I. Clinical Testing

Diagnosis is the process of discovering a patient’s underlying disease by ascertaining the patient’s signs and symptoms, choosing appropriate tests, interpreting the results and arriving at (hopefully correct) conclusions. This is often a highly complicated process and the manner in which experienced clinicians arrive at a diagnosis is not well understood.

Hypothetico-deductive reasoning refers to the diagnostic strategy that nearly all clinicians use most of the time. It is defined as the formulation from the earliest clues, of a short list of potential diagnoses or actions, followed by the performance of those clinical and laboratory tests that best reduce the length of the list to come up with a final diagnosis. The list of possibilities is reduced by considering the evidence for and against each, discarding those which are very unlikely and conducting further tests to increase the likelihood of the most plausible candidates. The process as used by experienced clinicians can be described in the following steps:

1. Formulate explanations (hypotheses) for the patient’s primary problem.
2. First consider those explanations that are most likely and/or those that are particularly harmful to miss (e.g., cerebral aneurysm for headache).
3. Simultaneously rule-out those that would be particularly harmful or catastrophic and try to rule-in those that are considered to be most likely.
4. Continue until the candidate list of explanations is shortened (i.e., 2-3) and/or one candidate disease is identified as having a very high likelihood (i.e., > 90% sure).

II. Clinical Test Characteristics

A. Sensitivity and Specificity

In clinical testing parlance a diagnostic test can be applied to any piece of clinical
information whether obtained from the patient's history, the physical examination or use of diagnostic procedures such as radiography or electrocardiography. To aid our discussion we will assume, initially, that both the disease and the diagnostic test have only two levels. So, the disease is either present (D+) or absent (D-) and the test is either positive (T+) (i.e., the test indicates that disease is present) or negative (T-) (i.e., the test indicates that the disease is absent).

There are 4 possible interpretations of these test results, two of which are correct - true positive (TP) and true negative (TN) and two of which are incorrect - false positive (FP) and false negative (FN). The relationship between these 4 test results is typically shown in the form of a two-by-two table (Figure 1).

Figure 1. Relationship between a Dichotomous Test Result and Disease Status

<table>
<thead>
<tr>
<th>DISEASE STATUS</th>
<th>PRESENT (D+)</th>
<th>ABSENT (D-)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TEST RESULT</strong></td>
<td><strong>TP</strong></td>
<td><strong>FP</strong></td>
</tr>
<tr>
<td><strong>POSITIVE (T+)</strong></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td><strong>NEGATIVE (T-)</strong></td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Note that because of the inherent variability in biological systems there is no perfect test - false positive and false negative results always occur. The interpretation of diagnostic test results is essentially concerned with comparing the relative frequencies of the two incorrect results - the FNs and FPs to the two correct results - the TPs and TNs.

While some diagnostic tests results are inherently either positive or negative, for many common diagnostic tests the results are expressed on a continuous scale. In such cases, it is common to divide the continuum of values into either ‘abnormal’ or ‘normal’ findings, where ‘abnormal’ means a test measurement that is indicative of having disease, and ‘normal’ means indicative of not having disease. The point at which values are determined to be either normal or abnormal is called the cut-point. Tests will seldom be perfect in separating out diseased from non-diseased patients - virtually all have some overlap between the two populations, as illustrated in Figure 2.

Figure 2. Results for a Typical Diagnostic Test Illustrating the Overlap between Diseased (D+) and Non-diseased (D-) Populations
To the extent that the two populations have similar measurements the test will not be able to discriminate between them. Conversely, the less the extent of overlap between the diseased and non-diseased populations the greater the discriminatory power of the test. The degree of overlap is therefore a measure of the test effectiveness and it is this that both sensitivity (Se) and specificity (Sp) quantify.

When reading an article about a diagnostic test, the presence or absence of disease has to be determined using some other source of information known as the "gold standard". The gold standard may involve obtaining a culture or biopsy, performing an elaborate diagnostic procedure such as a CAT scan, confirming the presence or absence of disease at surgery or post-mortem or simply determining the response to treatment or the results of long-term follow up.

**Sensitivity:** Sensitivity (Se) is defined as the proportion of individuals with disease that have a positive test result, or

\[
Se = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{a}{a + c}
\]

Se is the conditional probability of being test positive given that disease is present, or \(Se = P(T+ | D+)\). Se is also referred to as the true-positive rate. Note that Se is calculated solely from diseased individuals.

