EVALUATION of DIAGNOSTIC TESTS or Clinical Epidemiology

VEM5503 Epidemiology
D. O. Rae
Suggested Reading


Objectives

• Be able to define the **operating characteristics for diagnostic tests in words and algebraically**.

• Given the appropriate data, be able to **construct 2X2 tables and calculate sensitivity, specificity, predictive values and apparent prevalence**.

• Be able to **discuss test selection** with reference to a screening and/or diagnostic test and understand the effects of the **prevalence of disease** on the tests characteristics.
How do you arrive at a diagnosis?

- Subjective measures?
- Objective measures?
- Physical diagnosis?
- Laboratory testing?
- Classification of results...
Interpretation of Clinical Data

• Initial considerations:
  • Test results
    • Are results accurate? ...precise?
  • Patient status
    • Is the patient unusual? ...diseased? ...treatable?
  • Test values
    • Will there be a regression to the mean?
    • Are the results ‘abnormal’?
Accuracy and Precision

Do all test results provide equal confidence in their outcome value?
Accuracy and Precision

Accuracy of a test

- relates to its ability to give a **true measure** of the substance being evaluated (e.g. lipase, antibodies).
- To be 'accurate' a test need not always be close to the true value but if a number of tests are run, the average of the results should be close to the true value.

Precision of a test

- The precision of a test relates to how **consistent** the results from the test are.
- If a test always gives the same value for a sample (regardless of whether or not it is the correct value), it is said to have precision.
Example: Accurate and Precise

- Range of Values = Clumped and Close
- True Value =
Example: Inaccurate and Precise

- Range of Values = **Clumped but Not Close**
- True Value =
Example: Accurate but not Precise

- Range of Values =
- True Value =

Not clumped but on average True
Example: Inaccurate and not Precise

- Range of Values = Not clumped and Not True
- True Value =

*
Criteria for Abnormality

• Abnormal as Unusual… less frequent
  • What is 'unusual'? 
  • How is it measured? 
  • Statistical deviation from the mean?
Abnormal as Unusual

• What is usual/normal Somatic Cell Count for a dairy cow (>5 yrs)?

• 150-250,000 cells/ml often is classified as normal because older cows usually have counts higher than young cows (possibly due to chronic damage to the udder as a result of previous infections).

• **However**, some older cows have SCC <100,000 cells/ml.
What is 'unusual'?

- Golden retrievers (>4 yrs of age) are often considered normal when 5 kg overweight because it is common.
- Is it 'abnormal'?  
- Is it associated with disease?  
- Is it treatable?
Amylase value distribution in a canine population.

Is it associated with disease?
Can we treat elevated Amylase without disease?
So, do we consider it normal?
Regression to the Mean

- Due to normal biological variation, the values for a parameter will fluctuate within an individual.
- Abnormal values are often rechecked.
- Due to the normal variation, it is likely that the follow-up test will reveal a more 'normal' value.
- This is referred to as a regression to the mean.
- How does this influence our impressions on efficacy of treatment?
Example: Regression to the Mean

- A sick horse is presented with a potassium of 5.0 mEq/L, upper level of 'normal' is 3.5 mEq/L.
- The animal is placed on IV fluids with dextrose.
- A recheck is done at 12 hrs
- The potassium has dropped to 3.3 mEq/L.
- Was the change due to therapy or regression to the mean?
How do we establish normal and abnormal test results?
Clinical measurements:

• Interval scale data
  • heart rate, serum Ca+, K, BUN, total protein.

• Ordinal scale data
  • California mastitis test
  • urine dipstick (proteinuria, ketonuria)

• Nominal scale data
  • dichotomous response (yes/no)
  • culture results, pregnancy status.
Clinical measurements:

- Interval and ordinal scale data can be reduced to a nominal scale.
  - normal / abnormal dichotomy
- an ideal distribution clearly demarcates
  - normal/abnormal
  - healthy/diseased
  - infected/non-infected.
## Calf immunoglobulins / total proteins

<table>
<thead>
<tr>
<th>C#</th>
<th>IGG</th>
<th>TP</th>
<th>IG2000</th>
<th>TP5.5</th>
<th>TP6.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0</td>
<td>3.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>670</td>
<td>5.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1344</td>
<td>5.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>995</td>
<td>5.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>99</td>
<td>1545</td>
<td>5.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>593</td>
<td>5.6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1048</td>
<td>5.6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>149</td>
<td>1766</td>
<td>5.6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>49</td>
<td>1924</td>
<td>5.6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>61</td>
<td>1690</td>
<td>5.9</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>106</td>
<td>1690</td>
<td>5.9</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1545</td>
<td>6.0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>1545</td>
<td>6.0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>128</td>
<td>1924</td>
<td>6.5</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>69</td>
<td>2549</td>
<td>7.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>3414</td>
<td>7.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>3657</td>
<td>7.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>125</td>
<td>4176</td>
<td>7.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>3182</td>
<td>7.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Amylase value distribution in a canine population.

