Sensitivity, Specificity and Predictive Value

H.S. Teitelbaum, DO, PhD, MPH
DCOM

Pretest Questions

	Sick	Well	Total
Test +	a	b	a + b
Test -	c	d	c + d
Total	a + c	b + d	a + b + c + d

Either in terms of letters or words define the terms below:

1. Sensitivity
2. Specificity
3. Positive Predictive Value
4. Negative Predictive Value
5. Positive Likelihood Ratio
6. Negative Likelihood Ratio
7. Prevalence

Screening Tests

- Tests done among apparently well people to identify those at an increased risk of a disease or disorder
- Those identified are sometimes offered a subsequent diagnostic test or procedure, or in some instances, a treatment or preventative medication
- Can improve health, but inappropriate screening harms healthy individuals and squanders resources

Screening Tests

- Mammography
- Colonoscopy
- PSA levels
- BP
- Cervical PAP smears
- DEXA scans
- Chest X-Ray/CT scans
- HIV-1 testing in pregnancy
- Phenylalanine level testing in newborns

Ethical Implications

- *What are the potential harms of screening?*
- Screening engages apparently healthy individuals who are not seeking medical help
  - Might prefer to just be left alone
- Consumer-generated demand for screening might lead to expensive programs of no clear value

Ethical Implications

- Cost, injury and stigmatization must be considered
- Medical and ethical standards should be higher than with diagnostic tests
- Every adverse outcome of screening is iatrogenic and entirely preventable
- May be inconvenient, uncomfortable and expensive

Ethical Implications

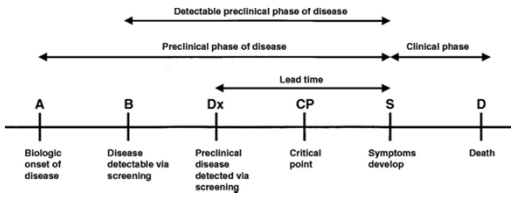
- Second wave of injury can arise after initial screening insult
  - False-positive results
  - True-positive results leading to dangerous interventions

Criteria for Screening

If a test is available, should it be used?

- Availability does not imply a test should be used for screening
- Before screening is done:
  - The disease should be medically important and clearly defined
  - The natural history should be known
    - Early detection should lead to a more favorable prognosis
    - Preclinical disease left untreated will lead to clinically evident disease with no spontaneous regression
  - An effective intervention should exist
  - Screening program should be cost-effective
  - Course of action after a positive result must be agreed on in advance and acceptable to those screened

—Diagram shows natural history of disease.



Herman C R et al. AJR 2002;179:825-831

AJR

Criteria for Screening

If a test is available, should it be used?

- The test should also "do its job"
  - Safe
  - Reasonable cut-off level defined
  - Be valid
    - Ability of the test to measure what it sets out to measure
    - Differentiates those with from those without the disease
  - Be reliable
    - Implies repeatability

Criteria for Screening

If a test is available, should it be used?

- Although early diagnosis has intuitive appeal, earlier might not always be better
  - ***Alzheimer's disease***
- What benefit might accrue, and at what cost from early (earlier) diagnosis?
  - Does early diagnosis really benefit those screened?
    - Survival
    - Quality of life
  - Will those diagnosed earlier comply with the proposed treatment?
  - Has the effectiveness of the screening strategy been objectively established?
  - Are the cost, accuracy and acceptability of the test clinically acceptable?

Assessment of test effectiveness

Is the test valid?

- **Sensitivity**
- **Specificity**
- **Positive predictive value**
- **Negative predictive value**

- Terminology used for over 50 years
- Clinically useful
  - Predicated on assumption that is often clinically unrealistic
  - All people can be dichotomized as ill or well
  - Do not fit all patients
  - Likelihood ratios used to refine clinician judgment about probability of disease
    - Incorporate varying degrees of test results
    - Not just positive or negative

Basic Notions

- Think of a proportion $\frac{x}{x+y}$
- Think of a 2 x 2 table.
- THINGS WORK THE WAY YOU WANT THEM TO WORK

Basic Structure

	True State of Affairs Sick	Well	
Test Results +	a	b	a + b
Test Results -	c	d	c + d
	a + c	b + d	a + b +c + d

Sensitivity

- If you knew someone was sick, what would you want their test result to be?

$$\frac{a}{a+c}$$

- We call this **sensitivity; The probability that a sick individual will have a positive test.**

Specificity

- If you knew someone was WELL what would you want their test to be?

$$\frac{d}{b+d}$$

We call this **SPECIFICITY**. The probability that a well individual will have a negative test.

