Structure of a DTA Review
Diagnostic Test Accuracy Reviews

- Question formulation
- Identification of studies
- Selection of relevant studies
- Quality assessment
- Data-extraction
- Data analysis
- Presentation and interpretation of results
Laboratory tests for caffeine addiction in PhD students

Review information

Authors
John Brown¹, Peter Smith¹, Sarah Jones¹

¹[Empty affiliation]

Citation example: Brown J, Smith P, Jones S. Laboratory tests for caffeine addiction in PhD students. Cochrane Database of Systematic Reviews [Year], Issue [Issue].

Add Author

Contact person
[Empty name]

Dates
Assessed as Up-to-date: 
Date of Search: 
Next Stage Expected: 
Protocol First Published: Not specified
Review First Published: Not specified
Last Citation Issue: Not specified
Intervention vs. Diagnostic
Intervention vs. Diagnostic
Titles intervention reviews
Titles DTA Reviews
Clinical assessment for diagnosing congenital heart disease in new born infants with Down syndrome
Exercise: the right title format?

- Diagnostic accuracy of image-guided biopsy in indeterminate renal masses.
- Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain.
- Cell-free fetal DNA for non-invasive prenatal diagnosis of sex.
Text of Review - Background

- Description of the condition
- Description of the intervention
- How the intervention might work
- Why is it important to do this review

Intervention review
- Title
- Review information
- Main text
  - Abstract
  - Plain language summary
- Background
  - Description of the condition
  - Description of the intervention
  - How the intervention might work
  - Why is it important to do this review

Diagnostic test accuracy review
- Title
- Review information
- Main text
  - Abstract
  - Plain language summary
- Background
  - Target condition being diagnosed
  - Index test(s)
  - Alternative test(s)
  - Rationale

- Objectives
- Methods
- Results
- Discussion
- Authors' conclusions
  - Acknowledgements
  - Contributions of authors
  - Declarations of interest
  - Differences between protocol and review
  - Published notes
- Tables
- Studies and references
  - Data and analyses
  - Figures
  - Sources of support
  - Feedback
  - Appendices
Background

- Target condition being diagnosed
  - A description of the target condition of interest (frequency, severity, prognosis and possible treatments).

- Index test
  - A description of the index tests that are being evaluated in this review, including current use and intended roles.

- Alternative test
  - Possible tests and strategies that are used in clinical practice, irrespective of whether they are evaluated in this review.

- Rationale
Text of Review

- Abstract
- Plain language summary
- Background
- Target condition being diagnosed
- Index test(s)
- Alternative test(s)
- Rationale
- Objectives
  - Secondary objectives
  - Investigation of sources of heterogeneity
- Methods
  - Criteria for considering studies for this review
  - Types of studies
Objectives (example)

Clinical assessment for diagnosing congenital heart disease in newborn infants with Down syndrome
- Protocol for Cochrane Review of Diagnostic Test Accuracy -

William McGuire, Peter W Fowlie, and Johannes B Reitsma

OBJECTIVES
Primary objectives
The objective of this review is to assess the accuracy of clinical assessment in diagnosing congenital heart disease in newborn infants with Down syndrome and specifically to determine how well the absence of abnormal findings on clinical assessment rules out a diagnosis in newborn infants with Down syndrome.
Objective of a DTA SR can be

- To make comparisons between tests concerning their global accuracy
- To estimate the accuracy of a test operating at a particular threshold
- To understand why results of studies vary
Components of a question

- For intervention reviews
  - Patients
  - Intervention
  - (Comparative intervention)
  - Outcome
Components of a question

- For diagnostic test accuracy reviews
  - Patients
  - Index test
  - (Comparator test)
  - Target disorder
Components of a question

- For diagnostic test accuracy reviews
  - Patients
  - Presentation
  - Prior tests
  - Index test
  - (Comparator test)
  - Purpose
  - Target disorder
  - Reference standard
Index and comparator tests

- The **index test** is the “new” test we wish to evaluate. A review may consider and compare several index tests.

- The **comparator test** is the alternative diagnostic management strategy which is standard practice and with which we would like to make comparisons.
Target condition and reference standard

- Target disease is the condition we are trying to diagnose

- Reference standard is the best way available of identifying target condition
  - May comprise several pieces of information
  - May only be available subsequently
Methods

Criteria for considering studies for this review
- Types of studies
- Participants
- Index tests
- Comparator tests
- Target conditions
- Reference standards

Search methods for identification of studies
- Electronic searches
- Searching other resources

Data collection and analysis
- Selection of studies
- Data extraction and management
- Assessment of methodological quality
- Statistical analysis and data synthesis
- Investigations of heterogeneity
- Sensitivity analyses
- Assessment of reporting bias

