Statistical Approaches

Patrick M Bossuyt
Key Messages

• Statistical approaches for comparing novel markers exist

• Comparative evaluations differ from single marker evaluations

• Should always consider true positives and false positives

• Statistical approach varies with phase of evaluation
## Phases of Evaluation

<table>
<thead>
<tr>
<th>Phase</th>
<th>Aim</th>
<th>Approach</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Established Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Early Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Screening Evaluation</td>
<td>Estimate performance</td>
<td>Intended use population</td>
</tr>
<tr>
<td>4</td>
<td>Program Evaluation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A novel test / marker
Comparator: FIT
Test Comparisons: Novel versus Existing

**Existing situation**
- Population
  - Initial tests
    - Existing test
      - +
      - -

**Replacement**
- Population
  - Initial tests
    - New test
      - +
      - -

**Triage**
- Population
  - New test
    - Existing test
      - +
      - -
      - +
      - -

**Add-on**
- Population
  - Initial tests
    - Existing test
      - +
      - -
      - +
      - -

Bossuyt et al. BMJ 2006;332;1089-1092
Comparator: FIT performance

Systematic review and bivariate/HSROC random-effect meta-analysis of immunochemical and guaiac-based fecal occult blood tests for colorectal cancer screening

Robert Launois, Jean-Gabriel Le Moine, Bernard Uzzan, Lucia I. Fiestas Navarrete and Robert Benamouzig

Background: Current literature evidences higher accuracy of immunochemical (FOBT) vs. guaiac-based (gFOBT) fecal occult blood tests for colorectal cancer (CRC) screening. Few well-designed head-to-head comparisons exist.

Aim: This meta-analysis assesses the performances of two FOBTs compared with an established gFOBT using colonoscopy as the gold standard.

Methods: We mobilized a bivariate and a hierarchical summary receiver operating characteristic (HSROC) model. Positive likelihood ratio (LR+) and negative likelihood ratio (LR−) and diagnostic odds ratios were back-calculated. We constructed bivariate credibility ellipses in the HSROC space and calculated areas under the curve to obtain a global measure of test performance. Estimates are presented at 95% credibility levels.

Results: We included and analyzed 21 studies. OC-Sensor was the best test for CRC screening, with high sensitivity (80.7%; 95% credibility interval: 74.3-86.0%) and specificity (83.5%; 95% credibility interval: 78.4-88.3%). The optimal LR+ (12.20) and LR− (0.14) and a high diagnostic odds ratio (83.05). Bivariate credibility ellipses showed OC-Sensor’s dominance over Hemoccult (sensitivity: 0.47; 95% credibility interval: 0.37-0.56; specificity: 0.83; 95% credibility interval: 0.91-0.93).

Conclusion: Our findings support the use of OC-Sensor for CRC detection. The diagnostic estimates obtained may be extended to derive model parameters for economic decision models and to offer insight for future clinical and public health decision making. Our findings could influence the future of FOBTs within the CRC screening arsenal.


Keywords: accuracy, advanced adenomas, biotest, colorectal cancer, diagnostic test, HSROC, literature review, meta-analysis, GAVADES

French Network for Evaluation in Health Economics, RESEDAM, Paris and
6 Service d’Hépato-Gastroenterologie, Hôpital Assurance AP’M, Bully, France

Correspondence to Robert Launois, PhD, French Network for Evaluation in Health Economics, 80, Boulevard de L’Alma, 75007 Paris, France

Received 16 April 2014 Accepted 16 June 2014

50 ng/ml cutoff
Advanced Neoplasia
Target Region in ROC space

Diamond: Performance FIT
In detecting advanced neoplasia

Green region: desired performance from new marker
Target Region in ROC space

Novel marker:
Better sensitivity – Better specificity
Target Region in ROC space

True Positive Rate : Sensitivity
False Positive Rate : 1 – Specificity
Acceptability Region in ROC space
More true positives – More false positives
Single Marker Evaluation

- CRC screening invitees
- Novel Marker
- Colonoscopy
- Cross-classification
Single Marker – One-sided hypothesis test

• Define joint hypothesis for novel marker
  
  Sensitivity  >  0.40 and  
  Specificity  >  0.90

• Define combined null hypothesis
  
  Sensitivity  ≤  0.40 and/or  
  Specificity  ≤  0.90
Comparative Marker Evaluation

- CRC screening invitees
- FIT
- Novel Marker
- Colonoscopy
- Cross-classification
Measures Comparative Marker Evaluation

• Novel Marker
  • Sensitivity: $\text{Sens}_{\text{NEW}}$
  • Specificity: $\text{Spec}_{\text{NEW}}$

• FIT
  • Sensitivity: $\text{Sens}_{\text{FIT}}$
  • Specificity: $\text{Spec}_{\text{FIT}}$

• Relative TPR / FPR
  • $r_{\text{TPR}}$: \[ \frac{\text{Sens}_{\text{NEW}}}{\text{Sens}_{\text{FIT}}} \]
  • $r_{\text{FPR}}$: \[ \frac{1 - \text{Spec}_{\text{NEW}}}{1 - \text{Spec}_{\text{FIT}}} \]
### Table 1. Sensitivity and Specificity of the Multitarget Stool DNA Test and the Focal Immunohistochemical Test (FIT) for the Most Advanced Findings on Colonoscopy.