Positive Predictive Value (PPV)

- If you knew someone had a positive test, what would you want the health status to be?

$$\frac{a}{a+b}$$

We call this **POSITIVE PREDICTIVE VALUE**; It is the probability that those who have a positive test are really sick

Negative Predictive Value (NPV)

- If someone had a negative lab test; what would you want their health status to be?

$$\frac{d}{c+d}$$

- We call this the **NEGATIVE PREDICTIVE VALUE.** This is the probability that those who have a negative test are really well.

Prevalence

- **Prevalence** – what is the probability of disease in the population you are studying?

$$\frac{a+c}{a+b+c+d}$$

Error Rates

- False positive error rate (Type I error)
  b / (b + d)
- False negative error rate (Type II error)
  c / (a + c)

A Second Look

	True State of Affairs		
	Sick	**Well**	
Test Results +	True Positive TP	False Positive FP	TP + FP
Test Results -	False Negative FN	True Negative TN	FN + TN
	TP + FN	FP + TN	TP + TN + FP + FN

True, False, Positive and Negative refer to the TEST RUSULTS

Sensitivity

- Detection rate
- Ability of a test to find those with the disease
- True positive implies that an individual with the disease will test positive

Specificity

- Ability of a test to identify those without the disease
- True negative implies that a person without disease will have a negative test

Sensitivity and Specificity

- Population measures
- Look backward at results gathered over time
- Generally not as valuable to clinicians
  - Must interpret test results to those tested
- Clinicians need to know the **predictive values** of the test

Predictive Values

- Individual measures
- Look forward
- Work horizontally in 2x2 tables as compared to sensitivity and specificity which works vertically in 2 x 2 tables.

Diagnostic accuracy

- Implies simplification of four indices of test validity
- No single term describes trade-offs between sensitivity and specificity that generally arise
- Sum of those correctly identified as ill and well divided by all those tested
- Essentially the proportion of correct results

(A + D) / (A + B + C + D)

Example

Screening Test	**Diastolic Hypertension** **Yes**	**No**	
Positive	36	25	61
Negative	9	230	239
	45	255	300

Please use the above formulae to calculate the following measures

Calculate

Sensitivity
Specificity
Positive Predictive Value
Negative Predictive Value
Prevalence
False Positive Rate
False Negative Rate

Calculations

Sensitivity

$$\frac{a}{a+c} = \frac{36}{45} = 0.80$$

Specificity

$$\frac{d}{b+d} = \frac{230}{255} = 0.90$$

Calculations

- Positive Predictive Value

$$\frac{a}{a+b} = \frac{36}{61} = 0.59$$

Negative Predictive Value

$$\frac{d}{c+d} = \frac{230}{239} = 0.96$$

Calculations

- Prevalence

$$\frac{a+c}{a+b+c+d} = \frac{45}{300} = 0.15$$

False Positive Rate

$$\frac{b}{b+d} = \frac{25}{255} = 0.098$$

Calculations

False Negative Rate

$$\frac{c}{a+c} = \frac{9}{45} = 0.20$$

Sensitivity and Specificity 2

H.S. Teitelbaum, DO, PhD, MPH
Department of Internal Medicine
DCOM

The Physician's Dilemma

		Disease +	Disease -
TEST	+	A True Positive	B False Positive
	-	C False Negative	D True Negative

This is what we want

		Disease
TEST	+	A + B All people with + tests
	-	C + D All people with - tests

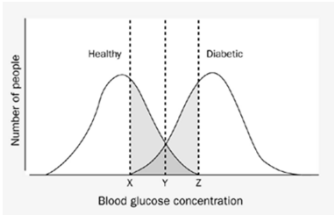
This is what we know

Trade-offs between sensitivity and specificity
Where should the cut-off for abnormal be?