**Specificity:** Specificity (Sp) is defined as the proportion of individuals without disease that have a negative test result, or

\[
Sp = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{d}{d + b}
\]

Sp is the conditional probability of being test negative given that disease is absent, or \(Sp = P(T- | D-)\). Sp is also referred to as the true-negative rate. Note that Sp is calculated solely from non-diseased individuals.

**B. Example Clinical Problem - Deep Vein Thrombosis**

Deep vein thrombosis (DVT) is a condition of active thrombosis in the deep venous system of one or both lower extremities that can lead to significant complications of pulmonary embolism, chronic venous insufficiency, and possibly death. The presence or absence of clinical signs and symptoms (pain, tenderness, swelling and edema) do not correlate well with the presence or absence of DVT. The gold standard diagnostic test is ascending functional venography, which at proficient centers, provides adequate visualization of the venous system in 95-98% of patients. The test is not perfect and can produce occasional false positive results. A negative or normal study however, virtually confirms that the patient is free of disease. There is a 2-4%
risk of the procedure actually inducing DVT in patients who were originally free of the condition. Because of the technical requirements of venography, the fact that it is an imperfect, and that it is associated with some complications, other non-invasive techniques have been developed.

One such non-invasive test is the *D-dimer assay*. D-dimer is a specific degradation product of cross-linked fibrin and is therefore a marker of endogenous fibrinolysis (which would be expected to be elevated in the presence of a DVT). Several studies have shown that D-dimer assays have high Se but have only moderate Sp. The example data we will use (Figure 3) is from a Canadian study (Wells PS et al, Circulation 1995;91:2184-87) of 214 patients seen at two hospitals, all of whom had a whole blood assay for D-dimer (SimpliRED) and underwent the gold standard test of contrast venography. The test characteristics were Se = 89% [95% CI = 77-96%], and Sp = 77% [95% CI = 63-80%].

**Figure 3.** Se & Sp of D-dimer whole blood assay (SimpliRED) for DVT (Ref: Wells PS, Circulation, 1995)

<table>
<thead>
<tr>
<th></th>
<th>PRESENT (D+)</th>
<th>ABSENT (D-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D- dimer test POS (T+)</td>
<td>47</td>
<td>37</td>
</tr>
<tr>
<td>D- dimer test NEG (T-)</td>
<td>6</td>
<td>124</td>
</tr>
</tbody>
</table>

Clinicians need to take into account the inherent attributes of a test (i.e., its Se and Sp) before a test is selected and used. Since no test is perfect (i.e., 100% sensitive and 100% specific), the usefulness of a particular test will depend on what information the clinician wants to get from it.

**Tests with High Sensitivity**

For a perfectly sensitive test (i.e., Se = 100%), all diseased patients test positive - there are no false negative results (the false negative rate is zero) (Figure 4). Note that all test negative patients are disease free but that in this example, a sizeable proportion of the disease-free population test positive (= false positive).

A perfectly sensitive tests almost never exists, more typically we are interested in tests that have a high sensitivity (i.e., > 90%). Here, false negative results among diseased individuals are few in number - the vast majority of test negative patients are disease free. Highly sensitive tests are most useful to rule-out disease because if the test is negative you can be confident that disease is absent since FN results are rare.
It is important to understand that a highly sensitive test does not tell you if disease is present, despite the fact that Se is calculated using only diseased individuals and that nearly all of these patients test positive! This is because Se provides no information regarding the number (or rate) of false positive results - this information is provided by the Sp of the test.

A highly sensitive test is therefore most helpful to a clinician when the test result is negative, because it rules out disease, whereas the interpretation of a positive result will depend on the rate of false positive results (see Specificity).

SnNout is a mnemonic designed to indicate that if a sign, symptom or other diagnostic tests has a sufficiently high Sensitivity, a Negative result rules out disease.

There are three clinical scenarios when tests with high sensitivity should be selected:

1) in the early stages of a work-up when a large number of potential diseases are being considered. If the test has a high Se, a negative result indicates that a particular disease is very unlikely and it can therefore be dropped from consideration (i.e., ruled out).

2) when there is an important penalty for missing a disease. Examples include tuberculosis and syphilis, which are dangerous but treatable conditions. We would not want to miss these cases, hence we want a test that has a low number of false negative results (i.e., a highly Se test).

