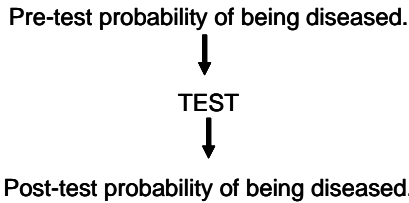


Value of a Diagnostic Test

Change from pre-test to post-test probabilities



Outcome	Disease	Test Result
True Positive TP	Present	Positive
True Negative TN	Absent	Negative
False Positive FP	Absent	Positive
False Negative FN	Present	Negative

	Gold Standard	
	Disease Present	Disease Absent
New Test +	TP	FP
New Test -	FN	TN

	Syphilia +	Syphilia -	total
VDRL +	800 TP	1,000 FP	1,800
VDRL -	200 FN	9,000 TN	9,200
total	1,000	10,000	11,000

$$\text{Prevalence} = \frac{\text{diseased}}{\text{population}} = \frac{(TP + FN)}{(TP + FN + FP + TN)} = \frac{1,000}{11,000} = .091$$

$$\text{Sensitivity} = \frac{TP}{\text{All_Cases}} = \frac{800}{(800 + 200)} = .80$$

	Syphilia +	Syphilia -	total
VDRL +	800 TP	1,000 FP	1,800
VDRL -	200 FN	9,000 TN	9,200
total	1,000	10,000	11,000

Figure 1 – Purpose of a Test

The goal of the diagnostic process is to determine if a person has or does not have a specified disease. A single test may serve this purpose, but often it only contributes information to this process by changing the probability of the disease being present. Either positive or negative test results can facilitate the diagnostic process.

Figure 2 – Possible outcomes of a test

A 'gold' standard is used to determine if the disease is present or absent and it is assumed to be a 100% reliable and valid. In cases in which the gold standard is not 100% reliable it may be referred to as a "silver" or "bronze standard" to reflect its inaccuracy. The results of a specific diagnostic test can then be compared to those of the gold standard resulting in correct (true) and incorrect (false) findings. (Note this is a simplification in which all results are either positive or negative.)

Figure 3 – Testing a test

A new test can be evaluated to determine if it produces the same results of a gold standard. A sample of people is selected and then have both the gold standard and the new test being assessed administered and the results are compared.

Figure 4 – Example

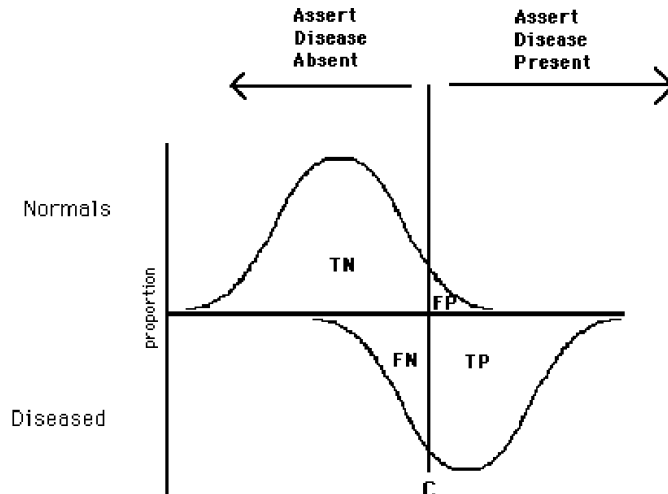
The VDRL, a test for syphilis, is being compared to the gold standard for diagnosing this disease. As determined by the gold standard, 1,000 (9% of the 11,000 studied) actually have the disease and 10,000 (91%) don't. Within those with syphilis, the VDRL test correctly classified 800 (80% true positives) and misclassified 200 (20% false negatives). Within those who don't have syphilis, 9,000 were correctly classified (90% true negatives) and 1,000 were misclassified (10% false positives.)