- Ideal test would perfectly discriminate between those with and those without the disorder
- The distribution of test results for the group would be bimodal and not overlap
- More commonly test values for those with and those without a disease overlap, sometimes widely
- Where one puts the cut-off defining normal versus abnormal determines the sensitivity and specificity

Trade-offs between sensitivity and specificity
Where should the cut-off for abnormal be?

- For any continuous outcome measurement, the sensitivity and specificity of a test will be inversely related
  - Blood pressure
  - Intraocular pressure
  - Blood glucose
  - Serum cholesterol
- Low cutoff will identify all with a condition, but many normals will be identified incorrectly



Cut-off at *x* produces perfect sensitivity

Identifies all those with diabetes

Cut-off at *y* appears compromise

Trade-off is poor specificity

Those in the healthy distribution (pink and purple) are incorrectly identified as having abnormal values

Cut-off at *z* produces perfect specificity

All healthy are correctly identified, but significant proportion of those with diabetes are missed

Trade-offs between sensitivity and specificity
Where should the cut-off for abnormal be?

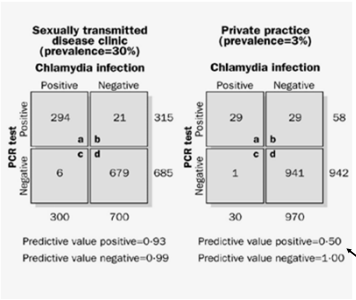
- Where the cut-off should be depends on the implications of the test
  - Receiver-operator characteristic curves useful in making this decision
- Screening for PKU in newborns places a premium on sensitivity rather than on specificity
  - The cost of missing a case is high
  - Effective treatment exists
  - Downside is a large number of false-positive tests
    - Causes anguish and further testing
- Screening for breast cancer should favor specificity
  - Further assessment of those tested will result in biopsies that are invasive and costly

Prevalence and Predictive values
Can test results be trusted?

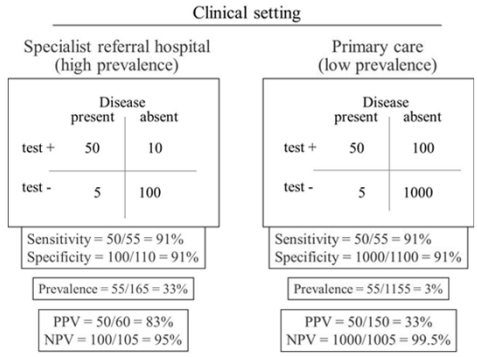
- Disease prevalence has **strong effect** on predictive values
- Clinicians must known approximate prevalence of condition of interest in population being tested
  - If not, reasonable interpretation is impossible

Consider new PCR test for chlamydia:

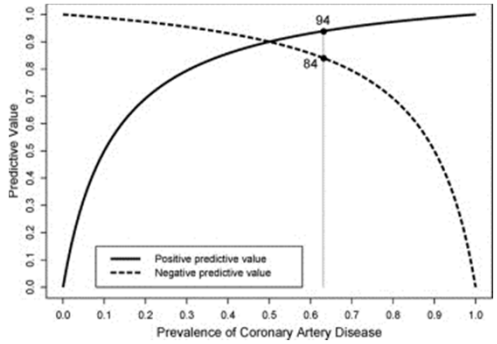
Sensitivity = 0.98; Specificity = 0.97



Flipping a coin has same positive predictive value and is cheaper than searching for bits of DNA



http://phprimer.afmc.ca/sites/default/files/primer_versions/57605/primer_images/image13.jpg?1321287867



http://ars.sciencedirect.com/content/image/1-s2.0-S0735109711048194-gr3.jpg

Prevalence and Predictive values

Can test results be trusted?

- When used in low-prevalence settings, even excellent tests have a poor positive predictive value
- The reverse is true for negative predictive values

Tests in Combination

Should a follow-up test be done?

- Clinicians rarely use tests in isolation
- Few tests have high sensitivity and specificity
- Common approach is to do tests in sequence
  - A sensitive, but not specific, test is the initial screen
  - Those who test positive will get a second, more specific, test
  - Only those who test positive on both are given a diagnosis

Prevalence and Predictive values

Can test results be trusted?

