

Screening and Diagnostic Testing

David Dayya, D.O., M.P.H.

Saint Barnabas Hospital

Department of Family Medicine

Overview

- Basic Definitions and Screening Goals
- Construct and Interpret a 2X2 table
- Defining and calculating sensitivity, specificity, predictive values
- Disease prevalence and predictive values
- Likelihood Ratios
- Accuracy and Reliability
- The Kappa Statistic
- Screening Examples

Screening

- Population is tested in order to identify those with asymptomatic disease or having a higher probability of developing a disease.

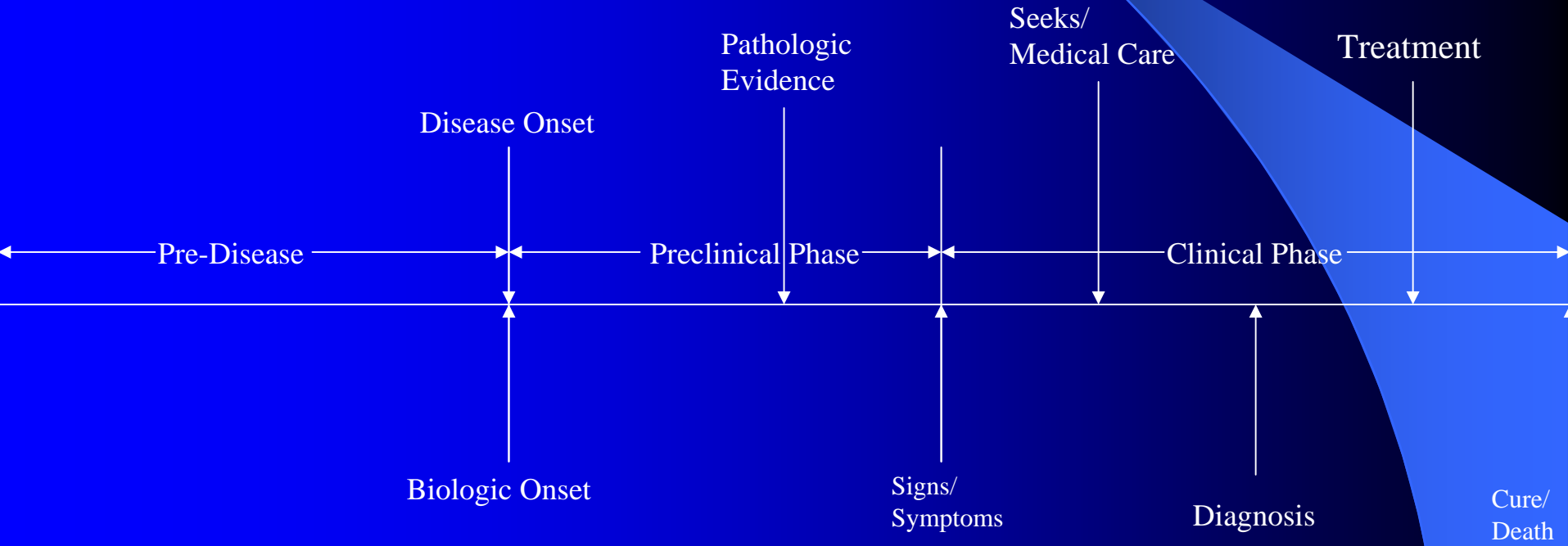
Diagnostic Testing

- Testing a patient with complaint or signs and symptoms of Disease in order to establish a diagnosis.

Goals of Screening

- Begin early treatment assuming one is available and is worth giving. Secondary Prevention.
- Prognostic considerations.
- Test is effective in identifying those with and without disease i.e. sensitivity, specificity, PPV, NPV?

Natural History of Disease in a Patient



Sensitivity and Specificity

- Proportion of those with the disease who have a positive test. True Positives Rate.
- Sensitivity is the ability of the test to find people who have the disease (i.e., new test is positive when gold standard is positive).
- Probability of a positive test among patients with disease.
- Ability to R/O Disease
- SnOut - highly Sensitive test, a Negative test rules out the disease.
- Proportion of those without the disease who have a negative test. True Negative Rate.
- Specificity is the ability of the test to identify people who do not have the disease (i.e., the new test is negative when the gold standard is negative).
- Probability of a negative test among patients without disease
- Ability to R/I Disease
- SpIn - highly Specific test, a Positive test rules in the disease

Sensitivity and Specificity

	Disease Present D+ (Gold Standard)	Disease Absent D- (Gold Standard)	
Positive Test T+	a	b	a+b
Negative Test T-	c	d	c+d
	a+c	b+d	a+b+c+d Population

Sensitivity and Specificity

	Disease Present (Gold Standard Truth?)	Disease Absent (Gold Standard Truth?)	
Positive Test	a True Positive (TP)	b False Positive (FP)	a+b
Negative Test	c False Negative (FN)	d True Negative (TN)	c+d
	a+c	b+d	a+b+c+d Population

Sensitivity

- Sensitivity = $\Pr(T+D+)$
 $a/(a+c) =$
 $TP/(TP+FN) =$
 $1 - FNR =$
 $1 - c/(a+c)$
- False negative rate =
Proportion of those with
the disease who have a
negative test.
- $FNR = c/(a+c) = \Pr(T-D+)$
 $FN/(FN+TN) =$
 $1 - \text{Sensitivity} =$
 $1 - a/(a+c)$
- Lower Sensitivity -> More
false negatives.
- Greater Sensitivity -> Less
false negatives.

