Diagnostic Studies

Study Design and Analysis

Steven Greenway, MSc, MD, FRCPC
Alberto Nettel-Aguirre, PhD

November 2, 2021
Objectives

• Understand the issues in the evaluation of a diagnostic test
• Appreciate components of evaluating test performance
  • precision and accuracy
  • sensitivity and specificity
  • likelihood ratio
  • positive and negative predictive values
  • receiver operating characteristic (ROC) curves
  • additional factors: cost, availability, acceptability, utility
Examples of Diagnostic Tests

• **Biochemical**
  • electrolytes, urea, creatinine

• **Imaging**
  • CXR, MRI

• **Genetic**
  • karyotype, array, WES

• **Microbiological**
  • blood culture

• **Physiological**
  • PFTs, exercise test, GTT

• **Clinical**
  • Lever sign to diagnose ACL tear

• **Patient-reported outcome measures**
  • questionnaire of symptoms to diagnose IBD
Purpose of a Diagnostic Test

• **Diagnose a disease or condition**
  • TSH
  • echocardiogram

• **Exclude a disease or condition**
  • HbA1C
  • Troponin

• **Estimate prognosis**
  • LDL cholesterol
  • BRCA1 mutation

• **Inform treatment decisions**
  • PSA
  • karyotype
Factors Affecting Diagnostic Test Performance

• Prevalence of the disease in the population
• Spectrum of the disease
• Often dependent on other factors
  • part of diagnostic pathway
  • test results may not be independent
  • often depend on prior knowledge
• Gold standard
  • established test which confirms the diagnosis

Barrett and Fardy, 2021
Types of Studies to Evaluate a Diagnostic Test

• **Precision (reproducibility)**
  - intra-observer (amount of variation for a single observer)
  - inter-observer (variation between 2 or more observers)

• **Accuracy**
  - cohort
  - case-control

• **Costs, Risks and Acceptability**
  - prospective
  - retrospective

• **Improvement of clinical outcome**
  - RCT
  - case-control
Precision

• Reproducibility or repeatability
• Agreement between repeated measures
• Intra-observer variability
  • agreement with your previous interpretation
• Inter-observer variability
  • agreement between observers
Accuracy

• Closeness of measurements to a specific value
• To what extent does the test give the right answer
• Requires a gold standard (definitive assessment)
• Measures of accuracy
  • sensitivity and specificity
  • positive and negative predictive values
  • receiver operating characteristic (ROC) curve
  • likelihood ratio
Precision vs. Accuracy

![Diagram showing targets with labels for low and high accuracy and precision](quora.com)
Sensitivity and Specificity

• **Sensitivity**
  • proportion of positive tests out of total disease
  • correctly identified positives
  • true-positive rate

• **Specificity**
  • proportion of negative tests out of total non-diseased
  • correctly identified negatives
  • true-negative rate
Dichotomous Outcome and Test Result
2x2 Contingency Table

<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Test</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Negative Test</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>
Calculating Sensitivity and Specificity

<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Test</td>
<td>a (TP)</td>
<td>b (FP)</td>
</tr>
<tr>
<td>Negative Test</td>
<td>c (FN)</td>
<td>d (TN)</td>
</tr>
</tbody>
</table>

sensitivity = \frac{a}{(a + c)}

specificity = \frac{b}{(b + d)}
Sensitivity and Specificity - Challenges

• Never consider these two parameters separately
• Trade off between sensitivity and specificity
  • As one increases, the other decreases
    • e.g. higher cutoff leads to increased specificity but decreased sensitivity
• A highly sensitive test is prone to false-positives
  • incorrectly label someone as having the disease
• A highly specific test is prone to false-negatives
  • fail to identify disease
• What is important to you?
• Avoid missing someone or avoid incorrectly labelling someone?

Barrett and Fardy, 2021
# Trade off Between Sensitivity and Specificity

**Table 3.** Test characteristics of the MIAH-CD, MIAH-UC and the combinations of the MIAH and calprotectin home tests compared with endoscopy.

