Diagnostic Principles

EBCP Module #10
Outline of Module

This module will introduce the following topics:

- Characteristics of diagnostic tests
  - Sensitivity and specificity
  - Positive and negative predictive values

- Context & interpretation of diagnostic tests
  - Prevalence and predictive values
  - Likelihood ratios
  - Receiver operating characteristic (ROC) curves
Objectives of Module

- Students who complete this module should be able to correctly define and apply the following terms as they pertain to evidence-based clinical practice.
  - Sensitivity and specificity
  - Positive/negative predictive value
  - Likelihood ratios for positive and negative test results
  - Nomogram
  - Receiver operating characteristic (ROC) curve
  - Area under the ROC curve (AUC)

- Students who complete this module should be able to:
  - Distinguish between sensitivity, specificity, and predictive values.
  - Calculate sensitivity, specificity, and positive and negative predictive values.
  - Use a likelihood ratio nomogram.
  - Interpret an ROC curve.
Characteristics of Diagnostic Tests
Role of Diagnostic Principles in EBCP

- In previous modules, we saw that epidemiologic data obtained from groups of patients or populations can influence a clinician’s differential diagnosis.
  - Examples of these statistics include prevalence and incidence.

- Once a differential diagnosis exists, a clinician may proceed with additional diagnostic testing in order to refine the diagnosis.

- Diagnostic data obtained from groups of patients or populations can provide very useful information about:
  - How well a diagnostic test works
    - How many cases of disease will it detect? How many will it miss?
    - How well does it distinguish cases from non-cases?
  - What a diagnostic result means for a given patient
    - Given that a patient has a positive (or negative) test result, what is the probability that the patient actually has the disease (or is free of the disease)?
Table Used for Diagnostic Studies

- Recall the basic 2 x 2 table:

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>No Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Exposed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- For diagnostic studies, this can be altered to form the following table.

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>No Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This table can then be filled to represent four important categories of patients in regard to diagnostic testing.

- **True positive**: Patients with disease who tested positive
- **False positive**: Patients without disease who tested positive
- **False negative**: Patients with disease who tested negative
- **True negative**: Patients without disease who tested negative

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>No Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Positive</td>
<td>True Positive (TP)</td>
<td>False Positive (FP)</td>
</tr>
<tr>
<td>Test Negative</td>
<td>False Negative (FN)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>
Terms Used for Diagnostic Studies

- Ideally, a test would perfectly distinguish between the presence and absence of disease.
  
  Of course, this is not a realistic outcome for diagnostic tests.*

- That is, every positive result would be true and every negative result would be true.
  - Of course, this is not a realistic outcome for diagnostic tests.*

- For this reason, terms such as sensitivity, specificity, and predictive value are used to describe the performance of diagnostic tests.

*We discussed in previous modules that treatments—even very good ones—may not work for all patients. This is also true for diagnostic tests.
Sensitivity and Specificity
Sensitivity

- Measures how well a test identifies the presence of disease when it is present.

- Defined as the proportion of positive results in patients with disease (TP) out of all patients with disease (TP+FN).
  - Sensitivity = TP / (TP+FN)

- Highly sensitive tests may be used to help rule out a diagnosis when the result is negative in a patient.
  - If a test turns positive for nearly everyone who has the disease, and our patient has a negative result, we can more confidently exclude the diagnosis.
  - This characteristic of a highly sensitive test is abbreviated as a “SnNOut” by some EBM practitioners.
Specificity

- Measures how well a test identifies the *absence* of disease when it is absent.

