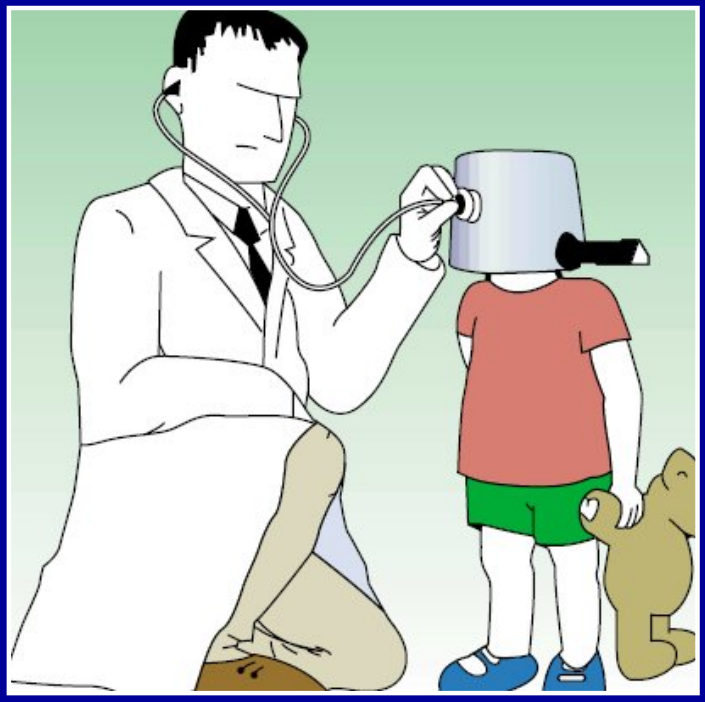
Critical appraisal of a diagnostic study

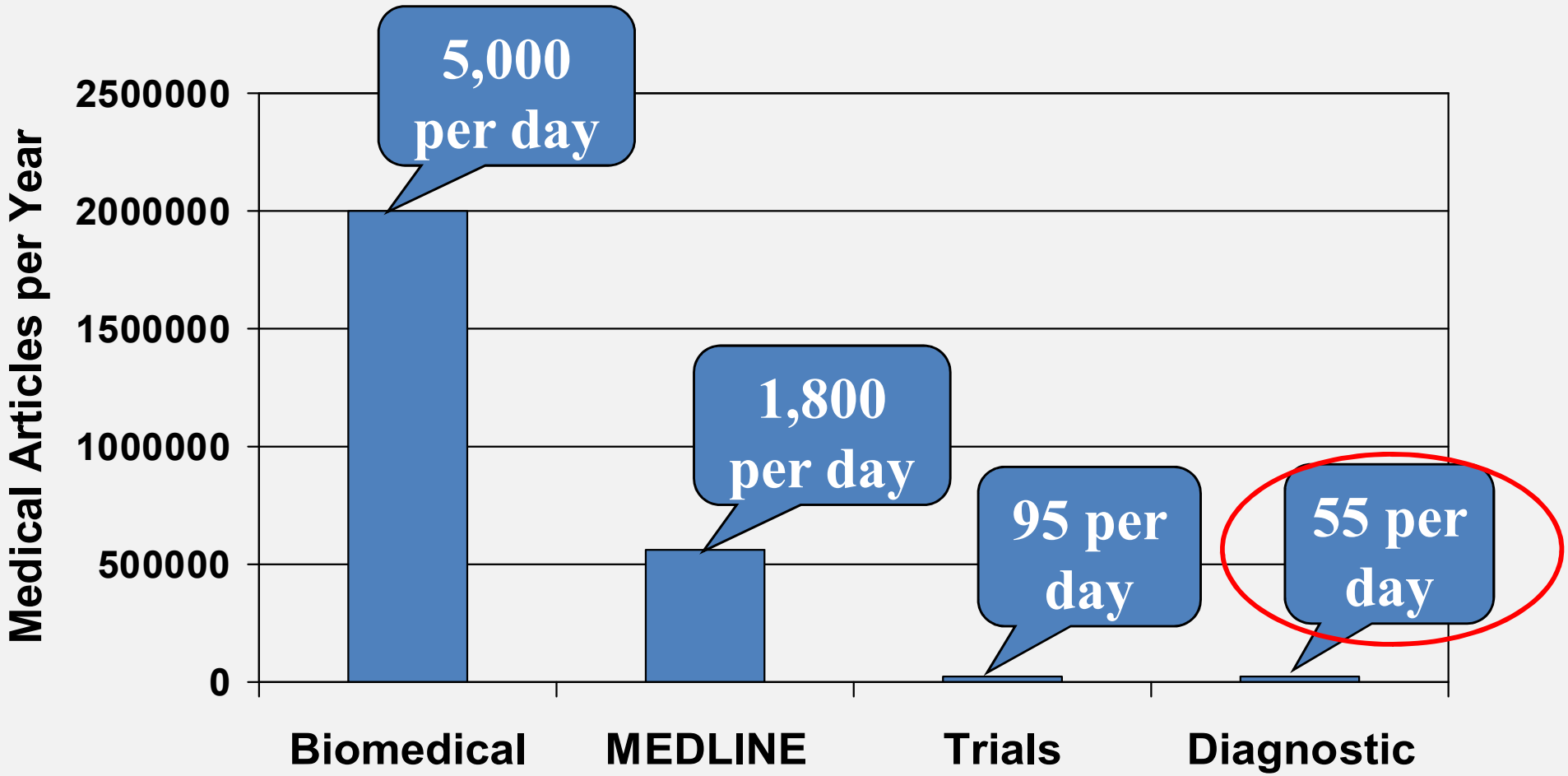


Samir Haffar M.D.

Assistant Professor of Gastroenterology

Al-Mouassat University Hospital – Damascus – Syria

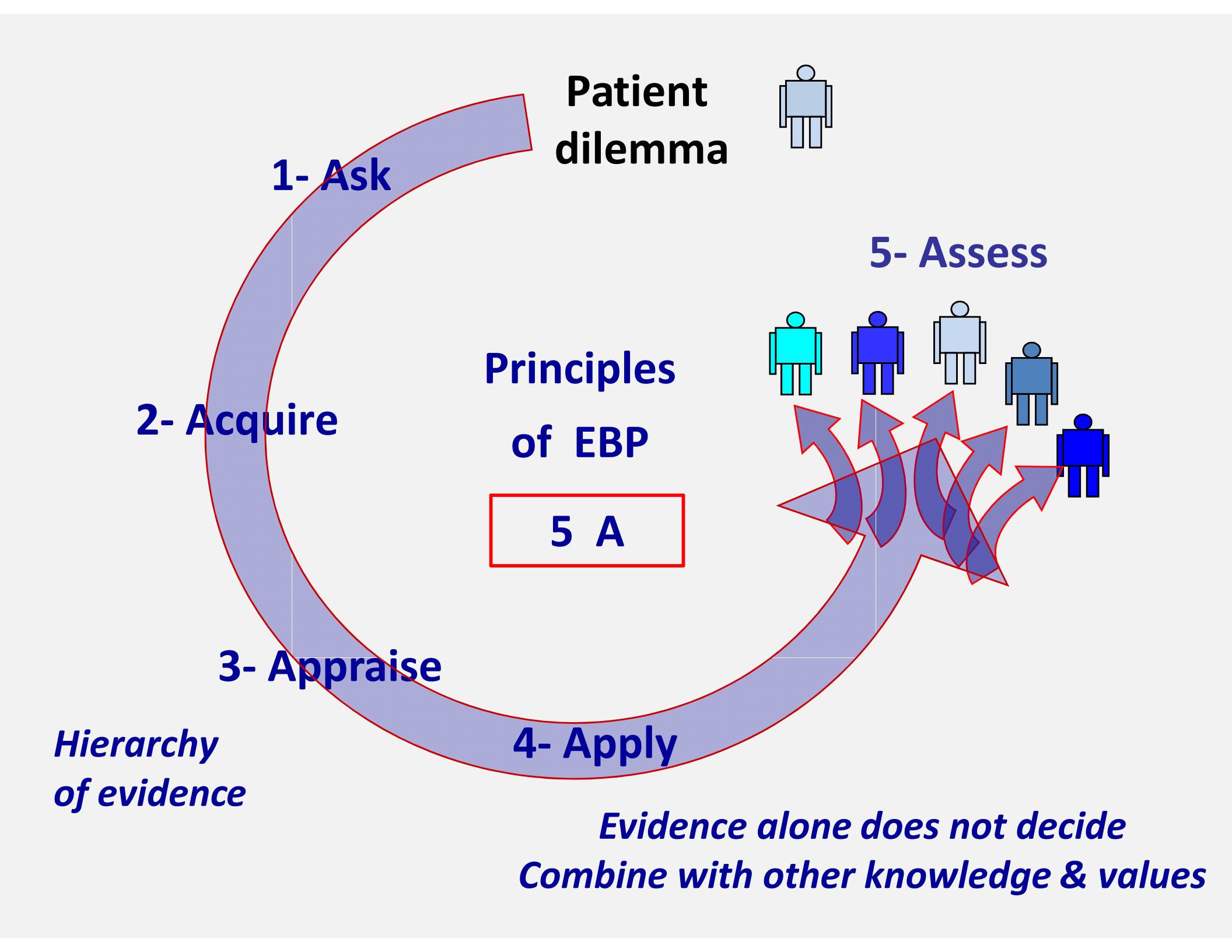
The amount of medical literature



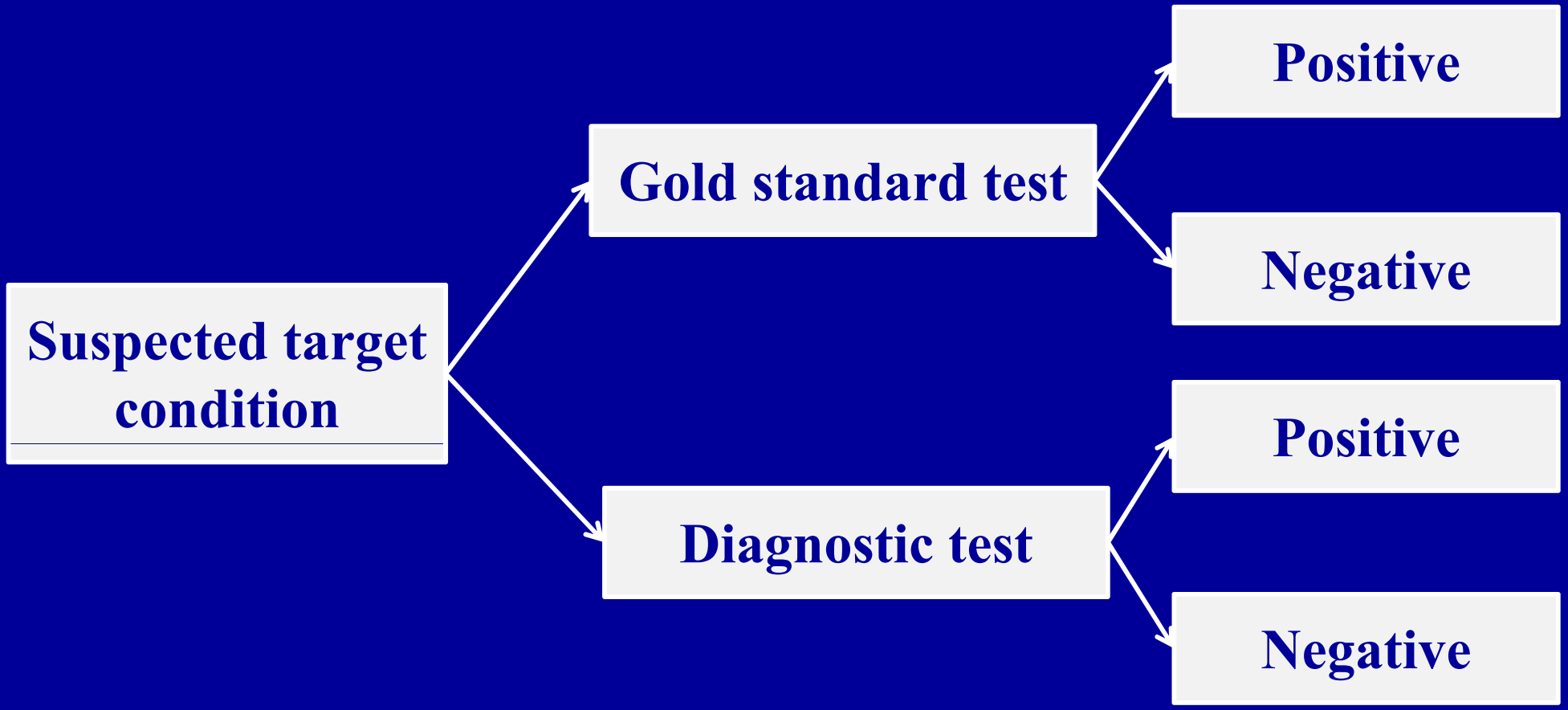
Glasziou P, Del Mar C. Evidence based practice workbook. Blackwell Publishing, 2nd edition, 2007.

Question type	Which study design will best answer my question?
Intervention	• RCT
Aetiology & risk factors	• Cohort or case control
Diagnosis	• Cross-sectional
Prognosis & incidence	• Cohort study
Prevalence	• Cross-sectional

In each case, SR of all available studies better than individual study


Patient dilemma
1- Ask
2- Acquire
3- Appraise
4- Apply
5- Assess
Principles of EBP
5 A
Hierarchy of evidence
Evidence alone does not decide
Combine with other knowledge & values

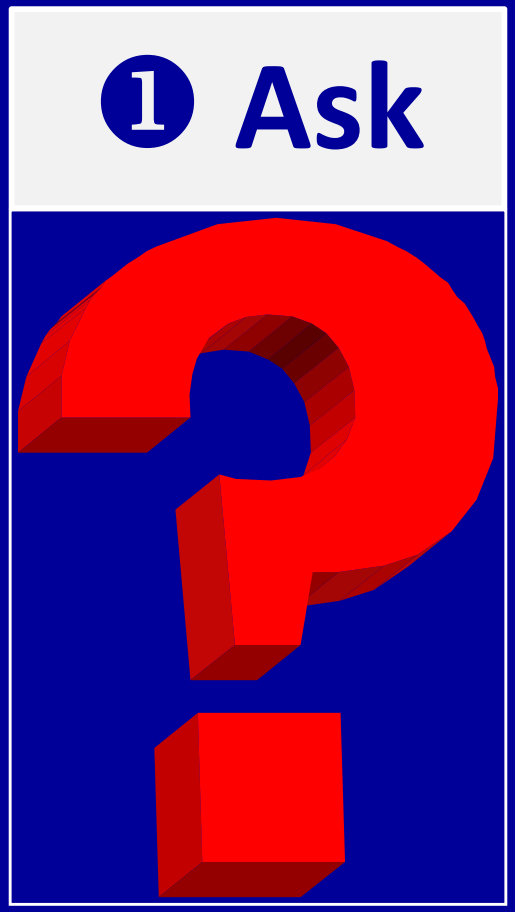
Structure for a study of diagnostic test



Accuracy of diagnostic test compared to gold standard

Guyatt G et all. Users' guides to medical literature: manual for EBP. McGraw-Hill, New York, USA, 2nd edition, 2008.

Steps of EBM



Clinical history

- 75-year-old woman with community-acquired pneumonia
  She responds nicely to appropriate antibiotics
  Hg:10 g/dL – MCV 80 – Blood smear: hypochromia
  Otherwise well – No incriminating medications
- Hg 10.5 g/dL 6 months ago – Ferritin 40 µg/L
  Never investigated before
- How to interpret ferritin result?
  How precise & accurate ferritin in diagnosis of IDA*?

* IDA: Iron Deficiency Anemia

Key components of your clinical question

Concept of PICO

P	**Patient**	Elderly patient with IDA
I	**Intervention**	Bone marrow aspirates
C	**Comparaison**	Ferritin
O	**Outcome**	Accuracy (**Sn** – **Sp** – **PPV** – **NPV** – **LR**)

* IDA: Iron Deficiency Anemia

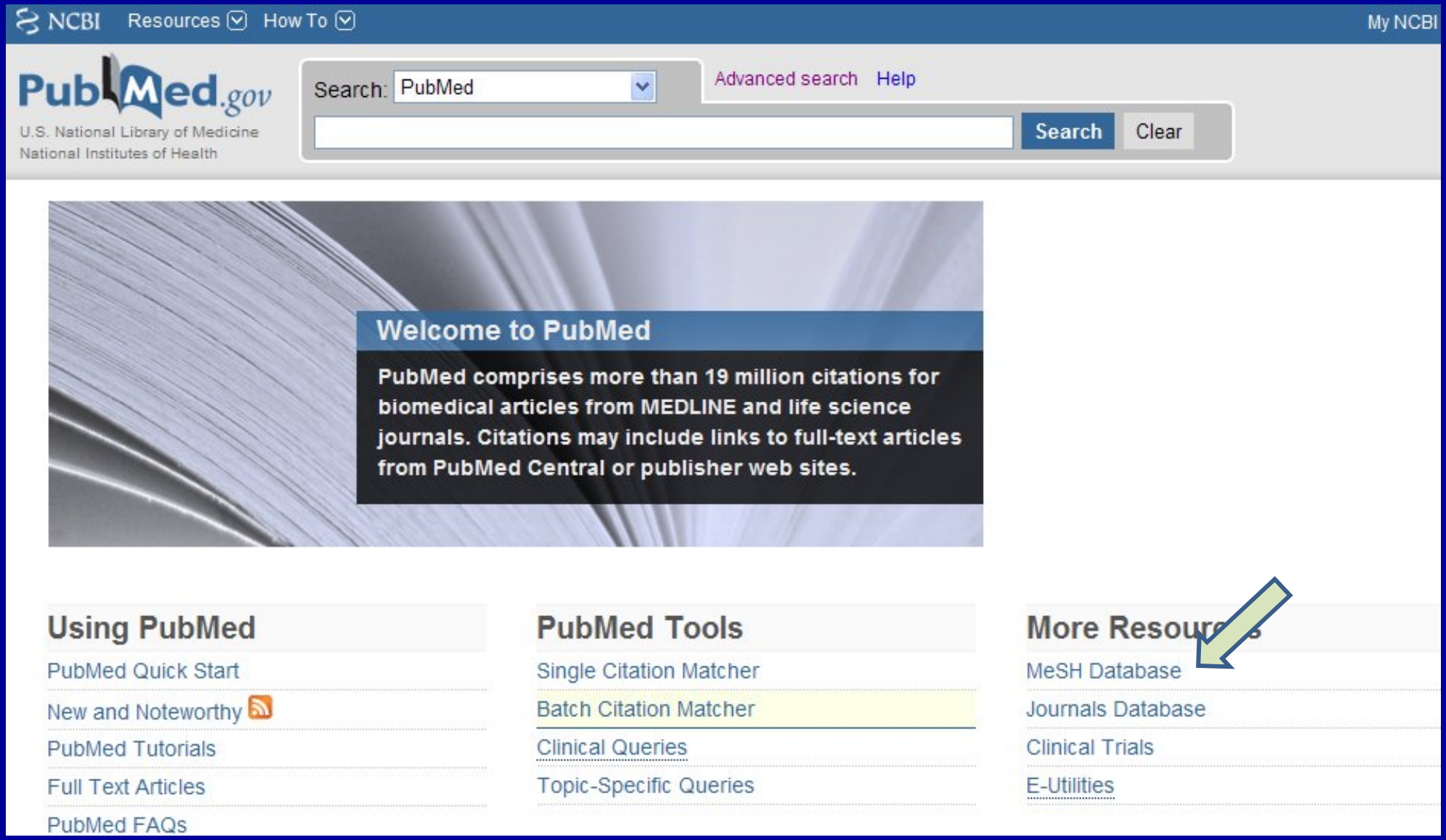
Formulation of the relevant question

In an elderly woman with hypochromic microcytic anemia, can a low serum ferritin level diagnose iron deficiency anemia?

