

## Assay Validation Methods

*This document contains educational materials developed by Dr. Larry Hammell, and is used with permission. This information was adapted from sections of the 'Applied Aquaculture Epidemiology' course, developed by the Atlantic Veterinary College and Canadian Aquaculture Institute. The USFWS acknowledges and thanks Dr. Hammell for the use of this material and his contributions to quality assurance in applied aquaculture.*

To fully assess or validate the performance of a laboratory test, it is important to understand that what we observe, as clinicians, may not always represent the actual disease status of an aquatic population undergoing an epizootic or sub-clinical infection. Apparent prevalence versus true prevalence, and the factors that affect how well these measures correlate is the topic of this section.

Similarly, if we wish to develop a new technique to detect or monitor disease, the first critical step is to determine the objective of the test in terms of sensitivity and specificity, and make sure we select the most appropriate test to meet the objective.

For the majority of diagnostic applications, the test threshold will be established with balance to ensure that multiple objectives are met without disproportionate number of false test results in either direction (false-positives, or false-negatives). In the following sections, we will discuss the difference between apparent prevalence and true prevalence, the factors that affect sensitivity and specificity of a test, and measuring agreement between two tests.

### Apparent versus True Prevalence

There is an important difference between apparent prevalence and true prevalence of disease in an aquatic animal population.

Apparent prevalence is the number of animals testing positive by a diagnostic test divided by the total number of fish in the sample tested.

True prevalence is the actual number of diseased animals divided by the number of individuals in the population.

Sampling techniques and Quality Assurance measures in the laboratory always strives to produce data that are accurate and bring the values for apparent and true prevalence as close as possible. The better our tests are in terms of sensitivity and specificity, the better our test results will represent the true prevalence of disease in a population. The more accurately we can detect pathogens or measures of abnormal pathology, the more effective we can be at taking necessary steps to manage disease, and control the spread of disease among aquatic populations.

True prevalence may never really be known for a population, unless all animals in a population are tested with an assay that is 100% accurate, which is extremely rare in human, or veterinarian medicine health tests. For diagnostic testing, or fish health inspections in fish populations, lethal samples are needed to assess disease prevalence. Testing the entire population would be counterproductive, and probably not worth the effort for the additional level of accuracy obtained, when often we are only interested in knowing IF the disease is present. Even when we are presented with a natural or aquaculture disaster, that produces mortality of all animals, the logistics of sampling and testing every fish is costly and daunting to the fish health laboratory.

Because these tests are not 100% accurate, it is important for the fish health specialist to know how to interpret test results using his/her knowledge of how the various tests perform. Given a set of test results, we need to know what proportion is truly positive, and what proportion are false positives. This can be just as important for negative results as well, what proportion are truly negative, and what proportion are false negatives.

## Interpreting Diagnostic Tests

Often the best way to understand these measures is to look at a hypothetical situation in which your test is 100% accurate to discern what proportion are expected to be positive and negative, then compare this data to what proportion actually tests positive and negative. First, let's look at the hypothetical 100% accurate test, and the test results we see.

**Table 1. A screening test only provides apparent prevalence.**

		DISEASE Status		
		D + Positive	D – Negative	
TEST Results	Test +			100
	Test –			900
		???	???	1000

When we are only presented with apparent prevalence, we have to make decisions based on this data. We often want to know what proportion of the individuals is expected to true-positives and what proportion will be false positive? Similarly, what proportion of the test negatives will be true-negatives versus false negatives?

**Predictive Value of a Positive Test** gives the proportion of the test-positives which are true-positives to the total number testing positive. In this table, 9 of the 100 reported positive tests represent truly-positive individuals, giving a low P.V. of 9%. We see the rate of false positives is high (as seen in the 91 disease-negative samples that tested positive).

**Table 2. Predictive Values of a Positive or Negative Test.**

		DISEASE Status			
		D + Positive	D – Negative		
Apparent Prevalence =	10%				
True Prevalence =	1%				
TEST Results	Test +	9	91	100	P.V.(+) = 9 %
	Test –	1	899	900	P.V.(–) = 99.8 %
		10	990	1000	

**Predictive Value of a Negative Test** gives the proportion of the test-negatives which are true-negatives to the total number testing negative. In the example, 899 of the 900 negative test results are correct, giving a P.V.– of 99.8%. Here, the rate of false negatives is low (less than 1%) for this test.