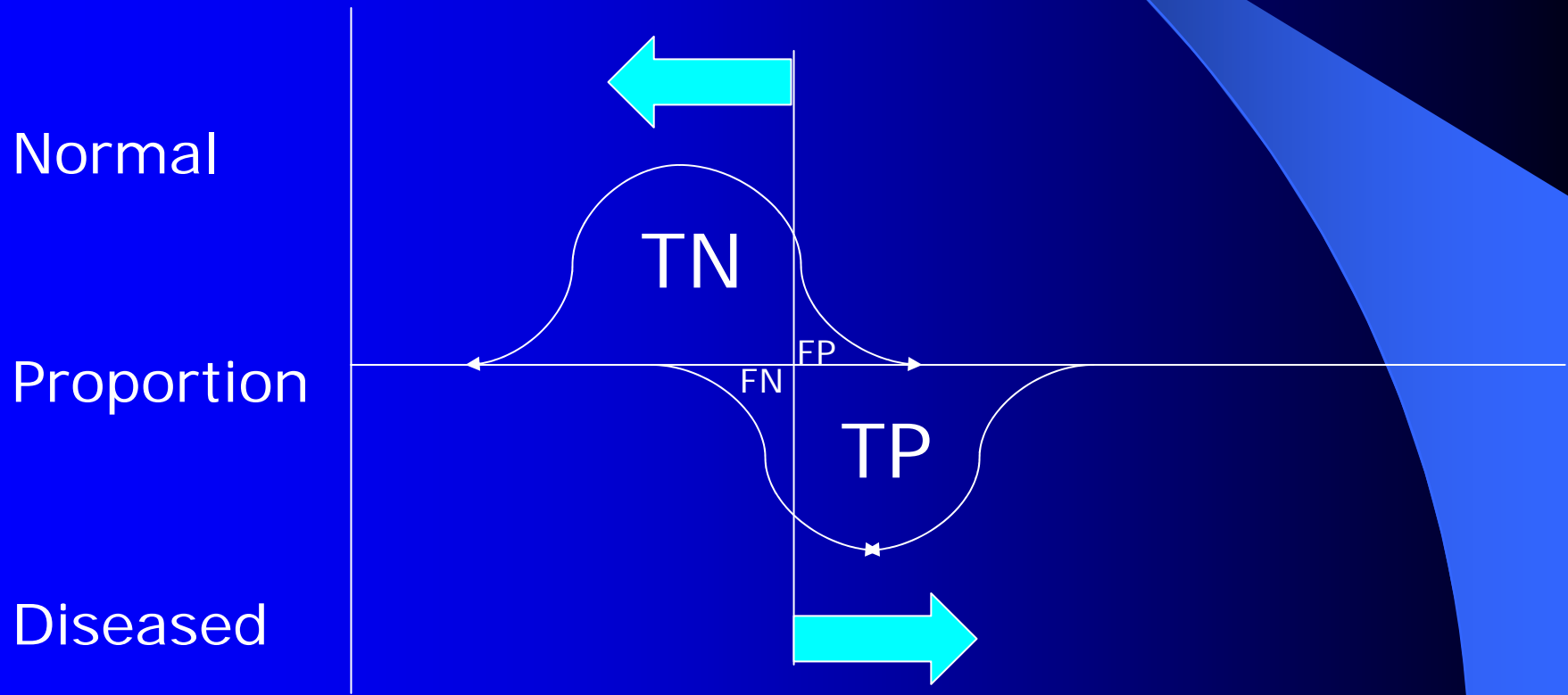
Specificity

- Specificity = $\Pr(T-D-)$
 $d/(b+d) =$
 $TN/(TN+FP) =$
 $1 - FPR =$
 $1 - b/(b+d) =$
- False positive rate =
Proportion of those
without the disease who
have a positive test.
- $FPR = b/(b+d) = \Pr(T+D-)$
 $FP/(FP+TN) =$
 $1 - \text{Specificity} =$
 $1 - d/(b+d)$
- Lower Specificity -> More
false positives.
- Greater Specificity ->
Less false positives.

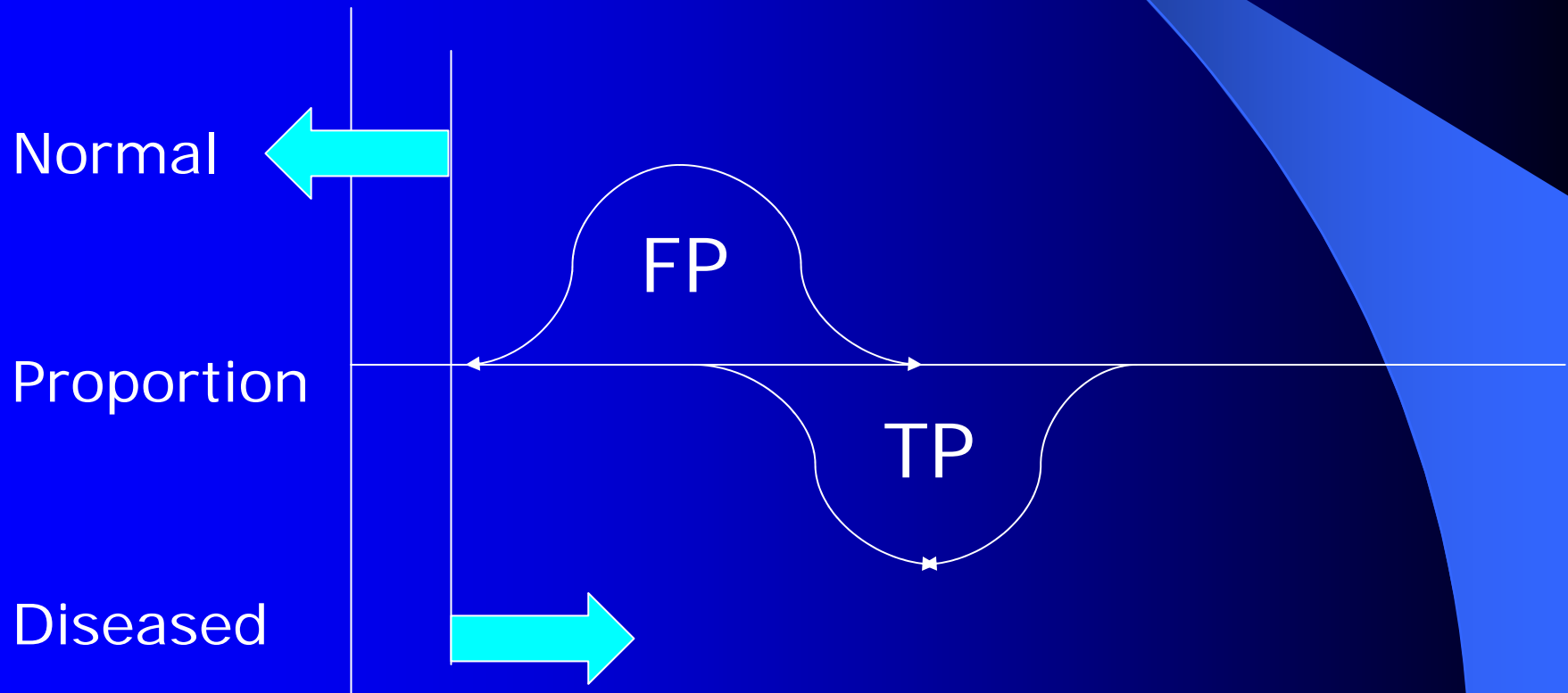
How big are the holes in your fishing net?



