Harmonisation of standards for the reporting of diagnostic accuracy studies in dementia

Leon Flicker, Craig Ritchie, Anna Noel-Storr, Rupert McShane
Background

- Why is it important for science as well as the art in diagnosis?
- Medical Practitioners and enthusiasts get it wrong
- E.g. Phrenology - sizes of brain areas are meaningful and can be inferred by examining the skull of an individual
Background
How clinically useful are tests?
Science 1994; Nov 11;266:1051.

A Potential Noninvasive Neurobiological Test for Alzheimer’s Disease

Leonard F. M. Scinto, Kirk R. Daffner, David Dressler, Bernard I. Ransil, Dorene Rentz, Sandra Weintraub, Marsel Mesulam, Huntington Potter

amide, placed in their eyes. It was possible to distinguish 18 of 19 individuals (95%) either clinically diagnosed with Alzheimer’s disease or classified as suspect Alzheimer’s individuals by neuropsychological screening from 30 of 32 normal elderly controls (94%).
Background

• Reporting standards drive better methodology

• Diagnostic tests are validated against a gold standard

• Claims about accuracy OFTEN specify a binary relationship
  – ie sensitivity and specificity
Why ‘conversion’?

- Gold or Reference standards
  - Cross-sectional vs NIA-AA ‘Alzheimer’s dementia’
  - Pathology
  - Longitudinal
    - (Decline in neuropsych test eg memory – slope)
    - Binary conversion from MCI to dementia
      » FDA specified for Dx test
      » Outcome measure for trials
Systematic review of biomarkers

• Weight of evidence
  – total numbers converting

• Quality of evidence
  – Quality of methodology
  – Quality of reporting
Methods (1)

• Stage 1 – Sensitive MEDLINE search from 2000 to 2010
  – 19,104 published abstracts/references

• Stage 2 - Abstract review
  – Inclusion criteria
    • Biomarker of interest (abeta, tau, PET, or structural MRI)
    • Longitudinal design

  – Team of 9 medical students
    • 2,000 references each
    • Standard sample of 100: kappa=0.62
    • Overlapping pairs of 100: kappa=0.62-0.75
    • 95% agreement
Results: search

MEDLINE (Ovid SP)
19104

1032
Cross-sectional

17572
Not relevant
(background/animal/review)

500
Longitudinal

77
abeta

64
tau

44
PET

124
MRI

202 references to studies for inclusion

142 primary papers

Inter-assessor agreement
Kappa: 0.62-0.7
# Numbers converting

## Table

<table>
<thead>
<tr>
<th></th>
<th>Ab</th>
<th>Tau</th>
<th>MRI</th>
<th>PET (FDG)</th>
<th>PET (PiB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCI at Baseline</td>
<td>2883</td>
<td>2527</td>
<td>4722</td>
<td>768</td>
<td>187</td>
</tr>
<tr>
<td>Dementia at Follow Up</td>
<td>1242</td>
<td>1069</td>
<td>1477</td>
<td>284</td>
<td>82</td>
</tr>
</tbody>
</table>
Results: STARD

- **CONSORT**: consolidated standards for reporting trials
- **QUORUM**: quality of reporting of meta-analyses
- **MOOSE**: meta-analysis of observational studies in epidemiology
- **STARD**: standards for reporting of diagnostic accuracy
STARD

Item 5: Participant sampling

Item 11: Blinding

Item 15: Characteristics of study population

Item 22: Handling of indeterminate/missing results

R %

Item 17: Time interval
Item 5: Participant sampling

Describe participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.
Item 5: Participant sampling

All tests

- Reported: 54%
- Partially reported: 11%
- Not reported: 35%
Item 11: Blinding

All tests

- Biomarker assessment made blind to conversion status: 32%
- Judgement about 'conversion' made blind to biomarker result: 23%
Item 15: Study population

• Duration of **MCI prior** to application of the biomarker test?
  – MCI for 10 years – less likely to convert?
  – ‘recent’ MCI diagnosis?

• **131 papers (92 %)** of papers did not report duration of MCI prior to test application

• 8% did report it, but very varyingly
Item 17: Time between tests

Report time interval from the index tests to the reference standard, and any treatment administered between
Item 17: Time between tests

% R

% studies reporting

<table>
<thead>
<tr>
<th></th>
<th>% R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abeta</td>
<td>51</td>
</tr>
<tr>
<td>Tau</td>
<td>42</td>
</tr>
<tr>
<td>sMRI</td>
<td>66</td>
</tr>
<tr>
<td>PET-FDG</td>
<td>73</td>
</tr>
<tr>
<td>PET-PiB</td>
<td>56</td>
</tr>
</tbody>
</table>
Item 17: Time between tests

- Mean AND sd: 53%
- Mean NO sd: 27%
- Range: 13%
- Median: 5%
- Other/not reported: 2%
Item 22: Indeterminate results

Report how indeterminate results, missing responses and outliers of the biomarker tests were handled
Item 22: Indeterminate results

- Reported: 29%
- Partially reported: 11%
- Not reported: 70%
Item 19 – reporting of binary data

<table>
<thead>
<tr>
<th></th>
<th>Ab</th>
<th>Tau</th>
<th>MRI</th>
<th>PET (FDG)</th>
<th>PET (PiB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversion</td>
<td>37</td>
<td>33</td>
<td>70</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>Conversion AND Sens and Spec</td>
<td>20</td>
<td>24</td>
<td>32</td>
<td>17</td>
<td>4</td>
</tr>
</tbody>
</table>
STARD-dem

- Dementia-specific
- Rendition of STARD reporting guideline
- Applies only and always when figures for sensitivity or specificity are reported
- Consensus
  - Researchers
  - Methodologists
  - Journal editors
STAGE ONE
Adjustment of original STARD items

• Iterative Process
  – Evaluation of biomarker papers
  – Adjustment of item content

• Criteria
  – Maximum inter-rater reliability of judgements about whether a reporting standard was met
  – Face validity
STAGE TWO

• Consensus generation
  – Commentary on draft STARD-dem
    • Web-site: www.starddem.org
    • 60 days
    • Open to all
    • Can be anonymous
  – Revisions in light of comments
  – Final consensus meeting at CTAD
  – Publication