Steps of EBM

❷ Acquire

PubMed



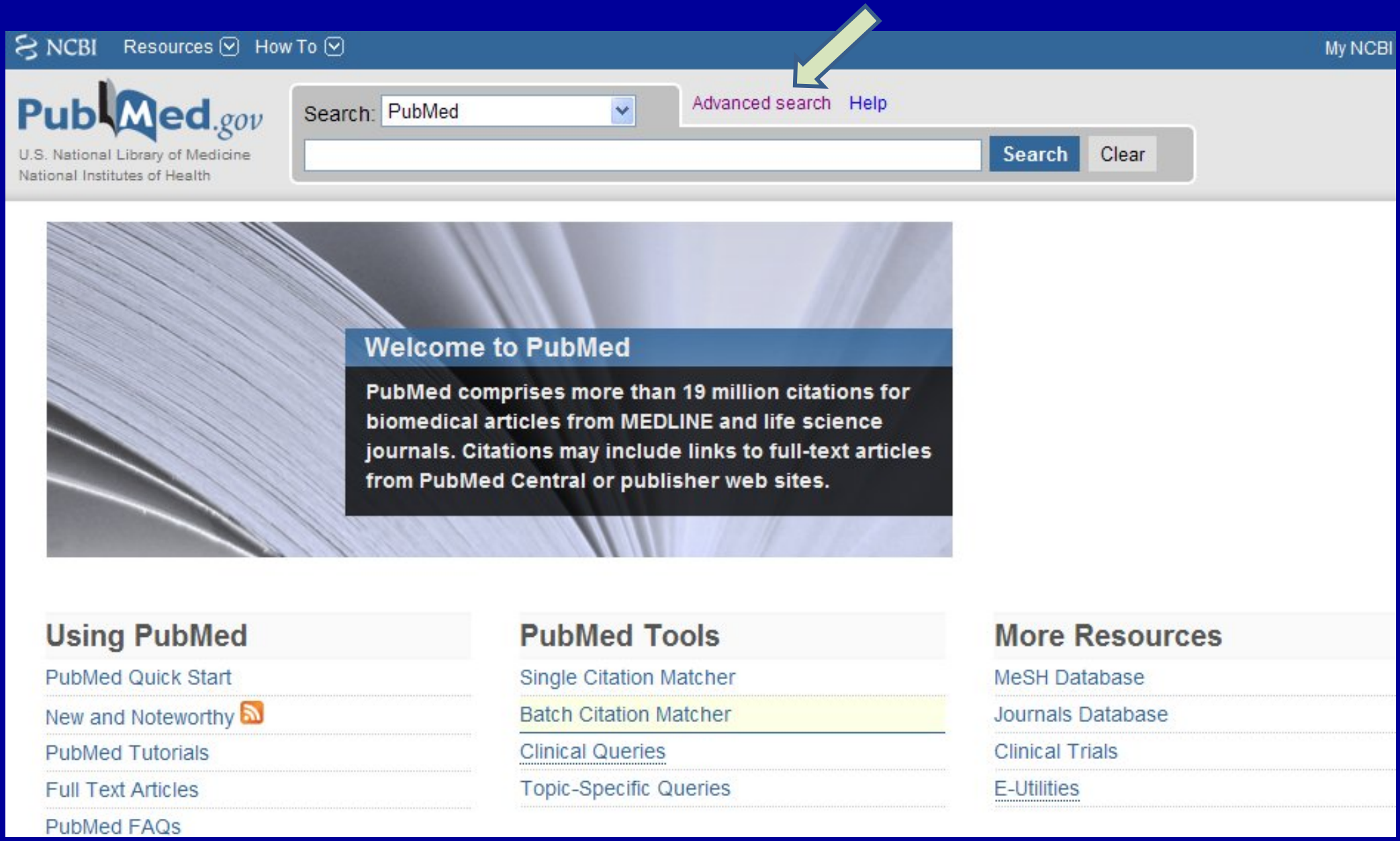
MeSH: **Me**dical **S**ubject **H**eadings in PubMed

PubMed translation of query into search terms

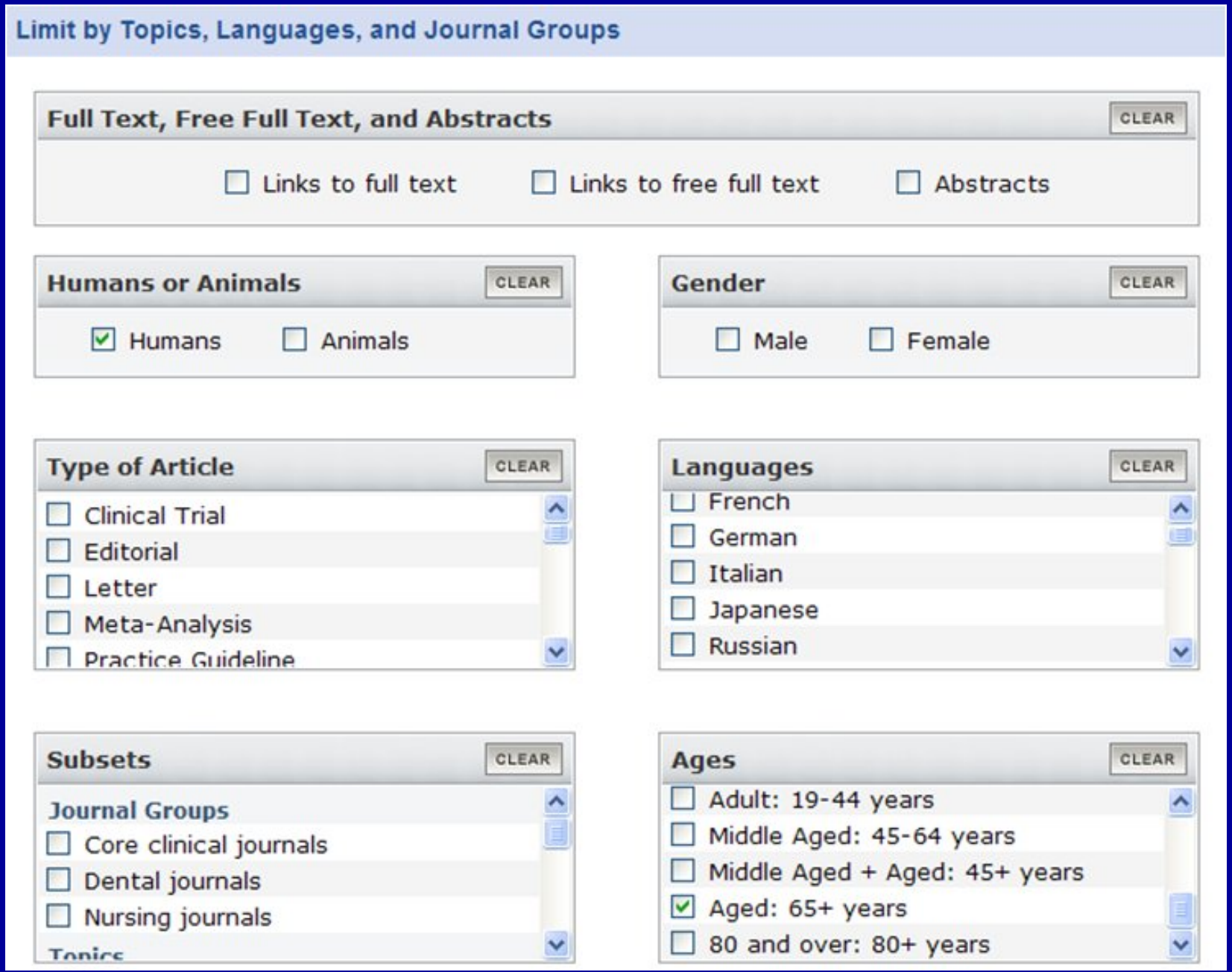
PICO	Element	Search terms for PubMed
P	Elderly Iron deficiency anemia	Limits to [aged: + 65 years] “iron deficiency anemia” [MeSH term]
I	Bone marrow aspirates	“bone marrow examination” [MeSH term]
C	Ferritin	“ferritins” [MeSH term]
O	Accuracy	“sensitivity” [MeSH term] “diagnosis” [MeSH term]

* **MeSH**: **Me**dical **S**ubject **H**eadings in PubMed

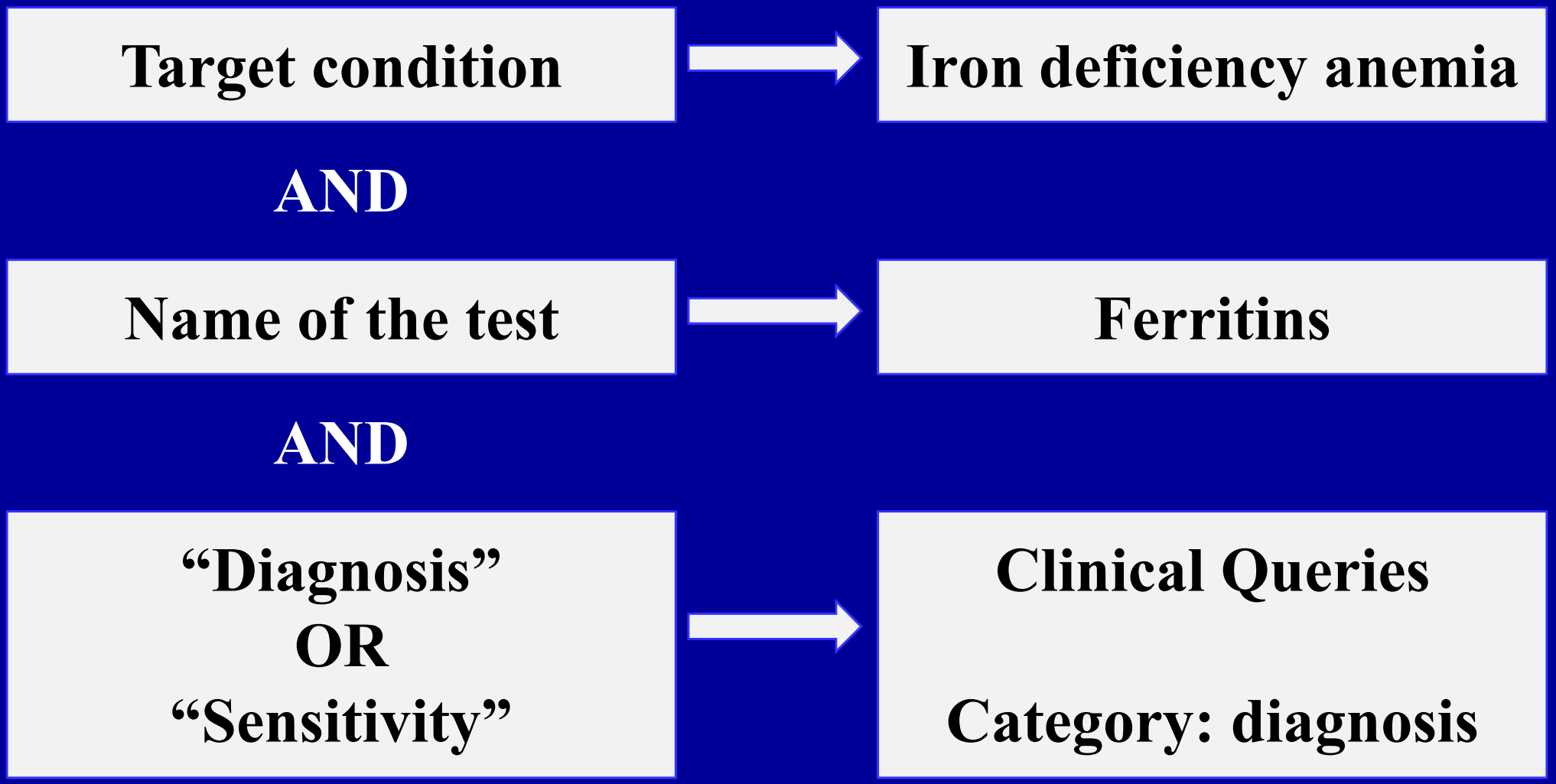
“Limits” link


NCBI
Resources
How To
My NCBI
PubMed.gov
U.S. National Library of Medicine
National Institutes of Health
Search: PubMed
Advanced search
Help
Search
Clear
Welcome to PubMed
PubMed comprises more than 19 million citations for biomedical articles from MEDLINE and life science journals. Citations may include links to full-text articles from PubMed Central or publisher web sites.
Using PubMed
PubMed Quick Start
New and Noteworthy
PubMed Tutorials
Full Text Articles
PubMed FAQs
PubMed Tools
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Topic-Specific Queries
More Resources
MeSH Database
Journals Database
Clinical Trials
E-Utilities

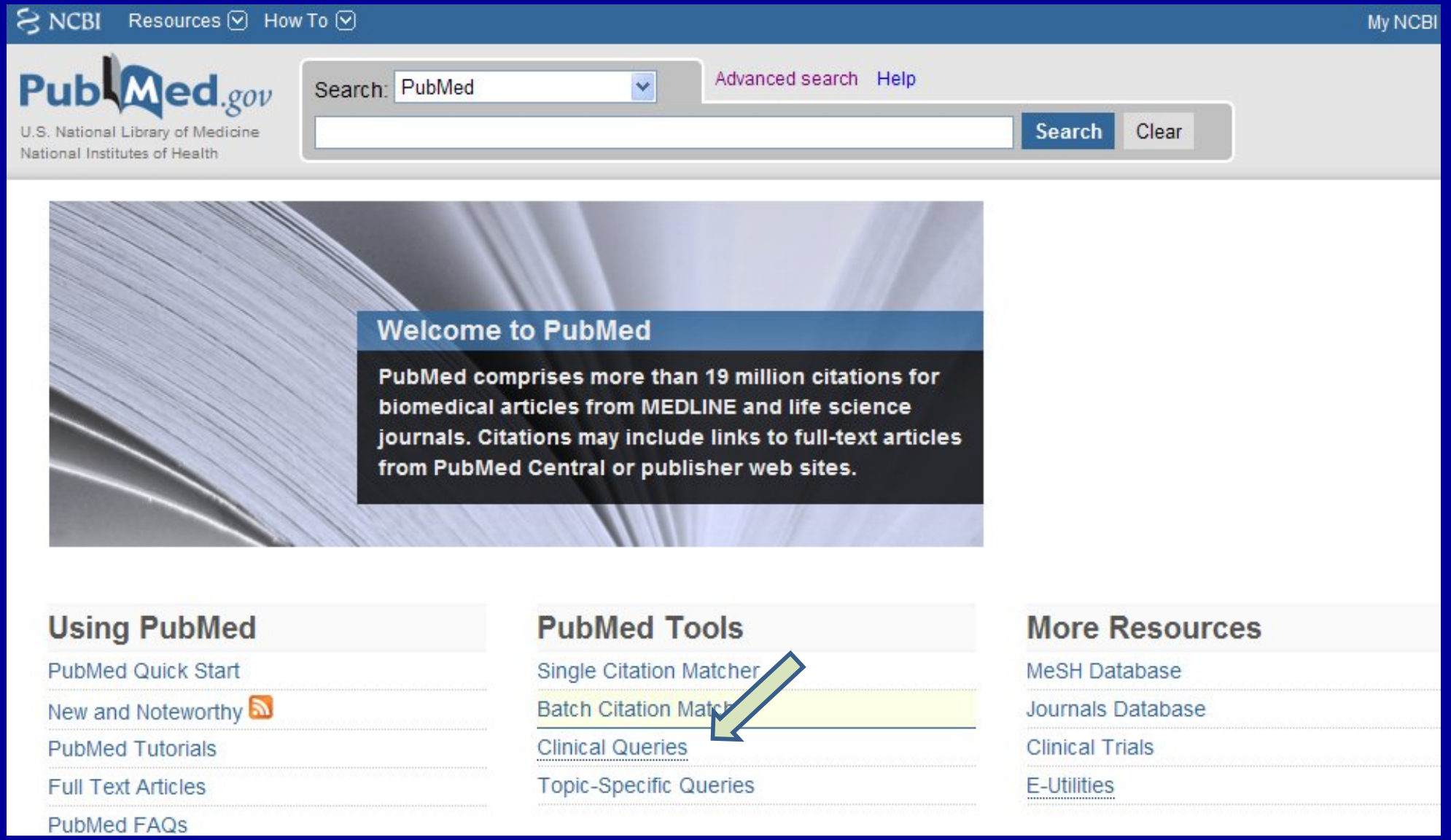
“Limits” link


Limit by Topics, Languages, and Journal Groups
Full Text, Free Full Text, and Abstracts
CLEAR
Links to full text
Links to free full text
Abstracts
Humans or Animals
CLEAR
Humans
Animals
Gender
CLEAR
Male
Female
Type of Article
CLEAR
Clinical Trial
Editorial
Letter
Meta-Analysis
Practice Guideline
Languages
CLEAR
French
German
Italian
Japanese
Russian
Subsets
CLEAR
Journal Groups
Core clinical journals
Dental journals
Nursing journals
Topics
Ages
CLEAR
Adult: 19-44 years
Middle Aged: 45-64 years
Middle Aged + Aged: 45+ years
Aged: 65+ years
80 and over: 80+ years

