Systematic reviews of diagnostic test accuracy

Karen R Steingart, MD, MPH

July 10, 2014

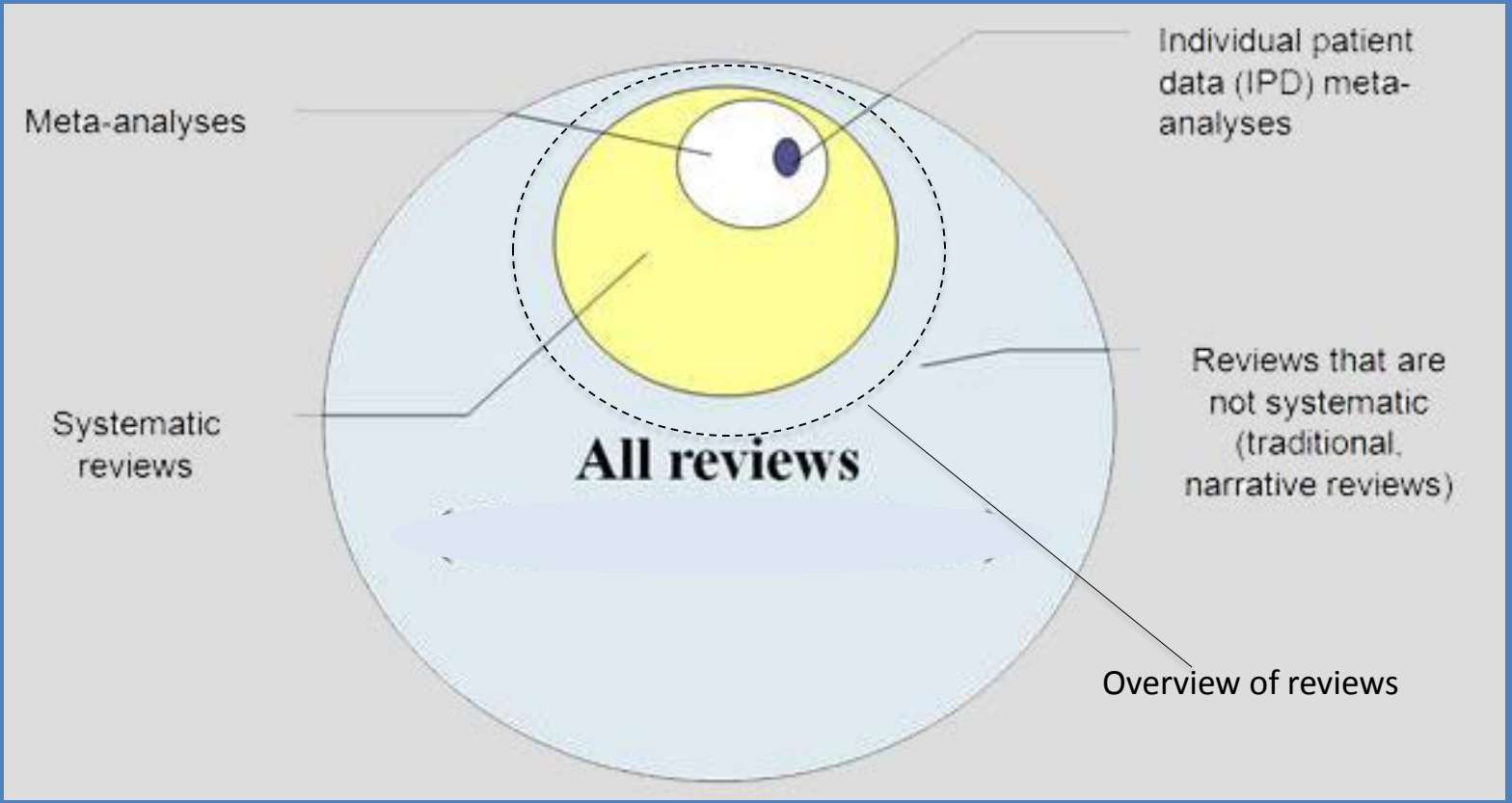
karen.steingart@gmail.com

Conflicts of interest

- Editor Cochrane Infectious Diseases Group
- Editor Cochrane Diagnostic Test Accuracy Working Group
- Member GRADE Working Group
- No financial interests to declare

Overview

- Describe key steps in a systematic review of diagnostic test accuracy
- Describe standard methods for meta-analysis of diagnostic test accuracy
- Describe resources for those who consider doing a systematic review of diagnostic test accuracy



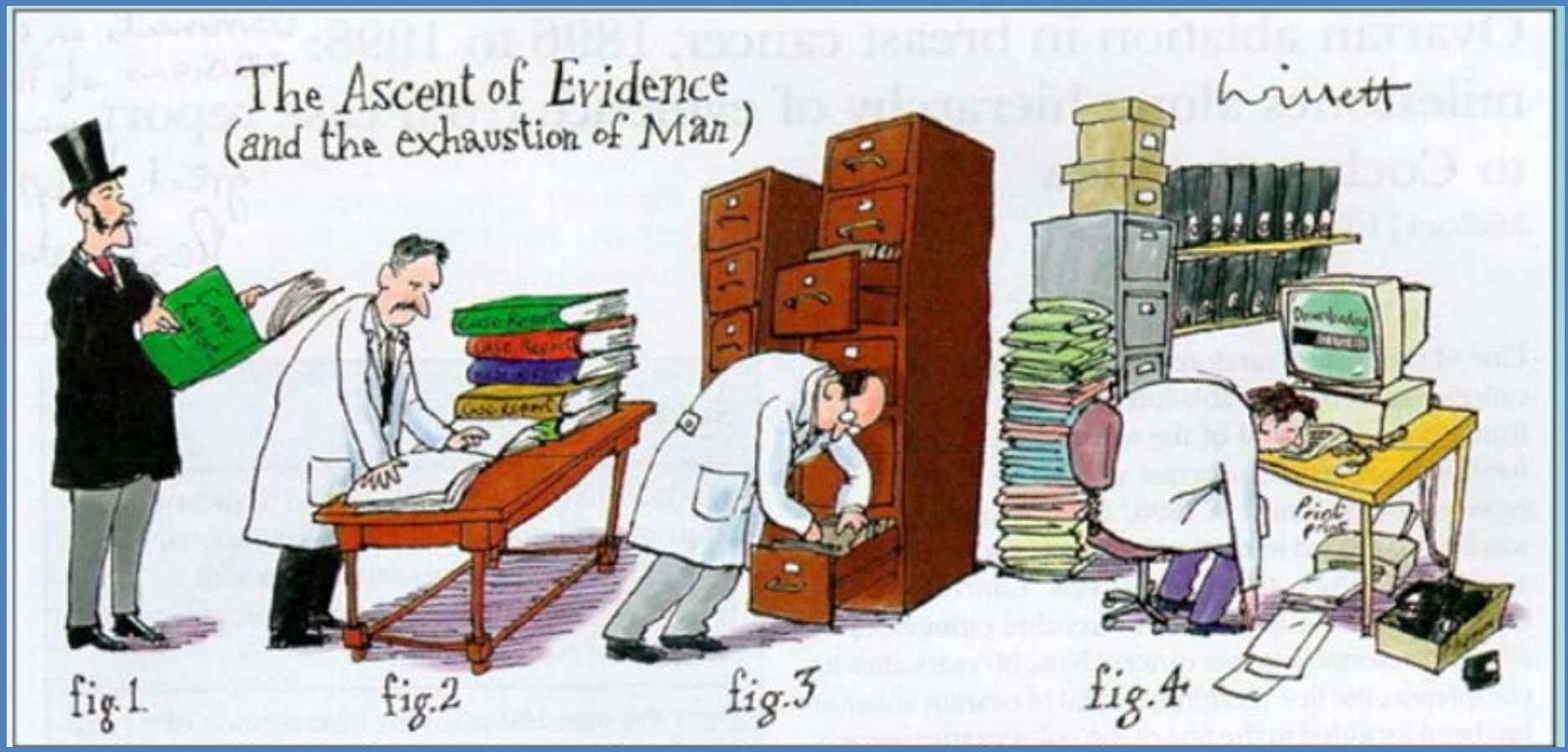
A systematic review starts with a clearly formulated question and uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review

Egg slide adapted from Madhu Pai

Why systematic reviews?

- Scientific summary of all available evidence
- Transparent and reproducible
- Minimize bias
- Studies can be formally compared to establish generalizability and consistency
- Heterogeneity can be identified and investigated
- Meta-analyses may increase the precision of the overall results

Why systematic reviews?



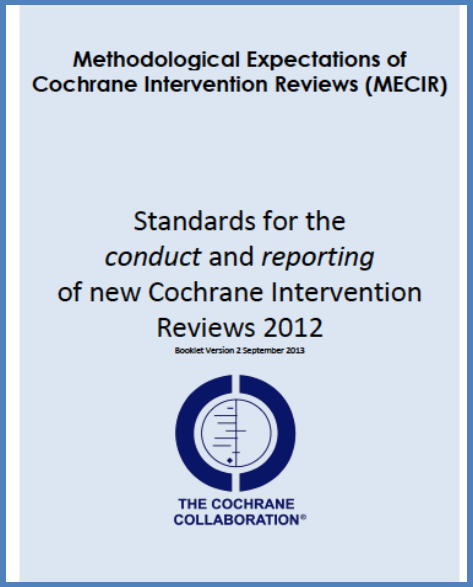
Standards for systematic reviews

- 21 standards (~ 80 elements of performance)

US Institute of Medicine

- 80 standards for conducting
- 108 standards for reporting

The Cochrane Collaboration



Selected Elements, Protocols and Reviews of Diagnostic Test Accuracy, Chapter 4, Handbook

Plain language summary			
Background	Target condition being diagnosed*		P
	Index test(s)*		P
	Clinical pathway	Prior test(s)*	P
		Role of index test(s)*	P
		Alternative test(s) *	P
	Rationale*		P
Objectives	Secondary objectives*		P
Methods	Criteria for considering studies for this review	Types of studies	P
		Participants	P
		Index tests	P
		Target conditions	P
		Reference standards	P
	Search methods for identification of studies	Electronic searches	P
		Searching other resources *	P
	Data collection and analysis	Selection of studies	P
		Data extraction and management	P
		Assessment of methodological quality	P
		Statistical analysis and data synthesis	P
		Investigations of heterogeneity *	P
		Sensitivity analyses *	P
		Assessment of reporting bias *	P
Results	Results of search		
	Methodological quality of included studies		
	Findings		
Discussion	Summary of main results		
	Strengths and weaknesses of review		
	Applicability of findings to review question		

Systematic review author

Diagnostic test accuracy - 1

- Diagnosis: does this person have this disease (more generally, target condition) at this point in time?

- Diagnostic test accuracy: refers to the ability of a test to distinguish between patients with disease and those without disease

Diagnostic test accuracy - 2

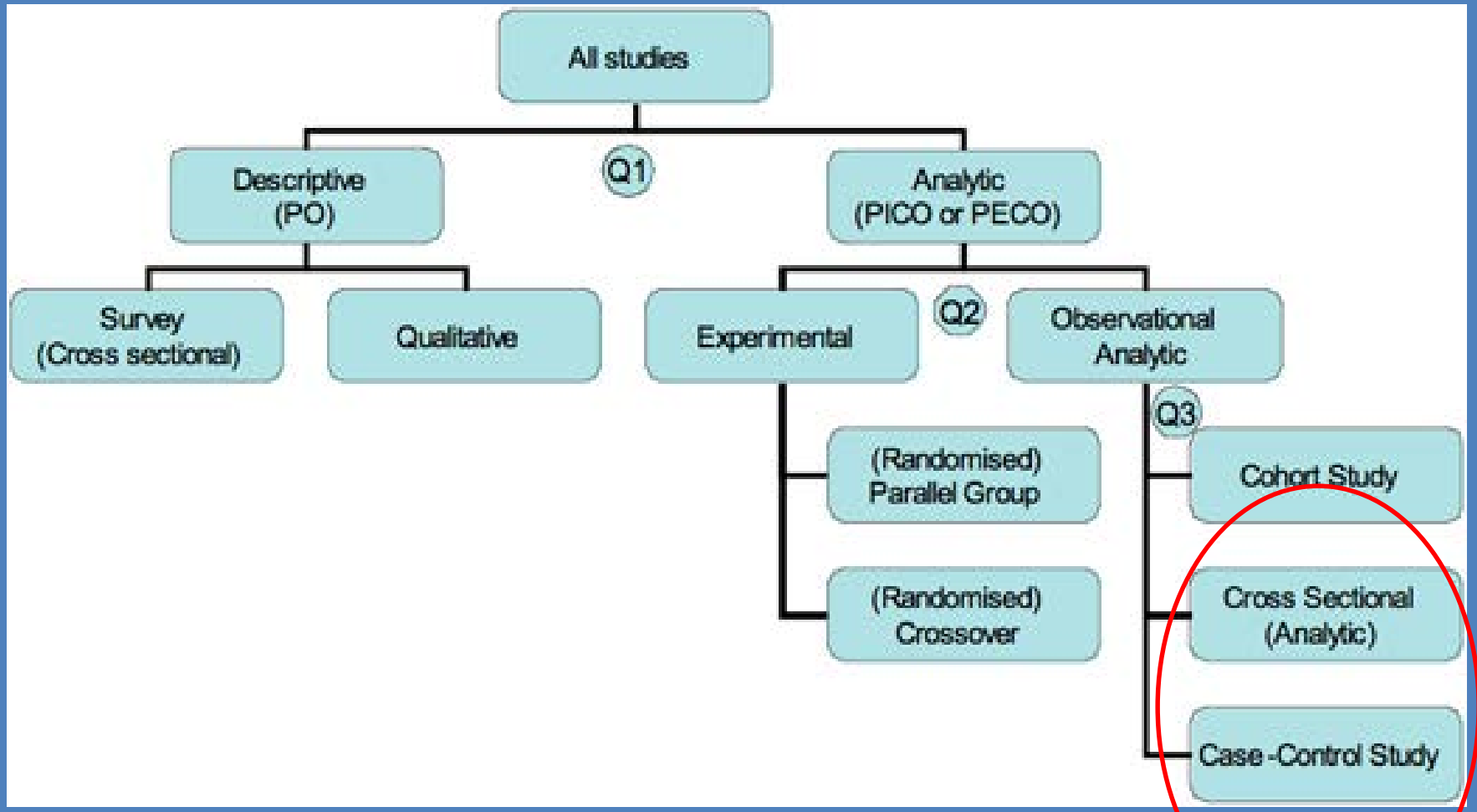
- Index test(s) - new (or simpler, cheaper, faster, less invasive) test; tests being compared
- Reference standard (gold standard) - an agreed-upon and accurate method for identifying patients who have the target condition