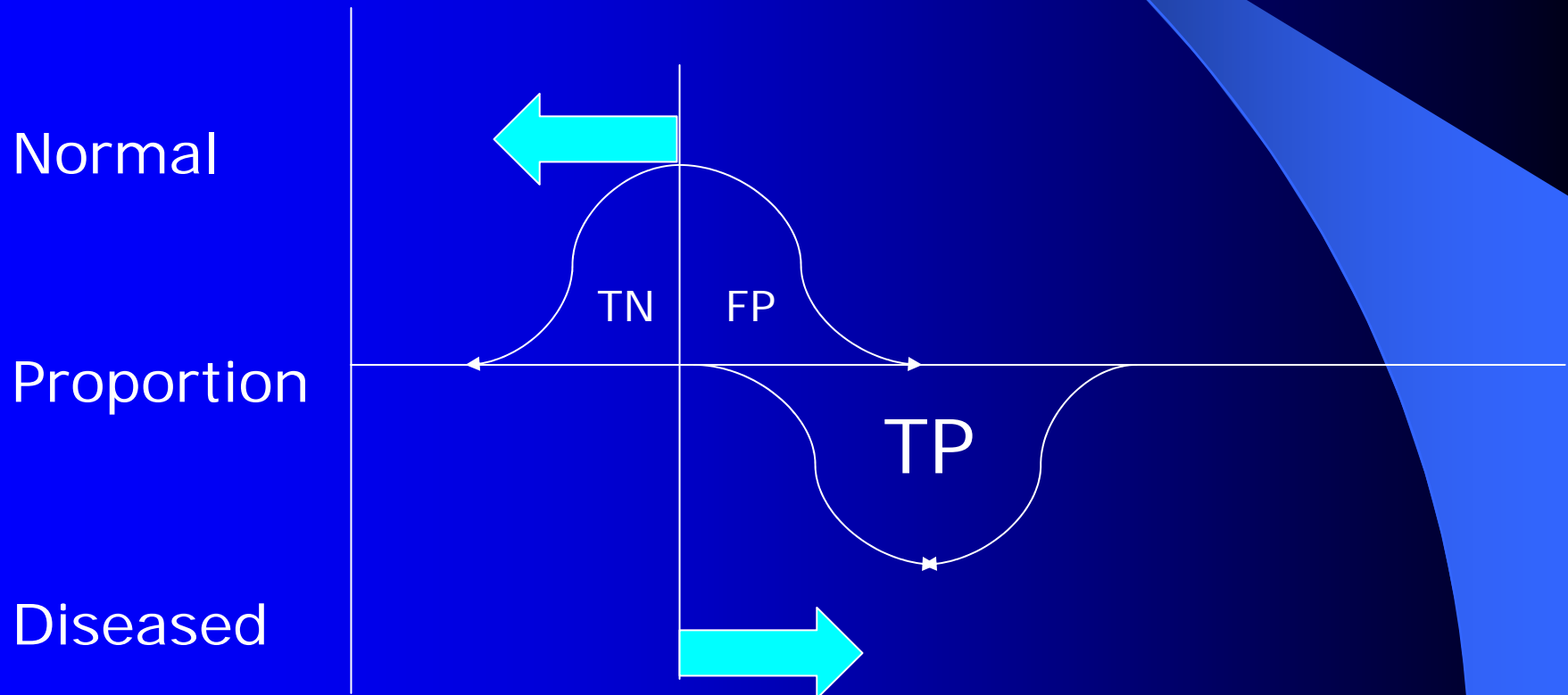
Disease Threshold



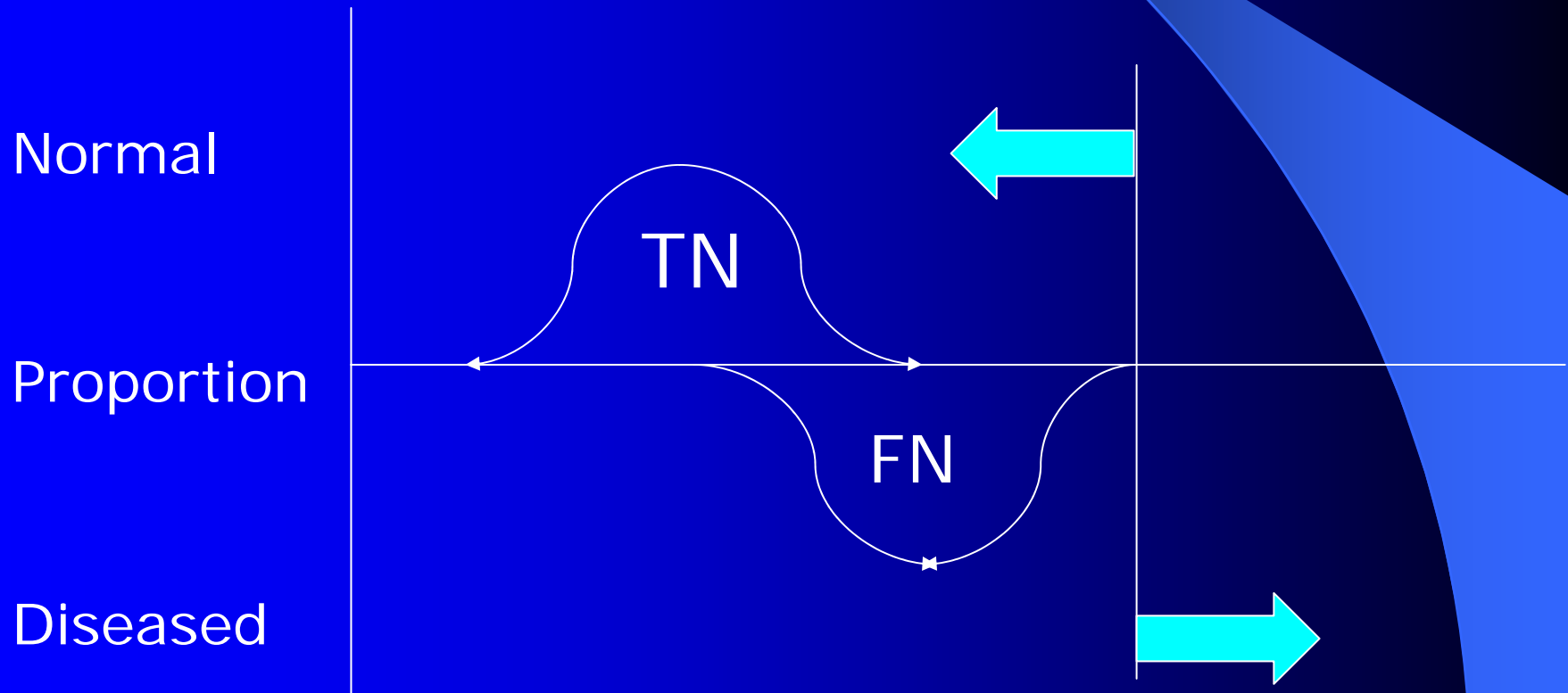
Disease Threshold



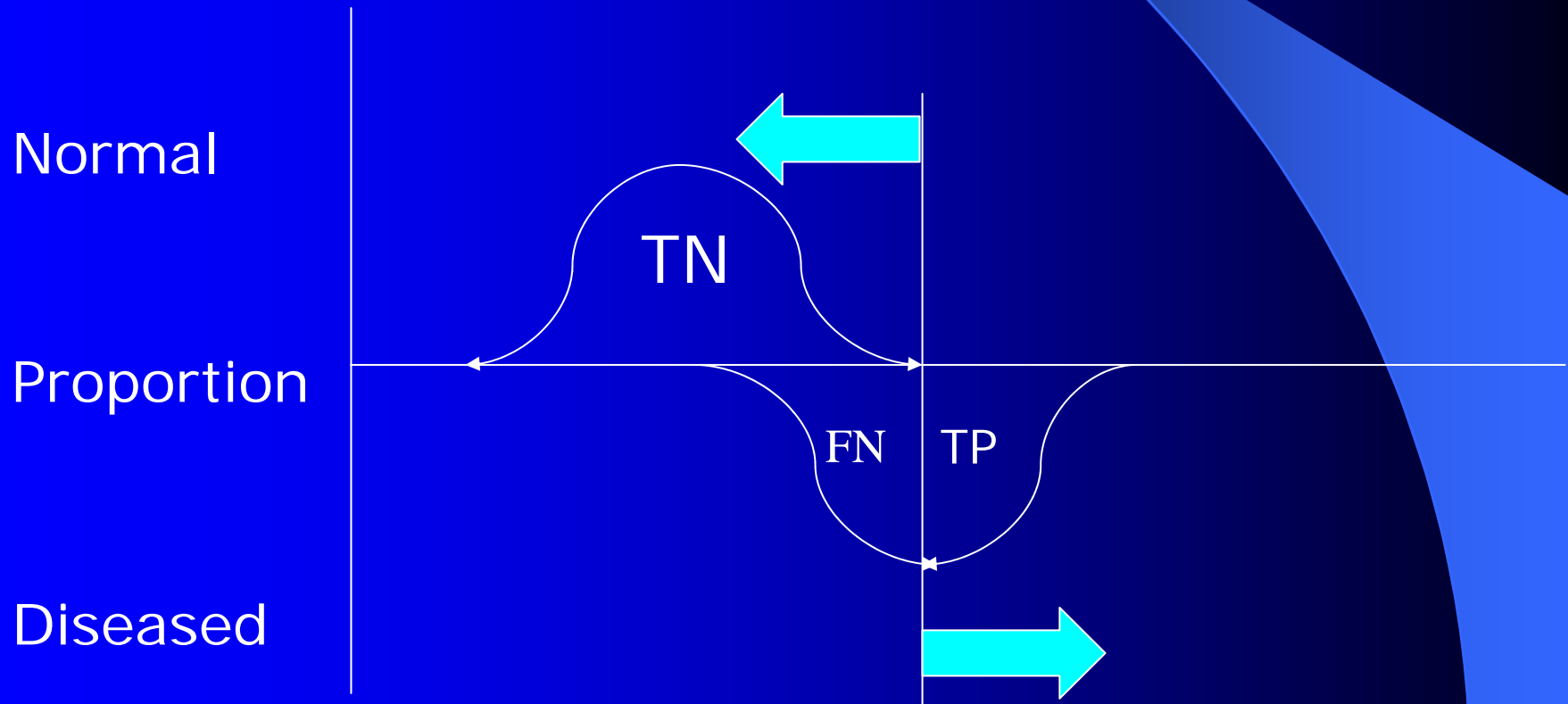
Disease Threshold



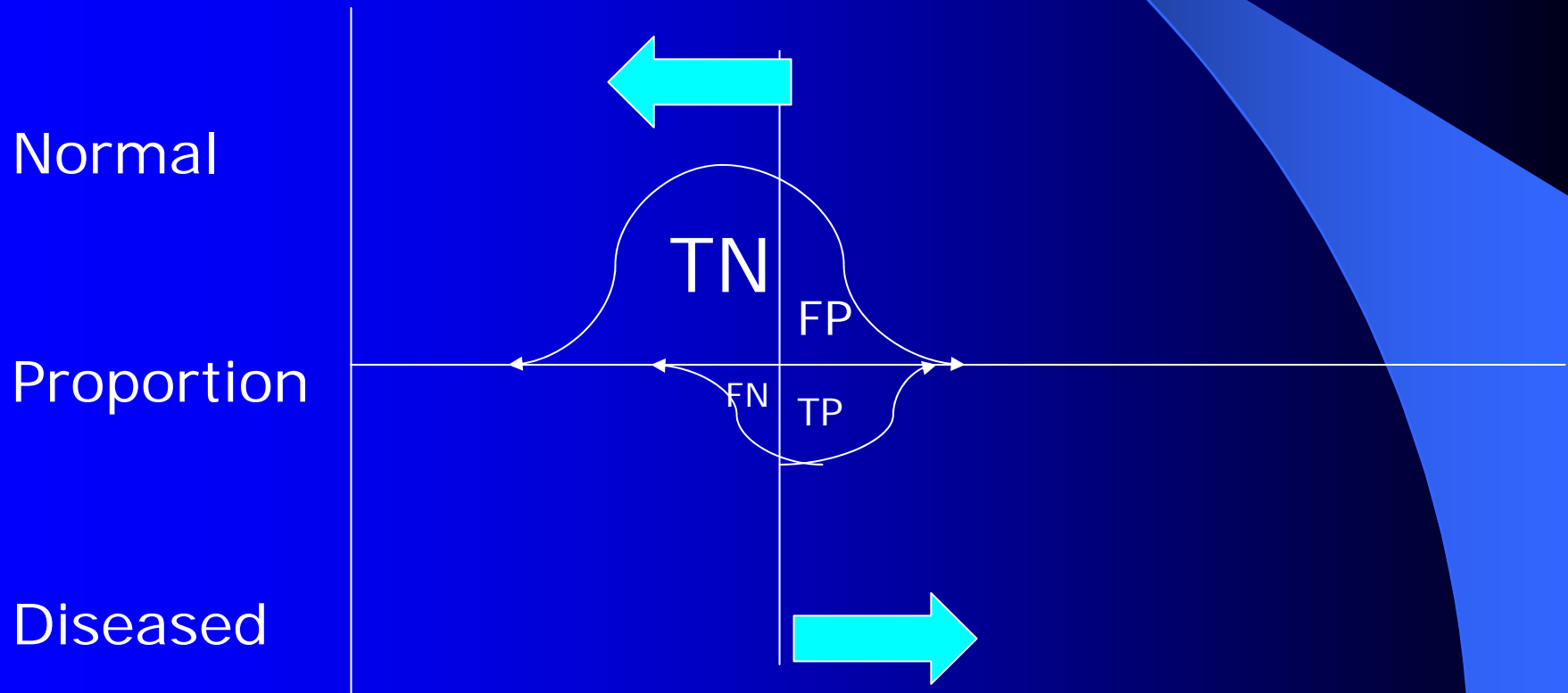
Disease Threshold



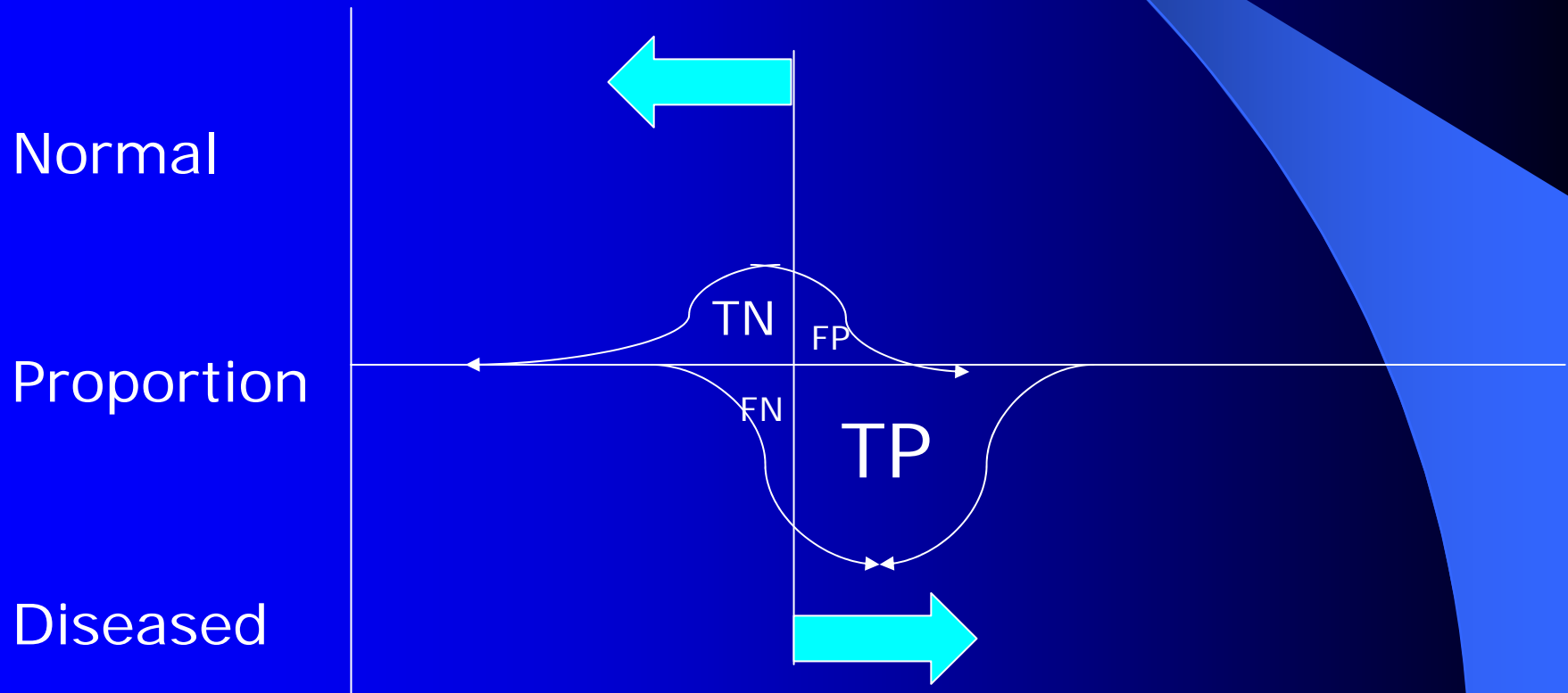
Disease Threshold



Disease Threshold