No RCTs!
Intervention Reviews vs DTA Reviews

- Methods
  - Criteria for considering studies for this review
  - Search methods for identification of studies
  - Data collection and analysis
    - Selection of studies
    - Data extraction and management
    - Assessment of risk of bias in included studies
    - Measures of treatment effect
    - Unit of analysis issues
    - Dealing with missing data
      - Assessment of heterogeneity
      - Assessment of reporting biases
    - Data synthesis
    - Subgroup analysis and investigation of heterogeneity
    - Sensitivity analysis
Identification of studies

Problems in indexing of DTA studies

- No study design terms (MeSH: sensitivity-and-specificity)
- Diagnostic search filters based on terms used to report results
- Filters don’t work (loss of relevant articles and not reducing NNR)

Search Strategy: include elements for

- target condition AND index test
- (more titles to screen)

Database of DTA studies is being developed in Sydney
Diagnostic Test Accuracy Reviews

- Question formulation
- Identification of studies
- Selection of relevant studies
- Quality assessment (later)
- Data-extraction
- Data analysis (later)
- Presentation and interpretation of results
Tables
Tables

- Characteristics of studies
  - Characteristics of included studies
  - Characteristics of excluded studies
  - Characteristics of studies awaiting classification
  - Characteristics of ongoing studies
- Summary of results tables
- Additional tables
Characteristics of studies

Characteristics of included studies

Dorlas 2006

<table>
<thead>
<tr>
<th>Clinical features and settings</th>
<th>Primary care; sleep deprivation, loss of concentration, hyperactivity.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>405 Dutch PhD students; no further information provided</td>
</tr>
<tr>
<td>Study design</td>
<td>Cross-sectional design; single group with equal suspicion of caffeine addiction; consecutive enrollments; every student underwent all tests.</td>
</tr>
<tr>
<td>Target condition and reference standard(s)</td>
<td>Caffeine addiction, measured by coffee consumption in liters per day</td>
</tr>
<tr>
<td>Index and comparator tests</td>
<td>Index test is spectrometry; comparator is questionnaire.</td>
</tr>
<tr>
<td>Follow-up</td>
<td>Two weeks follow up</td>
</tr>
<tr>
<td>Notes</td>
<td>Three students withdrew; reasons not explained.</td>
</tr>
</tbody>
</table>

Assessment of methodological quality table

<table>
<thead>
<tr>
<th>Item</th>
<th>Judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum?</td>
<td>Unclear</td>
<td>No information provided</td>
</tr>
<tr>
<td>Acceptable reference standard?</td>
<td>Yes</td>
<td>Coffee consumption</td>
</tr>
<tr>
<td>Acceptable delay between tests?</td>
<td>Yes</td>
<td>During two weeks, coffee consumption is daily measured; after one week</td>
</tr>
</tbody>
</table>
### Assessment of methodological quality table

<table>
<thead>
<tr>
<th>Item</th>
<th>Judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum?</td>
<td>Unclear</td>
<td>No information provided</td>
</tr>
<tr>
<td>Acceptable reference standard?</td>
<td>Yes</td>
<td>Coffee consumption</td>
</tr>
<tr>
<td>Acceptable delay between tests?</td>
<td>Yes</td>
<td>During two weeks, coffee consumption is daily measured; after one week and after two weeks, serum is measured; after two weeks questionnaire.</td>
</tr>
<tr>
<td>Partial verification avoided?</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>Differential verification avoided?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Incorporation avoided?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Reference standard results blinded?</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>Index test results blinded?</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>Relevant clinical information?</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Uninterpretable results reported?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Withdrawals explained?</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

---

### Egberts 2001

Clinical features and settings
Quality assessment: diagnosis

- Relation between quality items and bias is not as straightforward as it is for interventions

- Many more items: 11 mandatory and >10 facultative (QUADAS)

- Methodological Quality Table
  - “Risk of Bias Table”
Assessment of methodological quality
Methodological quality summary: Review authors' judgments about each methodological quality item for each included study.
Data and analyses
Data and Analysis

In diagnostics, less straightforward than RCT:

- Additional challenges are:
  - Pairs of sensitivity and specificity
  - Cut-off problems
  - Heterogeneity is rule rather than exception

NB: Apart from sources of heterogeneity known from the therapeutic field, we have to deal with the many different research designs used
## Data and analyses

### Data tables by test

<table>
<thead>
<tr>
<th>Test</th>
<th>Studies</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectrometry</td>
<td>5</td>
<td>1618</td>
</tr>
<tr>
<td>Short Questionnaire</td>
<td>5</td>
<td>1141</td>
</tr>
</tbody>
</table>

