Reliability of Screening Tests

Reliability (also known as precision):

- The extent to which the screening test will produce the same or very similar results each time it is administered

- A test must be reliable before it can be valid (an invalid test can demonstrate high reliability: the answer is wrong each time)

Sources of variability that can affect the reproducibility of results of a screening test:

1. Biological variation (e.g. blood pressure, cholesterol)
2. Reliability of the instrument itself
3. Intra-observer variability (differences in repeated measurement by the same screener)
4. Inter-observer variability (inconsistency in the way different screeners apply or interpret test results)
**Measuring the Performance (Yield) of a Test**

- The proportion of persons who screen positive on the test who actually have Disease
  - Predictive Value Positive (PV+ or PPV)

- The proportion of persons who screen negative on the test who are actually free of the Disease
  - Predictive Value Negative (PV- or NPV)

**Performance (Yield)**

<table>
<thead>
<tr>
<th>True Disease Status</th>
<th>Results of screening test</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Positive predictive value (PPV): Among persons who test positive, the probability that a person actually has the disease: \( PPV = \frac{a}{a + b} \)

Negative predictive value (NPV): Among persons who test negative, the probability that a person does not have the disease: \( NPV = \frac{d}{c + d} \)
### Performance (Yield)

#### True Disease Status

<table>
<thead>
<tr>
<th></th>
<th>Test+</th>
<th>Test-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>400</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>Negative</td>
<td>995</td>
<td>98,905</td>
<td>100,400</td>
</tr>
<tr>
<td>Total</td>
<td>1,395</td>
<td>99,005</td>
<td>100,400</td>
</tr>
</tbody>
</table>

**Sensitivity:** \( \frac{a}{a + c} = \frac{400}{400 + 100} = 80\% \)

**Specificity:** \( \frac{d}{b + d} = \frac{98,905}{995 + 98,905} = 99\% \)

**PPV:** \( \frac{a}{a + b} = \frac{400}{400 + 995} = 29\% \)

**NPV:** \( \frac{d}{c + d} = \frac{98,905}{100 + 98,905} = 99.9\% \)

### Factors that Influence PPV and NPV

1. The more specific the test, the higher the PPV
2. The more sensitive the test, the higher the NPV
3. The higher the prevalence of pre-clinical disease in the screened population, the higher the PPV (decreases NPV)

### Example: Performance (Yield)

#### True Disease Status

<table>
<thead>
<tr>
<th></th>
<th>Test+</th>
<th>Test-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>400</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>Negative</td>
<td>995</td>
<td>98,905</td>
<td>100,400</td>
</tr>
<tr>
<td>Total</td>
<td>1,395</td>
<td>99,005</td>
<td>100,400</td>
</tr>
</tbody>
</table>

**PPV:** \( \frac{a}{a + b} = \frac{400}{400 + 995} = 28.6\% \)

**NPV:** \( \frac{d}{c + d} = \frac{98,905}{100 + 98,905} = 99.9\% \)

Prevalence = 500/100,400 x 100 = 0.498 x 100 = 4.98 per 1000
Example 2: Performance (Yield)

<table>
<thead>
<tr>
<th>Results of Screening</th>
<th>True Disease Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>4000 954 4,954</td>
</tr>
<tr>
<td>-</td>
<td>1000 94,446 95,446</td>
</tr>
<tr>
<td></td>
<td>5000 95,400 100,400</td>
</tr>
</tbody>
</table>

PPV: \[
\frac{a}{a + b} = \frac{4000}{4000 + 954} = 80.7\% 
\]
NPV: \[
\frac{d}{c + d} = \frac{94,446}{1000 + 94,446} = 98.95\% 
\]
Prevalence = \[
\frac{5000}{100,400} \times 100 = 4.98 \text{ per 100}
\]

Performance (Yield)

<table>
<thead>
<tr>
<th>Prevalence (%)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>90%</td>
<td>95%</td>
<td>1.8%</td>
</tr>
<tr>
<td>1.0</td>
<td>90%</td>
<td>95%</td>
<td>15.4%</td>
</tr>
<tr>
<td>5.0</td>
<td>90%</td>
<td>95%</td>
<td>48.6%</td>
</tr>
<tr>
<td>50.0</td>
<td>90%</td>
<td>95%</td>
<td>94.7%</td>
</tr>
</tbody>
</table>

Articulate Quizmaker Quiz Placeholder - review1_lecture12_part2
Efficacy of Screening

- Efficacy of screening can be measured by the degree to which it benefits those who are screened compared to those who are not.

- Most convincing way to evaluate the efficacy of screening is through a RCT.

- However, it could be too expensive, not feasible or present ethical limitations.

Effectiveness of a Screening Test

Definition: How well the screening test performs under actual conditions. (Observational studies.)

- The Natural History of Disease (NHD) must provide adequate lead time to lead to an improved outcome (reducing morbidity and mortality) following early diagnosis.

- Medical support must be available to actually improve the outcome.

Effectiveness of a Screening Test

Effectiveness of screening

1. Overall shift in severity of disease at the time of diagnosis (disease will be detected earlier, so more people should be in early sub-clinical manifestations of disease)

2. Compare cause-specific mortality among those whose disease was picked up by screening versus those with a diagnosis related to symptoms
Effectiveness of Screening

Sources of bias in evaluating screening programs:

1. Self-selection bias (volunteer bias)
2. Lead time bias
3. Length bias

Self-selection Bias (Volunteer Bias)

• Volunteers for screening programs may be healthier, on average, than persons who do not participate in screening programs.

On the other hand,…

• The “worried well” may be more likely to participate and may be at overall higher risk due to family history or lifestyle characteristics

Lead Time Bias

Lead Time:

• The interval between detection of disease at screening and when it would have been diagnosed from clinical symptoms

• Survival may appear to be increased among screen-detected cases simply because diagnosis was made earlier in the course of the disease
Diagram shows how probability of detection is related to rate of disease progression. Length of each arrow represents length of detectable preclinical phase, from initial detectability to clinical diagnosis (Dx). Testing at a single moment detects four slowly progressive cases but only two rapidly progressive cases. Cases not detected by test (thin arrows) are diagnosed clinically either before or after time of testing. Thick arrows indicate detected cases.

**Length Bias**

- The overrepresentation among screen-detected cases of those with a long pre-clinical phase, and thus a more favorable prognosis

- Those with a long pre-clinical phase are more readily detectable by screening than more rapidly progressing cases with a short pre-clinical phase

---

**Summary of Screening for Disease Detection**

- Relatively quick way of detecting potential disease in a population before it manifests.

- Validity measured by sensitivity and specificity (depends on a Gold Standard measure).

- Precision has to do with consistency or stability of results (errors due to method, subject and/or observer variability).

- Have benefits and risks.

- Efficacy determined by RCT. (Effectiveness: volunteer, lead time and length time bias).
Articulate Quizmaker Quiz Placeholder - review2_lecture12_part2