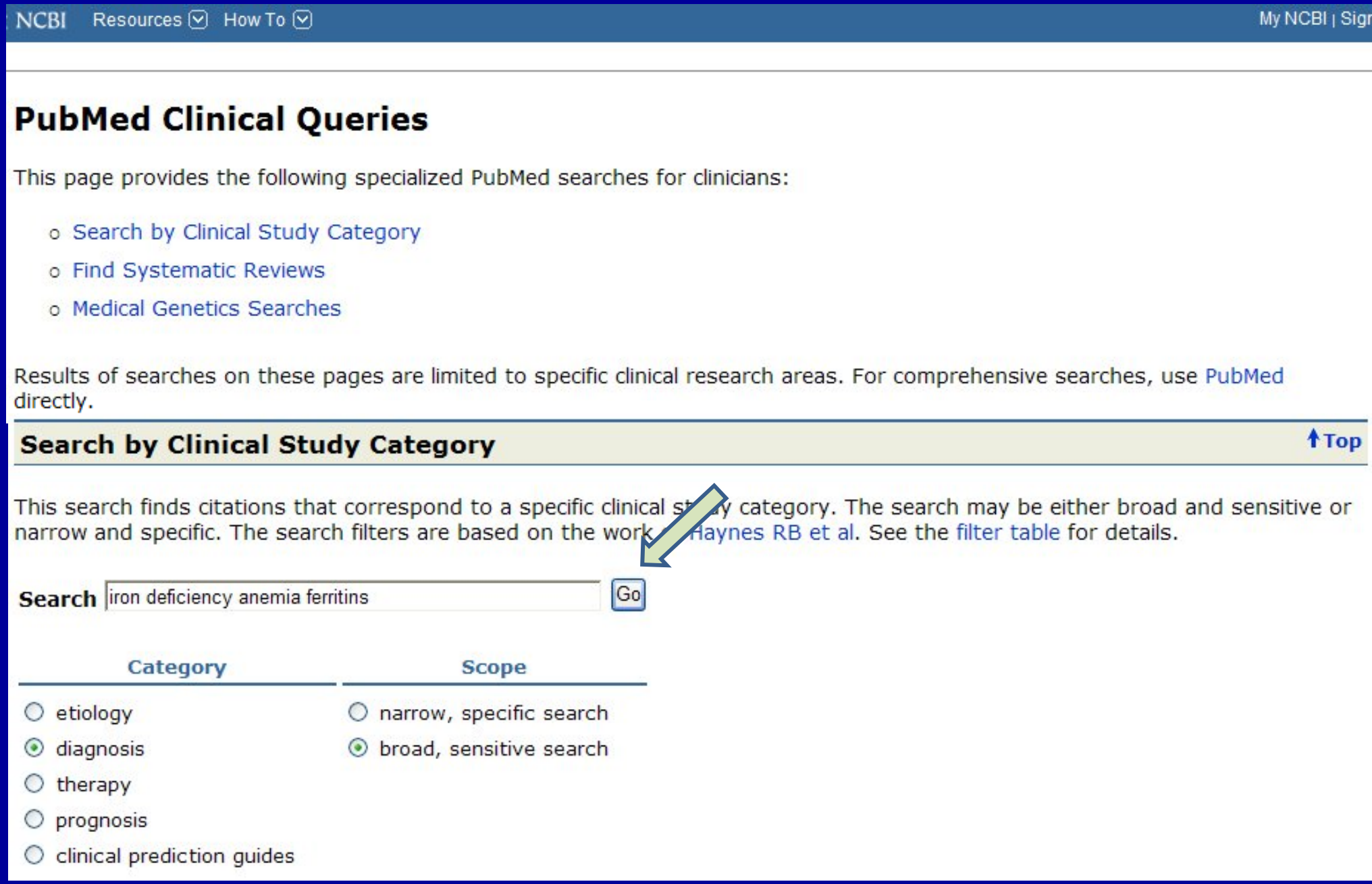
Terms used to search a diagnostic test in PubMed

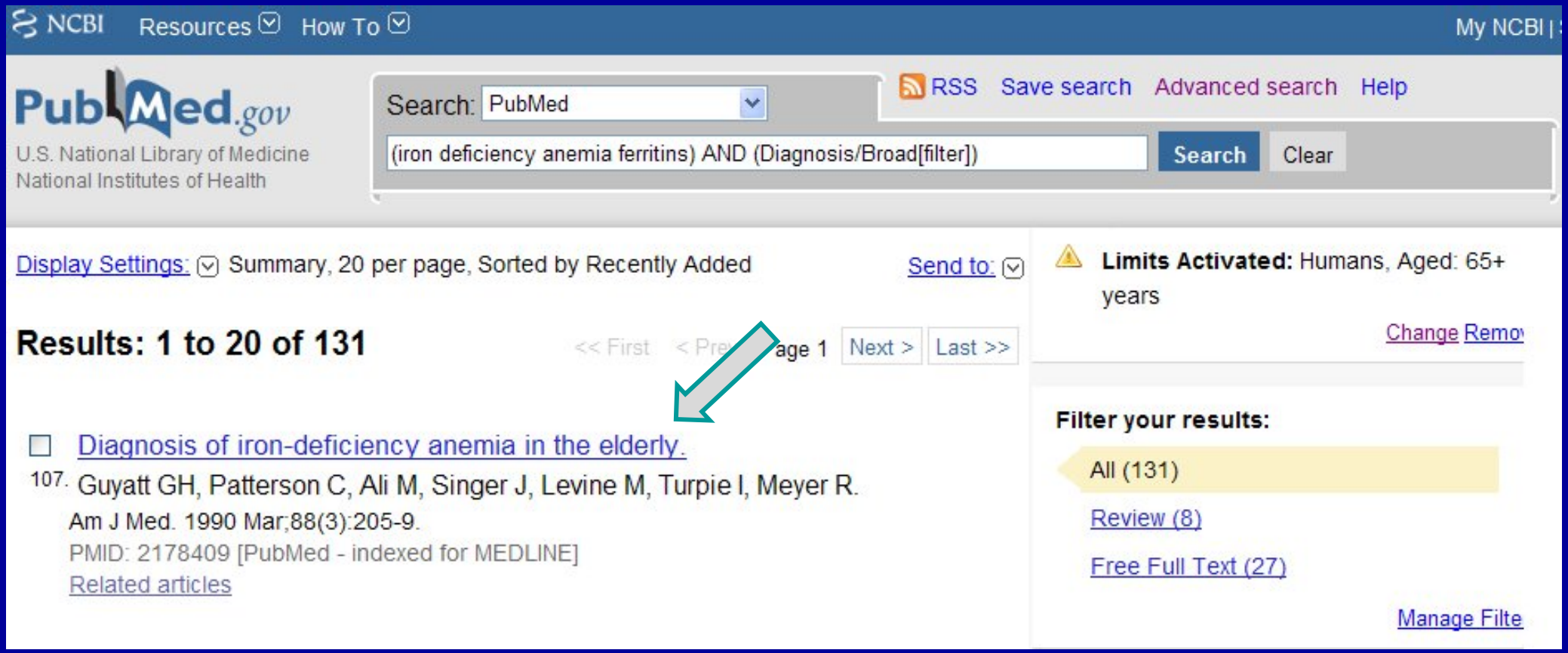


PubMed Clinical Queries



PubMed Clinical Queries




NCBI Resources How To My NCBI
PubMed.gov
U.S. National Library of Medicine
National Institutes of Health
Search: PubMed
RSS Save search Advanced search Help
(iron deficiency anemia ferritins) AND (Diagnosis/Broad[filter])
Search Clear
Display Settings: Summary, 20 per page, Sorted by Recently Added
Send to:
Limits Activated: Humans, Aged: 65+ years
Change Remov
Results: 1 to 20 of 131
<< First < Prev age 1 Next > Last >>
Diagnosis of iron-deficiency anemia in the elderly.
107. Guyatt GH, Patterson C, Ali M, Singer J, Levine M, Turpie I, Meyer R.
Am J Med. 1990 Mar;88(3):205-9.
PMID: 2178409 [PubMed - indexed for MEDLINE]
Related articles
Filter your results:
All (131)
Review (8)
Free Full Text (27)
Manage Filte

Steps of EBM

❸ Appraise

Key components of your clinical question

PICO

P	Target condition	Elderly patient with IDA
I	Gold standard test	Bone marrow aspirates
C	Comparaison	Ferritin Positive ≤ 45 µg/L Negative > 45 µg/L
O	Accuracy	Sn – Sp – PPV – NPV – LR – CIs

IDA: Iron Deficiency Anemia

Validity of study design of diagnostic study

❶ Blind comparison with the **g** **ndard test**

❷ Gold standard test **in all patients**

❸ Diagnos **luated in appropriate spectru** patients

If no → Stop here

Test evaluated in large spectrum of patients

Story of CEA

- CEA in **advanced CRC** & healthy controls[1]
  35 of 36 pts with advanced CRC had elevated CEA (98%)
  Almost all healthy controls had low levels of CEA
  CEA might be a useful screening tool for CRC
- CEA in **early-stage CRC** & other GI disorders[2]
  Test unable to differentiate early cancer from other disorders

Spectrum Bias

[1] Thomson DM et all. Proc Natl Acad Sci U S A 1969 ; 64 : 161.
[2] Bates SE. Ann Intern Med 1991 ; 115 : 623.

Accuracy of dichotomous diagnostic test

Only 2 results

- Sensibility (**Sn**)
- Specificity (**Sp**)
- Positive Predictive V)
- Negative Predi ue (**NPV**)
- Likeliho (**LRs**)
- Diagnos Odds Ratio (**OR**)

If no → Stop here

CIs

Newman TB & Kohn MA. Evidence-based diagnosis.
Cambridge University Press, Cambridge, UK, 1st edition, 2009.

Diagnosis of IDA in elderly patients

2 x 2 table

		Gold standard test Bone marrow aspirates		**Row totals**
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive**			
	Negative			
Column totals				

Diagnosis of IDA in elderly patients

2 x 2 table

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive**	**a** **True positive**		
	Negative			
Column totals				

Test positive & disease present

Diagnosis of IDA in elderly patients

2 x 2 table

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive**		**b** **False positive**	
	Negative			
Column totals				

Test positive & disease absent

Diagnosis of IDA in elderly patients

2 x 2 table

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive**			
	Negative	**c** **False negative**		
Column totals				

Test negative & disease present

Diagnosis of IDA in elderly patients

2 x 2 table

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive**			
	Negative		**d** **True negative**	
Column totals				

Test negative & disease absent

Diagnosis of IDA in elderly patients

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 μg/L	**70**	**15**	**85**
	Negative > 45 μg/L	**15**	**135**	**150**
Column totals		**85**	**150**	**235**

Guyatt G et all. Am J Med 1990;88:205.

Sensitivity

Percent of diseased individuals with positive test

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 µg/L	**a**		
	Negative > 45 µg/L	**b**		
Column totals		**(a + b)**		

$$\text{Sn} = \frac{a}{a + c}$$

Denominator = Column totals

Sensitivity

Percent of diseased individuals with positive test

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 µg/L	**70**		
	Negative > 45 µg/L	**15**		
Column totals		**85**		

$$\text{Sn} = \frac{70}{70 + 15} = 0.82$$

Denominator = Column totals

Sensitivity

- Mnemonic for sensitivity is "**PID**"
  "**P**ositive **I**n **D**isease"
  or **P**elvic **I**nflammatory **D**isease

- **SnNOUT**
  Highly **Sn**sitive test, when **N**egative, rules **OUT** disease
  Urine pregnancy test very sensitive for ectopic pergnancy
  Negative urine pregnancy test rules out ectopic pregnancy

Specificity

Percent of non-diseased individuals with negative test

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 µg/L		**b**	
	Negative > 45 µg/L		**d**	
Column totals			**b + d**	

$$\text{Sp} = \frac{\text{d}}{\text{b + d}}$$

Denominator = Column totals

Specificity

Percent of non-diseased individuals with negative test

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 µg/L		**15**	
	Negative > 45 µg/L		**135**	
Column totals			**150**	

$$\text{Sp} = \frac{135}{135 + 15} = 0.9$$

Denominator = Column totals

Specificity

- Mnemonic for specificity is "**NIH**"
  Negative **I**n **H**ealth
  or **N**ational **I**nstitutes of **H**ealth

- **SpPIN**
  Perfectly **Sp**ecific test, when **P**ositive, rules disease **IN**
  "Pathognomonic" findings
  Visualization of head lice for that infestation

Limitations of sensitivity & specificity

- Sn & Sp **work backwards** from clinical practice:
  Evaluate patients with known disease
  Data about presence or absence of certain dg test results

- Patients present with symptoms & diagnostic test results
  Work forwards to determine likelihood of the disease

PPV & NPV provide these data

Positive predictive value

Percent of positive tests that are truly positive

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 µg/L	**a**	**b**	**a + b**
	Negative > 45 µg/L			
Column totals				

$$\text{PPV} = \frac{\text{a}}{\text{a + b}}$$

Denominator = Row totals

Positive predictive value

Percent of positive tests that are truly positive

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 µg/L	**70**	**15**	**85**
	Negative > 45 µg/L			
Column totals				

$$\text{PPV} = \frac{70}{85} = 82\ \%$$

Denominator = Row totals

Negative predictive value

Percent of negative tests that are truly negative

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 µg/L			
	Negative > 45 µg/L	**c**	**d**	**c + d**
Column totals				

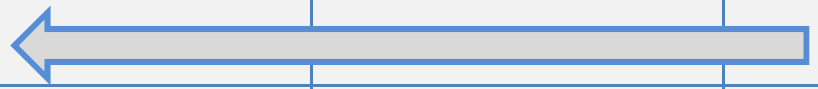
$$\text{NPP} = \frac{\mathbf{d}}{\mathbf{c + d}}$$

Denominator = Row totals

Negative predictive value

Percent of negative tests that are truly negative

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 μg/L			
	Negative > 45 μg/L	**15**	**135**	**150**
Column totals				



$$\text{NPV} = \frac{135}{15 + 135} = 90\ \%$$

Denominator = Row totals

Limitations of Sn/Sp & PPV/NPV

- **Sensitivity & specificity**

  Calculated around a cutoff point (45 μg/L for ferritin)

  Work where evaluated test has only 2 results (unusual)

- **PPV & NPV**

  Vary widely depending on disease's **prevalence**

  Work where evaluated test has only 2 results (unusual)

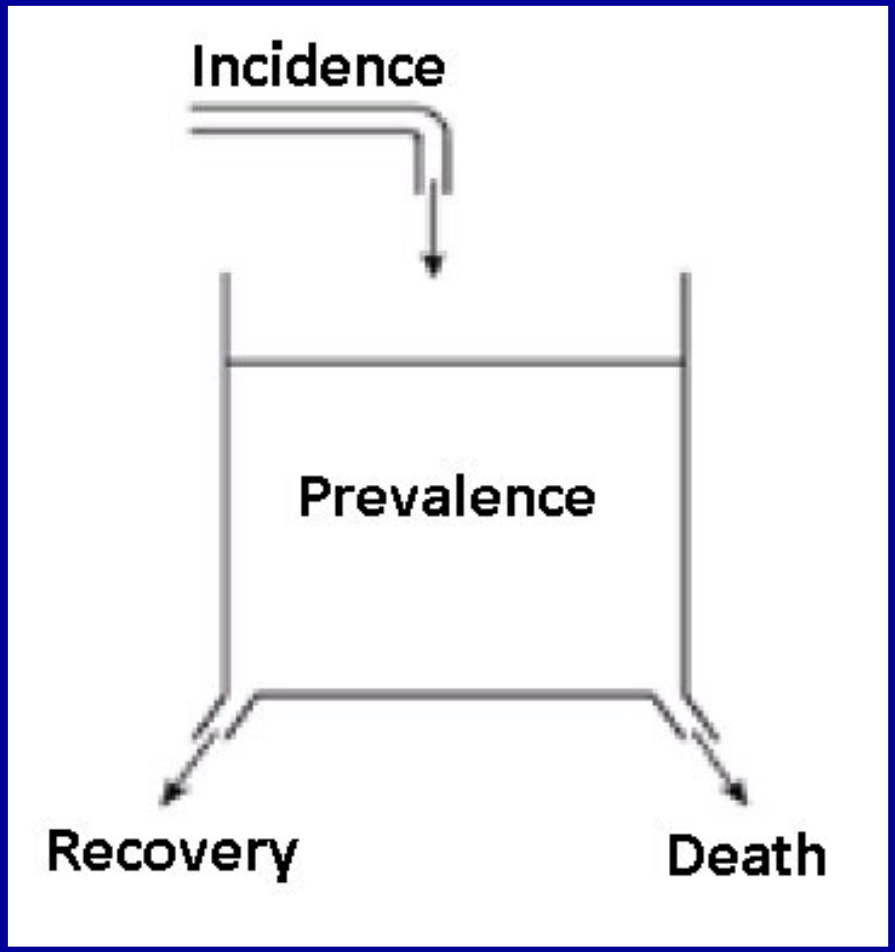
Likelihood ratios (LRs) overcome these weaknesses

Incidence & prevalence

- **Incidence:**
  Proportion of patients in the at-risk population who get the disease over a period of time (e.g.: 1 year)
- **Prevalence:**
  Proportion of patients in the at-risk population who have the disease at one point in time

Prevalence = Pre-test probability

Relationship between incidence & prevalence



Detels R et al. Oxford textbook of public health.
Oxford University Press, Oxford, 4th edition, 2002.

Diagnosis of IDA in the elderly

Prevalence

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 µg/L	**a**	**b**	
	Negative > 45 µg/L	**c**	**d**	
Column totals		**a + c**		**a + b + c + d**

$$\text{Prevalence} = \frac{\text{Cases}}{\text{Total population}} = \frac{a + c}{a + b + c + d}$$

Diagnosis of IDA in the elderly

Prevalence

		Gold standard test Bone marrow aspirates		Row totals
		Disease present	**Disease absent**	
Diagnostic test Ferritin	**Positive** ≤ 45 μg/L			
	Negative > 45 μg/L			
Column totals		**85**	**150**	**235**

$$\text{Prevalence} = \frac{85}{235} = 0.36$$

LRs & weaknesses of Sn, Sp, PPV & NPV

- **Weaknesses of Sn & Sp**

  LRs calculated for multiple ranges of dg test results

  LRs more useful to evaluate tests with > 2 results

- **Weaknesses of PPV & NPV**

  LRs don't change with different disease's prevalence

  LRs more useful to evaluate tests with > 2 results

Likelihood ratio (LR)

- Different way to interpret accuracy of a diagnostic test
- How much likelihood of disease changes given a test result
- Not important to comprehend formulae to calculate LR
- If LRs are not provided, you need to compute your own

LR for a positive test

LR + = Sensitivity / (1 – Specificity)

LR + = 0.82 / (1 – 0.9) = 8.24

Probability that the patient has true positive, rather than false positive test

Corresponds to clinically "ruling in disease"

LR for a negative test

LR – = (1 – Sensitivity) / specificity
LR – = (1 – 0.82) / 0.9 = 0.2
Probability that the patient has true negative rather than false negative test
Corresponds to clinically "ruling out disease"

Interpretation of LRs

Measure of changes in disease probability

- **Test result with LR > 1.0**

  Increase the likelihood of disease

  The higher the LR is, the closer you are to confirm dg

- **Test result with LR very close to 1.0**

  Little impact on estimates of likelihood of disease

- **Test result with LR < 1.0**

  Decrease the likelihood of disease

  The lower the LR is, the closer you are to rule out dg

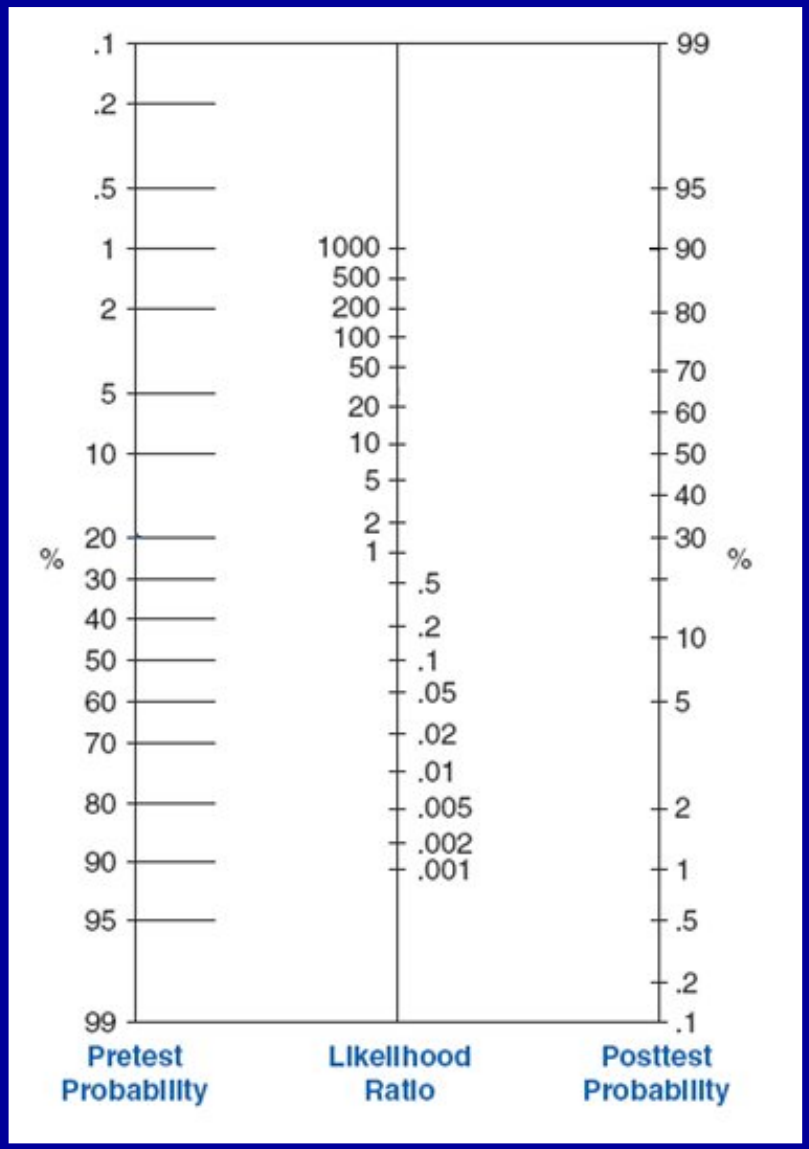
Interpretation of LRs

As a general rule of thumb

- **LR > 10 Rule in a diagnosis**
- **LR < 0.1 Rule out a diagnosis**
- **LR 0.5 – 2 Little effect on post-test probability**

Bayes nomogram

LR convert pre-test probability to post-test probability



Evid Based Med 2001 ; 6 : 164 – 166.

