Learning Objectives

1. Explain the Strength of Recommendation Taxonomy and Levels of Evidence
2. Calculate and interpret relative risk, absolute risk, and number needed to treat (harm)
3. Calculate and interpret sensitivity, specificity, positive predictive value, and negative predictive value
4. Summarize the importance of sensitivity and specificity as inherent characteristics of screening and diagnostic tests

Various "Sources" of Evidence

- Observational Studies
- Systematic Review
- Meta Analysis
- Clinical Experience
- Randomized Control Trial

Basic Research (test tubes, animals)

SORT

Strength of Recommendation Taxonomy

- **Category A**: Recommendation based on consistent and good quality patient-oriented evidence.
- **Category B**: Recommendation based on inconsistent or limited quality patient-oriented evidence.
- **Category C**: Recommendation based on consensus, usual practice, opinion, disease-oriented evidence-based series for studies of diagnosis, treatment, prevention, or screening

Levels of Evidence

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Meta Analysis of RCT</td>
</tr>
<tr>
<td>B</td>
<td>Systematic Review of RCT</td>
</tr>
<tr>
<td>B</td>
<td>Case control study</td>
</tr>
<tr>
<td>B</td>
<td>Case report or case series</td>
</tr>
<tr>
<td>C</td>
<td>Expert opinion</td>
</tr>
<tr>
<td>C</td>
<td>Anecdote</td>
</tr>
</tbody>
</table>

**What the U.S. Preventive Services Task Force Grades Mean**

The U.S. Preventive Services Task Force (USPSTF) grades its recommendations based on the strength of evidence and magnitude of net benefit (benefits minus harms):

- **A**: The USPSTF strongly recommends that clinicians provide [the service] to eligible patients. The USPSTF found good evidence that [the service] improves important health outcomes and concludes that benefits substantially outweigh harms.
- **B**: The USPSTF recommends that clinicians provide [the service] to eligible patients. The USPSTF found at least fair evidence that [the service] improves important health outcomes and concludes that benefits outweigh harms.
- **C**: The USPSTF makes no recommendation for or against routine provision of [the service]. The USPSTF found at least fair evidence that [the service] can improve health outcomes but concludes that the balance of benefits and harms is too close to justify a general recommendation.
- **D**: The USPSTF recommends against routinely providing [the service] to asymptomatic patients. The USPSTF found at least fair evidence that [the service] is ineffective or that harms outweigh benefits.
- **I**: The USPSTF concludes that the evidence is insufficient to recommend for or against routinely providing [the service]. Evidence that [the service] is effective is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.
1. Results of a clinical study show a relative risk reduction (RR = 0.67) of 33% and an absolute risk reduction (AR = 0.8) of 20%. There are 1000 patients each in the treatment and control groups. To help determine the potential benefit of the treatment, it is necessary to identify the number needed to treat (NNT).

Which one of the following is the NNT for this clinical study?

A. 3  
B. 5  
C. 13  
D. The number cannot be determined from the information provided

Calculations

Contingency Table (2X2 Table)

- Relative Risk (RR)
  - Increase
  - Decrease or Reduction
- Absolute Risk (AR)
  - Increase
  - Decrease or Reduction
- Number Needed to Treat/Harm (NNT)/(NNH)

Measures of Risk

- Relative risk
  - States by how many times the intervention or exposure increases (decreases) the risk of the outcome
  - The ratio of the incidence of the outcome among exposed persons to the incidence among unexposed
- Absolute risk
  - Event RATE in control group – Event RATE in experimental group

Calculating Relative Risk

- Consider a study that examines the risk factors for breast cancer among women participating in the first National Health and Nutrition Examination Survey (Prospective Cohort Study)
  - In a sample of 4,540 women who gave birth to their first child before the age of 25, 65 developed breast cancer
  - Of the 1,626 women who first gave birth at age 25 or older, 31 were diagnosed with breast cancer

What Is “Exposure” in This Case?

- We consider “exposure” to be the condition of having first given birth at age 25 or older
  - a. Construct a 2X2 contingency table for these data
  - b. Estimate the relative risk of developing breast cancer
The Relative Risk

\[ RR = \frac{P(\text{disease | exposed})}{P(\text{disease | unexposed})} \]

Calculating RR

\[ RR = \frac{\text{EER}}{\text{CER}} \]

Calculation - Example

<table>
<thead>
<tr>
<th>Disease Outcome</th>
<th>Risk(Exposure)</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast Cancer</td>
<td>31</td>
<td>1628</td>
</tr>
<tr>
<td>No Breast Cancer</td>
<td>1597</td>
<td></td>
</tr>
<tr>
<td>First birth ≥ 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Exposed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>4540</td>
</tr>
<tr>
<td>First birth &lt; 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Nonexposed)</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4475</td>
<td></td>
</tr>
</tbody>
</table>

Calculation

\[ RR = \frac{31}{1628} \]
\[ = \frac{65}{4540} \]
\[ = 1.33 \]

- A RR of 1.33 implies that women who gave birth at 25 years of age or older are 33% more likely to develop breast cancer than women who gave birth at a younger age.
- In general, a RR of 1.0 indicates that the probabilities of disease in the exposed and unexposed groups are identical; consequently, an association between the exposure and the disease does not exist.