➢ Determine agreement between results of index test and reference standard

➢ Focus on pairs of sensitivity and specificity

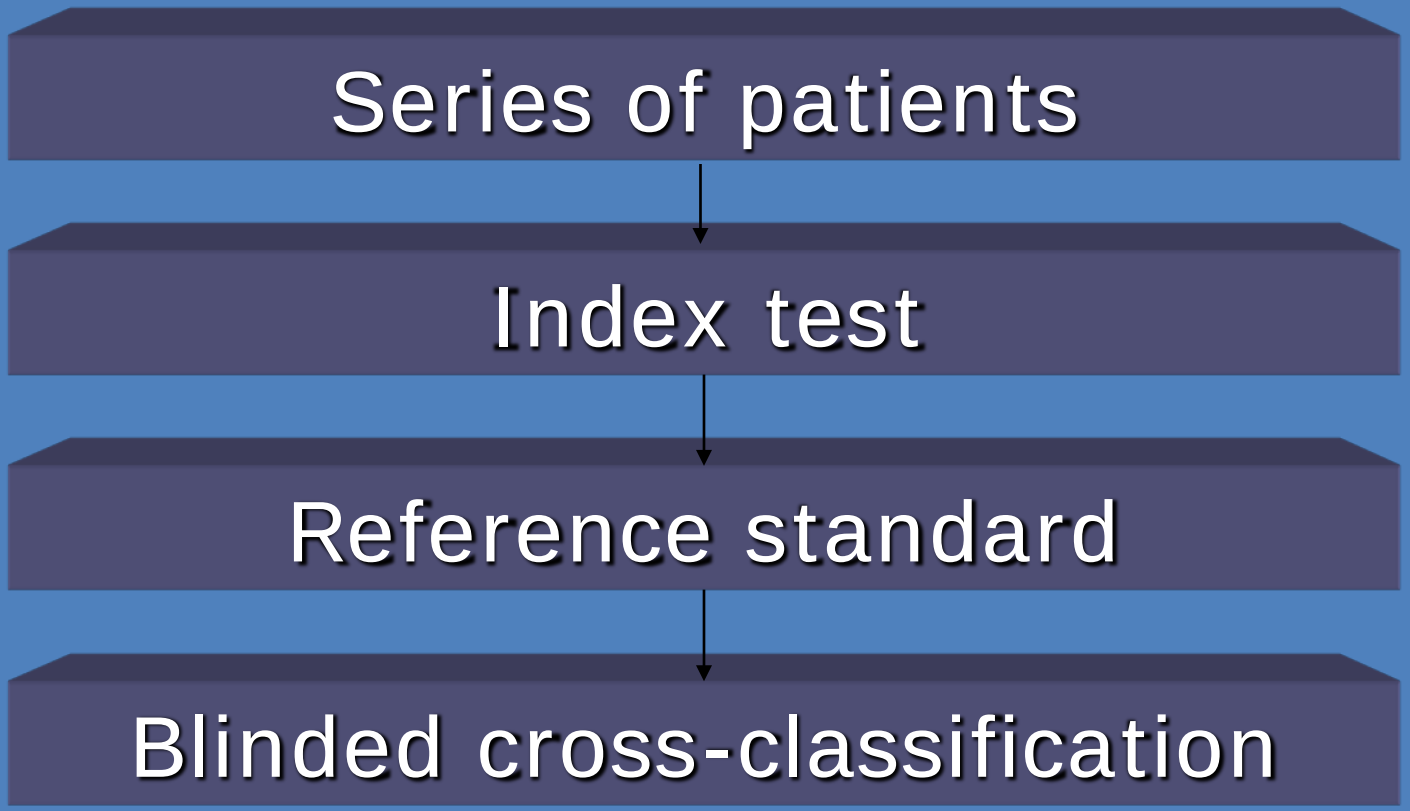
➢ Requires a 2×2 table

Study design tree



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

Diagnostic accuracy cross-sectional study design



P Bossuyt http://srdta.cochrane.org/presentations

2x2 Table - sensitivity and specificity

		Disease (Reference standard) +	Disease (Reference standard) -	
Index test	+	True Positive	False Positive	TP+FP
	-	False Negative	True Negative	FN+TN
		TP+FN	FP+TN	TP+FP+ FN+TN
		Sensitivity TP/(TP+FN)	Specificity TN/(TN+FP)	

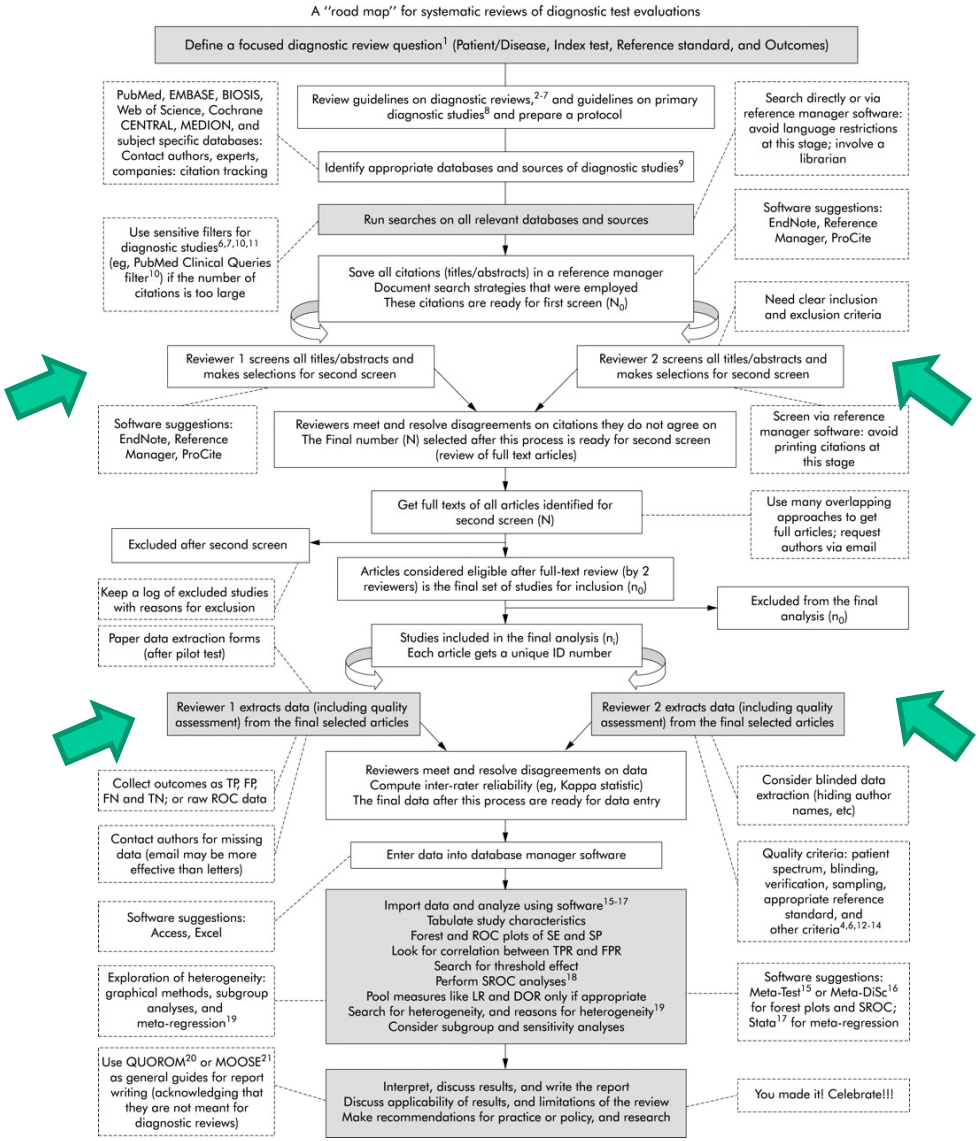
Strengths of diagnostic test accuracy

- Easy to do
- Feasible sample sizes
- Answers can be obtained quickly
- Results do not depend on factors such as treatment decisions, adherence, delivery of services

Limitations of diagnostic test accuracy

- Do not directly assess effect of the tests on patient outcomes
- Do not answer whether testing does more good than harm
- Only possible with adequate reference standard

Road map for diagnostic accuracy reviews



Two review authors work independently to screen citations, assess quality, and extract data

Pai M et al. Systematic reviews of diagnostic test evaluations: what's behind the scenes? Evid Based Med 2004;9:101-103

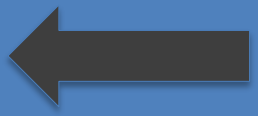
Key steps in a diagnostic test accuracy review

1. Framing focused questions
2. Searching for studies
3. Assessing study quality
4. Analyzing the data; undertaking meta-analyses
5. Interpreting and presenting results

1. Framing focused questions

Begin with a well-framed question - 1

- **P**articipants
- **I**ndex test(s)
- **(C**omparator test)
- (Outcomes, usually sensitivity and specificity)



PI(C,O)

Begin with a well-framed question - 2

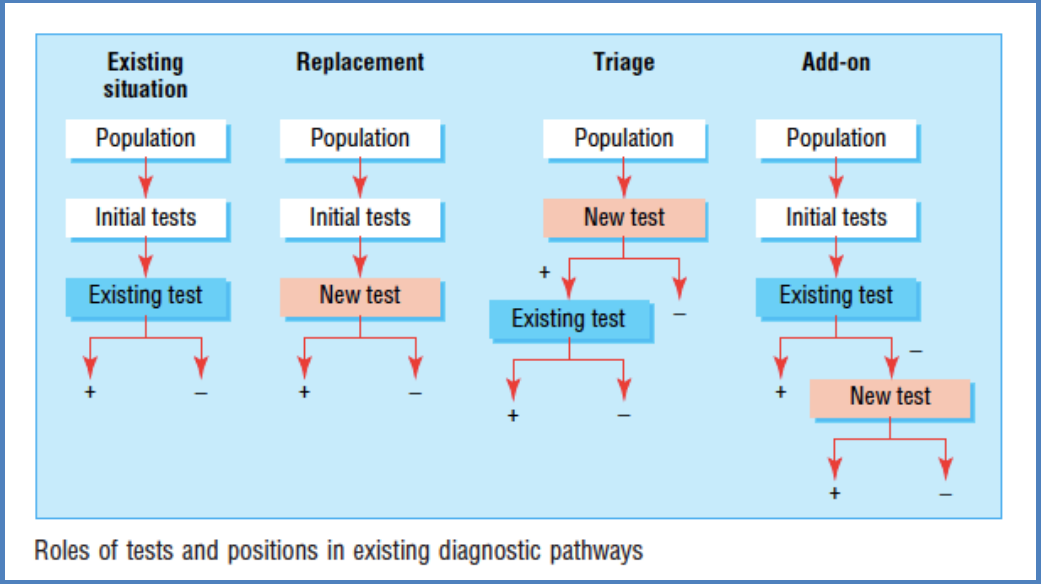
Clinical pathway

- Prior tests
  - Presentation
  - Setting
- Role of tests
  - Replacement
  - Triage
  - Add-on
- Alternative tests

