

# Meta-analysis of diagnostic studies

Karen R Steingart, MD, MPH  
Francis J. Curry International Tuberculosis Center  
University of California, San Francisco  
Berlin, 12 November 2010  
[karenst@uw.edu]



## Disclosure

- I serve as co-chair of the Evidence Synthesis subgroup of Stop TB Partnership's New Diagnostics Working Group
- I am a member of the GRADE (Grading of Recommendations Assessment, Development and Evaluation) Working Group

## 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 systematic reviews of diagnostic studies

## 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
  - ...calculates an overall summary; considers both dimensions of test performance
  - ...looks for and investigates possible reasons for inconsistency in results (heterogeneity)
  - ...evaluates the impact of quality and other study characteristics on diagnostic accuracy
  - ...stimulates new research questions
- Meta-analyses (pooling) can increase the precision of the overall result

## Measures of test performance

	Disease present	Disease absent	Total
Index test +	True positives (TP)	False positives (FP)	TP + FP
Index test -	False negative (FN)	True negatives (TN)	FN + TN
Total	TP + FN	TN + FP	TP + FP + FN + TN

Sensitivity =  $TP/(TP+FN)$

Specificity =  $TN/(FP + TN)$

Positive predictive value =  $TP/(TP + FP)$

Negative predictive value =  $TN/(FN + TN)$

Likelihood ratio positive =  $Sensitivity/(1 - Specificity)$

Likelihood ratio negative =  $(1 - Sensitivity)/Specificity$

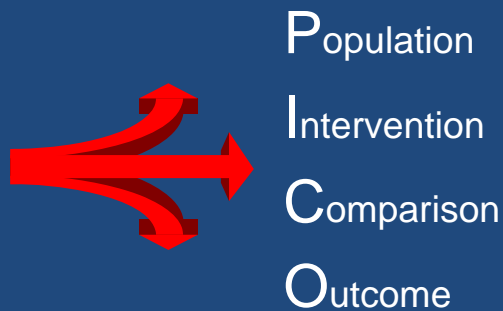
Prevalence (proportion of people with disease in population to whom the test has been applied) =  $TP + FN/(TP + FP + FN + TN)$

## 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

Leeflang. Ann Intern Med. 2008;149:889-897

## The objectives of the review



**+ 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

**P**opulation: In patients suspected of active TB

**I**ntervention: do commercial serological tests

**C**omparison: compared with sputum microscopy

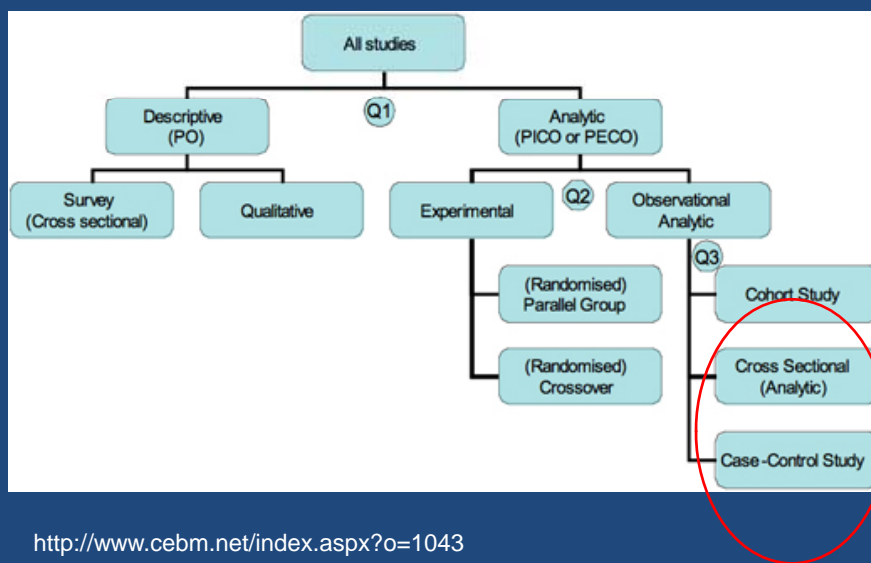
**O**utcomes: improve sensitivity and specificity

## 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

- Search 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 highly sensitive (broad) search strategy
- Reflect key concepts of the review (focus on index test and target condition) in search
- 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!

Bossuyt PM, Leeflang MM. Chapter 6: Developing Criteria for Including Studies. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4 [updated September 2008]*. The Cochrane Collaboration, 2008.