Predictive values are influenced by the sensitivity, specificity, and disease prevalence. PVs should not be confused with sensitivity and specificity. For example, in the above sample set, 9 of the 10 true-positive fish are detected, giving a test sensitivity of 90%. Alternately, 899 of the 990 true-negative fish are reported as negative, giving a specificity of 90%. Both of these values are very respectable for screening tests. We'll review sensitivity and specificity, and then address how prevalence influences Predictive Values.

## Sensitivity and Specificity

**SENSITIVITY** is the proportion of true-positives which actually test positive, and how well a test is able to detect positive individuals in a population. A sensitive test will rarely “miss” positive individuals, and should be used when the chance of missing disease poses a large penalty (i.e., introduces a serious or exotic disease).

**SPECIFICITY** is the proportion of true-negatives which actually test negative, and reflects how well an assay performs in a group of disease negative individuals. A specific test will not produce false positives, or misclassify the identity of a pathogen. A highly specific test should be employed when false-positive results would cause significant impacts to the program (i.e., erroneous reporting of a significant disease in humans, or eliminating rare animals from a broodstock program).

$$\text{SENSITIVITY} = \frac{a}{a + c}$$

$$\text{SPECIFICITY} = \frac{d}{b + d}$$

**Table 2. Sensitivity and Specificity formula.**

		DISEASE Status	
		D+	D –
TEST Results	Test +	a	
	Test –	c	
		a + c	

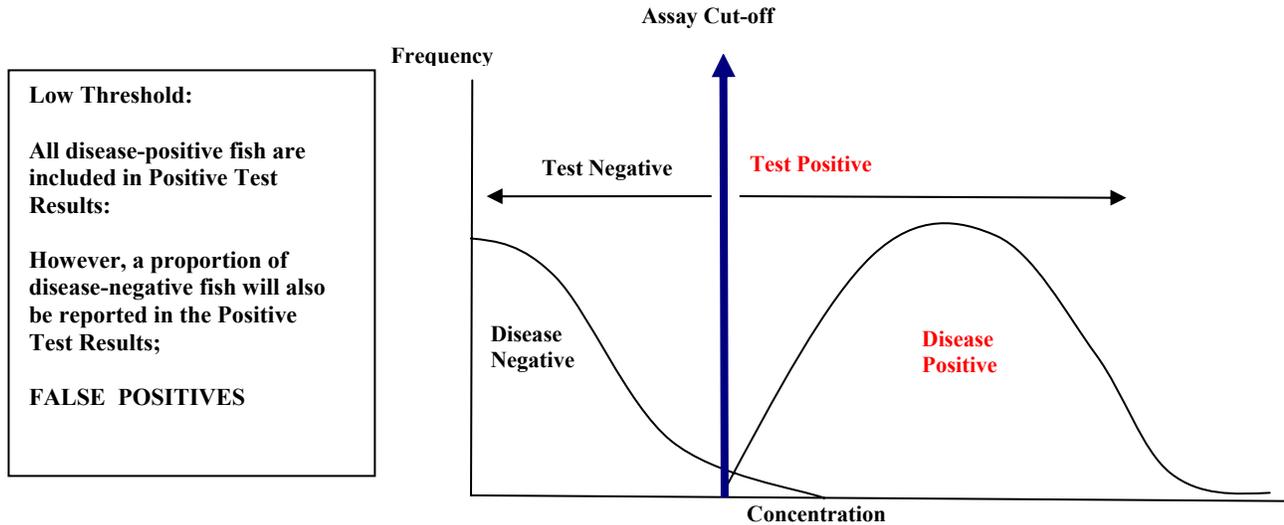
		DISEASE Status	
		D+	D –
TEST Results	Test +		b
	Test –		d
			b + d

It should be noted that the value for specificity is harder to calculate since defining a disease-free individual is more difficult than detecting a disease-positive individual.