<table>
<thead>
<tr>
<th>Test Characteristics</th>
<th>Sensitivity [95% CI]</th>
<th>Specificity [95% CI]</th>
<th>NPV [95% CI]</th>
<th>PPV [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIAH-CD</td>
<td>67.6 [55.5–78.2]</td>
<td>81.3 [69.5–89.9]</td>
<td>80.0 [70.1–87.2]</td>
<td>69.3 [61.3–76.4]</td>
</tr>
<tr>
<td>MIAH-UC</td>
<td>68.2 [52.4–81.4]</td>
<td>80.5 [70.6–88.2]</td>
<td>83.3 [76.2–88.6]</td>
<td>63.8 [52.4–73.9]</td>
</tr>
<tr>
<td>MIAH-CD score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with QOC</td>
<td>96.7 [82.8–99.9]</td>
<td>66.7 [46.0–83.5]</td>
<td>94.7 [72.1–99.2]</td>
<td>76.3 [65.3–84.7]</td>
</tr>
<tr>
<td>with CS</td>
<td>96.2 [80.4–99.9]</td>
<td>58.3 [36.6–77.9]</td>
<td>93.3 [66.5–99.0]</td>
<td>71.4 [60.8–80.2]</td>
</tr>
<tr>
<td>MIAH-UC score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with QOC</td>
<td>88.2 [63.6–98.5]</td>
<td>71.4 [53.7–85.4]</td>
<td>92.6 [77.0–97.9]</td>
<td>60.0 [46.4–72.3]</td>
</tr>
<tr>
<td>with CS</td>
<td>87.5 [61.7–98.5]</td>
<td>63.3 [43.9–80.1]</td>
<td>90.5 [71.6–97.3]</td>
<td>56.0 [43.4–67.8]</td>
</tr>
</tbody>
</table>

CD, Crohn’s disease; UC, ulcerative colitis; MIAH, Monitor IBD At Home questionnaire; QOC, QuantOn Cal home test; CS, Calpro Smart home test; NPV, negative predictive value; PPV, positive predictive value; CI, confidence interval.
Sensitivity and Specificity - Challenges

• **Affected by severity of disease**
  • results from a CXR for detection of lung cancer will depend on severity of illness and stage of the disease, size of the tumour etc.

• **Sensitivity and specificity describe how well a test performs**

• Don’t convey significance of the test result for an individual patient

Barrett and Fardy, 2021
Likelihood Ratio

• Assesses potential utility of a diagnostic test
• Assesses how likely the patient with a positive test has the disease
• Probability of positive test given disease relative to probability of positive test given no disease
• Answers question: How much more likely is a positive test result in the presence of disease compared with absence of disease?

\[
LR = \frac{\text{sensitivity}}{1 - \text{specificity}}
\]
Likelihood Ratio

• Has predictive value and stable with changes in prevalence
• Ranges from zero to infinity
• The higher the value, the more likely the patient has the condition
• $0 - 1 = \text{decreased evidence for disease}$
• $1 = \text{no diagnostic value}$
• $>1 = \text{increased evidence for disease}$
# Likelihood Ratio Example

<table>
<thead>
<tr>
<th>Serum Ferritin (mg/dL)</th>
<th>LR (of iron deficiency anemia)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;15</td>
<td>51.8</td>
</tr>
<tr>
<td>15-24</td>
<td>8.8</td>
</tr>
<tr>
<td>25-34</td>
<td>2.5</td>
</tr>
<tr>
<td>45-100</td>
<td>0.5</td>
</tr>
<tr>
<td>&gt;100</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Sloane 2008
Prediction

• Predictive values
• Ability of a diagnostic test to make a diagnosis in the future

• **Positive predictive value (PPV)**
  • proportion of diseased with positive test result
  • proportion of people with a positive test who have the disease

• **Negative predictive value (NPV)**
  • proportion of healthy individuals with a negative test result
  • proportion of people with a negative test who are free of disease
Prediction

- A test with a high positive predictive value makes the disease quite likely in a subject with a positive test
- A test with a high negative predictive value makes the disease quite unlikely in a subject with a negative test
<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Test</strong></td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td><strong>Negative Test</strong></td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Test</strong></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td><strong>Negative Test</strong></td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

**PPV** = \( \frac{a}{(a + b)} \)

**NPV** = \( \frac{d}{(c + d)} \)
Predictive values - Challenges

• **Cannot be used in case-control studies**
  - used for random samples or cohorts where observed prevalence is equivalent to true prevalence