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

- Search (tuberculosis[MeSH] OR mycobacterium tuberculosis[MeSH] OR tuberculosis[ti]) AND (microscopy[MeSH] OR (sputum[MeSH] AND smear\*) OR acid-fast[TI] OR (AFB[TIAB] AND smear\*) OR (AFB[TIAB] AND sputum) OR (sputum smear\*[TI]) OR (smear examination\*[TI]) OR ("sputum microscopy"[TI]) OR (bacteriolog\*[TI] AND tuberculosis[TI]) OR (direct microscop\*[TI]) OR (sensitivity[TI] AND microscopy[TI]) OR (microbiolog\*[TI] AND tuberculosis[TI]))

### 3. Assessment of study quality

#### **BMC Medical Research Methodology**



Research article

**Open Access**

#### **The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews**

Penny Whiting\*<sup>1</sup>, Anne WS Rutjes<sup>2</sup>, Johannes B Reitsma<sup>2</sup>,  
Patrick MM Bossuyt<sup>2</sup> and Jos Kleijnen<sup>1</sup>

Address: <sup>1</sup>Centre for Reviews and Dissemination, University of York, England, UK and <sup>2</sup>Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, University of Amsterdam, The Netherlands

Email: Penny Whiting\* - pfw2@york.ac.uk; Anne WS Rutjes - a.rutjes@amc.uva.nl; Johannes B Reitsma - j.reitsma@amc.uva.nl; Patrick MM Bossuyt - p.m.bossuyt@amc.uva.nl; Jos Kleijnen - jk13@york.ac.uk

\* Corresponding author

Published: 10 November 2003

Received: 14 July 2003

*BMC Medical Research Methodology* 2003, 3:25

Accepted: 10 November 2003

This article is available from: <http://www.biomedcentral.com/1471-2288/3/25>

© 2003 Whiting et al; licensee BioMed Central Ltd. This is an Open Access article; verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.



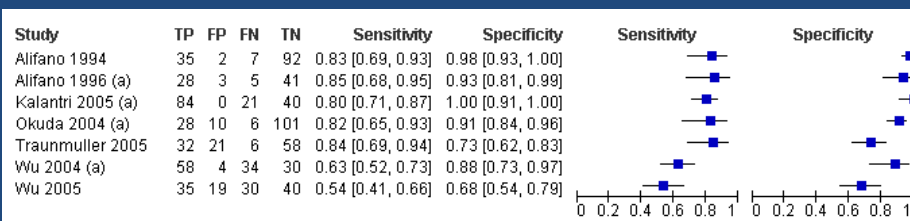
## 4. Data extraction, analysis, and presentation

- 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>

The screenshot displays the Cochrane IMS website. At the top, there is a navigation menu with links for Home, RevMan, Archive, Projects, Support, and Organisation. Below the menu, the main content area is titled "RevMan" and features a large blue "5" logo. The text describes RevMan as software used for preparing and maintaining Cochrane Reviews, highlighting its capabilities for data extraction, analysis, and presentation. A sidebar on the left contains links for "About RevMan 5", "Licensing", "Download", "New Releases", "Documentation", "Troubleshooting", and "Other Resources". Below the sidebar, there are sections for "Other Resources" and "News from the IMS".

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



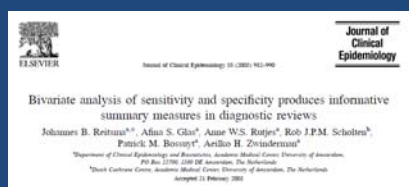
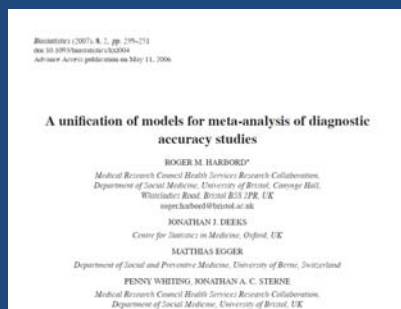
Steingart 2010 unpublished

Study	TP	FP	FN	TN	Sensitivity	Specificity	Sensitivity
Alifano 1994	35	2	7	92	0.83 [0.69, 0.93]	0.98 [0.93, 1.00]	
Alifano 1996 (a)	28	3	5	41	0.85 [0.68, 0.95]	0.93 [0.81, 0.99]	
Kalantri 2005 (a)	84	0	21	40	0.80 [0.71, 0.87]	1.00 [0.91, 1.00]	

