Systematic literature searching for evaluation of the accuracy of a new diagnostic test

Chien-Chang Lee MD, ScD

Department of Emergency Medicine,
National Taiwan University Hospital

Outline

1. Evaluate the accuracy of a diagnostic test
2. Design of a diagnostic test study
3. Systematic review of diagnostic test studies
4. Quality assessment
5. Interpretation and clinical application

Outline

1. Evaluate the accuracy of a diagnostic test
2. Systematic review of diagnostic test studies
3. Literature search
4. Quality assessment
5. Interpretation

Why we need a diagnostic test?

- We need “information” to make a decision
- “Information” is usually a result from a test

- Medical tests:
  - To screen for a risk factor (screening test)
  - To diagnose a disease (diagnostic test)
  - To estimate a patient’s prognosis (prognostic test)

- When and in whom, a test should be done?
  - When “information” from test result have a value.

Diagnostic test

- Dichotomous
  - DNA SNPs
  - HIV screening test
  - Physical exam, imaging test
- Ordered Categorical Scale
  - Charlson scale
  - Sequential Organ Failure Assessment (SOFA) scale
- Continuous
  - Biochemical tests: serum levels of creatinine, bilirubin or calcium
  - Biomarker tests: serum levels of biomarkers
  - Blood cell counts: WBC, RBC, Platelet count

Evaluate the accuracy of a new test

- **Validating tests against a gold standard:**

- New tests should be validated by comparison against an established gold standard in an appropriate subjects

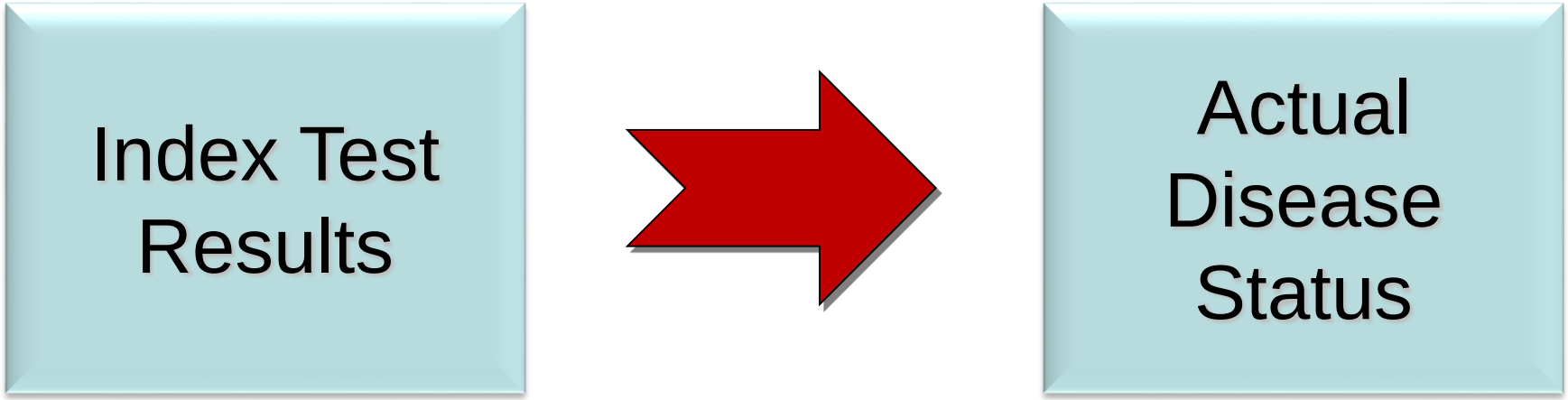
Binary Test Data Structure

	Case (Refernce test positive)	**Non-case (Reference test negative)**
Test Positive	True positive (a)	False positive (b)
Test Negative	False negative (c)	True negative (d)

- Measure of test performance
  - Sensitivity (true positive rate) ➔ a/(a+c)
  - Specificity (true negative rate) ➔ d/(b+d)
  - Positive predictive value ➔ a/(a+b)
  - Negative predictive value ➔ d/(c+d)

"Clinicians usually struggle with the interpretation of sensitivity and specificity, because positive/negative predictive value is the most straightforward measure .."

