected animals unnecessarily, which is an economic loss to the producer.

If this is a real eradication program, what do you do now to improve your predictive values?

* Confirm positive test results with a second test with higher specificity (confirmatory test).

## **Interpreting test results in a real-world situation without knowledge of true disease status**

## (No gold standard test available and unknown true prevalence)

### **Case 5.3. The same case in a more real-world situation.**

In the real-world, we unfortunately do not know the true prevalence of disease (or infection) in different populations. We can often estimate the sensitivity (Se) and specificity (Sp) of the test used from peer-reviewed scientific literature (or from the test manufacturer, which is often less reliable). After testing in the herd, you also can calculate the **Apparent Prevalence** (sometimes called **Test Prevalence** = the proportion of animals tested that test positive).

**How to estimate predictive values in this situation?** After testing the cattle in the herd (the original situation with 331 test-positive cattle above), the information you know at the outset:

* ELISA Se = 89%
* ELISA Sp = 97%
* Apparent prevalence = 331/1000 = 33.1%



You can estimate predictive values when you have an estimate of the Se and Sp of the test used and test results from using that test in the population. The next step is to calculate the **Estimated True Prevalence** of disease (or infection) using the following formula (called the **Rogen-Gladden calculator**):

#### Estimated true prevalence

 = (Apparent prevalence + Sp – 1) = (0.331 + 0.97 – 1)

 (Se + Sp – 1) (0.89 + 0.97 – 1

= 0.301 / 0.86 = 35%.

Remember 35% was the true prevalence given to you in that situation above? You now use this **Estimated true prevalence** value to fill in the column totals in the 2x2 table below (1,000 total animals x 35% = 350 animals True disease positive). You can use subtraction to calculate the total number of True disease negative animals (1000 - 350 = 650).



The next step is to fill in the table above using the Se and Sp estimates. By definition, 89% Se means that 89% of the True disease positive animals (n=350) will test positive (350 x 0.89 = 312), so place 312 in the upper left cell of the table (yellow highlighted cell), and use subtraction to fill in the other cells (subtract from the row or column totals).



Notice that these values are the same as those in the previous situation (early stages of eradication program), the difference being that this time you used the test Se and Sp along with the test results to fill in the cells of the 2x2 table (without an estimate of the true prevalence ahead of time). Now you can estimate the positive and negative predictive values.

Positive Predictive Value =

Negative Predictive Value =

Answer:

Positive Predictive Value = 312/331 = 94%

Negative Predictive Value = 631/669 = 94%

**Note**: This is the typical situation you will face in veterinary practice, as you typically don’t have a gold standard test (100% Se and 100% Sp) available and therefore don’t know the true prevalence in the animal population. You can instead estimate the true prevalence using the formula above (**Rogen-Gladden calculator**).

Two final points to mention:

* First, **you do need to have an estimate of the Test Se and Sp** to calculate predictive values. If these estimates are not available from the scientific literature, ask the test manufacturer for these values. You will not be able to interpret test results with confidence unless you understand how well the test performs.
* Second, there are some situations in which the formula to estimate true prevalence (Rogen-Gladden calculator) does not work. If the Apparent (test) prevalence is less than the expected proportion of false-positives in a population (1 – Sp), then the estimated true prevalence will result in a negative value (error). This makes sense if you consider that the expected proportion of animals testing positive in an uninfected population (0% true prevalence) should be (1 – Sp), the proportion of false-positives. So if the proportion of test-positives in your population (the apparent prevalence) is less than the expected proportion of false-positives, then the test Sp may not be correct. We expect some random variation of the Se and Sp values due to biologic variation, even if the values used are generally close to the true value, but remember that, in some situations, the values of Se and Sp may not allow estimation of true prevalence using the Rogen-Gladden calculator.

