Dr Mariska Leeflang
Dept. Clinical Epidemiology, Biostatistics and Bioinformatics
Academic Medical Center, University of Amsterdam
Room J1B – 210
PO Box 227700
1100 DE Amsterdam
m.m.leeflang@amc.uva.nl

Quality assessment of diagnostic test accuracy studies

Montreal, Monday May 25th, 2009

Mariska Leeflang

Steps in a systematic review

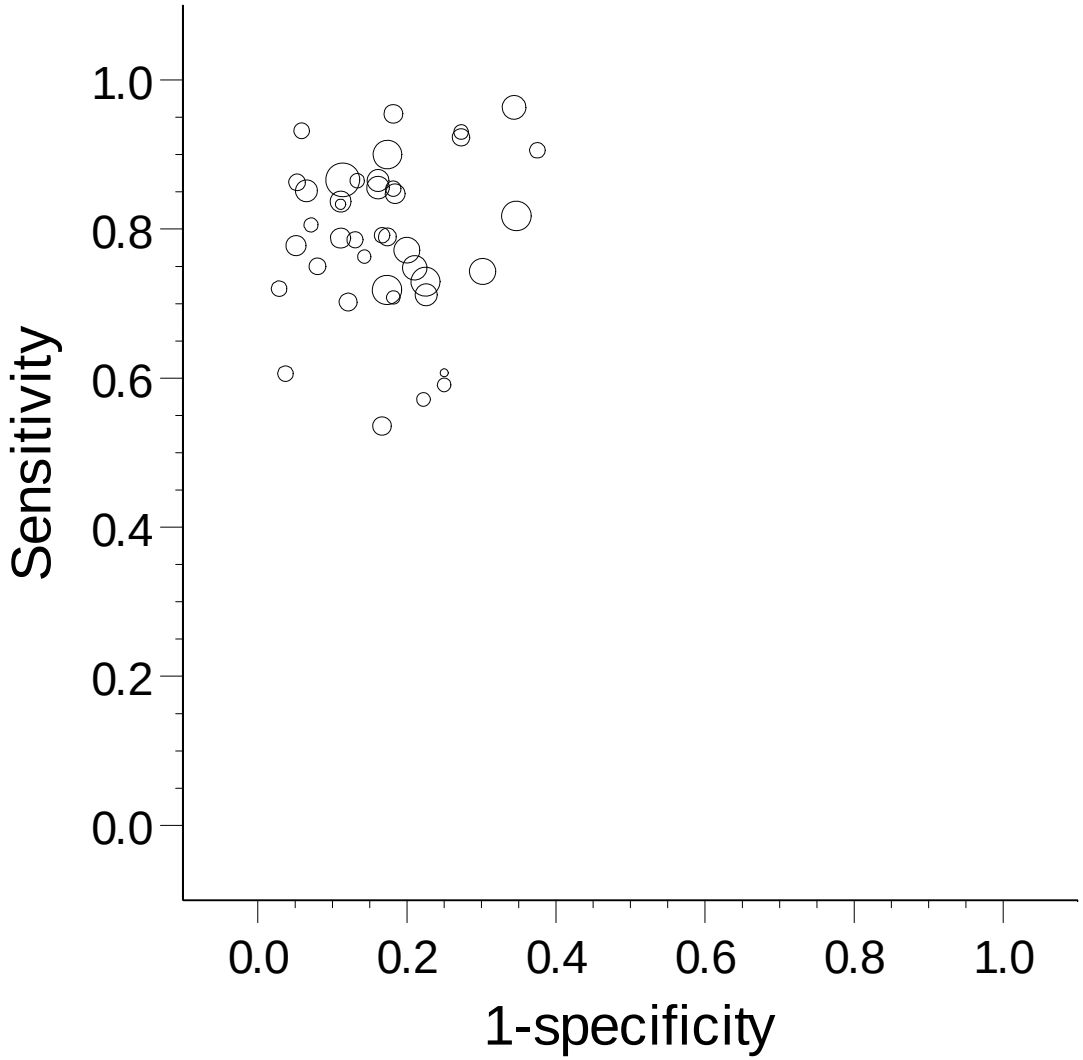
1. Formulating the question
   (and defining criteria for inclusion of studies)
2. Searching for studies
3. Selecting studies
4. Collecting data
5. Assessing methodological quality
6. Analysing and presenting results
7. Interpreting results

Quality assessment

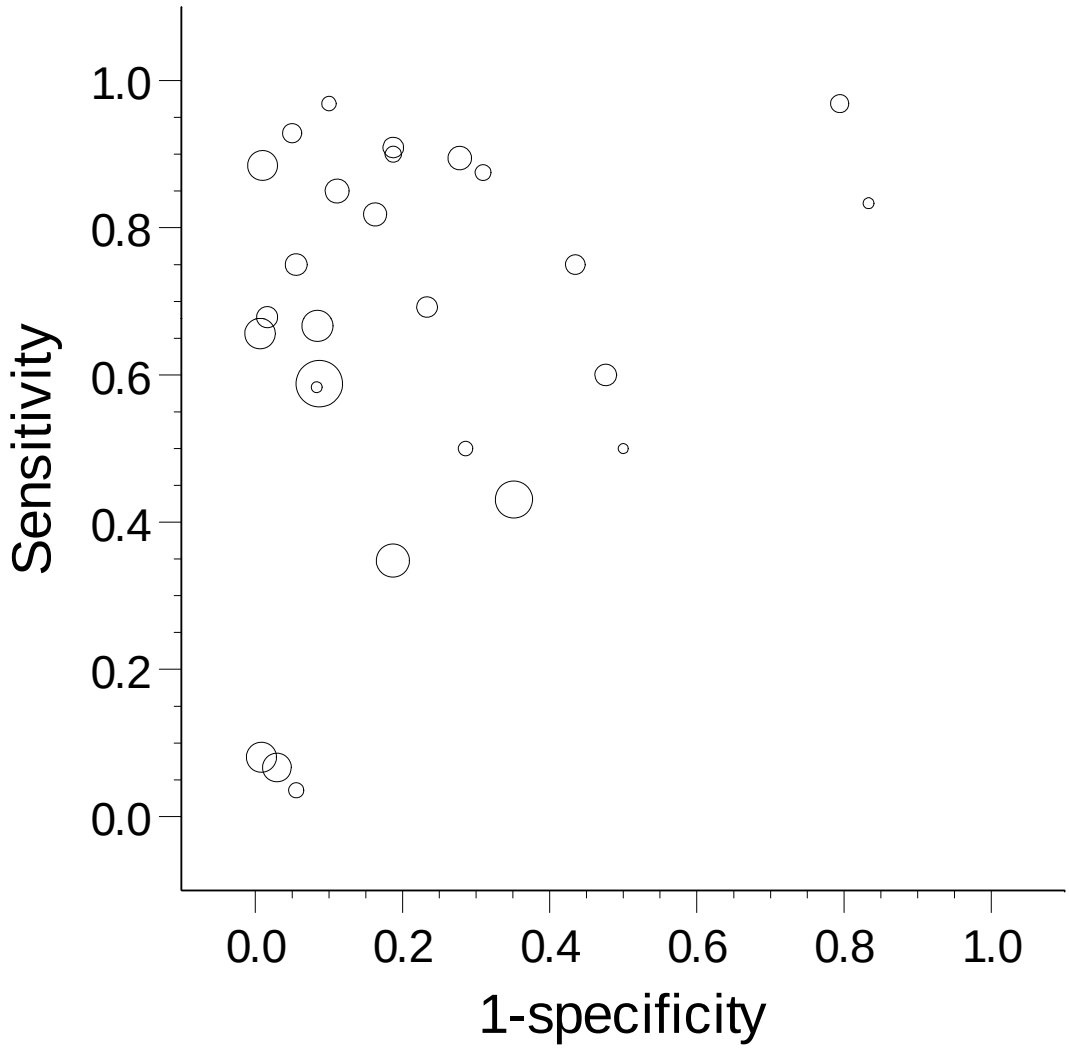
Why assess quality?

