Meta-Analysis
Methods used in Radiology Journals

Trevor McGrath
Medical Student
University of Ottawa

April 15, 2016
Disclosure Statement

• This project was funded by:
  
  – University of Ottawa Department of Radiology Research Stipend
  
  – University of Ottawa Faculty of Medicine Summer Studentship Program
Co-Investigators

• Matthew McInnes MD FRCPC
  – University of Ottawa
  – OHRI

• Daniel Korevaar MD
  – Academic Medical Center Amsterdam

• Patrick Bossuyt PhD
  – Academic Medical Center Amsterdam
Conclusions

• Recommended (hierarchical) methods for diagnostic accuracy meta-analysis in imaging journals are used in a minority of reviews
  – This trend has not changed over time

• Traditional (univariate) methods tend to overestimate diagnostic accuracy and provide narrower confidence intervals than recommended (hierarchical) methods for both sensitivity and specificity to a statistically significant degree
Meta-Analysis

• Statistical technique to combine results of two or more studies

• Meta-analysis of diagnostic test accuracy (DTA) studies provides:
  
  – Summary estimates of diagnostic performance
  
  – Uncertainty of these summary estimates (confidence intervals)
Diagnostic Test Accuracy Measures

- Specificity = 69%
- Sensitivity = 99%

Test measurement:
- TN (true negative)
- FN (false negative)
- FP (false positive)
- TP (true positive)

Diagnostic threshold
DTA Meta-Analysis Challenges

• Meta analysis of intervention typically only summarizes one measure (effect size)

• Diagnostic accuracy reviews have two summary statistics
  – Sensitivity and Specificity

• Pooling these independently does not account for the correlation between these two measures

• ‘Traditional’ random effects models are not appropriate for most diagnostic accuracy reviews
Hierarchical Models

- Hierarchical models are recommended for diagnostic accuracy meta-analysis

- Bivariate random effects model
  - Jointly models both sensitivity and specificity (logit of each) and estimates the correlation between them
Purpose

• Determine whether systematic reviews of diagnostic accuracy studies published in imaging journals are using recommended methods for meta-analysis

• Evaluate the impact of traditional methods versus hierarchical methods on summary estimates of sensitivity and specificity.
Methods

- Medline was searched for articles containing systematic reviews that included meta-analysis of test accuracy data, limited to imaging journals from 2005-2015

- One author (TM) reviewed retrieved titles and abstracts and if potentially relevant, assessed their full texts

- Two reviewers (TM, MM) independently extracted study data
  - First author, journal, year of publication, country, continent, subspecialty area, review title and method for meta-analysis
Methods

• Methods for meta-analysis were classified as:
  – traditional (univariate fixed or random effects pooling; summary ROC curve)
  – recommended (bivariate model; hierarchical summary ROC curve)

• Reviews performed meta-analysis with a traditional random effects model and provided 2x2 data were re-analyzed using a bivariate random effects model

• The analyses were performed using the ‘mada’ package in R
Results

- Overall, 118/300 (39%) reviews used recommended methods for meta-analysis
  - 109/300 reviews (36%) used a hierarchical model for meta-analysis
  - 201/300 reviews (67%) used a traditional model for meta-analysis
Results
# Impact of Univariate Methods

Summary statistics for bivariate random effects method recalculations versus results of published traditional univariate random effects pooling estimates

<table>
<thead>
<tr>
<th></th>
<th>Avg. Published Univariate</th>
<th>Avg. Calculated Bivariate</th>
<th>Avg. ( \Delta ) (Calc. – Pub.)</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower C.I. limit</td>
<td>78.4 (74.9, 81.8)</td>
<td>71.8 (68.1, 75.5)</td>
<td>-6.5 (-7.8, -5.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Summary estimate</td>
<td>83.3 (80.2, 86.4)</td>
<td>81.9 (79.2, 84.7)</td>
<td>-1.4 (-2.2, -0.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Upper C.I. limit</td>
<td>87.5 (84.7, 90.4)</td>
<td>88.7 (86.7, 90.7)</td>
<td>1.1 (-0.2, 2.5)</td>
<td>0.78</td>
</tr>
<tr>
<td>C.I. Range</td>
<td>9.2 (8.5, 9.8)</td>
<td>16.9 (15.1, 18.6)</td>
<td>7.7 (5.8, 9.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower C.I. limit</td>
<td>82.4 (79.3, 85.5)</td>
<td>73.5 (69.1, 77.8)</td>
<td>-8.9 (-11.3, -6.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Summary estimate</td>
<td>87.0 (84.3, 89.7)</td>
<td>84.5 (81.9, 87.2)</td>
<td>-2.5 (-3.4, -1.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Upper C.I. limit</td>
<td>90.3 (88.0, 92.6)</td>
<td>91.3 (89.6, 93.0)</td>
<td>1.0 (-0.2, 2.1)</td>
<td>0.98</td>
</tr>
<tr>
<td>C.I. Range</td>
<td>8.0 (7.2, 8.8)</td>
<td>18.0 (15.3, 20.7)</td>
<td>9.9 (7.0, 12.7)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Re-analyses were performed for 33 reviews, together performing 51 meta-analyses.

