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

	Disease Present	Disease Absent	Total
Test +	**True Positives (TP)**	**False Positives (FP)**	**TP + FP**
Test -	**False Negatives (FN)**	**True Negatives (TN)**	**FN + TN**
Total	**TP + FN**	**TN + FP**	**TP + FP + FN + TN**

Measures of test performance

	Disease Present	Disease Absent	Total
Test +	TP	FP	TP + FP
Test -	FN	TN	FN + TN
Total	TP + FN	TN + FP	TP + FP + FN + TN

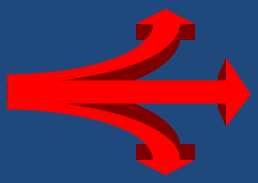
Sensitivity = TP/(TP+FN)
Specificity = TN/(FP + TN)
Positive predictive value = TP/(TP + FP)
Negative predictive value = TN/(FN + TN)
Likelihood ratio positive = Sensitivity/(1 – Specificity)
Likelihood ratio negative = (1 – Sensitivity)/Specificity
Prevalence (proportion of people with disease in population 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 review starts with a sensible clinical question



Population

Intervention

Comparison

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)

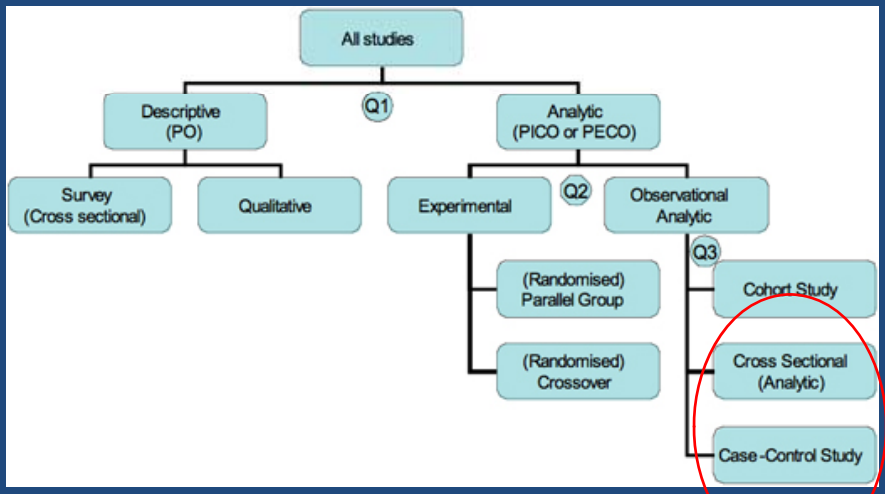
- **P**opulation: In adults and children with and without HIV infection suspected of having active tuberculosis
- **I**ntervention: do commercial serological tests
- **C**omparison: compared with sputum microscopy
- **O**utcomes: improve sensitivity and specificity?

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



http://www.cebm.net/index.aspx?o=1043

2. Study identification and selection

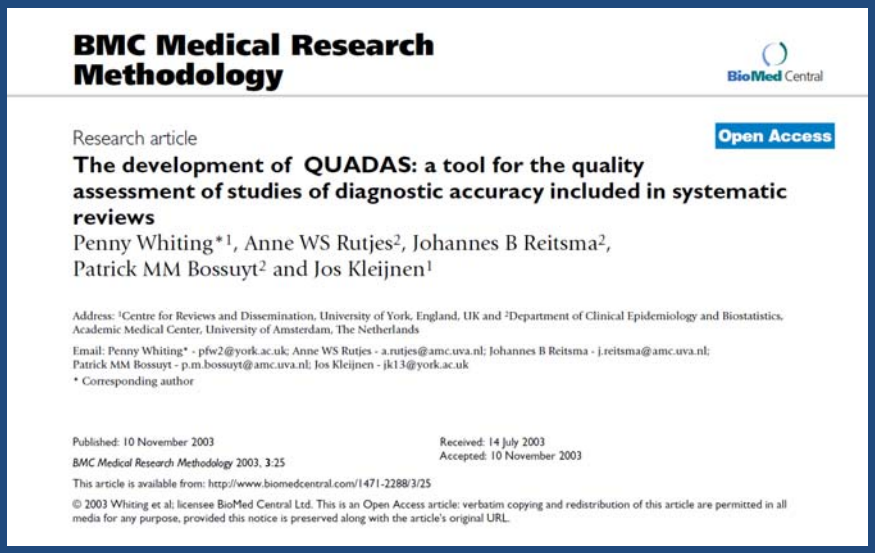
- MEDLINE, EMBASE, the Cochrane Register of Diagnostic Test Accuracy Studies (under development)
- Search related diagnostic test accuracy reviews (for example HTA database, DARE etc)
- Check references of relevant studies/reviews
- Use a highly sensitive (broad) search strategy
- Use a wide variety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