## **How can I increase my confidence in these test results? Methods for Improving Predictive Values of Test Results**

Remember: Predictive value is affected by sensitivity, specificity and prevalence. Examples of how to improve predictive value of a positive test could therefore include:

### Change the sensitivity or specificity of the test

* + **Select a different cut-point for considering a test positive**. Depending on your goals, you could change the test cut-point to either increase sensitivity (may decrease specificity) or to increase specificity (may decrease sensitivity). This can work for tests with numerical test results.
	+ **Change the conditions of the test**. As an example, when performing milk cultures to detect presence of mastitis pathogens, plating a larger volume (0.1 ml) of milk (instead of 0.01 ml of milk) increases the sensitivity for detecting *Staph. aureus* infections. This however also decreases specificity, meaning that more false positive results may occur.

### Increase the pre-test probability of disease

Pretest probability of disease = the true prevalence in the population of interest.By selectively testing only populations known to have a higher prevalence of disease, you will increase your predictive value of a positive test result (but will also decrease your predictive value of a negative test result). As an example, to increase the PPV of the serum ELISA tests for Johne’s disease, you could restrict testing to cows in 2nd or greater lactation, but not younger cattle (including calves and heifers). The expected prevalence is higher in older cattle, resulting in a higher positive predictive value in those cattle compared to younger cattle.

### Use two or more tests in combination

Predictive values are best if we use a very sensitive and very specific test. However, it may not be possible to apply the most highly sensitive and specific test as a screening test (too costly, inconvenient, invasive, or other). At other times, it is most logical to use multiple tests for economic reasons or time constraints. In that case, a choice can be made between applying **sequential testing** (one test followed by another) or **concurrent testing** (applying two or more tests at the same time).

#### **Sequential Testing**

##### a.1 Herd Retest: Repeated Testing Using the Same Test:

At the first testing, the test-positive animals are treated or culled. At the next testing, only the previous test-negative animals (those remaining in the herd) are re-tested. The retest is performed using the same test each year (may be more or less frequent depending on the disease). This strategy is often used used in federal or state eradication programs (e.g., test and slaughter of bovine tuberculosis test-positive cattle). It is also commonly used in monitoring and control programs to prove that a herd or animal is free of a particular disease (e.g., annual geriatric profile in older dogs or cats).

##### a.2 Sequential Testing Using Two or More Different Tests:

We apply a screening test first, then apply a diagnostic or confirmatory test next, but only to those animals which initially tested positive. Often the overall goal is to identify as many diseased animals as possible. In this case, the initial screening test should have high sensitivity (to minimize false negatives) and is often cheaper, but may have slightly lower specificity (a few false-positives will be detected) than another test. The second test may be more specific (can “rule-in” disease in those animals that initially tested positive, so avoiding false-positives) and is often more expensive (but used in fewer animals in this situation to manage the testing costs).

##### **Case 5.4. Example of sequential testing**

We will use the problem of detecting antibiotic residues in milk. Let’s say the prevalence of antibiotic residues in milk is 10% and there are two tests available for detecting milk antibiotic residues.

* The first test, called the BacT test, is based on bacterial growth inhibition (the Delvo-P test).
* The second test is based on spectrophotometric methods where the antibiotic is known to absorb light at a specific wavelength (the SpecT test).

 BacT SpecT

Sensitivity 95 % 96 %

Specificity 90 % 99.5 %

The SpecT test is obviously a better test, but is much more expensive to run, prohibiting its use on a large scale.

***Step 1.******Apply the BacT (screening test) to all samples:***



PPV = / = \_\_\_\_\_\_\_\_

NPV = / = \_\_\_\_\_\_\_\_

We have a problem with false positives (low predictive value of a positive test). Therefore, to improve the overall positive predictive value, we need to select a test that is more specific and perform it on the test-positive samples only.

***Step 2. Apply the SpectT (confirmatory test) to only the 185 samples positive on the BacT***:



PPV = / = \_\_\_\_\_\_\_\_\_

NPV = / = \_\_\_\_\_\_\_\_\_

**What is the overall sensitivity and specificity** after using the two tests sequentially?

91 truly affected samples tested pos. on both tests. The overall sensitivity = \_\_\_\_\_\_\_ = \_\_\_\_\_%.

We discarded 810 samples that were truly negative with the first test and 89 that were truly negative with the second test. The overall specificity = \_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_ %.