- Defined as the proportion of negative results in patients without disease (TN) out of all patients without disease (TN+FP).
  - Specificity = \( \frac{TN}{TN+FP} \)

- Highly specific tests may be used to help rules in a diagnosis when the result is positive in a patient.
  - If a test is known to turn negative in virtually everyone who doesn’t have the condition, and our patient has a *positive* result, we can more confidently *include* the diagnosis (that is, make the diagnosis).
  - This characteristic of a highly specific test is termed a “SpPlIn” by some EBM practitioners.
Sensitivity vs. Specificity – Summary

- Popular mnemonics
  - SnNOut (sensitivity, negative rules out)
  - SpPIn (specificity, positive rules in)

- Mathematical approaches
  - Sensitivity = TP/(TP+FN)
    - Remember: True positives out of all with disease, or “positivity in disease”
  - Specificity = TN/(TN+FP)
    - Remember: True negatives out of all without disease, or “negativity in health”
YouTube Video Link – Intuitive Sensitivity & Specificity

- Another video by Dr. Rahul Patwari – only 9 minutes long and well worth it!

https://www.youtube.com/watch?v=U4_3fditnWg
Clinical Applications

- An ideal test would have high sensitivity and high specificity, but this is not always possible.

- High sensitivities are desirable for screening tests, when we want to identify as many true positives as possible.
  - But we may well scare some people who have false positive results, especially if the test’s specificity isn’t very high.

- High specificities are desirable for confirmatory tests, when we want to verify a diagnosis.
  - But we may miss some people who have the disease (false negatives) when we use highly specific tests, especially if the sensitivity isn’t great.

- Pairs of diagnostic tests (e.g., a screening test and a confirmatory test) can be used in sequence to maximize these properties.
Standards: An Important Caveat

- All of the preceding slides imply that we can know whether or not a test measures what we want it to.
  - That is, we assume that we can know when positive results are “true” or “false.”
  - But, in practice, these measurements are not always so clear.

- All studies evaluating a new diagnostic test utilize a *standard* to determine when test results are true or false.
  - For example, a study evaluating whether a blood test can determine if a child has ADHD might use the clinical criteria published by a psychiatry professional organization (e.g., in the Diagnostic and Statistical Manual—DSM) defining ADHD.
  - Likewise, a study evaluating whether a new lab test can detect myocardial infarction might use an existing test for myocardial infarction as a standard.
    - In this case, many options for a standard exist, including troponin, creatine kinase, and EKG changes.
  - The **best** standard is often called the *gold standard*.

- When reading a study of a diagnostic test, it is always important to consider which standard is being used.
Refining the Differential Diagnosis

- Sensitivity and specificity describe the performance of diagnostic tests. These can aid in the selection of appropriate tests.

- Once a test result is received, however, how is it best interpreted?
  - If it is positive, what is the probability that the patient has the disease?
  - If it is negative, what is the probability that the patient is free of the disease?
Positive and Negative Predictive Values
Positive Predictive Value (PPV)

- How predictive is a positive test result?

- True positives out of all positives

  \[ \text{PPV} = \frac{TP}{TP + FP} \]
Negative Predictive Value (NPV)

- How predictive is a negative test result?

- True negatives out of all test negatives

  \[ \text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \]
By Dr. Rahul Patwari:
https://www.youtube.com/watch?v=zIn6d3umPGo

Sensitivity and specificity are a characteristic of a given test. PPV and NPV are a characteristic of a given group of patients.

Now, let’s see how all four of these—sensitivity, specificity, PPV, and NPV—can be calculated from the 2x2 table.
Calculation Example

<table>
<thead>
<tr>
<th></th>
<th>Disease (n=100)</th>
<th>No Disease (n=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Result Positive</td>
<td>TP = 90</td>
<td>FP = 20</td>
</tr>
<tr>
<td>Test Result Negative</td>
<td>FN = 10</td>
<td>TN = 80</td>
</tr>
</tbody>
</table>

The four quantities can be calculated as follows.
Sensitivity = $\frac{90}{90+10} = 90\%$
Specificity = $\frac{80}{80+20} = 80\%$
Positive Predictive Value = $\frac{90}{90+20} = 82\%$
Negative Predictive Value = $\frac{80}{80+10} = 89\%$

We will now explore how the context of diagnostic tests can influence our interpretation of these statistics.
Context & Interpretation of Diagnostic Tests
Prevalence and Predictive Values

- The predictive value of a diagnostic test depends on the prevalence of that diagnosis in the population.