AVC Diagnostic Lab- 1000 samples

Amylase (u/l)
$X$ value distribution in a canine population.
Classification of dichotomous observations

• **State of Nature**
  - If we were all-knowing and all-seeing,
  - what would the actual state of nature be?
  - (Diseased (D) or non-diseased (n)).

• **Observable/ Detectable classification**
  - A very human (doubt shrouded) method of "state" detection is employed.
  - (Test positive (T+) or negative (T-)).

• **How well does our classification match the true state of nature?**
State of Nature

n

D
Detection of “State”
Detection of “State”
Detection of “State”

- $a =$ true positive $T^+$ in $D$ (Diseased)
- $d =$ true negative $t^-$ in $n$ (Non-diseased)
- $b =$ false positive $T^+$ in $n$
- $c =$ false negative $t^-$ in $D$
Sensitivity (Se) is the probability of a test correctly identifying those animals that are diseased.
Sensitivity

• \( P(T+/D) = \frac{a}{a+c} \)

• The conditional probability of a positive test, given disease.

• The likelihood of a positive test in a diseased animal.

• The true positive rate (relative to all animals with disease).

• The proportion of animals with disease that have positive test for the disease.
Specificity (Sp) is the probability of a test correctly identifying those animals that are *not* diseased.
Specificity

- $P(t-/n) = \frac{d}{b+d}$
- The conditional probability of a negative test, given the absence of disease.
- The likelihood of a negative test in an animal without disease.
- The proportion of animals without disease that have a negative test.
- The true negative rate (relative to all animals without disease).
A sensitive test will rarely misclassify animals *with* the disease.
A specific test will rarely misclassify animals \textit{without} the disease.
Where do we set the cut-off?
Establishment of "Discriminant" Value

- **Screening:**
  - identification of unrecognized disease.
  - prevalence survey

- **Case finding:**
  - identification of disease in "high risk group"

- **Diagnosis:**
  - aid in disease confirmation.
What will the cost of false positives be?
What will the cost of false negatives be?
Detection of “State”
Detection of “State”
What type of test would you want ... 

...for HIV testing?

or

...for a disease where test positive animals are slaughtered?
# 2 X 2 Classification

<table>
<thead>
<tr>
<th>Test</th>
<th>Nature</th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td>t-</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
<td>N</td>
</tr>
</tbody>
</table>

The table above represents a 2 x 2 classification table, where:

- **D+**: Test positive
- **n-**: Test negative
- **T+**: True positive
- **t-**: True negative
- **a**: True positive
- **b**: False positive
- **c**: False negative
- **d**: True negative
- **N**: Total sample size

The diagram illustrates the distribution of samples based on the test results and nature, highlighting the categories of true positives, true negatives, false positives, and false negatives.
## 2 X 2 Classification

<table>
<thead>
<tr>
<th>Test</th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>t-</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
</tr>
</tbody>
</table>

- \( a = P(DT+) \)  \quad c = P(Dt-) 
- \( b = P(nT+) \)  \quad d = P(nt-) 
- Sensitivity =  \( P(T+|D) = \frac{a}{a+c} \) 
- Specificity =  \( P(t-|n) = \frac{d}{b+d} \)
### 2 X 2 Classification

<table>
<thead>
<tr>
<th>Nature</th>
<th>Test</th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>40</td>
<td>11</td>
<td>51</td>
</tr>
<tr>
<td>t-</td>
<td>4</td>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>56</td>
<td>100</td>
</tr>
</tbody>
</table>

- \(a = P(DT+)\) \(c = P(Dt-)\)
- \(b = P(nT+)\) \(d = P(nt-)\)
- Sensitivity = \(P(T+|D) = \frac{a}{a+c} = \frac{40}{44} = 0.90\)
- Specificity = \(P(t-|n) = \frac{d}{b+d} = \frac{45}{56} = 0.80\)
Misclassification

- false positive \( P(T+/n) = \frac{b}{b+d} \)
- false negative \( P(t-/D) = \frac{c}{a+c} \)

\[
\begin{array}{ccc}
D+ & n- \\
T+ & 40 & 11 & 51 \\
t- & 4 & 45 & 49 \\
& 44 & 56 & 100 \\
\end{array}
\]
Clinical Questions:

• How common is the disease in the population of animals?

• What is the likelihood of disease when the test is positive?

• What is the likelihood of not having disease when the test is negative?

• A knowledge of the operating characteristics of a test must be coupled with an estimate of disease probability before these questions can be answered.
Define Prevalence.

• The number of instances of a given disease or other condition at a designated time.

