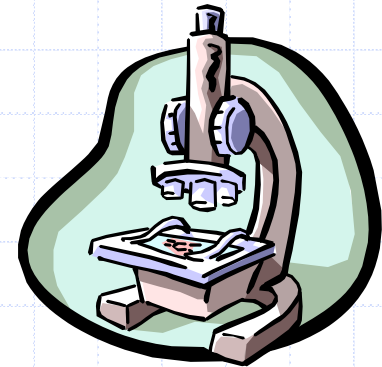
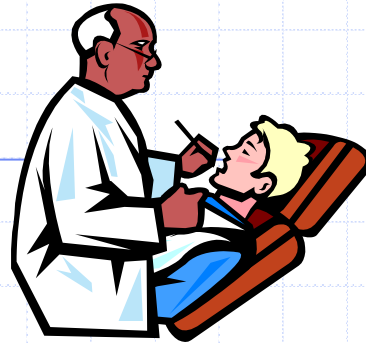
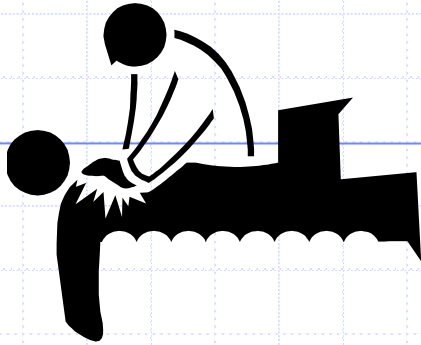


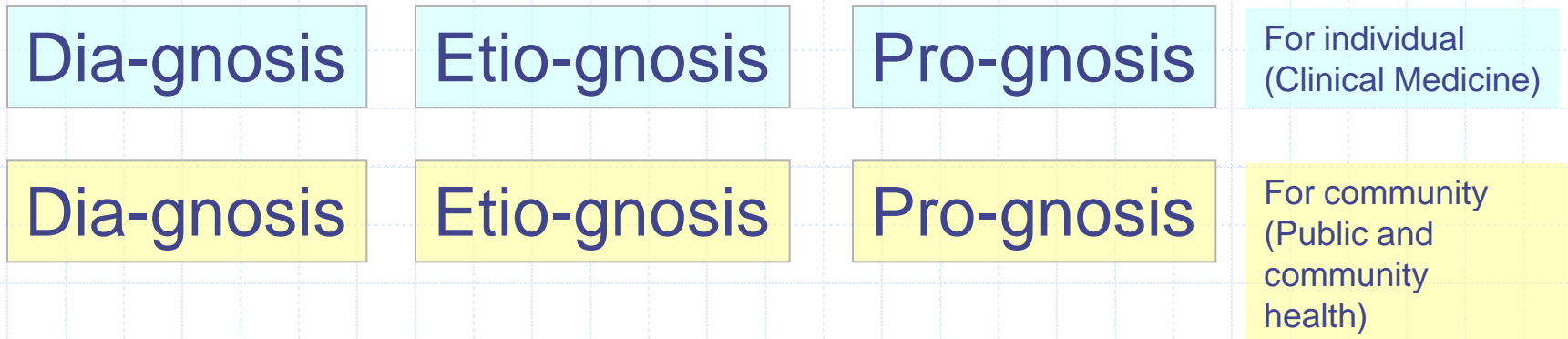
Diagnostic studies



Madhukar Pai, MD, PhD
Assistant Professor
Dept. of Epidemiology, Biostatistics & Occupational Health
McGill University
Montreal, Canada H3A 1A2
Email: madhukar.pai@mcgill.ca

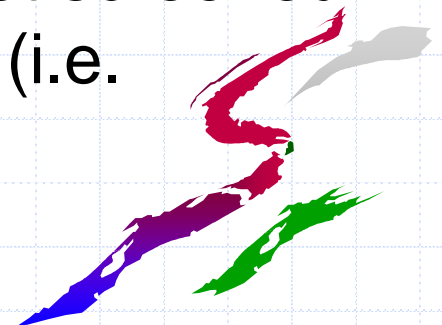
Diagnosis: why does it matter?

- ◆ To effectively practice medicine and public health, we need evidence/knowledge on 3 fundamental types of professional knowing “gnosis”:

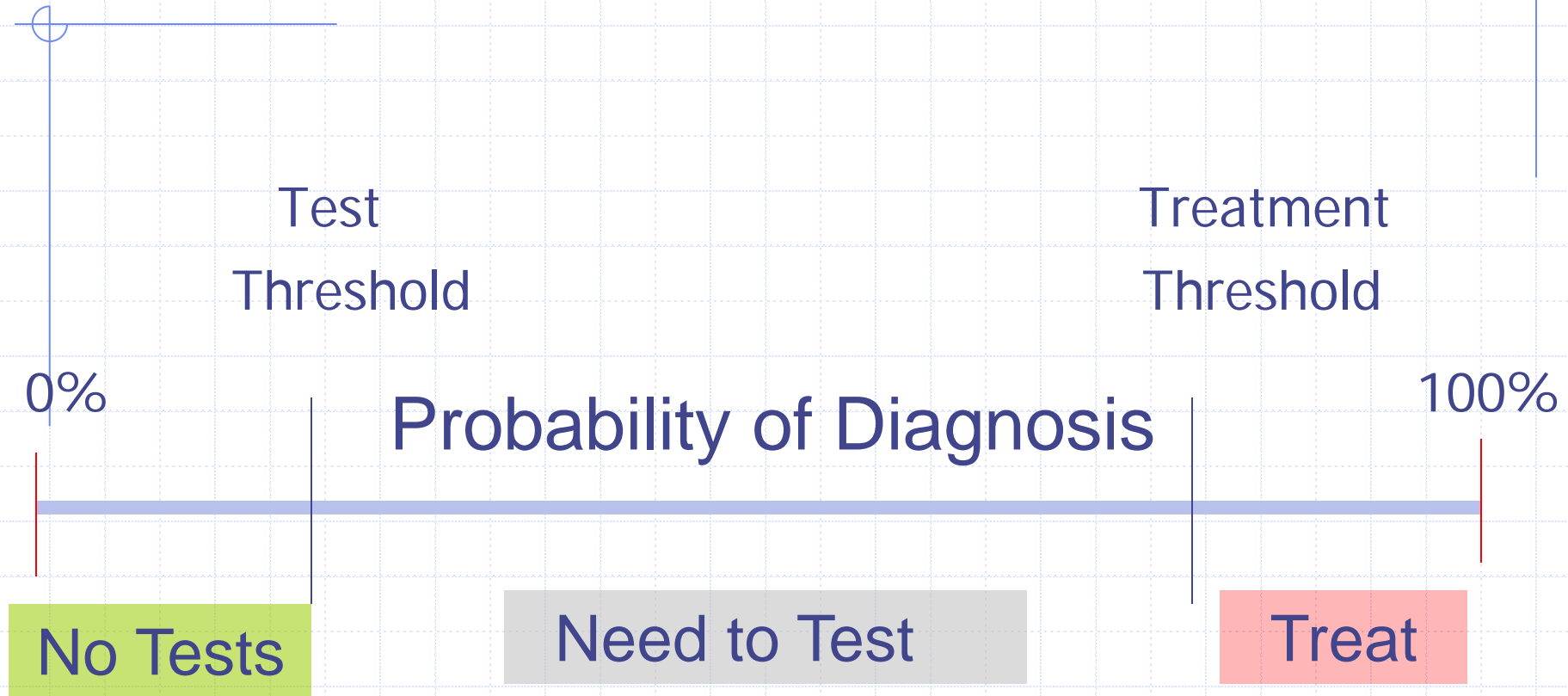


Diagnosis Vs Screening

- A diagnostic test is done on sick people
 - patient presents with symptoms
 - pre-test probability of disease is high (i.e. disease prevalence is high)
- A screening test is usually done on asymptomatic, apparently healthy people
 - healthy people are encouraged to get screened
 - pre-test probability of disease is low (i.e. disease prevalence is low)



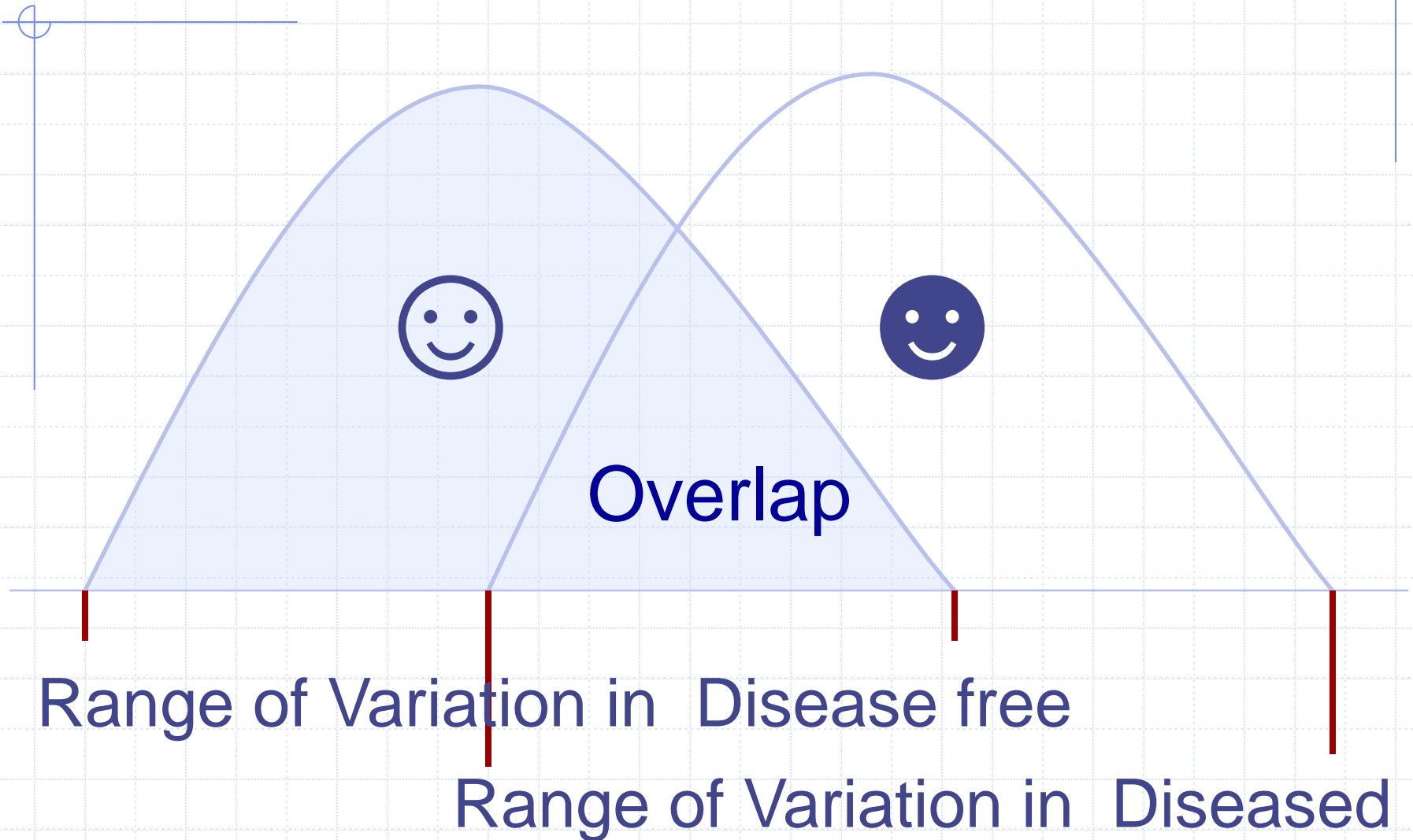
Process of diagnosis: all about probability!



The Perfect Diagnostic Test




Variations In Diagnostic Tests



There is no perfect test!

Thomas Bayes



Thomas Bayes (The correct identification of this portrait has been [1] questioned.)

Born	c. 1702 London
Died	April 17th 1761 Tunbridge Wells
Nationality	British

LII. *An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S.*

Dear Sir,

Read Dec. 23, 1763. I now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved. Experimental philosophy, you will find, is nearly interested in the subject of it; and on this account there seems to be particular reason for thinking that a communication of it to the Royal Society cannot be improper.

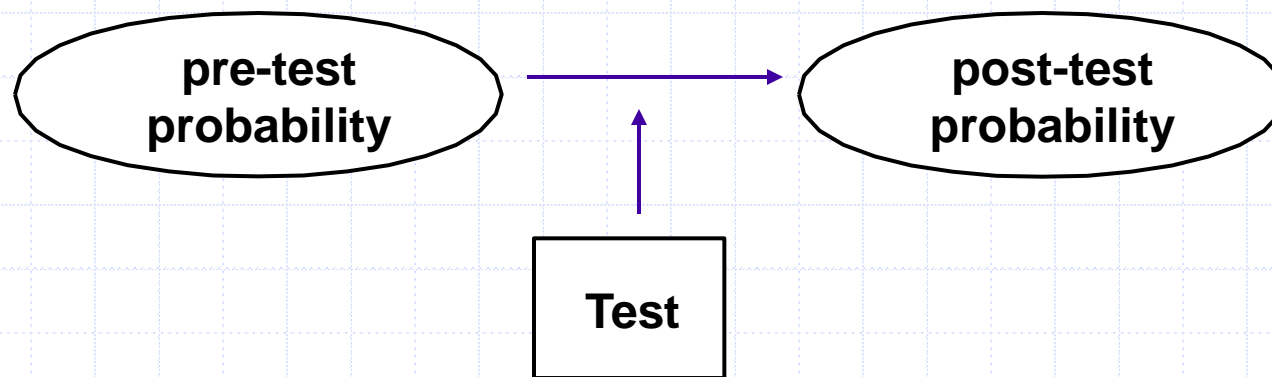
He had, you know, the honour of being a member of that illustrious Society, and was much esteemed by many as a very able mathematician. In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times. He adds, that he soon perceived that it would not be very difficult to do this, provided some rule could be found, according to which we ought to estimate the chance that the probability for the happening of an event perfectly unknown, should lie between any two named degrees of prob-

All we can hope to do is increase or decrease probabilities, and Bayes' theorem helps with this process

Bayes' theory

- Bayes' Theorem is a simple mathematical formula used for calculating conditional probabilities
- every test is done with a certain probability of disease - degree of suspicion [pre-test or prior probability]
- the probability of disease after the test result is the post-test or posterior probability