By using these two tests sequentially we have reduced our testing costs tremendously while maintaining overall good sensitivity and very high specificity. High specificity is very important here since a false-positive result could result in dumping an entire tanker load of milk, meaning significant financial losses to the producer.

**Answers:**

***Step 1.******Apply the BacT (screening test) to all samples:***



PPV = 95 / 185 = 51.4%

NPV = 810 / 815 = 99.4%

We have a problem with false positives (low predictive value of a positive test). Therefore, we need to select a test that is more specific and perform it on the test-positive samples only.

***Step 2. Apply the SpectT (confirmatory test) to only the 185 samples positive on the BacT***:



PPV = 91 / 92 = 98.9%

NPV = 89 / 93 = 95.7%

**What is the overall sensitivity and specificity** after using the two tests sequentially?

|  | Residue positive | Residue negative | Total |
| --- | --- | --- | --- |
| Positive on both tests | 91 | 1 | 92 |
| Negative on at least one test | 9 | 899 | 908 |
| Total | 100 | 900 | 1000 |

There were 92 samples that tested positive on both tests. Overall, we estimate that 91 samples of these samples were truly affected. . The overall sensitivity (from use of both tests in sequence) = 91/100 = 91%

We discarded 815 samples that were test-negative with the first test and 93 that were test-negative with the second test, for a total of 908 samples with at least one test-negative sample among the two tests. The overall specificity = 899 / 908 = 99.0%

By using these two tests sequentially, we reduced our testing costs tremendously while maintaining overall good sensitivity and very high specificity. High specificity is very important here since a false-positive result could result in dumping an entire tanker load of milk, meaning significant financial losses to the producer.

#### **Concurrent testing**

Both tests are used at the same time.

 Test1 (T1) Test2 (T2)

Sensitivity 70 % 66 %

Specificity 86 % 88 %



##### **b.1 Series Interpretation**:

 Only accept an animal as ‘diseased’ if the results are positive for **both** tests.

Result: Improves specificity, lowers the number of false-positives, and improves the predictive value of a positive test. Similar to Sequential testing but costs more money, since it involves running all tests at same time (compared to Sequential where you wait and only re-run second test only on samples positive from first test).



What is the overall test Se and Sp of the use of concurrent tests in this way?

What are the overall positive and negative predictive values in this situation?

Sensitivity = \_\_\_\_\_\_\_\_\_\_ PPV = \_\_\_\_\_\_\_\_\_\_

Specificity = \_\_\_\_\_\_\_\_\_\_ NPV = \_\_\_\_\_\_\_\_\_\_

**Answers:**

Sensitivity = 50% PPV = 92.6%

Specificity = 96% NPV = 65.8%

##### **b.2 Parallel Interpretation**:

 Accept the animal as ‘test positive’ if the results are positive for at least one test (either or both tests).

Result: Increases the test sensitivity, decreases the number of false negatives, and improves the predictive value of a negative test. Using the same data as the situation above:



What is the overall test Se and Sp of the use of concurrent tests in this way?

What are the overall positive and negative predictive values in this situation?

Sensitivity = \_\_\_\_\_\_\_\_\_\_ PPV = \_\_\_\_\_\_\_\_\_\_

Specificity = \_\_\_\_\_\_\_\_\_\_ NPV = \_\_\_\_\_\_\_\_\_\_

**Answers:**

Sensitivity = 86% PPV = 79.6%

Specificity = 78% NPV = 84.8%

#### **Summary from Using Tests in Combination**

1. We can use tests in combination to improve predictive values of test results.
2. Sequential and series testing are very similar – they both increase the Specificity of the two tests but decrease the Sensitivity.
3. Parallel Test interpretation increases the Sensitivity of the two tests but decreases the Specificity.
4. Which approach you choose depends on the relative consequence of false-positive and false-negative results, and which is more important for you to avoid:
	* If reducing the risk health risks of antibiotic residues to consumers is the most important, then we want to achieve maximal sensitivity to minimize the number of false-negatives (i.e., maximize predictive value of a negative test). In this case, using tests in parallel might be preferred.
	* If we want to minimize the risk of a false-positive resulting in a dumped load of milk, a substantial economic loss to the producer, then we would want to maximize specificity to minimize the number of false-positives (i.e., maximize the predictive value of a positive test). In this case, we could use sequential or series testing.