3) in screening tests where the probability of disease is relatively low (i.e., low disease prevalence) and the purpose is to discover asymptomatic cases of disease.
Table 1. Examples of Tests with High Sensitivities

<table>
<thead>
<tr>
<th>Disease/Condition</th>
<th>Test Result</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duodenal ulcer</td>
<td>History of ulcer, 50+ yrs, pain relieved by eating or pain after eating</td>
<td>95%</td>
</tr>
<tr>
<td>Favourable prognosis following non-traumatic coma</td>
<td>Positive Corneal reflex</td>
<td>92%</td>
</tr>
<tr>
<td>High intracranial pressure</td>
<td>Absence of spont. pulsation of retinal veins</td>
<td>100%</td>
</tr>
<tr>
<td>Deep vein thrombosis</td>
<td>Positive D-dimer</td>
<td>89%</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>Positive endoscopic retrograde cholangio-pancreatography (ERCP)</td>
<td>95%</td>
</tr>
</tbody>
</table>

Tests with High Specificity

For a perfectly specific test (i.e., Sp = 100%), all non-diseased patients test negative - there are no false positive results (the false positive rate is zero) (Figure 5). Also note that all test positive patients have disease but that in this example, a sizeable proportion of the diseased population test negative (= false negative).

Again, a perfectly specific tests almost never exists, more typically we are interested in tests that have a high specificity (i.e., > 90%). Here, false positive results among non-diseased individuals are few in number, so the vast majority of test positive patients have disease. Highly specific tests are most useful to **rule-in** disease, because if the test is positive you can be confident that disease is present since false positive results are rare.
It is important to understand that a highly specific test does not tell you if disease is absent, despite the fact that Sp is calculated using only non-diseased individuals and that nearly all of them test negative. This is because Sp provides no information regarding the number (or rate) of false negative results - this information is provided by the Se of the test.

*SpPin* is a mnemonic designed to show that if a sign, symptom or other diagnostic tests has a sufficiently high specificity, a Positive result rules in disease.

A highly specific test is therefore most helpful to a clinician when the test result is positive, since it rules-in disease. There are two clinical scenarios when tests with high specificity should be selected:

1) to rule-in a diagnosis that has been suggested by other tests - specific tests are therefore used at the end of a work-up to rule-in a final diagnosis e.g., biopsy, culture, CT scan.

2) when false positive tests results can harm the patient physically or emotionally. For example, the confirmation of HIV positive status or the confirmation of cancer prior to chemotherapy. A highly specific test is required when the clinician wants to be absolutely sure that a condition is present.
Table 2. Examples of Tests with High Specificities

<table>
<thead>
<tr>
<th>Disease/Condition</th>
<th>Test</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol dependency</td>
<td>No to 3 or more of the 4 CAGE questions</td>
<td>99.7%</td>
</tr>
<tr>
<td>Iron-deficiency anemia</td>
<td>Negative serum ferritin</td>
<td>90%</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Negative fine needle aspirate</td>
<td>98%</td>
</tr>
<tr>
<td>Strep throat</td>
<td>Negative pharyngeal gram stain</td>
<td>96%</td>
</tr>
</tbody>
</table>

C. Trade-off Between Sensitivity and Specificity

Because there is no such thing as a perfect test (a test that has no FP or FN results) there is an inherent trade-off between sensitivity and specificity. For clinical test results that have a continuous scale of measure, the location of the cut-point (the point on the continuum that divides normal and abnormal) is arbitrary and can be modified according to the purposes of the test. For example, in Figure 6 below, we can see that if the cut-point is made lower (i.e., moved to the left) there will be less FN results but more FP results - therefore Se will be increased at the expense of Sp. Figure 4 shows the extreme of this scenario, where the cut-point has been lowered to make Se = 100% at the expense of Sp. Conversely, Figure 5 shows the other extreme where the cut-point has been increased (moved to the right) to maximize Sp at the expense of Se. The trade-off between Sp and Se cannot be avoided and points to the fact that the ideal cut-point depends on what the purpose of the test is. A Receiver Operator Characteristic (ROC) Curve is a graphical way of illustrating the trade off between Se and Sp for various cut-points of a diagnostic test.

Fig 6. Trade-off: Lowering the Test Cut-point Increases Se but Decreases Sp
C. **Receiver Operator Characteristic (ROC) Curves**

There are two main uses of the ROC curve – to compare the accuracy of two or more tests, and to show the trade off between the Se and Sp as the cut-point is changed.

The ROC curve is constructed by plotting the sensitivity (or true positive rate) against the false positive rate (1 - Specificity), for a series of cut-points as illustrated in Figure 7 below which evaluates the utility of creatine kinase (CK) to diagnose acute myocardial infarction.