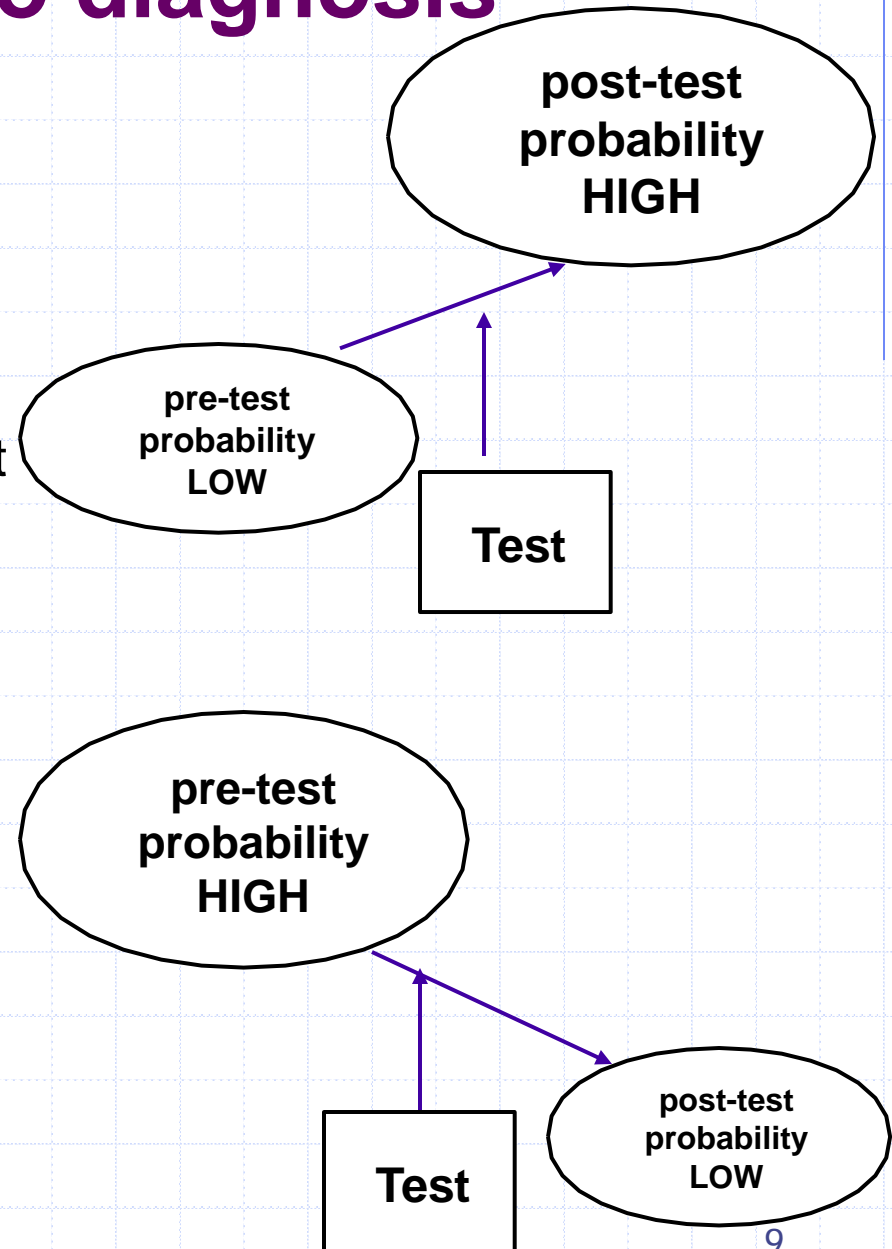
What you thought before + New information = What you think now



Post-test odds = Pre-test odds x Likelihood ratio

Bayesian approach to diagnosis

- An accurate test will help reduce uncertainty
- The pre-test probability is revised using test result to get the post-test probability
- Tests that produce the biggest changes from pretest to post-test probabilities are most useful in clinical practice [very large or very small likelihood ratios]



Evaluating a diagnostic test

- Define gold standard
- Recruit consecutive patients in whom the test is indicated (in whom the disease is suspected)
- Perform gold standard and separate diseased and disease free groups
- Perform test on all and classify them as test positives or negatives
- Set up 2 x 2 table and compute:
 - Sensitivity
 - Specificity
 - Predictive values
 - Likelihood ratios

Evaluating a diagnostic test

- Diagnostic 2 X 2 table*:

	Disease +	Disease -
Test +	True Positive	False Positive
Test -	False Negative	True Negative

*When test results are not dichotomous, then can use ROC curves [see later] 11

Sensitivity

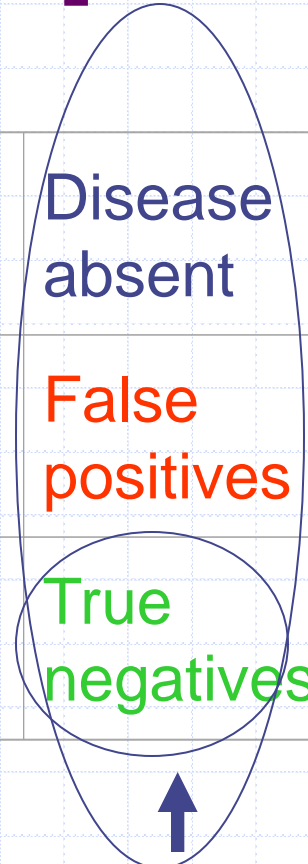
[true positive rate]

	Disease present	Disease absent
Test positive	True positives	False positives
Test negative	False negative	True negatives

The proportion of patients with disease who test positive = $P(T+|D+) = TP / (TP+FN)$

Specificity

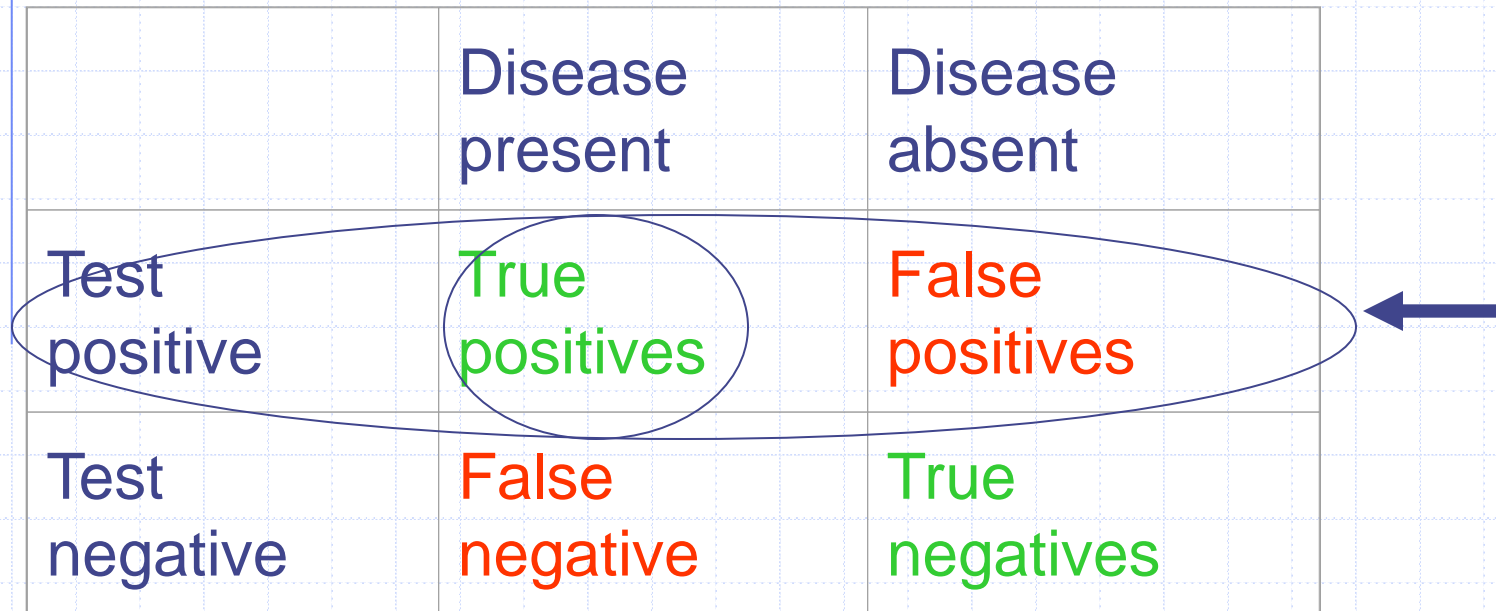
[true negative rate]



	Disease present	Disease absent
Test positive	True positives	False positives
Test negative	False negative	True negatives

The proportion of patients without disease who test negative: $P(T-|D-) = TN / (TN + FP)$.


Predictive value of a positive test



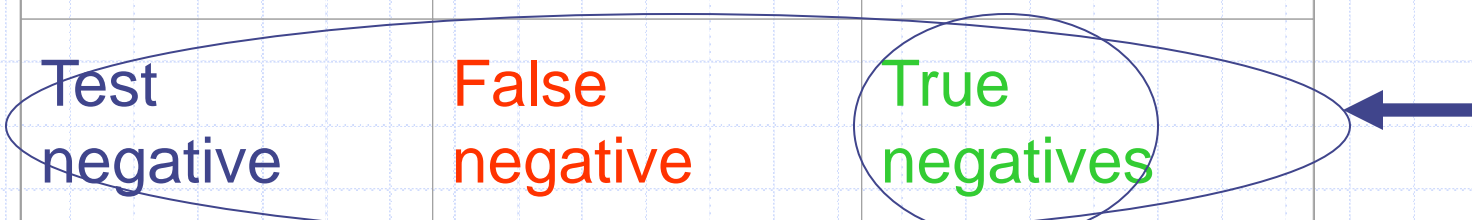
	Disease present	Disease absent
Test positive	True positives	False positives
Test negative	False negative	True negatives

Proportion of patients with positive tests who have disease = $P(D+|T+) = TP / (TP+FP)$

Predictive value of a negative test



	Disease present	Disease absent
Test positive	True positives	False positives
Test negative	False negative	True negatives



Proportion of patients with negative tests who do not have disease = $P(D^-|T^-) = TN / (TN+FN)$

Example: Serological test for TB

Culture (gold standard)

Yes

No

Serological Test

Positive

14

3

17

Negative

54

28

82

68

31

99

Sensitivity = 21%

Specificity = 90%

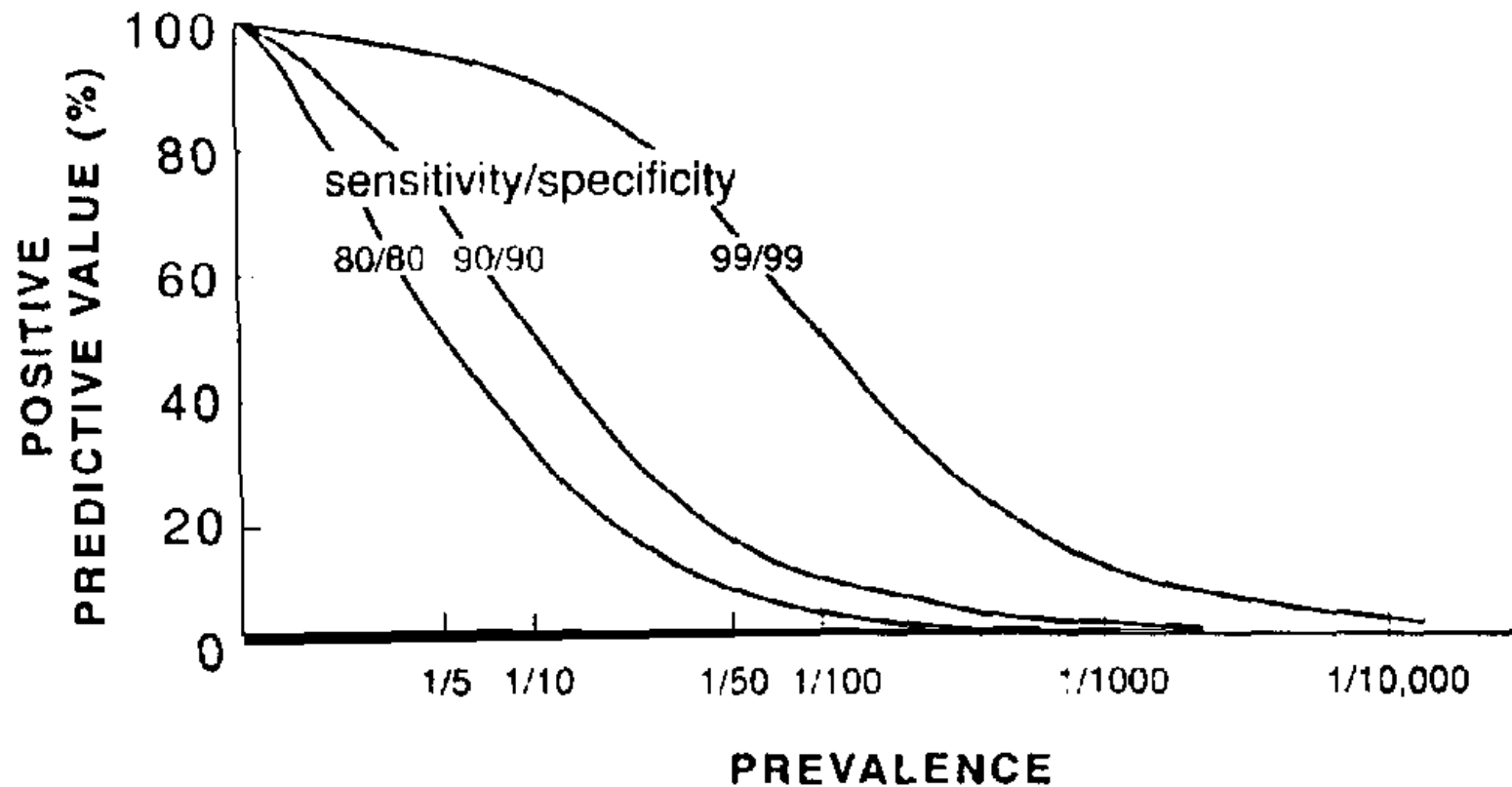
For a given test, predictive values will depend on prevalence

Effect of Prevalence on Predictive Value: Positive Predictive Value of Prostatic Acid Phosphatase for Prostatic Cancer (Sensitivity = 70%, Specificity = 90%) in Various Clinical Settings*

Setting	Prevalence (Cases/100,000)	Positive Predictive Value (%)
General population	35	0.4
Men, age 75 or greater	500	5.6
Clinically suspicious prostatic nodule	50,000	93.0

* From: Watson RA, Tang DB. *N Engl J Med*, 1980; 303:497-499.

For a given test, predictive values will depend on prevalence



Positive predictive value according to sensitivity, specificity, and prevalence of disease.

Likelihood Ratios (also called 'Bayes Factor')

- Likelihood ratio of a positive test: is the test more likely to be positive in diseased than non-diseased persons?

- $LR+ = TPR / FPR$

$$LR+ = \frac{\Pr(T+ | D+)}{\Pr(T+ | D-)}$$

- High LR+ values help in RULING IN the disease
- Values close to 1 indicate poor accuracy
- E.g. LR+ of 10 means a diseased person is 10 times more likely to have a positive test than a non-diseased person

Likelihood Ratio of a Positive Test

	Disease present	Disease absent
Test positive	True positives	False positives
Test negative	False negative	True negatives

How more often a positive test result occurs in persons with compared to those without the target condition

$$LR+ = TPR / FPR$$

$$LR+ = \frac{\Pr(T+ | D+)}{\Pr(T+ | D-)}$$

Likelihood Ratios

- Likelihood ratio of a negative test: is the test less likely to be negative in the diseased than non-diseased persons?