- Syphilis
  - Reagin test
    - Sensitive, but not specific
  - If positive, treponemal test
    - Specific
  - If both positive, then patient has syphilis
- HIV-1
  - ELISA
  - Western Blot

Sequential Testing

- Give one test, if + send them for another test, if again +, then declare the person as having the disease.
- I want to illustrate this.

Example[1]

Test 1 Blood Sugar
Sensitivity = 70%
Specificity = 80%

		Diabetes +	Diabetes -	
Test Results	+	350	1900	2250
	-	150	7600	7750
		500	9500	10,000

Assume: Disease Prevalence is 5% Population is 10,000

[1]Epidemiology, 3rd edition. Leon Gordis. Elsevier 2004, pg. 77

Example Cont'd

Test 2 (Glucose Tolerance Test)
Sensitivity = 90%
Specificity = 90%

		Diabetes +	Diabetes -	
Test Results	+	315	190	505
	-	35	1710	1745
		350	1900	2250

Net Sensitivity = 315/500 = 63% **Net Specificity = (7600 + 1710)/9500 = 98%**

Conclusion

- In sequential testing: Net Sensitivity decreases; Net Specificity increases.

Simultaneous Testing

Test A

TEST	Disease +	Disease -
+	160	320
-	40	**480**
	200	800

Sensitivity = 80%
Specificity = 60%

Test B

TEST	Disease +	Disease -
+	180	80
-	20	**720**
	200	800

Sensitivity = 90%
Specificity = 90%

Follow the logic

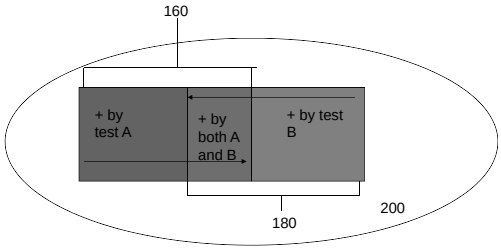
- We will administer both tests simultaneously. You will draw a vile of blood and analyze the sample with both tests.
- For a person to be negative (–) THEY MUST BE NEGATIVE ON BOTH TESTS
- For a person to be positive, they are positive on either test.

For NET SENSITIVITY

- Please follow on the previous tables
- Test A, with a sensitivity of 80%, identifies 160 of the 200 people as +.
- Test B, with a sensitivity of 90%, identifies 180 OF THE SAME 200 people as +.
- Thus, some of the people have tested + by both tests.
- Thus if we add those who tested + on test A to those who tested + on test B we will have counted some people twice.

Net Sensitivity cont'd

Diagrammatically



Think of it this way

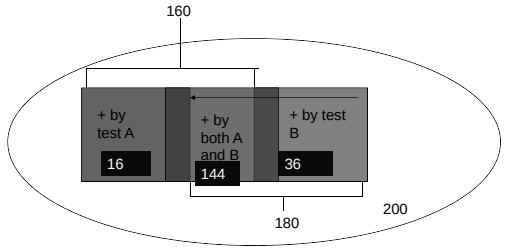
- OF the 160 people identified as + by A, the sensitivity of Test B would identify (.90 x 160) of them; or 144 people.
- We could go the other way and say of the 180 people identified as + by B, the sensitivity of Test A would identify (.8 x 180) of them; or 144 people.
- Thus 144 people are + by both A and B.

So now we have the components of the numerator for SIMULTANEOUS TESTING

- 160 – 144 would be those who test + by A alone. (16 people)
- 180 – 144 would be those who test + by B alone. (36 people)
- The numerator would be 16 +144 + 36 or 196.
- The denominator would be the 200 prevalent cases.
- NET SENSITIVITY = 196 / 200 = **98%**

Net Sensitivity cont'd

Diagrammatically



Net SPECIFICITY for Simultaneous Testing

- Use the same tables. Remember, we need to be (–) by BOTH TESTS.
- We will use the same logic as before.
- Test A with specificity of 60%, identified 480 of the 800 people as (-) (.6 x 800 = 480). These are true negatives by Test A.
- Test B with a specificity of 90% identified 720 of the 800 people as (–) (.9 x 800)

Net SPECIFICITY CONT'd

- SO to identify those who test (–) by both tests, we do the following:
- Test A identified 480 people. Test B with a specificity of .9 would identify (.9 x 480 = 432) of them as well.
- We could start with Test B. Test B identified 720 people. Test A with a specificity of .6 would identify (.6 x 720 = 432) of them as well.

