Meta-analysis of diagnostic research

Karen R Steingart, MD, MPH
karenst@uw.edu
Chennai, 15 December 2010

Overview

• Describe key steps in a systematic review/meta-analysis of diagnostic test accuracy studies

• Describe standard methods of meta-analysis of data from diagnostic studies

• Identify key references and tools for performing meta-analysis of diagnostic studies
Definitions

- **Systematic review**: A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review.

- **Meta-analysis**: The use of statistical techniques in a systematic review to integrate the results of included studies.

*Q: Can you do a systematic review without doing a meta-analysis? Can you do a meta-analysis without doing a systematic review?*
Challenges with meta-analysis of diagnostic studies

• Diagnostic accuracy cannot adequately be summarized by one measure

• Considerable between-study heterogeneity is the rule and models of meta-analysis must account for this

An individual study of the diagnostic accuracy of a test...

...estimates the ability of the test to distinguish between those with disease (condition) and those without disease
...compares results of the index test with best available reference for classifying patients as having/not having disease

• Most studies report pairs of sensitivity and specificity
A systematic review/meta-analysis of data from diagnostic studies...

...appraises the quality of primary studies
....synthesizes the information
...looks for and investigates possible reasons for inconsistency in results (heterogeneity)
……calculates an overall summary;* considers both dimensions of test performance
…stimulates new research questions

*Meta-analyses (pooling) can increase the precision of the overall result

---

The 2 x 2 Table

<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test +</strong></td>
<td>True Positives</td>
<td>False Positives</td>
<td>TP + FP</td>
</tr>
<tr>
<td></td>
<td>(TP)</td>
<td>(FP)</td>
<td></td>
</tr>
<tr>
<td><strong>Test -</strong></td>
<td>False Negatives</td>
<td>True Negatives</td>
<td>FN + TN</td>
</tr>
<tr>
<td></td>
<td>(FN)</td>
<td>(TN)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>TP + FN</td>
<td>TN + FP</td>
<td>TP + FP + FN + TN</td>
</tr>
</tbody>
</table>
### Measures of test performance

<table>
<thead>
<tr>
<th></th>
<th>Disease Present</th>
<th>Disease Absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test +</td>
<td>TP</td>
<td>FP</td>
<td>TP + FP</td>
</tr>
<tr>
<td>Test -</td>
<td>FN</td>
<td>TN</td>
<td>FN + TN</td>
</tr>
<tr>
<td>Total</td>
<td>TP + FN</td>
<td>TN + FP</td>
<td>TP + FP + FN + TN</td>
</tr>
</tbody>
</table>

- Sensitivity = $\frac{TP}{TP+FN}$
- Specificity = $\frac{TN}{FP + TN}$
- Positive predictive value = $\frac{TP}{TP + FP}$
- Negative predictive value = $\frac{TN}{FN + TN}$
- Likelihood ratio positive = $\frac{Sensitivity}{1 – Specificity}$
- Likelihood ratio negative = $\frac{1 – Sensitivity}{Specificity}$
- Prevalence (proportion of people with disease in population to whom the test has been applied) = $\frac{TP + FN}{TP + FP + FN + TN}$

### Key steps in a systematic review of diagnostic test accuracy

1. Definition of the objectives of the review
2. Study identification and selection
3. Assessment of study quality
4. Data extraction, analysis, and presentation
5. Interpretation of results

The review starts with a sensible clinical question

P: Population
I: Intervention
C: Comparison
O: Outcome

+ Purpose of the test/strategy
+ Study design
+ Reference standard

Richardson et al. The well-built clinical question: a key to evidence-based decisions. ACP Journal Club 1995;A-12

Sensible clinical question (PICO)

- Population: In adults and children with and without HIV infection suspected of having active tuberculosis
- Intervention: do commercial serological tests
- Comparison: compared with sputum microscopy
- Outcomes: improve sensitivity and specificity?
What is the purpose of the test?

- **Triage**
  - minimize use of invasive or expensive test
- **Add-on**
  - improve diagnosis beyond what is already done
- **Replacement**
  - replace test that is harmful or costly

Bossuyt et al. BMJ 2006

Overview of the study design tree

2. Study identification and selection

- MEDLINE, EMBASE, the Cochrane Register of Diagnostic Test Accuracy Studies (under development)
- Search related diagnostic test accuracy reviews (for example HTA database, DARE etc)
- Check references of relevant studies/reviews
- Use a highly sensitive (broad) search strategy
- Use a wide variety of search terms, both text words and database subject headings (MeSH terms)
- Routine use of search filters should generally be avoided!