- $LR- = FNR / TNR$

$$LR- = \frac{\Pr(T- | D+)}{\Pr(T- | D-)}$$

- Low LR- values help in RULING OUT the disease
- Values close to 1 indicate poor accuracy
- E.g. LR- of 0.5 means a diseased person is half as likely to have a negative test than a non-diseased person

Likelihood Ratio of a Negative Test

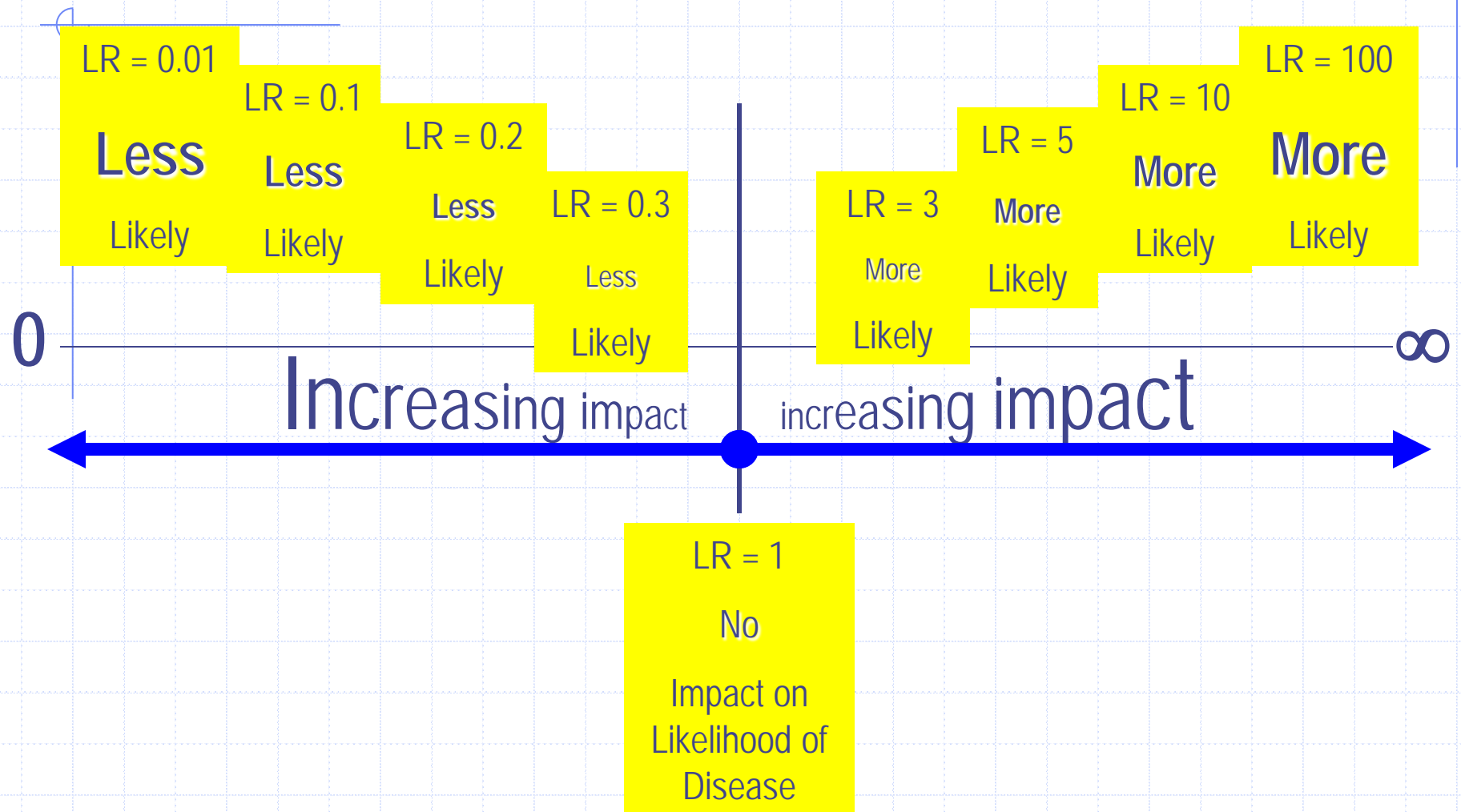
	Disease present	Disease absent
Test positive	True positives	False positives
Test negative	False negative	True negatives

How less likely a negative test result is in persons with the target condition compared to those without the target condition

$$LR- = FNR / TNR$$

$$LR- = \frac{\Pr(T- | D+)}{\Pr(T- | D-)}$$

LR: Impact on Likelihood of Disease



Example: Serological test for TB

Culture (gold standard)

Yes

No

Serological Test

Positive

14

3

17

Negative

54

28

82

68

31

99

$LR+ = 2$

$LR- = 0.9$

Quick review of odds vs. probability

◆ odds = probability / (1 – probability)

$$\text{Odds}(D+) = \frac{\text{Pr}(D+)}{1 - \text{Pr}(D+)}$$

◆ probability = odds / (1 + odds)

$$\text{Pr}(D+) = \frac{\text{Odds}(D+)}{1 + \text{Odds}(D+)}$$

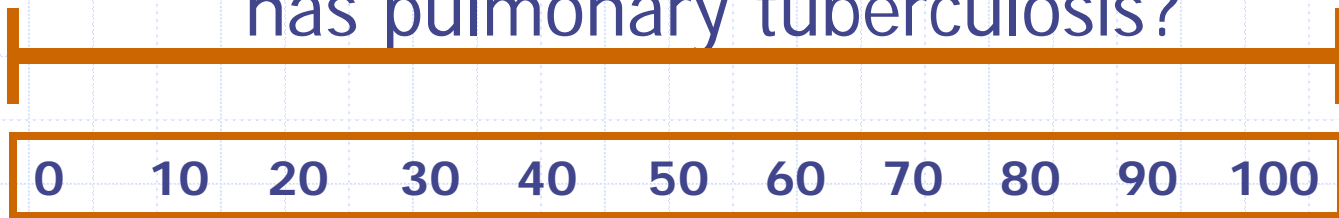
Using LRs in practice

◆ Scenario:

- Mr. A, a 27-year old man
- Recent immigrant from Vietnam
- Fever and productive cough for the past 2 weeks
- Lost weight

Assess the patient and estimate the baseline risk (pre-test probability)

Based on initial history, how likely is it that Mr. A has pulmonary tuberculosis?



Pre-Test Probability

How might the result of a serological test change the likelihood of TB in this patient?

Post-Test Probability

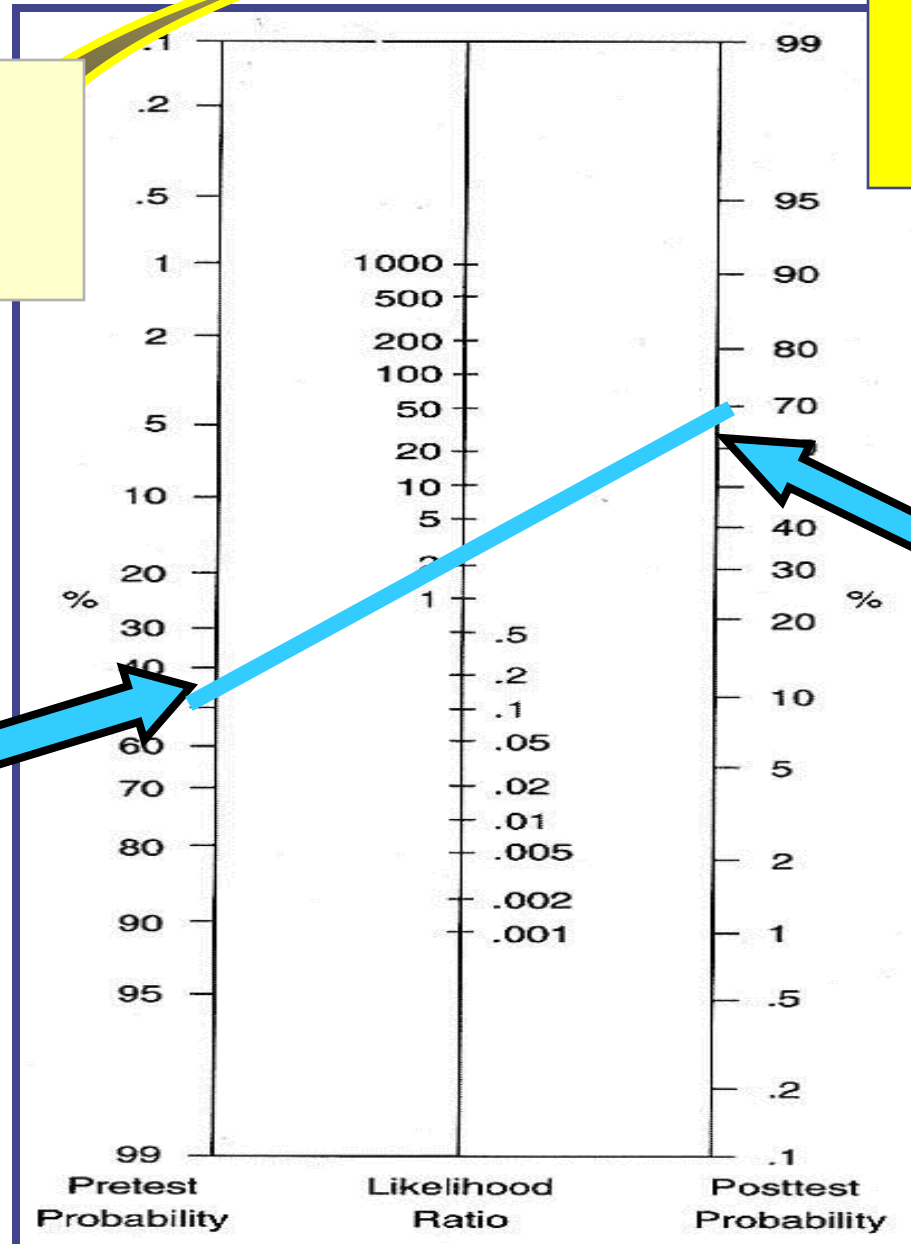
Likelihood Ratios

**Pre-Test
Probability**

**Post-Test
Probability**

**Mr. A
Pre-Test Prob.
50%**

**Post-Test
Prob. 70%**



Serological test
 $LR+ = 2$

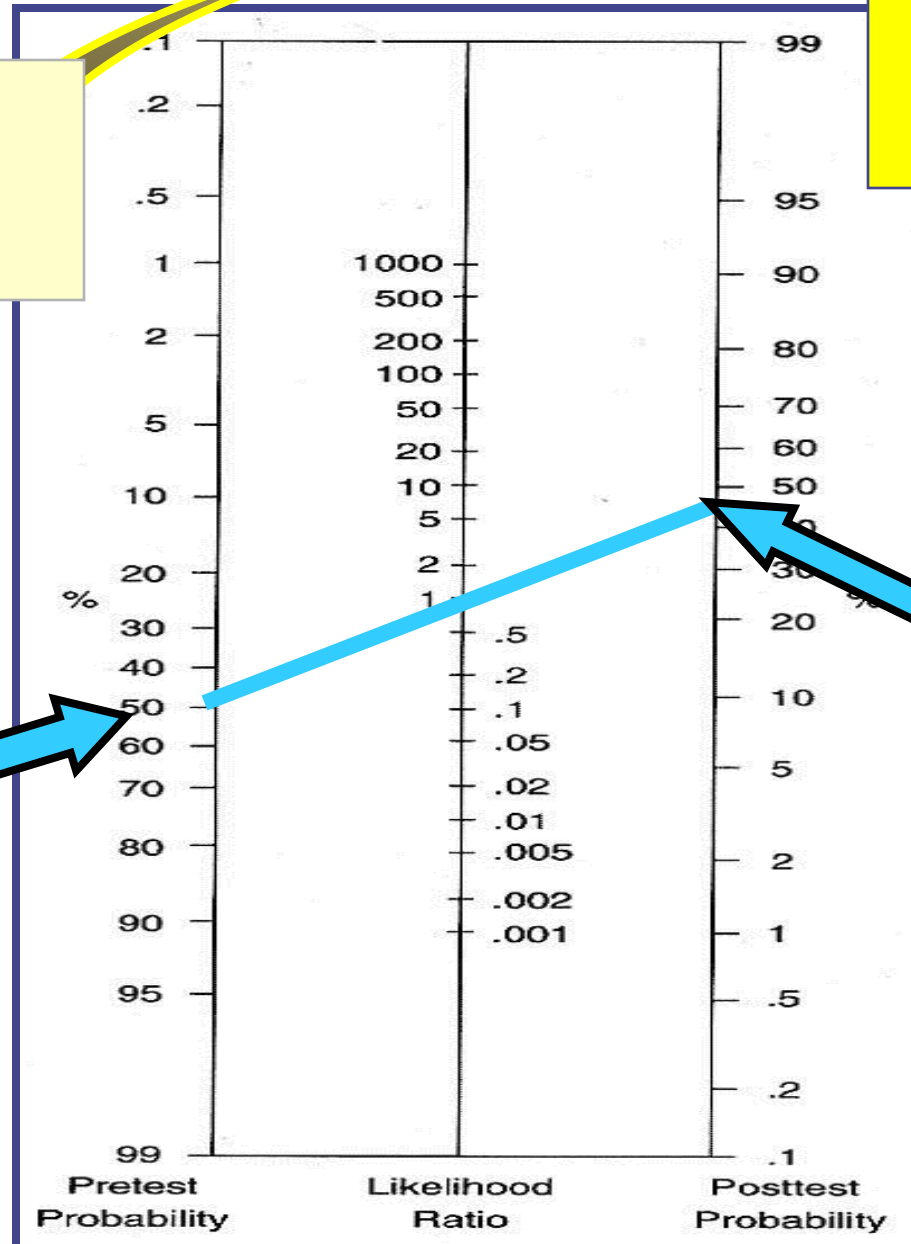
Likelihood Ratios

**Pre-Test
Probability**

**Post-Test
Probability**

**Mr. A
Pre-Test Prob.
50%**

**Post-Test
Prob. 45%**



Serological test
 $LR- = 0.9$

Using LRs in practice

◆ Scenario:

- Ms. B, a 18 year old college student
- Canadian born, no history of foreign travel
- Fever and productive cough for the past 1 week
- Nobody in the household has had TB