Pre-test probability/prevalence

- **Definition** Likelihood of disease before dg test result
- **Estimation**
  - History
  - Physical examination
  - Clinical experience

  } **Not accurate**

Prevalence studies **More accurate**

- **Order dg test:**
  - Very low → **No diagnostic test**
  - Intermediate? → **Diagnostic test**
  - Very high → **No diagnostic test**

Example of a low pre-test probability

- a 24-year-old female
- Occasional pain anterior chest wall not related to effort
- Physical findings unremarkable
- Probability of having a heart attack quite low
- Pre-test probability of heart attack to **0.1%**

Example of a high pre-test probability

- Prevalence of HP in PU in some countries ≈ 100%
- These patients automatically treated for HP after identification of peptic ulcer without testing for HP

What level of pretest probability is intermediate?

- Would it be 25 or 75% likelihood of disease?
- Use clinical judgment:
  - **Cost**
  - **Accuracy**
  - **Side effects** of diagnostic test
  - **Consequences of "missed" diagnosis**
    Lower threshold to order tests if fatal consequences
  - **Patient's preferences**
    Negative test reassure anxious patients

Estimating pre-test probability

- It is often difficult to know whether the pre-test probability you assign is correct
- Often, you must make a **guess**, which seems rather arbitrary given the complex calculations that ensue

Assigning pre-test probability is **both an art & a science**

Pines J & Everett W. Evidence-Based emergency care. Blackwell Publishing, Oxford, UK, 1st edition, 2008.

Likelihood ratio

Measure of changes in disease probability

- **Definition**

  How much likelihood of disease change given a test result

- **Estimation**

  Result(s) of study(ies) you reviewed

  LR varies depending on test result LR for a positive test

  LR for a negative test

Post-test probability

- **Definition**

  Likelihood of disease after a diagnostic test result

- **Estimation**

  Using Bayes nomogram

  For those who are numero-phobic

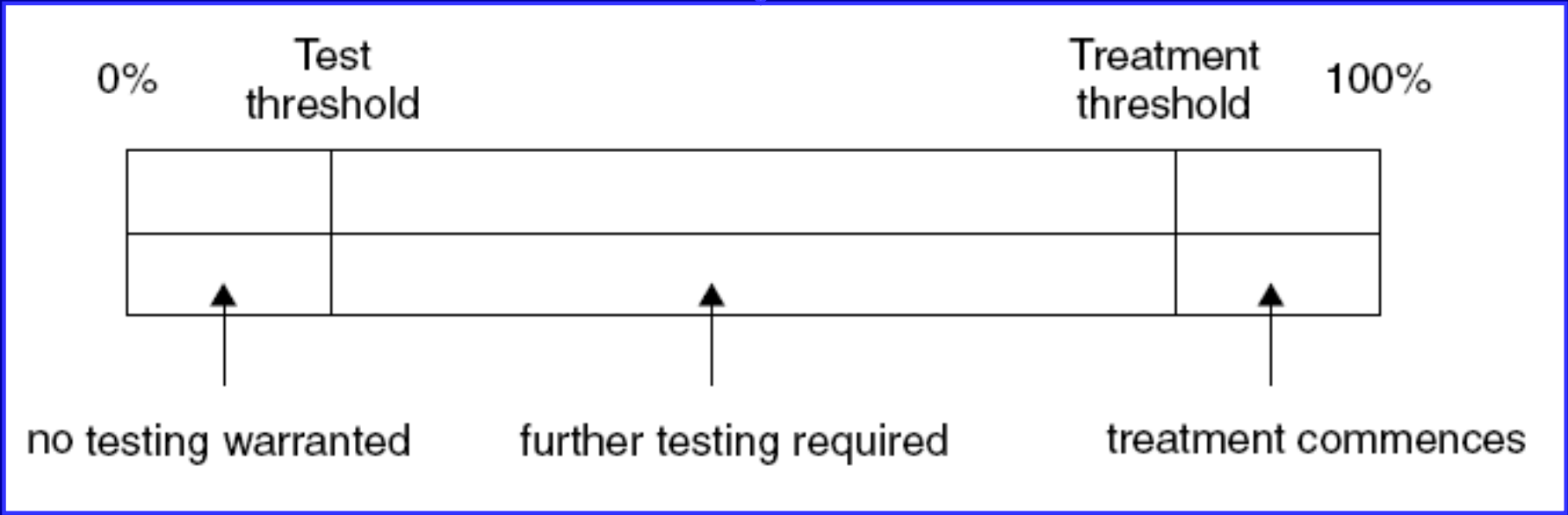
  Manual computation

  For those who are up to challenge of manual computation

Probability of diagnosis

After arriving at a post-test probability of disease

You may make a clinical decision



Guyatt G et all. Users' guides to medical literature: manual for EBP. McGraw-Hill, New York, USA, 2nd edition, 2008.

Use of likelihood ratio

75-year-old female with pneumonia

Hemoglobin level: 10 g/dL

MCV: 80

Ferritin level: 40 μg/L

Pre-test probability of IDA*: **36%**

LR for + & – test: **8.2 – 0.2**

Post-test probability: **?**

*IDA: iron deficiency anemia

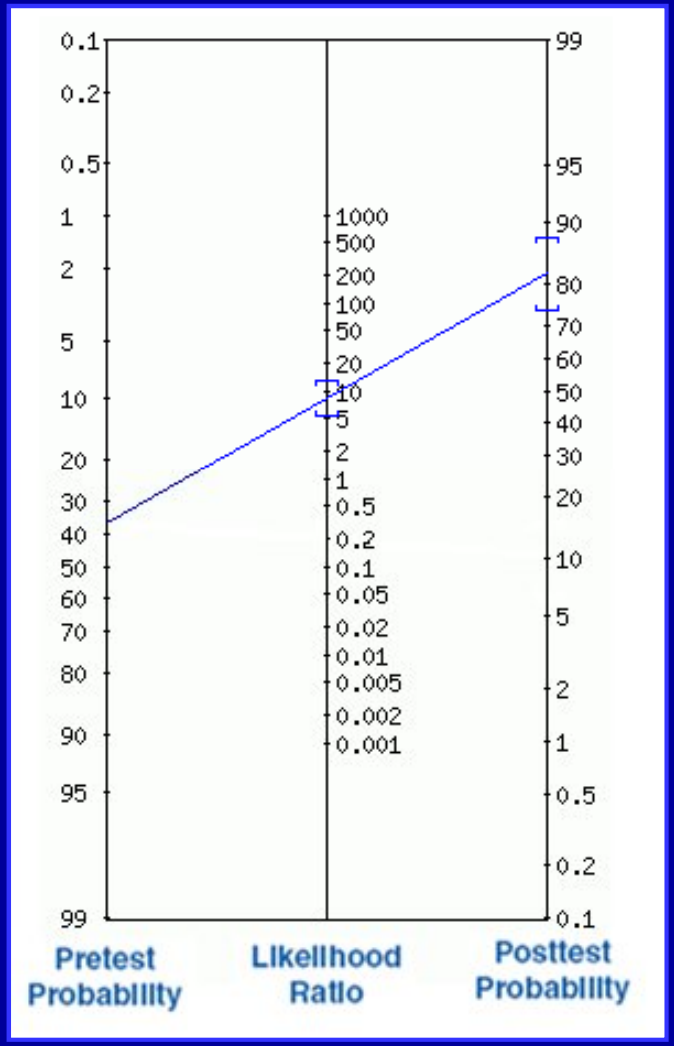
Fagan nomogram

Post-test probability for a positive test

Pre-test probability	**36%**
LR +	**8.2**
Post-test probability	**82 %**

Certainly high enough for us to want investigate her anemia

Corresponds to "**SpPin**"



http://araw.mede.uic.edu/

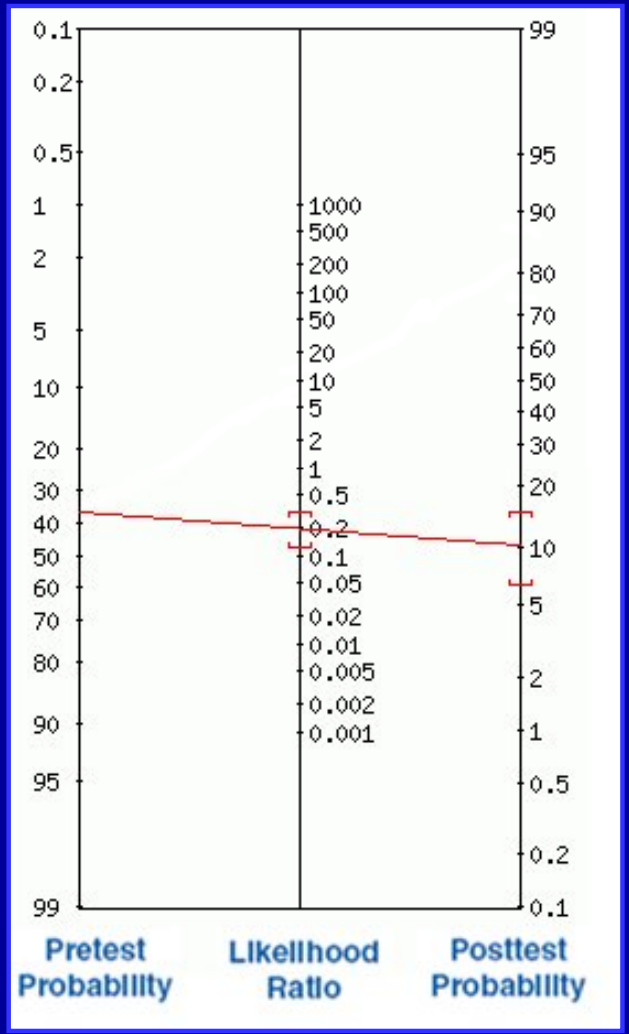
Fagan nomogram

Post-test probability for a negative test

Pre-test probability	**36%**
LR –	**0.2**
Post-test probability	**10 %**

Certainly low enough for us to exclude other causes of anemia

Corresponds to **SnNout**



http://araw.mede.uic.edu/

Diagnostic odds ratio

$$\text{Diagnostic OR} = \text{LR} + / \text{LR} -$$

$$\text{Diagnostic OR} = 8.2 / 0.2 = 41$$

Statistical significance

- **Confidence Interval**

  Reported around each measure of diagnostic accuracy:

  Sn – Sp – PPV – NPV – LRs – diagnostic OR

- **Interpretation of CI** in absence of 'line of no difference'

  Interpret results in same way at lower & higher end of CI?

Calculation of CI

$$95\%\ \mathrm{CI} = \mathrm{P} \pm 1.96 \times \sqrt{\mathrm{P}(1-\mathrm{P})/\mathrm{N}}$$

P Proportion of patients with etiology of interest

N Number of patients in the sample

Formula inaccurate if number of cases is 5 or fewer

Guyatt G et all. Users' guides to the medical literature.
Manual for evidence-based clinical practice. McGraw-Hill, 2nd edition, 2008.

Interpretation of 95% CI

When the data sampling is repeated many times, the 95% CI calculated from each sample will, on average, contain the "true" value of the proportion in 95% of the samples

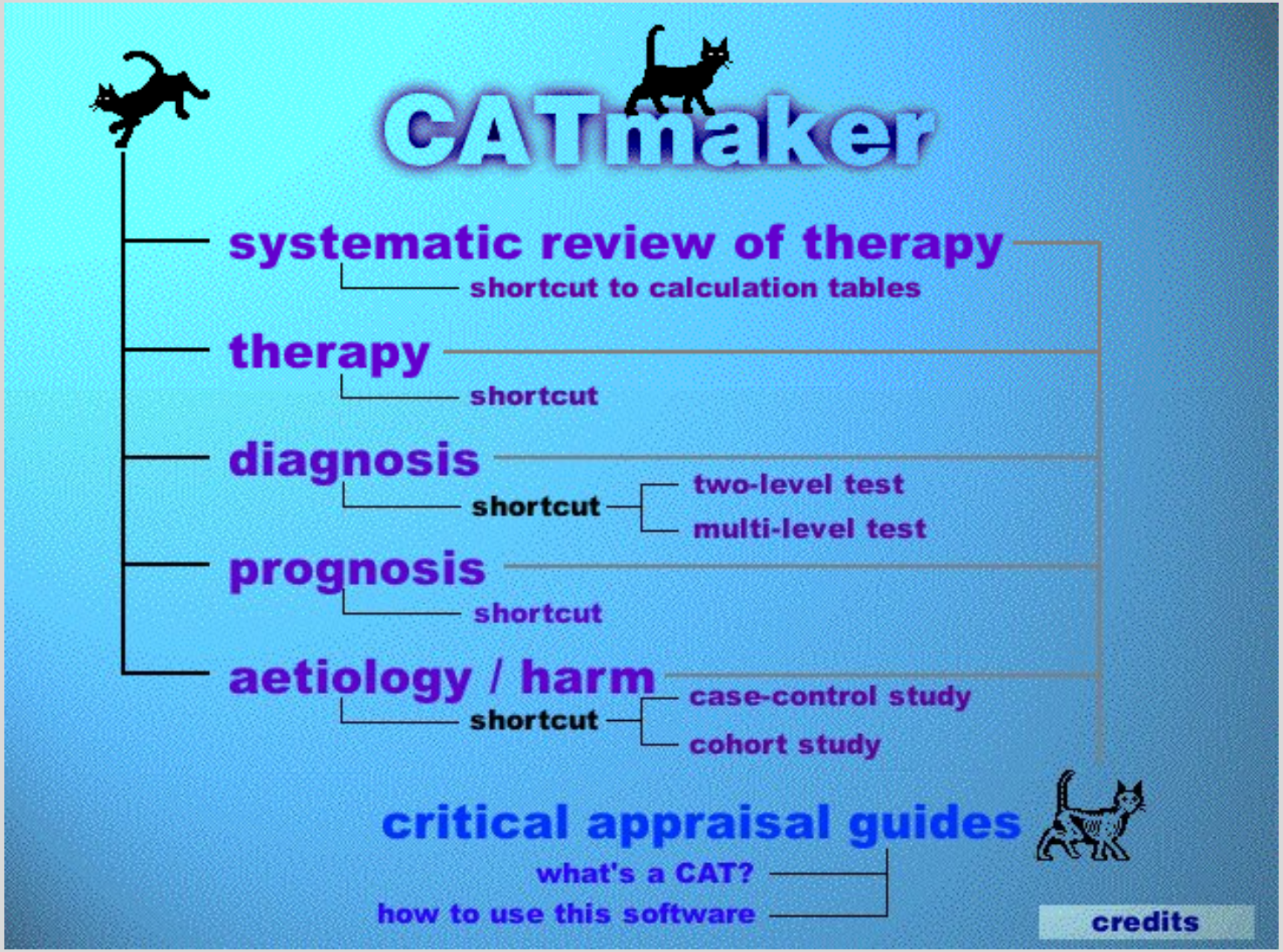
CI width smaller with increasing sample size