• **Affected by prevalence (proportion of subjects with disease)**
  - high prevalence
    - PPV increases and NPV decreases
  - low prevalence
    - PPV decreases, NPV increases

• **Less portable from population to population**
  - due to effect of prevalence
Effect of Prevalence on PPV and NPV

Assessment of a new diagnostic test when prevalence of disease is 50%

<table>
<thead>
<tr>
<th>New test</th>
<th>Gold standard</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>45 (a)</td>
<td>10 (b)</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>5 (c)</td>
<td>40 (d)</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Prevalence of disease = 50/100 = 50%

Sensitivity = a/(a + c) = 45/50 = 90%

Specificity = d/(b + d) = 40/50 = 80%

Positive predictive value = a/(a + b) = 45/55 = 82%

Negative predictive value = d /(c + d) = 40/45 = 88%

Barrett and Fardy, 2021
### Effect of Prevalence on PPV and NPV

**Assessment of a new diagnostic test when prevalence of disease is 10%**

<table>
<thead>
<tr>
<th>New test</th>
<th>Gold standard</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Positive</td>
<td>45 (a)</td>
<td>90 (b)</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>5 (c)</td>
<td>360 (d)</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>50</td>
<td>450</td>
<td>500</td>
</tr>
</tbody>
</table>

Prevalence of disease = 50/500 = 10%

Sensitivity = \( \frac{a}{a + c} = \frac{45}{50} = 90\% \)

Specificity = \( \frac{d}{b + d} = \frac{360}{450} = 80\% \)

Positive predictive value = \( \frac{a}{a + b} = \frac{45}{135} = 33\% \)

Negative predictive value = \( \frac{c}{c + d} = \frac{360}{365} = 98\% \)
<table>
<thead>
<tr>
<th>New test</th>
<th>Gold standard</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>45 (a)</td>
<td>10 (b)</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>5 (c)</td>
<td>40 (d)</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

Prevalence of disease = 50/100 = 50%
Sensitivity = a/(a + c) = 45/50 = 90%
Specificity = d/(b + d) = 40/50 = 80%
Positive predictive value = a/(a + b) = 45/55 = 82%
Negative predictive value = d /(c + d) = 40/45 = 88%

<table>
<thead>
<tr>
<th>New test</th>
<th>Gold standard</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>45 (a)</td>
<td>90 (b)</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>5 (c)</td>
<td>360 (d)</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>50</td>
<td>450</td>
<td>500</td>
</tr>
</tbody>
</table>

Prevalence of disease = 50/500 = 10%
Sensitivity = a/(a + c) = 45/50 = 90%
Specificity = d/(b + d) = 360/450 = 80%
Positive predictive value = a/(a + b) = 45/135 = 33%
Negative predictive value = c /(c + d) = 360/365 = 98%
Prevalence and Diagnostic Tests

• Diagnostic tests function best when prevalence is between 40-60%
• **Function poorly at extremes of prevalence**
• “When you are already pretty sure that the patient either does or does not have the diagnosis in question, additional testing may not alter that probability very much”
  • e.g. ECHO for endocarditis or chest CT for pulmonary embolus

Barrett and Fardy, 2021
Receiver Operating Characteristic (ROC) Curves

- Test result is not simply positive or negative
- Continuous test results
- Potentially multiple cutoffs
- Sensitivity (Y-axis) vs. 1-specificity (X-axis)
- Best cut-off maximizes sensitivity and specificity
  - 1 = perfect test
  - 0.5 = useless test (equivalent to random chance)
- Quantifies information gain for a test
- Provides summary estimate of the accuracy of the test
Examples of ROC Curves

Comparing ROC Curves

- **Worthless**
- **Good**
- **Excellent**

False positive rate

True positive rate
Area Under the ROC Curve (AUC)

- Values between 0.0 and 1.0
  - perfectly inaccurate to perfectly accurate
  - 0.5 = useless test

Chen and Boning 2019
Additional Considerations

• **Cost**
• **Availability**
• **Acceptability**
  • i.e. invasive test with potentially serious complications
• **Clinical utility**
  • ideally assessed using a RCT
    • assess outcomes
    • document adverse events
    • assess impact on decision-making
    • assess patient satisfaction and cost-effectiveness