SO

- Numerator = 432
- Denominator = 800
- Net SPECIFICITY = 432 / 800 = 54%.

CONCLUSION

- When simultaneous tests are used, there is a net GAIN in SENSITIVITY and net LOSS in SPECIFICITY.
- In sequential testing, there is a net LOSS in SENSITIVITY and a net GAIN in SPECIFICITY.
- In clinical medicine we do both.

OLD FRIENDS

	With Disease D+	Without Disease D-	Total
Test Positive T+	True Positive TP	False Positive FP	TP +FP
Test Negative T-	False Negative FN	True Negative TN	FN + TN
TOTAL	TP + FN	FP + TN	TP + FP +FN + TN

Sensitivity = TP/TP +FN
Specificity = TN/FP +TN
Prevalence = (TP + FN) / (TP +FP +FN + TN_
Positive Predictive Value = TP / All positive tests
Negative Predictive Value = TN / All negative tests

The likelihood ratio incorporates both the sensitivity and specificity of the test and provides a direct estimate of how much a test result will change the odds of having a disease. The likelihood ratio for a positive result (LR+) tells you how much the odds of the disease increases when a test is positive. The likelihood ratio for a negative result (LR-) tells you how much the odds of the disease decreases when a test is negative.

Equations for LR (+) and LR (-)

- LR (+) = (Sensitivity)/(1- Specificity)
- LR (-) = (1-Sensitivity)/ (Specificity)

LR (+) = (a/a+c)/ (1- (d/b+d))
LR (-) = (1- (a/(a+c))/ (d/(b+d))

Note: LR (+) > 10 are generally highly useful

Recall

- Sensitivity and Specificity are not effected by Prevalence.
- Predicted values are effected by prevalence.
- Combining these two statements we can infer the following (Sackett, 1992)

Sensitive signs when **N**egative help rule **out** disease (**SnNout**)

Specific signs when **P**ositive, help rule **in** the disease **(SpPin)**

•CTA had 83% sensitivity and 96% specificity
positive likelihood ratio 19.6 and negative likelihood ratio 0.18
positive predictive value (PPV) 86% (95% CI 79%-90%) overall

96% (95% CI 78%-99%) if high-clinical probability
92% (95% CI 84%-96%) if intermediate clinical probability
58% (95% CI 40%-73%) if low-clinical probability

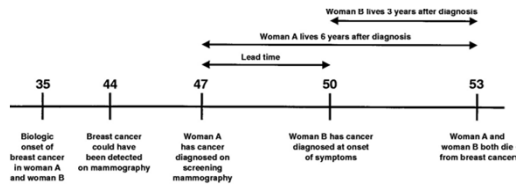
NPV 95% (95% CI 92%-96%) overall■96% (95% CI 92%-98%) if low-clinical probability
89% (95% CI 82%-93%) if intermediate clinical probability
60% (95% CI 32%-83%) if high-clinical probability

http://web.ebscohost.com.ezproxy.lmunet.edu/dynamed/detail?sid=a30a5efc-4447-4477-b970-eecba195740b%40sessionmgr11&vid=2&hid=18&bdata=JnNpdGU9ZHluYW1lZC1saXZlJnNjb3BlPXNpdGU%3d#db=dme&AN=115857&anchor=searchmatch_3

Benefit or Bias?

- Does a screening program really improve health?
- Lead-time bias
- Length bias
- Self-selection bias
- Over diagnosis bias

—Diagram depicts how lead-time bias can result in apparent increase in survival attributable to screening.