Forward Thinking

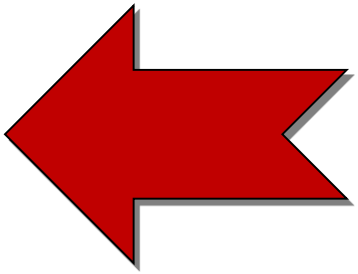


Positive predictive value
Negative predictive value

Influenced by Prevalence of Disease
Not comparable between studies

Reverse Thinking

Index Test Results

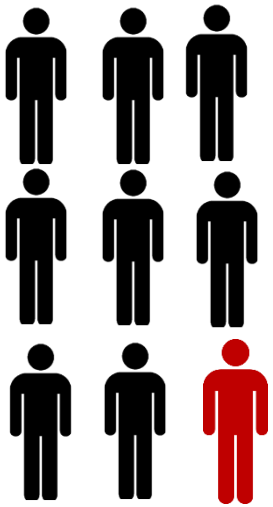


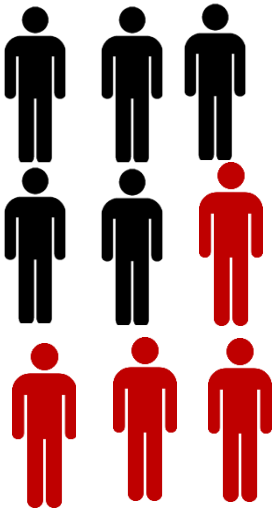
Actual Disease Status

Sensitivity
Specificity

Not Influenced by Prevalence of Disease
Comparable between studies

Same test in different populations

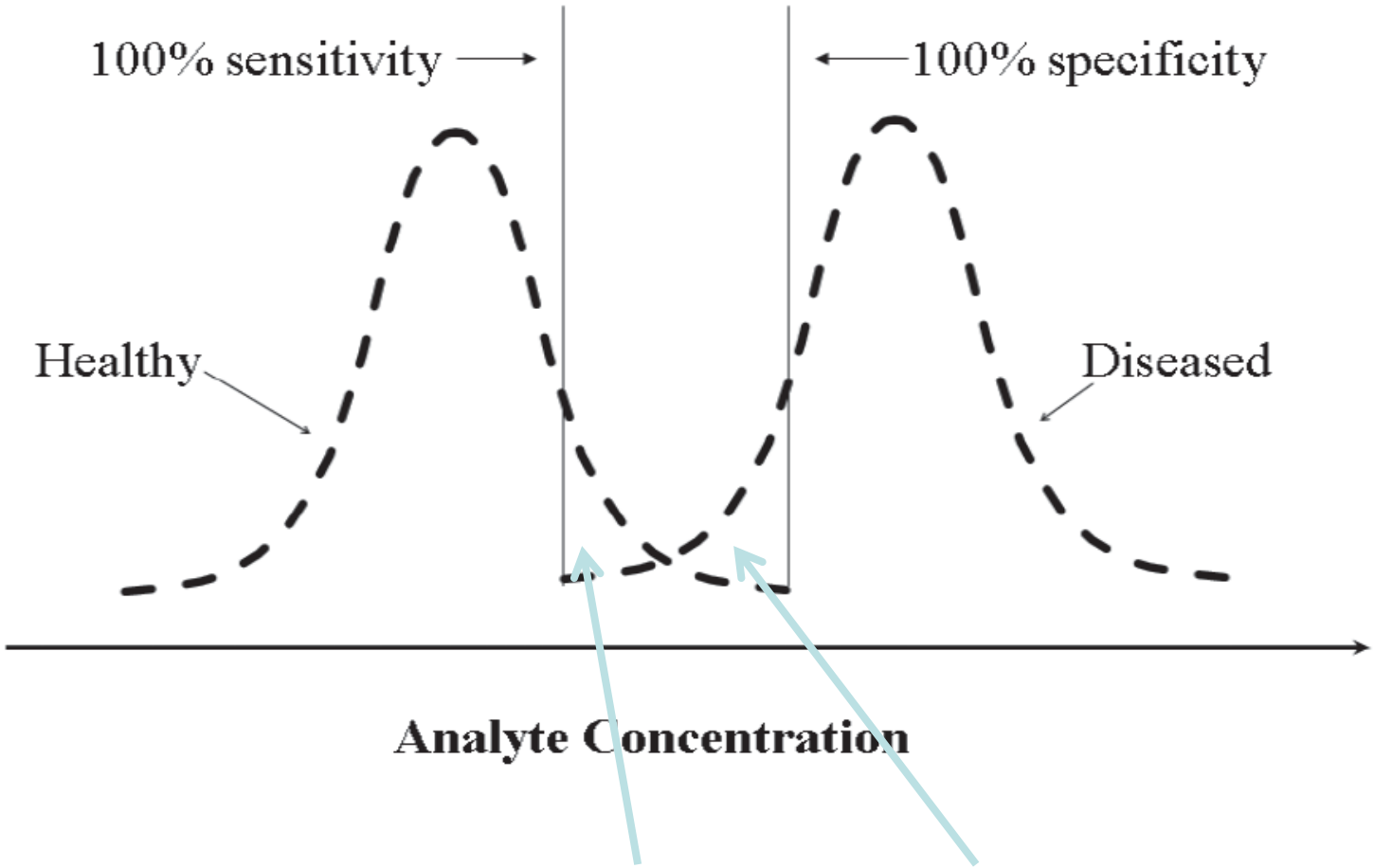




PPV low
Sen same

PPV high
Sen same

Determine the cutoff value



Sensitivity and specificity are negatively correlated, depending on the cutoff value selection

Choice of a cut-off point

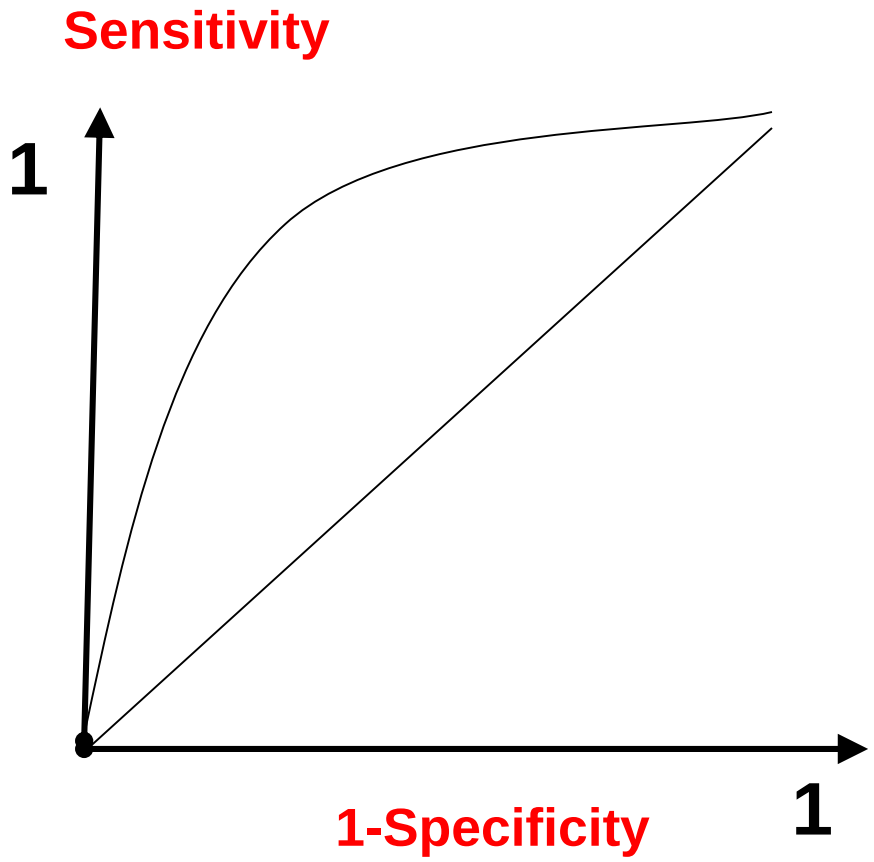
- If *false-positive* must be avoided, such as surgical decision, then the cutoff needs to be set to maximize the *specificity*

- If *false-negative* must be avoided, such as diagnosis of myocardial infarction, then the cutoff should be set to maximize the *sensitivity*

ROC curve

SGPT and Hepatitis

SGPT cutoff	Sen	Spe
< 50	95%	15%
100	80%	30%
150	70%	50%
200	60%	70%
250	30%	85%
≥300	10%	97%



ROC curve

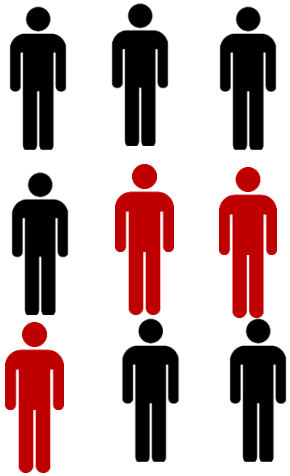
- Ccomplete description of performance
- Facilitate comparison and combination across studies of the same test
- Guide the choice of thresholds
- Enable comparisons between different non-binary tests

Outline

1. Evaluate the accuracy of a diagnostic test
2. Design of a diagnostic test study
3. Systematic review of diagnostic test studies
4. Quality assessment
5. Interpretation and clinical application