Figure 7. **An Example Receiver Operator Characteristic Curve (The Accuracy of the CK Test in the Diagnosis of Myocardial Infarction)**

The ROC curve is a good way of comparing the usefulness of different tests. The higher the sensitivity and specificity of a test the further the curve is pushed up into the top left hand corner of the box. Tests which discriminate best lie "further to the north-west", because they have both low FN rates (indicated by high Se) and low FP rates (indicated by a high Sp) (See Figure 3.5 in FF for an example of such a figure).

A test that has no discriminating ability has equal TP and FP rates which is indicated by the diagonal straight line in the above figure.\(^1\) The ability of different tests to

\[^{1}\text{This is equivalent to a likelihood ratio (LR) of 1.0 (we will discuss LR}s in Epi-547). Note that the slope of the ROC curve (i.e., the ratio of the TP Rate to the FP Rate) for any given cut-point is the likelihood ratio.}
discriminate between diseased and non-diseased individuals can be quantified by calculating the Area Under the ROC Curve (AUROCC), which varies from 0.5 (no discriminating ability) to 1.0 (perfect accuracy).

The ROC curve is also helpful in deciding on the best cut-point for a particular test. The choice of best cut-point is influenced by the likelihood of disease (i.e., its prevalence) and the relative costs (or risk-benefit ratio) associated with errors in diagnosis - both false positive and false negative. Advanced statistical decision theory can be applied to determine the optimal operating position on the ROC curve for a given test (the details of which are beyond this course), however, we can use the following intuition to understand the basic principle:

If the cost of missing a diagnosis (a false negative result) is high compared to the cost of falsely labeling a healthy individual as diseased (a false positive result), then one would want to operate along the horizontal part of the curve (e.g., a cut-point of 60 CK units in Figure 7), since at this point FN results are minimized at the expense of FP results. A CK cut point of about 60 therefore maximizes sensitivity (i.e., ~90%) while providing reasonable specificity (i.e., ~50%) 

On the other hand, if the cost of falsely labeling a healthy person as diseased (a false positive result) is high compared to the cost of missing a diagnosis (a false negative result), then one would want to operate along the vertical part of the curve (e.g., a cut-point of 320 CK units in Figure 7), since at this point FP results are minimized. A CK cut point of 320 therefore maximizes specificity (i.e., ~99%) while providing moderate sensitivity (i.e., ~40%)

III. Prevalence and Predictive Values
While Sensitivity and Specificity are important concepts to understand they, unfortunately, don't tell the full story of clinical testing. In terms of conditional probabilities Se and Sp can be defined as:

\[ Se = P(T+|D+) \]
\[ Sp = P(T-|D-) \]

The problem is that these measures can only be calculated if the true disease status is known i.e., they are "conditional" on the disease status being either positive (for Se) or negative (for Sp). However, the clinician is using a test precisely because the true disease status of the patient is unknown! The clinician wants to know the conditional probability of disease given the test result e.g., \( P(D+|T+) \), hence Se and Sp are apparently not much use!

In order to use diagnostic test data to infer the true disease status of a patient, the clinician needs to understand the concepts of predictive values and prevalence:

**Predictive Value Positive (PVP):** Predictive value positive is the probability of disease
in a patient with a positive (abnormal) test.

\[
PVP = \frac{\text{True Positives}}{\text{True Positives + False Positives}} = \frac{TP}{TP + FP} = a \quad \frac{a}{a + b}
\]

PVP is the conditional probability of being diseased given that the test was positive, or \( PVP = P(D+|T+) \). Note that Sp and PVP are "linked" in that they both provide information on the FP rate. A highly specific test helps to rule-in disease because PVP is maximized.

**Predictive Value Negative (PVN):** Predictive value negative is the probability of not having disease when the test result is negative (normal).

\[
PVN = \frac{\text{True Negatives}}{\text{True Negatives + False Negatives}} = \frac{TN}{TN + FN} = \frac{d}{d + c}
\]

PVN is the conditional probability of not being diseased given that the test was negative, or \( PVN = P(D-|T-) \). Note that Se and PVN are "linked" in that they both provide information on the FN rate. A highly sensitive test helps to rule-out disease because PVN is maximized.

From a clinical standpoint, we are actually more interested in the complement of the PVN or \( 1 - PVN \). This measure, which can also be expressed as \( P(D+|T-) \), tells the clinician what the probability is of having the disease despite testing negative (i.e., the rate of false negative test results among all negative test results). A high PVN means that there are few false negative results among all test negative results, so an alternative diagnosis should be sought.