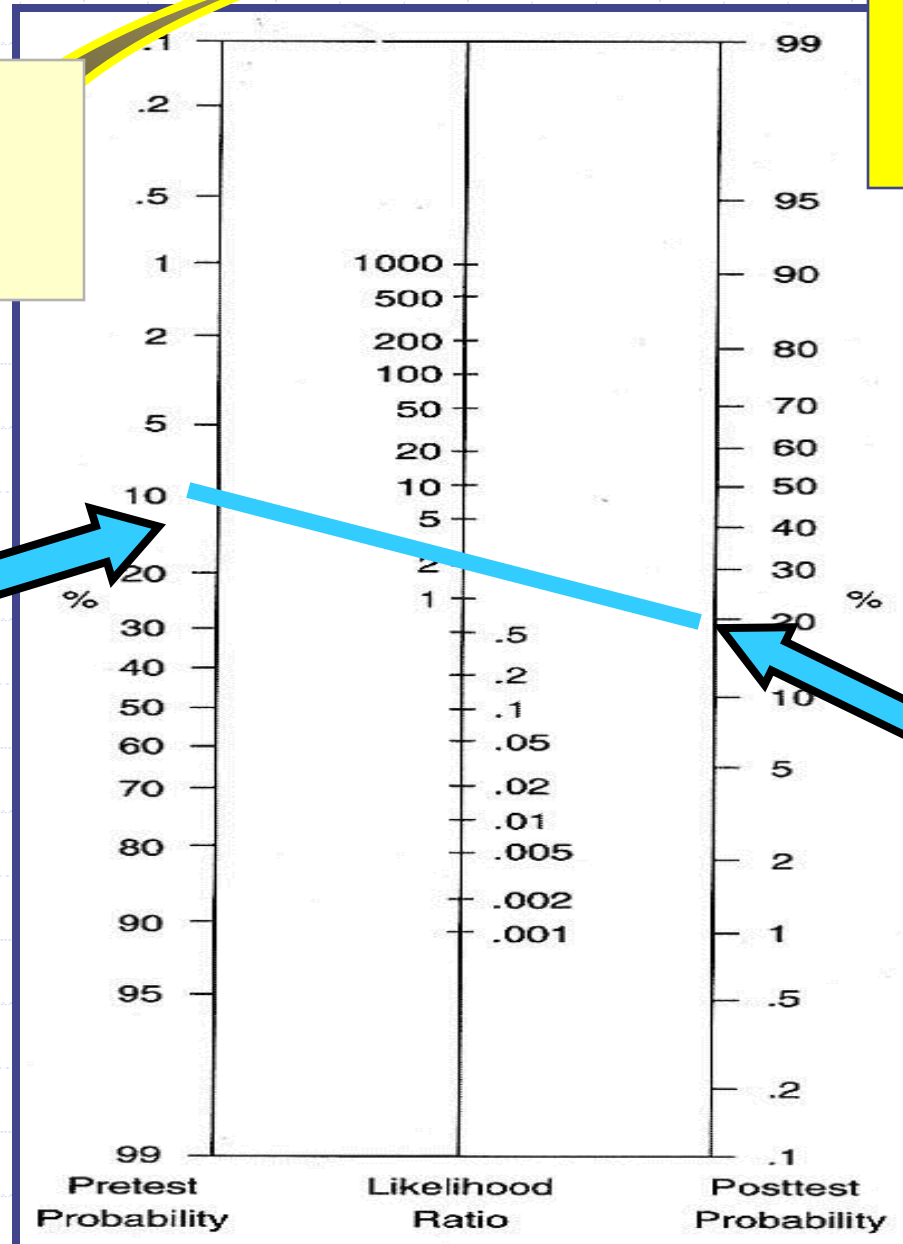
Likelihood Ratios

**Pre-Test
Probability**

**Post-Test
Probability**

**Ms. B
Pre-Test Prob.
10%**

**Post-Test
Prob. 20%**



Serological test
 $LR+ = 2$

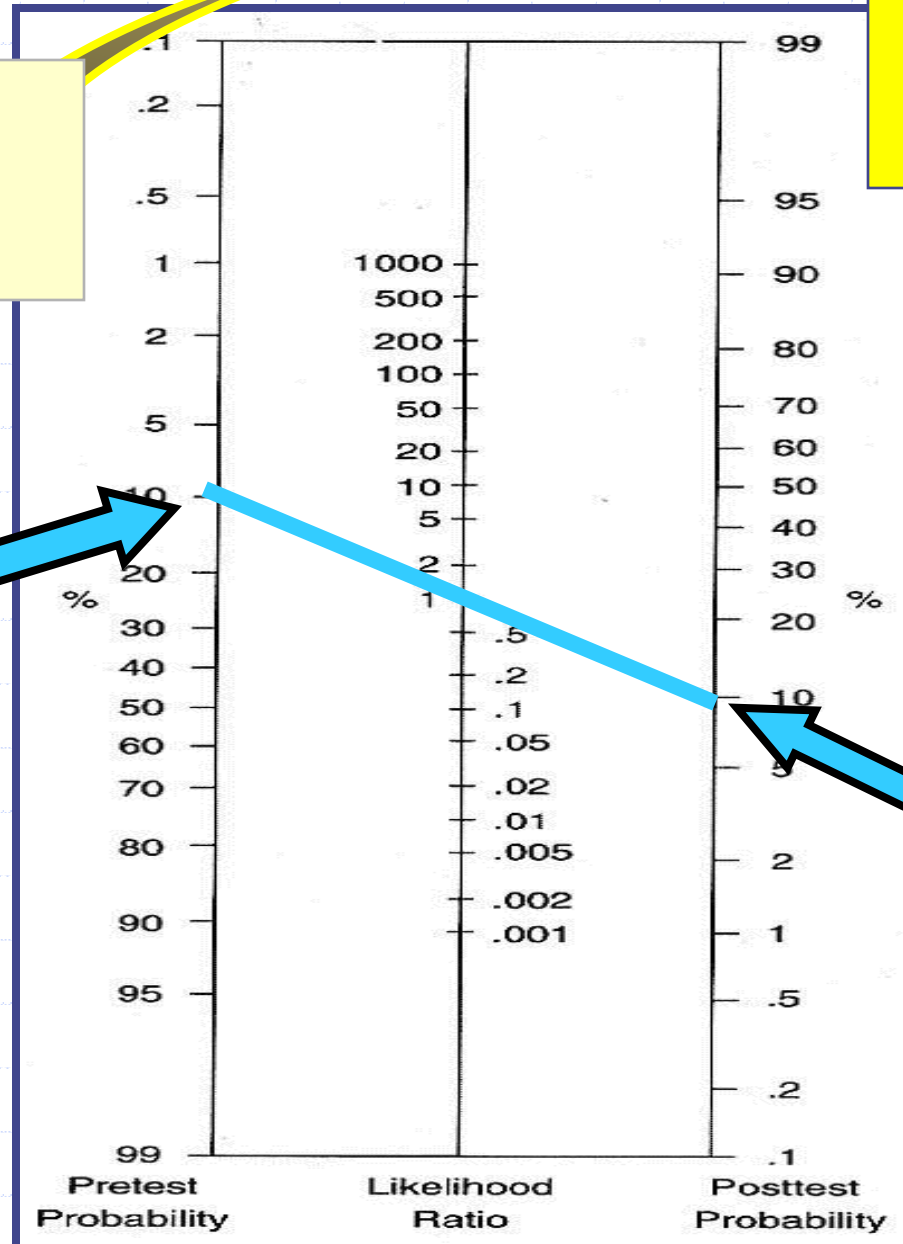
Likelihood Ratios

**Pre-Test
Probability**

**Post-Test
Probability**

**Ms. B
Pre-Test Prob.
10%**

**Post-Test
Prob. 10%**



Serological test
LR- = 0.9

Example: Ultrasonography for Down Syndrome



Another example: Nuchal fold & Down Syndrome

Down Syndrome

		Yes	No	
Nuchal fold	Positive	21	4	25
	Negative	7	188	195
		28	192	220

Sensitivity = 75%

Specificity = 98%

LR+ = 36

LR- = 0.26

DOR = 141

Using LRs in practice

◆ Scenario:

- Mrs. A, a 37-year old woman with a previous affected pregnancy, seen at a high-risk clinic in a tertiary, referral hospital
- What is the pretest probability of Down syndrome in this case?

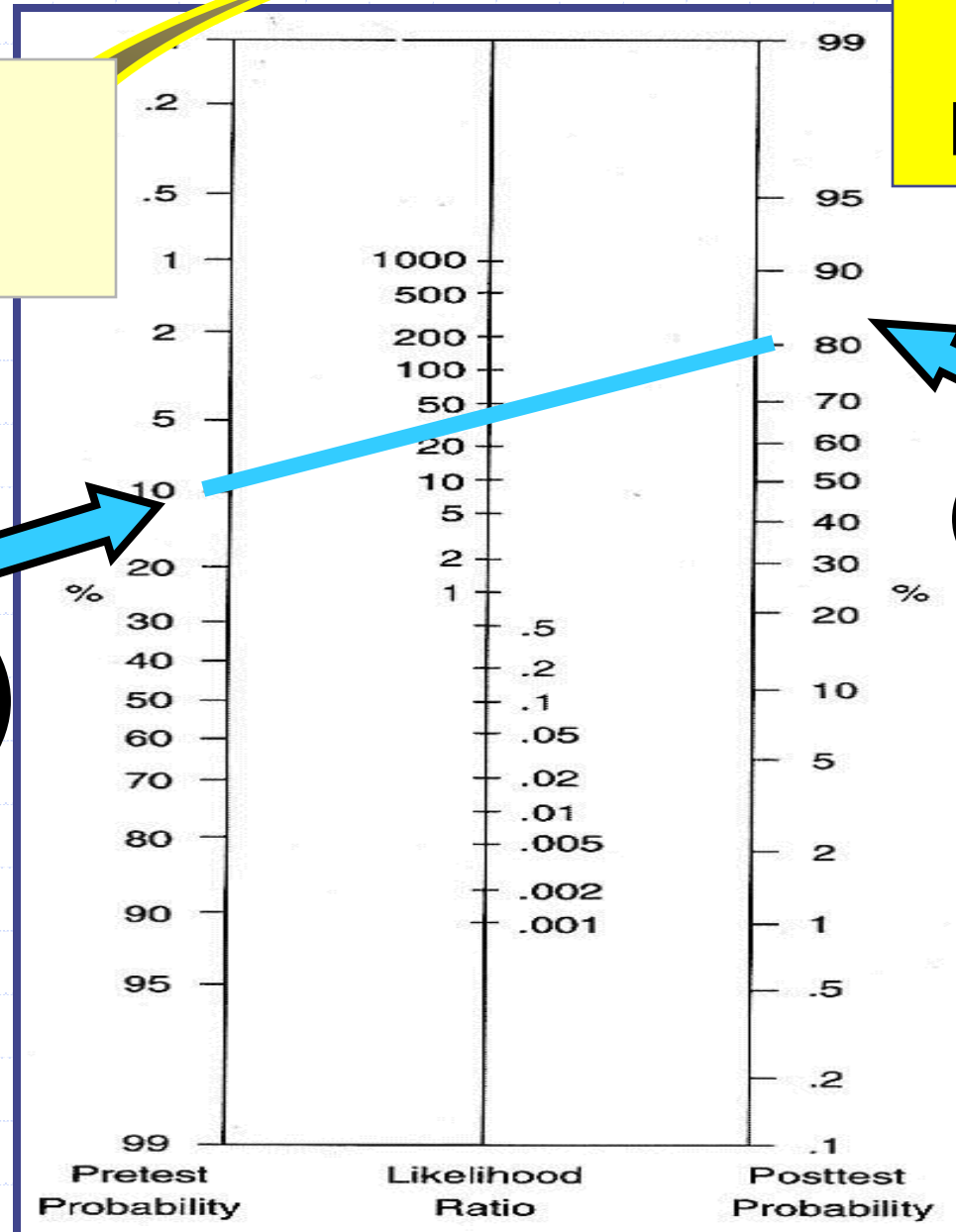
Likelihood Ratios

**Pre-Test
Probability**

**Post-Test
Probability**

**Mrs. A
Pre-Test Prob.
10%**

**Post-Test
Prob. 80%**



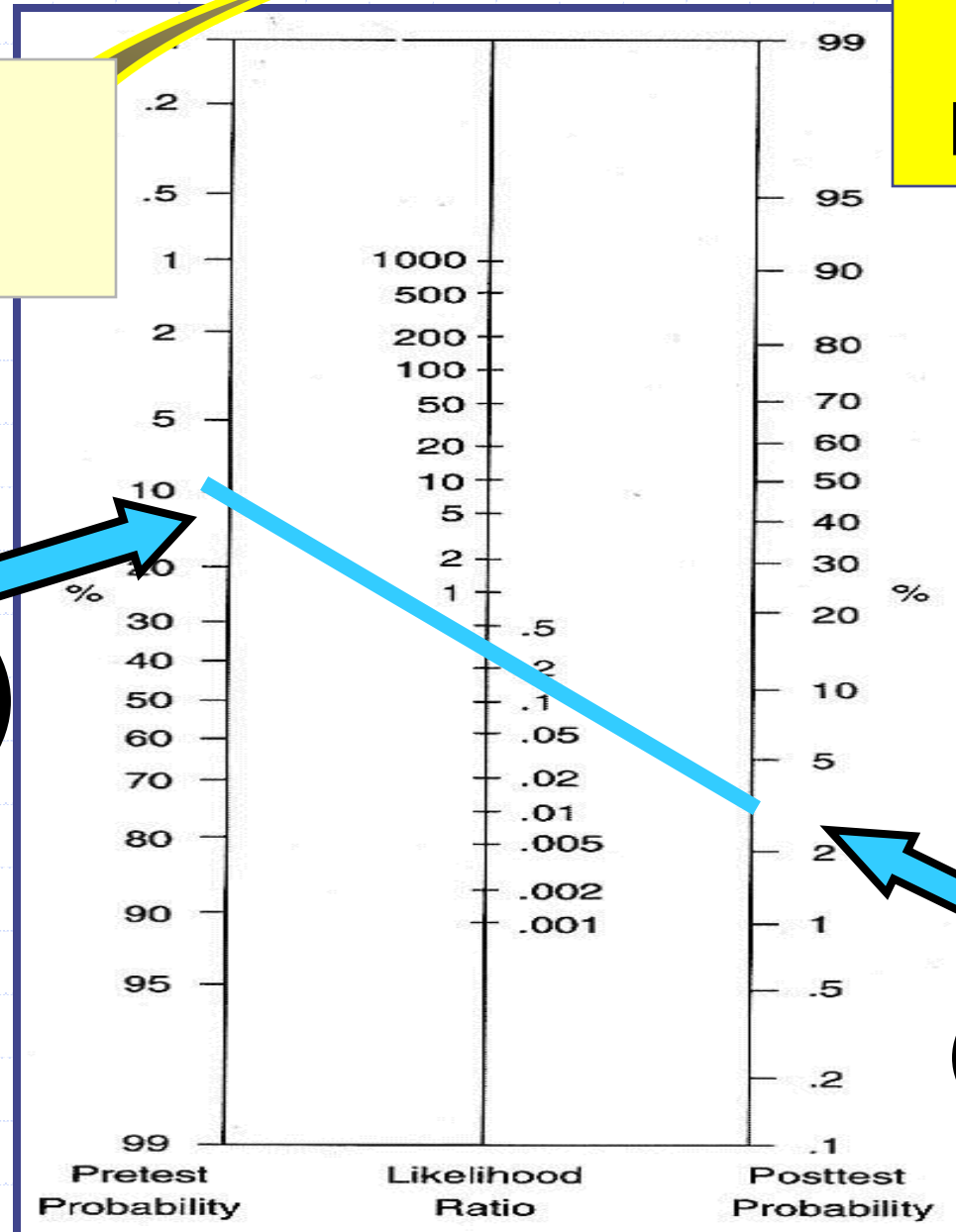
Nuchal fold abnormal
LR = 36

Likelihood Ratios

**Pre-Test
Probability**

**Post-Test
Probability**

**Mrs. A
Pre-Test Prob.
10%**



**Post-Test
Prob. 3%**

Nuchal fold normal
LR = 0.26

Using LRs in practice

◆ Scenario:

- Mrs. B, a 20-year old woman with a previous normal pregnancy, seen at a community hospital
- What is the pretest probability of Down syndrome in this case?