Figure 5 – Sensitivity

Sensitivity is the proportion of people with the disease that a test correctly classifies as having the disease. It is the rate of true positives and for this example it is .80 or 80%. Once sensitivity is known, the rate of false negatives is 1-sensitivity (1-.80 = .20 or 20%). High sensitivity is needed when you want to be sure you have identified everyone with the disease.

$$\text{Specificity} = \frac{TN}{\text{All_Healthy}} = \frac{9,000}{(9,000+1,000)} = .90$$

	Syphilis +	Syphilis -	total
VDRL +	TP 800	FP 1,000	1,800
VDRL -	FN 200	TN 9,000	9,200
total	1,000	10,000	11,000



$$\text{PPV} = \frac{TP}{\text{All_positives}} = \frac{800}{(800+1,000)} = .44$$

	Syphilis +	Syphilis -	total
VDRL +	TP 800	FP 1,000	1,800
VDRL -	FN 200	TN 9,000	9,200
total	1,000	10,000	11,000

$$\text{NPV} = \frac{TN}{\text{All_negatives}} = \frac{9,000}{(9,000+200)} = .98$$

	Syphilis +	Syphilis -	total
VDRL +	TP 800	FP 1,000	1,800
VDRL -	FN 200	TN 9,000	9,200
total	1,000	10,000	11,000

Figure 6 – Specificity

Specificity is the proportion of people without the disease that a test correctly classifies as not having the disease (i.e. healthy people classified as being healthy). It is the rate of true negatives and for this example it is .90 or 90%. Once specificity is known, the rate of false positives is 1-specificity (1-.90 = .10 or 10%). High specificity is needed when you want to be sure the person doesn't have the disease.

Figure 7 – Diagnostic Threshold

A cut-point, value, is often set to determine when a test's results will be categorized as "positive" or "negative". This threshold is sometimes under debate as with Prostate Specific Antigen (PSA). Depending on its value, the rates of TP, TN, FP and FN will change and therefore the sensitivity and specificity will change. In the example in the figure, if the diagnostic threshold is decreased such as has been recommended for PSA (the threshold moves to the left), then the rate of TN decreases, FP increases, FN decreases and TP increases. The reverse would happen with an increased threshold.

Figure 8 – Positive Predictive Value

The purpose of a test is to modify a pre-test probability of disease to a new post-test probability. A PPV represents the probability of the disease being present assuming a positive test result is obtained. It is the number of true positives divided by all test positives (which includes the false positives.) In this example, a pre-test probability of syphilis of .09 increased to .44 with a positive VDRL result. Syphilis is about 5 times more likely in this person after obtaining a positive test result.

Figure 9 – Negative Predictive Value

A NPV represents the probability of not having the disease assuming a negative test result is obtained. It is the number of true negatives divided by all test negatives (which includes the false negatives.) In this example, a pre-test probability of not having Syphilis was .91 (.09 have it) and it increased to .98 leaving only a .02 probability of having the disease. Another way to look at this is the probability of having the disease dropped from .09 to .02 or by over ¾.

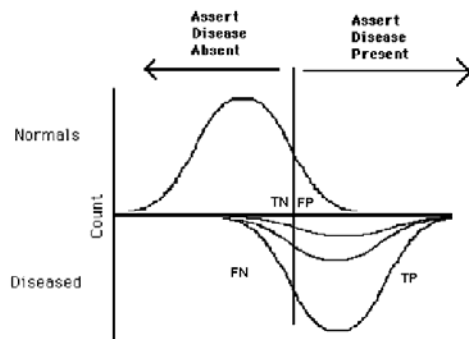
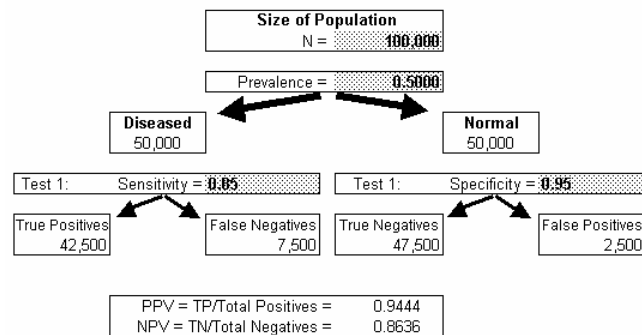