Disease Threshold



Example

Specificity/Sensitivity

	Disease Present (Gold Standard)	Disease Absent (Gold Standard)	
Positive Test	18	49	67
Negative Test	2	931	933
	20	980	1000

Sensitivity and Specificity

$$\begin{aligned}\text{Sensitivity} &= a/(a+c) = 18/20 \\ &= .9 \text{ or } 90\%\end{aligned}$$

$$\begin{aligned}\text{Specificity} &= d/(b+d) = \\ &931/980 \\ &= .95 \text{ or } 95\%\end{aligned}$$

Positive and Negative Predictive Value

	Disease Present (Gold Standard Truth?)	Disease Absent (Gold Standard Truth?)	
Positive Test	a-True Positive (TP)	b-False Positive (FP)	a+b
Negative Test	c-False Negative (FN)	d-True Negative (TN)	c+d
	a+c	b+d	a+b+c+d Population

Tests Predictive Value

- Positive Predictive Value (PPV)
- The proportion of those who test positive who do have the disease.
- $PPV = Pr(D+T+) = \frac{a}{a+b} = \frac{TP}{(TP+FP)}$
- Negative Predictive Value (NPV)
- The proportion who those who test negative who do not have the disease.
- $NPV = Pr(D-T-) = \frac{d}{c+d} = \frac{TN}{(TN+FN)}$