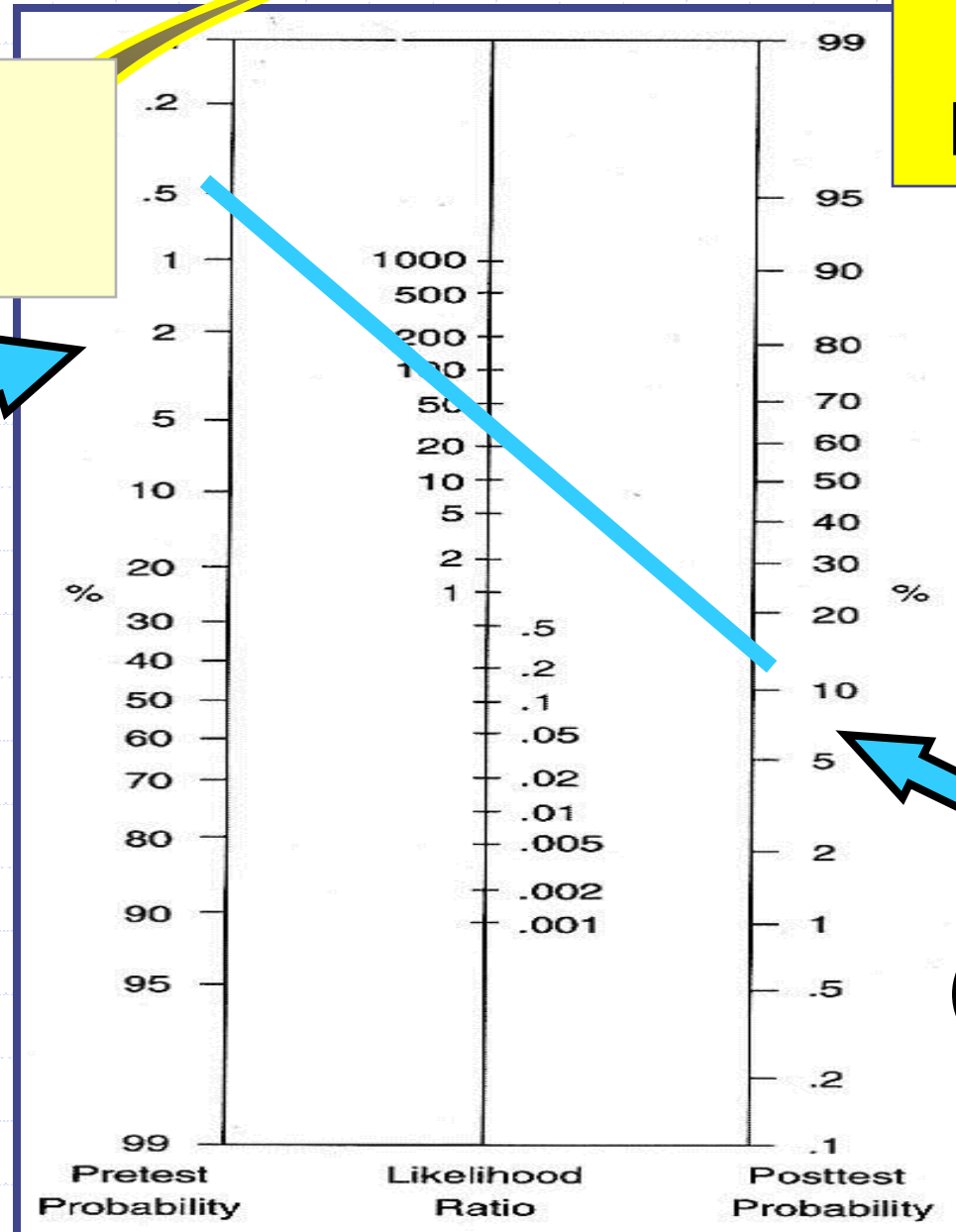
Likelihood Ratios

**Pre-Test
Probability**

**Post-Test
Probability**

**Mrs. B
Pre-Test Prob.
0.5%**

**Post-Test
Prob. 10%**



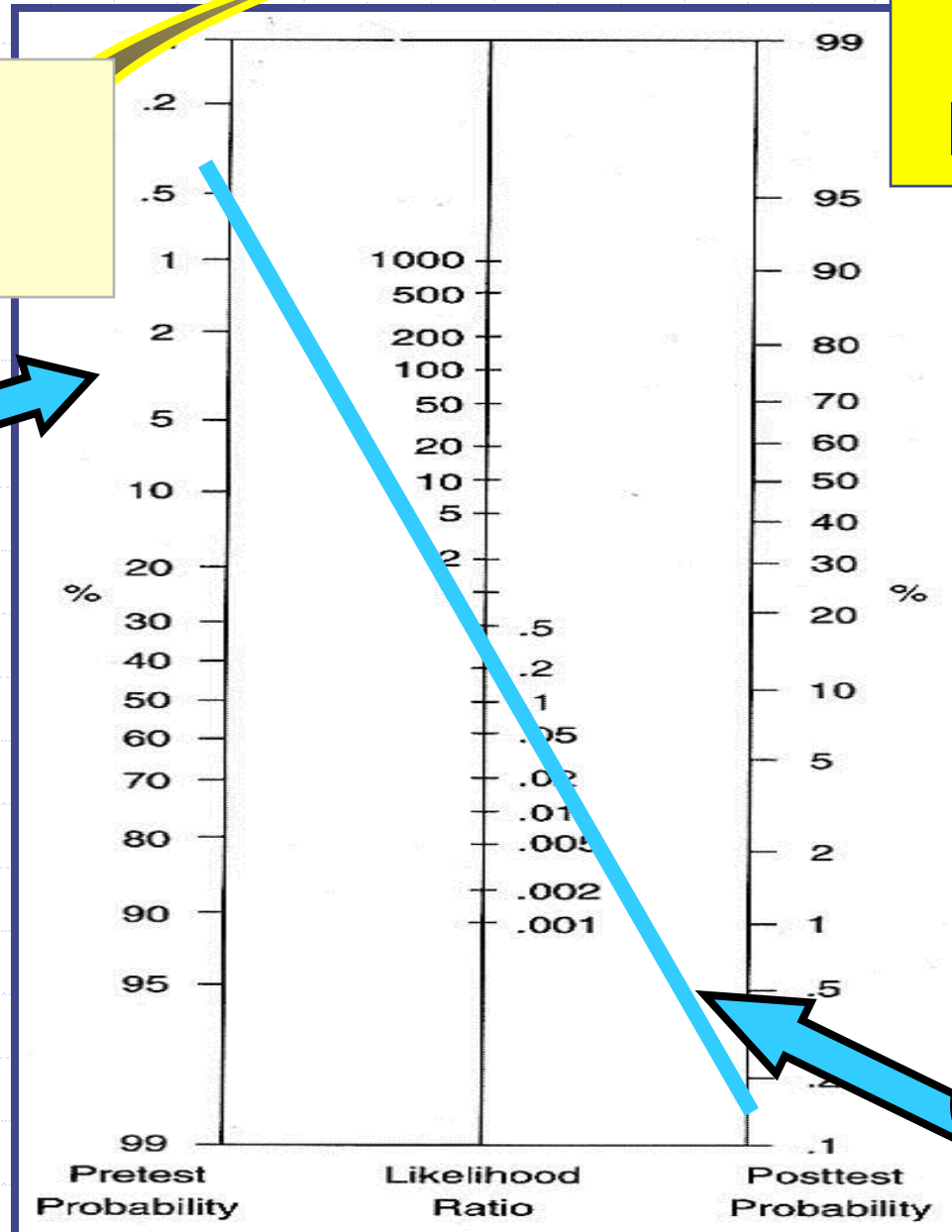
Nuchal fold abnormal
LR = 36

Likelihood Ratios

**Pre-Test
Probability**

**Post-Test
Probability**

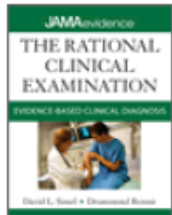
**Mrs. B
Pre-Test Prob.
0.5%**



**Post-Test
Prob. 0.1%**

Nuchal fold normal
LR = 0.26

Where do we get LRs from?



The Rational Clinical Examination: Evidence-Based Clinical Diagnosis >
Pretest Probabilities and Likelihood Ratios for Clinical Findings

Quick Reference **Online Only**

<http://jamaevidence.com>

Note: Large images and tables on this page may necessitate printing in landscape mode.

The Rational Clinical Examination > Pretest Probabilities and Likelihood Ratios for Clinical Findings >

Quick Reference

[+ Add to my saved tables](#)

Pretest Probabilities and Likelihood Ratios for Clinical Findings

	Prior Probability	Test/Finding	LR+	LR-
Chapter 1: Primer on Precision and Accuracy				
Chapter 2: Abdominal Aortic Aneurysm	Occur in 4% to 8% of older men. The prevalence in older women is less than 2%.	Physical examination for aneurysm > 4.0 cm	16 (8.6-29)	0.51 (0.38-0.67)
		Physical examination for aneurysm > 3.0 cm	12 (7.4-20)	0.72 (0.65-0.81)
Chapter 3:	Approximately 1% to 5% of the general population	Systolic-diastolic bruit	39 (10-145)	0.62 (0.49-0.73)

The Rational Clinical Examination

Copyright © American Medical Association. All rights reserved. | JAMA | The McGraw-Hill Companies, Inc.

Examples

	Prior Probability	Test/Finding	LR+	LR-
Chapter 41: Pneumonia, Infant and Child	15% to 35% prevalence of pneumonia given cough or respiratory symptoms	Grunting among children with wheezing, < 18 mo	2.8 (1.6–4.4)	0.7 (0.55–0.89)
		Retraction	2.7 (1.1–6.9)	0.97 (0.93–1.0)
		Rales	1.8–15	0.69–0.86
		Tachypnea (use WHO adjusted criteria)	1.6–8.0	0.32–0.91

	Prior Probability	Test/Finding	LR+	LR-
Chapter 51: Urinary Tract Infection, Women	48% among women with compatible symptoms	Dysuria	1.5 (1.2–2.0)	0.5 (0.3–0.7)
		Frequency	1.8 (1.1–3.0)	0.5 (0.4–1.0)
		Vaginal discharge	0.3 (0.1–0.9)	3.1 (1.0–9.3)
		Vaginal irritation	0.2 (0.1–0.9)	2.7 (0.9–8.5)
		Dipstick result	4.2	0.3

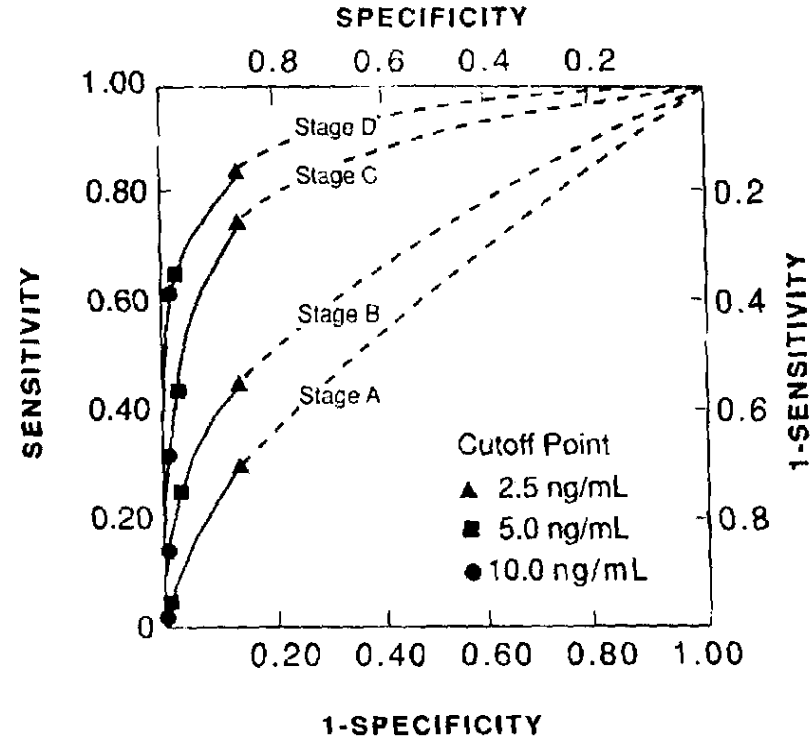
Are sens/spec and LRs inherent properties of a test?

- ◆ Most textbooks will say that sens and spec do not depend on disease prevalence
- ◆ This is partly true but oversimplified
- ◆ In reality, sens/spec and LRs vary across populations because of differences in disease spectra (case-mix) and several other factors
- ◆ This is equivalent to “effect modification” in epidemiology

Example

Sens and Spec across populations

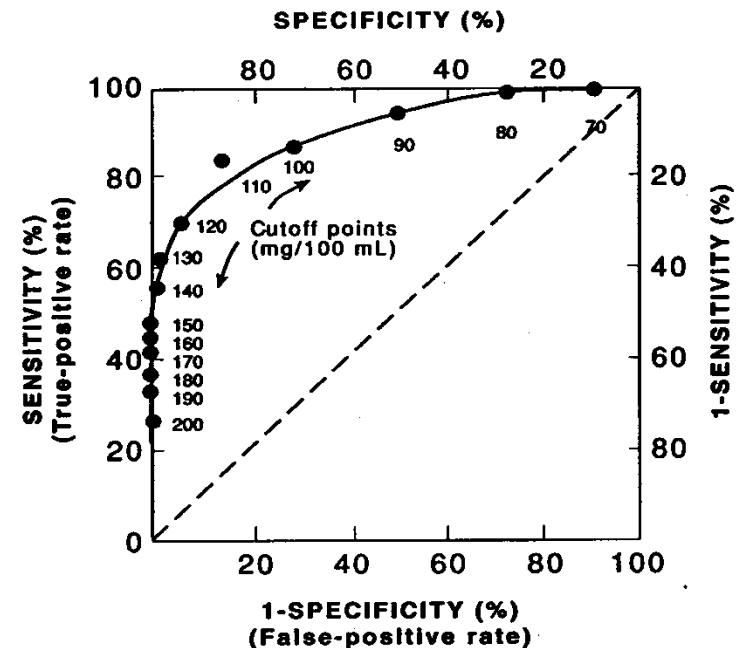
Ex:
Sensitivity+specificity
of serum CEA For
detection
of colorectal cancer,
across stages



ROC curve for CEA as a diagnostic test for colorectal cancer, according to stage of disease. The sensitivity and specificity of a test vary with the stage of disease. (Redrawn from Fletcher RH. Carcinoembryonic antigen. *Ann Intern Med* 1986;104:66-73.)