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

BioMed Central

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*[1], Anne WS Rutjes[2], Johannes B Reitsma[2], Patrick MM Bossuyt[2] and Jos Kleijnen[1]

Address: [1]Centre for Reviews and Dissemination, University of York, England, UK and [2]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

BMC Medical Research Methodology 2003, 3:25

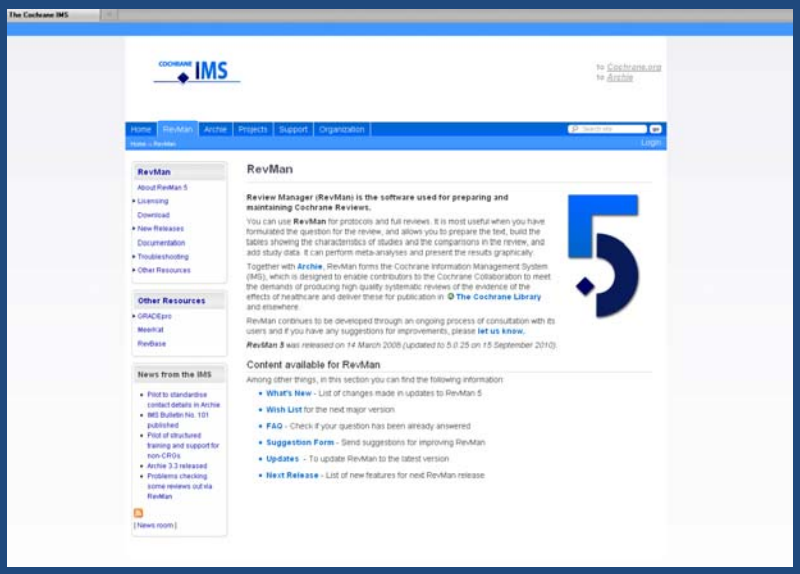
Received: 14 July 2003
Accepted: 10 November 2003

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

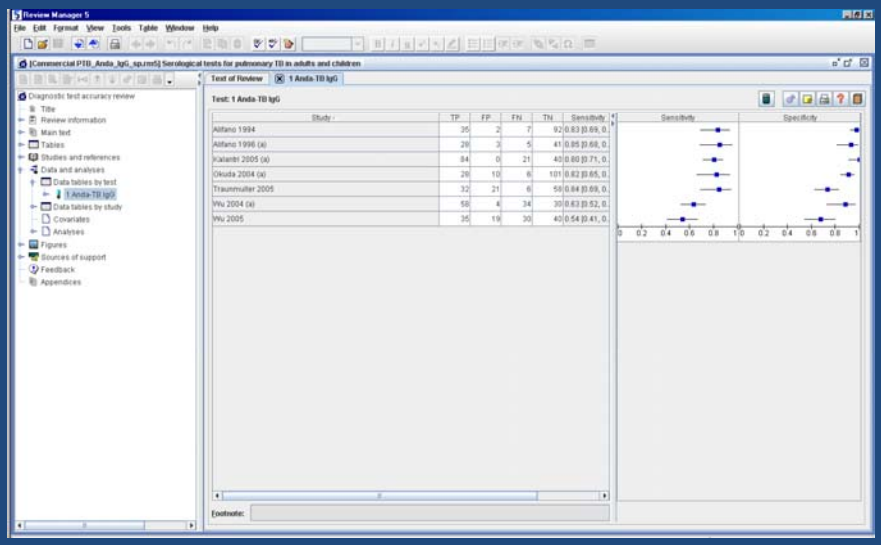
4. Data extraction, analysis, and presentation

- 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

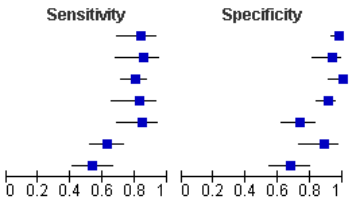


Enter data



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

Study	TP	FP	FN	TN	Sensitivity	Specificity
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]
Okuda 2004 (a)	28	10	6	101	0.82 [0.65, 0.93]	0.91 [0.84, 0.96]
Traunmuller 2005	32	21	6	58	0.84 [0.69, 0.94]	0.73 [0.62, 0.83]
Wu 2004 (a)	58	4	34	30	0.63 [0.52, 0.73]	0.88 [0.73, 0.97]
Wu 2005	35	19	30	40	0.54 [0.41, 0.66]	0.68 [0.54, 0.79]