Calculating Number Needed to Treat

\[ \text{NNT} = \frac{1}{\text{ARR}} \]

- NNT - "The number of patients needed to treat with the study treatment for the specified length of time in order to produce the good outcome specified."

Why Do You Need to Know the Difference – RR, ARR?

- Trial reports and ads often present results as RR, which can be misleading
- ARR and NNT give a much better appreciation of the magnitude of benefit and potential impact
- For most clinical trials, ARR and NNT are easily calculated
Calculating Number Needed to Harm

Control Event Rate \( CER = \frac{\# \text{ events in placebo group}}{\# \text{ subjects in placebo group}} \)

Experimental Event Rate \( EER = \frac{\# \text{ events in study group}}{\# \text{ subjects in study group}} \)

Relative Risk \( RR = \frac{EER}{CER} = \frac{IE}{IC} \)

Relative Risk Reduction (or Increase) - CHD

\[ \text{Relative Risk (RR)} = \frac{EER}{CER} \]

\[ \frac{0.019}{0.015} = 1.2666 \]

Event E/P Incidence Placebo Incidence RR

<table>
<thead>
<tr>
<th>Event</th>
<th>E/P</th>
<th>Incidence</th>
<th>Placebo</th>
<th>Incidence</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD</td>
<td>164</td>
<td>0.019</td>
<td>122</td>
<td>0.015</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Relative Risk (RR) = EER/CER

NNT(H) - CHD

\[ \text{NNT} = \frac{1}{ARR \times \%ARR} \]

\[ \text{ARR} = \frac{CER - EER}{EER} \]

\[ 0.015 - 0.019 = 0.004 \]

\[ 0.004 \times 100 = 4\% \]

\[ \text{NNT} = \frac{1}{0.004} \times 100 = 250 \]

Absolute Risk Reduction (or Increase) CHD

\[ \text{ARR} = \frac{CER - EER}{EER} \]

\[ = 0.015 - 0.019 \]

\[ = 0.004 \]

What Do the Numbers Mean?

Relative Risk Reduction (Increase) (RR)

Patients taking HRT for 5.2 years have 30% more MIs than patients taking placebo.

Absolute Risk Reduction (Increase) (ARR)

1.9% of patients taking HRT for 5.2 years had an MI compared with 1.5% not taking HRT.

Number Needed to Harm (NNH)

If 250 women took HRT for 5.2 years, the drug would cause one additional heart attack.

The Cochrane Library periodically reviews the evidence of the effectiveness of mammography screening for detection of breast cancer. One analysis from their 2006 review looked at deaths ascribed to breast cancer for women at least 50 years of age who had screening mammography, generally every 1–3 years, compared with women who did not. After 13 years of follow-up, 595 out of 146,284 women in the screened group died of breast cancer, compared with 701 of 122,590 unscreened women.

According to this data, the number needed to screen with mammography to prevent 1 death from breast cancer over 13 years of follow-up is approximately:

\[ \text{NNT} = \frac{1}{ARR \times \%ARR} \]

\[ \frac{1}{0.004} \times 100 = 250 \]

*ARR = CER - EER
**Calculating Number Needed to Screen**

<table>
<thead>
<tr>
<th>Control Event Rate</th>
<th>CER (= I_1)</th>
<th># events in placebo group</th>
<th># subjects in placebo group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Event Rate</td>
<td>EER (= I_2)</td>
<td># events in study group</td>
<td># subjects in study group</td>
</tr>
</tbody>
</table>

\[
RR = \frac{EER}{CER} = \frac{I_2}{I_1}
\]

**Absolute Risk Reduction**

\[
ARR = CER - EER
\]

\[
NNT = \frac{1}{ARR} \text{ OR } \frac{100}{\%ARR}
\]

**NNS** - "The number of patients needed to screen with the study test for the specified length of time in order to produce the good outcome specified"

**Calculation - Example**

<table>
<thead>
<tr>
<th>Disease Outcome</th>
<th>Breast Cancer</th>
<th>No Breast Cancer</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screened</td>
<td>595</td>
<td>145,689</td>
<td>146,284</td>
</tr>
<tr>
<td>NOT Screened</td>
<td>701</td>
<td>121,889</td>
<td>122,590</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1,296</td>
<td>267,578</td>
<td>278,874</td>
</tr>
</tbody>
</table>

**Absolute Risk Reduction**

**Breast Cancer**

\[
ARR = CER - EER
\]

\[
= 701/122,590 - 595/146,284
\]

\[
= 0.00572 - 0.00407
\]

\[
= 0.00165
\]

Accurate way of determining the benefit of screening is to examine the absolute (or attributable) risk reduction (ARR), which is the difference between the rates of occurrence of two conditions.