Role of tests

- Replacement: new test replaces existing test
- Triage: new test determines existing test
- Add-on: new test combined with existing test

Bossuyt et al. BMJ 2006



Roles of tests and positions in existing diagnostic pathways

(PICO) PPPIRTR for systematic review of diagnostic test accuracy

- **Participants**, **P**resentation, **P**rior tests
- **I**ndex test(s)
- **Role** of testing
- **T**arget condition
- **R**eference standard

Primary objective

- To assess the diagnostic accuracy of rapid diagnostic tests for detecting clinical *P. falciparum* malaria… in persons living in malaria endemic areas who present to ambulatory healthcare facilities with symptoms of malaria. Abba et al. The Cochrane Library 2012

- To assess the diagnostic accuracy of Xpert® MTB/RIF for pulmonary TB (TB detection) in adults thought to have TB, where Xpert® MTB/RIF was used as an initial test replacing microscopy or an add-on test following a negative smear microscopy result. Steingart et al. The Cochrane Library 2014

2. Searching for studies

Recommended standards for finding and assessing individual studies

Standard 3.1 Conduct a comprehensive systematic search for evidence

Required elements:

3.1.1 Work with a librarian or other information specialist trained in performing systematic reviews to plan the search strategy

3.1.2 Design the search strategy to address each key research question

3.1.3 Use an independent librarian or other information specialist to peer review the search strategy

3.1.4 Search bibliographic databases

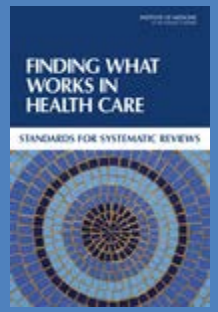
3.1.5 Search citation indexes

3.1.6 Search literature cited by eligible studies

3.1.7 Update the search at intervals appropriate to the pace of generation of new information for the research question being addressed

3.1.8 Search subject-specific databases if other databases are unlikely to provide all relevant evidence

3.1.9 Search regional bibliographic databases if other databases are unlikely to provide all relevant evidence



Searching for diagnostic studies - 1

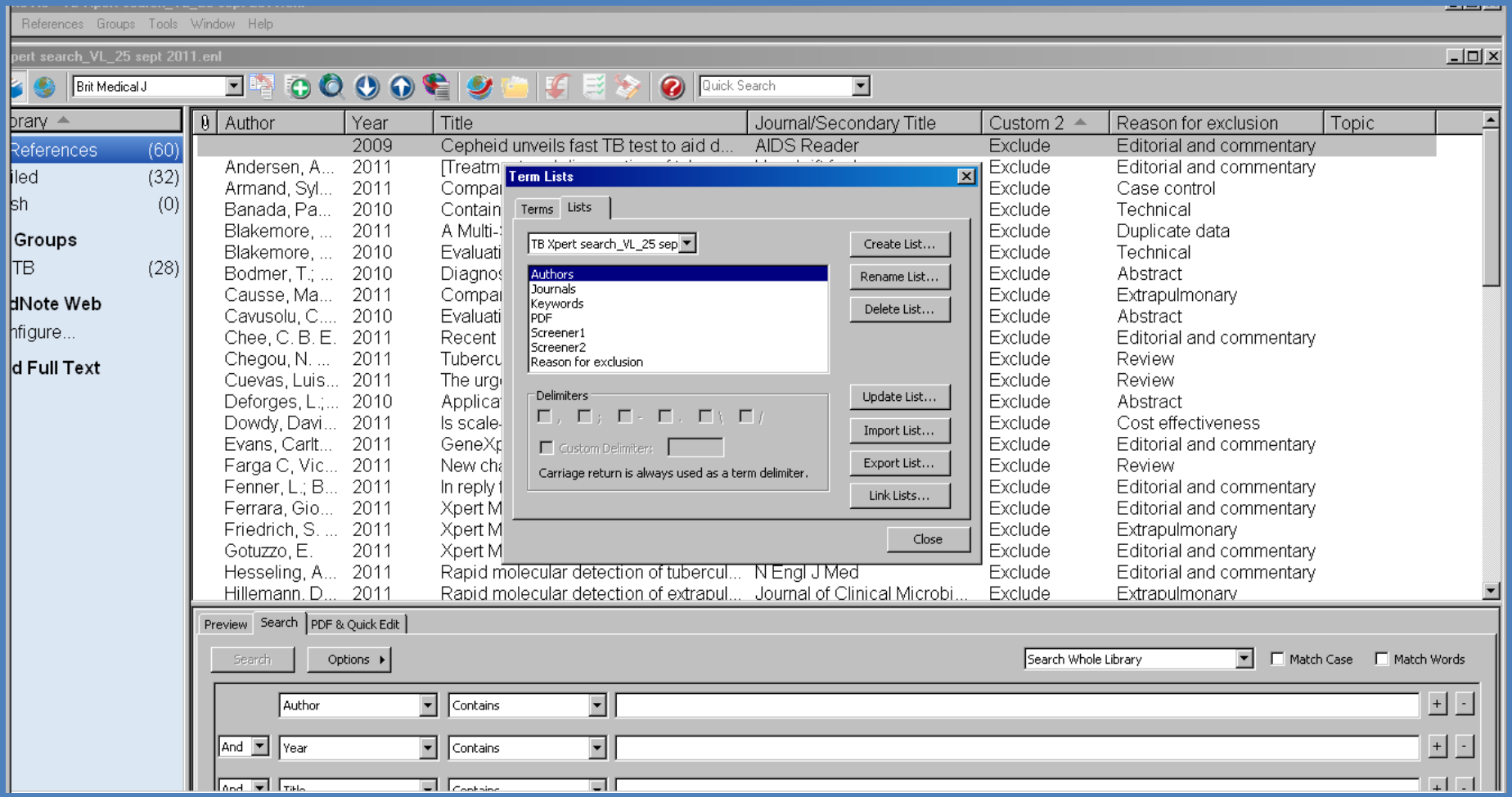
- The Cochrane Library
- The Cochrane Register of Diagnostic Test Accuracy Studies (under development)
- MEDLINE
- EMBASE
- ISI Web of Knowledge
- LILACS
- BIOSIS
- SCOPUS

- Specify day, month, and year of search
- Perform search without language or date restriction

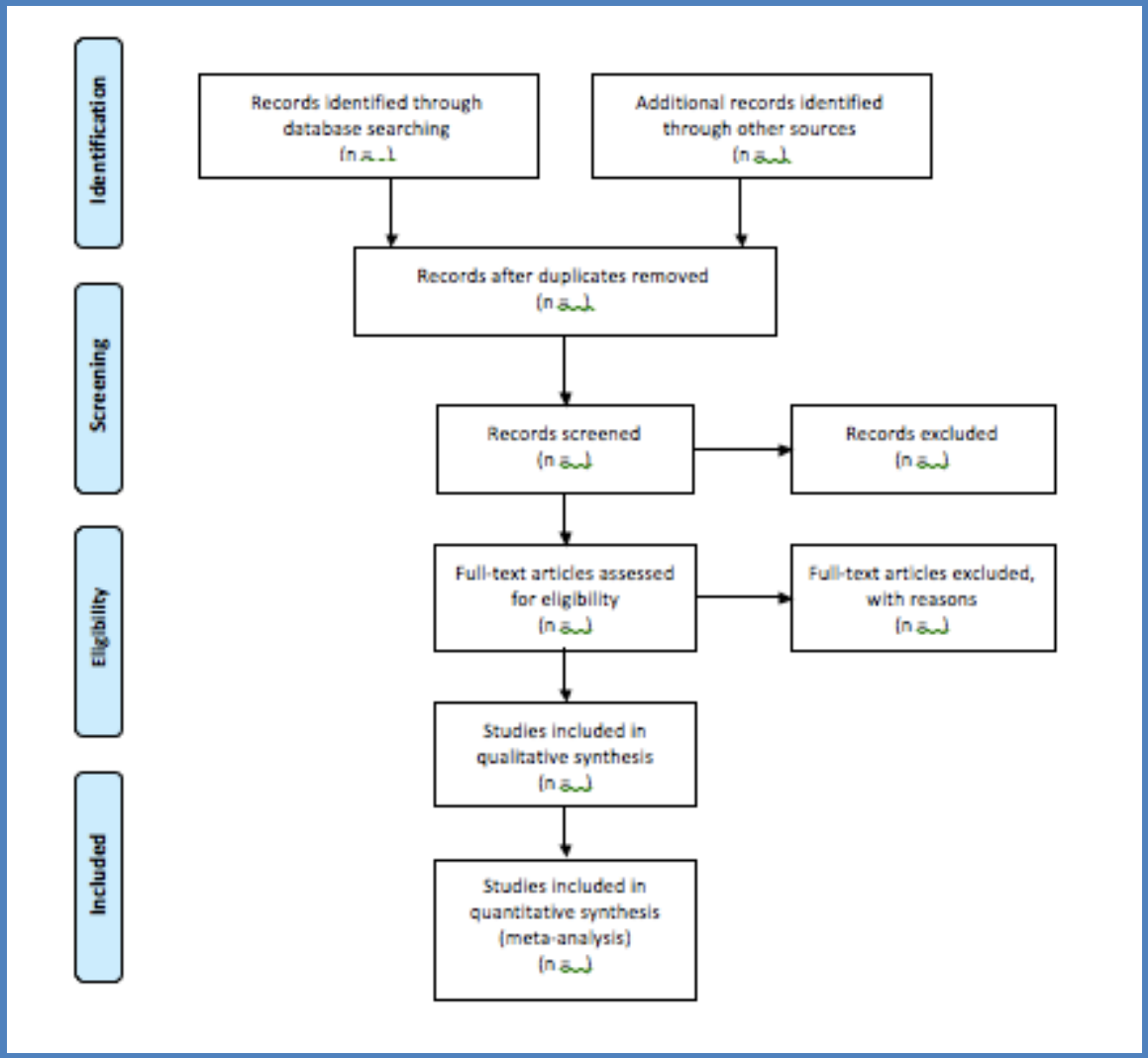
Searching for diagnostic studies - 2

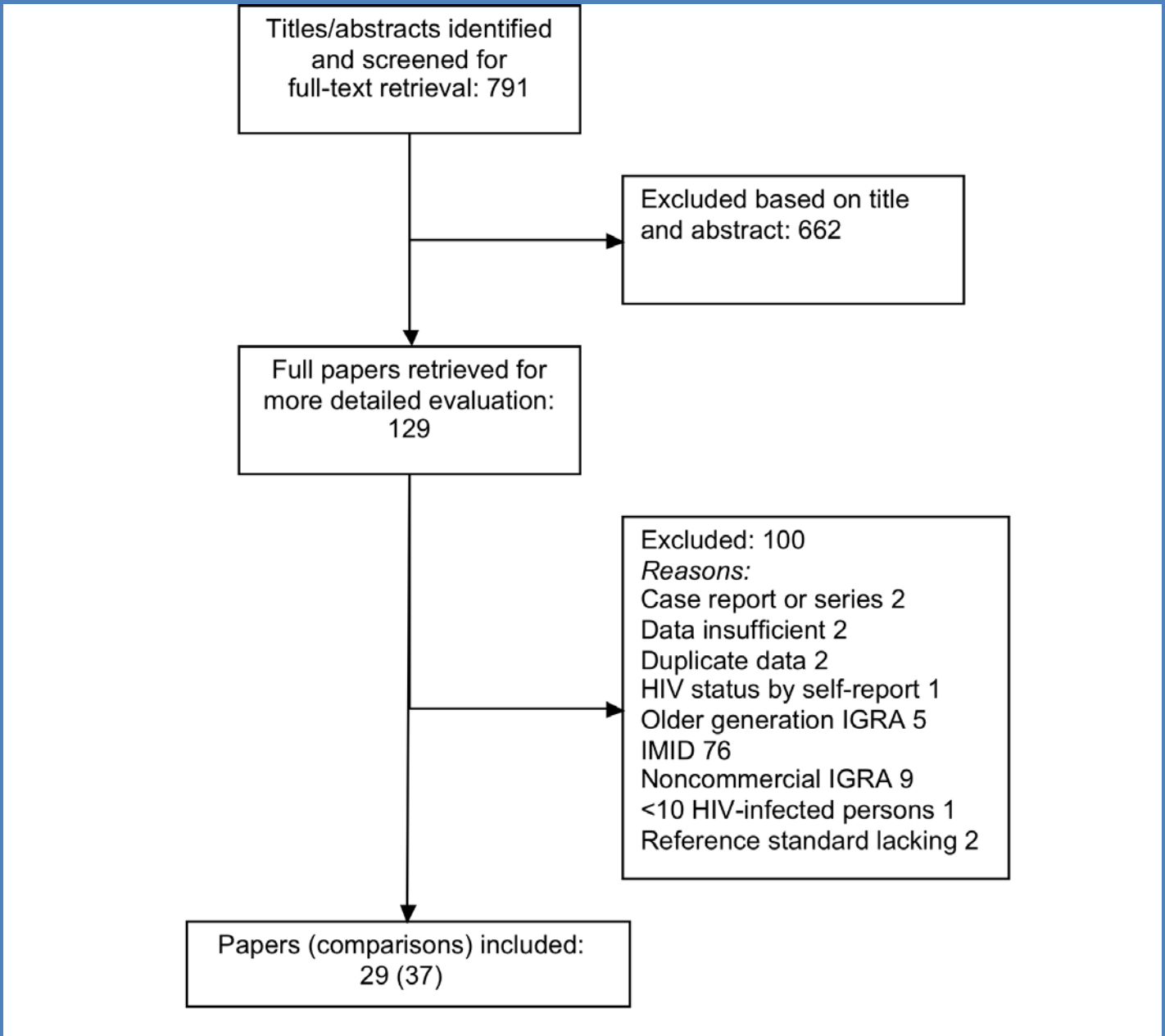
- Use a broad search strategy
- Use a wide variety of search terms, both text words and database subject headings (MeSH terms)
- Avoid use of search filters

Managing citations



PRISMA Flow Diagram
prisma-statement.org





Supplemental Figure 2. Flow of studies, Cattamanchi et al. JAIDS, 2011

The medical literature can be compared to a jungle. It is fast growing, full of deadwood, sprinkled with hidden treasure and infested with spiders and snakes. Morgan. Can Med Assoc J, 134,Jan 15, 1986

3. Assessing study quality

Definition of quality in systematic reviews of diagnostic test accuracy

Methodological quality of a study is the degree to which the design and conduct of a study match the study objectives

Two components

- Risk of bias
- Applicability

NATURE | COLUMN: WORLD VIEW

Beware the creeping cracks of bias

Evidence is mounting that research is riddled with systematic errors. Left unchecked, this could erode public trust, warns Daniel Sarewitz.