- Recall from the first module that prevalence represents the total disease burden in the population and in a population at equilibrium can be calculated as the incidence multiplied by the duration of disease.

- In general, the positive predictive value of a test is greater when the disease is more prevalent in the population.
  - This makes sense. When there is a lot of disease in the population, a positive test result is more likely to represent real disease.
    - For example: A positive HIV test in an area of high HIV prevalence, such as South Africa, is more likely to be truly positive than a positive HIV test in an area of low HIV prevalence, such as rural Minnesota.

- This is easily demonstrated by doubling the prevalence of disease in the example given earlier, as shown on the following slide.
Calculation Example with 2x Disease

<table>
<thead>
<tr>
<th></th>
<th>Disease (n=200)</th>
<th>No Disease (n=100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Result Positive</td>
<td>TP = 180</td>
<td>FP = 20</td>
</tr>
<tr>
<td>Test Result Negative</td>
<td>FN = 20</td>
<td>TN = 80</td>
</tr>
</tbody>
</table>

Sensitivity = $\frac{180}{180+20} = 90\%$
Specificity = $\frac{80}{80+20} = 80\%$

*Test is the same. Sensitivity and specificity do not change.*

PPV = $\frac{180}{180+20} = 90\%$

*However, PPV has increased from 82% (previous example) to 90%*

*Note: When prevalence increases, NPV decreases. In this example, it is 80%, down from 90%.*
Likelihood Ratios
Refining the Differential Diagnosis

- Suppose that a patient comes into the emergency room with fever and right lower quadrant abdominal pain.
  - A differential diagnosis is formed based on the most probable causes of this patient’s symptoms.
  - The most probable diagnoses are specific to a patient population and are affected by prevalence.
    - For example, the probability that lower abdominal pain is due to ectopic pregnancy is extraordinarily different between men and women who present with this complaint.
  - The baseline probability of a diagnosis is called the **pre-test probability** of a diagnosis.
    - Note that the **pre-test probability is also the prevalence of disease in that population**.
    - The pre-test probability that a patient presenting with fever and right lower quadrant pain has an ectopic pregnancy is very different for men and women.

- Suppose that diagnostic testing, such as CT imaging of the abdomen, is obtained for the patient in this example.
  - We can then refine our list of probable diagnoses based on the results of such imaging.
    - The probability of a diagnosis after testing results is called the **post-test probability** for a diagnosis.
Likelihood Ratios

- In diagnostic testing, the *likelihood ratio (LR)*, relates pre-test and post-test probabilities.
  - It is calculated using the sensitivity and specificity of a test.
  - Each test has two different likelihood ratios, one for a positive result and one for a negative result.
  - Likelihood ratios are a type of odds ratio

- **The LR+ refers to the likelihood ratio for a positive test result.**
  - It is calculated as sensitivity/(1-specificity)

- **The LR- refers to the likelihood ratio for a negative test result.**
  - It is calculated as (1-sensitivity)/specificity
Interpreting Likelihood Ratios

- When viewed alone, the LR can be interpreted as follows:
  - An LR of 1.0 does little to change our diagnostic conclusion, since pre-test probability = post-test probability.
  - A high value for LR (whether LR+ or LR-) increases the post-test probability relative to the pre-test probability (means the disease is more likely).
  - A low value for LR (whether LR+ or LR-) decreases the post-test probability relative to the pre-test probability (means the disease is less likely).

- The LR mathematically relates the pre-test and post-test probabilities.
  - You will not need to make these calculations yourself (they aren’t difficult, but involve converting back and forth between odds and probabilities).
  - However, you should be able to use a nomogram, which provides a graphical means to calculate post-test probability given pre-test probability and an LR.
Likelihood Ratio Nomogram

- The nomogram provides a visual way to relate pre- and post-test probabilities using LR data for a given test.