### Figures

### Sources of support

#### Internal sources
- No sources of support provided

#### External sources
- No sources of support provided

### Feedback
Laboratory tests for caffeine addiction in PhD students

Test: 1 Spectrometry

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egberts 2001</td>
<td>183</td>
<td>15</td>
<td>56</td>
<td>67</td>
<td>0.74 [0.68, 0.79]</td>
<td>0.82 [0.72, 0.89]</td>
</tr>
<tr>
<td>Dunnink 2003</td>
<td>22</td>
<td>6</td>
<td>1</td>
<td>9</td>
<td>0.96 [0.79, 1.00]</td>
<td>0.98 [0.87, 0.99]</td>
</tr>
<tr>
<td>Hanell and 2005</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Karris 2007</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Van Nelle 1999</td>
<td>18</td>
<td>8</td>
<td>32</td>
<td>2</td>
<td>0.38 [0.23, 0.51]</td>
<td>0.90 [0.85, 0.98]</td>
</tr>
<tr>
<td>Study</td>
<td>TP</td>
<td>FP</td>
<td>FN</td>
<td>TN</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Allan 2005</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>123</td>
<td>0.00 [0.00, 0.95]</td>
<td>0.99 [0.96, 1.00]</td>
</tr>
<tr>
<td>Becker 2003</td>
<td>6</td>
<td>12</td>
<td>7</td>
<td>62</td>
<td>0.46 [0.19, 0.75]</td>
<td>0.84 [0.73, 0.91]</td>
</tr>
<tr>
<td>Bretagne 1998</td>
<td>14</td>
<td>5</td>
<td>4</td>
<td>18</td>
<td>0.78 [0.52, 0.94]</td>
<td>0.78 [0.56, 0.93]</td>
</tr>
<tr>
<td>Challier 2004</td>
<td>20</td>
<td>9</td>
<td>6</td>
<td>35</td>
<td>0.77 [0.56, 0.91]</td>
<td>0.80 [0.65, 0.90]</td>
</tr>
<tr>
<td>Kawazu 2004</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>134</td>
<td>0.64 [0.31, 0.89]</td>
<td>0.97 [0.93, 0.99]</td>
</tr>
<tr>
<td>Maertens 2002</td>
<td>11</td>
<td>7</td>
<td>2</td>
<td>80</td>
<td>0.85 [0.55, 0.98]</td>
<td>0.92 [0.84, 0.97]</td>
</tr>
<tr>
<td>Marr 2004</td>
<td>13</td>
<td>11</td>
<td>11</td>
<td>32</td>
<td>0.54 [0.33, 0.74]</td>
<td>0.74 [0.59, 0.86]</td>
</tr>
<tr>
<td>Pereira 2005</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>29</td>
<td>1.00 [0.05, 1.00]</td>
<td>0.76 [0.60, 0.89]</td>
</tr>
<tr>
<td>Pinel 2003</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>756</td>
<td>0.50 [0.32, 0.68]</td>
<td>0.98 [0.97, 0.99]</td>
</tr>
<tr>
<td>Ulusakarya 2000</td>
<td>16</td>
<td>11</td>
<td>0</td>
<td>108</td>
<td>1.00 [0.83, 1.00]</td>
<td>0.91 [0.84, 0.95]</td>
</tr>
</tbody>
</table>
ROC Plot
Parameters from external software

### Externally Calculated Parameters

#### HSROC model parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambda</td>
<td></td>
</tr>
<tr>
<td>Thota</td>
<td></td>
</tr>
<tr>
<td>beta</td>
<td></td>
</tr>
<tr>
<td>Var(accuracy)</td>
<td></td>
</tr>
<tr>
<td>Var(threshold)</td>
<td></td>
</tr>
</tbody>
</table>

#### Bivariate model parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\log SE)$</td>
<td>1.1005</td>
</tr>
<tr>
<td>$E(\log Sp)$</td>
<td>2.2945</td>
</tr>
<tr>
<td>$Var(\log Se)$</td>
<td>0.3851</td>
</tr>
<tr>
<td>$Var(\log Sp)$</td>
<td>0.0203</td>
</tr>
<tr>
<td>$Cov(\log Se)$</td>
<td>0.2579</td>
</tr>
</tbody>
</table>

#### Confidence and prediction regions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SE(E(\log Se))$</td>
<td>0.3756</td>
</tr>
<tr>
<td>$SE(E(\log Sp))$</td>
<td>0.3133</td>
</tr>
<tr>
<td>Cov(ES)</td>
<td>0.02231</td>
</tr>
<tr>
<td>Studies</td>
<td>12</td>
</tr>
</tbody>
</table>

- Display summary curve
- Display summary point
- Display 96% confidence region
- Display 95% prediction region
When all the results are analyzed...

... then we can build the Summary of Results Table

- What are the results of our search, inclusion and quality assessment?
- What are the results of our meta-analysis?
- How sure are we about these results?
SoF versus SoR
<table>
<thead>
<tr>
<th>Intervention reviews</th>
<th>DTA reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Format for SoF table is clear and predefined</td>
<td>- There is no predefined format</td>
</tr>
<tr>
<td>- You can use GRADE software to build SoF table</td>
<td>- Authors are explicitly invited to be creative</td>
</tr>
<tr>
<td></td>
<td>- There is no existing software to help you</td>
</tr>
</tbody>
</table>
Summary

- No Plain Language Summary
- Methodological Quality (not RoB)
- Summary of Results (not SoF)
- Differences in background
- Differences in methods: see upcoming presentations