• The total number of all individuals who have an attribute or disease at a particular time divided by the population at risk of having the attribute or disease.
Apparent Prevalence- the proportion of all animals that give a positive test result.

- We determine ... apparent prevalence.

- $P(T+) = \frac{a+b}{N}$

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>40</td>
<td>11</td>
<td>51</td>
</tr>
<tr>
<td>t-</td>
<td>4</td>
<td>45</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>56</td>
<td>100</td>
</tr>
</tbody>
</table>
**We want to know ...**

true prevalence.

- \( P(D) = \frac{a+c}{N} \)

```
<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>40</td>
<td>11</td>
<td>51</td>
</tr>
<tr>
<td>t-</td>
<td>4</td>
<td>45</td>
<td>49</td>
</tr>
</tbody>
</table>

44 56 100
```
True Prevalence- the proportion of all animals that are "Diseased".

- **Do we know the true prevalence?**
- Unfortunately we never do.
- Apparent prevalence may be
  - more than
  - less than or
  - equal to the true prevalence.
Approximation of true prevalence:

- $P(D) = P(T+) + Se - 1 \over Sp + Se - 1$
- $Se = 0.90$
- $Sp = 0.80$
- $P(T+) = 0.51$
- $P(D) = 0.44$

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>40</td>
<td>11</td>
</tr>
<tr>
<td>t-</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>56</td>
</tr>
</tbody>
</table>
Predictive value (PV)

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>40</td>
<td>11</td>
</tr>
<tr>
<td>t-</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>56</td>
</tr>
</tbody>
</table>

\[
P(D/T+) = \frac{a}{a+b} \\
P(d/t-) = \frac{d}{c+d}.
\]

PV+ = 0.78
PV- = 0.92
PV of a positive test:

- the proportion of test positive animals which have the disease.
  - $P(D/T+) = \frac{a}{a+b}$

- Positive predictive value is:
  - The predictive value of a positive test.
  - The post-test likelihood of disease following a positive test.
  - The post-test probability of disease following a positive test.
PV of a negative test:

• the proportion of test negative animals which do not have the disease.
  • \[ P(n/t-) = \frac{d}{c+d} \]

• Negative predictive value is:
  • The post-test likelihood of no disease following a negative test.
  • The post-test probability of no disease following a negative test.
Effect of Se and Sp on PV:

- The more sensitive a test, the better the negative predictive value.
- The more specific a test, the better the positive predictive value.

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>40</td>
<td>11</td>
</tr>
<tr>
<td>t-</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>56</td>
</tr>
</tbody>
</table>
**Effect of *prevalence* on predictive value:**

- Positive results, even for *highly specific* tests, will be largely false positives if the pretest probability of disease is low.

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>t-</td>
<td>2</td>
<td>980</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>990</td>
</tr>
</tbody>
</table>
Effect of *prevalence* on predictive value:

- Negative results, even for a *highly sensitive* test, will be largely false negatives when applied to patients from a high prevalence group.

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T+</strong></td>
<td>85</td>
<td>1</td>
</tr>
<tr>
<td><strong>t-</strong></td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>10</td>
</tr>
</tbody>
</table>
Predictive value (PV) of positive test results vary directly with the prevalence of disease when Se and Sp are held constant.

**Example:**

<table>
<thead>
<tr>
<th>Prevalence (%)</th>
<th>50.0</th>
<th>5.0</th>
<th>2.0</th>
<th>1.0</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV+</td>
<td>95.0</td>
<td>50.0</td>
<td>27.9</td>
<td>16.1</td>
<td>1.9</td>
</tr>
</tbody>
</table>
Example 1:

- True Prevalence = 50%  \( \text{Se} = 0.95 \quad \text{Sp} = 0.99 \)

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>95</td>
<td>1</td>
</tr>
<tr>
<td>t-</td>
<td>5</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Predictive Value = \( \frac{95}{96} \approx 99\% \)
Example 2:

- True Prevalence = 1%  Se=0.95  Sp=0.99

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>n-</th>
</tr>
</thead>
<tbody>
<tr>
<td>T+</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>t-</td>
<td>1</td>
<td>1960</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1980</td>
</tr>
</tbody>
</table>

- Predictive value = $\frac{19}{39} \approx 50\%$
Predictive value decreases as prevalence decreases.
Prevalence change: Disease Eradication

• Disease eradication outset:
  • Disease prevalence- high
  • Positive predictive value high,
• With control, tail end of a eradication:
  • Disease prevalence- low (drops)
  • Positive predictive value decreases.
Example: TB test

- Caudal tail fold test, initial test
  - high Se

- Comparative cervical test, follow-up
  - high Sp
Example: Brucellosis test

- Card Test
  - high Se, low cost

- Tube Agglutination test
  - high Sp

- CF test
  - high Sp