**Prevalence:** Prevalence simply represents the proportion of the total population tested that have disease, or

\[
P = \frac{\text{Total Number of Diseased}}{\text{Total Population (N)}} = \frac{TP + FN}{TP+FN+FP+TN} = \frac{a + c}{a + b + c + d}
\]

Prevalence is very important since it has a dramatic influence on predictive value positive and negative. Prevalence is the "third force" - it is the player that often goes unnoticed only to reveal its influence in dramatic fashion! Other equivalent names for prevalence include the likelihood of disease, prior probability, prior belief, prior odds, pre-test probability and pre-test odds.

Let's go back and look at the example of D-dimer testing and DVT:
Figure 8. The PVP and PVN of D-dimer whole blood assay (SimpliRED assay) for DVT (Ref: Wells PS, Circulation, 1995) Prevalence = 25%

<table>
<thead>
<tr>
<th></th>
<th>PRESENT (D+)</th>
<th>ABSENT (D-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POS (T+)</td>
<td>47</td>
<td>37</td>
</tr>
<tr>
<td>NEG (T-)</td>
<td>6</td>
<td>124</td>
</tr>
</tbody>
</table>

\[ PVP = \frac{47}{84} = 56\% \]
\[ PVN = \frac{124}{130} = 95\% \]

From this 2-by-2 table, we can calculate the PVP as \( \frac{47}{84} = 56\% \) and the PVN as \( \frac{124}{130} = 95\% \). So, if the test is positive we are only 56% sure that the patient has the disease (about as sure as tossing a coin), whereas if the test is negative we are 95% sure that the subject is disease free. Obviously this test is extremely good for ruling out DVT but practically worthless at ruling in DVT! The other important piece of information to note is the high prevalence of disease in this population i.e., \( \frac{53}{214} = 25\% \).

The prevalence of DVT in this population was very high, since this group of patients had been admitted to one of 2 Hamilton, Ont., area referral hospitals participating in this research project. Now let’s look at the situation that a community-based physician might face. In a typical community-based hospital, the prevalence of DVT in a group of patients complaining of leg pain and swelling is likely to be lower. For illustration, let’s say its 5%. The primary care physicians at this community-based hospital are very excited to try out the new test that performed so well in Hamilton. The hospital used the same D-dimer test in another 214 sequential patients with clinical signs consistent with DVT. The Se and Sp of the test are still the same (i.e., 89% and 77%, respectively). The results obtained are shown below:

Figure 9. The PVP and PVN of D-dimer whole blood assay (SimpliRED assay) for DVT (Ref: Wells PS, Circulation, 1995). Prevalence of DVT = 5%

<table>
<thead>
<tr>
<th></th>
<th>PRESENT (D+)</th>
<th>ABSENT (D-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POS (T+)</td>
<td>10</td>
<td>47</td>
</tr>
<tr>
<td>NEG (T-)</td>
<td>1</td>
<td>156</td>
</tr>
</tbody>
</table>

\[ PVP = \frac{10}{57} = 18\% \]
\[ PVN = \frac{156}{157} = 99.4\% \]

Mathew Reeves, PhD
© Department of Epidemiology, MSU
The physicians are surprised to find out that only 18% of the patients who tested positive actually had DVT. Explanation?.... The lower disease prevalence meant that proportionally more patients were placed in the disease absent column (cells b and d), while fewer patients were placed in the disease present column (cells a and c). So, in the above example, 5% of the 214 subjects (i.e., 11) were put in the disease present column, while 203 were put in the disease absent column. Although the Sp remained at 77%, the 23% FP rate meant that 47 of the 203 patients that did not have DVT had FP results (cell b). The Se also remained the same (89%), so 10 of the 11 patients that had DVT tested positive (cell a). The PVP is lower because the relative size of cell a compared to cell b is now much smaller.