Does Bleach Processing Increase the Accuracy of Sputum Smear Microscopy for Diagnosing Pulmonary TB?

Medline search

3. Assessment of study quality

4. Data extraction, analysis, and presentation

- Extract paired estimates of sensitivity and specificity
- Visually examine results of individual studies
- Calculate overall summary estimates using HSROC/bivariate meta-analysis
- Look for and investigate possible reasons for heterogeneity
http://ims.cochrane.org/revman

Enter data
Forest plots of sensitivity and specificity, anda-TB IgG for the diagnosis of pulmonary TB

<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>Aflatoon 1994</td>
<td>35</td>
<td>2</td>
<td>7</td>
<td>62</td>
<td>0.83 [0.68, 0.93]</td>
<td>0.89 [0.93, 1.00]</td>
</tr>
<tr>
<td>Aflatoon 1996 (a)</td>
<td>28</td>
<td>3</td>
<td>5</td>
<td>41</td>
<td>0.69 [0.58, 0.81]</td>
<td>0.83 [0.91, 1.00]</td>
</tr>
<tr>
<td>Kalantri 2005 (a)</td>
<td>36</td>
<td>0</td>
<td>21</td>
<td>40</td>
<td>0.86 [0.71, 0.97]</td>
<td>1.00 [0.91, 1.00]</td>
</tr>
<tr>
<td>Ojeda 2004 (a)</td>
<td>28</td>
<td>19</td>
<td>0</td>
<td>101</td>
<td>0.62 [0.55, 0.69]</td>
<td>0.81 [0.84, 0.88]</td>
</tr>
<tr>
<td>Trauner 2006</td>
<td>33</td>
<td>21</td>
<td>0</td>
<td>59</td>
<td>0.84 [0.69, 0.94]</td>
<td>0.73 [0.82, 0.83]</td>
</tr>
<tr>
<td>Wu 2004 (a)</td>
<td>58</td>
<td>6</td>
<td>34</td>
<td>29</td>
<td>0.62 [0.52, 0.73]</td>
<td>0.89 [0.73, 0.97]</td>
</tr>
<tr>
<td>Wu 2005</td>
<td>35</td>
<td>19</td>
<td>30</td>
<td>40</td>
<td>0.54 [0.41, 0.68]</td>
<td>0.88 [0.54, 0.79]</td>
</tr>
</tbody>
</table>

- One row is displayed for each study
- Extracted data are presented: TP, FP, FN, TN
- Data shown in the graph are also displayed numerically
- Each study result is given a box for a point estimate
- Horizontal line = confidence interval (CI); measures how much we think the result of the study varies with chance
  - The wider the CI, the less confident we are in the result
- We can judge whether results are consistent depending if CIs overlap
Calculating an overall summary

• The hierarchical approach to SROC (HSROC) has emerged as the standard method.

The hierarchical approach to SROC (HSROC)

• Hierarchical model allows for both within and between study variability

• Random effects allows for heterogeneity between studies
Meta-analysis of diagnostic accuracy using hierarchical logistic regression

Abstract. Meta-analysis of diagnostic test accuracy presents many challenges. Even in the simplest cases, when the data are summarised by a 2 × 2 table from each study, a statistically rigorous analysis requires hierarchical (multilevel) models that respect the binomial data structure, such as hierarchical logistic regression. We present a Stata package, metandi, to facilitate the fitting of such models in Stata. The commands display the results in two alternative parameterisations and produce a customisable plot. metandi requires either Stata 10 or above (which has the new command xtlogit), or Stata 8.2 or above with glmans installed.

Keywords: st0151, metandi, metandiplot, diagnostics, meta-analysis, sensitivity and specificity, hierarchical models, generalised mixed models, glmans, atlogen, receiver operating characteristic (ROC), summary ROC, hierarchical summary ROC

Paste data from excel into Stata
Enter commands

Metandi output
Summary ROC plots for anda-TB IgG for diagnosis of TB: (A) smear+ and (B) smear- pulmonary TB patients. Red squares are pooled sensitivity and specificity values.

