Lecture 10

Prof Dr Saad Abed Madhi

Fundamental Concepts: The 2x2 Table

Before diving into the metrics, it's essential to understand the four possible outcomes when a test is applied to a population, often represented in a 2×2 contingency table:

	Disease Present (Actual)	Disease Absent (Actual)	Total
Test Positive	**True Positive (TP)**	**False Positive (FP)**	Test Pos Total
Test Negative	**False Negative (FN)**	**True Negative (TN)**	Test Neg Total
Total	Disease Present Total	Disease Absent Total	Grand Total (N)

- **True Positive (TP):** The test correctly identifies that the disease is present.
- **False Positive (FP):** The test incorrectly identifies that the disease is present (a "false alarm").
- **False Negative (FN):** The test incorrectly identifies that the disease is absent (misses a real case).
- **True Negative (TN):** The test correctly identifies that the disease is absent.

Key Metrics Explained

1. **Sensitivity (Recall / True Positive Rate)**

   - **Definition:** The ability of a test to correctly identify individuals who *have* the disease. It tells us, "Out of all the people who truly have the disease, how many did the test correctly find?"
   - **Formula:** $Sensitivity = \frac{TP}{TP+FN}$

2. **Specificity (True Negative Rate)**

   - **Definition:** The ability of a test to correctly identify individuals who *do not have* the disease. It tells us, "Out of all the people who truly do NOT have the disease, how many did the test correctly find?"
   - **Formula:** $Specificity = \frac{TN}{FP+TN}$

3. **Prevalence**

- **Definition:** The proportion of individuals in a given population who *actually have* the disease at a specific time or over a specific period. It represents the baseline probability of having the disease before any testing.
- **Formula:** $Prevalence = \frac{TP+FN}{N}$ (where N = Total Population)

4. **Accuracy**

- **Definition:** The overall proportion of test results that are correct (both true positives and true negatives). It measures how often the test makes the right call, regardless of whether the disease is present or absent.
- **Formula:** $Accuracy = \frac{TP+TN}{N}$

5. **Positive Predictive Value (PPV)**

- **Definition:** The probability that an individual *actually has* the disease, given that their test result was positive. It answers the question, "If my test is positive, what is the chance I really have the disease?"
- **Formula:** $PPV = \frac{TP}{TP+FP}$
- **Crucial Note:** PPV is highly influenced by the prevalence of the disease in the tested population. In low prevalence settings, even highly sensitive and specific tests can have low PPV.

6. **Negative Predictive Value (NPV)**

- **Definition:** The probability that an individual *does not have* the disease, given that their test result was negative. It answers the question, "If my test is negative, what is the chance I really do NOT have the disease?"
- **Formula:** $NPV = \frac{TN}{FN+TN}$
- **Crucial Note:** NPV is also highly influenced by prevalence. In high prevalence settings, even highly sensitive and specific tests can have low NPV.

Comparing Accuracy Between Two Tests

To compare the accuracy between two tests, you typically calculate the accuracy for each test using the formula $Accuracy = \frac{TP+TN}{N}$ and then compare the resulting percentages. A higher accuracy percentage indicates a better overall performance for that test in correctly classifying individuals.

Important Consideration: While accuracy provides a single overall measure, it can sometimes be misleading, especially if the prevalence is very low or very high. For instance, a test for a very rare disease might show high accuracy simply by classifying most people as "disease absent" (leading to many TNs), even if it misses a few actual cases (FNs). Therefore, it's always recommended to consider sensitivity, specificity, PPV, and NPV alongside accuracy for a complete picture.

Scenario for Examples 1-3:

A new screening test for Disease X is evaluated in a population of 1000 individuals.

- 300 individuals actually have Disease X.
- 700 individuals do not have Disease X.
- The test results show:
  - 270 tested positive and truly have Disease X (TP).
  - 30 tested negative but truly have Disease X (FN).
  - 50 tested positive but do not have Disease X (FP).
  - 650 tested negative and truly do not have Disease X (TN).

Let's set up the 2×2 table first:

	Disease Present	Disease Absent	Total
Test Positive	270 (TP)	50 (FP)	320
Test Negative	30 (FN)	650 (TN)	680
Total	300	700	1000

Example 1: Calculate Sensitivity

Problem: Using the data above, calculate the sensitivity of the test for Disease X.

Given:

- TP = 270
- FN = 30

Calculation:

$$Sensitivity = \frac{TP}{TP+FN} = \frac{270}{270+30} = \frac{270}{300} = 0.90$$

Answer: The sensitivity of the test is 90%. This means the test correctly identifies 90% of individuals who actually have Disease X.

Example 2: Calculate Specificity

Problem: Using the data above, calculate the specificity of the test for Disease X.

Given:

- TN = 650
- FP = 50

Calculation:

$$Specificity = \frac{TN}{FP+TN} = \frac{650}{50+650} = \frac{650}{700} \approx 0.9286$$

Answer: The specificity of the test is approximately 92.9%. This means the test correctly identifies approximately 92.9% of individuals who do not have Disease X.