Tests with continuous results: ROC curve analysis

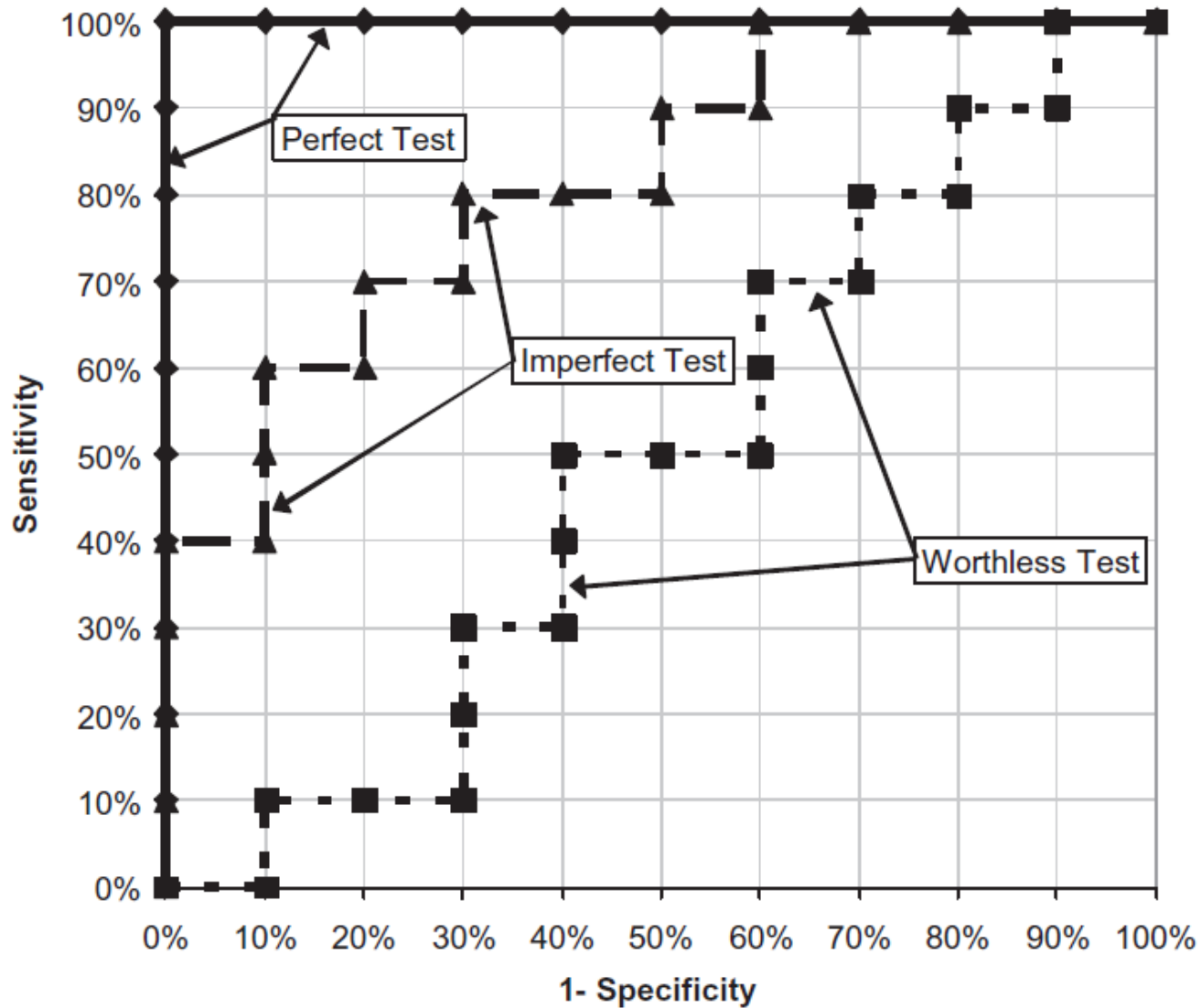
Blood sugar level (2-hour after food) in mg/100 ml	Sensitivity (%)	Specificity (100%)
70	98.6	8.8
80	97.1	25.5
90	94.3	47.6
100	88.6	69.8
110	85.7	84.1
120	71.4	92.5
130	64.3	96.9
140	57.1	99.4
150	50.0	99.6
160	47.1	99.8
170	42.9	100
180	38.6	100
190	34.3	100
200	27.1	100



A ROC curve. The accuracy of 2-hr postprandial blood sugar as a diagnostic test for diabetes mellitus.

The closer the curve follows the left-hand border and then the top-border of the ROC space, the more accurate the test. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.

Area under the curve (AUC) can range from 0.5 (random chance, or no predictive ability; refers to the 45 degree line in the ROC plot) to 1 (perfect discrimination/accuracy).



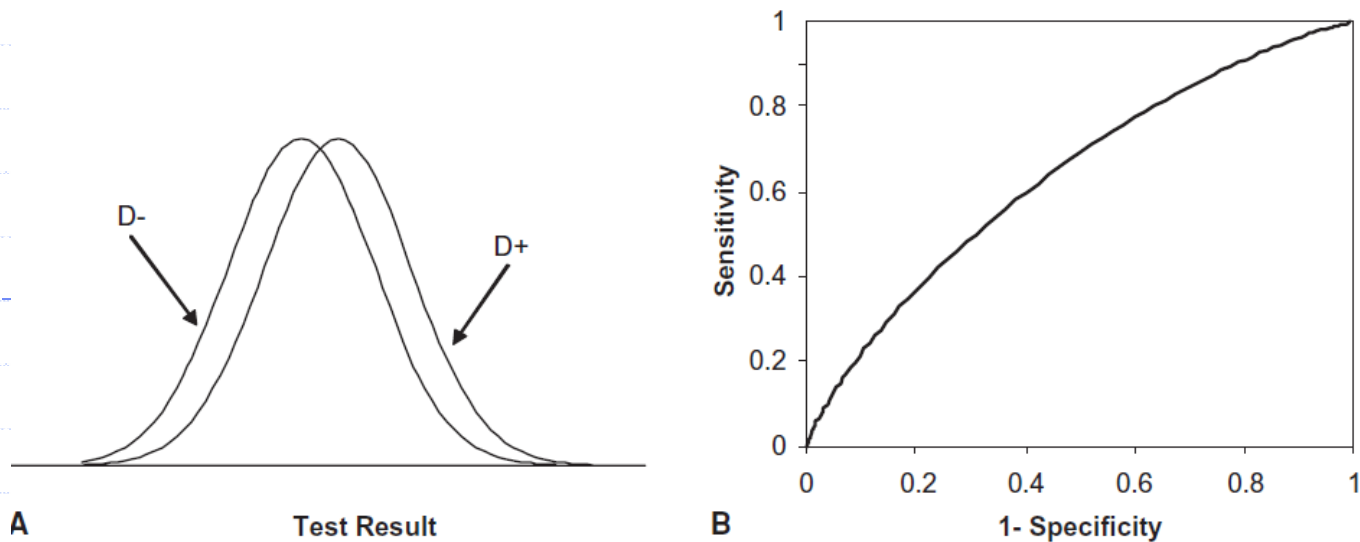


Figure 4.2 Test discriminates poorly between patients with disease (D+) and patient without disease (D-). (A) The distribution of test results in D+ patients is very similar to the distribution in D- patients. (B) This "bad" ROC curve approaches a 45-degree diagonal line.

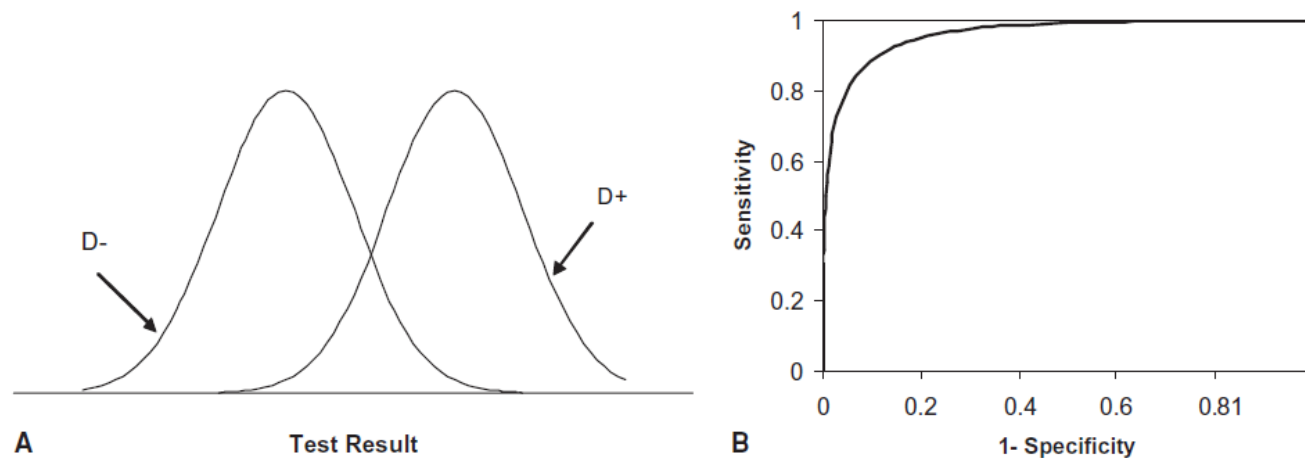


Figure 4.3 Test discriminates well between patients with the disease (D+) and patients without the disease (D-). (A) The distribution of test results in D+ patients differs substantially from the distribution in D- patients. (B) This "good" ROC curve nears the upper left corner of the grid.

Sources of bias in diagnostic studies

- ◆ Bias due to an inappropriate reference standard
- ◆ Spectrum bias
- ◆ Verification (work-up) bias
 - Partial verification bias
 - Differential verification bias
- ◆ Review bias (lack of blinding)
- ◆ Incorporation bias

Bias due to inappropriate or imperfect reference standard

- ◆ There is no such thing as a “gold” standard
- ◆ Imperfect reference standards are commonly used in diagnostic studies
 - Can lead to underestimation of test accuracy (under certain conditions)
- ◆ Examples: TB meningitis, Irritable bowel syndrome, tuberculosis in kids, migraine, depression



New gold standard: Phelps wins eighth medal

Michael Phelps won his record eighth gold medal at the Beijing Olympics as a member of the victorious U.S. 4x100-meter medley relay team, breaking a tie with Mark Spitz for most golds in a single games. [full story](#)

Spectrum bias

- ◆ Population used for evaluating the test:
 - Extreme contrast
 - ◆ Case-control design
 - Normal contrast (Indicated population)
 - ◆ Consecutively recruited patients in whom the disease is suspected
 - Extreme contrast (spectrum bias) can result in overestimation of test accuracy
 - Examples: Ultrasound for fluid in abdomen

Verification bias

◆ Verification bias in general:

- When the decision to perform the reference standard depends on the result of the index test
- When the type of reference standard used depends on the result of the index test

◆ Partial verification:

- Reference standard performed on test-positives, but not test-negatives

◆ Differential verification:

- Reference standard used for test-positives is different from that used for test-negatives

Review bias

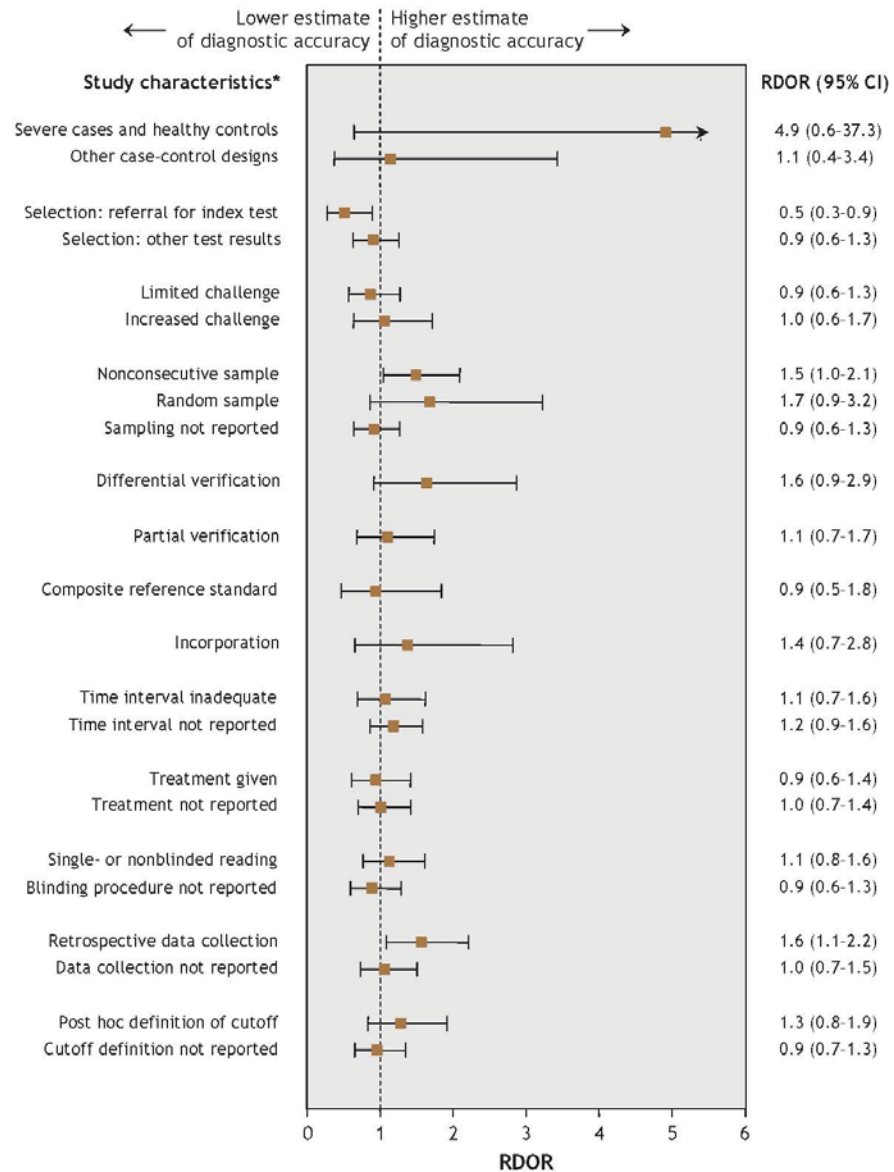
- ◆ Diagnostic studies may be:
 - Unblinded
 - Single blind (test or reference standard result is blinded)
 - Double blind (both test and ref. std results are blinded)
- ◆ Lack of blinding can lead to overestimation of test accuracy
- ◆ Examples: history and examination for hypothyroidism, touch and perception for fever

Incorporation bias

- ◆ If the test that is being evaluated is included in the reference standard
- ◆ Can lead to overestimation of test accuracy
- ◆ Examples: PCR for tuberculosis, clinical diagnosis of TB meningitis, Mantoux for TB among kids

Do design flaws affect study results?

487 diagnostic studies



*See Appendix 2 for descriptions of the study characteristics.

Fig. 2: Effects of study design characteristics on estimates of diagnostic accuracy. RDOR = relative diagnostic odds ratio (adjusted RDORs were estimated in a multivariable random-effects meta-epidemiologic regression model).

Do diagnostic trials lack methodologic rigor?