- Problem 1: Bias in primary studies can lead to misleading summary estimates of accuracy
- Problem 2: Results of primary studies may vary
- Quality assessment to guide the interpretation of results in terms of potential for bias and sources of heterogeneity

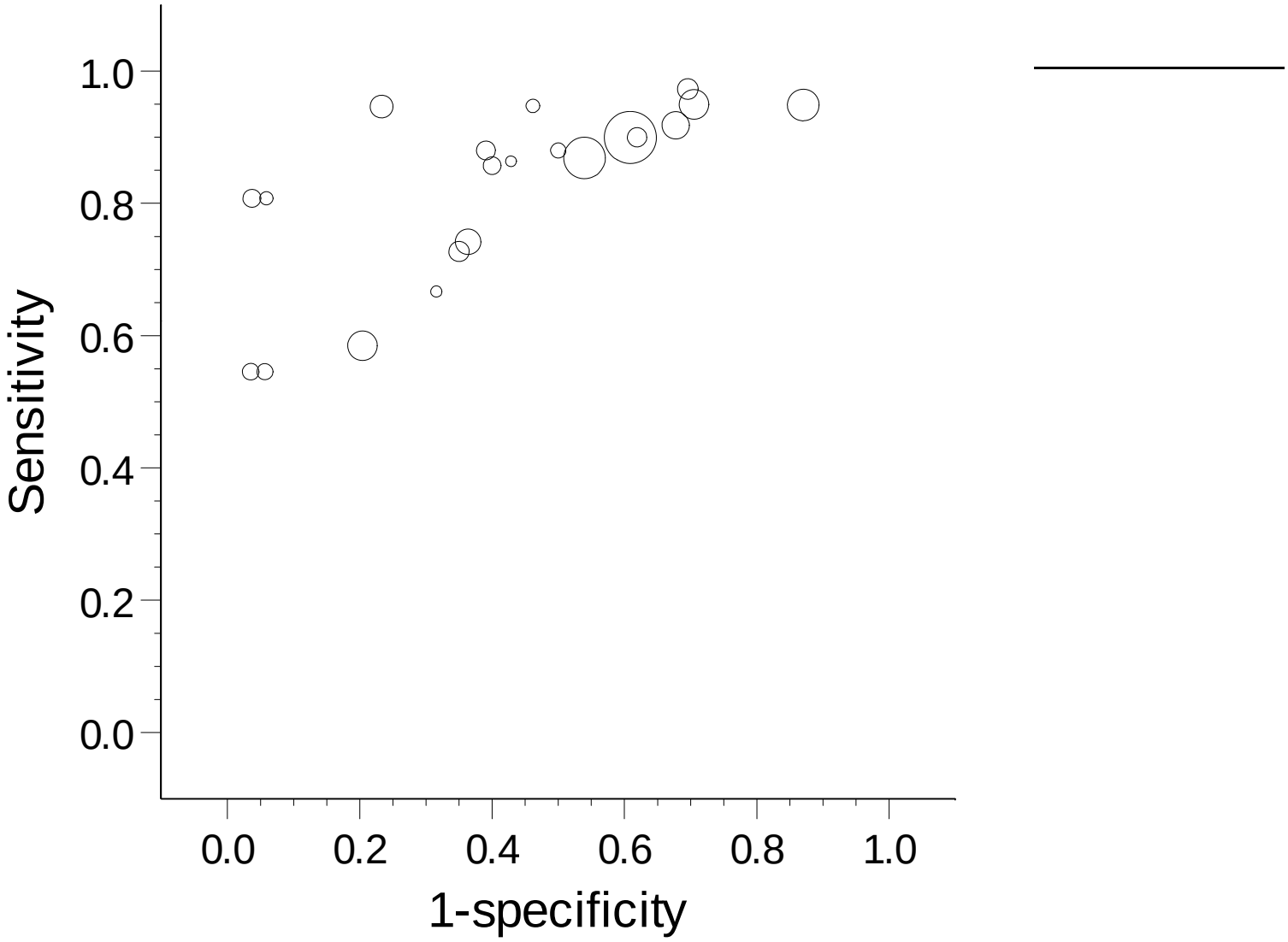
Echocardiography in Coronary Heart Disease



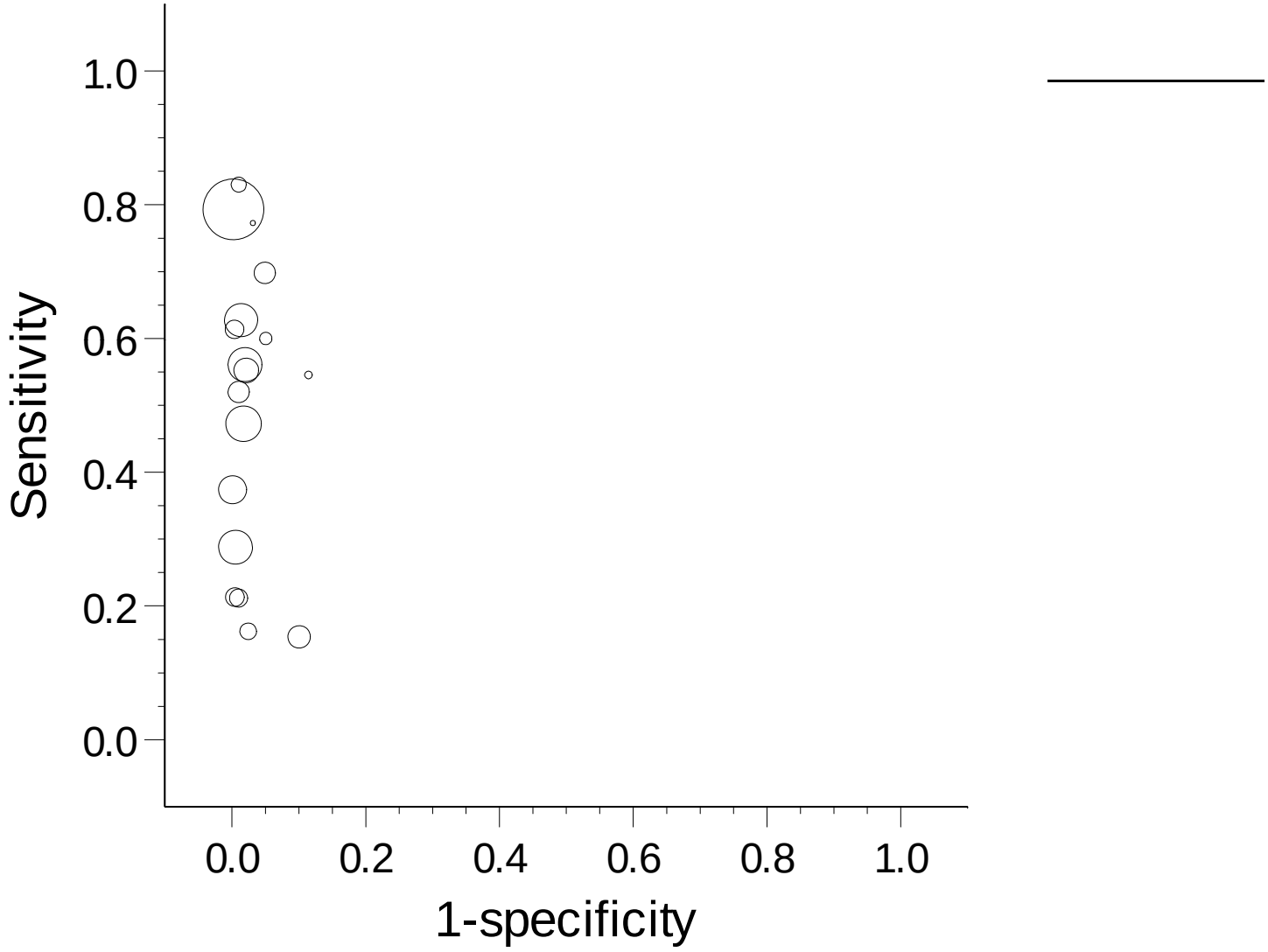
GLAL in Gram Negative Sepsis



F/T PSA in the Detection of Prostate cancer



Dip-stick Testing for Urinary Tract Infection



Cochrane definition of quality

"the methodological quality of a study; the degree to which the design and conduct of a study fit to the study objectives"

How to assess quality?

- Quality assessment tools:
  - Large number of different tools
  - Styles: Quality scores/levels of evidence/ component approach
  - Cochrane handbook recommends modified version of the QUADAS tool

QUADAS

- General tool
- Systematically developed based on empirical evidence and a formal consensus method
- Set of 11 required (strongly suggested) items, and a list of additional items to consider

QUADAS items

1	Was the spectrum of patients representative of the patients who will receive the test in practice?
2	Is the reference standard likely to correctly classify the target condition?
3	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
4	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
5	Did patients receive the same reference standard regardless of the index test result?
6	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
7	Were the index test results interpreted without knowledge of the results of the reference standard?
8	Were the reference standard results interpreted without knowledge of the results of the index test?
9	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
10	Were uninterpretable/ intermediate test results reported?
11	Were withdrawals from the study explained?

Sources of bias and variation

Whiting P et al. Ann Intern Med. 2004 ;140(3):189-202.