**NNS - Breast Cancer**

The number needed to screen is the reciprocal of the ARR.

\[
NNS = \frac{1}{ARR} \text{ OR } \frac{100}{\%ARR}
\]

\[
= \frac{1}{0.00165}
\]

\[
= 606
\]

*This data would suggest that approximately 606 women over age 50 would need to be screened with mammography every 1–3 years over 13 years of follow-up in order to prevent 1 death from breast cancer.*

2. The specificity of a screening test is best described as the proportion of persons:

A. with the condition who test positive
B. with the condition who test negative
C. without the condition who test positive
D. without the condition who test negative

A. 46%
B. 30%
C. 4%
D. 45%

© American Academy of Family Physicians. All Rights Reserved.
Sensitivity and Specificity
• Both measures of a test's validity – its ability to correctly detect people with or without the disease in question
• 4 logical possibilities in diagnostic testing

<table>
<thead>
<tr>
<th>Disease</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Type 1 error (α error)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
<tr>
<td>Type II error (β error)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity
• Ability to detect people who do have the disease
• The % of people with a disease that is correctly detected or classified:

\[
\text{Sensitivity} = \frac{TP}{TP+FN} \times 100
\]

Specificity
• Ability to detect people who do not have the disease
• The % of the disease-free people who are correctly classified or detected:

\[
\text{Specificity} = \frac{TN}{FP+TN} \times 100
\]
Sensitivity and Specificity

Clinical Practice

• In clinical practice, sensitivity and specificity are inversely related: an increase in one causes a decrease in the other.
  – This is because groups of patients with the disease and groups who are disease-free lie in a continuum, overlapping each other, rather than forming two distinctly different groups.
  – The tester therefore has to select a cut-off point to make a diagnostic decision.

Screening Tests

Sequence

• Although there are tests of relatively high sensitivity and specificity for some diseases, it is often best to use a combination of tests when screening for/diagnosing a particular disease.

Screening Tests

Sequence

• A highly sensitive (and usually relatively inexpensive) test should be used first, almost guaranteeing the detection of all cases of the disease (albeit at the expense of including a number of false-positive results).
• This should be followed by a more specific test (and usually more expensive test) to eliminate the false-positive results.
  – e.g. this is the usual sequence if testing for HIV, Hepatitis B, and many other common but serious diseases.

Predictive Values

• When the sensitivity of a test is known...
  – “Given that a patient has the disease, what is the ability of the test to discover this?”
• When the specificity of a test is known...
  – “Given that a patient is free of the disease, what is the ability of the test to discover this?”

Positive Predictive Value

• The proportion of positive results that are true positives
  – The likelihood that a person with a positive test actually has the disease
  \[
  PPV = \frac{TP}{TP+FP}
  \]

Negative Predictive Value

• The proportion of negative results that are true negatives
  – The likelihood that a person with a negative test truly does not have the disease
  \[
  NPV = \frac{TN}{FN+TN}
  \]
3. In a study to evaluate a test as a screen for the presence of a disease, 235 of the 250 people with the disease had a positive test and 600 of the 680 people without the disease had a negative test. Based on this data, the specificity of the test for the disease is:

A. $\frac{235}{250}=94\%$
B. $\frac{15}{250}=6\%$
C. $\frac{600}{680}=88\%$
D. $\frac{80}{680}=12\%$

### 2 x 2 Table

<table>
<thead>
<tr>
<th>Disease (+)</th>
<th>Disease (-)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test (+) for disease</td>
<td>235</td>
<td>80</td>
</tr>
<tr>
<td>Test (-) for disease</td>
<td>15</td>
<td>600</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>680</td>
</tr>
</tbody>
</table>

The specificity of a test for a disease is the proportion or percentage of those without the disease who have a negative test.

### Example: Test for DM

**Fasting Plasma Blood Glucose**

<table>
<thead>
<tr>
<th>Test Positive</th>
<th>Test Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+) DM</td>
<td>182</td>
<td>18</td>
</tr>
<tr>
<td>(-) DM</td>
<td>24</td>
<td>166</td>
</tr>
</tbody>
</table>

**Contingency Table**

- Sensitivity of the test $= \frac{TP}{TP+FN} = \frac{182}{182+18} = 91\%$
- Specificity of the test $= \frac{TN}{FP+TN} = \frac{166}{24+166} = 83\%$
- $PPV = \frac{TP}{TP+FP} = \frac{182}{182+24} = 88\%$
- $NPV = \frac{TN}{FN+TN} = \frac{166}{166+18} = 90\%$

### Sensitivity/Specificity

**PPV/NPV**

- As the prevalence of the disease in a population increases
  - PPV increases
  - NPV decreases
- As the prevalence of the disease in the population decreases
  - PPV decreases
  - NPV increases
Answers

1. B
2. D
3. C