CEBM
CENTRE FOR EVIDENCE BASED MEDICINE

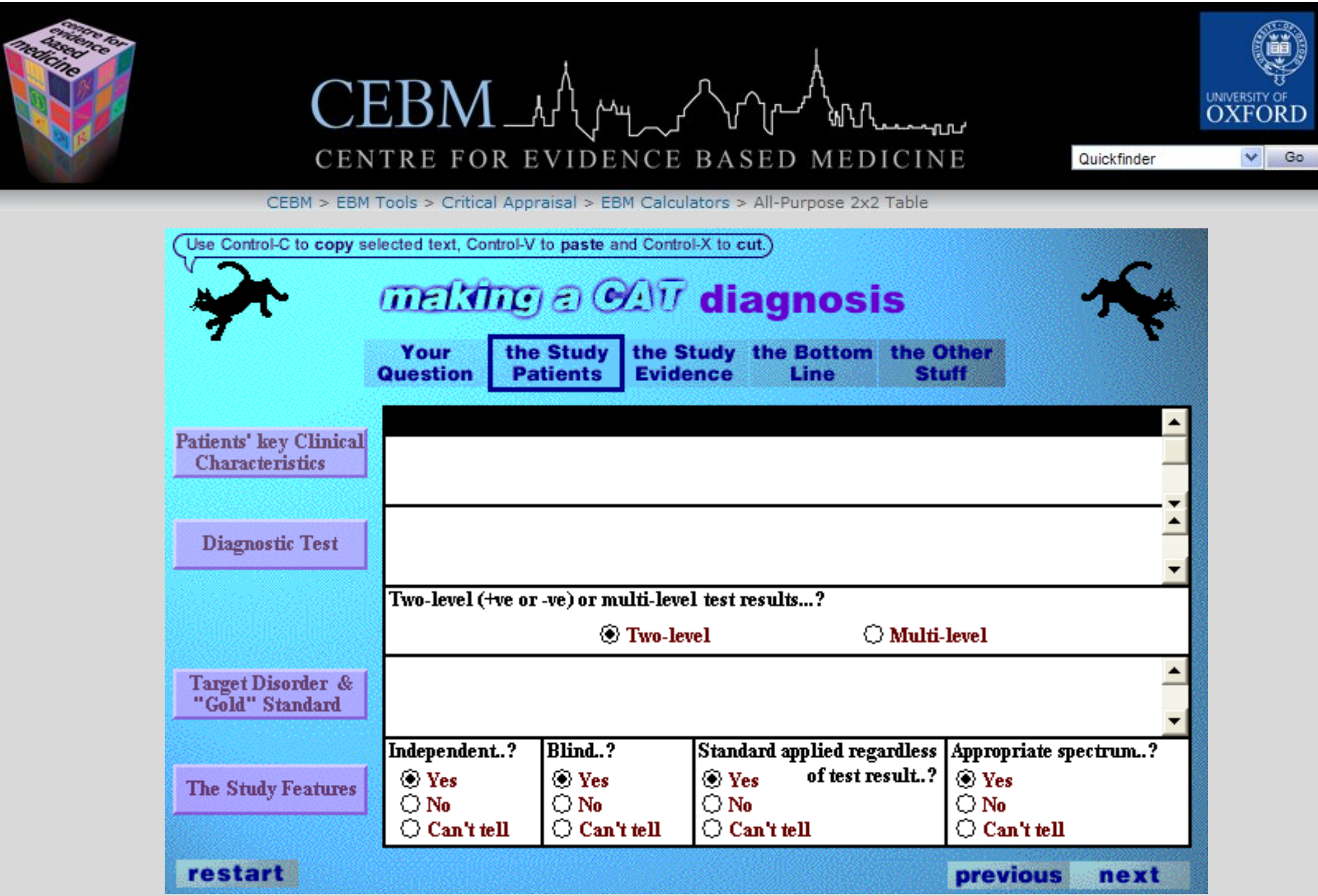
UNIVERSITY OF OXFORD

Quickfinder Go

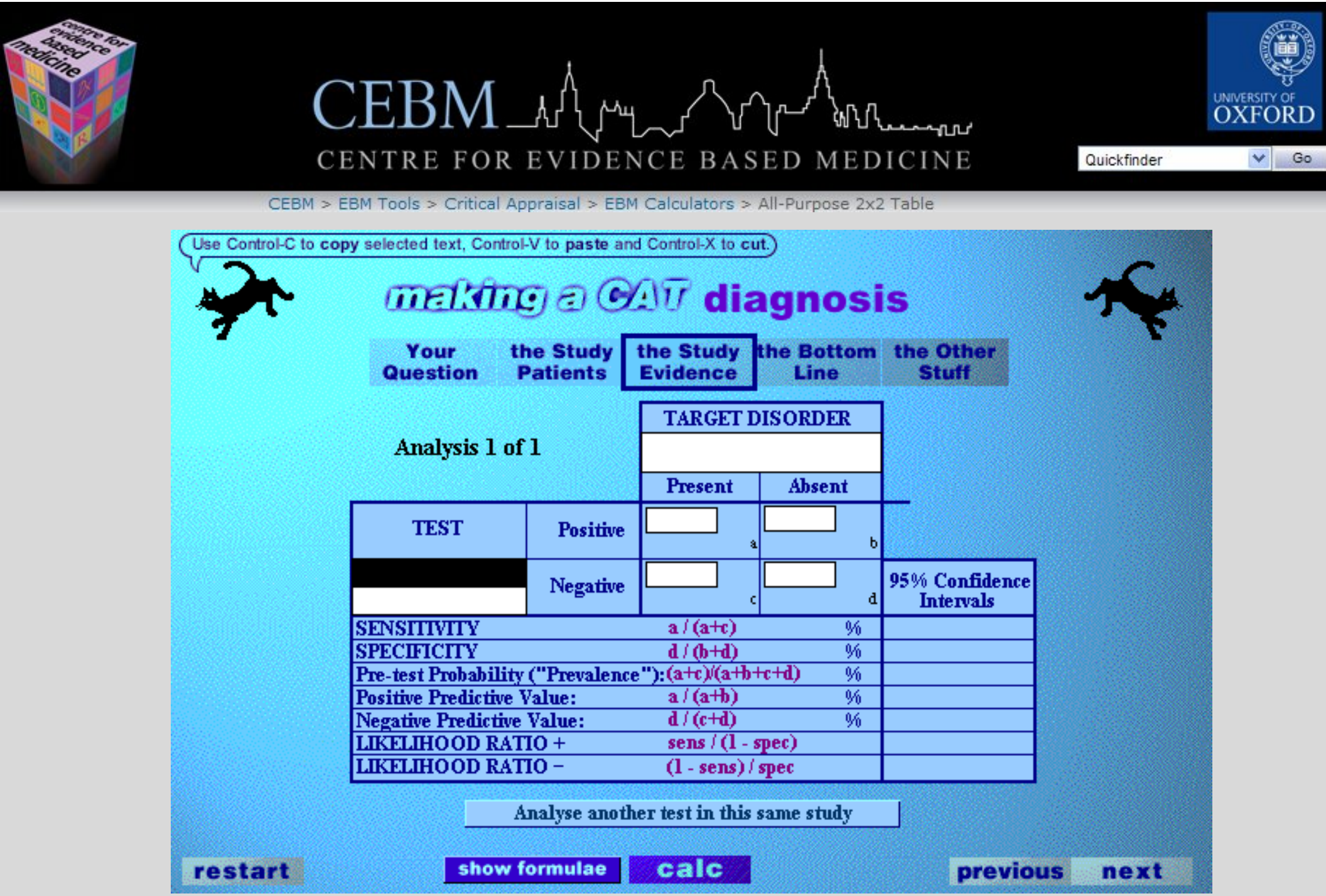
CEBM > EBM Tools > Critical Appraisal > EBM Calculators > All-Purpose 2x2 Table



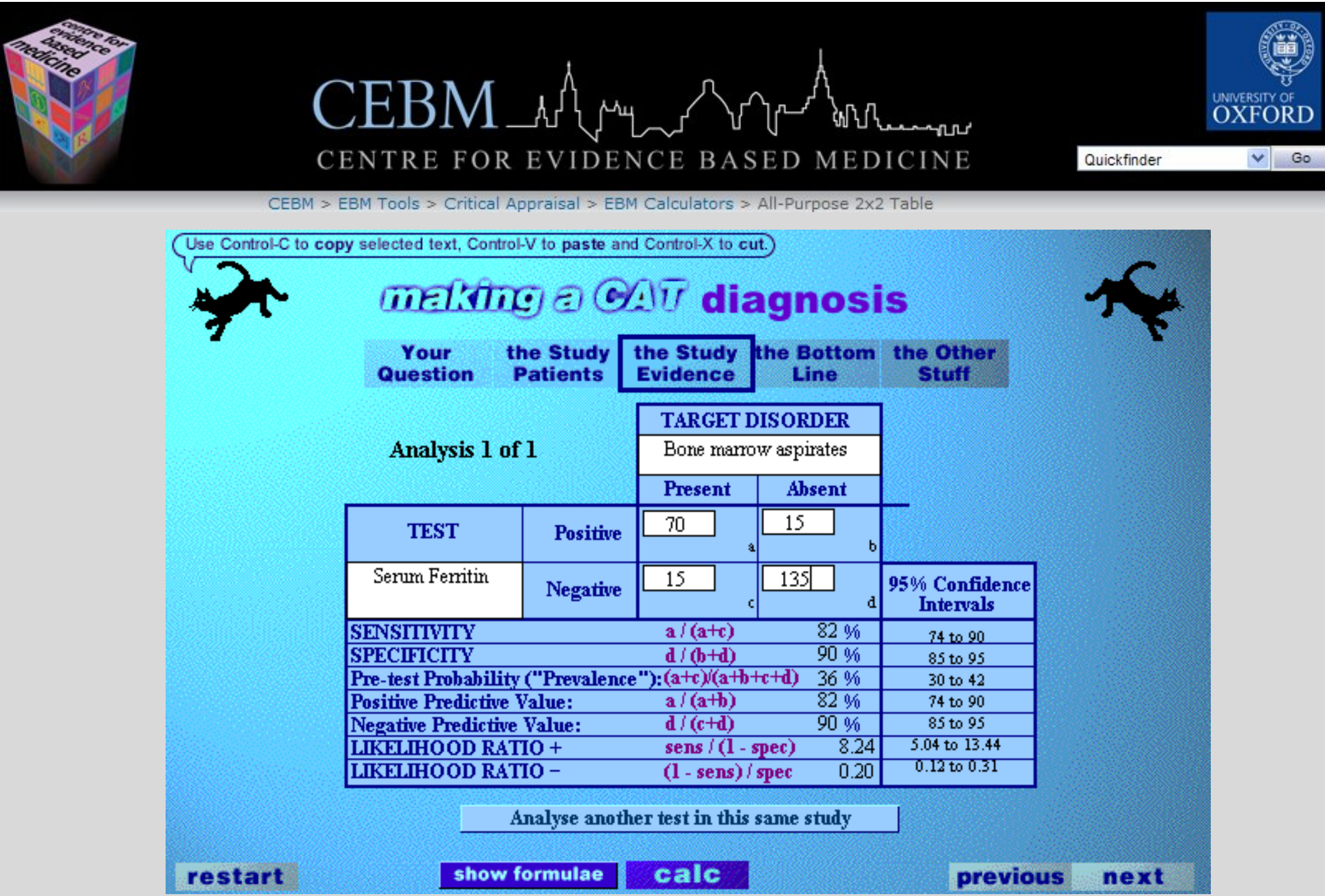
http://www.cebm.net/index



http://www.cebm.net/index



http://www.cebm.net/index



http://www.cebm.net/index

95% CIs around accuracy measures

Cut-off value 45 µg/L

Accuracy	Results (95% CIs)	Comments
Sensitivity	82% (74 – 90)	Good sensitivity
Specificity	90% (85 – 95)	Good specificity
PPV	82% (74 – 90)	Good PPV
NPV	90% (85 – 95)	Good NPV
Prevalence	36% (30 – 42)	–
LR+	8.24 (5.04 - 13.44)	Moderate changes
LR–	0.2 (0.12 - 0.31)	Moderate changes
Diagnostic OR	42	Good diagnostic OR

Accuracy of tests & number of results

- **Dichotomous test (only 2 results)**

  Sensibility & Specificity
  PPV & NPV
  LR + & –

  Diagnostic OR

  } CIs

- **Multilevel test (> 2 results)**

  Receiver Operating Characteristic (**ROC**)

  LR associated with each possible result or interval

  Make continuous test dichotomous: fixed cut-off value

Diagnosis of IDA in elderly patients

		Gold standard test Bone marrow aspirates		Sn	Sp
		Disease +	Disease –		
Dg test Ferritin (µg/L)	≤ 18	47	2	0.55 (47/85)	0.98 (148/150)
	≤ 45	70	15	0.82 (70/85)	0.90 (135/150)
	≤ 100	77	42	0.90 (77/85)	0.72 (108/150)
Column totals		85	150		

Guyatt G et all. Diagnosis of iron deficiency anemia in the elderly. Am J Med 1990 ; 88 : 205.

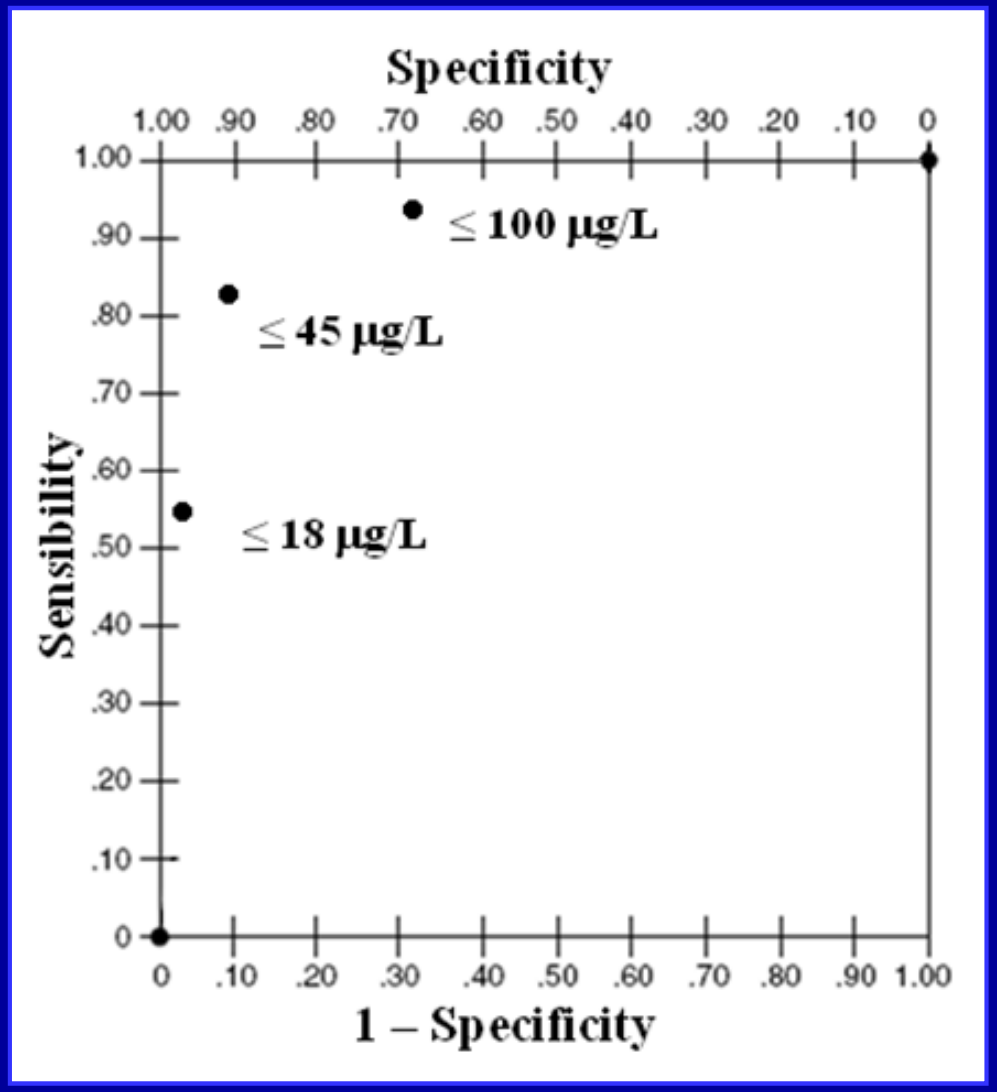
Receiver Operating Characteristic (ROC)

- Plot of test **sensitivity** on the y axis versus its **FPR** (or 1 – specificity) on the x axis

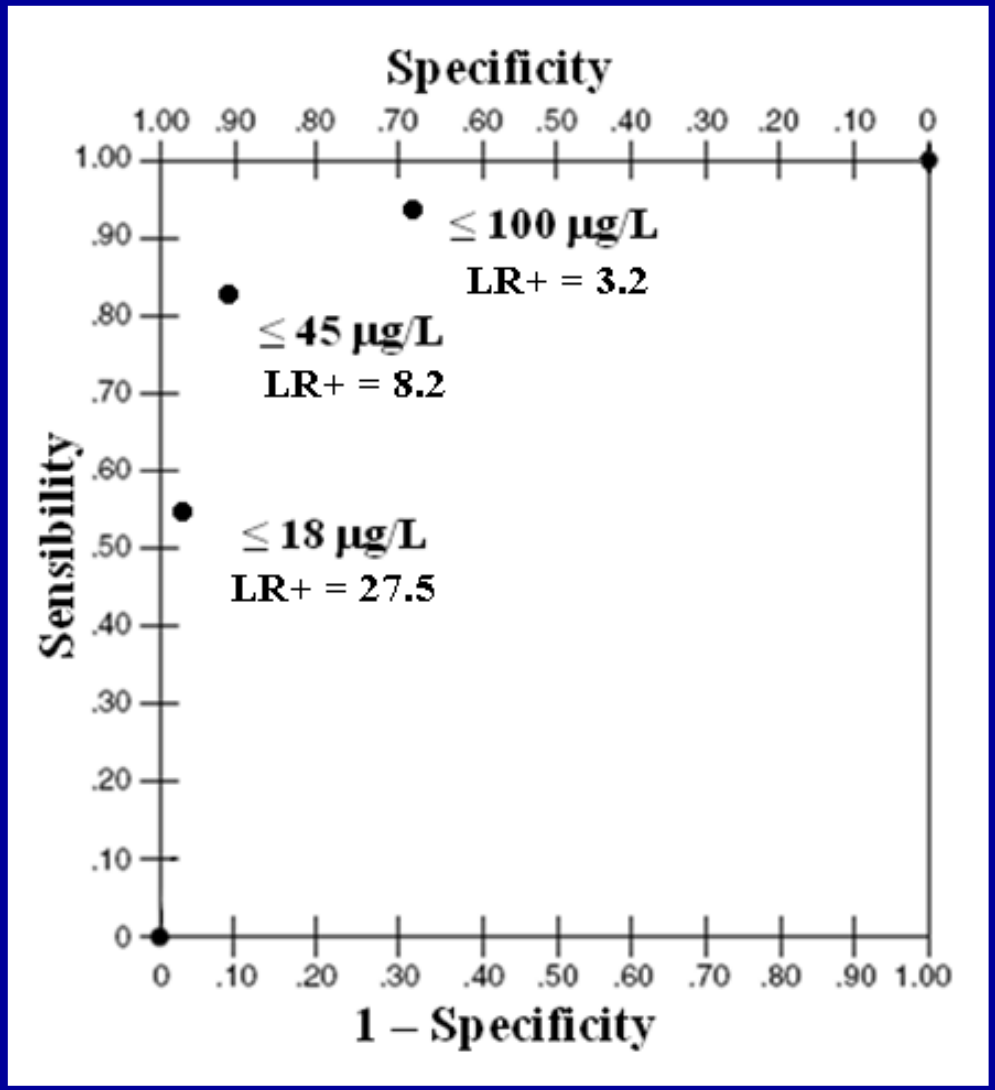
---

- Each discrete point on graph called **operating point**
- Curve illustrates how sensitivity & FPR vary together