09 May 2012

Alarming cracks are starting to penetrate deep into the scientific edifice. They threaten the status of science and its value to society. And they cannot be blamed on the usual suspects — inadequate funding, misconduct, political interference, an illiterate public. Their cause is bias, and the threat they pose goes to the heart of research.

Nothing will corrode public trust more than a creeping awareness that scientists are unable to live up to the standards that they have set for themselves. Useful steps to deal with this threat may range from reducing the hype from universities and journals about specific projects, to strengthening collaborations between those involved in fundamental research and those who will put the results to use in the real world.

What is bias?

Bias is **any process** at **any stage** of inference tending to produce **results** that **differ systematically** from the **true values**. *Murphy EA. The logic of medicine 1976*

Bias is **any trend** in the **collection, analysis, interpretation, publication or review** of data that can lead to **conclusions** that are **systematically different** from the **truth**. *Last J. A dictionary of epidemiology 2001*

This is the tendency of some (poor) study designs **systematically** to produce **results that are better** (rarely if ever worse) than those with a robust design. *Bandolier*

A Catalogue of Bias, M. Tevfik Dorak (adapted from David Sackett)
http://www.dorak.info/epi/bc.html

Literature Review

- Foreign language exclusion bias
- Literature search bias
- One-sided reference bias
- Rhetoric bias

Study Design

- Selection bias
- Sampling frame bias
  - Berkson (admission rate) bias
  - Centripetal bias
  - Diagnostic access bias
  - Diagnostic purity bias
  - Hospital access bias
  - Migrator bias
  - Prevalence-incidence (Neyman / selective survival; attrition) bias
  - Telephone sampling bias
- Nonrandom sampling bias
  - Autopsy series bias
  - Detection bias
  - Diagnostic work-up bias
  - Door-to-door solicitation bias
  - Previous opinion bias
  - Referral filter bias
  - Sampling bias
  - Self-selection bias
  - Unmasking bias
- Noncoverage bias
  - Early-comer bias
  - Illegal immigrant bias
  - Loss to follow-up (attrition) bias
  - Response bias
  - Withdrawal bias
- Noncomparability bias
  - Ecological (aggregation) bias
  - Healthy worker effect (HWE)
  - Lead-time bias
  - Length bias
  - Membership bias
  - Mimicry bias
  - Nonsimultaneous comparison bias
  - Sample size bias

Study Execution

- Bogus control bias
- Contamination bias
- Compliance bias

Data Collection

- Instrument bias
  - Case definition bias
  - Diagnostic vogue bias
  - Forced choice bias
  - Framing bias
  - Insensitive measure bias
  - Juxtaposed scale bias
  - Laboratory data bias
  - Questionnaire bias
  - Scale format bias
  - Sensitive question bias
  - Stage bias
  - Unacceptability bias
  - Underlying/contributing cause of death bias
  - Voluntary reporting bias
- Data source bias
  - Competing death bias
  - Family history bias
  - Hospital discharge bias
  - Spatial bias
- Observer bias
  - Diagnostic suspicion bias
  - Exposure suspicion bias
  - Expectation bias
  - Interviewer bias
  - Therapeutic personality bias
- Subject bias
  - Apprehension bias
  - Attention bias (Hawthorne effect)
  - Culture bias
  - End-aversion bias (end-of-scale or central tendency bias)
  - Faking bad bias
  - Faking good bias
  - Family information bias
  - Interview setting bias
  - Obsequiousness bias
  - Positive satisfaction bias
  - Proxy respondent bias
- Recall bias
  - Reporting bias
  - Response fatigue bias
  - Unacceptable disease bias
  - Unacceptable exposure bias
  - Underlying cause (rumination bias)
  - Yes-saying bias
- Data handling bias
  - Data capture error
  - Data entry bias
  - Data merging error
  - Digit preference bias
  - Record linkage bias

Analysis

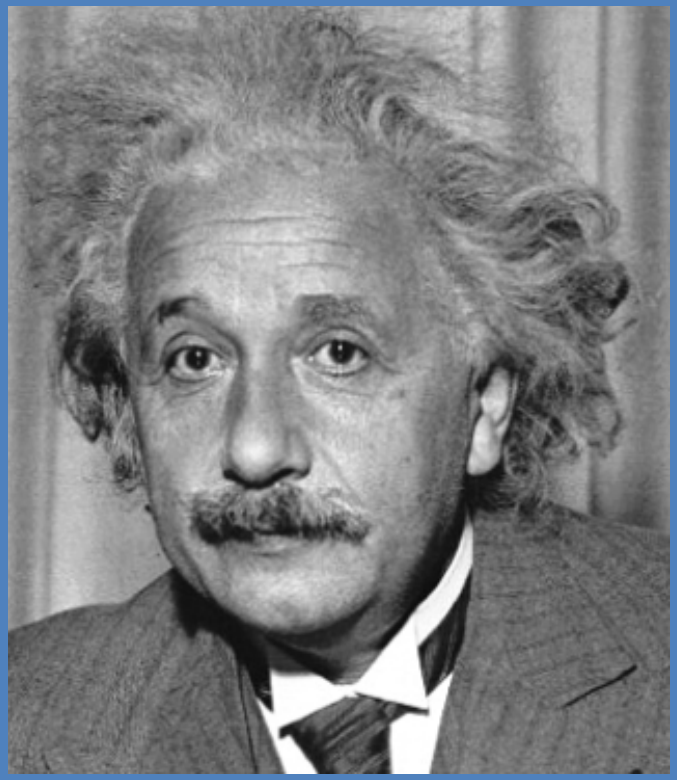
- Confounding bias
  - Latency bias
  - Multiple exposure bias
  - Nonrandom sampling bias
  - Standard population bias
  - Spectrum bias
- Analysis strategy bias
  - Distribution assumption bias
  - Enquiry unit bias
  - Estimator bias
  - Missing data handling bias
  - Outlier handling bias
  - Overmatching bias
  - Scale degradation bias
- Post hoc analysis bias
  - Data dredging bias
  - Post hoc significance bias
  - Repeated peeks bias

Interpretation of Results

- Assumption bias
- Cognitive dissonance bias
- Correlation bias
- Generalization bias
- Magnitude bias
- Significance bias
- Underexhaustion bias

Publication

- All's well literature bias
- Positive result bias
- Hot topic bias



“Everything should be made as simple as possible but not simpler.”

Three key sources of bias in diagnostic studies

1. Inclusion of right spectrum of patients
2. Verification of patients
    - choice of reference standard
    - complete verification
3. Independent interpretation of index test and reference standard results (blinding)

1. Inclusion of right spectrum of patients

- Ideally, included patients should resemble those in whom the test will be used in practice

- Spectrum bias = study uses selected patients, *perhaps those whom you suspect have the disease*

Ann Intern Med. 1992 Jul 15;117(2):135-40.

Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection.

Lachs MS[1], Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS.

⊕ Author information

Abstract

OBJECTIVE: To determine if the leukocyte esterase and bacterial nitrite rapid dipstick test for urinary tract infection (UTI) is susceptible to spectrum bias (when a diagnostic test has different sensitivities or specificities in patients with different clinical manifestations of the disease for which the test is intended).

DESIGN: Cross-sectional study.

PATIENTS: A total of 366 consecutive adult patients in whom clinicians performed urinalysis to diagnose or exclude UTI.

SETTING: An urban emergency department and walk-in clinic.

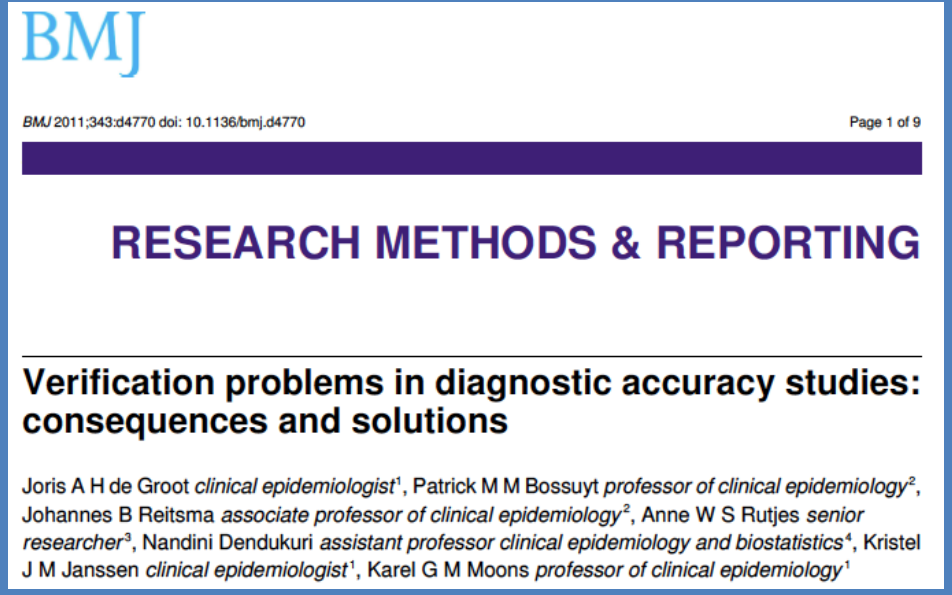
MEASUREMENTS: After the patient encounter, but before dipstick test or culture was done, clinicians recorded the signs and symptoms that were the basis for suspecting UTI and for performing a urinalysis and an estimate of the probability of UTI based on the clinical evaluation. For all patients who received urinalysis, dipstick tests and culture were done in the clinical microbiology laboratory by medical technologists blinded to clinical evaluation. Sensitivity for the dipstick was calculated using a positive result in either leukocyte esterase or bacterial nitrite, or both, as the criterion for a positive dipstick, and greater than 10(5) CFU/mL for a positive culture.

RESULTS: In the 107 patients with a high (greater than 50%) prior probability of UTI, who had many characteristic UTI symptoms, the sensitivity of the test was excellent (0.92; 95% CI, 0.82 to 0.98). In the 259 patients with a low (less than or equal to 50%) prior probability of UTI, the sensitivity of the test was poor (0.56; CI, 0.03 to 0.79).

CONCLUSIONS: The leukocyte esterase and bacterial nitrite dipstick test for UTI is susceptible to spectrum bias, which may be responsible for differences in the test's sensitivity reported in previous studies. As a more general principle, diagnostic tests may have different sensitivities or specificities in different parts of the clinical spectrum of the disease they purport to identify or exclude, but studies evaluating such tests rarely report sensitivity and specificity in subgroups defined by clinical symptoms. When diagnostic tests are evaluated, information about symptoms in the patients recruited for study should be included, and analyses should be done within appropriate clinical subgroups so that clinicians may decide if reported sensitivities and specificities are applicable to their patients.