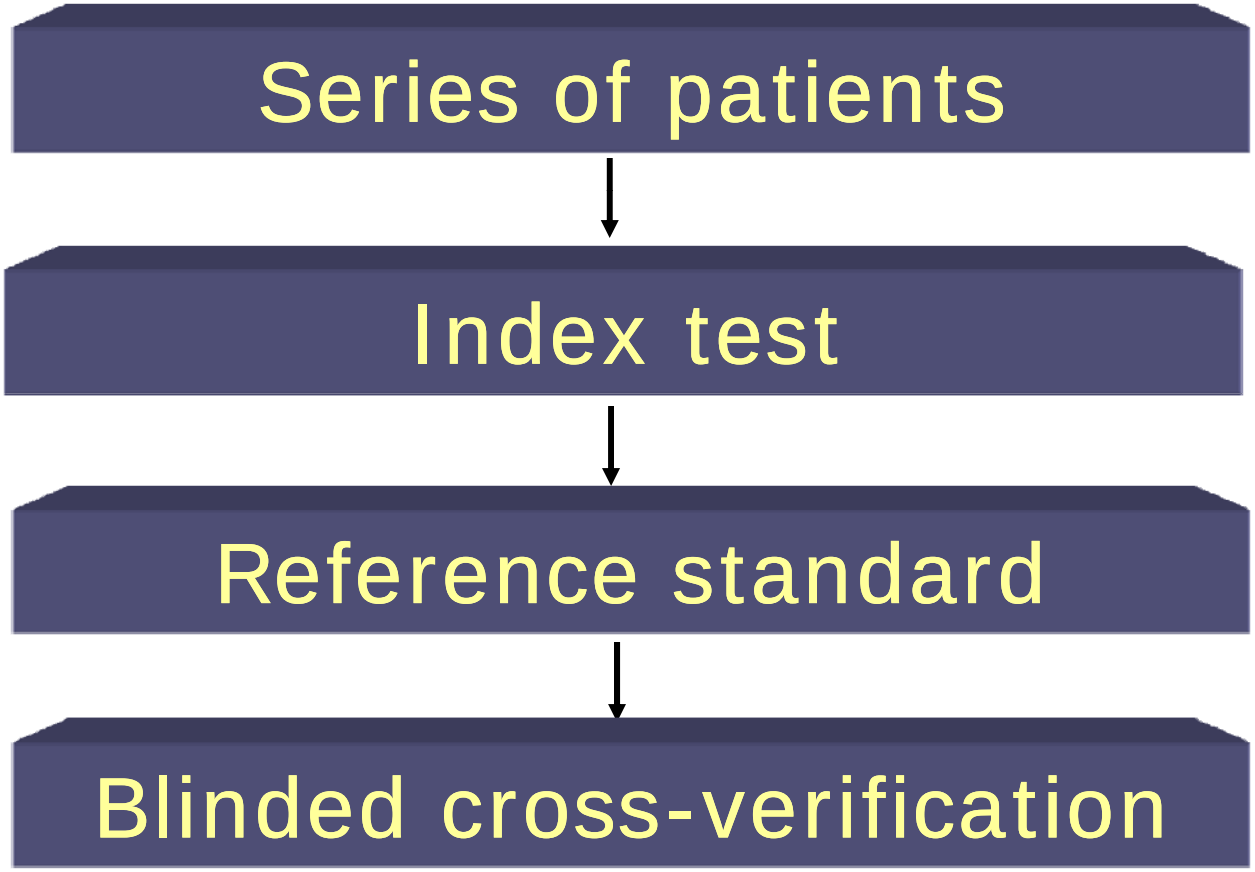
- *overview of sources of bias and variation*

Whiting P et al. BMC Med Res Methodol. 2003;3:25.

- *development of QUADAS*

Cochrane Handbook for Reviews of Diagnostic Test Accuracy

Basic Test Accuracy Study

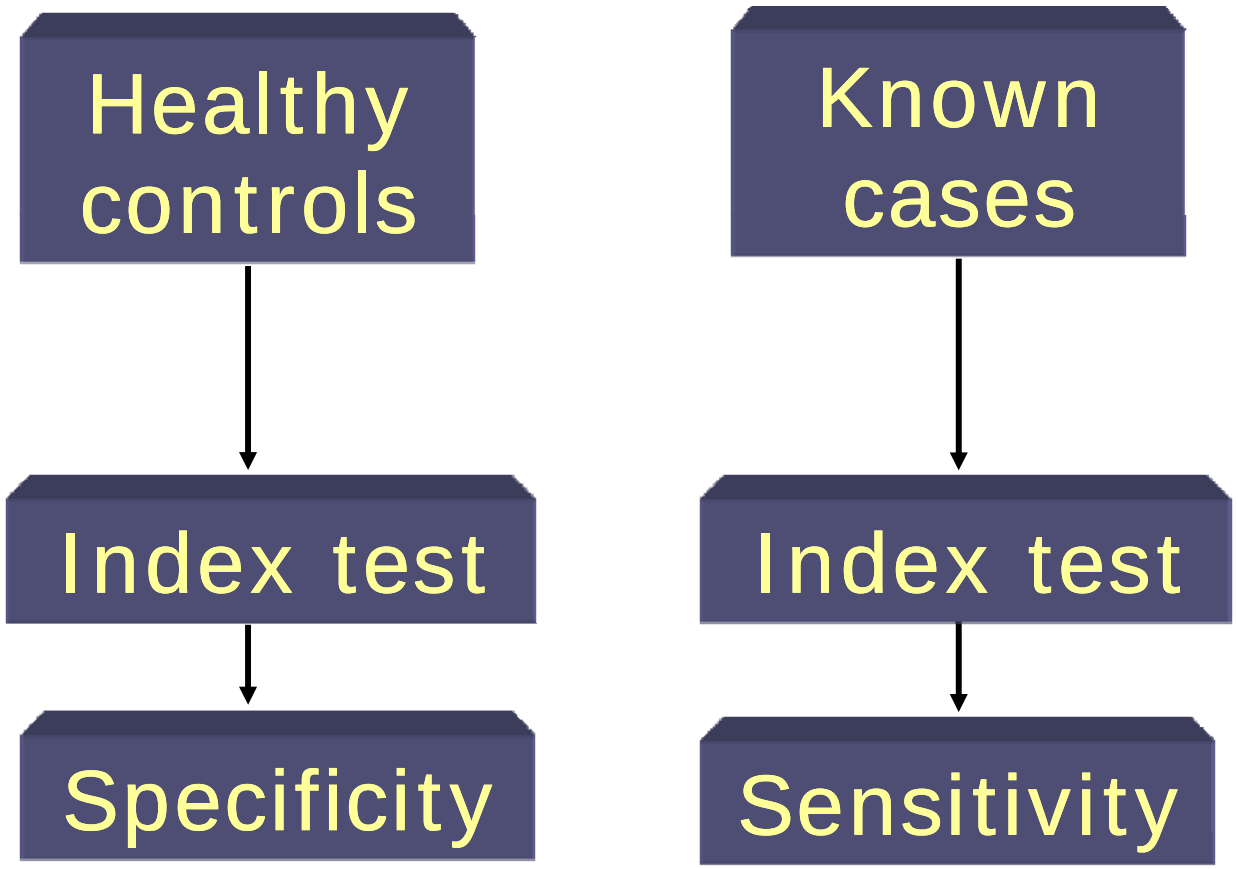


Problems with spectrum

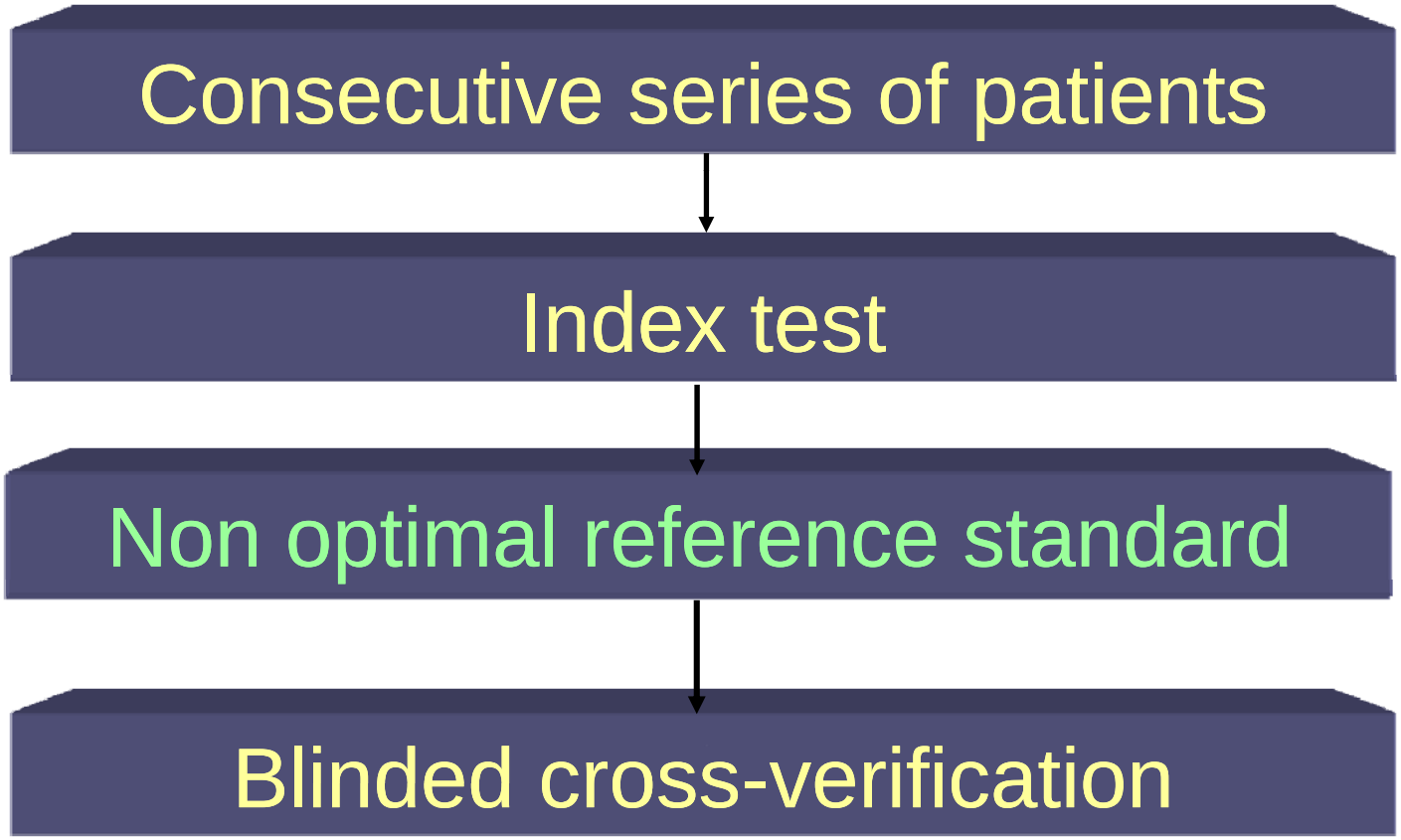
Measures of accuracy vary across patient groups:

- Patient characteristics e.g. age
- Patient selection/Study design
- Setting

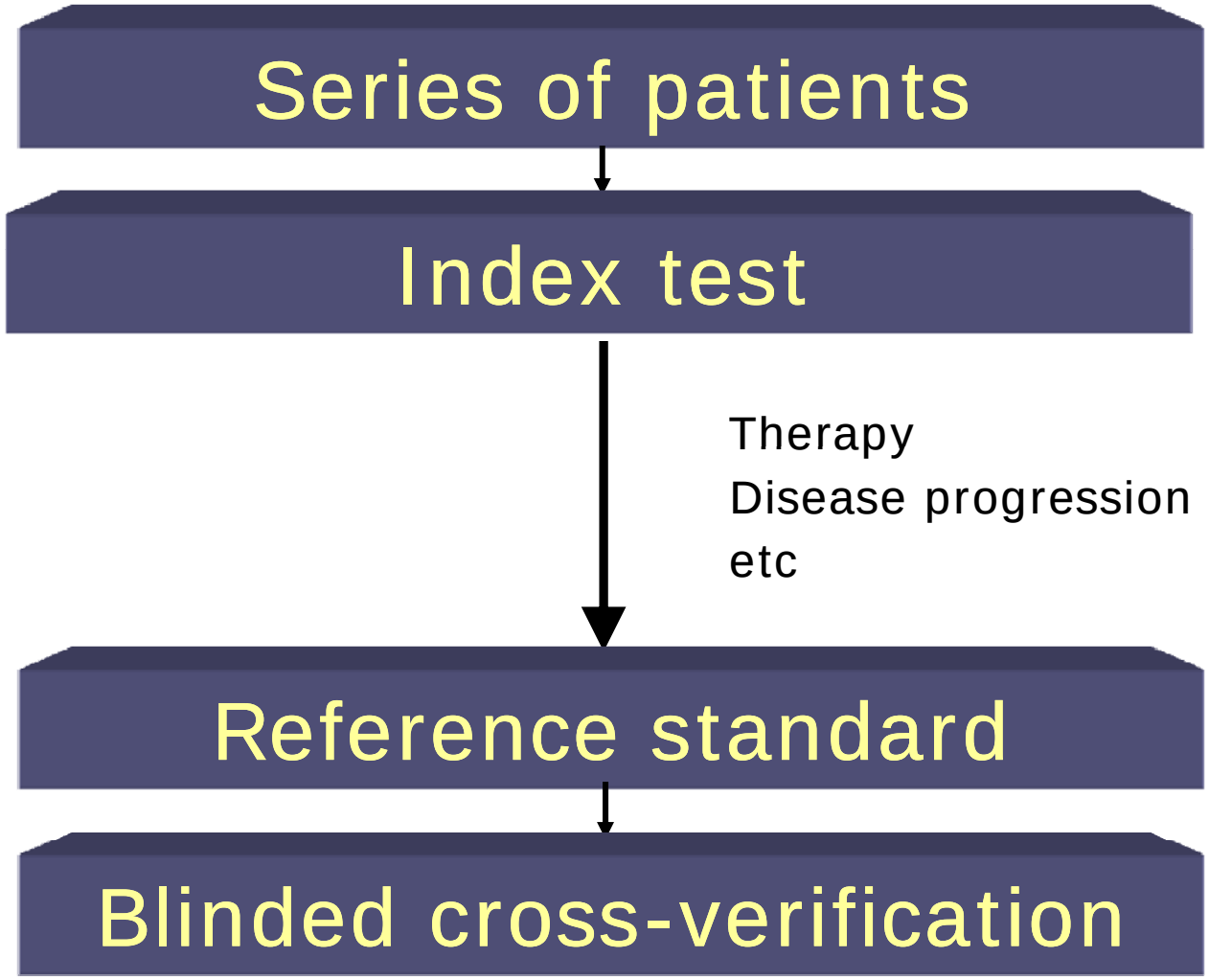
Diagnostic case-control design



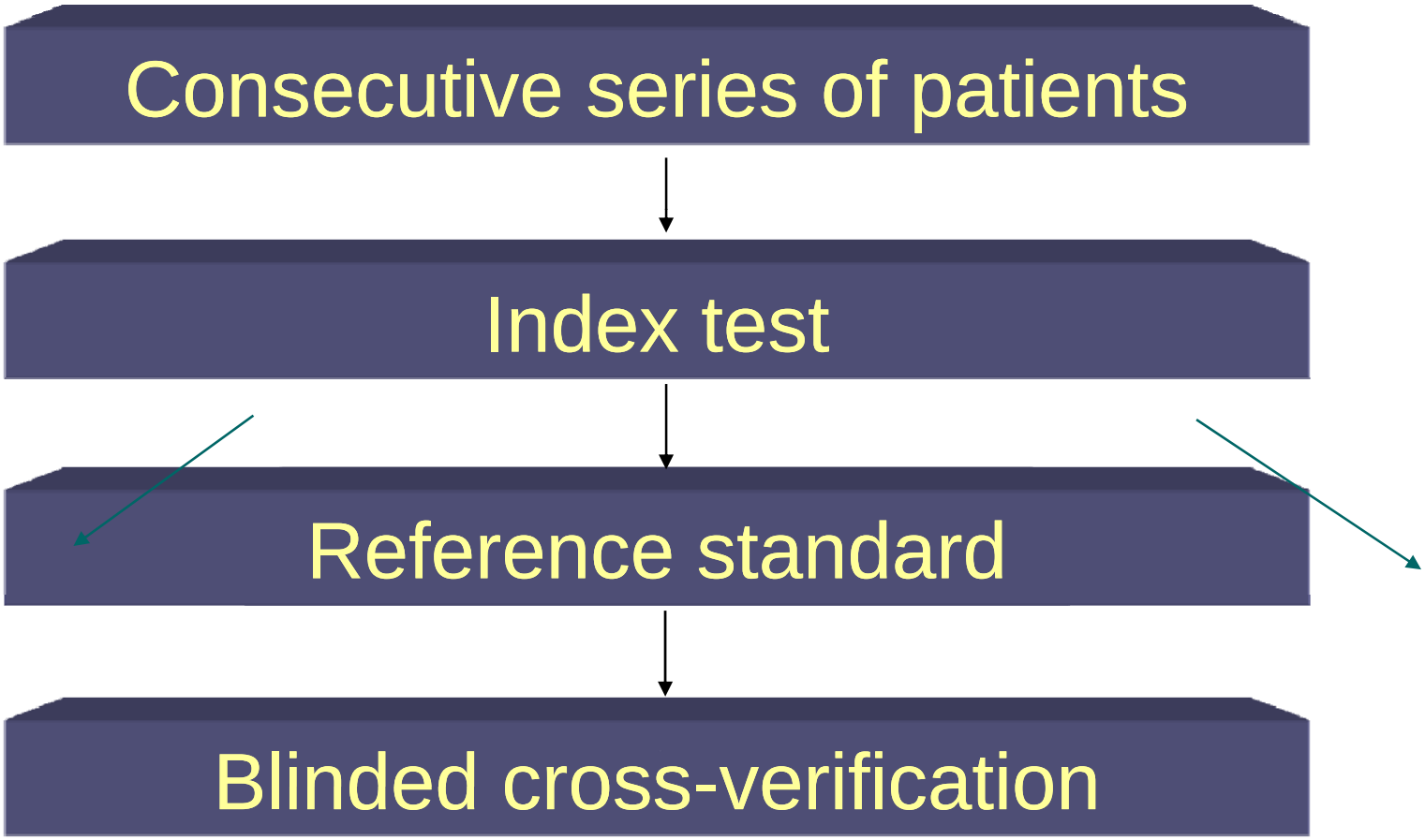
Reference standard bias



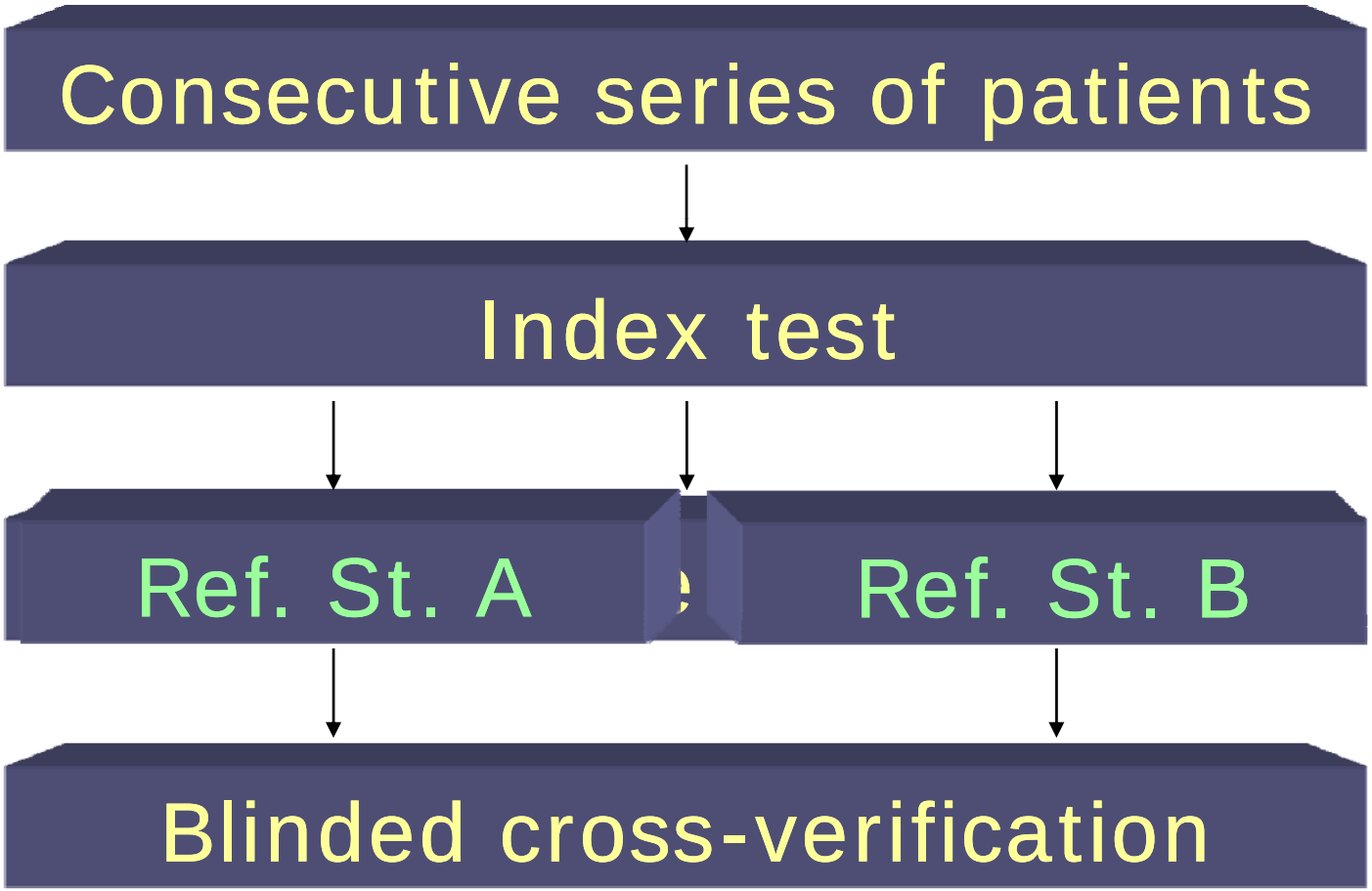
Time between index test and reference standard