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

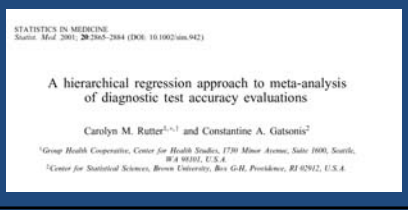
A unification of models for meta-analysis of diagnostic accuracy studies

ROGER M. HARBORD*

JONATHAN J. DEEKS

MATTHIAS EGGER

PENNY WHITING, JONATHAN A. C. STERNE



STATISTICS IN MEDICINE

A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations

Carolyn M. Rutter and Constantine A. Gatsonis

ELSEVIER

Journal of Clinical Epidemiology

Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews

Johannes B. Reitsma, Afina S. Glas, Anne W.S. Rutjes, Rob J.P.M. Scholten, Patrick M. Bossuyt, Aeilko H. Zwinderman

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

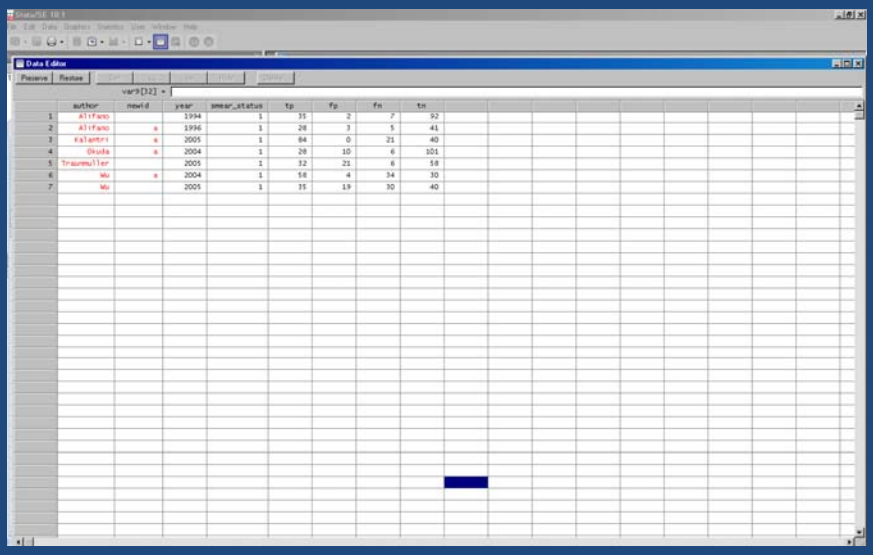
Roger M. Harbord
Department of Social Medicine
University of Bristol
Bristol, UK
roger.harbord@bristol.ac.uk

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

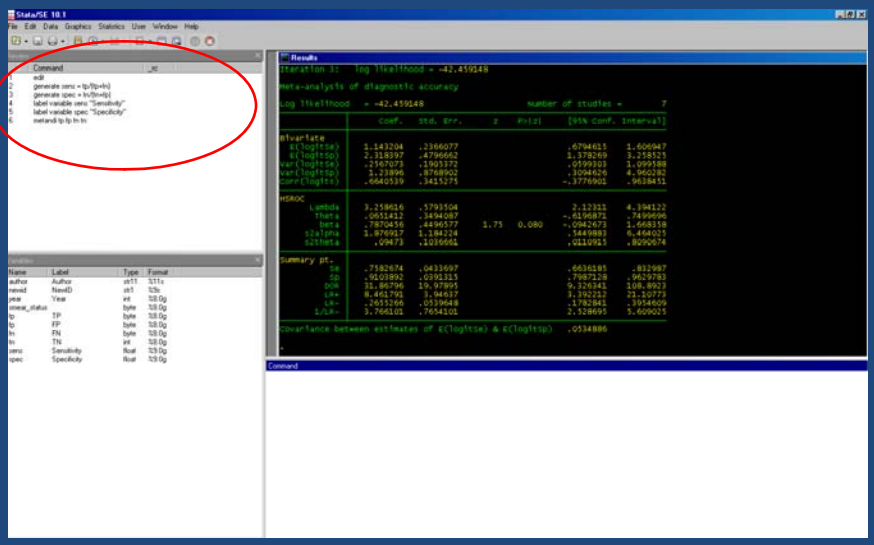
Abstract. Meta-analysis of diagnostic test accuracy presents many challenges. Even in the simplest case, when the data are summarized by a 2×2 table from each study, a statistically rigorous analysis requires hierarchical (multilevel) models that respect the binomial data structure, such as hierarchical 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