Example

Positive and Negative Predictive Value

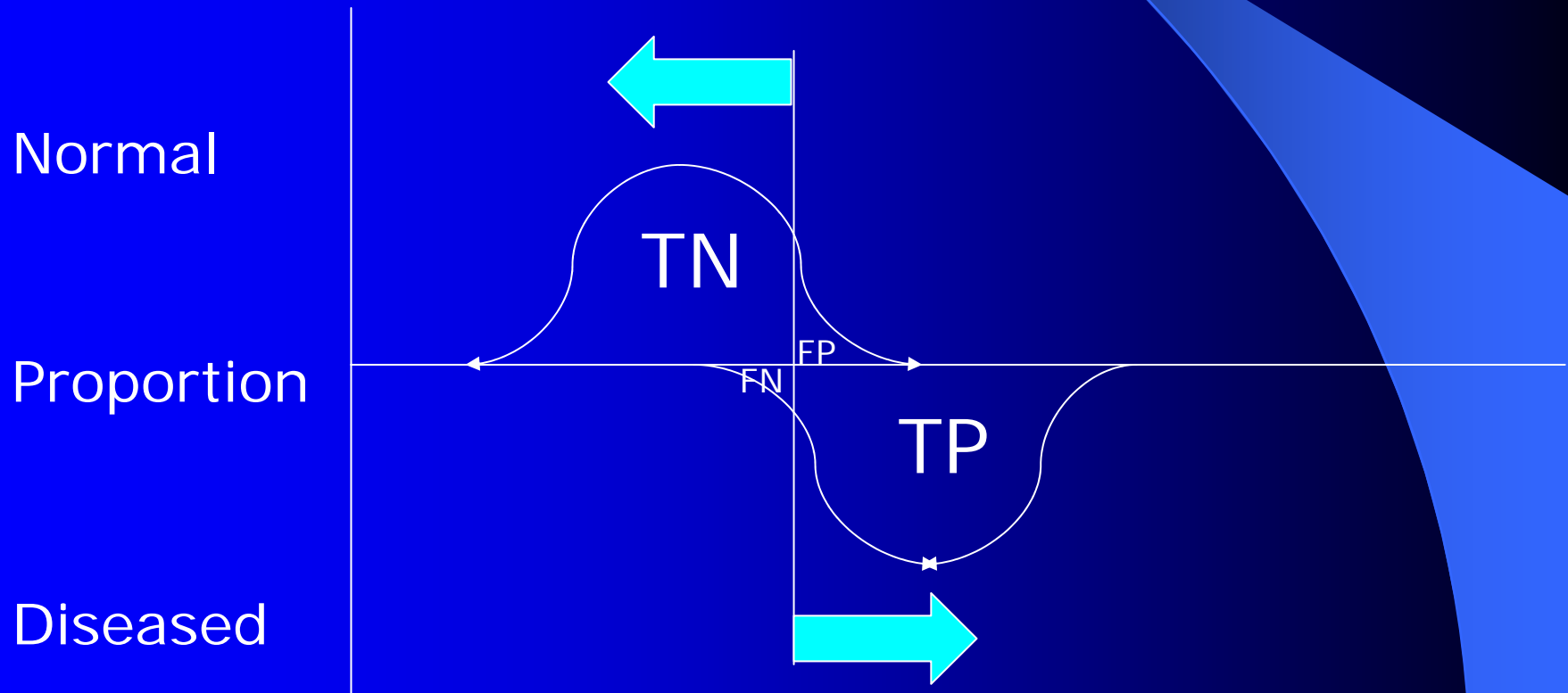
	Disease Present	Disease Absent	
Positive Test	18	49	67
Negative Test	2	931	933
	20	980	1000

Positive and Negative Predictive Value

$$\begin{aligned} \text{PPV} &= a/(a+b) = 18/67 \\ &= .27 \text{ or } 27\% \end{aligned}$$

$$\begin{aligned} \text{NPV} &= d/(c+d) = 931/933 \\ &= .998 \text{ or } 99.8\% \end{aligned}$$

Disease Threshold



Example 1

Population = N = 100,000

Prevalence = 0.50 or 50%

Diseased
50,000

Nondiseased
50,000

Test 1

Sens. = 0.85

Spec. = 0.95

True Positives	False Negatives	True Negatives	False Positives
42500	7500	47500	2500

$PPV = TP / \text{Total Positives} = 0.9444$

$NPV = TN / \text{Total Negatives} = 0.8636$

Example 2

Population = N = 100,000

Prevalence = 0.05 or 5%

Diseased
5,000

Nondiseased
95,000

Test 1

Sens.= 0.85

Spec. = 0.95

True Positives	False Negatives	True Negatives	False Positives
4250	750	90250	4750

PPV = TP/Total Positives = 0.4722 (prev = .9444)

NPV = TN/Total Negatives = 0.9918 (prev.=.8636)

Example 3

Population = N = 100,000

Prevalence = 0.01 or 1%

Diseased
1000

Nondiseased
99000

Test 1

Sens. = 0.85

Spec. = 0.95

True Positives	False Negatives	True Negatives	False Positives
850	150	94050	4950

PPV = TP/Total Positives = 0.1466 (init. = .9444)

NPV = TN/Total Negatives = 0.9984 (init. = .8636)

Example 4

Population = N = 100,000

Prevalence = 0.0001 or 0.01%

Diseased
12

Nondiseased
99988

Test 1

Sens. = 0.995

Spec. = 0.996

True Positives	False Negatives	True Negatives	False Positives
12	0	99588	400

PPV = TP/Total Positives = 0.029 (init. = .9444)

NPV = TN/Total Negatives = 1 (init. = .8636)

Example 5

Population = N = 100,000

Prevalence = 0.0001 or 0.01%

Diseased
12

Nondiseased
99988

Test 1

Sens. = 0.955

Spec. = 0.907

True Positives	False Negatives	True Negatives	False Positives
11	1	90689	9299

PPV = TP/Total Positives = 0.0012

NPV = TN/Total Negatives = 1

Effect of Prevalence

- Increased Prevalence
- Increased PPV
- Decrease NPV
- Increased false positives
- Decreased false negatives
- *SnOut*, NPV
- Decreased Prevalence
- Decreased PPV
- Increased NPV
- Decreased false positives
- Increased false negatives
- *SpIn*, PPV

Pre and Post Test probability

- Pre-test Probability = Prevalence
- Post-test Probability of a + test = Probability of having the disease after having a positive test
PPV.
- Post-test Probability of a - test = Probability of having the disease after having a negative test (1-NPV) =?
- Post-test Odds + = Pre-test Odds*LR+
- Post-test Odds - = Pre-test Odds*LR -

Likelihood Ratio LR

- A Probability Ratio like the Odds p/q or $p/(1-p)$ (ratio of the probability of success over the probability of failure, ratio of the probability of event occurring over the probability of the event not occurring).
- Pre-test Probability = Prevalence (relative)
- Post-test Probability = Pre-Test Odds * LR
- High LR = Strong Evidence to R/I Disease
- Low LR = Strong Evidence to R/O Disease

Odds and Odds Ratio

- $p/q = p/(1-p) = a/(a+b) / b/(a+b) = a/b$
- $p/q = p/(1-p) = c/(c+d) / d/(c+d) = c/d$
- $OR = a/b / c/d = ad/bc$
- $p/q = p/(1-p) = a/(a+c) / c/(a+c) = a/c$
- $p/q = p/(1-p) = b/(a+d) / d/(a+d) = b/d$
- $OR = a/c / b/d = ad/bc$