*p values were determined by logit transforming the univariate and bivariate values for each study. Differences between these logit transformed values for each study were compared to zero using a One-Tailed T-Test. Studies reporting values of 100% had a -0.5% continuity correction applied.
## Impact of Univariate Methods

Frequency of calculated deviations from literature values

<table>
<thead>
<tr>
<th></th>
<th>Datasets with Calculated Deviations of 0-4% n (%)</th>
<th>Datasets with Calculated Deviations of 5-10% n (%)</th>
<th>Datasets with Calculated Deviations of 11-20% n (%)</th>
<th>Datasets with Calculated Deviations ≥ 20 n (%)</th>
<th>Largest Calculated Deviation %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower C.I.</td>
<td>19 (37)</td>
<td>20 (39)</td>
<td>11 (22)</td>
<td>1 (2)</td>
<td>22</td>
</tr>
<tr>
<td>Summary estimate</td>
<td>43 (84)</td>
<td>7 (14)</td>
<td>1 (2)</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Upper C.I.</td>
<td>41 (80)</td>
<td>6 (12)</td>
<td>4 (8)</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>C.I. Range</td>
<td>20 (39)</td>
<td>14 (27)</td>
<td>13 (26)</td>
<td>4 (8)</td>
<td>33</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower C.I.</td>
<td>14 (27)</td>
<td>21 (41)</td>
<td>10 (20)</td>
<td>6 (12)</td>
<td>39</td>
</tr>
<tr>
<td>Summary estimate</td>
<td>39 (76)</td>
<td>11 (22)</td>
<td>1 (2)</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Upper C.I.</td>
<td>44 (86)</td>
<td>5 (10)</td>
<td>2 (4)</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>C.I. Range</td>
<td>19 (37)</td>
<td>18 (35)</td>
<td>5 (10)</td>
<td>9 (18)</td>
<td>47</td>
</tr>
</tbody>
</table>

Re-analyses were performed for 33 reviews, together performing 51 meta-analyses.
Role of fluorodeoxyglucose-PET versus fluorodeoxyglucose-PET/computed tomography in detection of unknown primary tumor: a meta-analysis of the literature
Meng-jie Dong, Kui Zhao, Xiang-ting Lin, Jun Zhao, Ling-xiang Ruan, and Zhen-feng Liu

Reported univariate sensitivity: 0.78 (0.74, 0.84)
Calculated bivariate sensitivity: 0.70 (0.65, 0.76)

Diagnostic Performance of Dual-time $^{18}$F-FDG PET in the Diagnosis of Pulmonary Nodules: A Meta-analysis

Richard L. Barger, Jr., MD, Kiran R. Nandalur, MD

Reported univariate specificity: 0.77 (0.72, 0.81)
Calculated bivariate specificity: 0.65 (0.46, 0.80)
Discussion

• Traditional univariate pooling methods continue to dominate meta-analyses published in imaging journals.

• 2012 release of the freeware ‘mada’ in R
  – Technical and economic barriers to hierarchical methods of data pooling should be considerably reduced for research groups.

• 120 reviews that used traditional univariate pooling methods for meta-analysis used the freeware package ‘Meta-DiSc’, representing nearly two thirds of reviews in this category.

• Cochrane Handbook for diagnostic test accuracy reviews.
Conclusions / Implications for Practice

• The use of traditional meta-analysis methods in imaging journals is associated with higher summary estimates and more narrow confidence intervals for sensitivity and specificity than those derived using recommended methods.

• If these erroneous estimates are applied in clinical practice or used to inform policy, they could falsely estimate diagnostic performance which could negatively impact patient care; leading to missed diagnoses or superfluous further testing.
Thank You

Questions/comments: tmcgr043@uottawa.ca