2. Verification of patients

- Ideally, the appropriate reference standard is used
- Ideally, all patients receive the reference standard
- Verification bias = only some patients receive the reference standard, *perhaps those whom you suspect have the disease*



BMJ

BMJ 2011;343:d4770 doi: 10.1136/bmj.d4770

RESEARCH METHODS & REPORTING

Verification problems in diagnostic accuracy studies: consequences and solutions

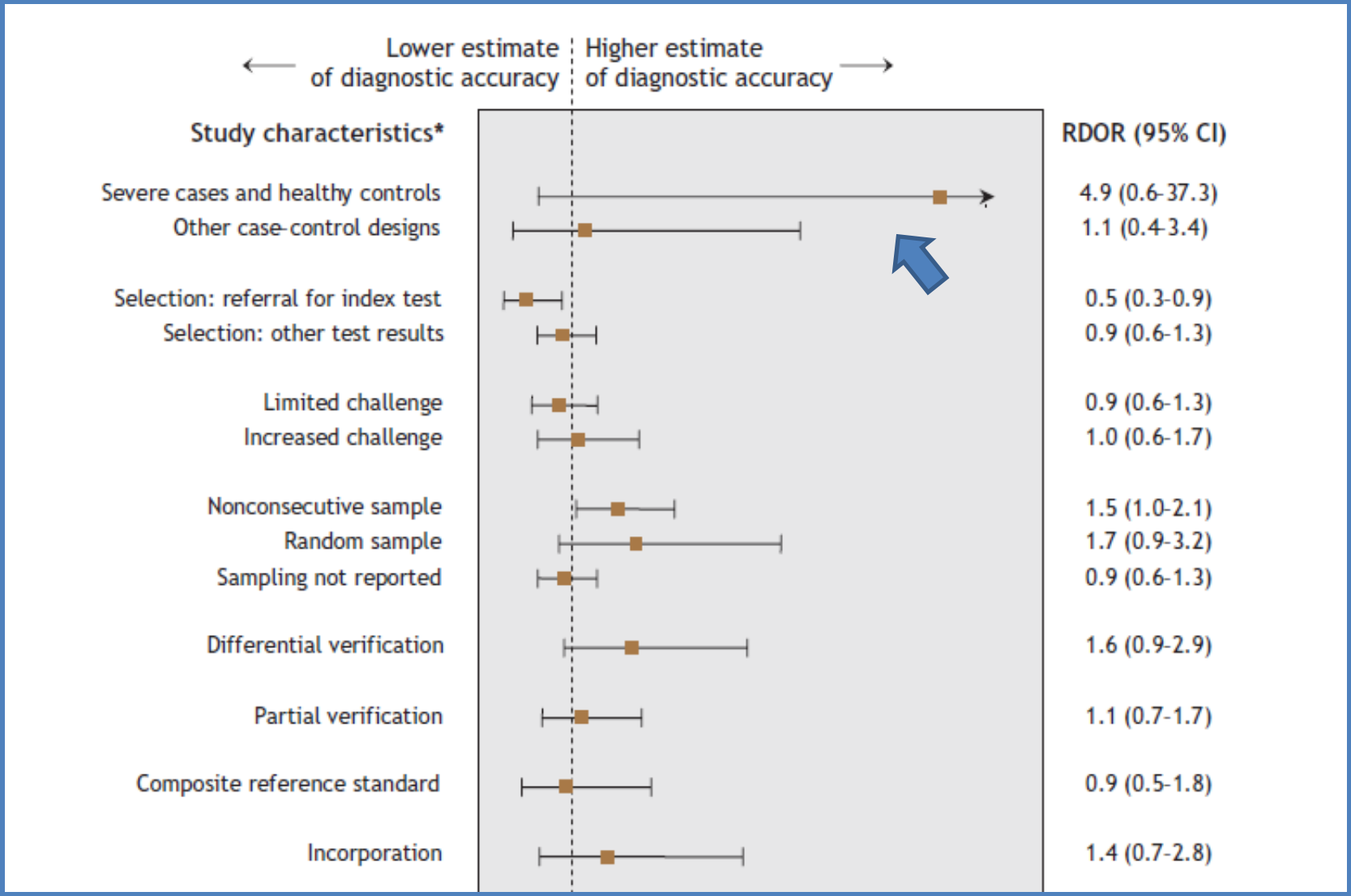
Joris A H de Groot *clinical epidemiologist*[1], Patrick M M Bossuyt *professor of clinical epidemiology*[2], Johannes B Reitsma *associate professor of clinical epidemiology*[2], Anne W S Rutjes *senior researcher*[3], Nandini Dendukuri *assistant professor clinical epidemiology and biostatistics*[4], Kristel J M Janssen *clinical epidemiologist*[1], Karel G M Moons *professor of clinical epidemiology*[1]

"In a study evaluating the accuracy of digital rectal examination and prostate specific antigen for the early detection of prostate cancer, 145 out of 1000 men fulfilled the criterion for verification by the reference standard (transrectal ultrasound combined with biopsy). However, 54 of these men did not undergo the reference standard, for unknown reasons…."

3. Blinding

- Ideally the results of the index test are interpreted without knowing the results of the reference standard and vice versa
- Test review bias = interpretation of the results of the index test is influenced by knowledge of the results of the reference standard
- Diagnostic review bias = interpretation of the results of the reference standard is influenced by knowledge of the results of the index test

Which type of bias matters the most?



Effects of study design on estimates of diagnostic accuracy. Rutjes CMAJ 2006

Applicability relates to the extent to which the primary study was relevant to the review

- Compared with the review question…

- the patient group had similar demographic or clinical features
- the index test was applied or interpreted in a similar manner
- the definition of the target condition was similar

Quality Assessment of Diagnostic Accuracy Studies, QUADAS-2

- Domain list
  - patient selection
  - index test
  - reference standard
  - flow and timing
- Signalling questions are used for judgments of risk of bias
- First 3 domains are also assessed for applicability

Whiting et al. Annals Internal Medicine. 2011

Patient selection - Risk of Bias

Describe methods of patient selection and included patients (prior testing, presentation, role of index test, and setting)

- Q1:Was a consecutive or random sample of patients enrolled?
- Q2: Was a case-control study design avoided?
- Q3: Did the study avoid inappropriate exclusions?

Signalling questions: Yes, No, Unclear

Risk of Bias: Yes, No, Unclear

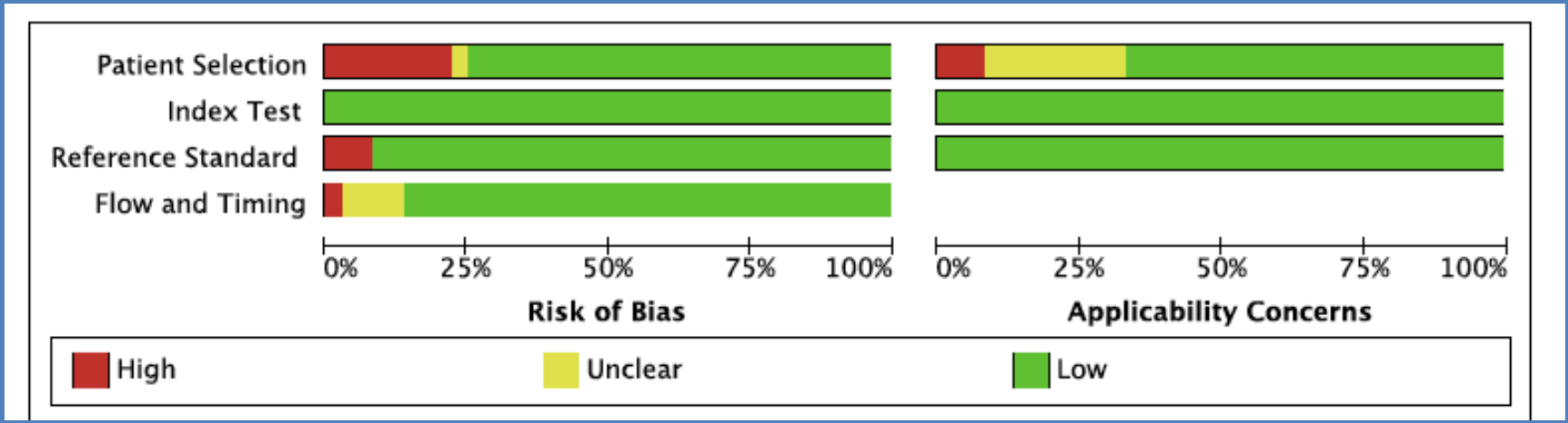
Patient selection - Applicability

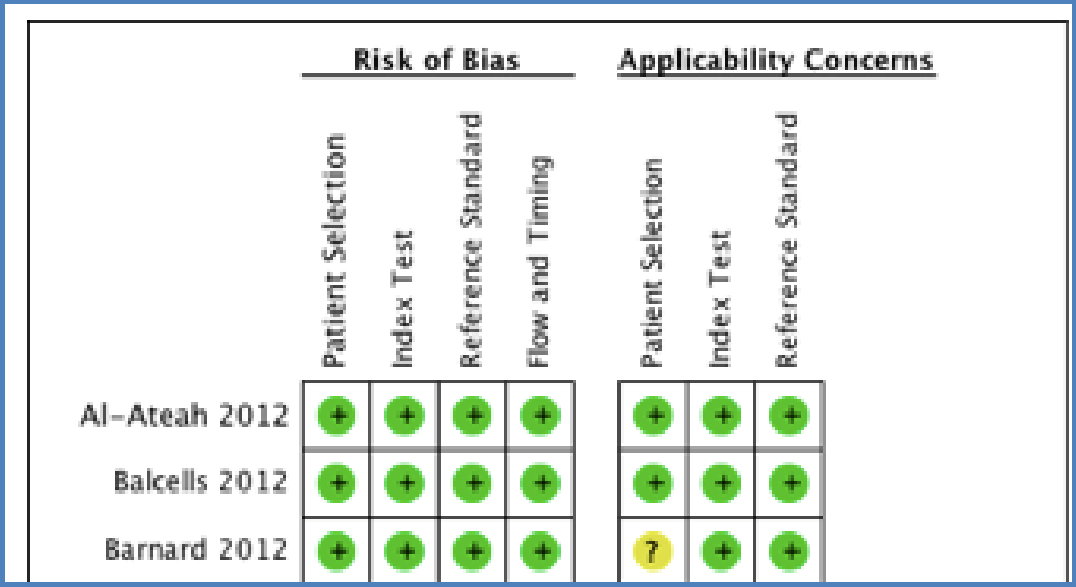
Describe methods of patient selection and included patients (prior testing, presentation, role of index test, and setting)

- Do the included patients and setting match the question?

Applicability: High, Low, or Unclear

QUADAS-2 graphs





Steingart et al. The Cochrane Library 2014

Quality assessment - limitations

- Information about quality is often not reported
- Quality assessment is subjective
- No summary scores and cut-offs for ‘high’ quality
- Statistical incorporation of quality may not be possible with a limited number of studies

STARD: STAndards for the Reporting of Diagnostic accuracy studies

Clinical Chemistry 49:1
1–6 (2003)

STARD Initiative

Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy: The STARD Initiative

Patrick M. Bossuyt,[1*] Johannes B. Reitsma,[1] David E. Bruns,[2,3]
Constantine A. Gatsonis,[4] Paul P. Glasziou,[5] Les M. Irwig,[6] Jeroen G. Lijmer,[1]
David Moher,[7] Drummond Rennie,[8,9] and Henrica C.W. de Vet,[10] for the STARD Group

Background: To comprehend the results of diagnostic accuracy studies, readers must understand the design, conduct, analysis, and results of such studies. That goal can be achieved only through complete transparency from authors.

Results: The search for published guidelines on diagnostic research yielded 33 previously published checklists, from which we extracted a list of 75 potential items. The consensus meeting shortened the list to 25 items, using evidence on bias whenever available. A prototyp-

STARD checklist for the reporting of studies of diagnostic accuracy.

First official version, January 2003.