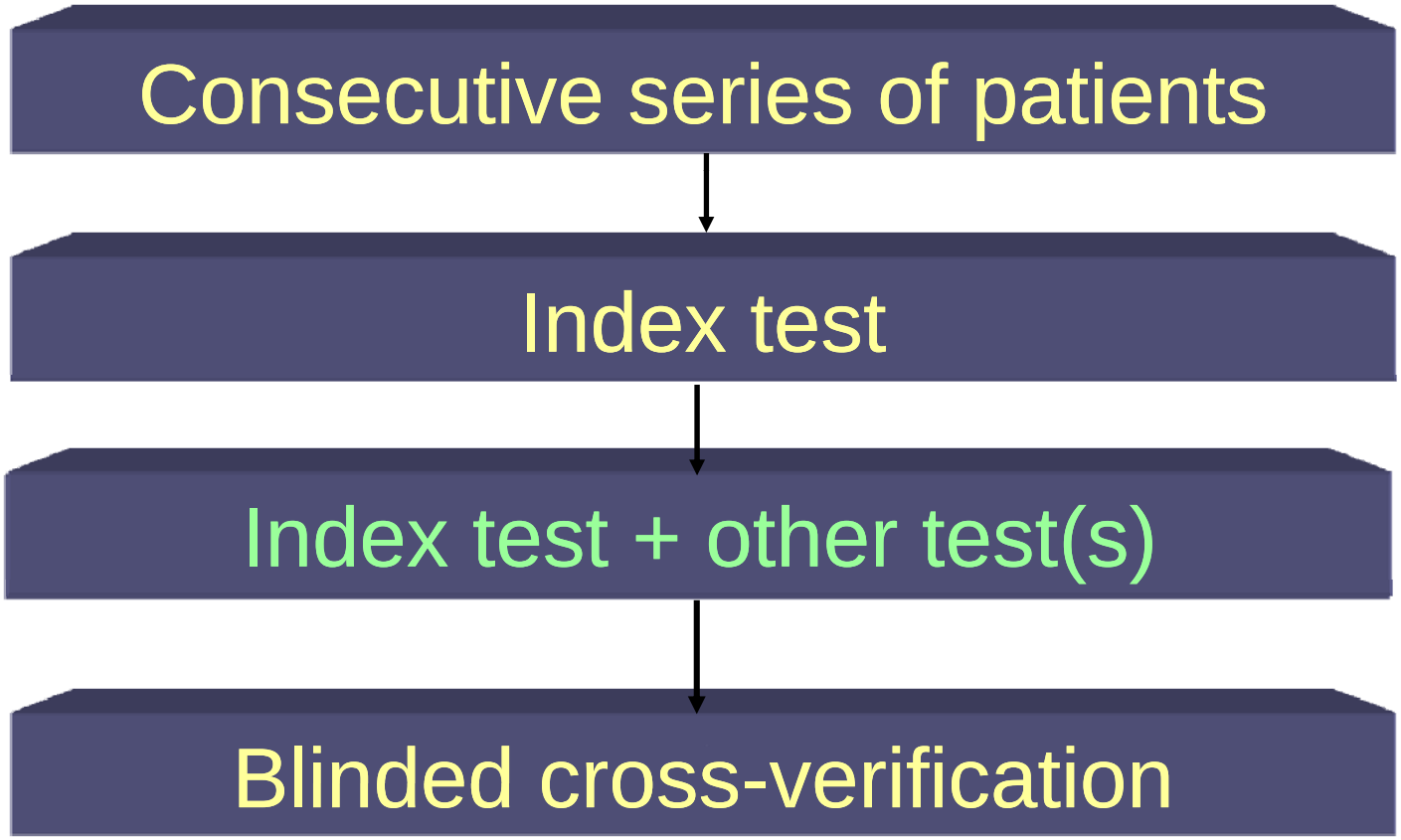
Partial verification bias



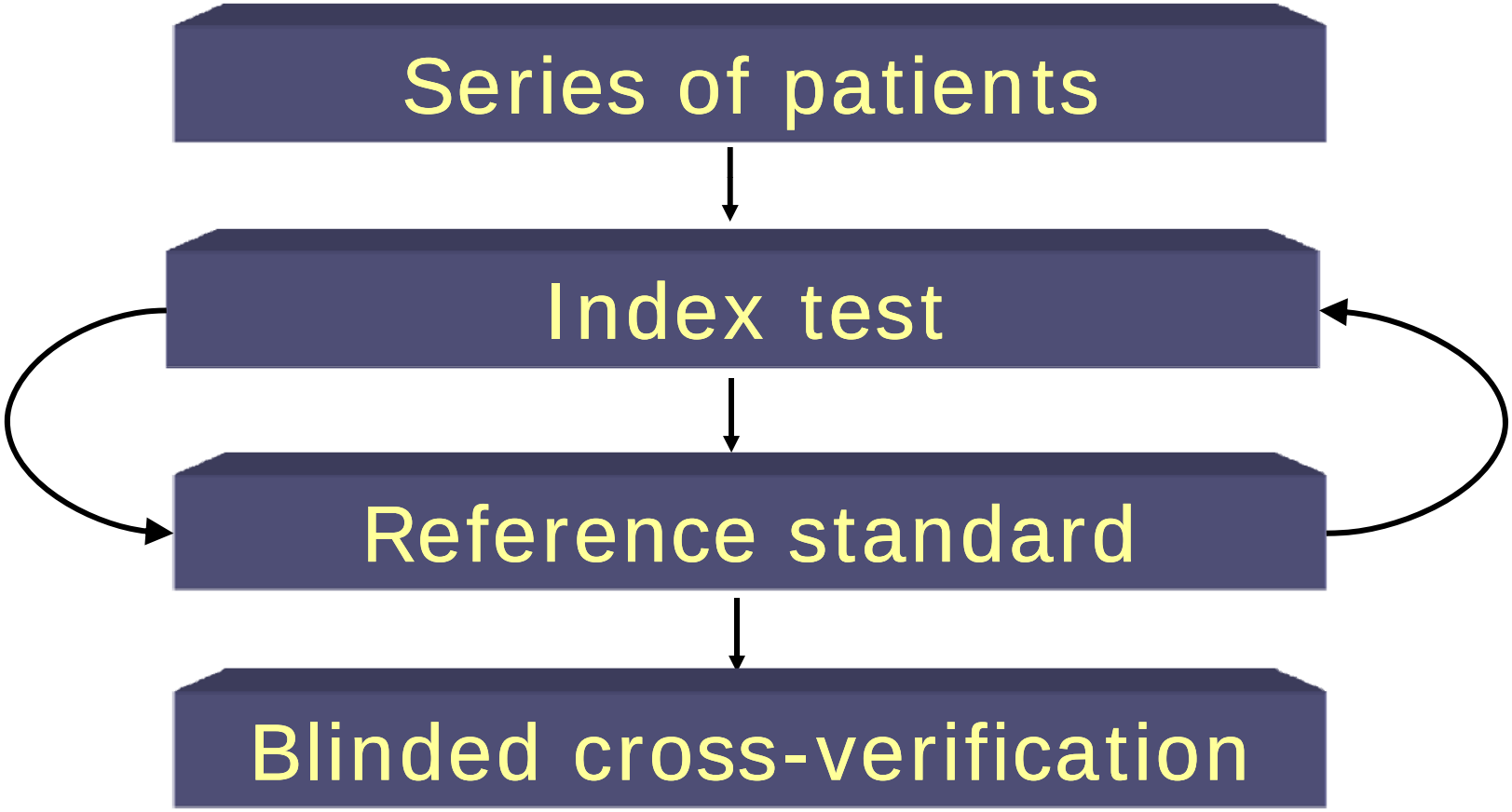
Differential verification bias



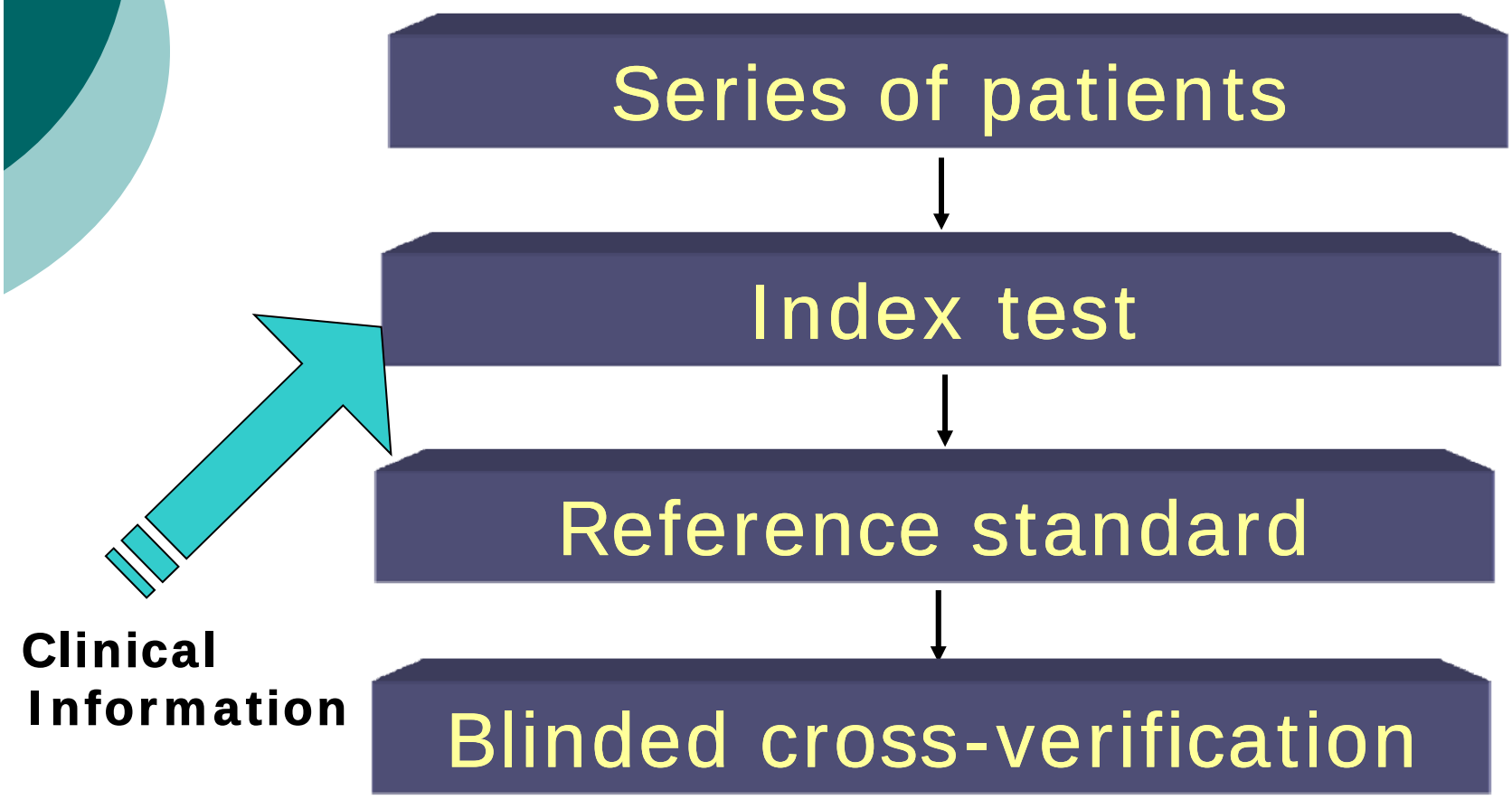
Incorporation bias



Blinding


Series of patients
Index test
Reference standard
Blinded cross-verification

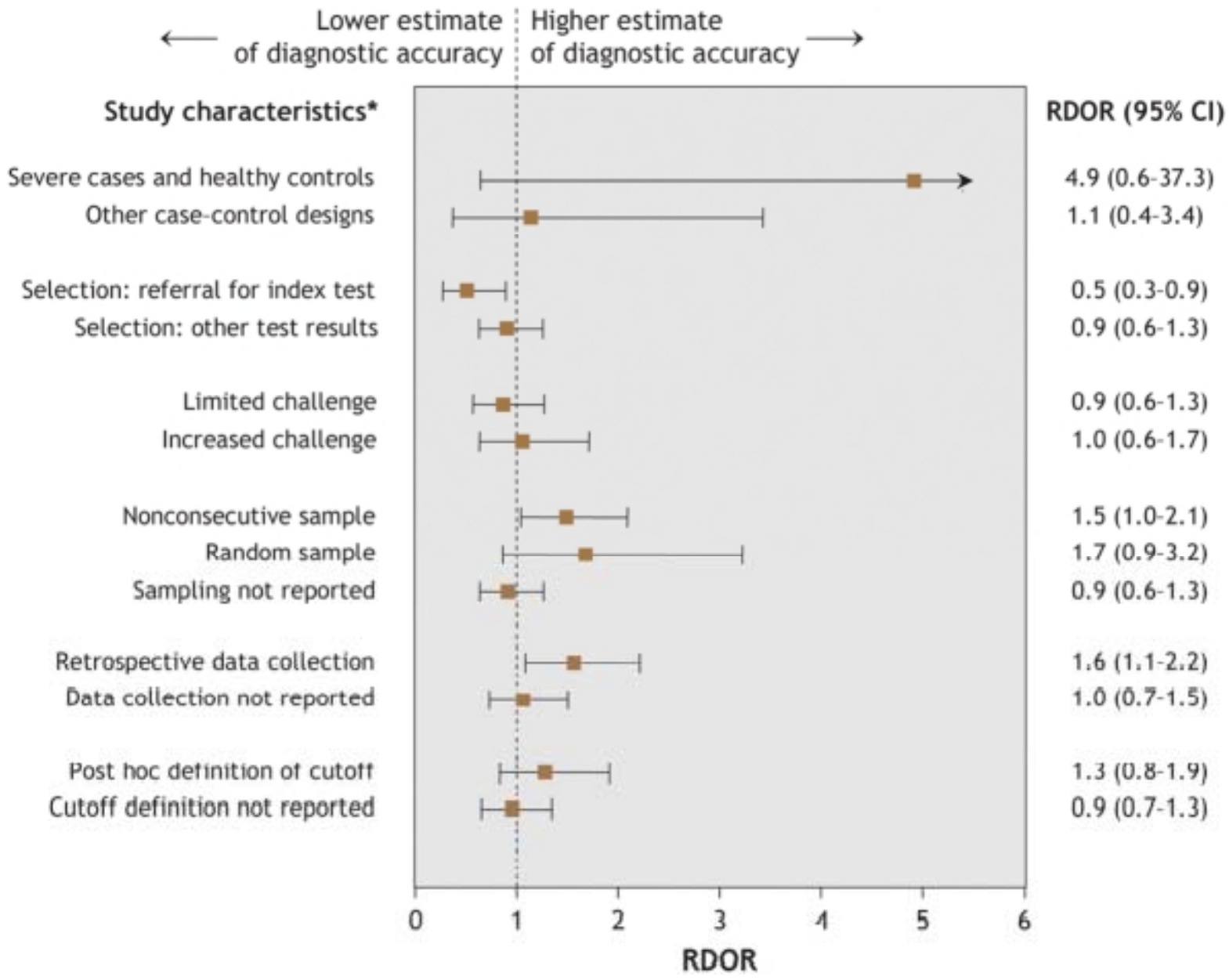
Blinding


Series of patients
Index test
Reference standard
Blinded cross-verification
Clinical Information

Two important reporting items

- Reporting of uninterpretable/ intermediate test results
- Explanation of withdrawals

	biopsy		
	+	-	
HPV+	45	68	113
+/-	20	25	45
HPV-	7	161	198
	72	254	356


Lower estimate of diagnostic accuracy
Higher estimate of diagnostic accuracy
Study characteristics*
RDOR (95% CI)
Severe cases and healthy controls
4.9 (0.6-37.3)
Other case-control designs
1.1 (0.4-3.4)
Selection: referral for index test
0.5 (0.3-0.9)