Diagnostic studies in 4 general medical journals

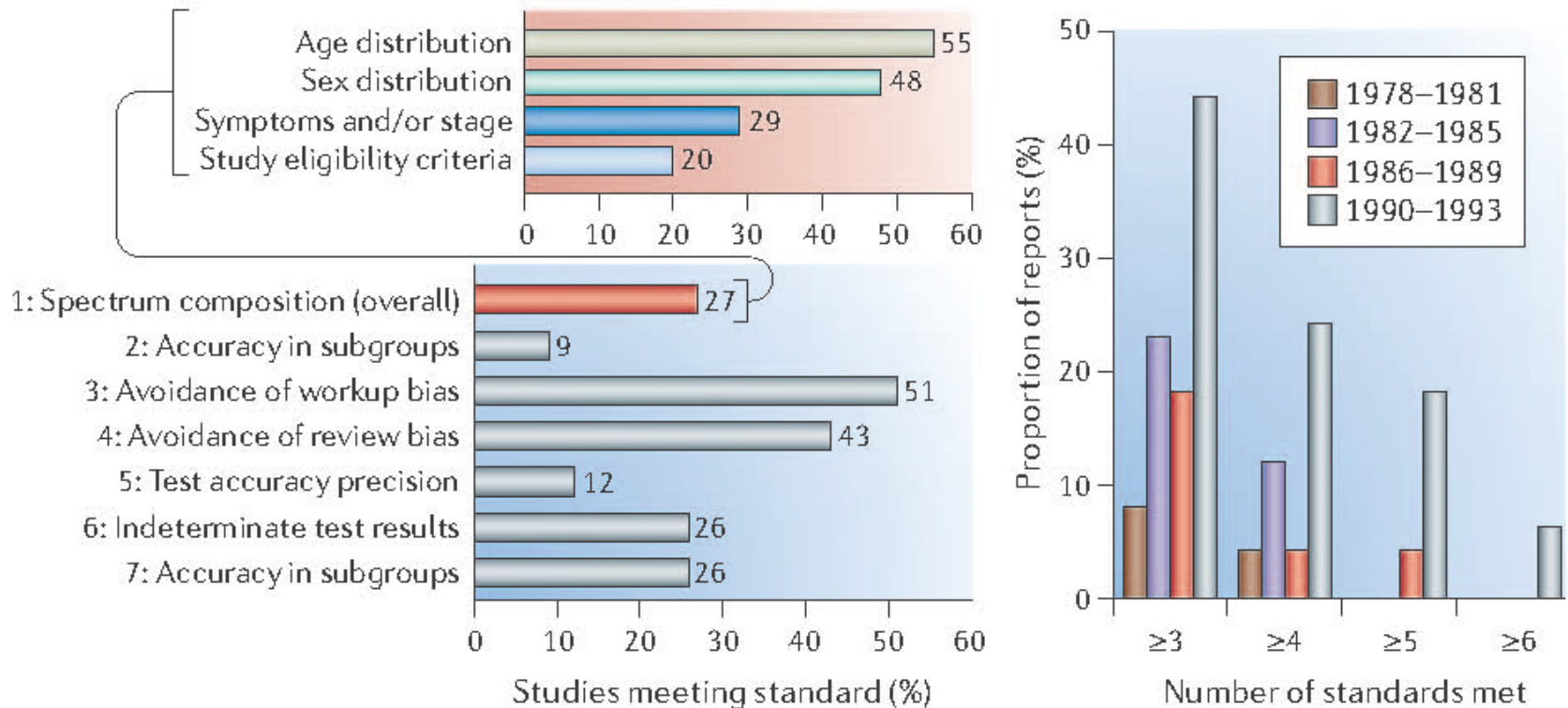


Figure 4 | **Proportion of diagnostic evaluations meeting accepted standards.** The seven standards are shown on the left. The data are taken from REF. 10.

Do diagnostic trials lack methodologic rigor?

OPEN ACCESS Freely available online

PLOS one

Quality and Reporting of Diagnostic Accuracy Studies in TB, HIV and Malaria: Evaluation Using QUADAS and STARD Standards

Patricia Scolari Fontela¹, Nitika Pant Pai², Ian Schiller², Nandini Dendukuri², Andrew Ramsay³, Madhukar Pai^{1,4*}

1 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada, **2** Department of Medicine, Division of Clinical Epidemiology, McGill University, Montreal, Canada, **3** Special Programme for Research and Training in Tropical Diseases, World Health Organization, Geneva, Switzerland, **4** Respiratory Epidemiology and Clinical Research Unit, Montreal Chest Institute, Montreal, Canada

Abstract

Background: Poor methodological quality and reporting are known concerns with diagnostic accuracy studies. In 2003, the QUADAS tool and the STARD standards were published for evaluating the quality and improving the reporting of diagnostic studies, respectively. However, it is unclear whether these tools have been applied to diagnostic studies of infectious diseases. We performed a systematic review on the methodological and reporting quality of diagnostic studies in TB, malaria and HIV.

Methods: We identified diagnostic accuracy studies of commercial tests for TB, malaria and HIV through a systematic search of the literature using PubMed and EMBASE (2004–2006). Original studies that reported sensitivity and specificity data were included. Two reviewers independently extracted data on study characteristics and diagnostic accuracy, and used QUADAS and STARD to evaluate the quality of methods and reporting, respectively.

Findings: Ninety (38%) of 238 articles met inclusion criteria. All studies had design deficiencies. Study quality indicators that were met in less than 25% of the studies included adequate description of withdrawals (6%) and reference test execution (10%), absence of index test review bias (19%) and reference test review bias (24%), and report of uninterpretable results (22%). In terms of quality of reporting, 9 STARD indicators were reported in less than 25% of the studies: methods for calculation and estimates of reproducibility (0%), adverse effects of the diagnostic tests (1%), estimates of diagnostic accuracy between subgroups (10%), distribution of severity of disease/other diagnoses (11%), number of eligible patients who did not participate in the study (14%), blinding of the test readers (16%), and description of the team executing the test and management of indeterminate/outlier results (both 17%). The use of STARD was not explicitly mentioned in any study. Only 22% of 46 journals that published the studies included in this review required authors to use STARD.

Conclusion: Recently published diagnostic accuracy studies on commercial tests for TB, malaria and HIV have moderate to low quality and are poorly reported. The more frequent use of tools such as QUADAS and STARD may be necessary to improve the methodological and reporting quality of future diagnostic accuracy studies in infectious diseases.

What can be done to improve quality and reporting of diagnostic studies?

- ◆ Report better using standardized reporting formats (e.g. STARD)
- ◆ Improve study design using guidelines specific for diagnostic trials
 - E.g. QUADAS, DEEP

STARD reporting standards

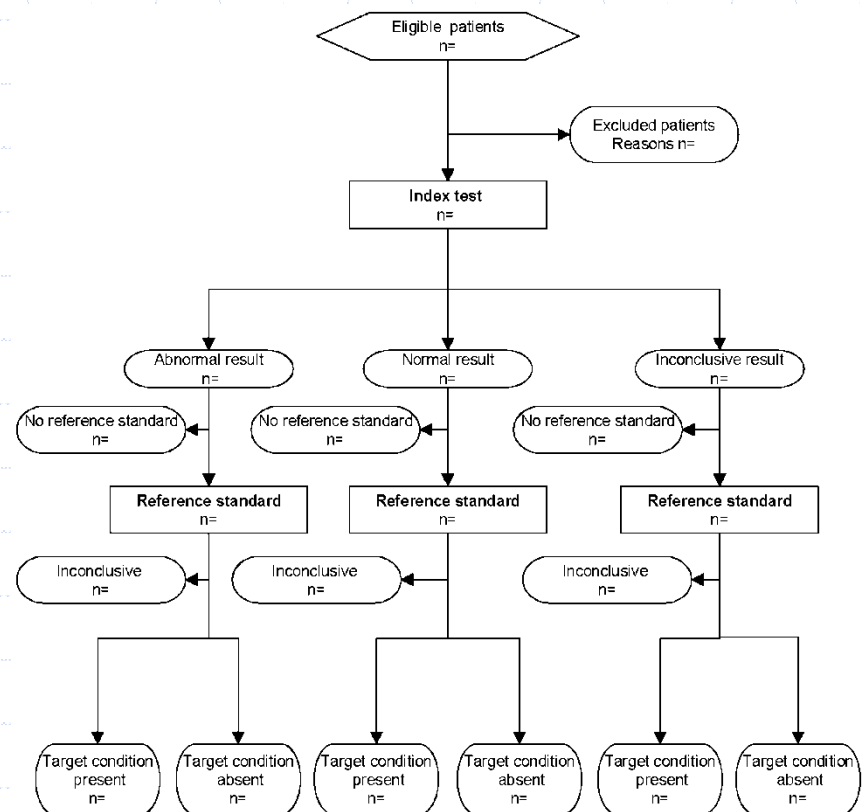
STARD checklist

APPENDIX 1 | STANDARDS FOR REPORTING OF DIAGNOSTIC ACCURACY (STARD) CHECKLIST

Section and topic	Item #	On page #
Title/abstract/keywords	1	
Introduction	2	
Methods	Describe:	
Participants	3	
	4	
	5	
	6	
Test methods	7	
	8	
	9	
	10	
	11	
Statistical methods	12	
	13	
Results	Report:	
Participants	14	
	15	
	16	
Test results	17	
	18	
	19*	
	20	
Estimates	21	
	22	
	23	
	24	
	25	

* This entry has been modified from the original.

STARD flow diagram



QUADAS tool for quality assessment of diagnostic studies

**BMC Medical Research
Methodology**



Research article

Open Access

The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews

Penny Whiting*¹, Anne WS Rutjes², Johannes B Reitsma²,
Patrick MM Bossuyt² and Jos Kleijnen¹

Although designed for quality assessment in systematic reviews, it can be used to improve study design

QUADAS tool for quality assessment of diagnostic studies

Table I: QUADAS

Item #	Description
1.	Was the spectrum of patients representative of the patients who will receive the test in practice?
2.	Were selection criteria clearly described?
3.	Is the reference standard likely to correctly classify the target condition?
4.	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? (disease progression bias)
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis? (partial verification bias)
6.	Did patients receive the same reference standard regardless of the index test result? (differential verification bias)
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)? (incorporation bias)
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?
10.	Were the index test results interpreted without knowledge of the results of the reference standard? (test review bias)
11.	Were the reference standard results interpreted without knowledge of the results of the index test? (diagnostic review bias)
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? (clinical review bias)
13.	Were uninterpretable/ intermediate test results reported?
14.	Were withdrawals from the study explained?

TDR/WHO Diagnostics Evaluation Expert Panel (DEEP) guidelines



EVALUATING DIAGNOSTICS

Evaluation of diagnostic tests for infectious diseases: general principles

The TDR Diagnostics Evaluation Expert Panel

Are sensitivity and specificity the most meaningful measures?

Table 1. Hierarchy of Diagnostic Evaluation and the Number of Studies Available for Different Levels of Diagnostic Test in a Technology Assessment of Magnetic Resonance Spectroscopy for Brain Tumors*

Level	Description	Examples of Study Purpose or Measures	Studies Available, <i>n</i>	Patients, <i>n</i>
1	Technical feasibility and optimization	Ability to produce consistent spectra	85	2434
2	Diagnostic accuracy	Sensitivity and specificity	8	461
3	Diagnostic thinking impact	Percentage of times clinicians' subjective assessment of diagnostic probabilities changed after the test	2	32
4	Therapeutic choice impact	Percentage of times therapy planned before MRS changed after the test	2	105
5	Patient outcome impact	Percentage of patients who improved with MRS diagnosis compared with those without MRS (e.g., survival, quality of life)	0	0
6	Societal impact	Cost-effectiveness analysis (e.g., use to detect tumor in asymptomatic population)	0	0

* MRS = magnetic resonance spectroscopy.

Redundancy of Single Diagnostic Test Evaluation

Karel G.M. Moons,^{1,2,3} Gerri-Anne van Es,⁴ Bowine C. Michel,⁵ Harry R. Büller,⁶
J. Dik F. Habbema,³ and Diederick E. Grobbee¹

Moons et al. Epidemiology 1999

Diagnostic research

Diagnostic studies as multivariable,
prediction research

K G M Moons, D E Grobbee

Patient outcomes in diagnostic research

Moons et al. JECH 2002

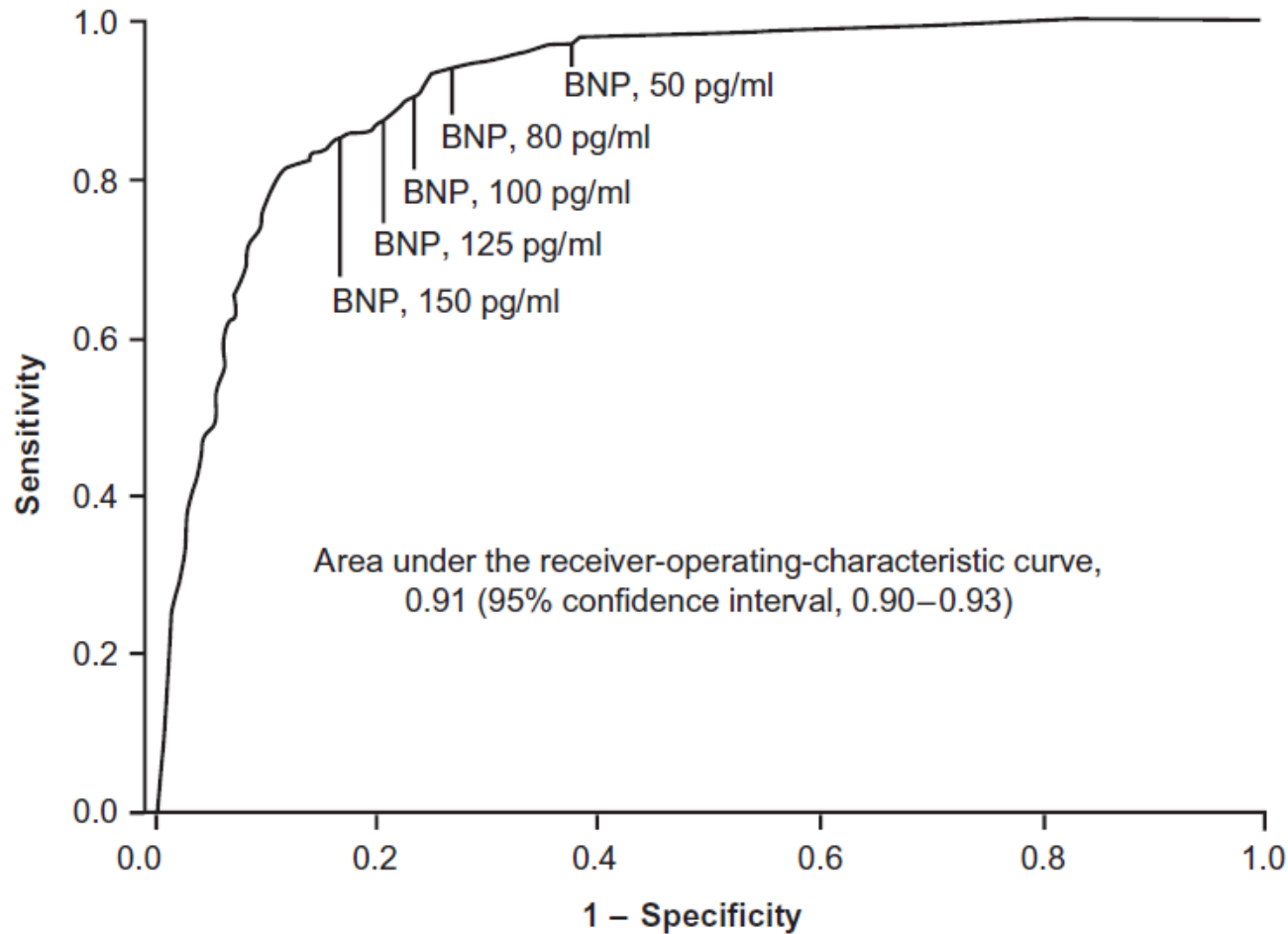
Opinion

Test Research versus Diagnostic Research

Moons et al. Clin Chem 2004

Accuracy vs Impact:

Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure



B-Type Natriuretic Peptide Testing, Clinical Outcomes, and Health Services Use in Emergency Department Patients With Dyspnea

A Randomized Trial

Hans-Gerhard Schneider, MBBS, MD; Louisa Lam, MPH; Amaali Lokuge, MBBS; Henry Krum, MBBS, PhD; Matthew T. Naughton, MBBS; Pieter De Villiers Smit, MBBS; Adam Bystrzycki, MBBS; David Eccleston, MBBS, PhD; Jacob Federman, MBBS; Genevieve Flannery, MBBS; and Peter Cameron, MBBS, MD

Background: B-type natriuretic peptide (BNP) is used to diagnose heart failure, but the effects of using the test on all dyspneic patients is uncertain.