- A line connecting the pre-test probability and the LR will intersect with the post-test probability.

Figure from Center for EBM (http://www.cebm.net/index.aspx?o=1043).
Example 1

LR=1

Post-test probability = Pre-test probability

Figure adapted from Center for EBM (http://www.cebm.net/index.aspx?o=1043).
Example 2

LR=20 (LR>1)

Post-test probability
> Pre-test probability

Figure adapted from Center for EBM (http://www.cebm.net/index.aspx?o=1043).
Example 3

LR=0.05 (LR<1)

Post-test probability < Pre-test probability

Figure adapted from Center for EBM (http://www.cebm.net/index.aspx?o=1043).
Notes on LR+ and LR-

- Note that the previous three examples used the general term “LR.”
  - However, these examples apply for both LR+ (likelihood ratio for a positive test result) and LR- (likelihood ratio for a negative test result).

- It is important to understand the difference between LR+ and LR-.
  - An example on the following slide illustrates the difference between these two statistics.
Imagine that a young man was involved in a minor car accident and presents to the ER with shortness of breath and shaking. The ER physician is trying to decide whether the shortness of breath is due to anxiety or whether the patient sustained lung trauma (e.g., hemothorax) as a result of the accident.

The ER physician knows that chest pain is likely to accompany shortness of breath due to lung trauma. However, it is likely to be absent in shortness of breath due to anxiety.

- For lung trauma, the LR+ for chest pain is >1. That is, the presence of chest pain increases the likelihood of lung trauma as an explanation for shortness of breath.
- For anxiety, the LR+ for chest pain is <1. That is, the presence of chest pain decreases the likelihood of anxiety as an explanation for shortness of breath.

The same relationship can also be illustrated with LR-.

- For lung trauma, the LR- for chest pain is <1. That is, the absence of chest pain decreases the likelihood of lung trauma as an explanation for shortness of breath.
- For anxiety, the LR- for chest pain is >1. That is, the absence of chest pain increases the likelihood of anxiety as an explanation for shortness of breath.

Of course, clinical diagnosis is usually based on more than a single symptom. (In this example, if the ER physician maintains suspicion of lung trauma, he or she should obtain a more detailed history and exam and, if warranted, obtain diagnostic tests.)

The nomogram illustrates how the presence or absence of symptom may increase or decrease the probability of any diagnosis.

Likelihood ratios and nomograms provide a useful way to think about the clinical history, signs, symptoms, and tests.
Receiver Operating Characteristic Curves
So far, we have described test results as being “positive” and “negative.”

However, many real-world test results are continuous.

Consider, for example, the measurement of hemoglobin concentration in a patient’s blood.

In order to use hemoglobin (which may take any value in the range of 0 to >20 g/dL) to determine whether a patient is anemic (yes or no), we have to specify a cut-off.

How do we know what the right cut-off is? What is the effect of changing the cut-off?

The receiver operating characteristic (ROC) curve is a useful tool to understand how continuous variables are used in diagnostic testing.
The ROC curve plots sensitivity (y-axis) against 1-specificity (x-axis) for a range of cut-off values.

Remember that 1-specificity is equal to 1-(TN/[TN+FP]), which is the same as ([TN+FP-TN]/[TN+FP]) or FP/(TN+FP).

Generally, larger values on the x-axis reflect a higher false positive rate.

Remember that sensitivity is equal to TP/(TP+FN).

Generally, larger values on the y-axis reflect a higher true positive rate.
ROC Curve Example

- The figure below shows the ROC curve from a study of the Mayo End-Stage Liver Disease (MELD) score, which is used to measure disease severity in people with chronic alcoholic liver disease.
- The investigators wanted to know whether the MELD score could be used to predict 30-day mortality for adults hospitalized with alcoholic liver disease.
- The sensitivity (true positive rate) and 1-specificity (false positive rate) of different MELD score cut-offs are shown as the stepped line.