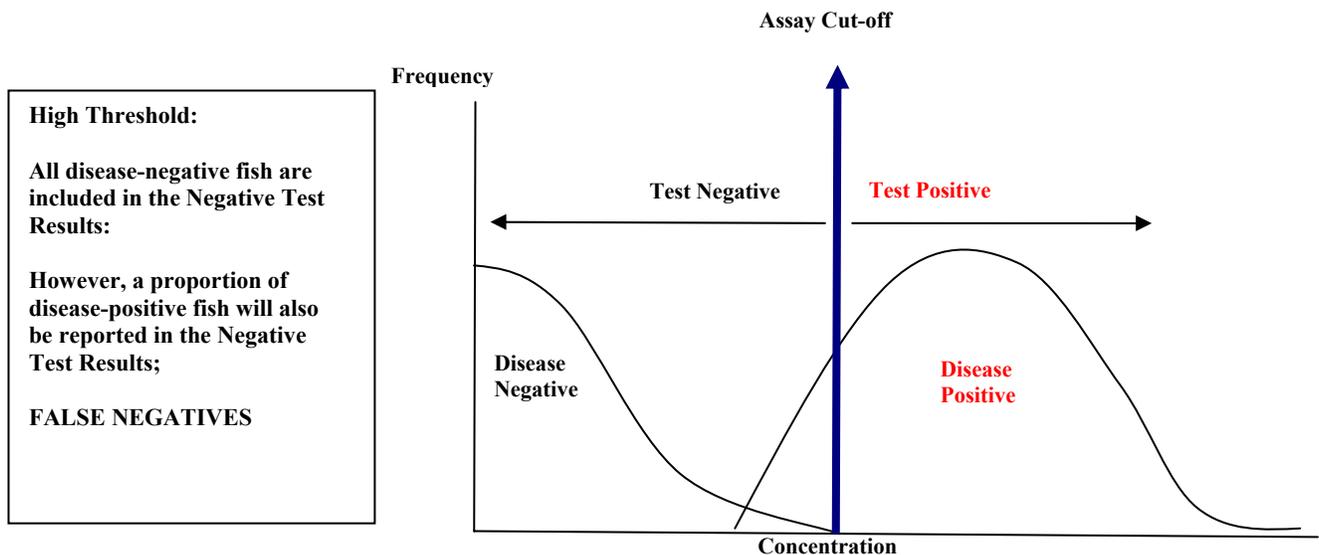
The ideal test is of course both highly sensitive and highly specific. However, the majority of diagnostic tests have continuous outcomes (i.e., quantity of protein, antibody titer, etc.), with the assignment of a positive/negative cut-off value decided by the diagnostician, or authoritative body conducting the testing. This setting of the positive/negative threshold makes these tests appear to produce dichotomous outcomes (i.e., positive or negative).

## What occurs when sensitivity or specificity is altered?

When either test sensitivity or specificity is altered, it will almost always cause a corresponding change in the other. For example, if in a given test, if the cut-off threshold is set lower, to ensure inclusion of all positive fish, it will also include a larger proportion of false-positive test results.



On the other hand, if the threshold is set very high to exclude all negative fish in the positive test results, then the test will also identify a large proportion of true-positive fish, reported as negative test results (false-negatives).



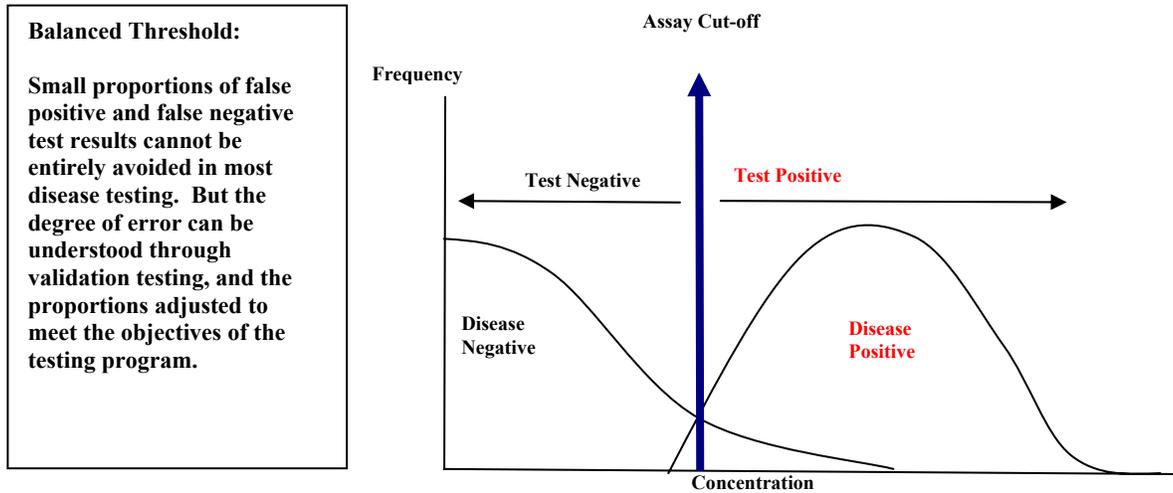
Determining where to set the cut-off value is very important, and often is driven by the level of error one is willing to accept in either direction. For example, let's look at the case of a captive breeding program, where every fish is extremely valuable, and as many eggs as possible are needed for the restoration program (i.e., restoration is a priority over disease status of a single pathogen).