Selection: other test results
0.9 (0.6-1.3)
Limited challenge
0.9 (0.6-1.3)
Increased challenge
1.0 (0.6-1.7)
Nonconsecutive sample
1.5 (1.0-2.1)
Random sample
1.7 (0.9-3.2)
Sampling not reported
0.9 (0.6-1.3)
Retrospective data collection
1.6 (1.1-2.2)
Data collection not reported
1.0 (0.7-1.5)
Post hoc definition of cutoff
1.3 (0.8-1.9)
Cutoff definition not reported
0.9 (0.7-1.3)
0
1
2
3
4
5
6
RDOR

Assessment of items

- All items scored as yes/no/unclear
- Items phrased so that yes indicates absence of bias
- Background document describes how items should be scored

Practical Issues

- Number of assessors
- Background of assessors
- Resolving disagreement
- Piloting the assessment process
- Develop your quality assessment tool

Your quality assessment tool

- Items to include
  - Core items
  - Additional items
    - select from suggested items
    - add your own if other items are important for your review topic
- Produce scoring guidelines specific to your review

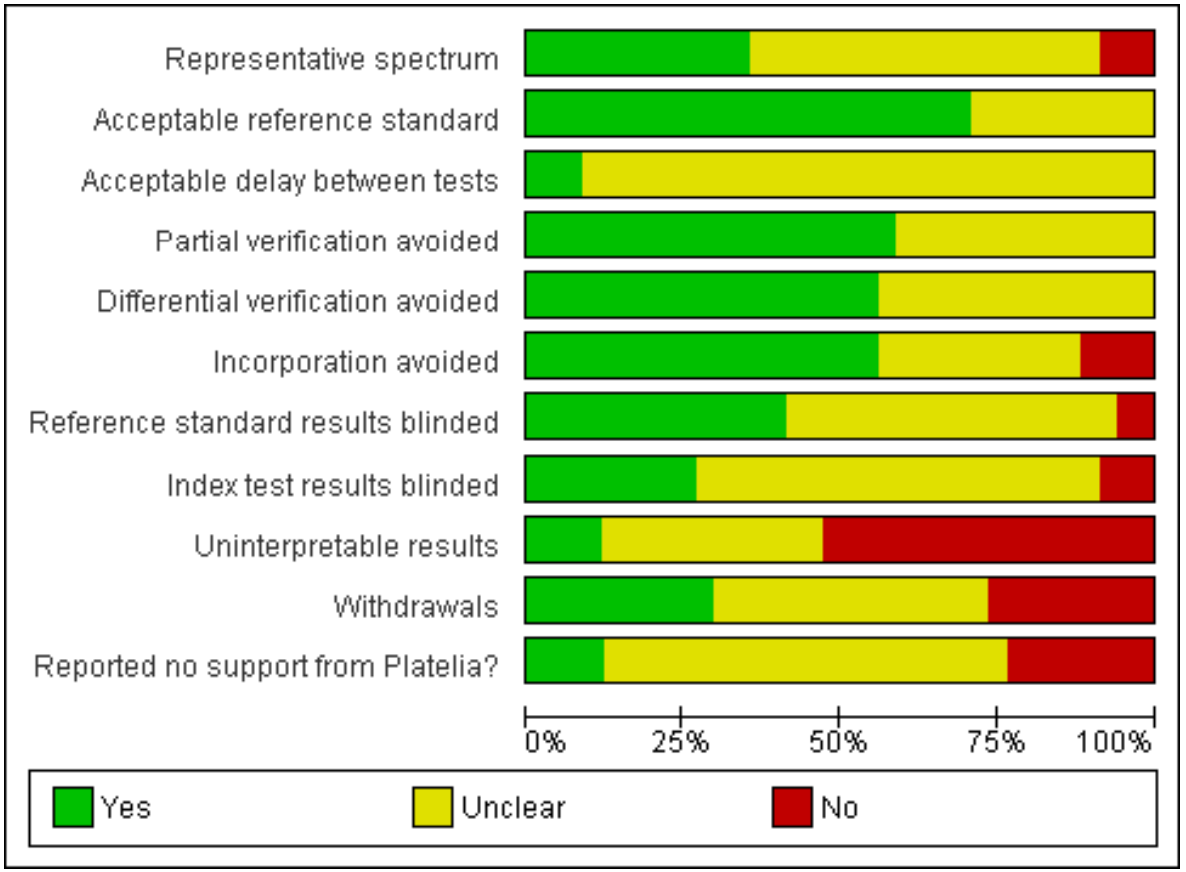
Additional items

12.	If a cut-off value has been used, was it established before the study was started (pre-specified cut-off value)?
13.	Is the technology of the index test likely to have changed since the study was carried out?
14.	Did the study provide a clear definition of what was considered to be a “positive” result?
15.	Was treatment started after the index test was carried out but before the reference standard was performed?
16.	Was treatment started after the reference standard was carried out but before the index test was performed?
17.	Were data on observer variation reported?
18.	Were data on instrument variation reported?
19.	Were data presented for appropriate patient sub-groups?
20.	Was an appropriate sample size included?
21.	Were objectives pre-specified?