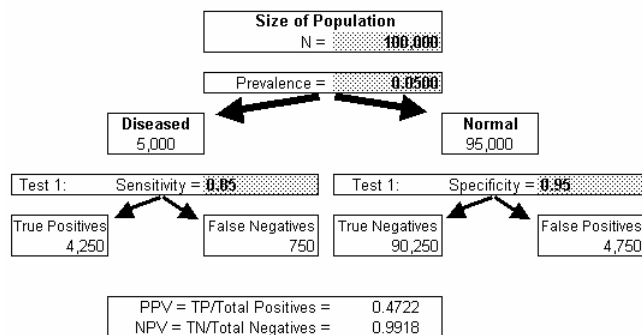
Figure 10 – Effect of Prevalence

If the proportion of the population who have the disease (prevalence) increases, the number of people with TP and FN test results will increase (but not the ratio between them.). Even though there is no change in the sensitivity and specificity of the test, this will result in an increase in PPV and a decrease in NPV. Therefore, the prevalence in the groups being tested actually change how useful the test may be.

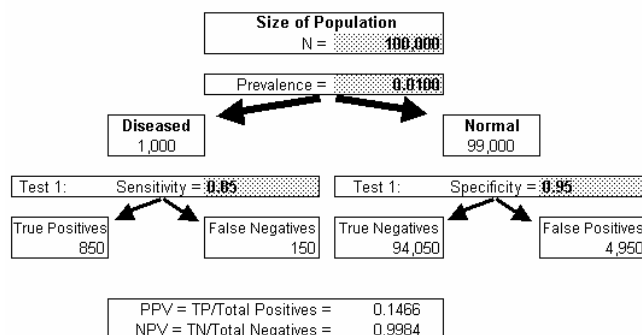


Figures 11-13 - Prevalence Changes

In this example a hypothetical population of 100,000 people is tested. In the first figure the prevalence is assumed to be .50 (i.e. half the population has the disease and half doesn't.) That means 50,000 have it and 50,000 don't. In the next two slides the prevalence is changed to .05 (5,000 with the disease and 95,000 without) and .01 (1,000 with the disease and 99,000 without.)



Assuming a diagnostic test with a sensitivity of .85, the true positive counts are 42,500, 4,250, and 850 respectively for these three different prevalence rates. (Remember that sensitivity is the proportion of those who have the disease that are correctly identified as having the disease.) The resulting false negative counts are 7,500, 750, and 150.



Assuming a specificity of .95 then the true negative counts are 47,500, 90,250, and 94,050 respectively. (Remember that specificity is the proportion of those without disease who are correctly classified as not having the disease.) The corresponding false positives counts are 2,500, 4,750, and 4,950.

Therefore the PPVs, $TP/(TP + FP)$, are .94, .47, and .14. The PPV is higher for diseases with higher prevalence and lower for diseases with low prevalence.) The NPVs, $TN/(TN + FN)$ are .86, .992, and .998 demonstrating a lower NPV with high prevalence and a higher NPV with lower prevalence. Did you notice that NPV doesn't change as much as PPV? That is because in most situations the prevalence is small (i.e. most people who are being tested don't have the disease.)



H - Healthy

E - Exposed

A - Asymptomatic

S - Symptomatic

R - Resolution: cure, death, remission

Figure 14 –Definition of Screening

A diagnostic process in which an apparent healthy population is tested in order to detect the presence of disease in its asymptomatic phase in order to begin an early intervention that will benefit the individual or the community.

Since screening tests attempt to detect disease when it is in the asymptomatic phase, the longer this phase is, the more successful is the screening program.

Figure 15 – Characteristics of a Good Screening Test

1. Disease is worth the effort
2. Intervention is worth giving
3. Test is effective in correctly identifying those with and without the disease
4. Test is applicable and acceptable

For more information:

Scott L. Wetstone, M.D.
Associate Professor
Depart. of Community Medicine & Health Care
University of Connecticut School of Medicine
(860) 679-4440 wetstone@nso.uchc.edu