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	
METHODS			
Participants	3	Describe the study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	
	4	Describe participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	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.	
	6	Describe data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	Describe the reference standard and its rationale.	
	8	Describe technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	

www.stard-statement.org

Use of STARD in TB diagnostic studies

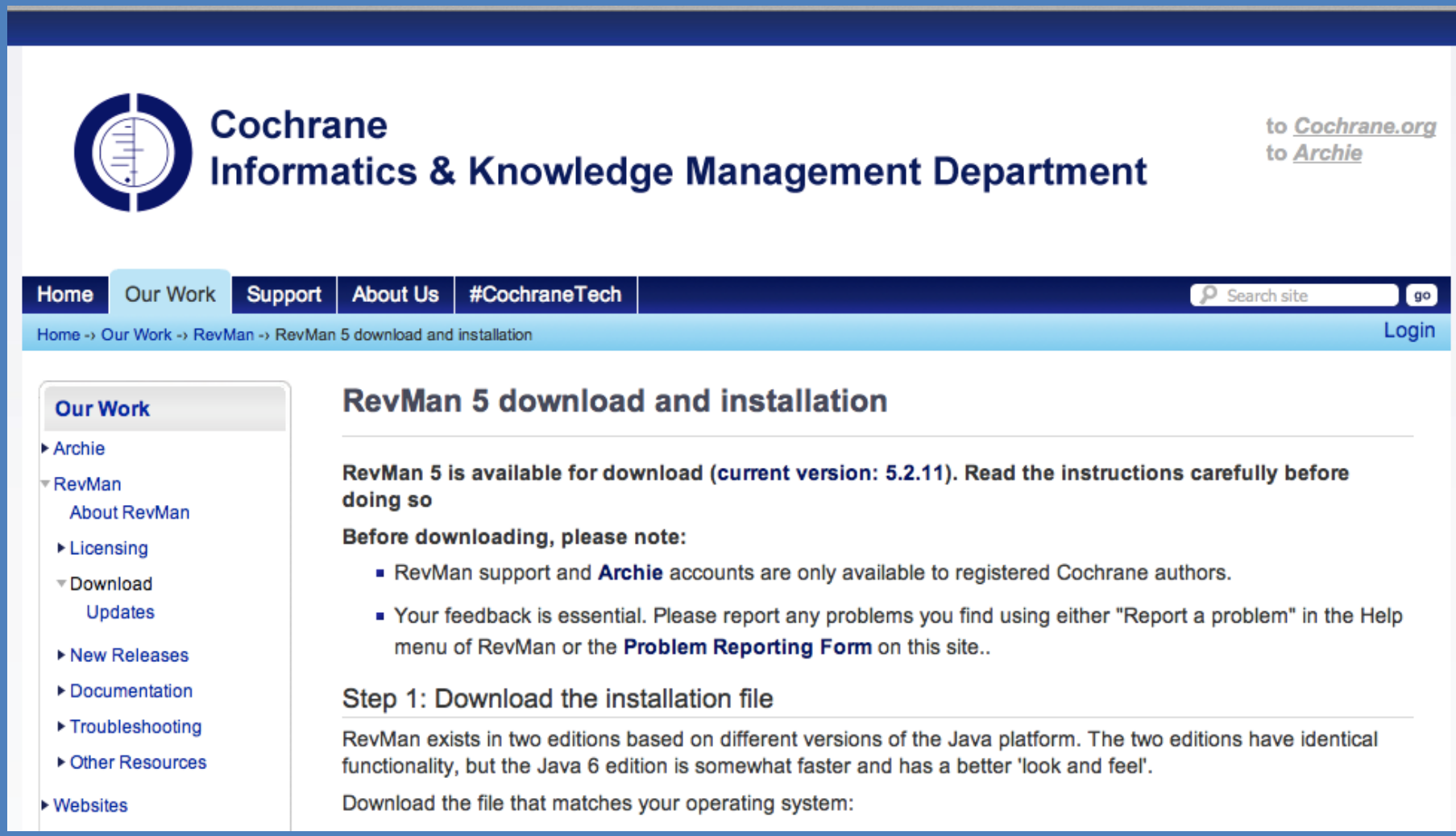
- “To evaluate the MTBDR*sl*, we compared the LiPA performance to DST performed by MGIT and to direct sequencing as reference standards following the STARD recommendations.” Miotto et al. ERJ 2012

- “This study conforms to the STARD initiative guidelines …for reporting of studies of diagnostic accuracy.” Lawn et al. PLoS Med 2011

4. Analyzing the data; undertaking meta-analyses

Steps in data analysis

- Extract true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values
- Calculate paired estimates of sensitivity and specificity
- Visually examine results of individual studies
- Calculate pooled accuracy estimates using recommended methods for meta-analysis
- Look for and investigate possible reasons for heterogeneity



ims.cochrane.org/revman

Extract true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values
Calculate paired estimates of sensitivity and specificity
Visually examine results of individual studies

HIV positive

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Balcells 2012	11	1	1	47	0.92 [0.62, 1.00]	0.98 [0.89, 1.00]
Boehme 2010a	7	0	0	2	1.00 [0.59, 1.00]	1.00 [0.16, 1.00]
Boehme 2010b	0	0	1	1	0.00 [0.00, 0.97]	1.00 [0.03, 1.00]
Boehme 2010c	60	0	6	81	0.91 [0.81, 0.97]	1.00 [0.96, 1.00]
Boehme 2010d	27	2	6	141	0.82 [0.65, 0.93]	0.99 [0.95, 1.00]

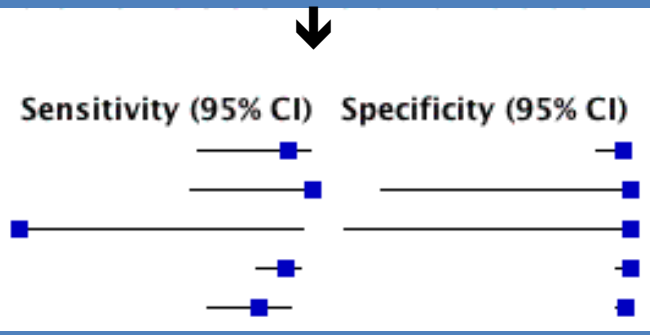
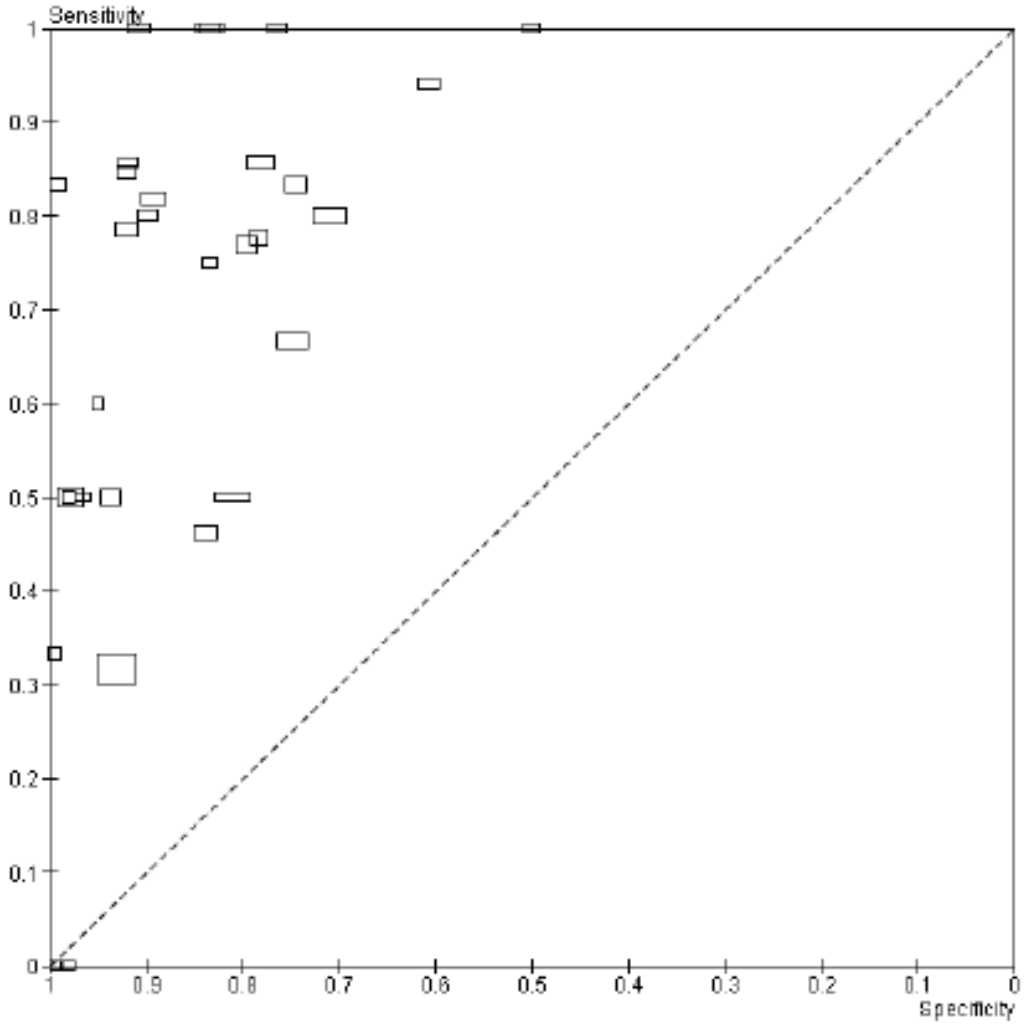


Figure 5. Plot of sensitivity versus specificity for all 30 studies, irrespective of cut-off value. The width of the blocks is proportional to the inverse standard error of the specificity in every study and the height of the blocks is proportional to the inverse standard error of the sensitivity.

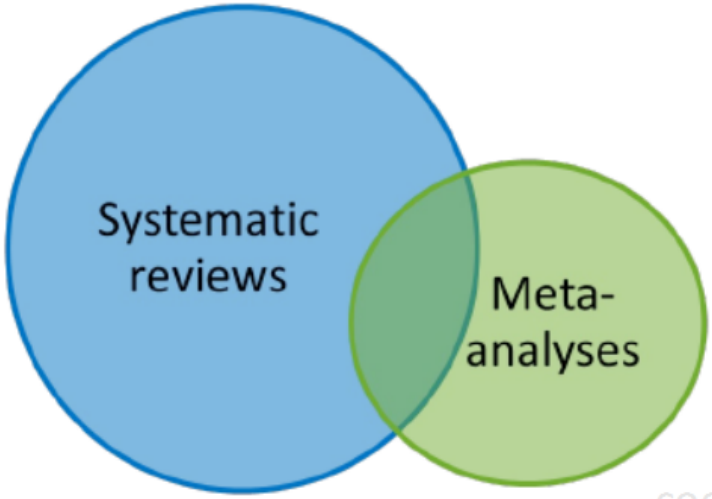


Leeflang. Cochrane Library 2009

Calculate pooled accuracy estimates using recommended methods for meta-analysis

What is a meta-analysis?

- combines the results from two or more studies
- estimates an 'average' or 'common' effect
- optional part of a systematic review



Source: Julian Higgins

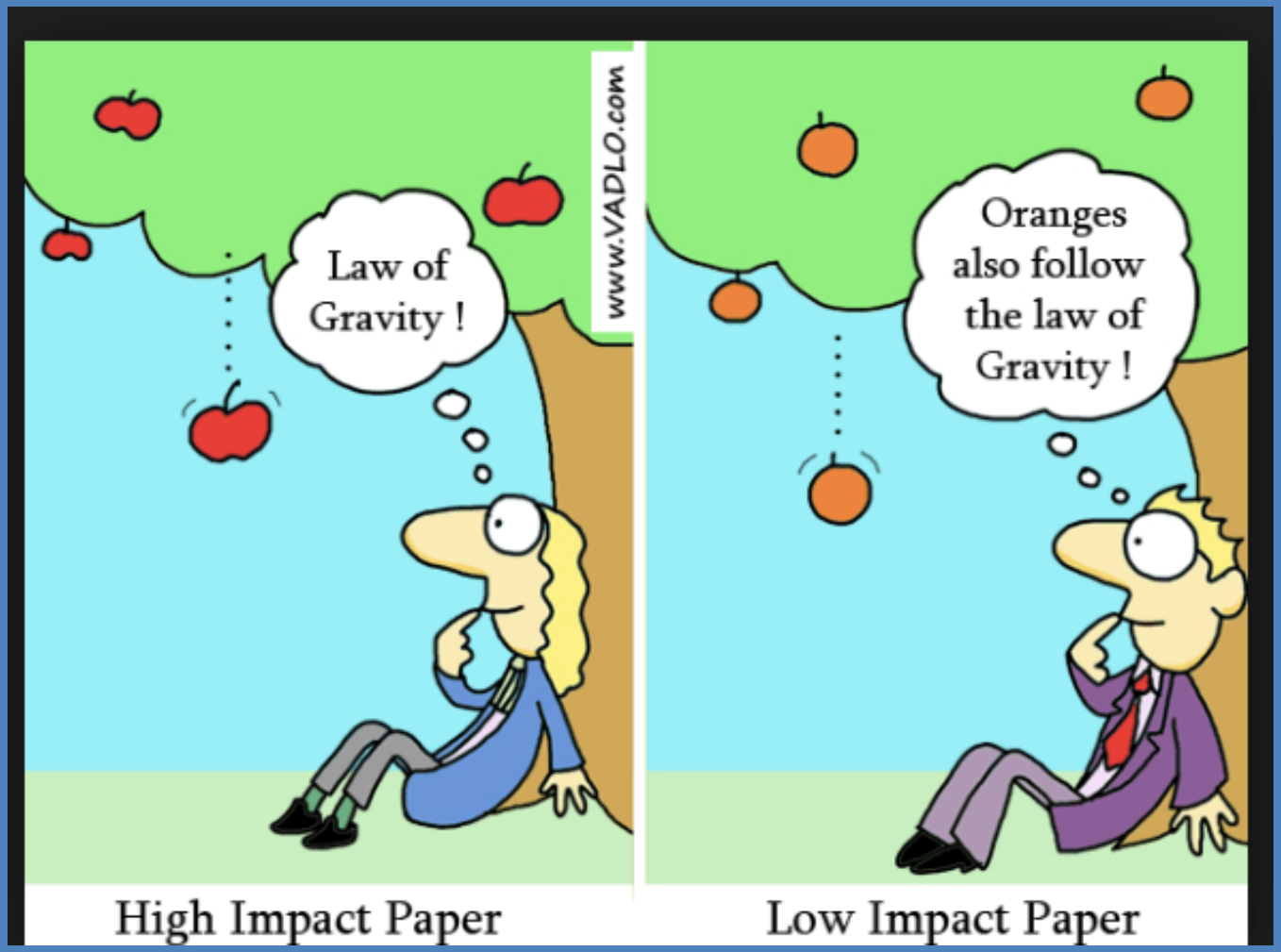
Meta-analysis of diagnostic test accuracy - 1