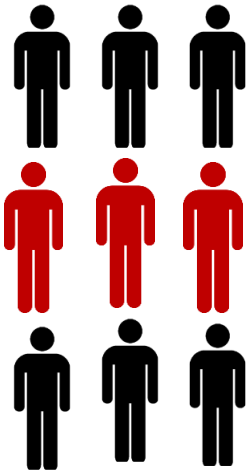
Cross sectional design

Index Test Results



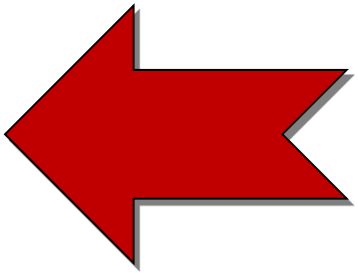
Sen: 67%
Spe: 83%

Reference standard

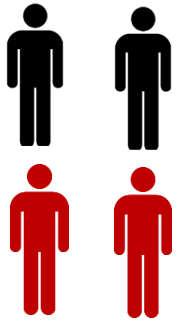


Case control design

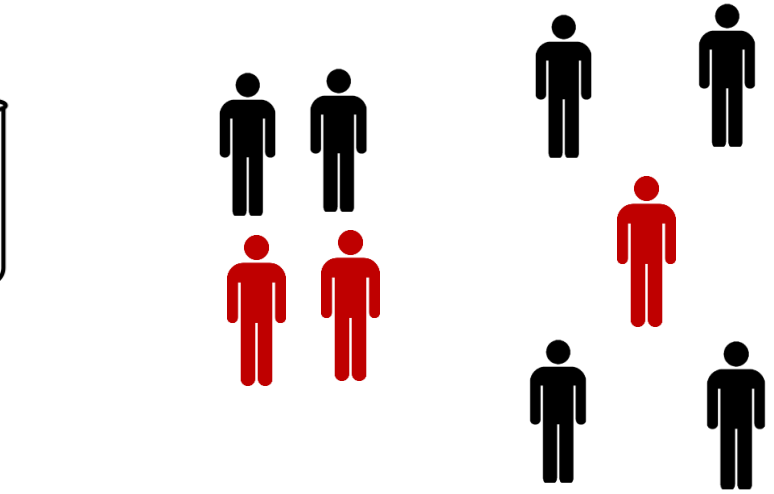
Index Test Results



Reference standard



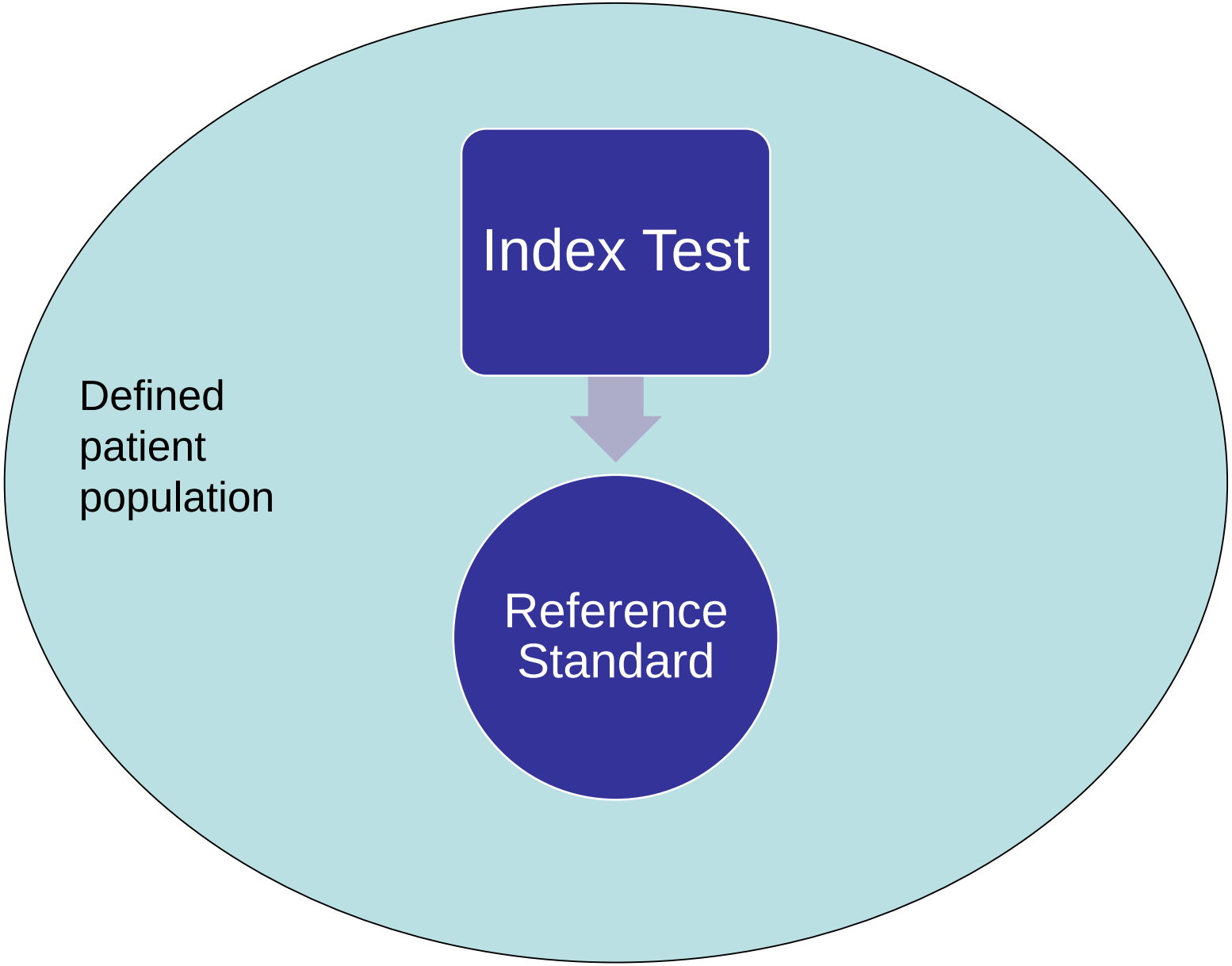
Sen: 100%
Spe: 100%

