Assessing Biomarker Performance and Sample Size Calculations for Biomarkers

Heather Thiessen Philbrook, MMath, AStat
London Health Sciences Centre
London, Canada
Objectives

- Identify strengths and weaknesses of measures to assess biomarker performance
- Describe considerations in sample size calculations
- How biomarkers influence required sample sizes
Setting

- Adults undergoing cardiac surgery
  - 6 sites in North America
- Biomarkers of acute kidney injury
  - Urine IL-18
- Acute Kidney Injury (Outcome)
  - RIFLE Stage 1 or higher
I have the ability to quantify the unquantifiable.

That is why they call me Dogbert the quantifier.

Who calls you that?

Eight people.
## Measure of Association – Odds Ratio

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Acute Kidney Injury</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>TOTAL</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Odds Ratio = odds of AKI in biomarker positive group relative to odds of AKI in biomarker negative group
### Measure of Association – Odds Ratio

<table>
<thead>
<tr>
<th>Urine IL-18</th>
<th>Acute Kidney Injury</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>TOTAL</td>
</tr>
<tr>
<td>Elevated</td>
<td>29</td>
<td>211</td>
<td>240</td>
</tr>
<tr>
<td>Not elevated</td>
<td>3</td>
<td>237</td>
<td>240</td>
</tr>
<tr>
<td>TOTAL</td>
<td>32</td>
<td>448</td>
<td>480</td>
</tr>
</tbody>
</table>

Odds Ratio = \( \frac{(29/211)}{(3/237)} \)

= 10.9
# Measures of Association – Odds Ratio

<table>
<thead>
<tr>
<th>Urine IL-18</th>
<th>Acute Kidney Injury</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>TOTAL</td>
<td></td>
</tr>
<tr>
<td>Elevated</td>
<td>29</td>
<td>211</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>Not Elevated</td>
<td>3</td>
<td>237</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>32</td>
<td>448</td>
<td>480</td>
<td></td>
</tr>
</tbody>
</table>

Confidence Interval =
Point Estimate ± Confidence Coefficient * Standard Error

SE (In Odds Ratio) = \( \sqrt{\frac{1}{3} + \frac{1}{29} + \frac{1}{237} + \frac{1}{211}} \)

CI for OR = \( \exp(\text{In (Odds Ratio)} ± 1.96 \text{SE(In Odds Ratio)}) \)
Advantages / Disadvantages - Odds Ratio

- **Advantages**
  - Can be directly estimated from logistic regression models
  - Logistic regression techniques are well established, flexible and familiar to clinicians
  - Can compare across studies
  - Provides a measure of association (causal relationship)

- **Disadvantages**
  - Not a measure of diagnostic performance.
  - Assumes a log-linear relationship between biomarker and outcome.
  - Outcome is dichotomous
Measure of Discrimination – ROC Curve

- **True Positive Rate (TPR)**: Also known as Sensitivity
- **False Positive Rate (FPR)**: Also known as 1-Specificity
Measures of Discrimination – TPR, FPR

<table>
<thead>
<tr>
<th>Urine IL-18</th>
<th>Acute Kidney Injury</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>TOTAL</td>
<td></td>
</tr>
<tr>
<td>&gt;60 pg/mL</td>
<td>29</td>
<td>211</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>≤60 pg/mL</td>
<td>27</td>
<td>933</td>
<td>960</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>56</td>
<td>1144</td>
<td>1200</td>
<td></td>
</tr>
</tbody>
</table>

TPR = 29 / 56 = 0.52
FPR = 211 / 1144 = 0.18

Sensitivity 1-Specificity
Measure of Discrimination - ROC Curve

True Positive Rate (TPR) vs. False Positive Rate (FPR)

Sensitivity

False Positive Rate (FPR)
(1-Specificity)

AUC = 0.724
AUC = 0.767
Quantifies discrimination performance
- Rank based method so outliers do not affect ROC curve
- Comparable across biomarkers or risk prediction models with different measurement units

Difficult to quantify ROC curve with a single metric
- AUC is problematic
- Balance of TPR, FPR

ROC curve analysis is well defined for continuous exposures and binary outcomes but not ordinal outcomes or time to event
Relationship between measures of association and discrimination

Odds Ratio 10.9 (3.3, 36)