SROC curve recombinant proteins, Steingart, Laal et al CVI 2009.
Enter data from excel
Select plot and characteristics

Export plot
Heterogeneity

- Refers to variation in results among studies
- May be caused by variation in
  - test thresholds (unique to meta-analyses of diagnostic tests)
  - prevalence of disease
  - patient spectrum
  - study quality
  - chance variation
- *When significant heterogeneity is present, summary estimates from meta-analyses may not be meaningful*

Exploring heterogeneity

- Subgroup (stratified) analyses
- Meta-regression analysis

### Diagnostic Odds Ratio (DOR) and Relative DOR

- **DOR** = odds of a positive result in diseased individuals versus odds of a positive result in non-diseased individuals
- Combines both likelihood ratios  \( \text{DOR} = \frac{LR_+}{LR_-} \)
- **DOR** = 1 means the test cannot discriminate between people with and without disease
- **RDOR** (relative DOR) = ratio of 2 DORs
- **RDOR** = 1 means a particular covariate (e.g. blinded study design) does not affect the overall DOR
Subgroup analyses. The results show a high degree of variability in accuracy across studies, Ling 2008.

<table>
<thead>
<tr>
<th>Study Characteristic (n)</th>
<th>DOR</th>
<th>Ch² test of heterogeneity</th>
<th>P value for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DIRECTION</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prospective (108)</td>
<td>255.63</td>
<td>(199.23, 328.01)</td>
<td>678.67</td>
</tr>
<tr>
<td>Retrospective (9)</td>
<td>315.65</td>
<td>(99.64, 996.57)</td>
<td>150.21</td>
</tr>
<tr>
<td>Both (8)</td>
<td>371.42</td>
<td>(161.83, 852.49)</td>
<td>31.49</td>
</tr>
<tr>
<td><strong>STUDY DESIGN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross Sectional (126)</td>
<td>269.56</td>
<td>(212.30, 342.26)</td>
<td>869.08</td>
</tr>
<tr>
<td><strong>RECRUITMENT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consecutive (45)</td>
<td>220.90</td>
<td>(154.41, 287.60)</td>
<td>180.24</td>
</tr>
<tr>
<td>Convenient (26)</td>
<td>347.98</td>
<td>(225.63, 536.67)</td>
<td>91.71</td>
</tr>
<tr>
<td>Both (35)</td>
<td>298.50</td>
<td>(90.73, 982.18)</td>
<td>40.04</td>
</tr>
<tr>
<td>Random (2)</td>
<td>278.72</td>
<td>(3.12, 24901.4)</td>
<td>9.73</td>
</tr>
<tr>
<td>Not Reported (51)</td>
<td>284.91</td>
<td>(184.02, 441.13)</td>
<td>529.38</td>
</tr>
<tr>
<td><strong>VERIFICATION</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete (1123)</td>
<td>264.79</td>
<td>(208.66, 336)</td>
<td>865.88</td>
</tr>
<tr>
<td><strong>BLINDING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both (8)</td>
<td>163.93</td>
<td>(69.91, 384.42)</td>
<td>75.49</td>
</tr>
</tbody>
</table>

Meta-regression

- Is a form of linear regression in which studies are the unit of analysis
- Aims to relate the size of effect to one or more characteristics of the studies involved
- DOR is the dependent variable
- Covariates that might be associated with the variability in DOR are the independent variables
- Tip: Specify covariates that you want to explore in advance
The threshold effect (-0.21) was significant ($p = 0.01$). This was also seen in the SROC plot, Ling 2008.

5. Interpretation of results

- What are the consequences of using the test in terms of the numbers of TP, FP, FN, and TN?
- How applicable are the results?
- To what extent were the primary studies biased? If serious study limitations were identified, could these impact the results?
- What are the implications for research?
References and tools for meta-analysis

- Zamora. BMC Medical Research Methodology 2006, 6:31
- Cochrane Diagnostic Test Accuracy Working Group http://srdta.cochrane.org/
- http://www.teachepi.org/ Dr Pai’s website for learning and teaching epidemiology
- http://www.tbevidence.org/ Evidence-based TB diagnosis
- RevMan http://ims.cochrane.org/revman
In summary

• Described key steps in a systematic review/meta-analysis of diagnostic test accuracy

• Demonstrated HSROC/bivariate meta-analysis of data from diagnostic studies

• Identified key references and tools for performing systematic reviews of diagnostic test accuracy

With special thanks to

• Mariska Leeflang
• Madhu Pai
• Many others

Workshop on Meta-analyses of Diagnostic Test Accuracy, Montreal, May 2009