Example 3: Calculate Accuracy, Prevalence, PPV, and NPV

Problem: Using the data above, calculate the accuracy, prevalence, positive predictive value (PPV), and negative predictive value (NPV) of the test.

Given:

- TP = 270, FP = 50, FN = 30, TN = 650
- N = 1000

Calculations:

- **Accuracy:**
  $$Accuracy = \frac{TP+TN}{N} = \frac{270+650}{1000} = \frac{920}{1000} = 0.92$$
- **Prevalence:**
  $$Prevalence = \frac{TP+FN}{N} = \frac{270+30}{1000} = \frac{300}{1000} = 0.30$$
- **PPV:**
  $$PPV = \frac{TP}{TP+FP} = \frac{270}{270+50} = \frac{270}{320} \approx 0.8438$$
- **NPV:**
  $$NPV = \frac{TN}{FN+TN} = \frac{650}{30+650} = \frac{650}{680} \approx 0.9559$$

Example 4: Impact of Low Prevalence on PPV

Problem: A highly accurate test for a very rare genetic condition has a sensitivity of 99% and a specificity of 98%. If this test is used in a population where the prevalence of the condition is only 0.1% (1 in 1000), what is the PPV?

Given:

- Sensitivity = 0.99
- Specificity = 0.98
- Prevalence = 0.001 (0.1%)
- Assume a population N = 100,000 for easier calculation.

Step 1: Calculate the number of individuals with and without the disease.

- Disease Present (TP + FN) = $N \times Prevalence = 100,000 \times 0.001 = 100$
- Disease Absent (FP + TN) = $N - (TP + FN) = 100,000 - 100 = 99,900$

Step 2: Calculate TP and FN using Sensitivity.

- $TP = Sensitivity \times (TP + FN) = 0.99 \times 100 = 99$
- $FN = (TP + FN) - TP = 100 - 99 = 1$

Step 3: Calculate TN and FP using Specificity.

- $TN = Specificity \times (FP + TN) = 0.98 \times 99,900 = 97,902$
- $FP = (FP + TN) - TN = 99,900 - 97,902 = 1,998$

Step 4: Construct the 2×2 table:

	Disease Present	Disease Absent	Total
Test Positive	99 (TP)	1,998 (FP)	2,097
Test Negative	1 (FN)	97,902 (TN)	97,903
Total	100	99,900	100,000

Step 5: Calculate PPV.

$$PPV = \frac{TP}{TP+FP} = \frac{99}{99+1,998} = \frac{99}{2,097} \approx 0.0472$$

Answer: The PPV is approximately 4.72%. Despite the test having very high sensitivity and specificity, if you test positive for this rare condition, there's only about a 4.7% chance you actually have it due to the low prevalence. This highlights the importance of prevalence in interpreting positive results.

Example 5: Impact of High Prevalence on NPV

Problem: Consider a different scenario: a disease with a very high prevalence of 80% in a certain high-risk group. A new rapid diagnostic test has a sensitivity of 90% and a specificity of 70%. What is the NPV of this test in this group?

Given:

- Sensitivity = 0.90
- Specificity = 0.70
- Prevalence = 0.80 (80%)
- Assume a population N = 1000

Step 1: Calculate the number of individuals with and without the disease.

- Disease Present (TP + FN) = $N \times Prevalence = 1000 \times 0.80 = 800$
- Disease Absent (FP + TN) = $N - (TP + FN) = 1000 - 800 = 200$

Step 2: Calculate TP and FN using Sensitivity.

- $TP = Sensitivity \times (TP + FN) = 0.90 \times 800 = 720$
- $FN = (TP + FN) - TP = 800 - 720 = 80$

Step 3: Calculate TN and FP using Specificity.

- $TN = Specificity \times (FP + TN) = 0.70 \times 200 = 140$
- $FP = (FP + TN) - TN = 200 - 140 = 60$

Step 4: Construct the 2×2 table:

	Disease Present	Disease Absent	Total
Test Positive	720 (TP)	60 (FP)	780
Test Negative	80 (FN)	140 (TN)	220
Total	800	200	1000

Step 5: Calculate NPV.

$$NPV = \frac{TN}{FN+TN} = \frac{140}{80+140} = \frac{140}{220} \approx 0.6364$$

Answer: The NPV is approximately 63.6%. Even with a negative test, in this high-prevalence group, there's still a significant chance (around 36.4%) of having the disease. This is because there are many diseased people in the population, and the specificity isn't perfect, leading to a noticeable number of false negatives relative to true negatives.

Example 6: Finding FN given Sensitivity, TP, and TN

Problem: A medical test has a sensitivity of 95%. In a study, 190 people truly had the disease and tested positive (TP). 800 people truly did not have the disease and tested negative (TN). How many false negatives (FN) were there?