<table>
<thead>
<tr>
<th>Urine IL-18</th>
<th>Acute Kidney Injury</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Positive</td>
<td>29</td>
</tr>
<tr>
<td>Negative</td>
<td>3</td>
</tr>
</tbody>
</table>

TPR = 0.91  FPR = 0.47

<table>
<thead>
<tr>
<th>Urine IL-18</th>
<th>Acute Kidney Injury</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Positive</td>
<td>26</td>
</tr>
<tr>
<td>Negative</td>
<td>6</td>
</tr>
</tbody>
</table>

TPR = 0.81  FPR = 0.29
Reclassification metrics

- Reclassification percent
- Net Reclassification Indices
- Integrated Discrimination Improvement
Sample Size Considerations
Utility of Biomarkers in Clinical Trials of AKI
1. Biomarkers as entry criteria
   - High risk trial defined by biomarker risk score
2. Biomarkers as surrogate outcomes
   - Acute kidney injury defined by biomarkers
Patients undergoing Cardiac Surgery

- Low biomarkers
  - Measure biomarkers
  - No enrollment
- High biomarkers
  - Randomize
  - Standard of Care
  - Intervention
Patients undergoing Cardiac Surgery N=1200
AKI 5%

Low biomarker IL-18 ≤60 pg/mL N=960
No enrollment

Measure IL-18

High biomarker IL-18 > 60 pg/mL N=240
Screen Positive Rate 20%

Randomize

Standard of Care AKI 10%

Intervention

51% of Possible AKI Events Included
Simple formula for difference in proportions

Sample size in each group (assumes equal sized groups)

\[ n = \frac{2(\bar{p})(1 - \bar{p})(Z_\beta + Z_{\alpha/2})^2}{(p_1 - p_2)^2} \]

Desired Power
For 80% power 0.84

Significance Level
Typically 1.96

Measure of variability

Effect Size
(The difference in proportions)
## Entry criteria examples

<table>
<thead>
<tr>
<th>Possible Enrollment Criteria</th>
<th>N</th>
<th>Screen Positive Rate</th>
<th>% Events included</th>
<th>% AKI Events in Controls</th>
<th>Sample size require to detect a relative risk of at least 0.7 §</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine IL-18 &gt; 60 pg/mL</td>
<td>240</td>
<td>20</td>
<td>51</td>
<td>10.4</td>
<td>3125</td>
</tr>
<tr>
<td>Urine IL-18 &gt; 60 &amp; CPB &gt; 120 min</td>
<td>148</td>
<td>12</td>
<td>46</td>
<td>12.2</td>
<td>2661</td>
</tr>
</tbody>
</table>
Trade off Between Screen Positive Rate and Event Rate

- Screen Positive Rate
- Control Event Rate

- Plasma NGAL > 200
- CPB > 120
- Plasma NGAL > 200 & CPB > 120
- Urine IL-18 > 60
- Urine IL-18 > 60 & CPB > 120
- Delta SCr > 0.17 mg/dL
Biomarkers as Surrogates
Biomarkers as Surrogates

- **Randomize**
- **Standard of Care**
- **Intervention**

**Cardiac Surgery**

**Day 1**
- Measure Biomarkers 0-6 Hours
  - AKI > 0.3mg/dL

**Day 2**
- AKI > 50%

**Day 3**
- AKI > 100% or dialysis
### Biomarkers as Surrogates Example

**Randomize**

**Controls:**
- CPB Time > 120 min
- N=153

**Treatment:**
- CPB Time < 80 min
- N=153

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Controls (N=153)</th>
<th>Treatment (N=153)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGAL &gt; 102 ng/mL</td>
<td>33%</td>
<td>8%</td>
</tr>
<tr>
<td>AKI &gt; 50%</td>
<td>26%</td>
<td>12%</td>
</tr>
<tr>
<td>AKI &gt; 100% or dialysis</td>
<td>7%</td>
<td>3%</td>
</tr>
</tbody>
</table>

RR = 0.24 (N=95)  RR = 0.46 (N=273)  RR = 0.43 (N=1030)
Conclusion

- Use performance metrics related to research question
- Potential for biomarkers used in clinical trials
  - Trade off between screening and event rates
  - Biomarkers as surrogate outcomes may reduce sample sizes