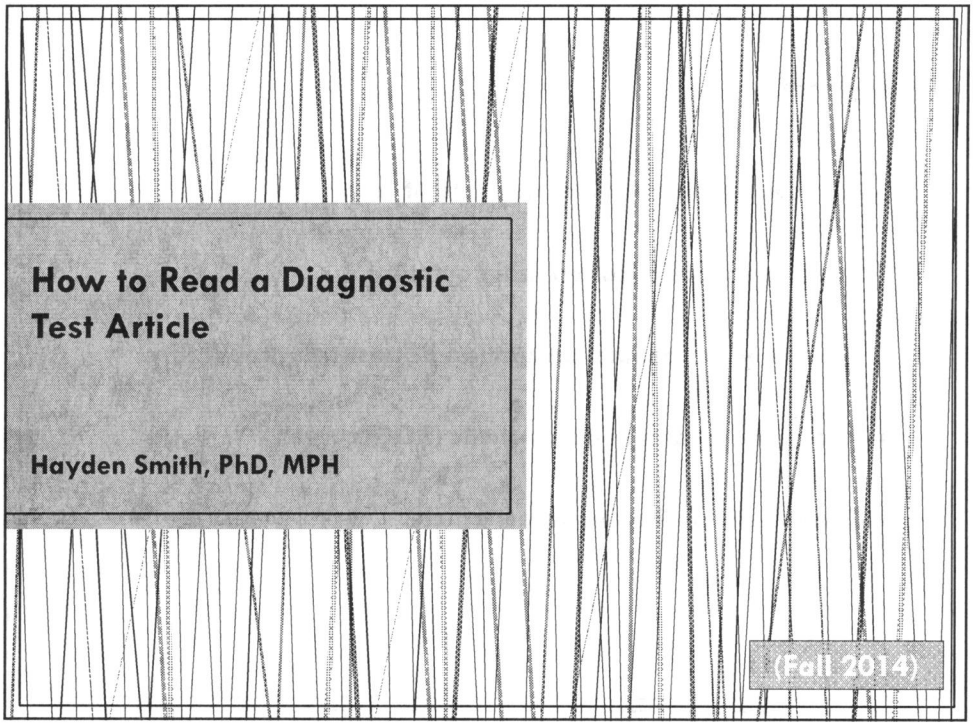


Information and clinical examples provided in this presentation are solely for educational purposes, and should not be substituted for clinical guidelines or up-to-date medical information.

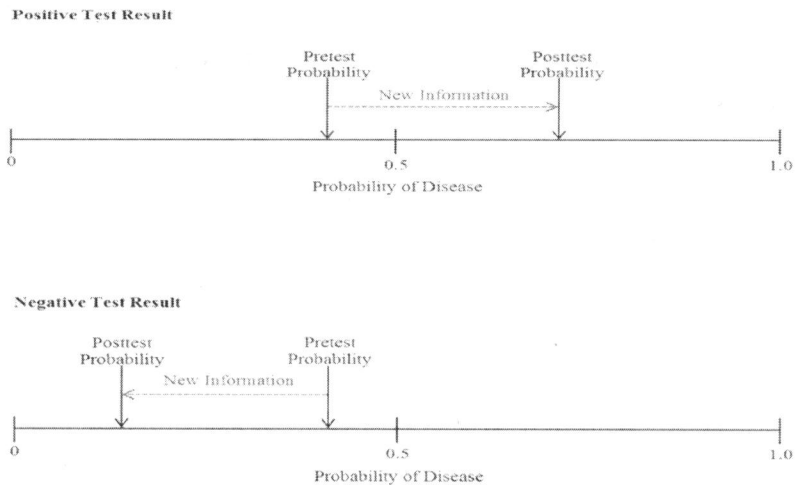
Outline

1. Clinical decision making and role of test
2. Two x two classification table notation
3. Likelihood ratios and calculation of post-test probability
4. Receiver operating characteristic (ROC) curve
5. Evaluation of diagnostic test article

Clinical Decision Making Steps

1. Clinical assessment of symptoms, signs, risk factors of disease
2. Estimation of disease probability...pre-test probability
3. If reasonable probability - order test
4. Test may increase or decrease post-test probability

Accounting for New Information



Case Vignette

An otherwise healthy 51 year-old woman presents to her physician with pleuritic right posterior chest pain, without dyspnea or hemoptysis.

Her temperature is 38.2° and her pulse is 102 bpm. Physical exam discloses a pleural friction rub over the posterior right hemithorax but is otherwise unremarkable. Chest radiograph is normal.

She is treated with an anti-inflammatory agent for presumed viral pleurisy. Three days later, she returns reporting dyspnea and slight hemoptysis. How should she be evaluated?

Question?