Empirical ROC/ Diagnosis of IDA in elderly



Empirical ROC/ Diagnosis of IDA in elderly



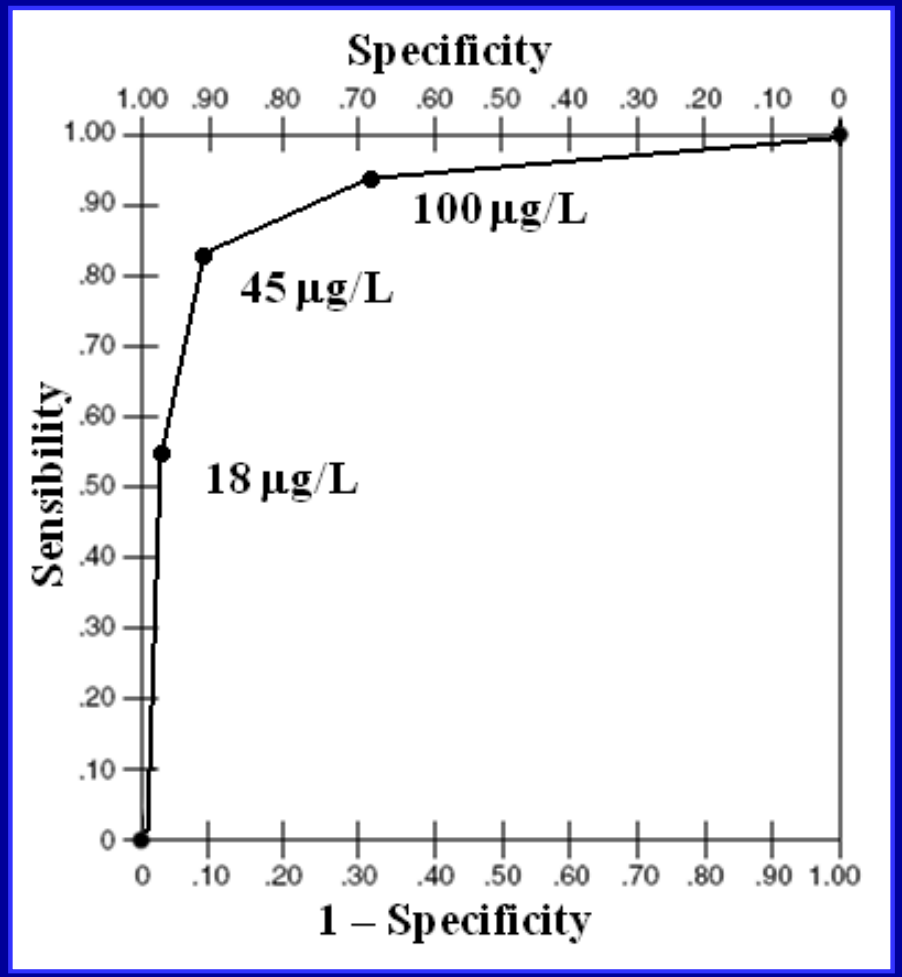
Empirical ROC/ Diagnosis of IDA in elderly

Connect all the points obtained

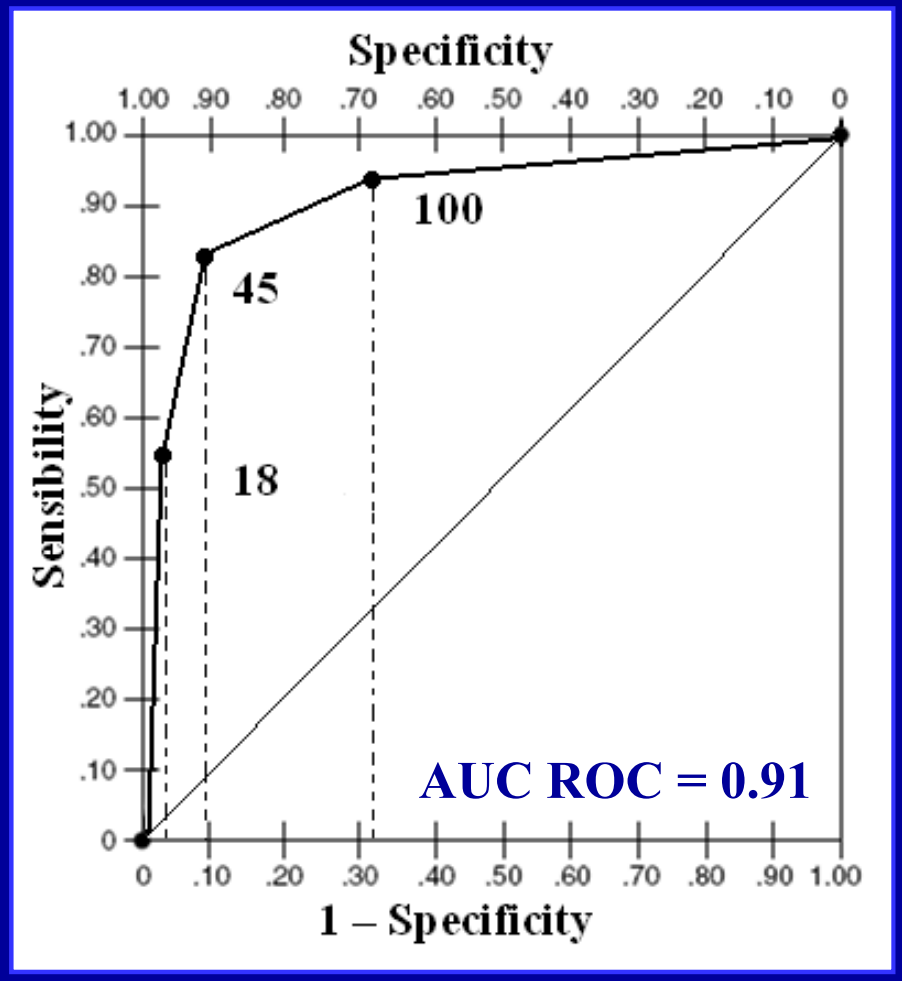
at all the possible cutoff levels:

- 3 values of FPR & sensibility

- 2 endpoints on curve: 0,0 & 1,1



Empirical ROC curve/Area under the curve



Summation of areas of trapezoids formed by connecting points on ROC curve

Useful properties of ROC curve

❶ Accuracy of binary diagnostic test for a cut-point value

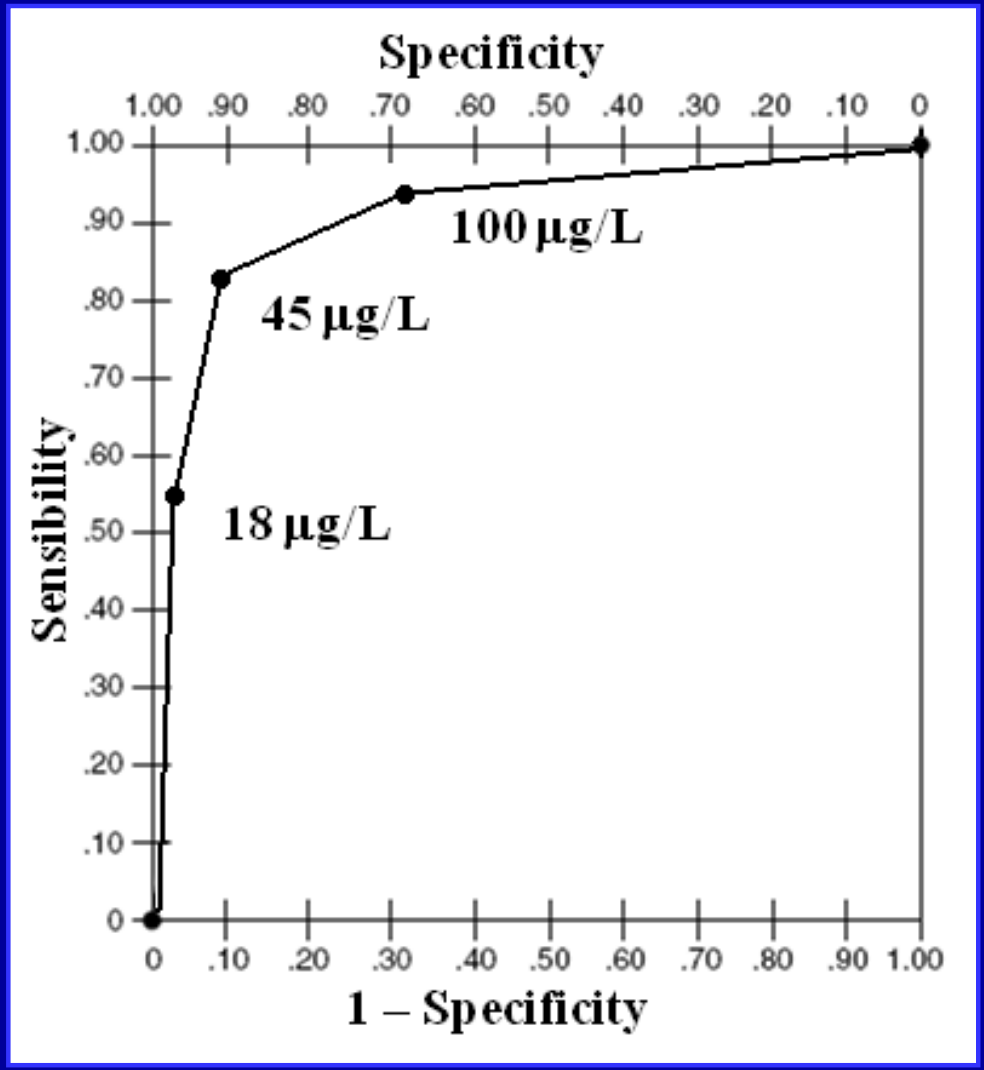
❷ AUC provides an overall measure of a test's accuracy

❸ Slope of tangent at cut-point gives LR for that value

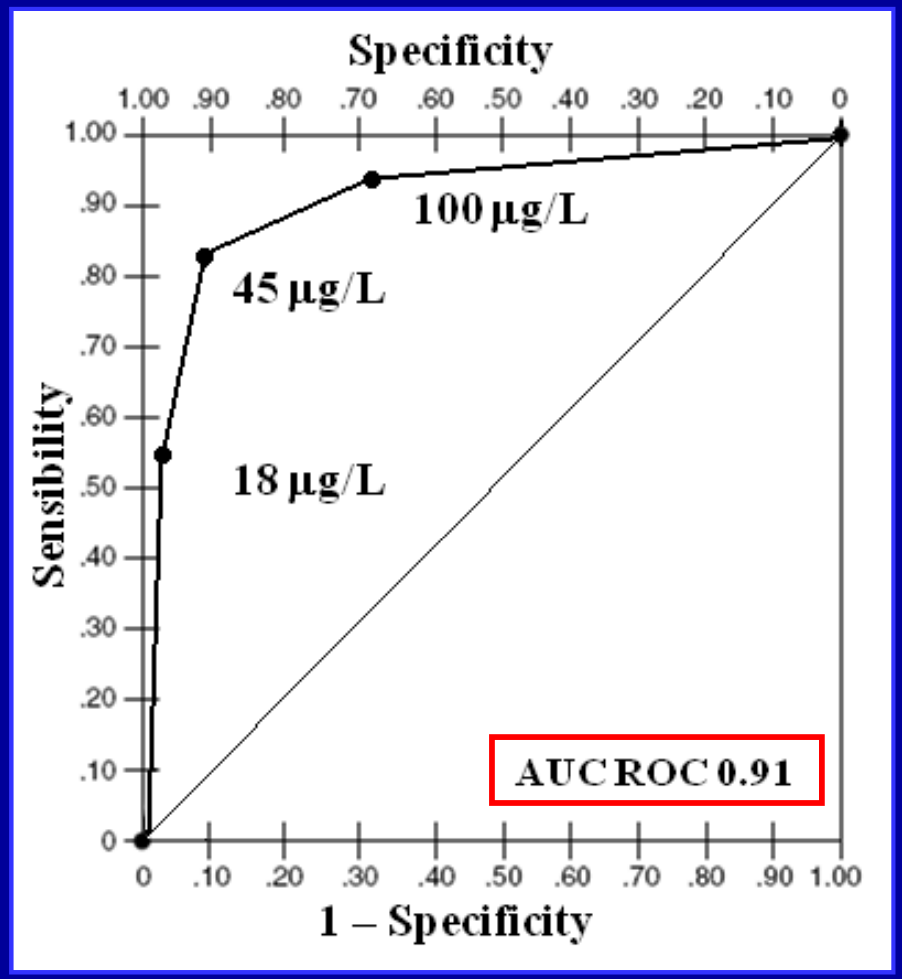
❹ Determination of cut-off point to distinguish D + & D –

❺ Comparison of different tests for dg of a target disorder

❶ Accuracy of binary dg test for a cut-point value



❷ Area under the ROC curve in IDA



If we select 2 patients at random one with IDA & one without

Probability is 0.91 that patient with IDA will have abnormal ferritin

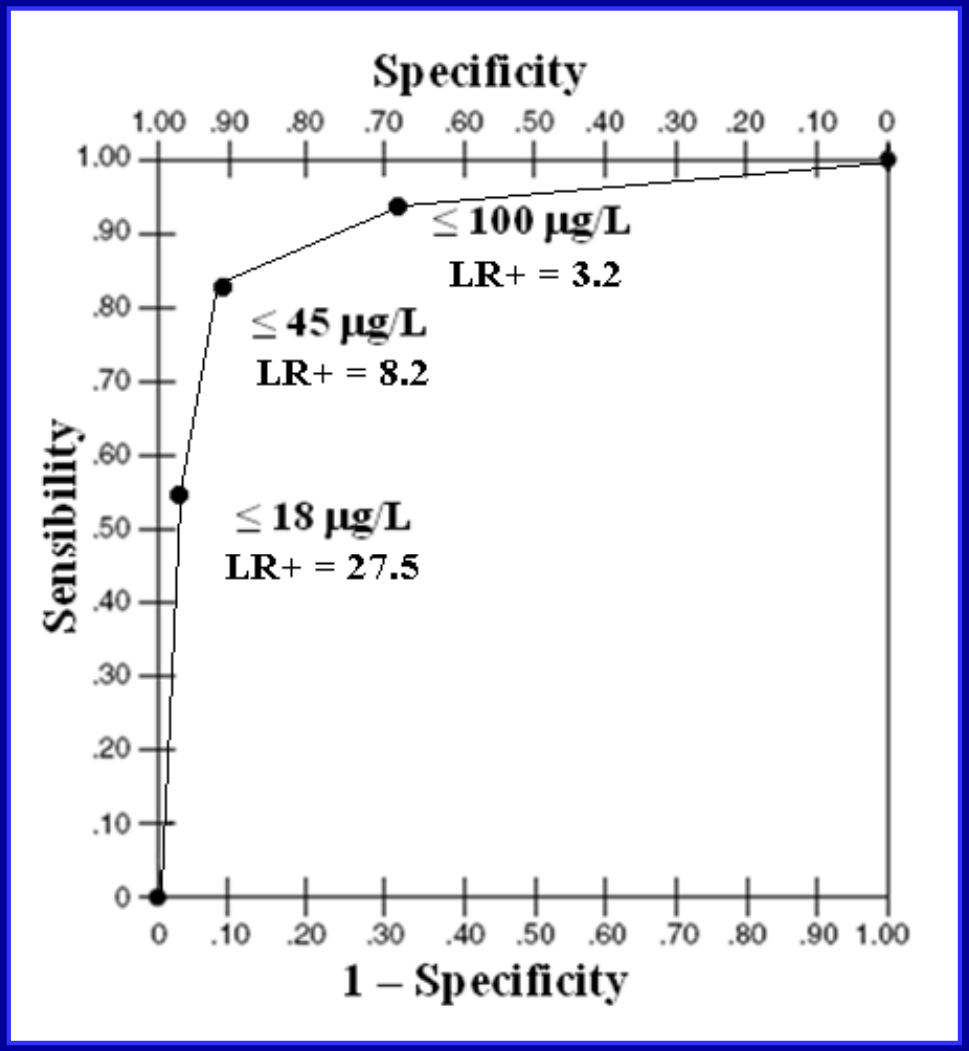
Accuracy of diagnostic test using AUC of ROC

Value	Accuracy
0.90 - 1.00	Excellent
0.80 - 0.90	Good
0.70 - 0.80	Fair
0.60 - 0.70	Poor

The higher AUC the better the overall performance of the test

Pines JM & Everett WW. Evidence-Based emergency care: diagnostic testing & clinical decision rules. Blackwell's publishing – West Sussex – UK – 2008.

❸ Slope of tangent at cut-point gives LR

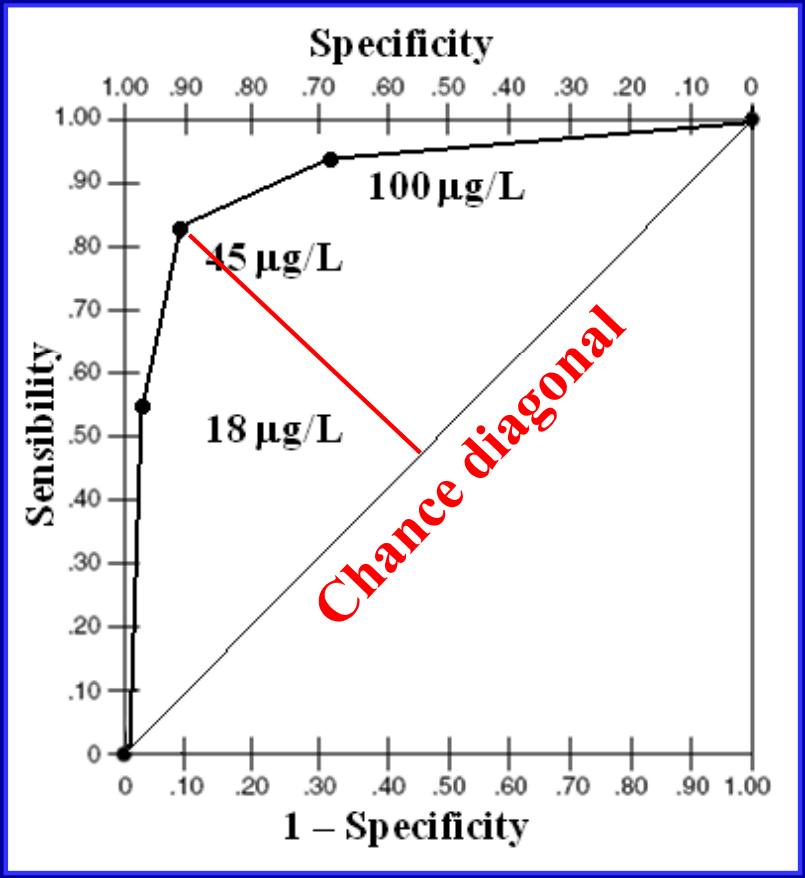


Steeper tangent for cutoff of 18 than it is for cutoff of 45

❹ Determination of cut-off point to distinguish D + & D –

Cut-off point discriminates between subjects with or without disease

Indicated by the point on curve that is far away from chance diagonal



Peat JK. Health science research. Allen & Unwin, Australia, 1st edition, 2001.