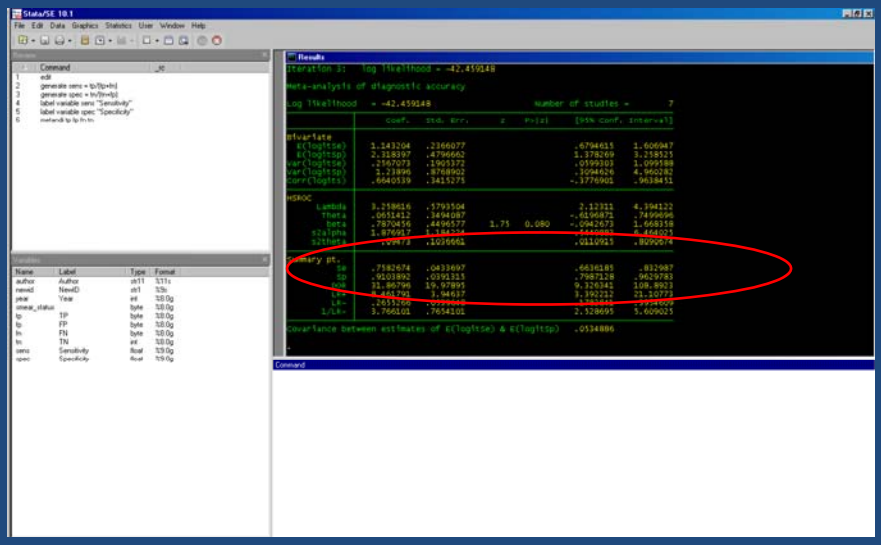
Paste data from excel into Stata

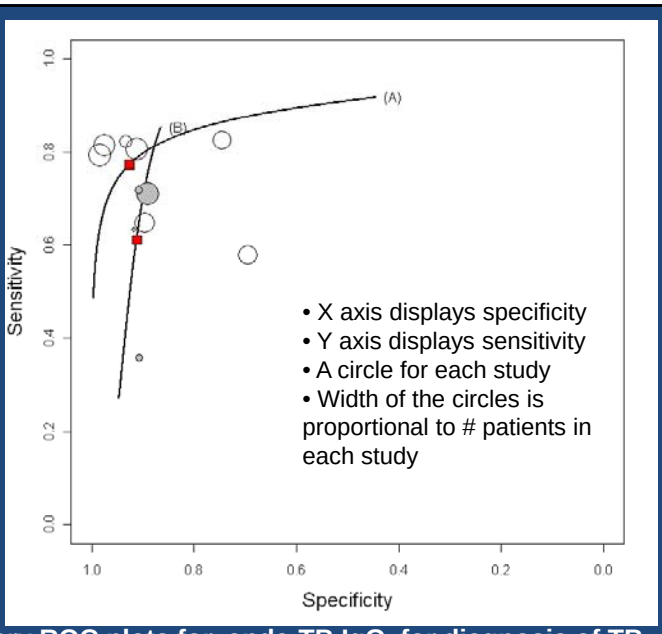


Enter commands

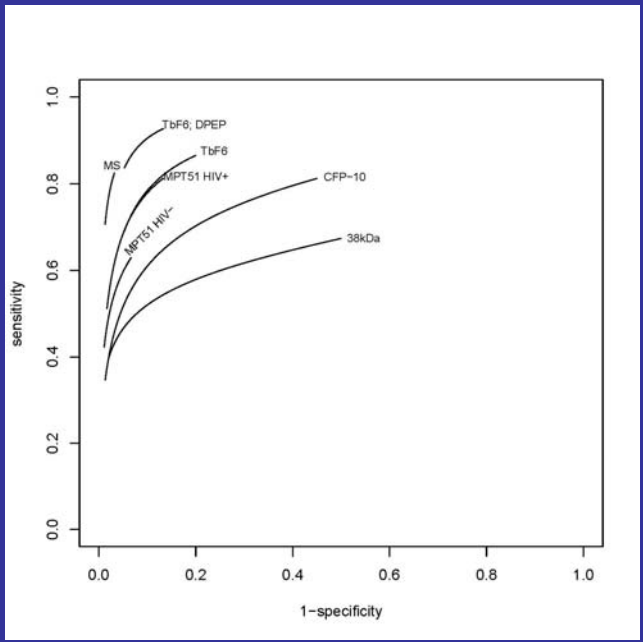


Metandi output





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

BMC Medical Research Methodology

BioMed Central

Software

Open Access

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

Javier Zamora*[1], Victor Abraira[1], Alfonso Muriel[1], Khalid Khan[2] and Arri Coomarasamy[2]