- Calculate the diagnostic accuracy of a test
- Compare the diagnostic accuracy of two or more tests
- Investigate the variability of results between studies (heterogeneity)

Meta-analysis of diagnostic test accuracy - 2

- Methods have been devised to account for two summary statistics (sensitivity and specificity) simultaneously rather than one
- Heterogeneity is to be expected in meta-analyses of diagnostic test accuracy
- No equivalent to the I^2 statistic is currently available for meta-analysis of diagnostic test accuracy

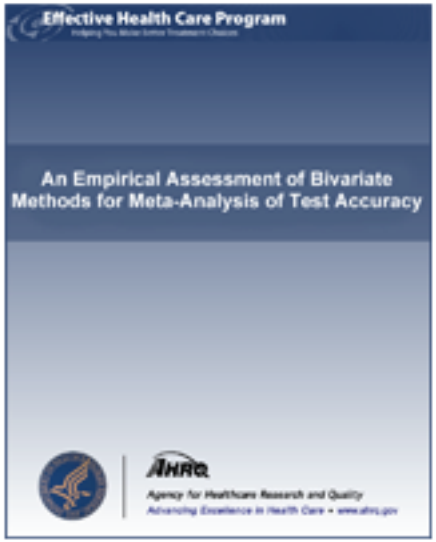
A common criticism of meta-analysis is that it combines apples with oranges by mixing together estimates from studies that differ in important ways



When should you NOT do a meta-analysis?

- When studies differ in terms of the participants they recruit and the tests that they evaluate
- When studies do not address the review question

How is a meta-analysis performed -1?



An Empirical Assessment of Bivariate Methods for Meta-Analysis of Test Accuracy

Methods Research Reports

Investigators: Issa J Dahabreh, MD, MS,[*] Thomas A Trikalinos, MD,[*] Joseph Lau, MD, and Christopher Schmid, PhD[*].

Tufts Evidence-based Practice Center, Tufts Medical Center

Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Nov.
Report No.: 12(13)-EHC136-EF

"In our empirical comparison, bivariate meta-analyses produced point estimates that were largely similar to those of separate univariate analyses….however, bivariate models have stronger theoretical motivation for most common diagnostic test meta-analysis scenarios."
ncbi.nlm.nih.gov/books/NBK115736/

How is a meta-analysis performed - 2?

- The bivariate model
  - gives average sensitivity and specificity
  - use when studies report a *common threshold* for a positive result
- Hierarchical summary ROC curve
  - gives a summary ROC curve
  - use when studies report *several different thresholds*
- Both models use random effects

Macaskill P et al. Chapter 10. Cochrane Handbook, Diagnostic Test Accuracy, Version 1.0.2010.

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
  - unexplained

Variation due to threshold differences

- Explicit threshold differences
  – studies may use different values to define positive test results
- Implicit threshold differences
  – differences in observers
  – differences in equipment

Investigating heterogeneity

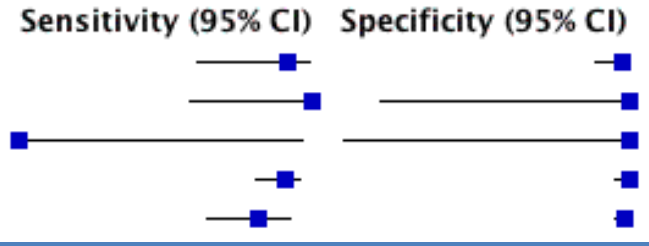
- Visual inspection
- Subgroup analysis
- Meta-regression

Visual inspection, do the confidence intervals overlap?

A

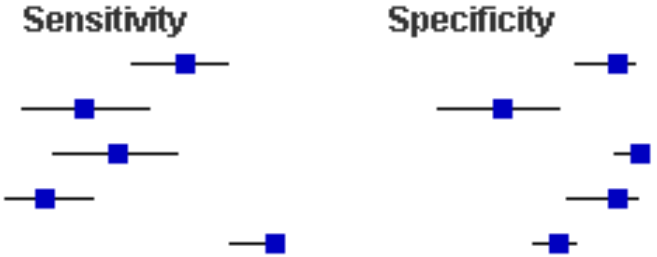
HIV positive

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Balcells 2012	11	1	1	47	0.92 [0.62, 1.00]	0.98 [0.89, 1.00]
Boehme 2010a	7	0	0	2	1.00 [0.59, 1.00]	1.00 [0.16, 1.00]
Boehme 2010b	0	0	1	1	0.00 [0.00, 0.97]	1.00 [0.03, 1.00]
Boehme 2010c	60	0	6	81	0.91 [0.81, 0.97]	1.00 [0.96, 1.00]
Boehme 2010d	27	2	6	141	0.82 [0.65, 0.93]	0.99 [0.95, 1.00]



B

Study	TP	FP	FN	TN	Sensitivity	Specificity
Alifano 1998b	31	3	11	41	0.74 [0.58, 0.86]	0.93 [0.81, 0.99]
Banerjee 2003a	13	13	17	19	0.43 [0.25, 0.63]	0.59 [0.41, 0.76]
Caminero 1993	16	0	14	48	0.53 [0.34, 0.72]	1.00 [0.93, 1.00]
Caminero 1994	18	2	38	29	0.32 [0.20, 0.46]	0.94 [0.79, 0.99]
Gevaudan 1992b	26	47	0	147	1.00 [0.87, 1.00]	0.76 [0.69, 0.82]



Subgroup analysis

Covariate (Number of studies)	Median pooled sensitivity (95% credible interval)	Median pooled specificity (95% credible interval)
Smear status		
Smear + (21)	98% (97, 99)	***
Smear - (21)	67% (60, 74)	99% (98, 99)
HIV status		
HIV- (7)	86% (76, 92)	99% (98, 100)
HIV+ (7)	79% (70, 86)	98% (96, 99)

Steingart et al. Xpert MTB/RIF. Cochrane Library 2014

Meta-regression

Number of studies (participants)	Sensitivity 95% CI	Specificity 95% CI	Sensitivity 95% CI	Specificity 95% CI	Overall accuracy, P value
	FQ, culture, indirect		***FQ, culture, direct***		
23 (2799)	0.831 0.787,0.867	0.977 0.943,0.991	0.851 0.719,0.927	0.982 0.968,0.990	0.5485
	SLID, culture, indirect		***SLID, culture, direct***		
20 (2584)	0.769 0.611,0.876	0.995 0.971,0.999	0.944 0.252,0.999	0.982 0.889,0.997	0.0121

Theron et al. MTBDR*s*/. submitted

Publication bias

- Publication bias exists when studies included in the review have results that differ systematically from relevant studies that are missed

- Formal assessment of publication bias is usually not performed for diagnostic test accuracy studies

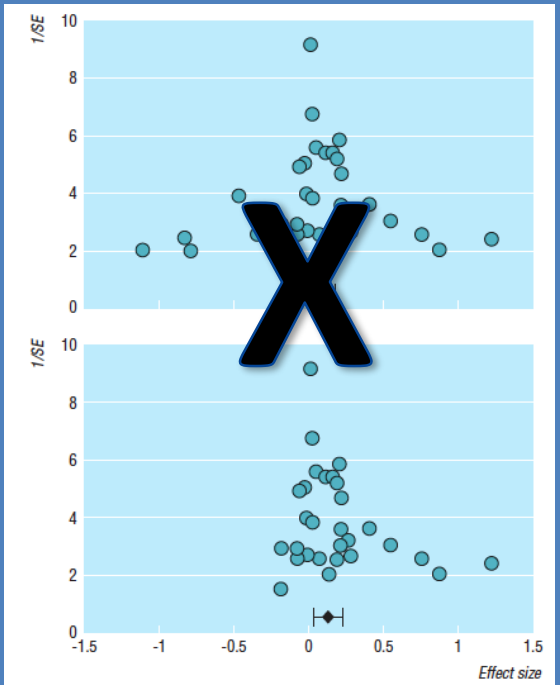


Fig 1 Typical funnel plot generated from 35 simulated studies (top) and same data with five missing studies showing a typical manifestation of publication bias (bottom)

5. Interpreting and presenting results

Discussion Section

- Summary of main results
- Strengths and weaknesses of the review
- Applicability of findings to the review question
- Implications for practice (clinical and policy)
- Implications for research (what and how, avoid saying ‘more research is needed)

Table. Summary of Findings

Abdominal ultrasound for the diagnosis of pancreatic cancer

Patients or population: symptomatic patients in primary care with suspicion of pancreatic cancer
Setting: mainly outpatients
New Test: abdominal ultrasound[1]
Reference Test: endoscopic ultrasound with biopsy[2]
Threshold: Proven or probable pancreatic cancer

Test result	Number of results per 1000 patients tested[3] (95% CI)		Number of participants (studies)	Quality of the evidence (GRADE)
	Prevalence 20 per 1000[4]: Which is typically seen in otherwise healthy adults presenting with symptoms of jaundice, fatigue, pain of the abdomen, and dark urine.	Prevalence 400 per 1000[4]: Which is typically seen in older adults presenting with symptoms of jaundice, fatigue and pain, with a family history of pancreatic cancer, history of chronic pancreatitis, who have diabetes, and are current or past smokers.		
Sensitivity (95% CI): **0.64** (0.50 to 0.77)			2777 (18 studies)	⊕⊕⊕⊕ High[5]
True positives	**13** per 1000 (9 to 15 per 1000)	**282** per 1000 (220 to 339 per 1000)		
False negatives	**7** per 1000 (4 to 10 per 1000)	**158** per 1000 (101 to 220 per 1000)		
Specificity (95% CI): **0.95** (0.91 to 0.97)				
True negatives	**931** per 1000 (901 to 960 per 1000)	**532** per 1000 (510 to 543 per 1000)		
False positives	**49** per 1000 (30 to 89 per 1000)	**28** per 1000 (17 to 50 per 1000)		

CI: Confidence interval

Footnotes:

[1] A diagnostic test for pancreatic cancer needs to be less invasive than the current reference standard and lessen the burden to patients

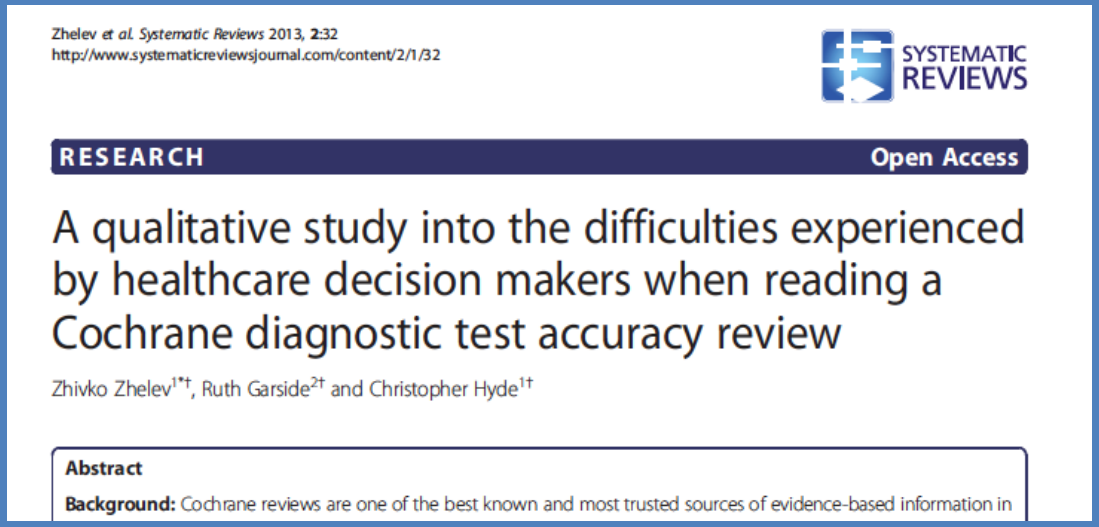
Reem Mustafa, Cochrane Colloquium 2013

Abdominal ultrasound for the diagnosis of pancreatic cancer

Patients or population: symptomatic patients in primary care with suspicion of pancreatic cancer
Setting: mainly outpatients
New Test: abdominal ultrasound[1]
Reference Test: endoscopic ultrasound with biopsy[2]
Threshold: Proven or probable pancreatic cancer

Test result	Number of results per 1000 patients tested[3] (95% CI)		Number of participants (studies)	Quality of the evidence (GRADE)
	Prevalence 20 per 1000[4]: Which is typically seen in otherwise healthy adults presenting with symptoms of jaundice, fatigue, pain of the abdomen, and dark urine.	**Prevalence 400 per 1000[4]:** Which is typically seen in older adults presenting with symptoms of jaundice, fatigue and pain, with a family history of pancreatic cancer, history of chronic pancreatitis, who have diabetes, and are current or past smokers.		
Sensitivity (95% CI): **0.64** (0.50 to 0.77)				
True positives	**13** per 1000 (9 to 15 per 1000)	**282** per 1000 (220 to 339 per 1000)	2777	⊕⊕⊕⊕
False negatives	**7** per 1000 (4 to 10 per 1000)	**158** per 1000 (101 to 220 per 1000)		

- The most common error was to misinterpret the role of the reference standard as if it was a comparator
- Participants could not recall the definitions for sensitivity
- and specificity
- The Summary of Findings format was understood
- Take home message from this qualitative study "…in the real world, many readers, especially clinicians with little interest in the methodology of the reviews, will read only abstract."