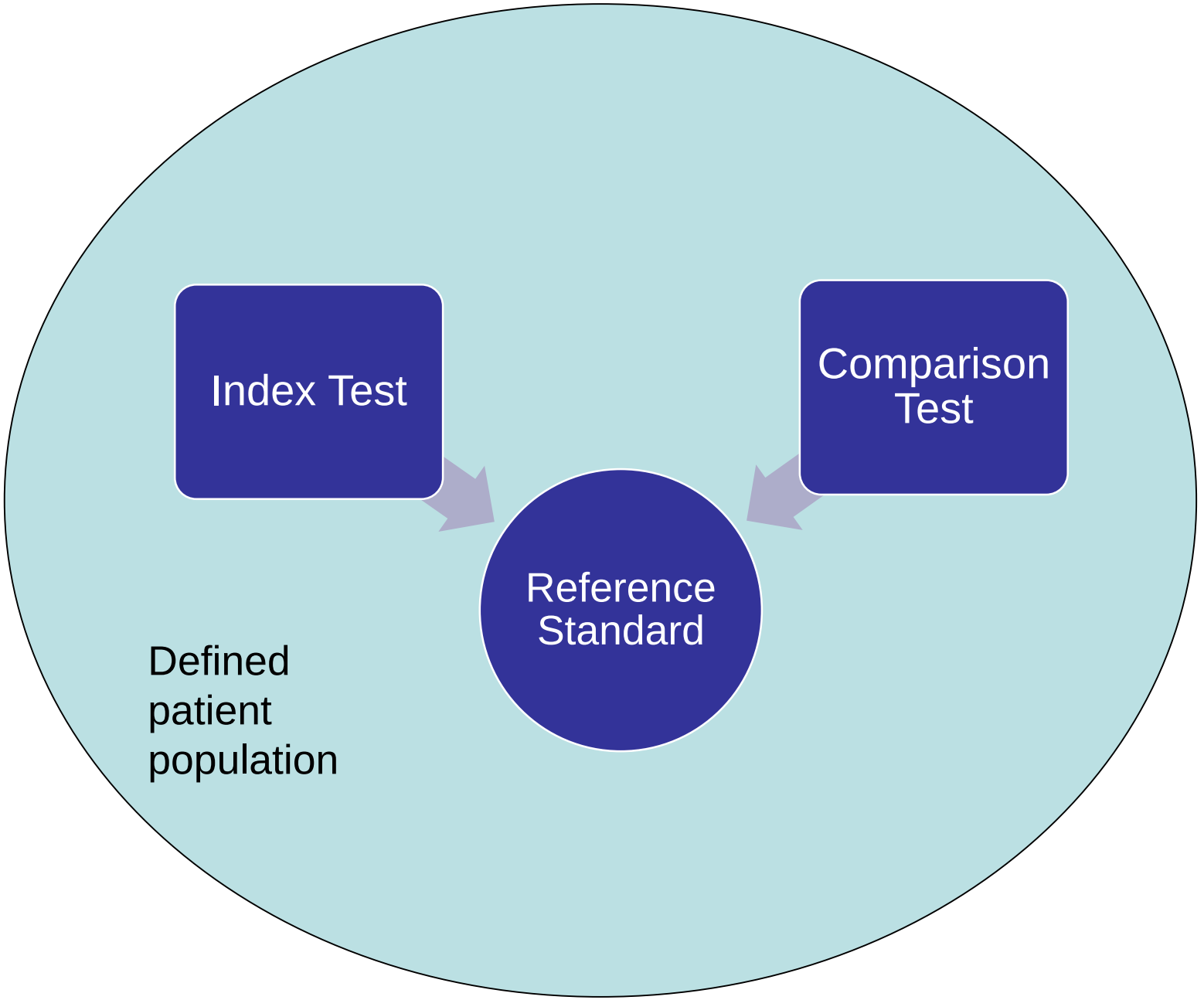


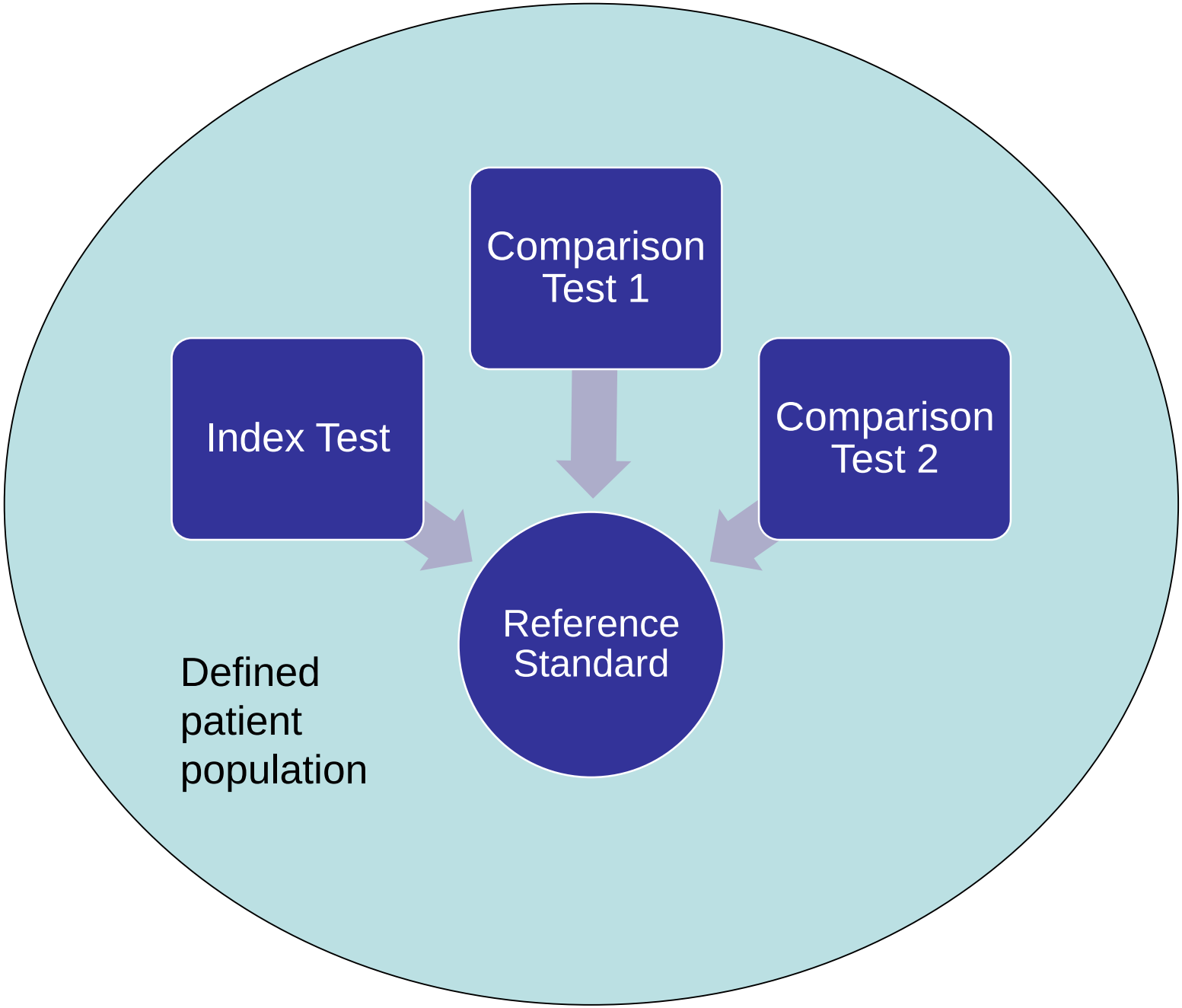
Case-control vs. Cross sectional

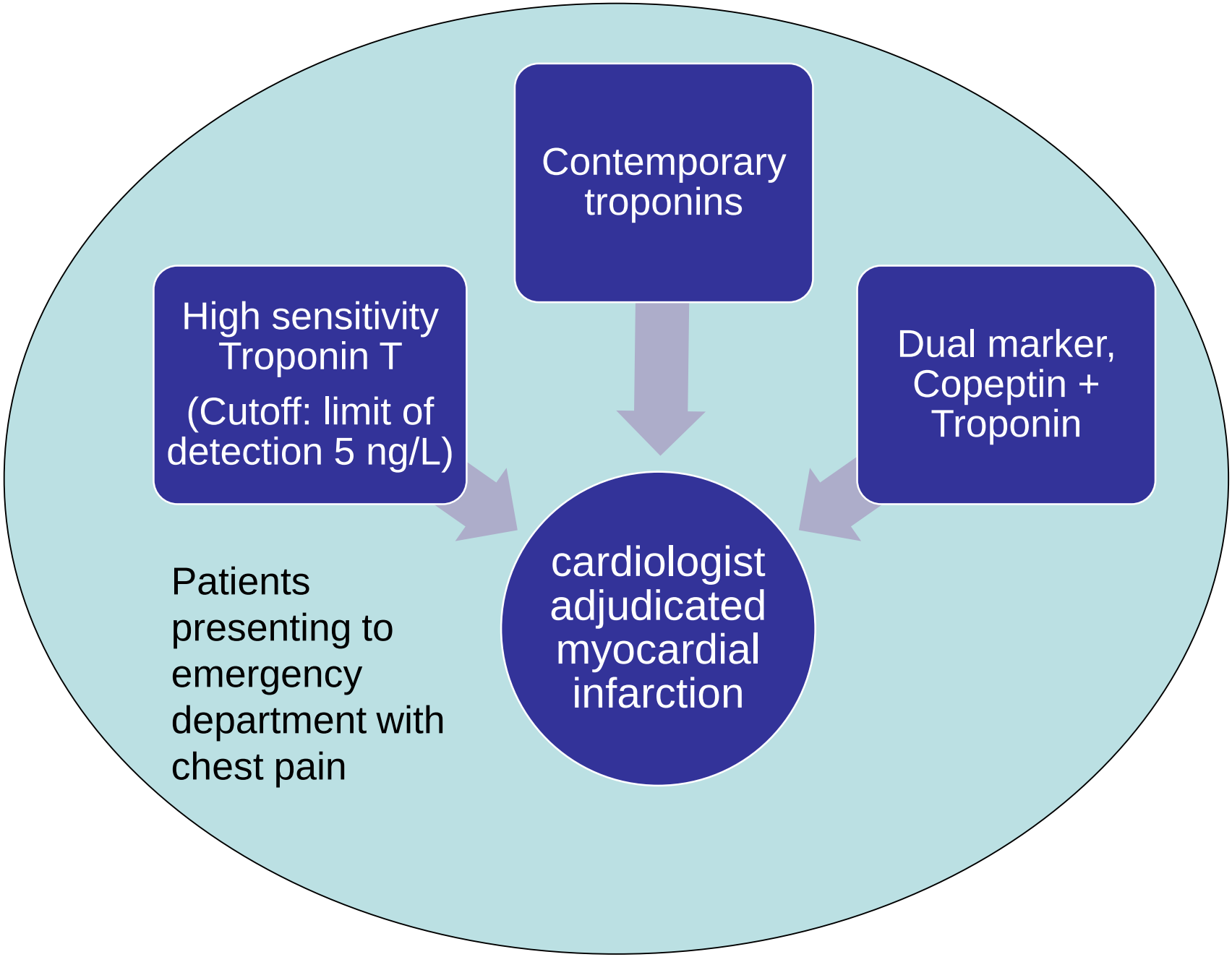
- Case-control design
  - Provide an indication of maximal accuracy of a test
  - Valuable in the technical validation
  - Prevalence or predictive values cannot be estimated
  - Not representative of accuracy in clinical practice