What is the probability of pulmonary embolism (PE) in patient:

low (0-20%);

intermediate (20-80%); or

high (80% or higher)

Clinical Assessment of Symptoms, Signs, Risk Factors for PE:

Simplified Wells Scoring System	
Findings	**Score***
Clinical signs and symptoms of deep venous thrombosis (minimum of leg swelling and pain with palpation of the deep veins of the leg)	3.0
No alternate diagnosis likely or more likely than PE	3.0
Heart rate > 100 beats/mins	1.5
Immobilization or surgery in last 4 weeks	1.5
Previous history of deep venous thrombosis or PE	1.5
Hemoptysis	1.0
Cancer actively treated within last 6 months	1.0
*Scores calculated by summing weights of predictor variables; Totaled scores are as follows: Low < 2; moderate 2-6; and high > 6.	

Designing a Diagnostic Test Study

1. Enrollment of patients with a clinically suspected diagnosis – inclusion and exclusion criteria

2. Adoption of gold standard to verify disease status – determines actual probability of disease in study population

3. Actual probability = prevalence = pretest probability of all participants

Designing a Diagnostic Test Study

4. Study test compared to gold in determining accuracy of study test.

5. Accuracy is total number of true positives and true negatives for test, divided by total number of tests.

6. A 100% accurate test would contain no false positives or false negatives:

$$\frac{TP + TN}{TP + (0)FP + TN + (0)FN} \quad \frac{TP + TN}{TP + TN} = 100\%$$

PULMONARY EMBOLISM

Table Accuracy of Pretest Probability Assessment for Pulmonary Embolism Using Clinical Gestalt

Source	No. of Patients	Prevalence of Pulmonary Embolism, %	Category	Probability Estimate, %	No. of Patients	Actual Probability, %
PIOPED,[9] 1990	887	28	Low	0-19	228	9
			Moderate	20-79	569	30
			High	80-100	90	68

PIOPED Study (1990)[1]

Purpose to evaluate usefulness of V/Q scan for PE

Actual probability of PE = 28%

Scans read as:

- **High probability V/Q scan**
- **Intermediate probability V/Q scan**
- **Low probability V/Q scan**
- **Normal/near normal V/Q scan**

Outline

1. Clinical decision making and role of test
2. Two x two table notation – 4 test properties
3. Likelihood ratios and calculation of post test probabilities
4. Receiver operating characteristic (ROC) curve
5. Evaluating an article on a diagnostic test

Two x Two (2 x 2) Table Notation

		Defined by "Gold Standard"	
		Disease +	Disease -
Study Test	+	True Positive	False Positive
	–	False Negative	True Negative

2 X 2 Table Notation: Sensitivity

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

Sensitivity = True positive rate or proportion of those with disease who test positive

$$= \frac{TP}{TP + FN}$$

Table —Comparison of the Results of Diagnostic Test (Ventilation-Perfusion Scan) With the Result of Reference Standard (Pulmonary Angiogram) Assuming Only High-Probability Scans Are Positive (Truly Abnormal)*

	Angiogram	
Scan Category	Pulmonary Embolus Present	Pulmonary Embolus Absent
High probability	102	14
Others	149	616
Total	251	630

2 X 2 Table Notation: Specificity

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

Specificity = True negative rate or proportion of those without disease who test negative

$$= \frac{TN}{TN + FP}$$

A very specific test has a very low false positive rate

Table –Comparison of the Results of Diagnostic Test (Ventilation-Perfusion Scan) With the Result of Reference Standard (Pulmonary Angiogram) Assuming Only High-Probability Scans Are Positive (Truly Abnormal)*

	Angiogram	
Scan Category	Pulmonary Embolus Present	Pulmonary Embolus Absent
High probability	102	14
Others	149	616
Total	251	630

2 X 2 Notation: Positive Predictive Value

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

$$PPV = \frac{TP}{TP + FP}$$

- PPV = probability patient has disease if test is positive
- If there are 0 false positives, a test has a positive predictive value of 100%
- Increased specificity (lower false positive rate) increases PPV

Table —Comparison of the Results of Diagnostic Test (Ventilation-Perfusion Scan) With the Result of Reference Standard (Pulmonary Angiogram) Assuming Only High-Probability Scans Are Positive (Truly Abnormal)*

	Angiogram	
Scan Category	Pulmonary Embolus Present	Pulmonary Embolus Absent
High probability	102	14
Others	149	616
Total	251	630

What Can We Conclude About High Probability V/Q Scan?

- Not very sensitive for PE
- If positive has high PPV (because specificity is high)

→ What can we conclude about a normal V/Q Scan?