Address: [1]Clinical Biostatistics Unit, Ramón y Cajal Hospital, Madrid, Ctra. Colmenar km 9.100 Madrid 28034, Spain and [2]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 - arricoomar@blueyonder.co.uk

* Corresponding author

Published: 12 July 2006

BMC Medical Research Methodology 2006, 6:31 doi:10.1186/1471-2288-6-31

Received: 31 March 2006
Accepted: 12 July 2006

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

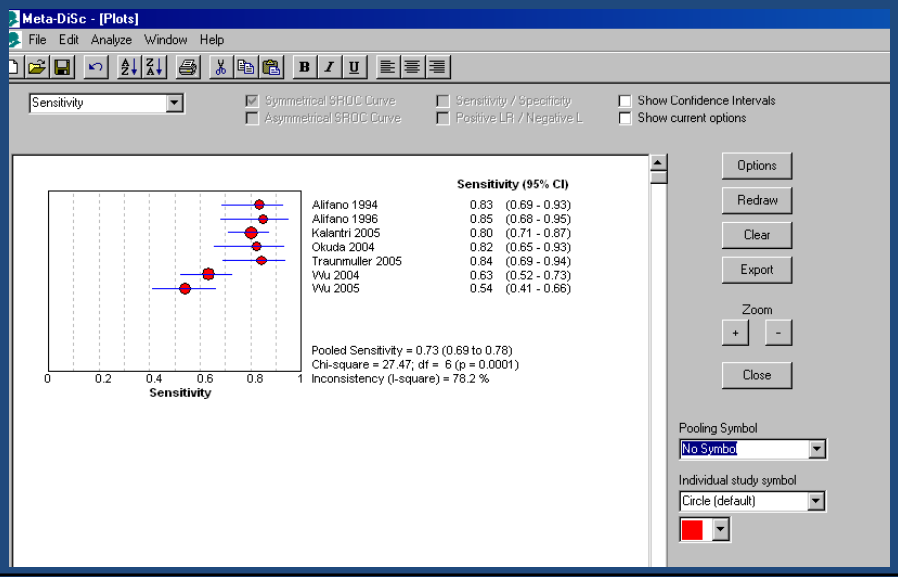
Enter data from excel

Meta-DiSc - [Data - Untitled]

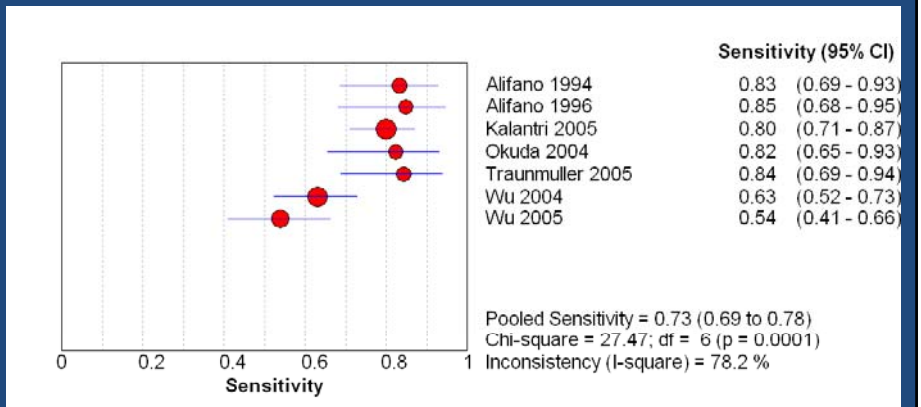
File Edit Analyze Window Help

No.	Author	StudyId	Smear	TP	FP	FN	TN
1	Alifano 1994		1	35	2	7	92
2	Alifano 1996	a	1	28	3	5	41
3	Kalantri 2005	a	1	84	0	21	40
4	Okuda 2004	a	1	28	10	6	101
5	Traunmuller 2005		1	32	21	6	58
6	Wu 2004	a	1	58	4	34	30
7	Wu 2005		1	35	19	30	40
8							
9							
10							
11							
12							
13							
14							