- Cross sectional design
  - Provide valid estimates of diagnostic accuracy in the real world settings
  - Prevalence or predictive values cannot be estimated


Index Test
Defined patient population
Reference Standard


Index Test
Comparison Test
Reference Standard
Defined patient population


Comparison Test 1
Index Test
Comparison Test 2
Reference Standard
Defined patient population


Contemporary troponins
High sensitivity Troponin T
(Cutoff: limit of detection 5 ng/L)
Dual marker, Copeptin + Troponin
cardiologist adjudicated myocardial infarction
Patients presenting to emergency department with chest pain

Outline

1. Evaluate the accuracy of a diagnostic test
2. Design of a diagnostic test study
3. Systematic review of diagnostic test studies
4. Quality assessment
5. Interpretation and clinical application

Systematic Review

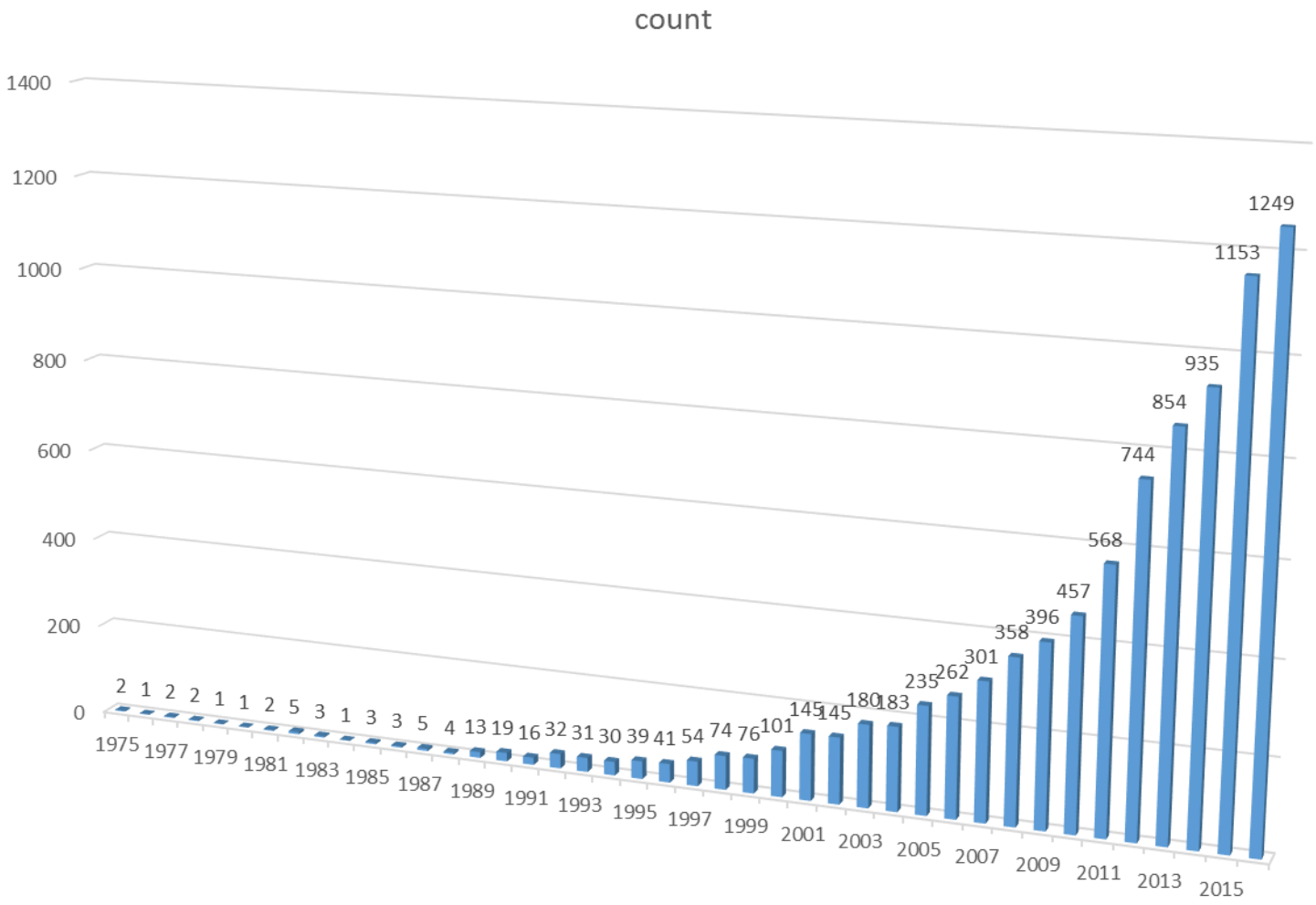
- Systematic approach
- Minimizing bias and random errors

- Comparison with narrative review
  - Complete collection evidence
  - Transparency of methods allowing replication
  - Less subjectivity