Two x Two Notation: Negative Predictive Value

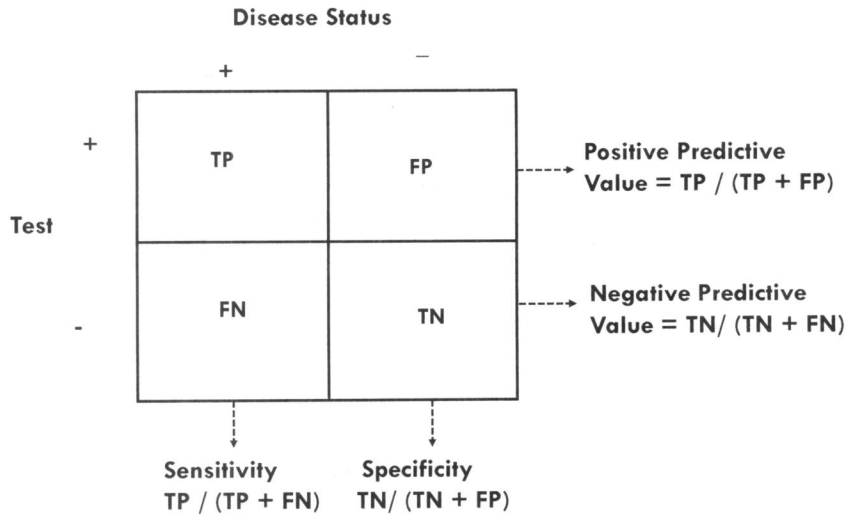
Study Test	Disease +	Disease -
+	True Positive	False Positive
–	False Negative	True Negative

→ $NPV = \frac{TN}{TN + FN}$

- NPV = probability patient does not have disease if test is negative
- If there are 0 false negatives, a test has a negative predictive value of 100%

Table -Comparison of the Results of Diagnostic Test (Ventilation-Perfusion Scan) With the Result of Reference Standard (Pulmonary Angiogram) Assuming Only Normal/Near-Normal Scans Are Negative (Truly Normal)*

	Angiogram	
Scan Category	Pulmonary Embolus Present	Pulmonary Embolus Absent
High, intermediate, and low probability	246	504
Near normal/normal	5	126
Total	251	630

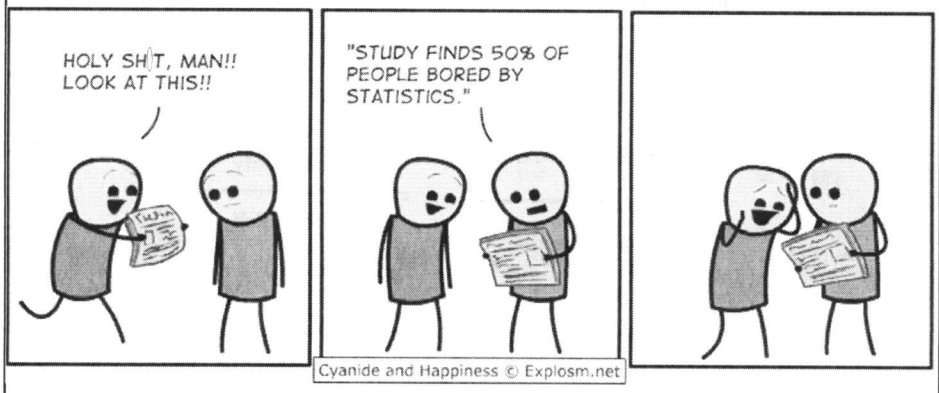


Examples

Disorder	Test	(Low False Positive) High PPV High Specificity	(Low False Negative) High NPV High Sensitivity
SLE	ds DNA	( )	( )
SLE	ANA	( )	( )

Questions

- **Elevator problem…**



Outline

1. Clinical decision making and the role of a test
2. Two x two table notation – 4 test properties
3. Likelihood ratios and calculation of post test probabilities
4. Receiver operating characteristic curve (ROC)
5. Evaluating an article on a diagnostic test

Likelihood Ratio of a Positive Test

What are the odds that a positive test would be found in a person with the condition compared to a person without the condition?

$$LR(+) = \frac{\text{True positive rate}}{\text{False positive rate}} = \frac{\text{Sensitivity}}{\text{1 - Specificity}}$$

Table —Test Properties of Ventilation-Perfusion (V/Q) Scanning

	Pulmonary Embolism				
	Present		Absent		
V/Q Scan Result	No.	Proportion	No.	Proportion	Likelihood Ratio
High probability	102	102/251 = 0.406	14	14/630 = 0.022	18.3
Intermediate probability	105	105/251 = 0.418	217	217/630 = 0.344	1.2
Low probability	39	39/251 = 0.155	273	273/630 = 0.433	0.36
Normal/near normal	5	5/251 = 0.020	126	126/630 = 0.200	0.10
Total	251	...	630	...	...