Given:

- Sensitivity = 0.95
- TP = 190
- TN = 800

Calculation:
We know that $Sensitivity = \frac{TP}{TP+FN}$.
We can rearrange this formula to find FN.
$0.95 = \frac{190}{190+FN}$
$0.95 \times (190 + FN) = 190$
$180.5 + 0.95 \times FN = 190$
$0.95 \times FN = 190 - 180.5$
$0.95 \times FN = 9.5$
$FN = \frac{9.5}{0.95} = 10$

Answer: There were 10 false negatives.

Example 7: Comparing Accuracy Between Two Tests

Problem: Two different diagnostic tests for Hepatitis C, Test A and Test B, are evaluated in the same population of 5000 individuals.

- **Test A:**
  - TP = 450
  - FP = 100
  - FN = 50
  - TN = 4400
- **Test B:**
  - TP = 480
  - FP = 200
  - FN = 20
  - TN = 4300

Which test has higher overall accuracy?

Given:

- Total Population (N) = 5000 for both tests.

Calculations:

- **Accuracy for Test A:**
  $Accuracy_A = \frac{TP_A + TN_A}{N} = \frac{450+4400}{5000} = \frac{4850}{5000} = 0.97$

- **Accuracy for Test B:**
  $Accuracy_B = \frac{TP_B + TN_B}{N} = \frac{480+4300}{5000} = \frac{4780}{5000} = 0.956$

Answer: Test A has an accuracy of 97%, while Test B has an accuracy of 95.6%. Therefore, **Test A has a higher overall accuracy** compared to Test B. (Note: Test B has higher sensitivity (480/500 = 96%) but lower specificity (4300/4500 = 95.6%) compared to Test A's sensitivity (450/500 = 90%) and specificity (4400/4500 = 97.8%), demonstrating that overall accuracy is a balance.)

Example 8: Calculating FP given Specificity and TN

Problem: A blood test for diabetes has a specificity of 85%. In a cohort of 600 people without diabetes, 510 correctly tested negative (TN). How many false positives (FP) did the test yield?

Given:

- Specificity = 0.85
- TN = 510
- Total without disease (FP + TN) = 600

Calculation:
We know that $Specificity = \frac{TN}{FP+TN}$.
We can substitute the known values:
$0.85 = \frac{510}{FP+510}$
$0.85 \times (FP + 510) = 510$
$0.85 \times FP + (0.85 \times 510) = 510$
$0.85 \times FP + 433.5 = 510$
$0.85 \times FP = 510 - 433.5$
$0.85 \times FP = 76.5$
$FP = \frac{76.5}{0.85} = 90$

Answer: The test yielded 90 false positives.

Example 9: Calculating Prevalence from TP, FN, and Total Population

Problem: In a study of 500 patients, 120 were diagnosed with a specific condition. Out of these 120, a new diagnostic test correctly identified 100 as positive (TP). The remaining 20 were missed by the test (FN). What is the prevalence of this condition in the study population?

Given:

- TP = 100
- FN = 20
- Total Population (N) = 500

Calculation:
First, find the total number of people with the disease: $TP + FN = 100 + 20 = 120$.
Then, use the prevalence formula:
$Prevalence = \frac{TP+FN}{N} = \frac{120}{500} = 0.24$

Answer: The prevalence of the condition in this study population is 24%.

Example 10: Calculating PPV and NPV with a New Test's Characteristics (Sensitivity, Specificity, and Population Prevalence)

Problem: A new rapid test for flu has a sensitivity of 92% and a specificity of 88%. It is used in a clinic where the prevalence of flu during the season is estimated to be 15%. What are the PPV and NPV for this test?

Given:

- Sensitivity = 0.92
- Specificity = 0.88
- Prevalence = 0.15
- Assume a population N = 1000 for calculation.

Step 1: Determine true disease status in the population.

- Disease Present (TP + FN) = $1000 \times 0.15 = 150$
- Disease Absent (FP + TN) = $1000 - 150 = 850$

Step 2: Calculate TP, FN, FP, TN based on test characteristics.

- $TP = Sensitivity \times (TP + FN) = 0.92 \times 150 = 138$
- $FN = (TP + FN) - TP = 150 - 138 = 12$
- $TN = Specificity \times (FP + TN) = 0.88 \times 850 = 748$
- $FP = (FP + TN) - TN = 850 - 748 = 102$

Step 3: Construct the 2×2 table:

	Disease Present	Disease Absent	Total
Test Positive	138 (TP)	102 (FP)	240
Test Negative	12 (FN)	748 (TN)	760
Total	150	850	1000

Step 4: Calculate PPV and NPV.

- $PPV = \frac{TP}{TP+FP} = \frac{138}{138+102} = \frac{138}{240} = 0.575$
- $NPV = \frac{TN}{FN+TN} = \frac{748}{12+748} = \frac{748}{760} \approx 0.9842$

Answer:

- The PPV for the flu test is 57.5%. (If you test positive, there's a 57.5% chance you have the flu.)
- The NPV for the flu test is approximately 98.4%. (If you test negative, there's a 98.4% chance you do NOT have the flu.)