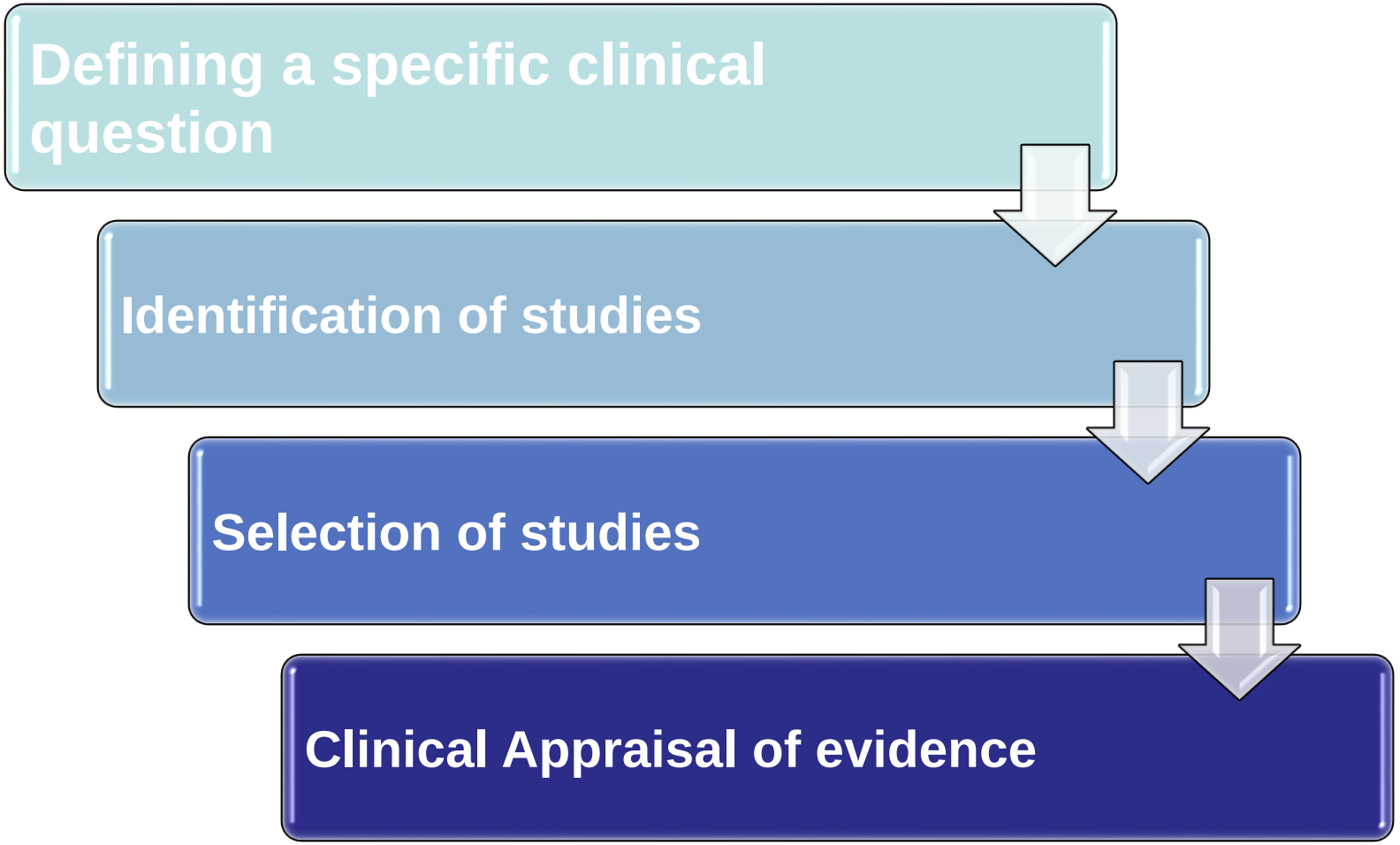
Aim

- To investigate whether a test is sufficiently specific or sensitive to fit its role in practice
- To compare the accuracy of two or more diagnostic tests
- To investigate where existing variation in results comes from

Number of diagnostic test systematic reviews searched by PubMed



Systematic Approach



Literature Search

- More difficult than searching for randomized trials
- No indexing term for a diagnostic study
- Medical Subheading (MeSH) terms “ sensitivity and specificity” can be used but may miss some studies
- Use broad term and manual screening reference lists

Review question for an efficacy study P.I.C.O. model

Population	• *Target population*
Intervention	• *Treatment group*
Comparison	• *Control group*
Outcome	• *Relative risks*

Review question for an diagnostic test P.I.C.O. model

Population	• *Target population*
Index test	• *New diagnostic test of interest*
Comparison	• *Conventional test for comparison*
Outcome	• *Accuracy measure: sensitivity or specificity*

High Sensitivity Troponin T for Early Diagnosis of Myocardial Infarction

Population	• ED patients with suspected MI
Test	• *High Sensitivity Troponin T test*
Comparison	• *Conventional Troponin Test*
Outcome	• *Accuracy measure: sensitivity or specificity*

Design Key Words

Population	• Chest pain/ discomfort • Emergency room/department
Index Test	• *High Sensitivity/sensitive Troponin T/Troponin*
Comparison	
Outcome	• *Acute coronary syndrome* • *Myocardial Infarction*

EMBASE search function for diagnostic test accuracy studies

- PICO tools: can modify the “Intervention” to “Index test”
- Study types filter : has a “diagnostic test accuracy study” filter
- These search tools will enhance the specificity of search results at the cost of reduced sensitivity (may miss some studies)

PICO Search

Note: Filling any search line is optional

Default search strategy

/mj /de /exp /br

Population

emergency ward /exp ▼ | + 4 synonyms:all ▼ | and ▼ | chest pain :all ▼ | or ▼ | chest dyscomfort :all ▼

Clear field

Intervention

high sensitivity troponin t /exp ▼ | or ▼ | high sensitive troponin :all ▼

Clear field

Outcome

acute coronary syndrome /exp ▼ | + 2 synonyms:all ▼ | or ▼ | heart infarction /exp ▼ | + 19 synonyms:all ▼

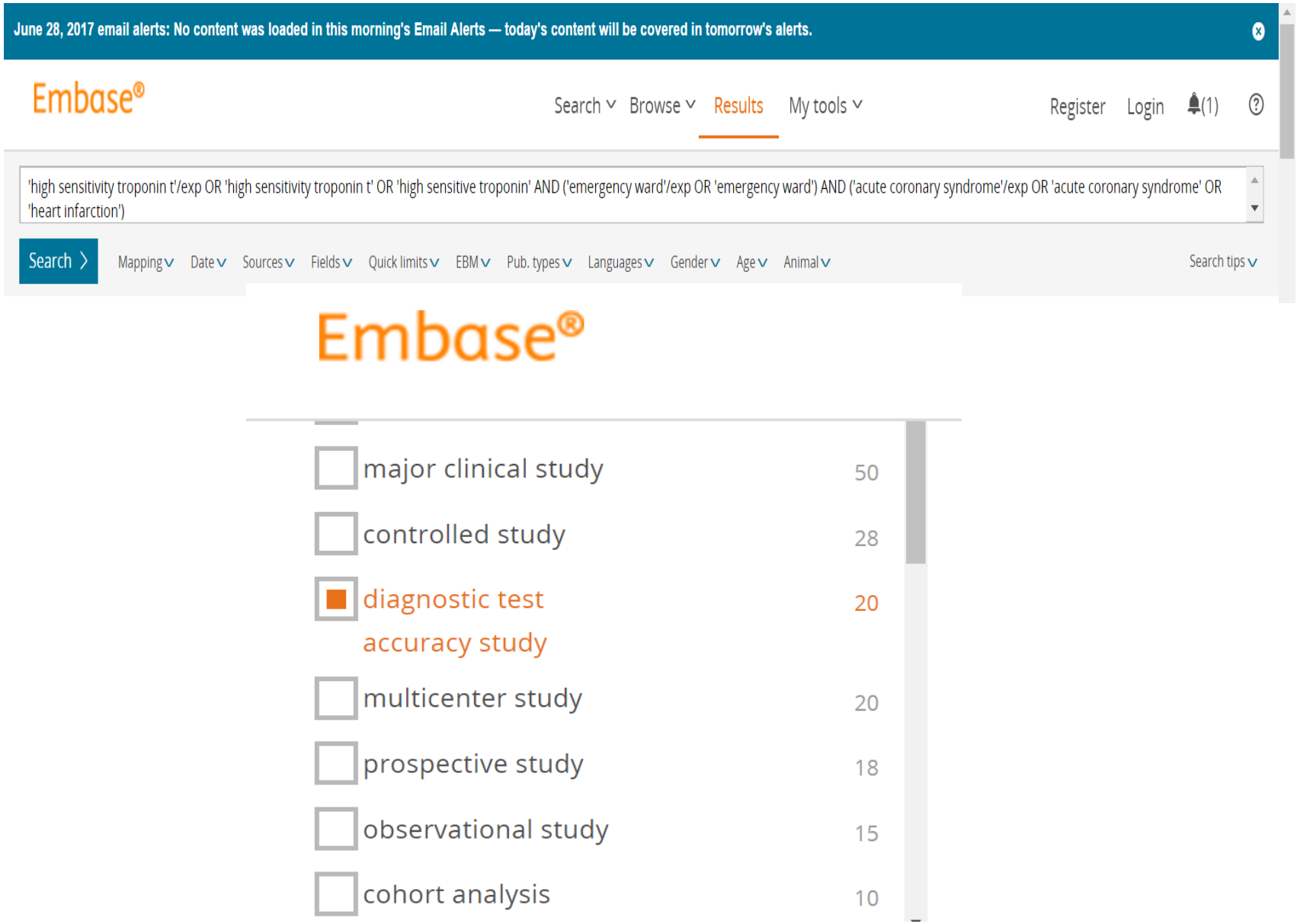
Clear field

Study design (or miscellaneous)

e.g. randomized controlled trial

Reset query | Info

Show 56 results


June 28, 2017 email alerts: No content was loaded in this morning's Email Alerts — today's content will be covered in tomorrow's alerts.
Embase®
Search
Browse
Results
My tools
Register
Login
(1)
'high sensitivity troponin t'/exp OR 'high sensitivity troponin t' OR 'high sensitive troponin' AND ('emergency ward'/exp OR 'emergency ward') AND ('acute coronary syndrome'/exp OR 'acute coronary syndrome' OR 'heart infarction')
Search
Mapping
Date
Sources
Fields
Quick limits
EBM
Pub. types
Languages
Gender
Age
Animal
Search tips
Embase®
major clinical study 50
controlled study 28
diagnostic test accuracy study 20
multicenter study 20
prospective study 18
observational study 15
cohort analysis 10