**In general, as: Sensitivity increases (lower threshold), false-positive test results will increase, Specificity increases (higher threshold), false-negative test results increase.**

Example: Hatchery practice for salmonid broodstock is to discard all eggs from females testing positive for Bacterial Kidney Disease due to vertical transmission of this disease. In setting up a captive breeding program for recovery of a threatened and endangered (T&E) broodstock, the decision might be made to set the assay threshold higher for a group of fish already at critically low numbers. By setting the positive/negative threshold higher, you would be ensuring that none of the truly-negative fish are discarded as false-positive fish. The downside of this approach would be the acceptance of some truly-positive fish that will test negative (false negatives). But, as stated previously, this may a risk you are willing to take, versus

potentially discarding valuable eggs from negative fish. Genetic diversity (maintenance of as many genotypes in the next generation) usually outweighs disease management in these types of scenarios.

When an assay is validated, through testing the sensitivity and specificity using a known positive population, or spiked positive sample set, the threshold can be set in a balanced manner, or to meet specific objectives. A confidence interval can be constructed (see Evaluating Diagnostic Tests) and the clinician will be aware of the test errors and accepted trade-offs in the testing program.



In review, we have covered the following definitions and looked at examples of how sensitivity and specificity are altered when thresholds are established for a diagnostic test.

**Table 4. Review of Testing Terms and Formulas.**

		DISEASE Status			
		D +	D -		
TEST Results	T +	a	b	a + b	Apparent Prevalence $(a+b) / n$
	T -	c	d	c + d	True Prevalence $(a+c) / n$
		a + c	b + d	n	Predictive Value of Positive test $a / (a + b)$
					Predictive Value of Negative test $d / (c + d)$
					Sensitivity $a / (a + c)$
					Specificity $d / (b + d)$

Let's return now to the predictive values for positive and negative tests and see how true prevalence in the population affects these values. In this example, the true prevalence is 50%, meaning one would expect that any individual, randomly selected, has a 50% chance of being positive for the disease.

**Table 5. Apparent and True Prevalence – Predictive Values of a Test.**

True Prevalence (%): 50		DISEASE Status		Apparent Prevalence (%): 50	
		D + Positive	D – Negative		% Correct
TEST Results	Test +	475	25	500	95 %
	Test –	25	475	500	95 %
Total Disease Status (+/-)		500	500	1,000	

Sensitivity 95 %

Specificity 95 %

After applying this test, we are 95% sure that test-positive fish truly have the disease, and 95% of the test-negative fish are free of the disease. Also the apparent prevalence and true prevalence are the same. If we use the same screening assay to test a population with a true prevalence of 10%, the predictive values change.

At 10% prevalence, the predictive value of a positive test is significantly lower at 67%.

**Table 6. Apparent and True Prevalence (10%) – Predictive Values of a Test.**

True Prevalence (%): 10		DISEASE Status		Apparent Prevalence (%): 14	
		D +	D –		% Correct
TEST Results	Test +	95	45	140	67.9%
	Test –	5	855	860	99.4%
		100	900	1,000	

Sensitivity 95 %

Specificity 95 %

At 1% true prevalence, the predictive value of a positive is 15%. This means that 85% of the test-positive individuals are incorrectly identified.

**Table 7. Apparent and True Prevalence (1%) – Predictive Values of a Test.**

True Prevalence (%):		DISEASE Status		Apparent Prevalence (%):	
1		D +	D –	5.9	
					% Correct
TEST Results	Test +	9	49.5	59	15.3%
	Test –	1	940	941	99.9%
		10	990	1,000	
Sensitivity	90 %				
Specificity	95 %				

At 0.1% true prevalence, the predictive value of a positive is <2%, meaning 98% of the test-positive individuals are incorrectly identified.

**Table 8. Apparent and True Prevalence (0.1%) – Predictive Values of a Test.**

True Prevalence (%):		DISEASE Status		Apparent Prevalence (%):	
0.1		D +	D –	5.1	
					% Correct
TEST Results	Test +	1	50	51	1.9%
	Test –	0	949	949	100%
		1	999	1,000	
Sensitivity	100 %				
Specificity	95 %				

In summary, the predictive value decreases as the true prevalence of a population decreases. So, at very low prevalence levels (0.1 -1%), as in carrier or sub-clinical infections, the predictive value of a positive test becomes very small and is not very informative. At higher prevalence levels (i.e., during a severe epizootic), the predictive values will be much stronger.