Barrett and Fardy, 2021
Example

• Evaluation of a new diagnostic test
• Cell-free DNA for the diagnosis of rejection after heart transplantation
• De Vlaminck et al. Sci Transl Med 2014
Acute Cellular Rejection Grades

- 0R: no rejection
- 1R: mild rejection
- 2R: moderate rejection
- 3R: severe rejection

Current (flawed) Gold Standard

• **Endomyocardial biopsy**
  • invasive
  • expensive
  • GA required for children
  • potential for serious complications
  • sampling error
  • subjective interpretation
Non-Invasive Assay for Rejection
Cell-Free DNA

Genome Transplant

Healthy

Acute Cellular Rejection

1. Collect cell-free DNA from plasma

2. Perform shotgun sequencing; identify reads with Donor and Recipient SNP calls to calculate % Donor DNA

3. Monitor Donor DNA level over time to detect onset of rejection

Snyder et al. PNAS 2011
Genome Transplant Dynamics

De Vlaminck et al. Sci Transl Med 2014
Snyder et al. determined by endomyocardial biopsy-determined rejections, our results clearly establish unambiguously that donor-specific signatures can be detected in plasma (14, 15), our data establish of organ rejection. Although the existing cell-free DNA literature has presented con of elevated levels of donor DNA can be used as an indication.

In this study we aimed to demonstrate that donor-derived cell-free DNA exists in the plasma of organ transplant recipients, and we have shown evidence here that G... and May 2005 at our institution. This cohort, funded by the National Institutes of Health (5P01AI050153-02), was assembled prospectively to study the re...

**Materials and Methods**

DNA from the transplanted organ and graft injury. As GTD is not particularly dependent on... the blood, and likely have different sources for false positives/

**Discussion**

In comparing GTD to noninvasive expression analysis tests, such as AlloMap, one observes some similarities and differences. It is not possible donor and recipient pair.

Using single nucleotide polymorphisms that can be used for any organs, here we have also demonstrated a generalizable strategy...

Collectively these results establish that donor-derived DNA in the plasma is a promising biomarker for the onset of, and recovery from, heart transplant rejection. Whereas most earlier studies focused on the limited cases of females receiving male hearts, here we have also demonstrated a generalizable strategy.

As GTD is not particularly dependent on...

...onset of rejection before biopsy. Whereas GTD showed some sensitivity for grade 2 and grade 3R biopsy. Neither GTD nor AlloMap have been shown to dis...

The level of donor DNA signal rises in correlation with the endomyocardial biopsy results, with mean values increasing to 3% of the total.

Having established that the mean value of >4% of the total...
Prospective Study Design

Heart transplant recipients
- n = 21 pediatric; n = 44 adult

Pretransplant
- Collect whole blood samples
- Genotype donor and recipient (65 pairs)

Posttransplant
- Collect serial blood samples (565)
- Purify and sequence cell-free DNA in plasma
- Calculate fraction of donor-derived DNA in plasma
- Compare to biopsy grade

De Vlaminck et al. Sci Transl Med 2014
sensitivity 0.58
specificity 0.87

Biopsy
0.58
0.96
3R vs. 0R
2R/3R vs. 0R
2R/3R vs. 1R
1R vs. 0R

De Vlaminck et al. Sci Transl Med 2014
Other Issues

• Flawed gold standard
  • makes development/evaluation of new diagnostics challenging

• Cost

• Availability
  • requires DNA sequencing infrastructure

• Acceptability
  • peripheral blood test vs. cardiac catheterization

• Clinical utility
  • low pre-test probability of serious rejection
  • confirm positive results with invasive procedure
Summary

• **Multiple metrics to evaluate a diagnostic test**
  • Test performance
    • precision (reproducibility) and accuracy
    • sensitivity or specificity
    • likelihood ratio

• **Positive and negative predictive values**
  • affected by disease prevalence
  • function poorly at the extremes

• **ROC curves**
  • estimate accuracy of the test for different cutoff values
  • summarized with AUC

• **Impact and non-clinical factors**