**Methods for Algorithm Comparison Assessment**

by

Nancy A. Obuchowski, PhD

Cleveland Clinic Foundation

---

**Outline**

- TAVR study description
- Study objectives
- Analysis when ground truth is known
- Absence of ground truth
  - test for interchangeability

---

**TAVR Study**

- 110 patients undergoing transcatheter aortic valve replacement (TAVR)
- Pre-surgical 3D CT used to match prosthesis size to annulus size
- Radiologist manually measures annulus from CT image = clinical standard
TAVR Study

- Automated analysis – faster, potentially more reproducible than manual

- Two Automated Methods:
  - Fully-automated method
  - Semi-automated (insertion points of measurements reviewed and corrected, as needed)

Study Design for TAVR Study

<table>
<thead>
<tr>
<th></th>
<th>Expert Radiologist</th>
<th>Less experienced Radiologist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual measurements</td>
<td>2x</td>
<td>1x</td>
</tr>
<tr>
<td>Fully-Automated</td>
<td>2x</td>
<td>2x</td>
</tr>
<tr>
<td>Semi-automated</td>
<td>2x</td>
<td>2x</td>
</tr>
</tbody>
</table>

Study Goals:

1. Compare agreement in area measurements of automated methods with manual
   - Bias
   - Precision
   - Summary measure
2. Assess effect of automated methods on patient management:
   - Test for interchangeability
### Bias
(mean difference between Automated method and Expert’s Manual measurement)

<table>
<thead>
<tr>
<th></th>
<th>Reader 1</th>
<th>Reader 2</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully-automated</td>
<td>0.715</td>
<td>0.677</td>
<td>0.696</td>
</tr>
<tr>
<td>Semi-automated</td>
<td>0.158</td>
<td>0.267</td>
<td>0.212</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### Precision

- Intra-reader variability
- Inter-reader variability
Intra-Reader Agreement

Coverage Probability

Agreement Criterion

Manual
Auto with Correction
Auto w/o Correction

Inter-Reader Agreement

Coverage Probability

Agreement Criterion

Manual
Auto with Correction
Auto w/o Correction

Results from Disaggregate Analysis

<table>
<thead>
<tr>
<th></th>
<th>Least Bias</th>
<th>Most Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully Automated</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Semi-Automated</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Summary Measures
(aggregate analysis)

- ICC (intraclass correlation coefficient): its magnitude depends on between-subject variability

- CP (Coverage Probability): simple proportion of cases with agreement as the definition of agreement becomes more lax

Goal #2: Effect of Automated Method on Patient Management
**Three Valve Sizes:**

- 23 valve: Annulus area 3.3-4.0
- 26 valve: Annulus area 4.2-5.0
- 29 valve: Annulus area 5.3-6.3

<3.3 or >6.3 → no replacement performed

---

Can clinicians switch between manual and semi-automated methods without affecting patient management?

**Test for Interchangeability**

<table>
<thead>
<tr>
<th>Method</th>
<th>Expert Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Manual measurements by Expert</td>
<td>0.945</td>
</tr>
<tr>
<td>1 Manual and 1 Semi-Automated by Expert</td>
<td>0.876</td>
</tr>
</tbody>
</table>

---

**Proportion of cases where Expert Reader agrees about valve size**
Proportion of cases where Different Readers agree about valve size

<table>
<thead>
<tr>
<th></th>
<th>Manual measurement by different readers</th>
<th>Manual measurement by one reader and Semi-Automated by another</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Manual measurements</td>
<td>0.594</td>
<td>0.824</td>
</tr>
<tr>
<td>by different readers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Are Manual and Automated methods Interchangeable?

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Expected Outcome of Switching between Manual and Automated:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert Reader</td>
<td>1.7-12.2% increase in frequency of disagreement in valve size</td>
</tr>
<tr>
<td>Different Readers</td>
<td>14.8-31.3% decrease in frequency of disagreement in valve size</td>
</tr>
</tbody>
</table>

Conclusion

- Semi-automated less bias, but also less precise than fully-automated method.
- Semi-automated method agrees with manual more often.
- Semi-automated method can be used interchangeably with manual in typical clinical scenario.