Outline

1. Evaluate the accuracy of a diagnostic test
2. Design of a diagnostic test study
3. Systematic review of diagnostic test studies
4. Quality assessment
5. Interpretation and clinical application

Study Quality Assessment QUADAS-2

- QUADAS-2
  - Quality Assessment of Diagnostic Accuracy Studies-2 checklist
  - Assesses the quality of studies over four domains

QUADAS-2

Patient selection	• *Study design* • *Sample selection*
Index test	• *Blinding* • *Threshold effect*
Reference standard	• *Incorporation bias* • *Independence/blinding*
Patient Flow and Timing	• *Appropriate time interval* • *Verification bias*

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Describe methods of patient selection:

- ❖ Was a consecutive or random sample of patients enrolled? Yes/No/Unclear
- ❖ Was a case-control design avoided? Yes/No/Unclear
- ❖ Did the study avoid inappropriate exclusions? Yes/No/Unclear

Could the selection of patients have introduced bias? **RISK: LOW/HIGH/UNCLEAR**

B. Concerns regarding applicability

Describe included patients (prior testing, presentation, intended use of index test and setting):

Is there concern that the included patients do not match the review question? **CONCERN: LOW/HIGH/UNCLEAR**

Whiting PF *Ann Intern Med* 2011

DOMAIN 2: INDEX TEST(S)

If more than one index test was used, please complete for each test.

A. Risk of Bias

Describe the index test and how it was conducted and interpreted:

- Were the index test results interpreted without knowledge of the results of the reference standard? Yes/No/Unclear
- If a threshold was used, was it pre-specified? Yes/No/Unclear

Could the conduct or interpretation of the index test have introduced bias? **RISK: LOW /HIGH/UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question? **CONCERN: LOW /HIGH/UNCLEAR**

Whiting PF *Ann Intern Med* 2011

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Describe the reference standard and how it was conducted and interpreted:

- ❖ Is the reference standard likely to correctly classify the target condition? Yes/No/Unclear
- ❖ Were the reference standard results interpreted without knowledge of the results of the index test? Yes/No/Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: LOW /HIGH/UNCLEAR**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question? **CONCERN: LOW /HIGH/UNCLEAR**

Whiting PF *Ann Intern Med* 2011

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):

Describe the time interval and any interventions between index test(s) and reference standard:

- Was there an appropriate interval between index test(s) and reference standard? Yes/No/Unclear
- Did all patients receive a reference standard? Yes/No/Unclear
- Did patients receive the same reference standard? Yes/No/Unclear
- Were all patients included in the analysis? Yes/No/Unclear

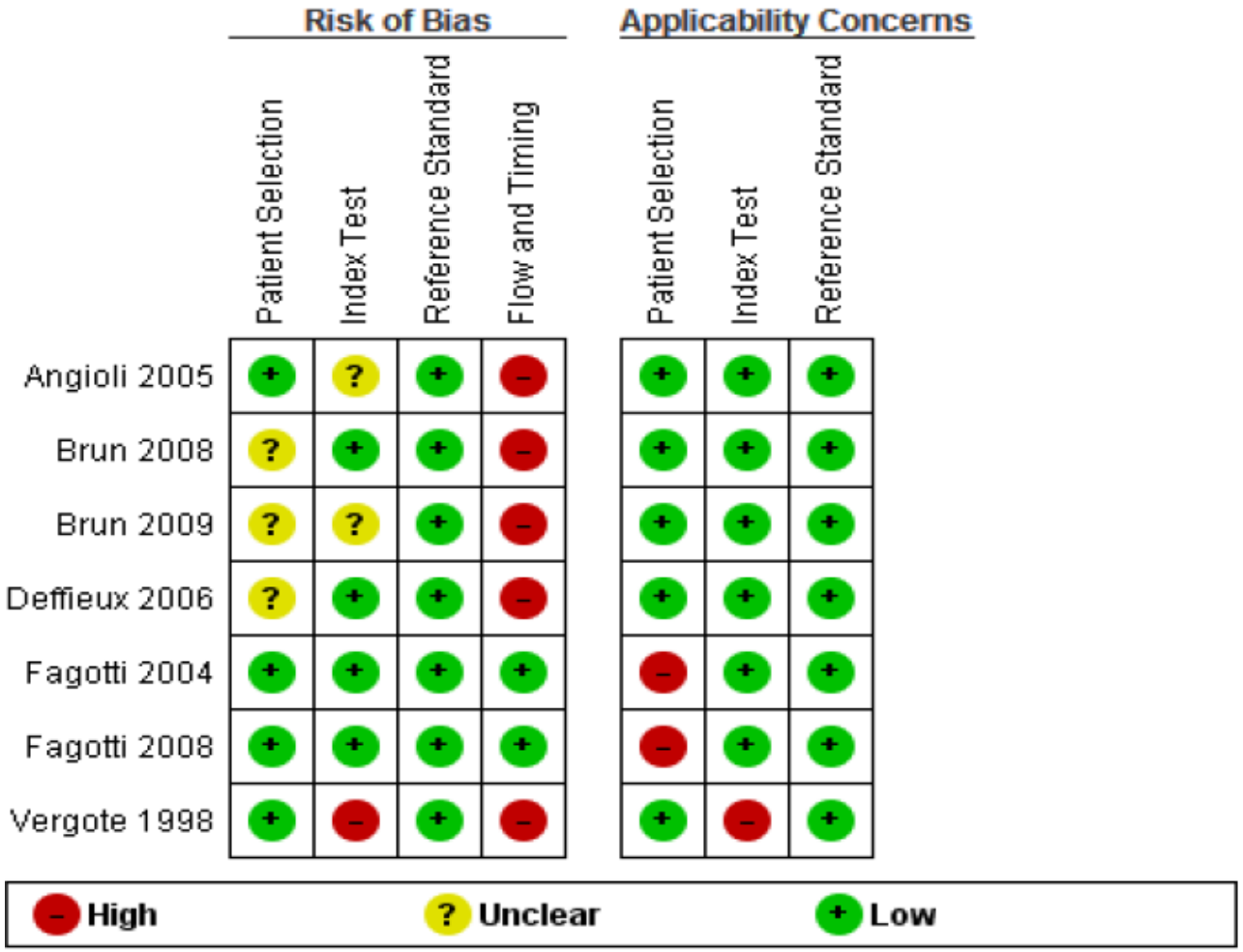
Could the patient flow have introduced bias? **RISK: LOW /HIGH/UNCLEAR**

Whiting PF *Ann Intern Med* 2011

Common Study Design Flaws

- Case-control design
  - Exaggerate the accuracy of the test
- Incorporation bias
  - The reference standard includes the index test
- Verification bias
  - Not all participants received same reference standard evaluation