Zhelev et al. Systematic Reviews 2013, 2:32
http://www.systematicreviewsjournal.com/content/2/1/32

SYSTEMATIC REVIEWS

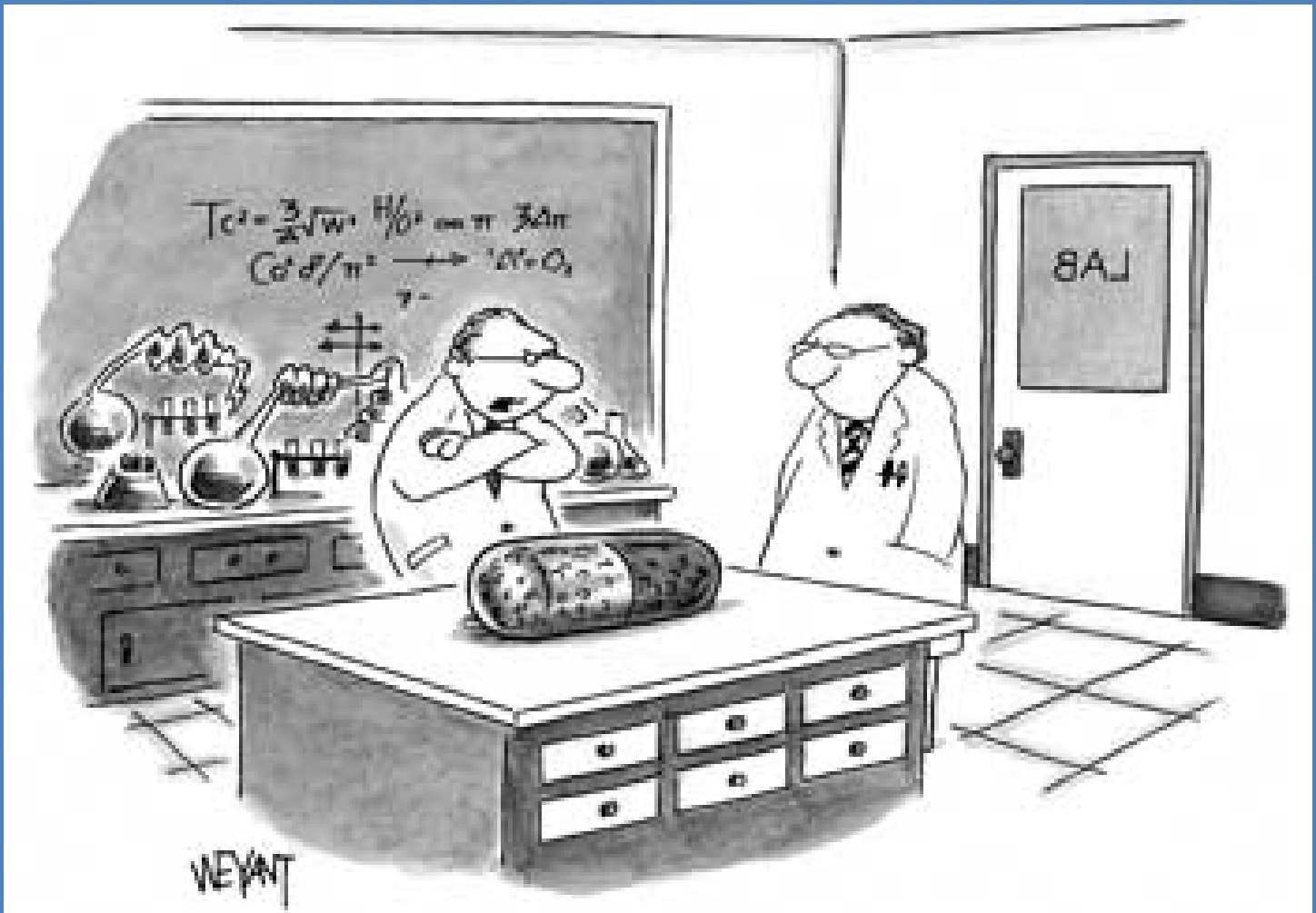
RESEARCH Open Access

A qualitative study into the difficulties experienced by healthcare decision makers when reading a Cochrane diagnostic test accuracy review

Zhivko Zhelev[1*†], Ruth Garside[2†] and Christopher Hyde[1†]

Abstract

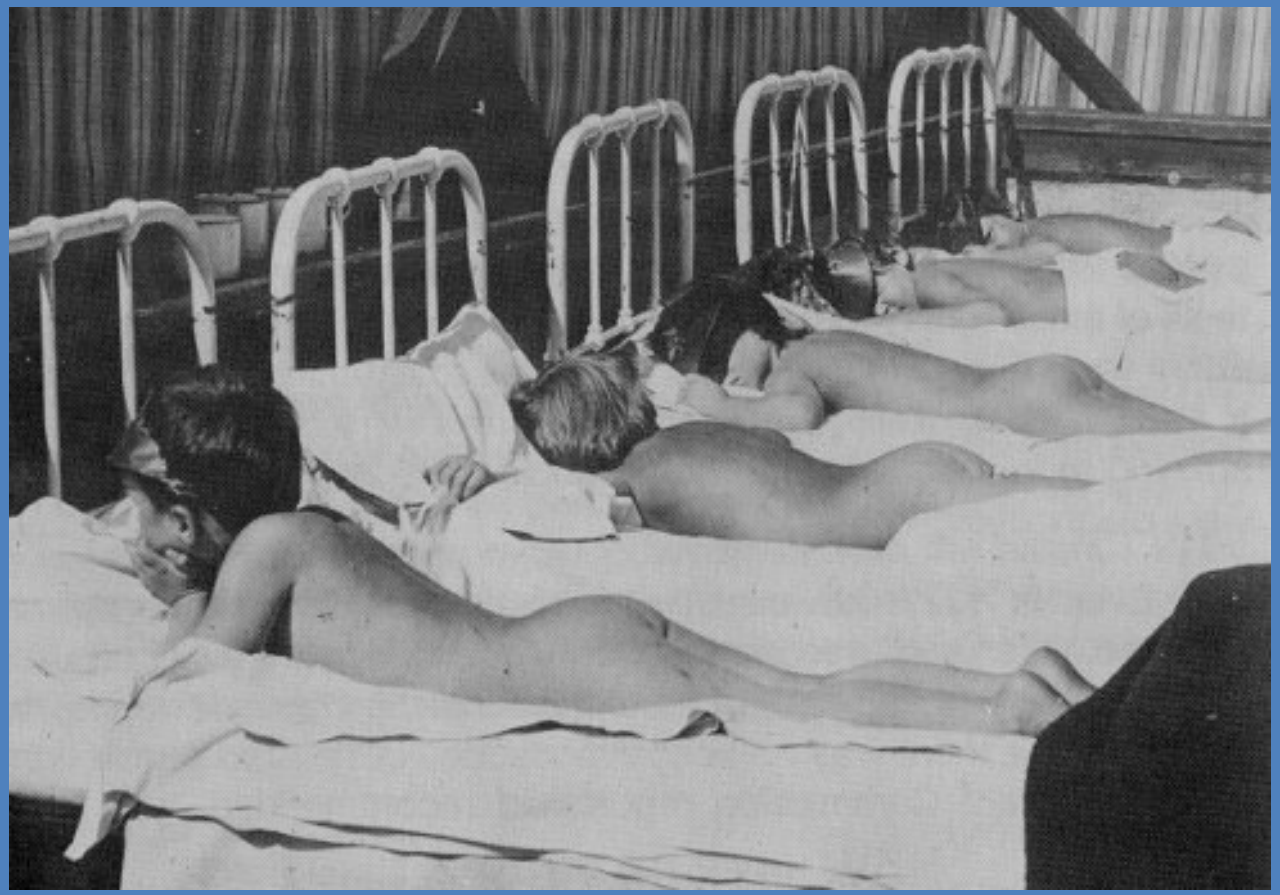
Background: Cochrane reviews are one of the best known and most trusted sources of evidence-based information in



"That's just great. I discover the cure for the common cold and all you can do is criticize."

Diagnostic tests do not make patients better

Diagnostic accuracy ≠ patient outcome



http://www.lung.ca/tb/images/full_archive/007_sun_treatment.jpg

Table 1| **Attributes of the test-treatment pathway that affect patient health**

Pathway component and mechanism	Definition
(1) Diagnostic test delivered	
Timing of test	Speed with which a test is performed within the management strategy
Feasibility	Completion of test process. Reasons for non-completion are: patient acceptability (patient's refusal to have test), test was contraindicated (clinical reason not to administer test), and technical failure (ability of diagnostic equipment to produce data)
Test process	Patients' interaction with test procedure, potentially causing physical or psychological harms or benefits
(2) Test result produced	
Interpretability	Degree to which test data can be used to inform a diagnostic classification
Accuracy	Ability of a test to distinguish between patients who have disease and those who do not
Timing of results	Speed with which test results are available
(3) Diagnosis made	
Timing of diagnosis	Speed with which a diagnostic decision is made
Diagnostic yield	Degree to which the test contributes to a patient diagnosis in any form, including: provision of a definitive diagnosis, confirmation of a suspected diagnosis, ruling out a working diagnosis, and distinguishing between alternative diagnoses with different treatment implications. Diagnostic yield is different from accuracy because it also incorporates any other information used by a doctor to make a diagnosis (such as previous test results)
Diagnostic confidence	Degree of confidence that doctors and patients have in the validity or applicability of a test result
(4) Management decided	
Therapeutic yield	Degree to which diagnostic decisions affect treatment plans
Therapeutic confidence	Certainty with which doctors and patients pursue a course of treatment
(5) Treatment implemented	
Timing of treatment	Speed with which patients receive treatment
Treatment efficacy	Ability of the treatment intervention to improve patient outcomes
Adherence	Extent to which patients participate in the management plan, as advised by their doctor, to attain therapeutic goal

di Ruffano, BMJ 2012

Test result produced	
Accuracy	Ability of the test to distinguish between patients who have disease and those who do not

di Ruffano, BMJ 2012

Int J Tuberc Lung Dis. 2013 Sep;17(9):1160-9. doi: 10.5588/ijtld.13.0050. Epub 2013 Jun 27.

Methodological and reporting quality of systematic reviews on tuberculosis.

Nicolau I[1], Ling D, Tian L, Lienhardt C, Pai M.

- 137 systematic reviews on TB published between 2005 and 2010

"Overall, none of the 137 systematic reviews fulfilled all of the 11 AMSTAR quality items…Furthermore, as Cochrane reviews tend to be more rigorous and better reported, the TB field may benefit from a larger number of Cochrane systematic reviews."

AMSTAR: assessing methodological quality of systematic reviews

1. an a priori design
2. duplicate study selection and data extraction
3. a comprehensive literature search
4. the use of status of publication as an inclusion criteria
5. a list of included/excluded studies
6. characteristics of included studies
7. documented assessment of the scientific quality of included studies
8. appropriate use of the scientific quality in forming conclusions
9. the appropriate use of methods to combine findings of studies
10. assessment of the likelihood of publication bias
11. documentation of conflict of interest

Shea et al. BMC Medical Research Methodology 2007, 7:10

Summary

1. Review question – PPPIPTR (PICO)
2. Study selection – involve information specialist
3. Quality assessment - use QUADAS-2
4. Data-analysis - Bivariate or HSROC random-effects model for meta-analysis
5. Interpretation and presentation - pay special attention to the abstract; use Summary of Findings tables

References

- Cochrane Diagnostic Test Accuracy Working Group srdta.cochrane.org/
- Leeflang. Ann Intern Med. 2008;149:889-897
- Whiting PF et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.
- RevMan ims.cochrane.org/revman

With special thanks

- Chris Hyde
- Mariska Leeflang
- Reem Mustafa
- Hans Reitsma
- Penny Whiting