Objective: To assess whether BNP testing alters clinical outcomes and health services use of acutely dyspneic patients.

Design: Randomized, single-blind study. Patients were assigned to a treatment group through randomized numbers in a sealed envelope. Patients were blinded to the intervention, but clinicians and those who assessed trial outcomes were not.

Setting: 2 Australian teaching hospital emergency departments.

Patients: 612 consecutive patients who presented with acute severe dyspnea from August 2005 to March 2007.

Intervention: BNP testing ($n = 306$) or no testing ($n = 306$).

Measurements: Admission rates, length of stay, and emergency department medications (primary outcomes); mortality and readmission rates (secondary outcomes).

Results: There were no between-group differences in hospital admission rates (85.6% [BNP group] vs. 86.6% [control group]; dif-

ference, -1.0 percentage point [95% CI, -6.5 to 4.5 percentage points]; $P = 0.73$), length of admission (median, 4.4 days [interquartile range, 2 to 9 days] vs. 5.0 days [interquartile range, 2 to 9 days]; $P = 0.94$), or management of patients in the emergency department. Test discrimination was good (area under the receiver-operating characteristic curve, 0.87 [CI, 0.83 to 0.91]). Adverse events were not measured.

Limitation: Most patients were very short of breath and required hospitalization; the findings might not apply for evaluating patients with milder degrees of breathlessness.

Conclusion: Measurement of BNP in all emergency department patients with severe shortness of breath had no apparent effects on clinical outcomes or use of health services. The findings do not support routine use of BNP testing in all severely dyspneic patients in the emergency department.

Primary Funding Source: Janssen-Cilag.

Ann Intern Med. 2009;150:365-371.

For author affiliations, see end of text.

ClinicalTrials.gov registration number: NCT00163709.

www.annals.org

Evaluation of Diagnostic Accuracy, Feasibility and Client Preference for Rapid Oral Fluid-Based Diagnosis of HIV Infection in Rural India

Nitika Pant Pai^{1*}, Rajnish Joshi², Sandeep Dogra³, Bharati Taksande², S. P. Kalantri², Madhukar Pai⁴, Pratibha Narang², Jacqueline P. Tulsy⁵, Arthur L. Reingold⁶

1 Immunodeficiency Service, Montreal Chest Institute, McGill University Health Center, Montreal, Canada, **2** Mahatma Gandhi Institute of Medical Sciences, Sevagram, Maharashtra, India, **3** Acharya Shri Chander College of Medical Sciences, Jammu, India, **4** Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada, **5** Department of Internal Medicine, University of California at San Francisco, San Francisco, California, United States of America, **6** Division of Epidemiology, University of California at Berkeley, Berkeley, California, United States of America

Background. Oral fluid-based rapid tests are promising for improving HIV diagnosis and screening. However, recent reports from the United States of false-positive results with the oral OraQuick® ADVANCE HIV1/2 test have raised concerns about their performance in routine practice. We report a field evaluation of the diagnostic accuracy, client preference, and feasibility for the oral fluid-based OraQuick® Rapid HIV1/2 test in a rural hospital in India. **Methodology/Principal Findings.** A cross-sectional, hospital-based study was conducted in 450 consenting participants with suspected HIV infection in rural India. The objectives were to evaluate performance, client preference and feasibility of the OraQuick® Rapid HIV-1/2 tests. Two Oraquick® Rapid HIV1/2 tests (oral fluid and finger stick) were administered in parallel with confirmatory ELISA/Western Blot (reference standard). Pre- and post-test counseling and face to face interviews were conducted to determine client preference. Of the 450 participants, 146 were deemed to be HIV sero-positive using the reference standard (seropositivity rate of 32% (95% confidence interval [CI] 28%, 37%)). The OraQuick test on oral fluid specimens had better performance with a sensitivity of 100% (95% CI 98, 100) and a specificity of 100% (95% CI 99, 100), as compared to the OraQuick test on finger stick specimens with a sensitivity of 100% (95% CI 98, 100), and a specificity of 99.7% (95% CI 98.4, 99.9). The OraQuick oral fluid-based test was preferred by 87% of the participants for first time testing and 60% of the participants for repeat testing. **Conclusion/Significance.** In a rural Indian hospital setting, the OraQuick® Rapid- HIV1/2 test was found to be highly accurate. The oral fluid-based test performed marginally better than the finger stick test. The oral OraQuick test was highly preferred by participants. In the context of global efforts to scale-up HIV testing, our data suggest that oral fluid-based rapid HIV testing may work well in rural, resource-limited settings.

Citation: Pant Pai N, Joshi R, Dogra S, Taksande B, Kalantri SP, et al (2007) Evaluation of Diagnostic Accuracy, Feasibility and Client Preference for Rapid Oral Fluid-Based Diagnosis of HIV Infection in Rural India. PLoS ONE 2(4): e367. doi:10.1371/journal.pone.0000367

Impact of Round-the-Clock, Rapid Oral Fluid HIV Testing of Women in Labor in Rural India

Nitika Pant Pai^{1*}, Ritu Barick², Jacqueline P. Tulsy³, Poonam V. Shivkumar², Deborah Cohan³, Shriprakash Kalantri², Madhukar Pai⁴, Marina B. Klein¹, Shakuntala Chhabra²

1 Division of Infectious Diseases and Immunodeficiency Service, Montreal Chest Institute, McGill University Health Center, Montreal, Canada, **2** Mahatma Gandhi Institute of Medical Sciences, Sevagram, Wardha, Maharashtra, India, **3** Positive Health Program, Division of Internal Medicine, University of California San Francisco, San Francisco, California, United States of America, **4** Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada

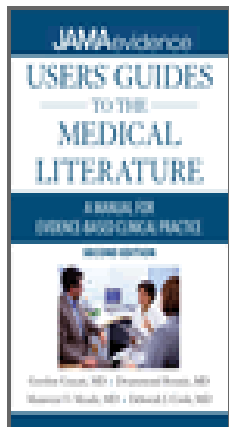
Methods and Findings

After they provided written informed consent, women admitted to the labor ward of a rural teaching hospital in India were offered two rapid tests on oral fluid and finger-stick specimens (OraQuick Rapid HIV-1/HIV-2 tests, OraSure Technologies). Simultaneously, venous blood was drawn for conventional HIV ELISA testing. Western blot tests were performed for confirmatory testing if women were positive by both rapid tests and dual ELISA, or where test results were discordant. Round-the-clock (24 h, 7 d/wk) abbreviated prepartum and extended postpartum counseling sessions were offered as part of the testing strategy. HIV-positive women were administered PMTCT interventions. Of 1,252 eligible women (age range 18 y to 38 y) approached for consent over a 9 mo period in 2006, 1,222 (98%) accepted HIV testing in the labor ward. Of these, 1,003 (82%) women presented with either no reports or incomplete reports of prior HIV testing results at the time of admission to the labor ward. Of 1,222 women, 15 were diagnosed as HIV-positive (on the basis of two rapid tests, dual ELISA and Western blot), yielding a seroprevalence of 1.23% (95% confidence interval [CI] 0.61%–1.8%). Of the 15 HIV test-positive women, four (27%) had presented with reported HIV status, and 11 (73%) new cases of HIV infection were detected due to rapid testing in the labor room. Thus, 11 HIV-positive women received PMTCT interventions on account of round-the-clock rapid HIV testing and counseling in the labor room. While both OraQuick tests (oral and finger-stick) were 100% specific, one false-negative result was documented (with both oral fluid and finger-stick specimens). Of the 15 HIV-infected women who delivered, 13 infants were HIV seronegative at birth and at 1 and 4 mo after delivery; two HIV-positive infants died within a month of delivery.

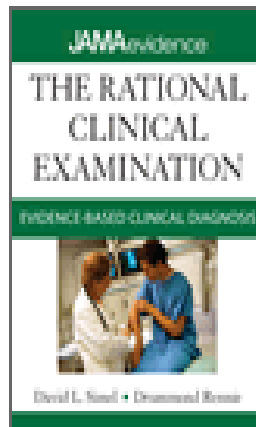
Conclusions

In a busy rural labor ward setting in India, we demonstrated that it is feasible to introduce a program of round-the-clock rapid HIV testing, including prepartum and extended postpartum counseling sessions. Our data suggest that the availability of round-the-clock rapid HIV testing resulted in successful documentation of HIV serostatus in a large proportion (82%) of rural women who were unaware of their HIV status when admitted to the labor room. In addition, 11 (73%) of a total of 15 HIV-positive women received PMTCT interventions because of round-the-clock rapid testing in the labor ward. These findings are relevant for PMTCT programs in developing countries.

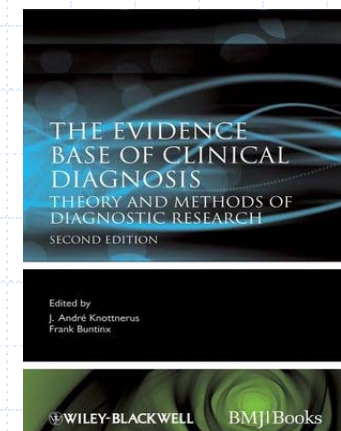
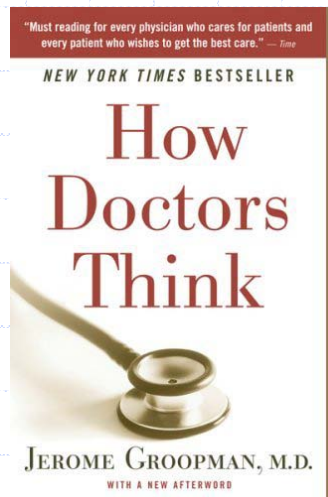
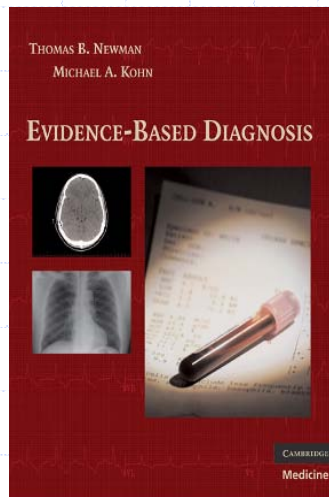
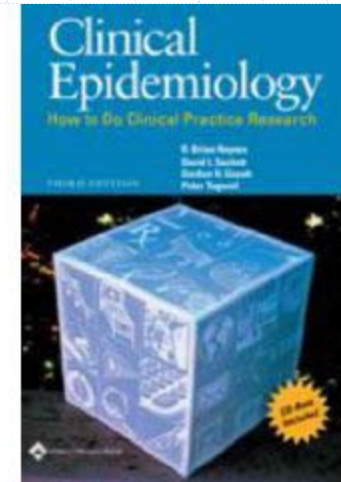
Relevant books



Users' Guides to the Medical Literature
A Manual for Evidence-Based Clinical Practice, 2nd Edition



The Rational Clinical Examination
Evidence-Based Clinical Diagnosis
Includes online-only content



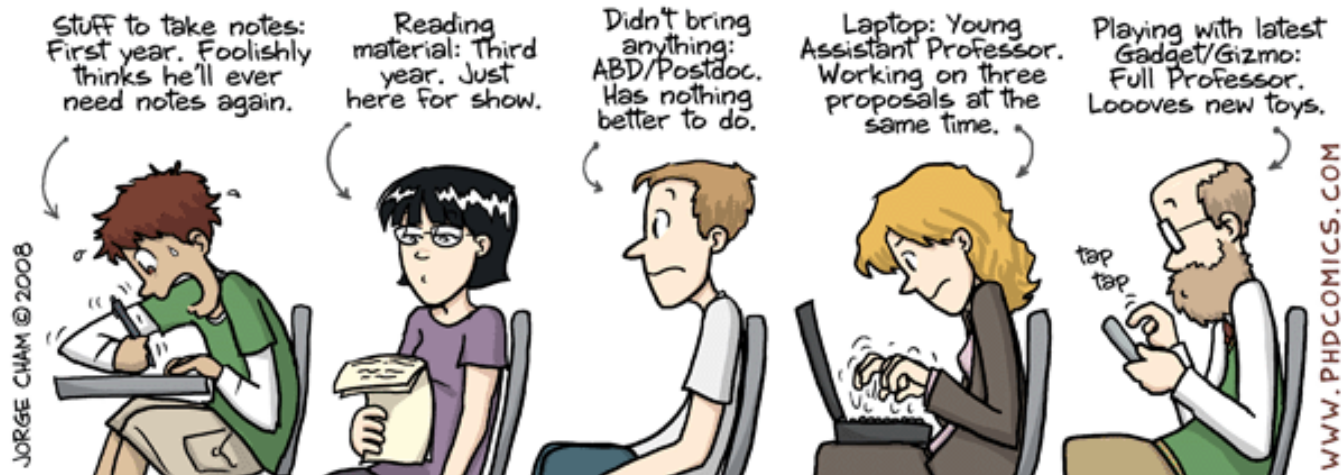
Readings

- ◆ Rothman text:
 - Chapter 11: Epidemiology in clinical settings
- ◆ Gordis text:
 - Chapter 5



"Your urine sample contained traces of malaise and cynicism."

WHAT YOU BROUGHT TO SEMINAR AND WHAT IT SAYS ABOUT YOU:



***footnote: Thanks to Zoe from EPFL for this comic idea!*

THE AUTHOR LIST: GIVING CREDIT WHERE CREDIT IS DUE

The first author
Senior grad student on the project. Made the figures.

The third author
First year student who actually did the experiments, performed the analysis and wrote the whole paper. Thinks being third author is "fair".

The second-to-last author
Ambitious assistant professor or post-doc who instigated the paper.

Michaels, C., Lee, E. F., Sap, P. S., Nichols, S. T., Oliveira, L., Smith, B. S.

The second author
Grad student in the lab that has nothing to do with this project, but was included because he/she hung around the group meetings (usually for the food).

The middle authors
Author names nobody really reads. Reserved for undergrads and technical staff.

The last author
The head honcho. Hasn't even read the paper but, hey, he got the funding, and his famous name will get the paper accepted.

JORGE CHAM © 2005

www.phdcomics.com