Select plot and characteristics



Export plot

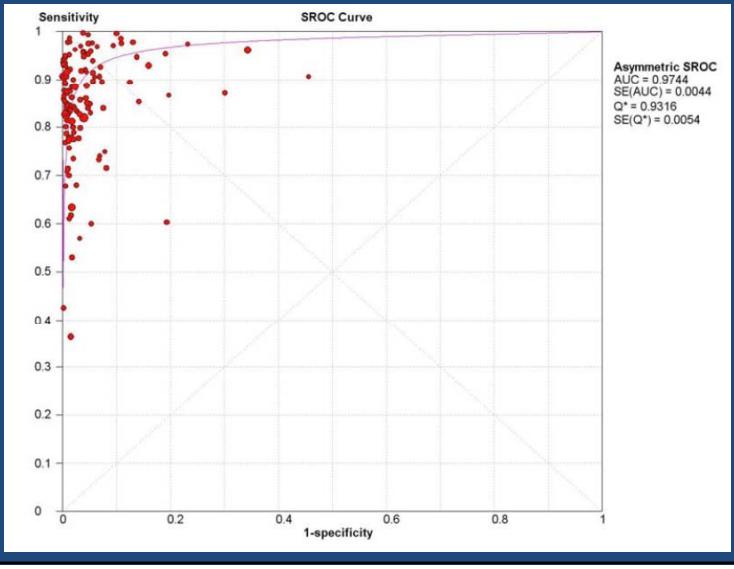


Heterogeneity

- 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*

35

Commercial Nucleic-Acid Amplification Tests for Diagnosis of Pulmonary Tuberculosis in Respiratory Specimens… Ling et al PLoS One 2008



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 DOR = 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 4. Diagnostic Odds Ratio (DOR) Estimates from Subgroup Analysis

Study Characteristic (n)	DOR	Chi^2 test of heterogeneity	P value for heterogeneity
DIRECTION			
Prospective (108)	255.63 (199.23, 328.01)	678.67	<.001
Retrospective (9)	315.65 (99.68, 999.57)	150.21	<.001
Both (8)	371.42 (161.83, 852.49)	31.40	<.001
STUDY DESIGN			
Cross Sectional (124)	269.56 (212.30, 342.26)	869.08	<.001
RECRUITMENT			
Consecutive (43)	220.90 (154.41, 316.00)	180.24	<.001
Convenient (24)	347.98 (225.63, 536.67)	91.71	<.001
Both (5)	298.50 (90.72, 982.18)	40.54	<.001
Random (2)	278.72 (3.12, 24901.4)	9.73	0.002
Not Reported (51)	284.91 (184.02, 441.13)	529.38	<.001
VERIFICATION			
Complete (123)	264.79 (208.66, 336)	863.88	<.001
BLINDING			
Both (8)	163.93 (69.91, 384.42)	25.49	0.001

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.

Table 6. Results from Meta-Regression Analysis Using the Restricted Maximum Likelihood Method

Comparison	Model Coefficient	Relative Diagnostic Odds Ratio (95% CI)	P value
Threshold Effect (S)	−0.21	—	0.01
Retrospective/Both (17) vs Prospective Design (108)	0.13	1.14 (0.56, 2.33)	0.71
Some Convenient Sampling/NR (80) vs Consecutive/Random Sampling (45)	0.38	1.46 (0.87, 2.43)	0.15
No Blinding/NR (105) vs Any Blinding (20)	0.25	1.29 (0.65, 2.58)	0.47
FDA-Approved NAATs (92) vs Not FDA-Approved NAATs (33)	−0.06	0.95 (0.53, 1.68)	0.85
Respiratory Specimens (95) vs Sputum Specimens (30)	0.64	1.89 (1.01, 3.52)	0.05
Culture Reference Standard (105) vs Clinical Reference/Both (20)	0.34	1.40 (0.70, 2.81)	0.34
Resolved Data (37) vs Unresolved Data (88)	−0.05	0.95 (0.54, 1.66)	0.86

doi:10.1371/journal.pone.0001536.t006

Determined using 'Metareg' command in Stata

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 Subgroups	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
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 ⊕○○○

Metcalfe unpublished

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.tbevidence.org/ Evidence-based TB diagnosis
- RevMan http://ims.cochrane.org/revman
- Meta-analysis in Stata… Ed. Jonathan Sterne 2009

In summary

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