<table>
<thead>
<tr>
<th>Most Advanced Finding</th>
<th>Colonoscopy (N=9989)</th>
<th>Multitarget DNA Test (N=9989)</th>
<th>FIT (N=9989)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive Results</td>
<td>Sensitivity (95% CI)</td>
<td>Positive Results</td>
</tr>
<tr>
<td></td>
<td>n1</td>
<td>%</td>
<td>n2</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any</td>
<td>63</td>
<td>60</td>
<td>93.3 (83.6–94.0)</td>
</tr>
<tr>
<td>Stage I to III</td>
<td>60</td>
<td>56</td>
<td>93.3 (83.6–94.0)</td>
</tr>
<tr>
<td>Colorectal cancer and high-grade dysplasia</td>
<td>104</td>
<td>87</td>
<td>83.7 (75.1–90.2)</td>
</tr>
<tr>
<td>Advanced precancerous lesions†</td>
<td>757</td>
<td>331</td>
<td>42.4 (38.9–46.8)</td>
</tr>
<tr>
<td>Nonadvanced adenoma</td>
<td>2893</td>
<td>498</td>
<td>17.2 (15.5–18.8)</td>
</tr>
<tr>
<td>All nonadvanced adenomas, non-neoplastic findings, and negative results on colonoscopy</td>
<td>9167</td>
<td>1211</td>
<td>86.6 (85.9–87.7)</td>
</tr>
</tbody>
</table>

|                           |      |     |      |     |     |
|                           |      |     |      |     |     |
|                           |      |     |      |     |     |
|                           |      |     |      |     |     |
|                           |      |     |      |     |     |

**TPR**<sub>NEW</sub>: \( \frac{60 + 321}{757 + 65} = 0.46 \)

**TPR**<sub>FIT</sub>: \( \frac{48 + 180}{757 + 65} = 0.28 \)

**rTPR**: \( \frac{60 + 321}{48 + 180} = 1.7 \)

**FPR**<sub>NEW</sub>: \( \frac{1231}{9167} = 0.13 \)

**FPR**<sub>FIT</sub>: \( \frac{472}{9167} = 0.05 \)

**rFPR**: \( \frac{1231}{472} = 2.6 \)
Comparison – Double Superiority

• Define joint hypothesis

\[ 1 < r_{TPR} \quad \text{and} \quad r_{FPR} < 1 \]

• Define combined null hypothesis

\[ r_{TPR} \leq 1 \quad \text{and/or} \quad 1 \leq r_{FPR} \]

• Sample size depends on study design (and discordance)
Superiority in true positives

- **New marker inferior**
- **Inconclusive**
- **Inconclusive**
- **New marker superior**

**rTPR:**
- **sens** \(_{NEW}\)
- **sens** \(_{FIT}\)

**rTPR:**
Non-inferiority for false positives

- New marker superior
- New marker not inferior
- Inconclusive
- New marker inferior

**rFPR**

$$rFPR: \frac{1 - \text{Spec}_{\text{NEW}}}{1 - \text{Spec}_{\text{FIT}}}$$

**Margin**

- New marker better
- rFPR
- New marker worse
Comparison – Non-Inferiority & Superiority

• Define joint hypothesis

\[ 1 < r_{TPR} \text{ and } r_{FPR} < \delta \text{ (margin } 1 < \delta) \]

• Define combined null hypothesis

\[ r_{TPR} \leq 1 \text{ and/or } \delta \leq r_{FPR} \]

• Sample size depends on study design (and discordance)
Caveat

• Always must consider
  • True Positives and False Positives

• Can be
  • Superiority for both
  • Superiority for one; Non-inferiority for other
  • Non-inferiority for both
<table>
<thead>
<tr>
<th>Phase</th>
<th>Aim</th>
<th>Approach</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Established Disease</td>
<td>Study distribution of test results</td>
<td>Compare to other groups No estimates of sensitivity and specificity yet</td>
</tr>
<tr>
<td>2</td>
<td>Early Disease</td>
<td>Study distribution of test results</td>
<td>Compare to more groups No estimates of sensitivity and specificity yet</td>
</tr>
<tr>
<td>3</td>
<td>Screening Evaluation</td>
<td>Estimate / Test Performance</td>
<td>Intended use population</td>
</tr>
<tr>
<td>4</td>
<td>Program Evaluation</td>
<td>Validate Performance Estimate Effectiveness</td>
<td>Intended use population</td>
</tr>
</tbody>
</table>
Key Messages

• Statistical approaches for comparing novel markers exist

• Comparative evaluations differ from single marker evaluations

• Should always consider true positives and false positives

• Statistical approach varies with phase of evaluation
STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies


STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies

OPEN ACCESS

Incomplete reporting has been identified as a major source of avoidable waste in biomedical research. Essential information is often not provided in study reports, impeding the identification, critical appraisal, and replication of studies. To improve the quality of reporting of diagnostic accuracy studies, the Standards for Reporting Diagnostic Accuracy (STARD) statement was developed. Here we present STARD 2015, an updated list of 30 essential items that should be included in every report of a diagnostic accuracy study. This update incorporates recent evidence about sources of bias and variability in diagnostic accuracy and is intended to facilitate the use of STARD. As such, STARD 2015 may help to improve completeness and transparency in reporting of diagnostic accuracy studies.