Presentation of Quadas-2 results



Outline

1. Evaluate the accuracy of a diagnostic test
2. Design of a diagnostic test study
3. Systematic review of diagnostic test studies
4. Quality assessment
5. Interpretation and clinical application

Principle of Interpretation

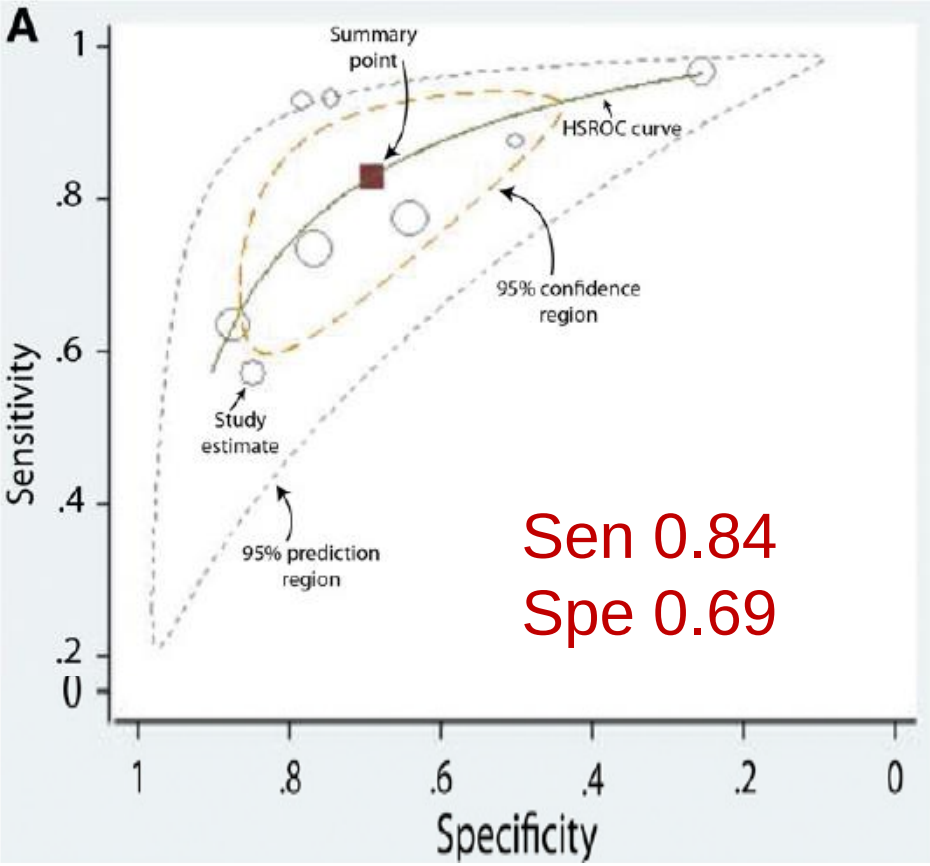
- The clinical meaning of the estimated sensitivity and specificity is usually vague

- Interpret the potential consequences of a positive test result and a negative test result in the clinical practice

Background

- Fever is a very common reason for pediatric visits to the emergency department (ED).
- Of these, about 8% may have an occult serious bacterial infection, such as bacteremia, urinary tract infection (UTI), pneumonia, or meningitis.
- Procalcitonin (PCT) has been shown to distinguish bacterial from viral infections

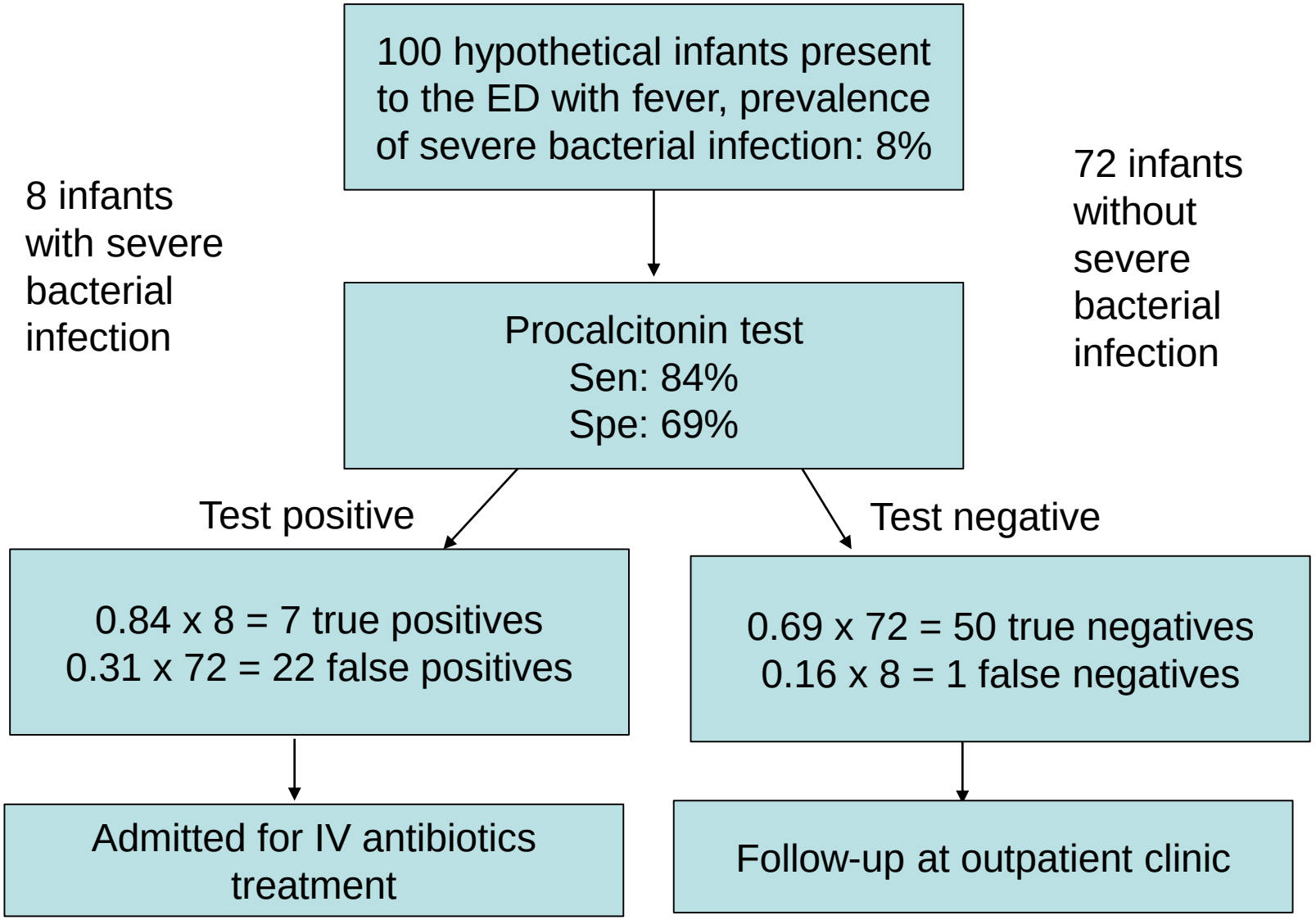
Comparison of the Test Characteristics of Procalcitonin to C-Reactive Protein and Leukocytosis for the Detection of Serious Bacterial Infections in Children Presenting With Fever Without Source: A Systematic Review and Meta-analysis



Summary ROC

Procalcitonin test AUC: 0.85

Lee CC Ann Emerg Med. 2012 Nov;60(5):591-600



Lee CC et al. *Annals Emerg 2012*

Conclusion

- Systematic reviews of diagnostic test accuracy summarize the accuracy, e.g. the sensitivity and specificity, of diagnostic tests in a systematic and transparent way.