Interpreting a Likelihood Ratio:

LR	Interpretation
>10	Strong evidence to rule in disease
5-10	Moderate evidence to rule in disease
2-5	Weak evidence to rule in disease
0.5-2	No significant change in likelihood of disease
0.2-0.5	Weak evidence to rule out disease
0.1-0.2	Moderate evidence to rule out disease
<0.1	Strong evidence to rule out disease

18.3 (95% Confidence Interval: 10.7, 31.4)

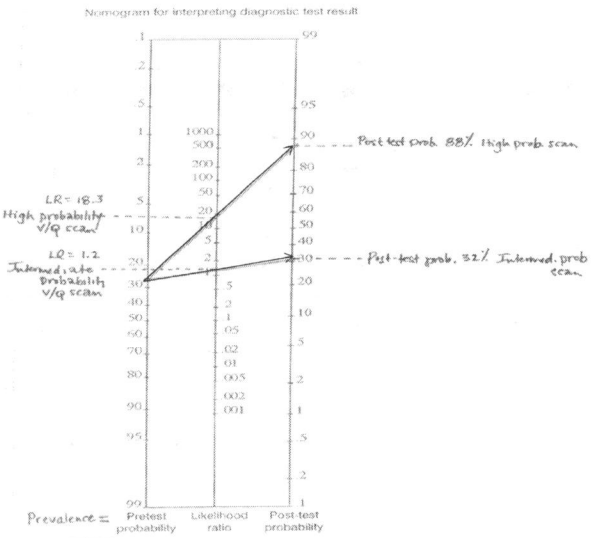


Figure [ ] A likelihood ratio nomogram. Adapted from Fagan T J 1975 Nomogram for Bayes's Theorem (c). New England Journal of Medicine 293: 257

Post-Test Probability = PPV

What is the probability of the condition given a positive test?

Can be answered in two equivalent ways:

1. **Post-test probability using LR and pre-test probability or using Bayes' Theorem**
2. **Positive predictive value (PPV) using 2 x 2 table notation**

Likelihood Ratio of a Negative Test

What are the odds that a negative test would be found in a person with the condition compared to a person without the condition?

$$LR(-) = \frac{\text{False negative rate}}{\text{True negative rate}} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

Table —Test Properties of Ventilation-Perfusion (V/Q) Scanning

	Pulmonary Embolism				
	Present		Absent		
V/Q Scan Result	No.	Proportion	No.	Proportion	Likelihood Ratio
High probability	102	102/251 = 0.406	14	14/630 = 0.022	18.3
Intermediate probability	105	105/251 = 0.418	217	217/630 = 0.344	1.2
Low probability	39	39/251 = 0.155	273	273/630 = 0.433	0.36
Normal/near normal	5	5/251 = 0.020	126	126/630 = 0.200	0.10
Total	251	...	630	...	...

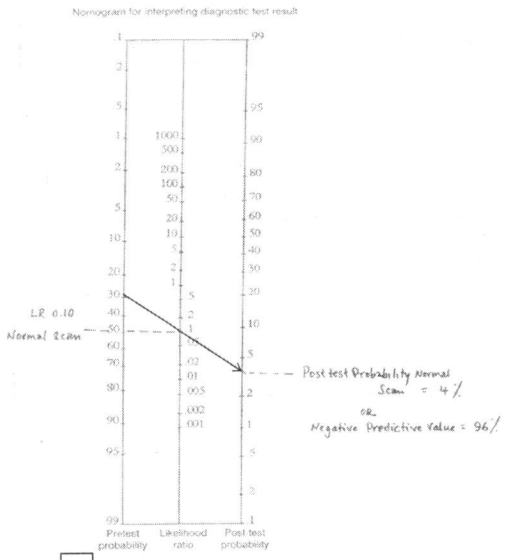


Figure [redacted] A likelihood ratio nomogram. Adapted from Fagan TJ 1975 Nomogram for Bayes's Theorem (c). New England Journal of Medicine 293: 257

Limitations of V/Q scan for diagnosis PE:

High probability (116) or normal scan (131) = useful in 247 patients

Intermediate (332) or low probability scan (312) = indeterminate in 634 patients

Table –Test Properties of Ventilation-Perfusion (V/Q) Scanning

	Pulmonary Embolism				
	Present		Absent		
V/Q Scan Result	No.	Proportion	No.	Proportion	Likelihood Ratio
High probability	102	102/251 = 0.406	14	14/630 = 0.022	18.3
Intermediate probability	105	105/251 = 0.418	217	217/630 = 0.344	1.2
Low probability	39	39/251 = 0.155	273	273/630 = 0.433	0.36
Normal/near normal	5	5/251 = 0.020	126	126/630 = 0.200	0.10
Total	251	...	630	...	...