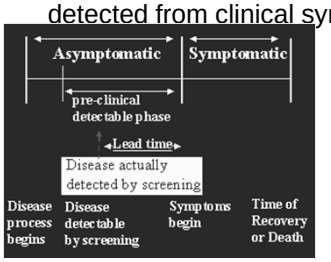


Herman C R et al. AJR 2002;179:825-831

AJR

Lead-time

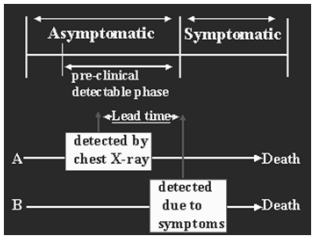
- The interval between "diagnosis" of a disease and when it would have been detected from clinical symptoms



Lead time is the amount of time gained by earlier detection of a cancer by screening than by later detection with the appearance of symptoms. This can be seen in the Figure above. If this lead time is not associated with a decrease in mortality, then lead time bias is present.

70

Lead-time



In this example, using lung cancer for which clinical trials have demonstrated no efficacy for screening, the principle of lead time is illustrated. Despite person A being diagnosed with disease earlier than person B, they both die at the same time. Thus, no decrease in mortality was gained by person A, only the length of time during which he knew he was sick was increased. Time with disease is extended which leads to the false impression that early detection improves total survival

HST 2012 AOCOPM 71

Lead-time bias

- Refers to a spurious increase in longevity associated with screening simply because diagnosis was made earlier in the course of the disease
- Assume mammography screening leads to cancer detection 2 years earlier than would have ordinarily occurred, yet screening did not prolong life
- On average women with breast cancer detected through screening live 2 years longer than those with cancers detected by traditional means

HST 2012 AOCOPM 72

Lead-time bias

- The gain in longevity is apparent and not real
- This hypothetical screening allows women to live 2 years longer with the knowledge that they have cancer, but does not prolong survival
- Example of lead-time shift

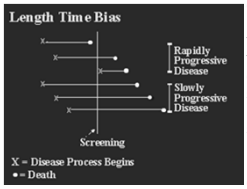
HST 2012 AOCOPM 73

Length bias
Prognostic selection

- More subtle than lead-time bias
- Longevity association is real, but indirect
- Assume mammography screening in a community in 10 year intervals
- Women whose cancers were detected through screening live 5 years longer on average from cancer initiation to death than those whose cancers were detected by usual means

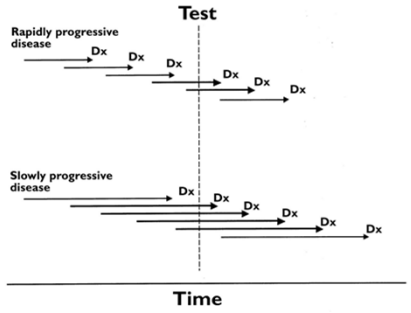
HST 2012 AOCOPM 74

Length-time Bias



Slowly progressing tumors have more opportunity than faster ones to be detected by screening. In addition, slowly progressive tumors take longer to lead to death than faster ones. Therefore, the screen-detected cancers will appear to have an increased survival after diagnosis, giving the mistaken impression that screening leads to improved survival. In reality, the improved survival is a result of these cancers being more slowly progressing. Thus, the survival rate of a group of people with screen-detected cancers will be artificially increased due to length time bias compared with the survival rate of those with non screen-detected cancers.

—Length-time bias.



Stanley R J AJR 2001;177:989-992

AJR

Length bias

- That screening is associated with longer survival seems to impart clear benefits
- The benefit may be just the inherent variability in cancer growth rates and not a benefit of screening
  - Reflects long preclinical phase as compared to patients with more aggressive illness and short preclinical phase
- Women with indolent, slow-growing tumors are more likely to live long enough to be identified in a 10 year screening
- Those with rapidly progressing tumors are less likely to survive until screening

Lead-time and Length-time Bias

- Because of lead and length time biases, survival with a disease cannot be used to assess the efficacy of screening.
- The ultimate evaluation outcome of the efficacy of a screening test is a comparison of the mortality rate of the population screened with the mortality rate of the non-screened population.

Self-selection bias

- Volunteers for screening programs may be healthier, on average, than persons who do not participate in screening programs
- The “worried well” may also be more likely to participate and may be at overall higher risk because of family history or lifestyle

Over-Diagnosis Bias

- Persons who screen positive and are truly disease free, yet are erroneously diagnosed as having the disease (false positives)
- Since these persons are truly disease-free, we expect a more favorable long-term outcome
  - Gives the appearance of a very effective screening program