#### **Summary of Testing Strategies** (Smith, Chapter 4, Pg. 58)



## **How do I decide when to apply a test? - The Pre-test and Post-test Probability of Disease**

A test should be useful to both the clinician and the patient, or it isn’t worth doing. The application of an appropriate test should be based on whether the test result will alter the probability that a disease does or does not exist. The test result must have the potential to impact or alter the clinician’s decision or action.

***Ask yourself “If I perform this test, will the result alter my diagnosis or course of action?” If the answer is ‘NO’, the test may not be worth performing.***

**Pre-test (or prior) probability of disease:** The probability that an individual animal has a disease before a test is applied. For an individual randomly selected from the target population, this will be the same as the **true prevalence of the disease** in the target population. For an animal presenting to you with a given signalment and history (which are actually tests in themselves), the pre-test probability will be much higher for certain diseases. For example, the pre-test probability of a clinical intestinal parasite infestation in a dog picked at random from your practice population is low. However the pre-test probability of an intestinal parasite infestation in a 15 week old puppy that presents to you with loose stools is much higher.

**Post-test probability of disease** given a positive test is the same as the **predictive value of a positive test** (a / (a+b)). (e.g., the probability that our puppy with loose stools truly has a clinical intestinal parasite infestation, given a positive fecal test).

## **How we decide when to apply a diagnostic test?**

In general, we:

* Withhold treatment when the pre-test probability of disease is low.
* Treat empirically when the pre-test probability of disease is high.
* Let test results guide us in managing the case when the pre-test probability of disease is
 intermediate.

But, how low is low, and how high is high? Factors to consider in deciding on these thresholds include:

* What you believe the pre-test probability of the disease to be?
* The benefit of appropriate therapy (vs. the risk if left untreated = risk of a false-negative).
* The risk or cost of inappropriate therapy if treated for the wrong disorder (= risk of a false-positive).
* The time involved, risk, convenience, and cost of the test procedure.
* The sensitivity and specificity of the test.
* What you believe the predictive values of the test results will be (post-test probability of disease).

***In most cases, tests are MOST useful to clinicians when there is a large difference between the pre-test and the post-test probability of disease.***

The greatest difference between the pre- to post-test probability of disease will generally occur when the pre-test probability is between 40-60%. Put another way, a sign, symptom, or lab test is of greatest diagnostic use to us (has the greatest chance to change our diagnosis or course of action) when we are in a 50:50 dilemma and cannot decide whether or not the patient has the target disorder. We get to a pre-test probability of 40-60% through collecting information as to the signalment (age, sex, stage of lactation, etc.), history, and physical exam.

**Note**: The questions you ask when obtaining a signalment and history and the data you assemble while completing a physical exam are also (each of them) tests, as each new bit of information collected is quite possibly refining or directing your diagnosis.

## **Summary: Guidelines for Selection of a Diagnostic Test (Bonnett, 1990)**

### **Assessment of Test Accuracy and Consistency**

* Has the test been compared to an appropriate gold standard?
* Is the test (sample) population appropriate relative to prevalence of the target disorder, spectrum of disease (target disorder), and is there an appropriate range of other conditions among the group?
* Has test consistency (precision) been determined?
* Have Test sensitivity and specificity been determined, preferably at all levels of the test result? The effect of cut-points should have been determined and discussed. An appropriate definition of normal and abnormal should have been specified.
* Have predictive values and/or post-test likelihoods been computed for appropriate levels of prevalence or prior probability?
* Have procedures/conditions required to perform the test been specified?
* If tests are intended for use with other tests, the method of interpretation and contribution of individual tests to overall sensitivity and specificity should be described.

### **Assessment of Utility of a Test**

* Has the test been assessed on animals similar to those you would use it on? Have the sample population characteristics been described?
* Does the test offer advantages over the gold standard or currently used tests in terms of biological or practical costs? The consequences of misclassifications should be discussed (false-positives and false-negatives).
* Does the test have the potential to influence case management (diagnosis, prognosis, therapeutic, or culling decisions)?