Evaluation of Pulmonary Embolism

	High PPV	High NPV
V/Q Scan High probability	√	
V/Q Scan Normal		√
Pulmonary Angiogram	√	√
Spiral CT Scan	( )	( )
D-Dimer	( )	( )
Leg Vein Ultrasonogram	( )	( )

Measuring Diagnostic Procedures

Definitions

Sensitivity: The proportion of subjects with a disease/condition who are positive by the test being studied. Sensitivity = (number of true positives by test)/(number with disease) x 100. Sensitivity determines how good a diagnostic test is for detecting the condition it is testing for and thus being positive in patients who actually have the condition. A test that is highly sensitive has a low false-negative rate. **SnNout: If a highly sensitive (Sn) test is negative (N), the disease is ruled out.**

Specificity: The proportion of those without the disease/condition who are negative by the test being studied. Specificity = (number of true negatives by test)/(number without disease) x 100. Specificity determines how well the diagnostic test correctly identifies those patients who do not have the condition. A test that is highly specific has a low false-positive rate. **SpPin: If a highly specific (Sp) test is positive (P), the disease is ruled in.**

Positive Predictive Value: The chance that an individual will have the characteristic of interest if the test for that characteristic is positive.

Negative Predictive Value: The chance that an individual will not have the characteristic of interest if the test for that characteristic is negative.

Accuracy: The total number of true positive and true negative values for a test, divided by the total number of tests.

Likelihood Ratio: The likelihood ratio, a measure of the accuracy of a diagnostic test, determines the odds that the test result occurs in patients with the disease versus those without the disease. **The likelihood ratio for a positive test is the true-positive rate (sensitivity) divided by the false-positive rate (1 - specificity). The likelihood ratio for a negative test is the false-negative rate (1 – sensitivity) divided by true-negative rate (specificity).**

Outline

1. Clinical decision making and the role of a test
2. Two x two table notation – 4 test properties
3. Likelihood ratios and calculation of post test probabilities
4. Receiver operating characteristic (ROC) curve
5. Evaluating an article on a diagnostic test

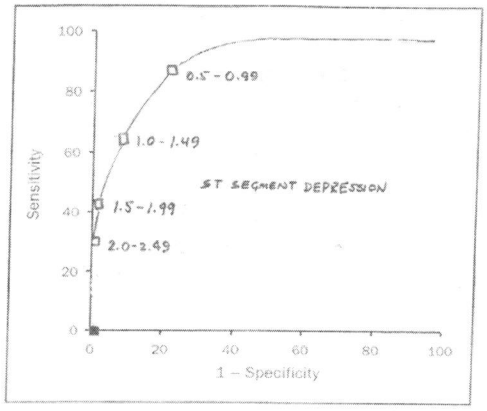
Receiver Operating Characteristic (ROC) Curve

- ROC Curve
- Area Under the Curve (AUC)
- C-Statistic (C: concordance)
- C-Index
- Discrimination

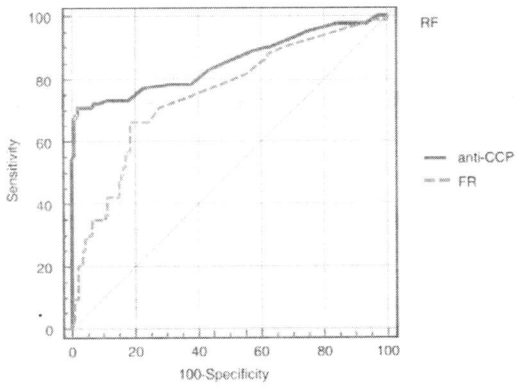
ROC Curve

- Graph that correlates true positives (sensitivity) and false positive rates (1 - specificity)
- Used in an individual study when dealing with a test that is quantitative
- Used when pooling a number of studies in a meta-analysis
- Greater area under the curve, the more accurate the test
- Provides accuracy

ROC Curve: Exercise Electrocardiography for Angiographic Coronary Artery Disease



Which diagnostic test[2] – anti-CCP antibodies or RF– is more accurate in the diagnosis of rheumatoid arthritis?



A test with an 85% specificity means:

- **(A) 85% of patients testing positive have the disease**
- **(B) 15% of normal subjects test falsely positive**
- **(C) 85% of normal subjects test negative**
- **(D) B & C are both correct**

A test which is 100% accurate means that area under the ROC curve is 100%:

- **(A) True**
- **(B) False**

If a test is 95% sensitive it would contain which of the following:

- **(A) 5% False positives**
- **(B) 5% True positives**
- **(C) 95% True positives**
- **(D) 5% True Negatives**

Outline

1. Clinical decision making and the role of a test
2. Two x two table notation – 4 test properties
3. Likelihood ratios and calculation of post test probabilities
4. Receiver operating characteristic (ROC) curve
5. Evaluating an article on a diagnostic test

Evaluating an Article on a Diagnostic Test

Important questions about validity of a study:

1. Has the test been compared with a true gold standard?
2. What is the actual probability or prevalence of the disease in the study?
3. What are the properties of the test as derived from the study?
4. Are there potential sources of bias and variation?
5. Is the test potentially relevant to my practice or are more validation studies necessary?