❹ Comparing different tests for target disorder

Diagnosis of IDA	AUC of the ROC
Seurm ferritin	0.91
Transferrin saturation	0.79
MCV	0.78
RCP	0.72

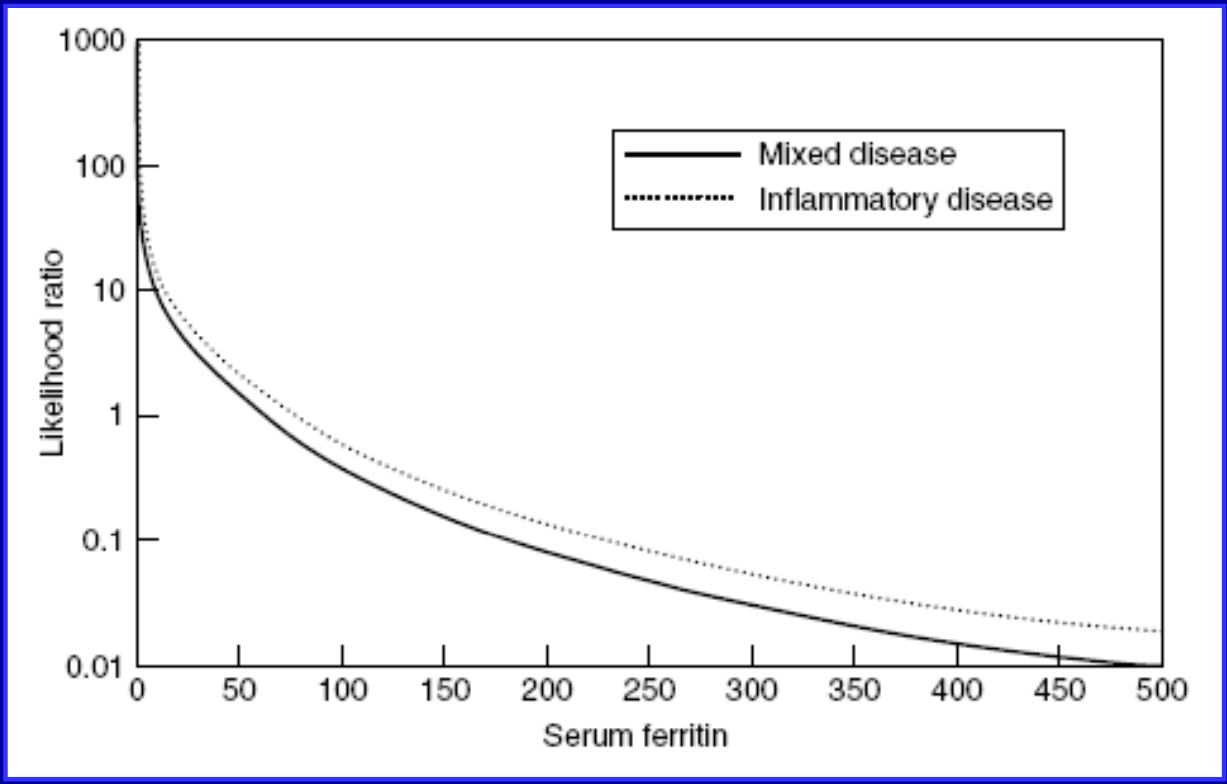
* RCP: Red Cell Protoporphyrin

Guyatt GH et al. J Gen Intern Med 1992 ; 7 : 145 – 153.

Likelihood Ratio Line

If you have sufficient observations, you can go beyond the multiple cut approach & construct LR line describing the relation between the test result & LR across the entire range of test values

LR line for serum ferritin in dg of IDA



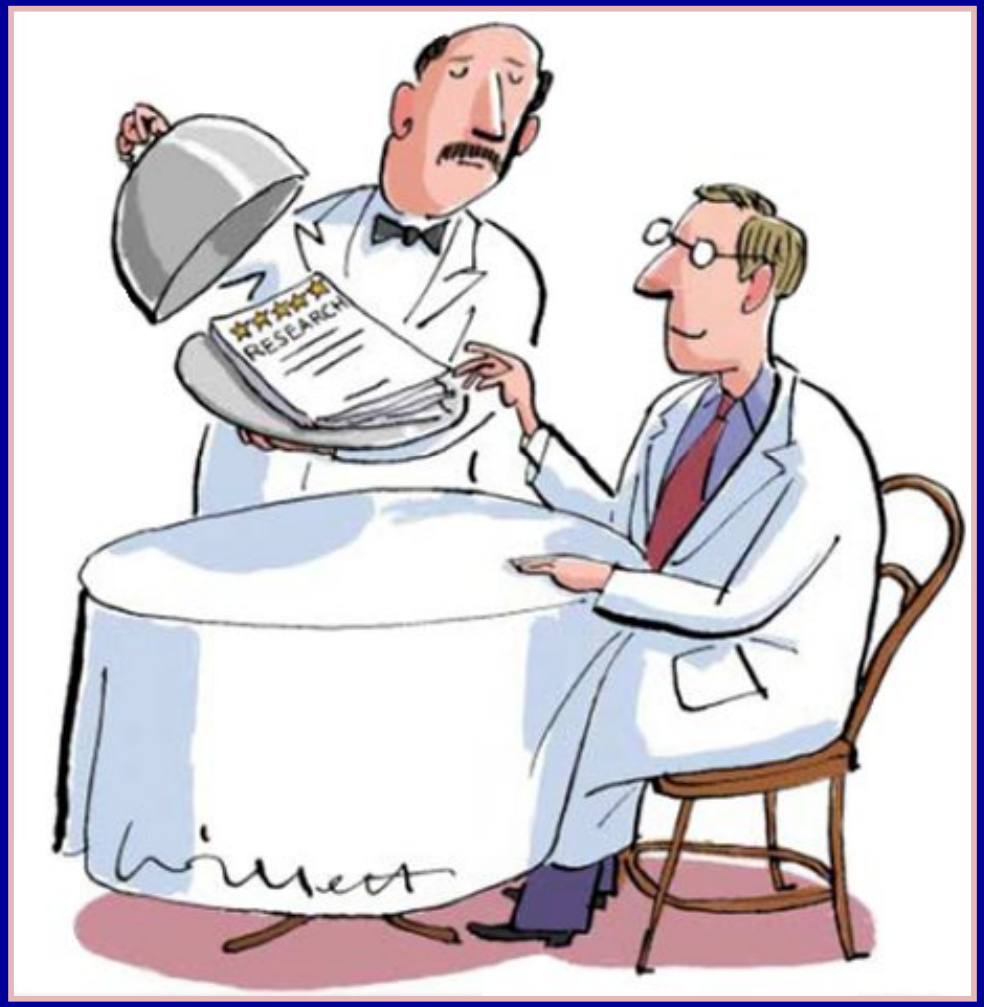
Systematic review – **55 eligible studies** – **2,579 patients**

Completely smooth curve (No difference in LRs between categories)

Need help of a statistician

Guyatt GH et al. J Gen Intern Med 1992 ; 7 : 145 – 153.

Critical appraisal of an article on dg test



Glasziou P et al. BMJ 2004 ; 328 : 39 – 41.

Furberg BD & Furberg CD. Evaluating clinical research.
Springer Science & Business Media – First Edition – New York – 2007.

Appraising a diagnostic study – 1

❶ Are the results valid (study design)?

a. Blind comparison with **gold standard test**	**Yes**
b. Gold standard test performed in **all patients**	**Yes**
c. Diag test done in appropriate **spectrum** of pts	**Yes**

Straus C et al. Evidence-based medicine: How to practice & teach & EBM. Elsevier – 3rd edition - 2009

Appraising a diagnostic study – 2

❷ Are the valid results accurate?

a. **Accuracy**: Sn, Sp, PPV, NPV, LR,dg OR	**Good**
b. **Precision** of the results	**Yes** (narrow CI)
c. **Dichotomizing** continuous test scores: Sn, Sp, LR+, LR–	**Yes**

Straus C et al. Evidence-based medicine: How to practice & teach & EBM. Elsevier – 3rd edition – 2009.

Steps of EBM

❹ Apply

Appraising a diagnostic study – 3

❸ Can we apply valid accurate results to our patient

a. Test available & reproducible in my setting	**Yes**
b. Generate sensible pre-test probability	**Yes** (36%)
c. Post-test probability affects management	**Yes** (82%)
d. Consequences of the test help our patient	**Yes**

Straus C et al. Evidence-based medicine: How to practice & teach & EBM. Elsevier – 3rd edition - 2009

Reproducibility

- **Test results should not vary if repeated by:**
  - Same observer — **Intra-rater** reproducibility
  - Different observer — **Inter-rater** reproducibility
  - Same or different locations

- **Reproducibility measurement**
  - Dichotomous results — **Kappa & its variants**
  - Continuous results — Within-subject standard deviation
    - Within-subject coefficient variation
    - Correlation coefficients
    - **Bland-Altman plot**

Newman TB & Kohn MA. Evidence-based diagnosis.
Cambridge University Press, Cambridge, UK, 1st edition, 2009.

Interpretation of different values of kappa

Kappa from Greek letter κ

Value of kappa	Strength of agreement
0 – 0.20	Poor
0.21– 0.40	Fair
0.41– 0.60	Moderate
0.61– 0.80	Good
0.81–1.00	Very good

kappa score of 0.6 indicates good agreement

Perera R, Heneghan C & Badenoch D. Statistics toolkit.
Blackwell Publishing & BMJ Books, Oxford, 1st edition, 2008.

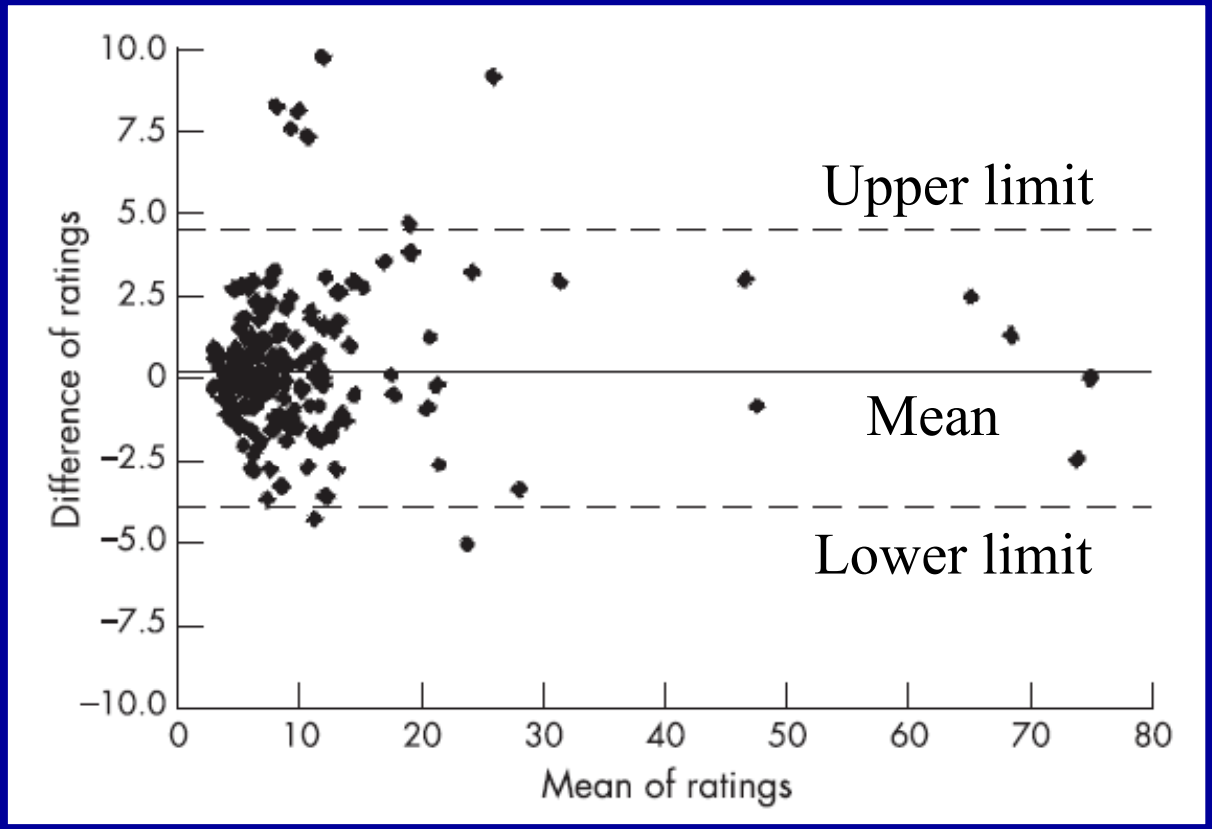
Grading & staging systems for chronic hepatitis

IASL[1]	Batts–Ludwig[2]	Metavir[3]
Grading system (kappa 0.2 – 0.6)		
Minimal activity	Grade 1	A1
Mild activity	Grade 2	A2
Moderate activity	Grade 3	A3
Marked activity	Grade 4	A3
Marked activity & bridging	Grade 4	A3
Staging system (kappa 0.5 – 0.9)		
No fibrosis	Stage 0	F0
Fibrous portal expansion	Stage 1	F1
Few bridges or septa	Stage 2	F2
Numerous bridges	Stage 3	F3
Cirrhosis	Stage 4	F4

[1] Desmet VJ et all. Hepatology 1994;19:1513-1520.
[2] Batts KP et all. Am J Surg Pathol 1995;19:1409-1417.
[3] Bedossa P et all. Hepatology 1996;24:289-293.

Reproducibility of TE in assessing hepatic fibrosis.

Bland Altman Plot

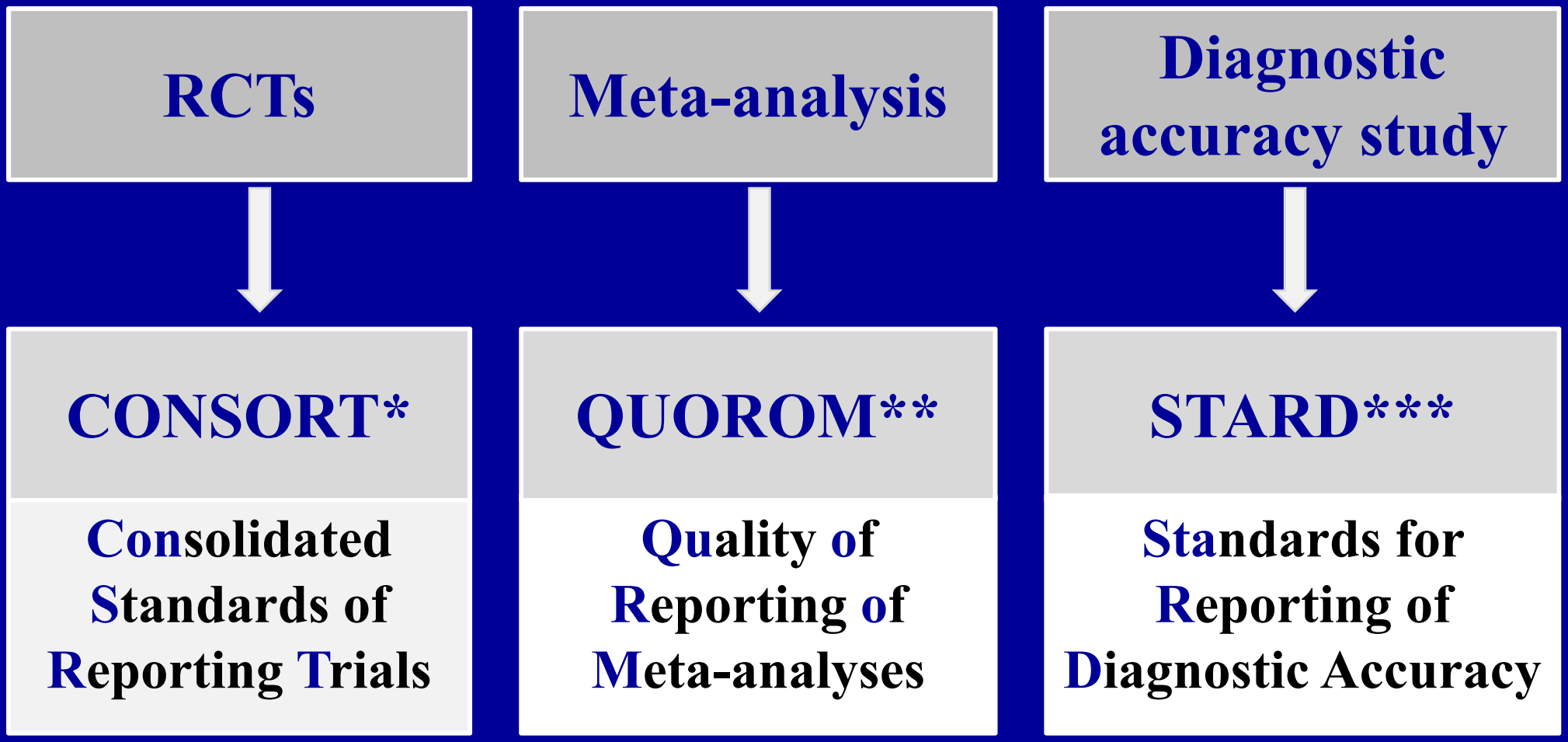


95% limit of agreement

200 patients with CLD & varying etiologies
TE performed twice by 2 different operators within 3 days
8 patients scored outside limits of agreement

Fraquelli M et al. Gut 2007 ; 56 : 968 – 973.

Improving quality of reports



* Altman DG et al. Ann Intern Med 2001 ; 134 : 663 - 94.

** Moher D et al. Lancet 1999 ; 354 : 1896 - 900.

*** Bossuyt PM et all. BMJ 2003; 326 : 41 – 44.

STARD
Standards for Reporting of Diagnostic Accuracy

- **Date** 2 days consensus meeting (sep16-17, 2000)
- **Invited experts** Researchers – editors – organizations
- **Literature** 33 checklists for diagnostic research
  75 different items
- **Aim** Reduce extended list of potential items
- **Results** **Checklist** with 25 items based on evidence
  Flow diagram

Bossuyt PM et all. BMJ 2003 ; 326 : 41 – 4.

Challenges of a diagnostic test study

- Conducting a high-quality diagnostic test study is very challenging at every step of the way:
  - Formulating the proposal
  - Obtaining funding
  - Carrying out the proposal
- To have an idea of the challenges involved ACP Journal Club in 2003 published **86 RCTs** & only **7 diagnostic test studies**

Accuracy of a diagnostic test -1

- **Dichotomous test (only 2 results)**
  - Sensibility (**Sn**) & Specificity (**Sp**)
  - Positive Predictive Value (**PPV**)
  - Negative Predictive Value (**NPV**)
  - Likelihood Ratios + & – (**LRs**)
  - Diagnostic Odds Ratio (**OR**)

  } **CIs**

- **Multilevel test (> 2 results)**
  - Receiver Operating Characteristic (**ROC**)
  - **LR line**

Newman TB & Kohn MA. Evidence-based diagnosis.
Cambridge University Press, Cambridge, UK, 1st edition, 2009.