- 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

## Metandi in Stata

The Stata Journal (2009)  
9, Number 2, pp. 211–229

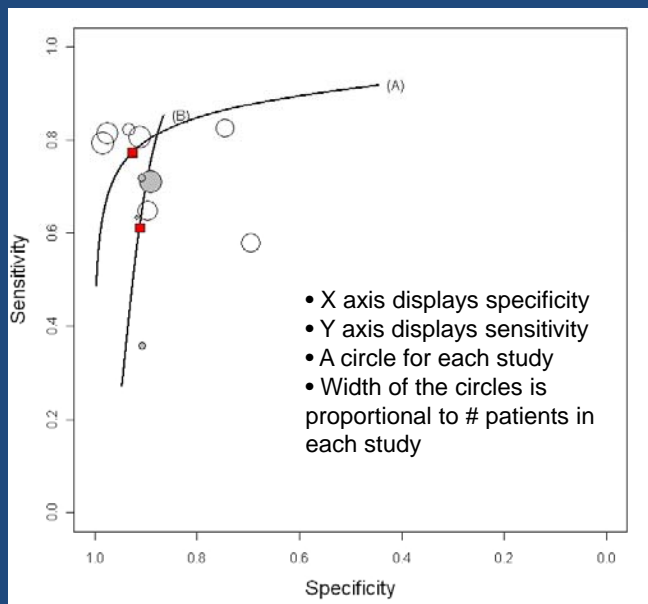
### metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression

Roger M. Harbord  
Department of Social Medicine  
University of Bristol  
Bristol, UK  
roger.harbord@bristol.ac.uk

Penny Whiting  
Department of Social Medicine  
University of Bristol  
Bristol, UK

**Abstract.** Meta-analysis of diagnostic test accuracy presents many challenges. Even in the simplest case, when the data are summarized by a  $2 \times 2$  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 parameterizations and produce a customizable plot. `metandi` requires either Stata 10 or above (which has the new command `xtmelogit`), or Stata 8.2 or above with `gllamm` installed.

**Keywords:** `st0163`, `metandi`, `metandiplot`, diagnosis, meta-analysis, sensitivity and specificity, hierarchical models, generalized mixed models, `gllamm`, `xtmelogit`, receiver operating characteristic (ROC), summary ROC, hierarchical summary ROC



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. Steingart unpublished

## BMC Medical Research Methodology

Software
Open Access

### Meta-DiSc: a software for meta-analysis of test accuracy data

Javier Zamora\*<sup>1</sup>, Victor Abraira<sup>1</sup>, Alfonso Muriel<sup>1</sup>, Khalid Khan<sup>2</sup> and Arri Coomarasamy<sup>2</sup>

Address: <sup>1</sup>Clinical Biostatistics Unit, Ramón y Cajal Hospital, Madrid, Ctra. Colmenar km 9.100 Madrid 28034, Spain and <sup>2</sup>University of Birmingham and Birmingham Women's Hospital, Edgbaston, Birmingham, UK

Email: Javier Zamora\* - javier.zamora@hrc.es; Victor Abraira - victor.abraira@hrc.es; Alfonso Muriel - Alfonso.muriel@hrc.es; Khalid Khan - k.s.khan@bham.ac.uk; Arri Coomarasamy - arricomar@blueyonder.co.uk

\* Corresponding author

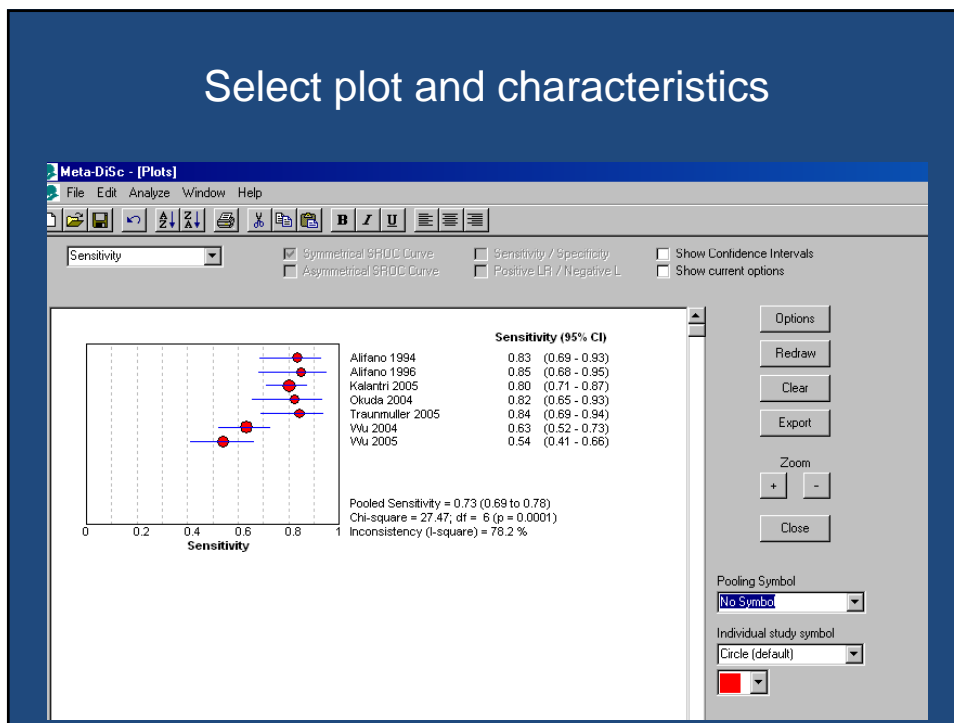
Published: 12 July 2006 Received: 31 March 2006