Sources of Bias and Variation[3]

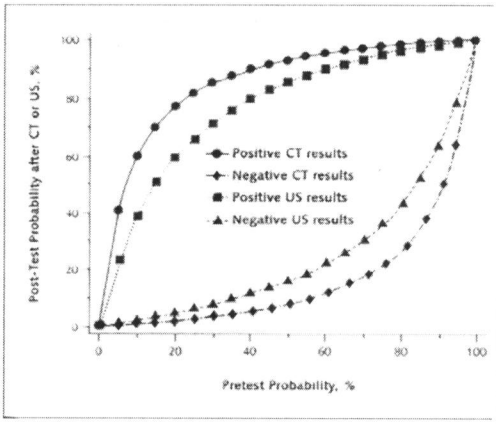
Population:	**Demographic features**
	Disease severity
	Disease prevalence
Test Protocol:	**Test technology variation**
Reference Standard:	**Partial verification bias**
Interpretation:	**Clinical review bias**
	Observer variability

***Confident about internal and external validity**

Importance of Prevalence

- **Very low or high prevalence will affect test performance; i.e., more false positives and false negatives**
- **Knowledge or awareness of prevalence of a disease in your clinical setting is important in application of test(s)**
- **Prevalence or pretest probabilities in 20-80% range can generate reasonable shifts in post-test probabilities**

Knowledge **Gained**[4]



- **Example: CT and US to detect acute appendicitis in adults and adolescents.**

Limitations of EBM

- **Studies looks at a test within a specific context**
- **Generalizing findings related to a diagnostic test beyond the specific context of a study is problematic**
- **Need to take into account how much variation exists between your patients and study population**
- **Estimating pre-test probability takes practice and experience**

5

EDITORIALS

Editorials represent the opinions of the authors and JAMA and not those of the American Medical Association.

Evidence-Based Medicine Requires Appropriate Clinical Context

Robert A. McNutt, MD, PhD
Edward H. Livingston, MD

WHAT IF A PATIENT—AFTER DIAGNOSTIC TESTS have been performed and there is no more certainty to obtain—still has a 1 in 100 chance of having venous thromboembolism (VTE)? Should the patient's physician engage the patient in a discussion of the harm and benefit of anticoagulation? What if the chance of VTE was 1 in 20? Or even 1 in 10?

Most clinicians may have difficulty answering these questions because of the need for judgment about treatment when certainty is impossible. This judgment has to balance the absolute benefit against the absolute harm for treating a disease when there is uncertainty that the disease exists. This concept reflects a treatment threshold¹ representing the probability of a disease, such as VTE, above which treatment would provide more benefit than harm, and below which treatment would produce more harm than benefit. If this probability is not known, and if variations in the individual patient's probabilities after testing is completed are not considered within the clinical practice context, physicians will have difficulty applying evidence-based medicine to patient care.

Any recommendation about the use of diagnostic testing must begin with a judgment about the value of treatment once the diagnosis is made. Diagnosis and treatment are so intimately linked that both should be part of the analysis of the utility of any diagnostic test. For example, testing for some diagnoses for which treatment is not helpful and, additionally, carries significant harm, is not a good idea. An example might be testing to find cancer in a patient with a poor prognosis from another condition that would preclude treatment for the cancer. Also, testing for a diagnosis that has an effective and beneficial treatment but with little associated harm or cost is also not good clinical care, because in this case, treatment may always be best. An example might be testing for causes of a sore throat in a patient with a high probability of a treatable infection for which no set of negative tests would dissuade a clinician from recommending treatment.[2]

See also p 438.

In this issue of JAMA, Johnson et al[3] report the results of a meta-analysis of 7 studies assessing the occurrence of VTE in 4731 patients with a single negative whole-leg compression ultrasound (CUS) result who did not receive anticoagulation; among these patients, there were 34 VTE events or suspected VTE-related deaths. The authors concluded, "withholding anticoagulation following a single negative whole-leg CUS result was associated with a low risk of venous thromboembolism during 3-month follow-up."[3]

A summary estimate from a meta-analysis reflects an "average" risk, obscuring the clinical reality of variations in probabilities. It is unlikely that any single patient in any of the studies included in this analysis had exactly the probability of VTE after a CUS of 0.37% (95% confidence interval [CI], 0.23%-0.80%). Accordingly, using this average probability for clinical decision making in some clinical contexts may do more harm than good.

Among the 7 studies included in the meta-analysis by Johnson et al,[3] the range of probabilities for developing VTE within 3 months following a negative CUS result varied from 0% to 3.6% (see Figure 2 in the article). Three of the 7 studies had a greater than average probability. In meta-analysis, weighting techniques, studies that have small standard errors, usually from large sample sizes, have greater influence on the summary estimated effect size. For example, one study by Severitt et al,[4] which contributed 1243 patients to the analysis and reported a VTE incidence of 0.48% (95% CI, 0.18%-1.05%), appears to heavily influence the overall pooled estimate of posttest probability for VTE, but included only ambulatory patients. In contrast, the authors also included data from another study by Severitt et al, which contributed 313 patients to the meta-analysis; in that report, the probability of VTE after a negative CUS result was 1.45% (95% CI, 0.34%-3.96%), but the study was performed in hospitalized patients. While it is not clear from the overall report, it is likely that the risk for developing VTE in hospitalized patients was higher than that for ambulatory patients.