Accuracy of a diagnostic test -2

Sensitivity	Percent of diseased individuals who have + test
Specificity	Percent of non-diseased who have negative test
PPV	Percent of positive tests that are truly positive
NPV	Percent of negative tests that are truly negative
Pre-test probability	Likelihood of disease before dg test result
LRs	Change of disease probability given a test result
Post-test probability	Likelihood of disease after dg test result
ROC curve	Accuracy of dg test with multiple levels
LR line	Accuracy of dg test if large no. of observations

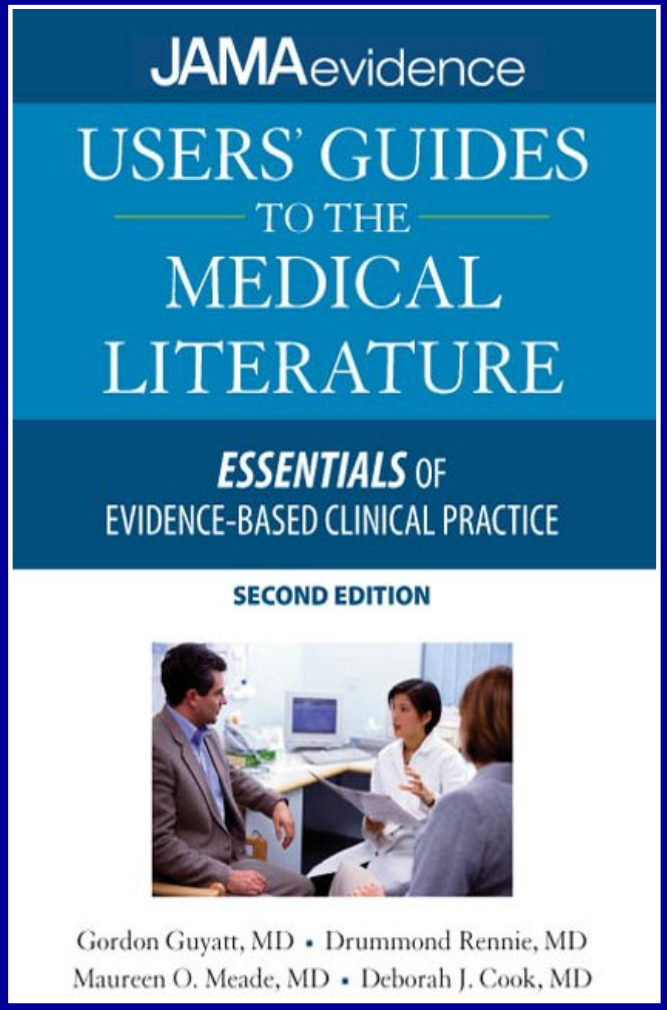
ERIC notebook in epidemiology. http://hrsd.durham.med.va.gov/eric/

Critical appraisal of a diagnostic study

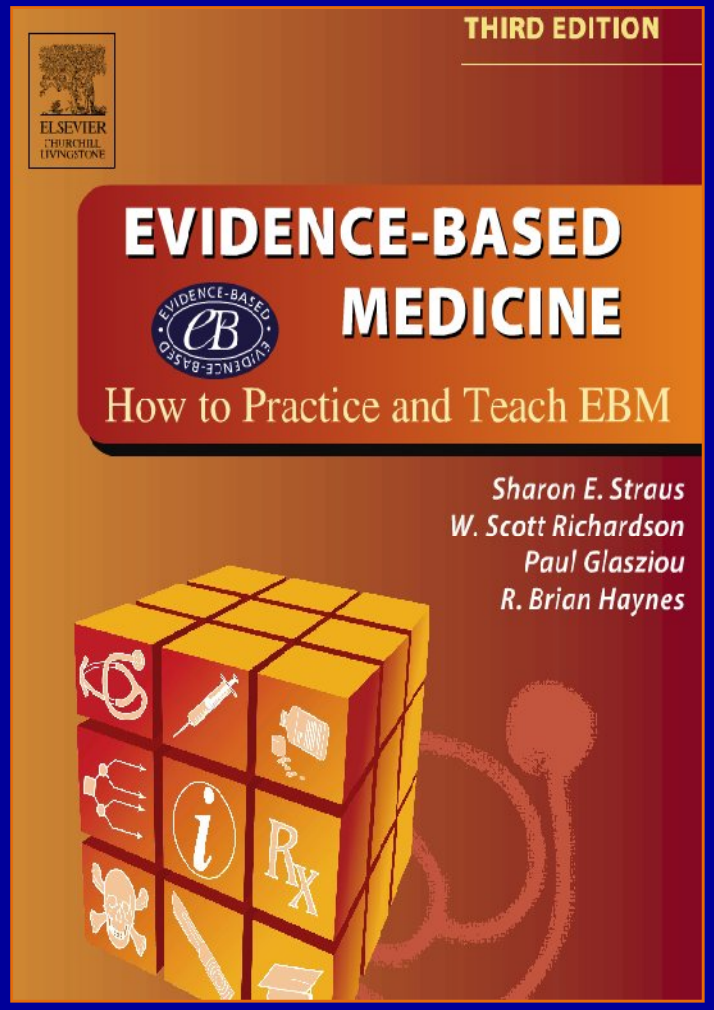
Valid	Blind comparison with gold standard test Gold standard test performed in all patients Dg test done in appropriate spectrum of pts
Accurate	Accuracy: Sn, Sp, PPV, NPV, LR, diagnostic OR Precision of the results
	Dichotomizing continuous test scores
Apply	Test available & reproducible in my setting Generate sensible pre-test probability Post-test probability affects management Consequences of test help my patient

Straus C et al. Evidence-based medicine: How to practice & teach & EBM. Elsevier – 3rd edition – 2009.

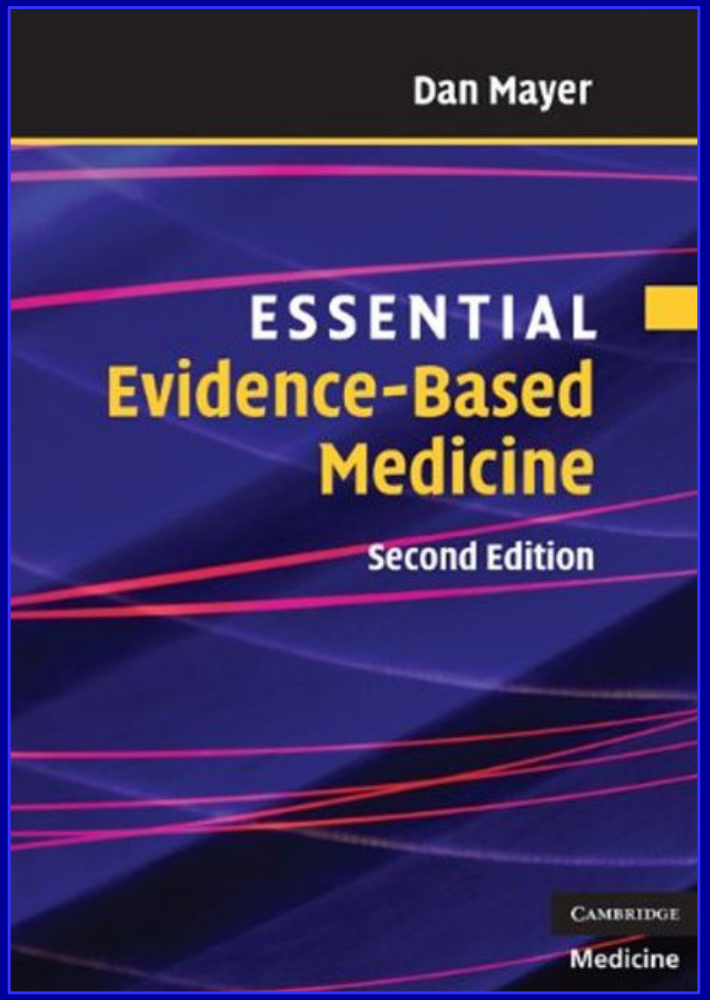
References



Mc Graw Hill
2008

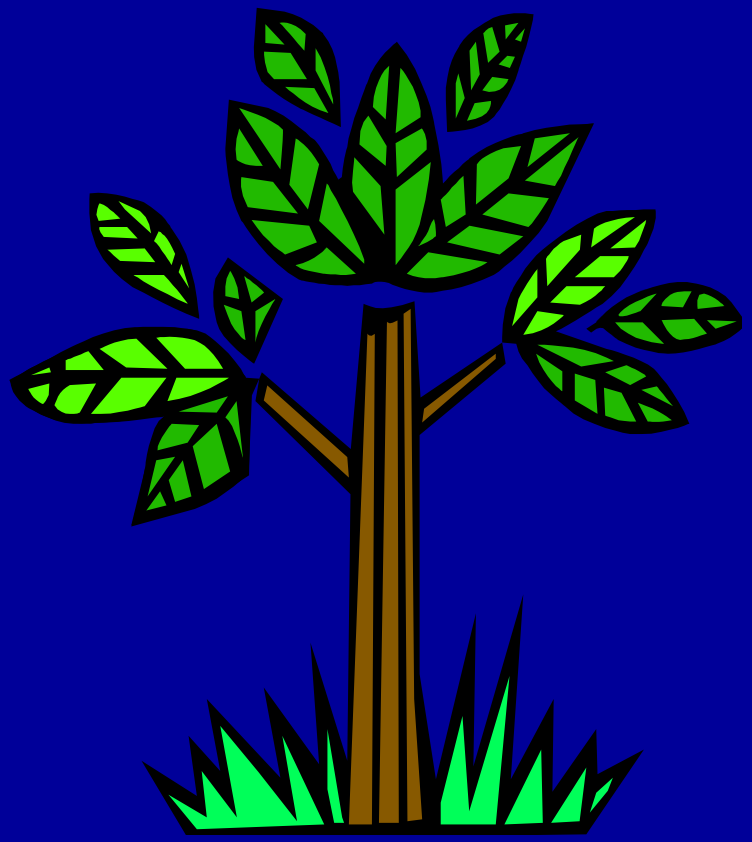


Elsevier
2009



Cambridge University Press
2010

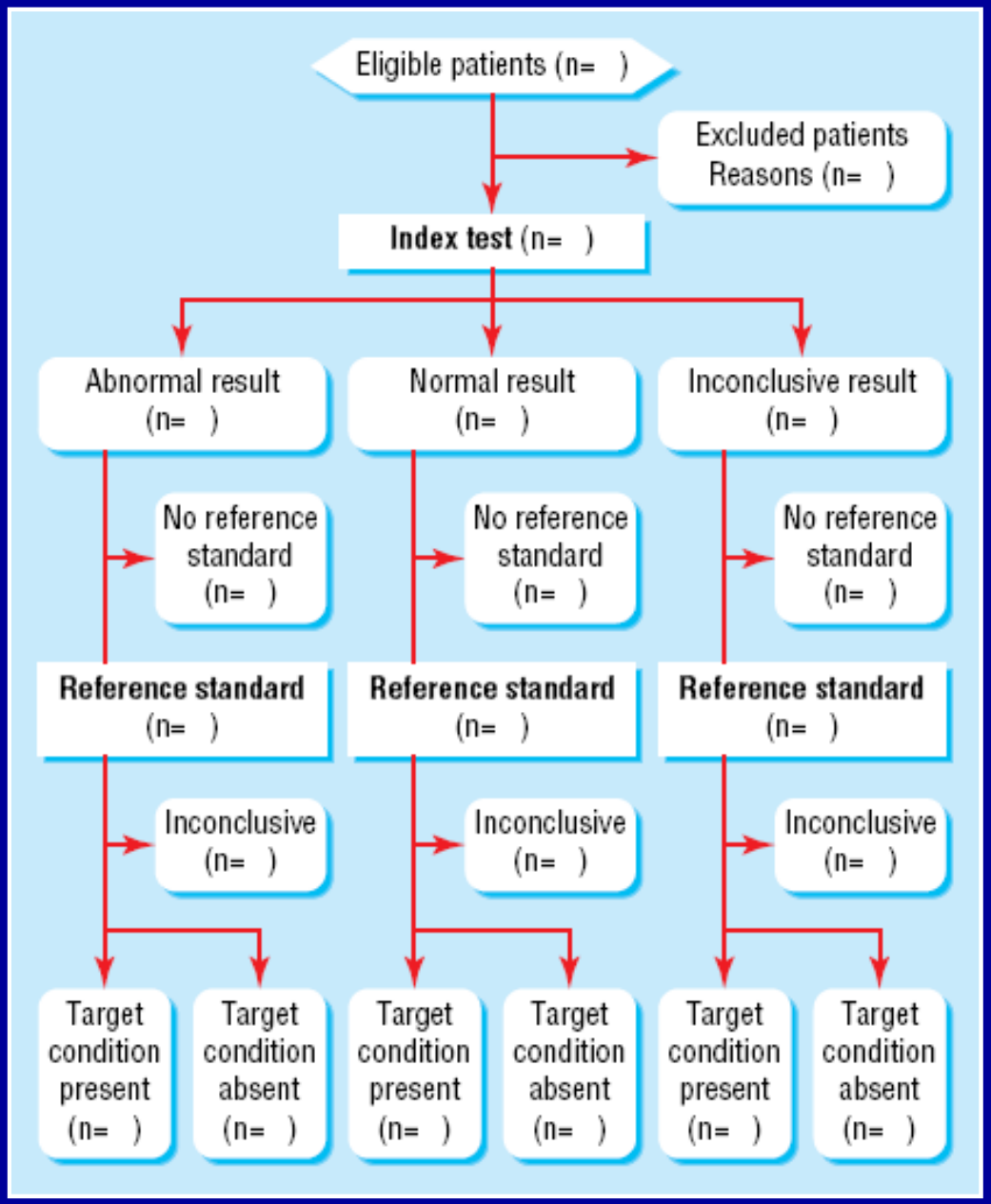
Thank You



STARD checklist for reporting diagnostic accuracy studies

Section and topic	Item	Description
Title, abstract, & keywords	**1**	Identify article as a study of diagnostic accuracy (**sensibility, specificity**)
Introduction	**2**	State research questions (estimate **dg accuracy** or accuracy between tests)
Methods: Participants	**3**	Inclusion & exclusion criteria, settings, & locations
	4	Describe participant recruitment: presenting symptoms, previous tests
	5	Describe participant sampling: consecutive series of participants or not
	6	Describe data collection: prospective or retrospective
Test methods	**7**	Describe reference standard & its rationale
	8	Describe **technical specifications of material & methods**
	9	Describe definition & rationale for units: index tests & reference standard
	10	Describe number, training, & expertise of persons executing index tests
	11	Were readers of the index tests & reference standard **blind** ?
Statistical methods	**12**	Describe methods for calculating **dg accuracy** & quantify uncertainty (**CI**)
	13	Describe methods for calculating test **reproducibility**, if done
Results: Participants	**14**	Report when study was done including begin & end dates of recruitment
	15	Report clinical & demographic characteristics (age, sex, symptoms, treatment)
	16	Report how many participants undergo index tests or reference standard, or both
Test results	**17**	Report time interval from index tests to reference standard & treatment between
	18	Report **distribution of severity of disease** (define criteria)
	19	Report cross tabulation of results of the index tests by reference standard
	20	Report **adverse events** from performing index test or reference standard
Estimates	**21**	Report estimates of **dg accuracy** & measures of **statistical uncertainty (CI)**
	22	Report how indeterminate results, missing responses, & outliers were handled
	23	Report estimates of **variability of dg accuracy between readers**, centers,…
	24	Report estimates of test **reproducibility**, if done
Discussion	**25**	Discuss clinical **applicability** of findings

Flow diagram for study on diagnostic accuracy



Bossuyt PM et all. BMJ 2003 ; 326 : 41 – 4.

Look at things CRITICALLY

Thank You

Example of a diagnostic test

- **Target condition** **Choledocholithiasis**
- **Gold standard test** **ERCP (costlier, riskier)**

- **Diagnostic test** **MRCP**
- **Accuracy** **Sn – Sp – PPV – NPV – LRs – CI**

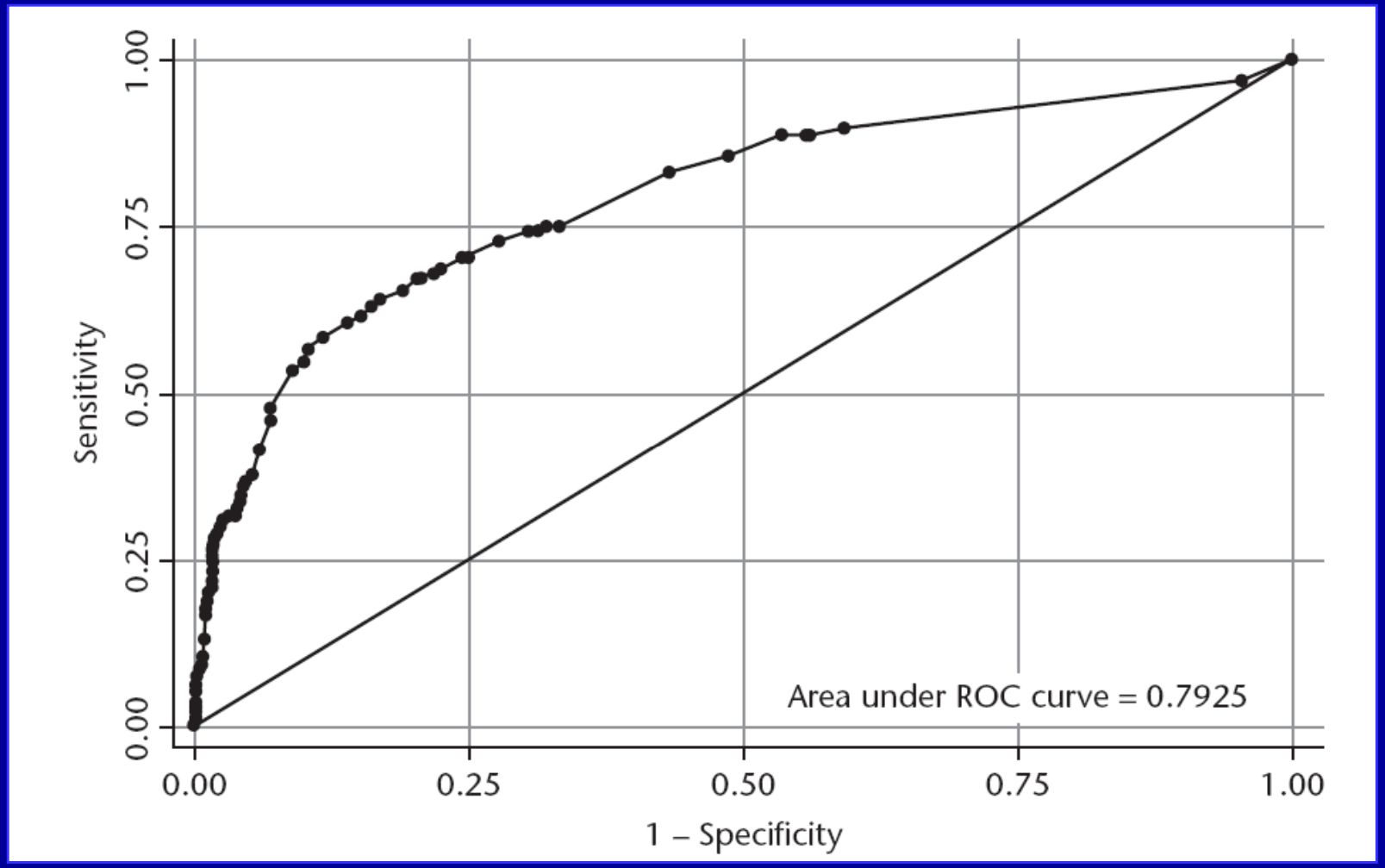
Blinding in diagnostic test

- **Target condition** **Choledocholithiasis**
- **Gold standard test** **ERCP**
- **Diagnostic test** **MRCP**
- **Accuracy** **Sn – Sp – PPV – NPV – LRs – CIs**

Biased interpretation of MRC if radiologist know that

ERCP shows choledocholithiasis

Hypothetical ROC curve



Pines JM & Everett WW. Evidence-Based emergency care: diagnostic testing & clinical decision rules. Blackwell's publishing, West Sussex, UK, 2008.