*BMC Medical Research Methodology* 2006, 6:31 doi:10.1186/1471-2288-6-31 Accepted: 12 July 2006

This article is available from: <http://www.biomedcentral.com/1471-2288/6/31>

© 2006 Zamora et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



## 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*

27

## Exploring heterogeneity

- Subgroup (stratified) analyses
- Meta-regression analysis

## 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?

**Table 2. GRADE Summary of Findings – Role of IGRAs for evaluation of patients with pulmonary TB in low- and middle-income countries**

**Review question:** What is the diagnostic accuracy of commercial IGRAs for pulmonary tuberculosis?  
**Patients/population:** Adult pulmonary TB suspects and confirmed cases in low- and middle-income countries  
**Setting:** outpatients and inpatients  
**Index test:** Commercial Interferon-gamma Release Assays (QuantiFERON-TB Gold In-Tube [QFT-GIT], Cellestis, Australia and T-SPOT.TB [T-SPOT], Oxford Immunotec, United Kingdom)  
**Importance:** Rapid, accurate, simple test could supplement microscopy and expand testing to peripheral health centers  
**Reference standard:** Microbiologic (culture or smear-microscopy) or clinical diagnosis of pulmonary TB  
**Studies:** Cross-sectional or cohort

Outcomes: TP, TN, FP, FN	Effect % (95% CI)	No. of participants (studies)	What do these results mean given 10% prevalence among suspects being screened for TB?	What do these results mean given 30% prevalence among suspects being screened for TB?	Quality of Evidence
Subgroups					
T-SPOT.TB, HIV-infected	Sensitivity 78% (56, 91) Specificity 55% (45, 64)	549 (5)	With a prevalence of 10%, 100/1000 will have TB. Of these, 78 (TP) will be identified; 22 (FN) will be missed by T-SPOT.TB. Of the 900 patients without TB, 495 (TN) will not be treated; 405 (FP) will be unnecessarily treated.	With a prevalence of 30%, 300/1000 will have TB. Of these, 234 (TP) will be identified; 66 (FN) will be missed by T-SPOT.TB. Of the 700 patients without TB, 385 (TN) will not be treated; 315 (FP) will be unnecessarily treated.	Very Low ⊕○○○

Metcalf unpublished

## Summary

- Described key steps in a systematic review/ meta-analysis of diagnostic test accuracy
- Described HSROC/bivariate meta-analysis of data from diagnostic studies
- Identified key references and tools for performing systematic reviews of diagnostic test accuracy





## Going beyond diagnostic accuracy...

- Link evidence on diagnostic test accuracy to clinical practice
- Frame review questions so that influences of the clinical context on test performance are identified
- Use results of systematic reviews of diagnostic test accuracy as inputs into decision analyses
- Go beyond summary ROC curves to describe test performance in terms of the expected downstream benefits and harms of using a test

Diagnostic Test Accuracy and Clinical Decision Making *Ann Intern Med.* 2008;149:904-906

## References and tools for meta-analysis

- Leeflang. *Ann Intern Med.* 2008;149:889-897
- Rutter and Gatsonis. *Stat Med.* 2001; 20:2865–2884
- Reitsma. *J Clin Epidemiol.* 2005; 982–990
- 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.thevidence.org/> Evidence-based TB diagnosis
- RevMan <http://ims.cochrane.org/revman>
- Making Sense of Likelihood Ratios, McMaster University  
[http://ebm.mcmaster.ca/video/making\\_sense\\_of\\_likelihood\\_ratios.html](http://ebm.mcmaster.ca/video/making_sense_of_likelihood_ratios.html)
- Meta-analysis in Stata... Ed. Jonathan Sterne 2009

*With special thanks to*

- Mariska Leeflang
- Madhu Pai
- Many others



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