One will also note from this example that PVN increased with the lower prevalence - again this is a result of the change in relative sizes of cells c and d. The influence of prevalence on PVP and PVN is demonstrated in the following figure:

Figure 10. The PVP and PVN as a Function of Prevalence for a Typical Diagnostic Test

The effect of prevalence can be summarized as follows:

*As prevalence falls, positive predictive value must fall along with it, and negative predictive value must rise. Conversely, as prevalence increases, positive predictive value will increase and negative predictive value will fall.*
IV. Bayes’ theorem - the calculation of predictive values for any combination of Se, Sp and Prevalence values using a 2 x 2 table

Bayes theorem is essentially the process by which disease probabilities are revised in face of new test information. We will learn much more about the application of Bayes theorem in diagnostic testing in Epi-547, but for now our task is to be able to calculate predictive values for any combination of Se, Sp and prevalence values using a 2 x 2 table. A step-by-step approach (using an example with Se = 90%, Sp = 80% and prevalence = 10%) follows:

1. First pencil out a 2 x 2 table with disease status (present or absent) along the top (columns) and test status (positive or negative) along the left-hand side (or rows).
2. Next fix the total number of subjects (N) to be included in the table. You can use any number but obviously it makes it easier to use a whole number such as 100 or 1000. Lets pick 1,000.
3. Now calculate the expected number of diseased individuals by applying the disease prevalence rate of 10% to the 1,000 subjects (= 100), and place them at the bottom of the left hand (disease +) column.
4. Place the 900 disease-free subjects at the bottom of the right hand column.
5. Calculate the number of subjects in the top left cell (cell “a”) by multiplying 100 by the sensitivity (i.e., 0.90 x 100 = 90) and place the remaining 10 in the lower left cell (cell “c”) – these are the false negative subjects.
6. Likewise use the Sp of 80% and the 900 subjects to calculate the numbers of subjects in cells “b” and “d” (180 and 720, respectively).
7. Now use the top row (cells a and b) to calculate the PPV (90/270 = 33%) 
8. And use the lower row of cells (cells c and d) to calculate the PVN (720/730 = 98.6%). (N.B. you can also calculate PVP and PVN directly using the two Bayes’ equations shown in the lecture).
9. Your table should look like below.
10. Practice doing this using the table of Se, Sp and Prevalence values on Angel.

<table>
<thead>
<tr>
<th>Disease</th>
<th>PRESENT (D+)</th>
<th>ABSENT (D-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POS (T+)</td>
<td>90</td>
<td>180</td>
</tr>
<tr>
<td>NEG (T-)</td>
<td>10</td>
<td>720</td>
</tr>
</tbody>
</table>

N = 100  N = 900  N = 1000
Se = 90%  Sp = 80%

PVP = 90/270  = 33%
PVN = 720/730  = 98.6%

V. Multiple testing strategies
Diagnostic tests that have sufficiently high sensitivity and specificity that they can simultaneously “rule out” and “rule-in” are very rare. Generally the physician has access to an array of imperfect tests. However, armed with a good understanding of these diagnostic test principles and Bayes’ theorem, she will be able to squeeze out more information from the available tool box by combining tests. There are two ways of doing this:

Parallel testing
This describes the situation where several test are run simultaneously (i.e., a panel of tests) and any one positive test leads onto further evaluation (see Figure 3.12 in FF). The net effect is to increase the likelihood of detecting disease i.e. sensitivity increases (because there are multiple opportunities for a positive test result). However, as is always the case, there is a price to pay - the probability of false positive results increases so specificity declines. Parallel testing is typically used in the early phases of the work up where you are trying to quickly rule out several conditions – by running a panel of related tests - if they all come back negative - then the condition(s) can be “ruled out” (think SnOut – parallel testing maximizes Se, so PVN is maximized). However, when using this strategy a positive test means little (other than more testing is required). The parallel testing strategy can be easily abused by using it as a screening tool for “anything and everything”. This approach, often favoured by neophyte interns and residents, is very costly, highly inefficient, dangerous to the patient (who has to undergo unnecessary follow-up tests because of the false positive results), and is ultimately bad medicine. As explained in the FF text (page 53-55) this strategy works best when a highly sensitive test strategy is required but you are armed with 2 or more relative insensitive tests. If these tests measure different clinical phenomenon, then combining them in parallel maximizes your chance of identifying diseased subjects.

Serial testing
This describes the situation where several tests are run in order and each subsequent test is only run if the first test was positive (see Figure 3.12 in FF). In this approach any negative test leads to the suspension of the work-up. The net effect is to increase specificity and positive predictive value because each case has to test positive to multiple tests (so false positives are rare). However, again there is a price to pay - the probability of false negative results increases so sensitivity declines. Serial testing is typically used when one wants to be sure that a disease is ruled in with certainty, and there is no rush to do so. It is also used when a particular definitive test is expensive, difficult, or invasive – to avoid over-using such a test, a cheaper and/or less invasive test is run first and only those testing positive go on to have the definitive test. An example would be the use of a colonoscopy following a positive fecal occult blood test.