The authors of the study suggest that a MELD score of >11 has reasonable sensitivity and specificity to predict 30-day hospital mortality.

- A lower cut-off would result in increased sensitivity and decreased specificity (increased “1-specificity).
  - That is, a lower cut-off would increase both the number of true and false positives. It would move towards the right along the curve.
- A higher cut-off would result in decreased sensitivity and increased specificity.
  - That is, a higher cut-off would decrease both the number of true and false positives. It would move towards the left along the curve.

Interpreting ROC Curves

- The figure to the right shows three theoretical ROC curves.
  - **Curve A** represents the best possible curve.
    - With a cutoff just greater than 0, the test has a perfect true positive rate with no false positives.
    - The test has 100% sensitivity and 100% specificity.
  - **Curve C** represents the worst possible curve.
    - The true and false positive rates remain equal regardless of cut-off.
    - Curve C is similar to flipping a coin to predict the outcome. If heads comes up, the patient has the outcome. If tails comes up, the patient does not.
      - Here, the chance of predicting the true outcome is always equal to the chance of predicting the false one.
  - **Curve B** is a realistic curve. It shows that changing the test cut-off value involves trade-offs between sensitivity and specificity.
    - Importantly, where such trade-offs exist, the determination of the “best” cut-off is somewhat subjective.
    - The “best” cut-off should reflect the purpose of the test.
      - For example, it may be important for a screening test for a serious disease to have a high true positive rate in order not to miss patients with the disease.
      - Similarly, specificity (a decreased false positive rate) may be a priority with a confirmatory test.
Interpreting ROC Curves

- Besides providing a graphical way to examine the effects of changes in cut-off for a continuous test, the ROC curve can also be used to indicate how well a test predicts an outcome.

- The area under the ROC curve (AUC) indicates how well a test predicts an outcome.
  - The AUC has values of 0.5 to 1.0.
  - An AUC of 1.0 means that the test is fully predictive. It corresponds to Curve A on the previous slide.
  - An AUC of 0.5 means that the test is not at all predictive. It corresponds to Curve C on the previous slide.
AUC Example

- The figure below is from a study of how well red cell distribution width (RDW) predicts anemia in children.
  - In this study, anemia is defined as hemoglobin concentration ≤ 10 g/dL.
    - Here, hemoglobin concentration is the “gold standard.”

- The authors find an AUC of 0.83.
  - This can be interpreted as showing that RDW is moderately predictive of anemia in children.

YouTube Video for ROC Curves

- By Dr. Rahul Patwari; about 12 minutes long:

  https://www.youtube.com/watch?v=2Ilgj5Pr6u4
History & physical exam findings can be thought of as diagnostic “tests”

- Answers to questions during a patient history, or findings on physical exam, can be thought of as diagnostic tests

- Clinicians use them “instinctively” to determine approximate post-test probabilities

- Sensitivities, specificities and likelihood ratios for specific H&P findings are available in the medical literature and in a variety of smartphone apps!
Key Points for Module 10

- A highly sensitive test may be used to rule out diagnoses when test results are negative. Sensitivity is a desirable quality for screening tests.

- A highly specific test may be used to rule in diagnoses when test results are positive. Specificity is a desirable quality for confirmatory tests.

- The positive predictive value (PPV) reflects the proportion of positive results which represent true disease. The PPV increases as the prevalence of disease increases.

- Sensitivity and specificity may be used to calculate a likelihood ratio, which can relate the pre-test and post-test probabilities of disease using a nomogram.

- The ROC curve and AUC provide important information about the predictive value of continuous variables, and the overall accuracy of tests.
Please complete the Module 10 quiz

Module Design Team: Martha Carvour, Matthew Rysavy, Timothy Bahr
Revisions & Updates by T. Hegmann
YouTube videos by Dr. Rahul Patwari