Author Affiliations: Department of Medicine, Rush University Medical Center, Chicago, Illinois (Dr McNutt); and Division of Gastrointestinal and Endocrine Surgery, University of Texas Southwestern Medical Center, Dallas (Dr Livingston); Drs Moffatt and Livingston are Contributing Editors, JAMA.
Corresponding Author: Edward H. Livingston, MD, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Room F7-126, Dallas, TX 75390 (edward.livingston@utsouthwestern.edu).

Key Points

- **Actual Probability = prevalence = pretest probability**
- **Multiple test properties can be confusing**
- **ROC and area under the curve (AUC) = accuracy**
- **Prevalence or pretest probability range of 20-80% (intermediate) is where tests best applied**
- **Limitations of EBM**

6

Identifying depression in primary care: a comparison of different methods in a prospective cohort study

Verena Henkel, Roland Mergl, Ralf Kohnen, Wolfgang Maier, Hans-Jürgen Möller, Ulrich Hegerl

Department of Psychiatry, Ludwig-Maximilians-University Munich, Nußbaumstr 7, D-80336 Munich, Germany
Verena Henkel *psychiatrist*
Roland Mergl *psychologist*
Hans-Jürgen Möller *professor*
Ulrich Hegerl *professor*
Institute for Medical Research Management and Biometrics (IMBERM), Scheurlstr 21, D-90478 Nuremberg, Germany
Ralf Kohnen *professor*
Department of Psychiatry, University of Bonn, Sigmund-Freud-Str 25, D-53105 Bonn, Germany
Wolfgang Maier *professor*
Correspondence to: V Henkel verena.henkel@psymed.uni-muenchen.de

BMJ 2006;326:200–1

Depressive disorders are a major health problem in primary care, and at least half of these disorders remain undetected.[1] There are two recommended approaches to diagnosing depression in primary care: one is to perform routine screening, and the other is to evaluate patients only when the clinical presentation triggers the suspicion of depression. Our aim was to compare these two approaches, and to compare three different screening tools in order to evaluate which would be most appropriate for use in primary care. From among the many available screening tools, we selected three brief, self rating instruments: one disorder-specific (the depression module of the brief patient health questionnaire (B-PHQ, 9 items)),[2] one broad based (the general health questionnaire (GHQ-12, 12 items)),[3] and one that is less restricted to both issues (WHO-5 wellbeing index (WHO-5, 5 items)).[4]

Methods and results

Eighteen primary care facilities participated in our prospective cohort study. The study protocol was approved by our local ethics committee. On one given day, all patients who presented in one of the practices were asked to complete the three screening questionnaires before seeing a doctor. The doctors who treated the patients remained blind to the questionnaire results until they had completed a brief "physician's encounter form" to indicate their clinical assessment of their patient's current diagnoses.

Within a period not exceeding six days after they had completed the questionnaires, the patients were contacted by telephone for a fully structured, standardised psychiatric interview (composite international diagnostic interview (CIDI)) conducted by a trained psychologist blind to the screening results. We chose the composite international diagnostic interview as the reference standard because its reliability and validity have been established.[5] The interviewing psychologists met a high standard of inter-rater reliability.

The main outcome measures were, firstly, the family doctors' performance in detecting depression without any tool to help guide diagnosis decisions and, secondly, the test accuracy of the screening questionnaires. We calculated sensitivity, specificity, and predictive values using two-by-two tables. We used two statistical tests to compare differences of characteristics of test accuracy (table).

For 431 patients, all screening questionnaires, the composite international diagnostic interview, and the physician's encounter form were completed. Of these patients, 17% suffered from any depressive disorder and 83% did not.

Comment

The sensitivity of the family doctors' unaided clinical diagnoses was 65%. With standard cut-off points, the briefest screening questionnaire (and therefore the most practical to use), the WHO-5, produced significantly greater sensitivity (93%) and a better negative predictive value (98%) than the other questionnaires (see table). However, the brief patient health questionnaire and unaided clinical diagnosis produced better specificity. The brief patient health questionnaire also produced the best positive predictive value. However, since screening tools are designed to identify all patients at risk for a disorder, sensitivity and negative predictive value are the most important operating characteristics.