Presenting study quality

- Present the results of the quality assessment:
  - Graphically

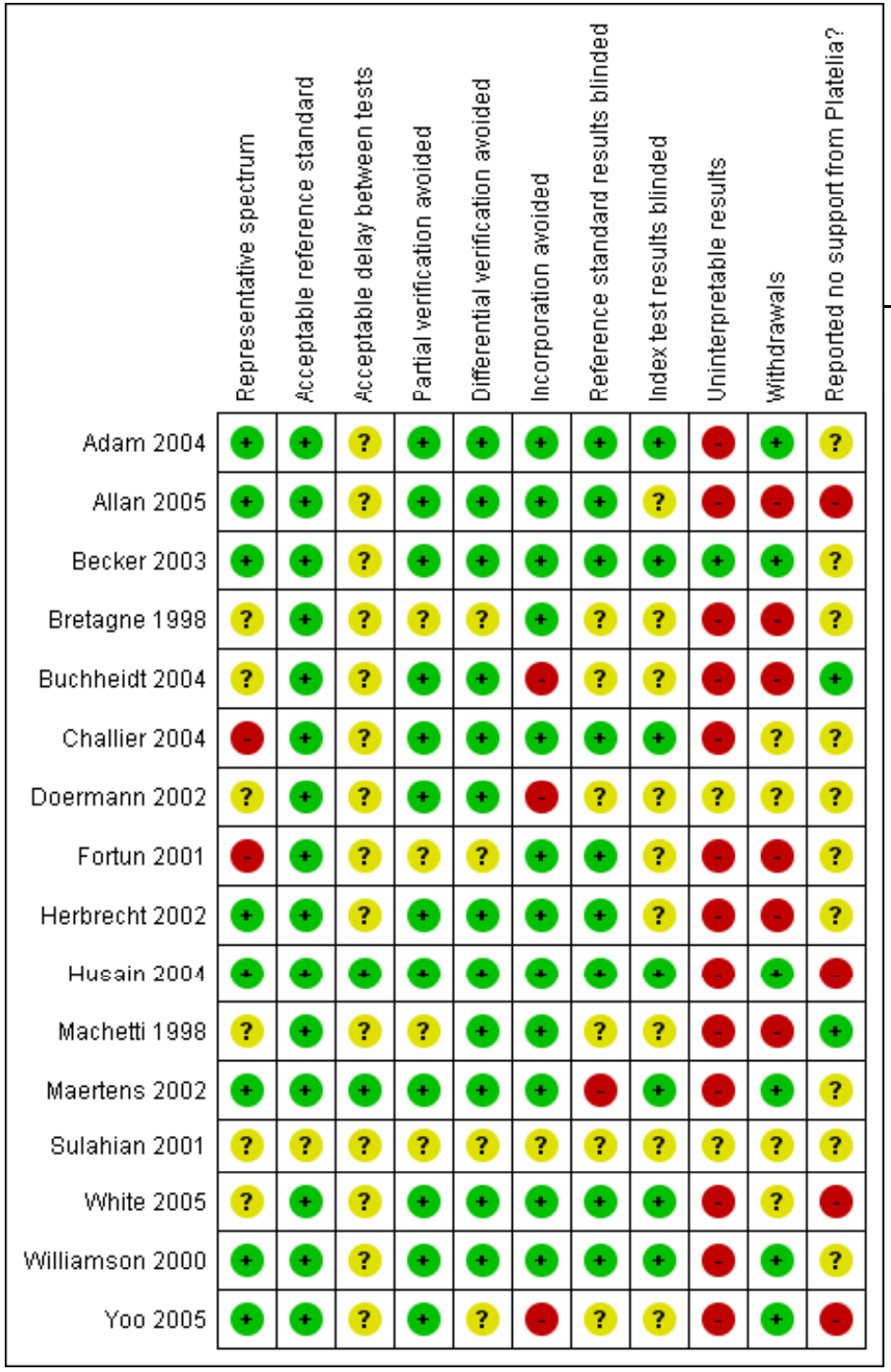
Methodological Quality Graph



Review authors' judgments about each methodological quality item presented as percentages across all included studies.

Incorporating study quality

- Present the results of the quality assessment:
  - In a table
  - Graphically



Methodological quality summary.

Review authors' judgments about each methodological quality item for each included study.

Using study quality

- Present the results of the quality assessment:
  - In a table
  - Graphically
- Investigate individual quality items as potential sources of heterogeneity
- Basis for recommendations for future research

Formal incorporation of study quality

- Restricting the analysis to high quality studies
- Stratified analysis according to presence/absence of specific quality criteria
- Sensitivity analyses to investigate robustness of results
- Investigate several features simultaneously using meta-regression analysis

Always: define methodological criteria a priori

Problems with quality assessment

- Not as straightforward as it might sound!
- Hampered by poor reporting
- Quality assessment is subjective
- Quality scores are not recommended
- Statistical incorporation of quality problematic with limited studies

Now it’s your turn!

Example: BNP for heart failure

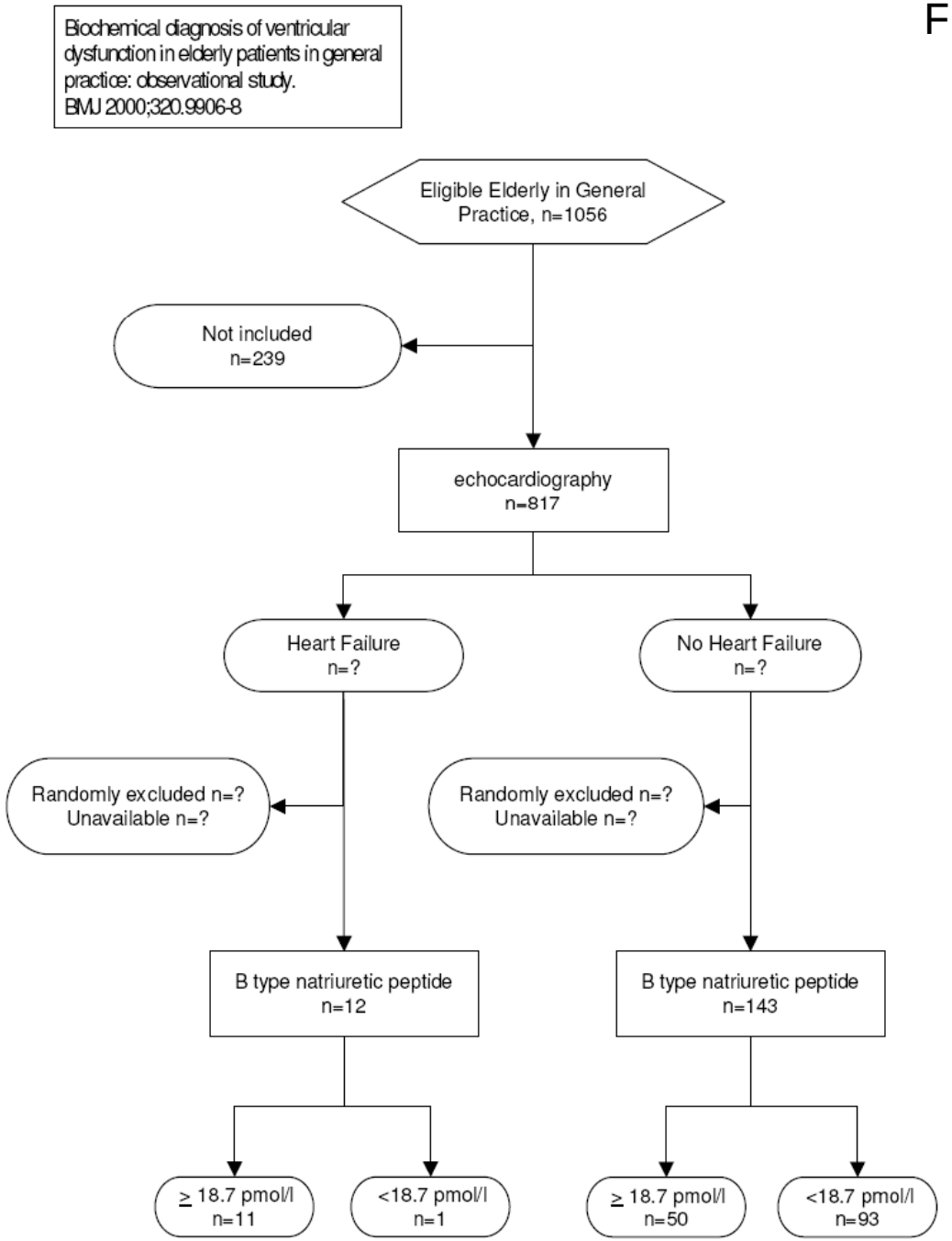
- **Aim:** To assess the accuracy of BNP for the diagnosis of heart failure
- In small groups:
  1. Produce a flow diagram for the study
  2. Discuss (attention to what has been done, what is missing and possible consequences):

     QUADAS item 1 (spectrum)
     QUADAS items 2, 4 and 5 (verification)
  3. Discuss the conclusion of the authors

QUADAS items

1	Was the spectrum of patients representative of the patients who will receive the test in practice?
2	Is the reference standard likely to correctly classify the target condition?
3	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
4	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
5	Did patients receive the same reference standard regardless of the index test result?
6	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
7	Were the index test results interpreted without knowledge of the results of the reference standard?
8	Were the reference standard results interpreted without knowledge of the results of the index test?
9	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
10	Were uninterpretable/ intermediate test results reported?
11	Were withdrawals from the study explained?

Flowchart


Biochemical diagnosis of ventricular dysfunction in elderly patients in general practice: observational study. BMJ 2000;320.9906-8
Eligible Elderly in General Practice, n=1056
Not included n=239
echocardiography n=817
Heart Failure n=?
No Heart Failure n=?
Randomly excluded n=? Unavailable n=?
Randomly excluded n=? Unavailable n=?
B type natriuretic peptide n=12
B type natriuretic peptide n=143
≥ 18.7 pmol/l n=11
<18.7 pmol/l n=1
≥ 18.7 pmol/l n=50
<18.7 pmol/l n=93

Conclusions

- Quality assessment is essential, but exact effects not (yet) known
- The QUADAS tool should be used as a starting point
- Study quality should be incorporated into all reviews
- The quality items and scoring guidelines should be tailored to your review question
- The results of the quality assessment should be presented
- No quality scores and cut-offs for ‘good’ quality