Comparison of test accuracy of screening questionnaires for depression and family doctors' unaided clinical diagnosis. Values are means (95% confidence intervals) unless stated otherwise

Measures of test accuracy	Screening questionnaires			Unaided clinical diagnosis (UCD)	Significant differences ($P \leq 0.05$, one sided tests)§
	WHO-5*	GHQ-12†	B-PHQ‡		
Sensitivity (%)	93 (85 to 98)	85 (74 to 92)	78 (66 to 87)	65 (53 to 76)	WHO-5>GHQ-12, B-PHQ>UCD
Negative predictive value (%)	98 (95 to 99)	95 (92 to 98)	95 (92 to 97)	91 (88 to 94)	WHO-5>B-PHQ>UCD, GHQ-12>UCD
Specificity (%)	64 (59 to 69)	62 (57 to 67)	85 (81 to 89)	74 (69 to 79)	B-PHQ>UCD>WHO-5, UCD>GHQ-12
Positive predictive value (%)	34 (28 to 41)	31 (25 to 38)	51 (42 to 61)	34 (26 to 42)	B-PHQ>WHO-5>GHQ-12, B-PHQ>UCD

*WHO-5 wellbeing index (scoring procedure as indicated in *World Health Organization info package*[4]).
†General health questionnaire (scoring procedure as indicated in Goldberg 1978[3]).
‡Brief patient health questionnaire (scoring procedure as indicated in Spitzer et al 1999[2]).
§McNemar's test to compare sensitivities and specificities, analogue of McNemar's test to compare predictive values.

Our results suggest that the use of WHO-5 could improve family doctors' ability to detect depression, supporting the World Health Organization's recommendation that every patient in primary care should participate in a screening process with the completion of WHO-5 as a standard first step, done in the waiting room.[4] The questionnaire can easily be scored by hand. Patients who score positively for depression should be examined by their doctor in order to confirm a diagnosis of depression or to rule out normal distress or physical causes of depression. At this stage, doctors could use the brief patient health questionnaire as a checklist.

We hope that our results favouring such a simple, two stage screening process for depression in primary care, starting with the questionnaire WHO-5, will encourage further research in other countries.

We thank Simone Braun, Kathrin Allgaier, Petra Ohlendorf, Isabelle Seidscheck, and Evelyn Puth for data collection. We thank Jan Stefanek and Simone Braun for conducting the ROC-analyses presented in an earlier draft of this paper.

Contributors: VH had the idea for this paper and drafted the paper. RM analysed the data. RK, WM, H-JM, and UH commented on the study protocol and the text of the paper. UH is the speaker of the "German Research Network on Depression." VH and UH are guarantors for the study.

Funding: The study was funded by grants from the German Federal Research Ministry within the programme "German Research Network on Depression" and by additional funds from Pfizer and Novartis.

Competing interests: None declared.

1 Paykel ES, Tylee A, Wright A, Priest RG, Rix S, Hart D. The defeat depression campaign: psychiatry in the public arena. *Am J Psychiatry* 1997;154(6 suppl):59–65.
2 Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. *JAMA* 1999;282:1737–44.
3 Goldberg DG. *Manual of the general health questionnaire*. Windsor: NFER Publishing, 1978.
4 *World Health Organization info package: Mastering depression in primary care*. Frederikaborg: World Health Organization, Regional Office for Europe, Psychiatric Research Unit, 1998.
5 Andrews G, Peters L. The psychometric properties of the composite international diagnostic interview. *Soc Psychiatry Psychiatr Epidemiol* 1998;33:80–8.

(Accepted 15 August 2002)

Questions

- **Gold standard?**
- **Prevalence?**
- **Best screening questionnaire?**
- **Limitations/Bias?**
- **Applicable to your patients?**

References

- 1. Division of Lung Diseases, National Heart, Lung, and Blood Institute. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED) The PIOPED Investigators. *JAMA*. 1990;263(20):2753-9.
- 2. Tampoia M, Brescia V, Fontana A, et al. Proteomic: new advances in the diagnosis of rheumatoid arthritis. *Clinica Chimica Acta*. 2005;357:219-225.
- 3. Whiting P, Rutjes AWS, Reitsma JB, et al. Sources of variation and bias in studies of diagnostic accuracy. *Ann Intern Med*. 2004;140:189-202.
- 4. Terasawa T, Blackmore CC, Bent S, et al. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. *Ann Intern Med*. 2004;141:537-546.
- 5. McNutt RA, Livingston EH. Evidence-based medicine requires appropriate clinical context. JAMA. 2010;303:454-455.
- 6. Identifying depression in primary care: a comparison of different methods in prospective cohort study. BMJ. 2003;326:200-201.

Extra Information:
Post-Test Probability using Bayes

Probability of condition given a positive test:

Bayes' Theorem: $P(A \mid B) = \dfrac{P(A)\,P(B \mid A)}{P(B)}$

Example:

P(A) = probability of PE, prevalence 0.28

P(B | A) = probability of patient with PE getting a 'High' classification, sensitivity 0.40

P(B) = probability of a "High" classification, 0.13

$$\frac{(0.28)\,(0.40)}{(0.13)} = 0.86$$, *some rounding error included*