Odds and Odds Ratio

- One die
- $p = 1/6, q=5/6$
- $p/q = 1/6 / 5/6 = 1/5$
- $p/(1-p) = 1/6 / (1-1/6) = 1/6 / 5/6 = 1/5$
- $p = \text{odds} / 1+\text{odds} = 1/5 / (1+1/5) = 1/5 / 6/5 = 1/6$

Likelihood Ratio

Probability of patient *with* disease having a given test result

Probability of patient *without* disease having a given test result

Positive Likelihood Ratio (LR+)

Probability of patient with disease having a positive test result

Probability of patient without disease having a positive test result

Negative Likelihood Ratio (LR-)

Probability of patient with disease having a negative test result

Probability of patient without disease having a negative test result

Likelihood Ratios

- LR+

$$\frac{\text{sensitivity}}{1 - \text{specificity}}$$

TPR

FPR

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

$$1 - \frac{d}{b+d} = \frac{b}{b+d}$$

- LR-

$$\frac{1 - \text{sensitivity}}{\text{specificity}}$$

FNR

TNR

$$1 - \frac{a}{a+c} = \frac{c}{a+c}$$

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

Sensitivity Specificity and LR

- As sensitivity increases LR- decreases and NPV increases (SnOut)
- As specificity increases LR+ increases and PPV increases (SpIn)

Let's say you have 2 tests

You can do 2 tests to screen your patients:
FBS (sens 95%, spec 60%), HbA1C (sen 75%, spec 75%) and 3-hr OGTT (sen 60%, spec 95%). Which of the following would you use in which sequence?

A. HbA1c then FBS

C. OGTT then FBS

B. FBS then HbA1C

D. FBS then OGTT

Kappa statistic

- Percent agreement between independent observers solely as a function of chance.
- To what extent would they agree beyond that would be expected by chance alone?
- Kappa =
$$\frac{(\text{Percent agreement observed}) - (\text{Percent agreement expected by chance alone})}{100\% - (\text{Percent agreement expected by chance alone})}$$

Kappa $>.75$ = excellent agreement

Kappa $<.40$ = poor agreement

Observer 1

Positive

Negative

Positive

a

b

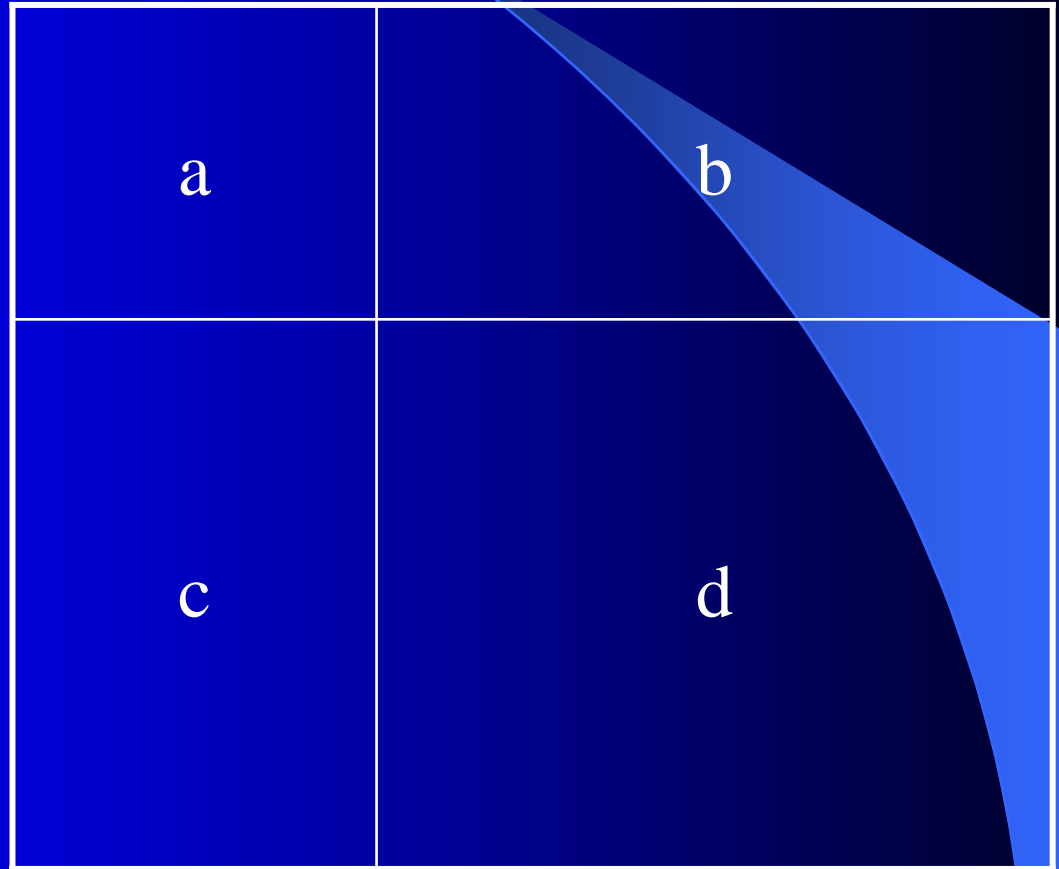
Observer 2

Negative

c

d

Percent agreement =
 $a + d / (a + b + c + d)$



Positive	a	b
Negative	c	d

Reject The Null

No

Yes

True

True Positive

False Positive
Type 1 Error
alpha

Null

False

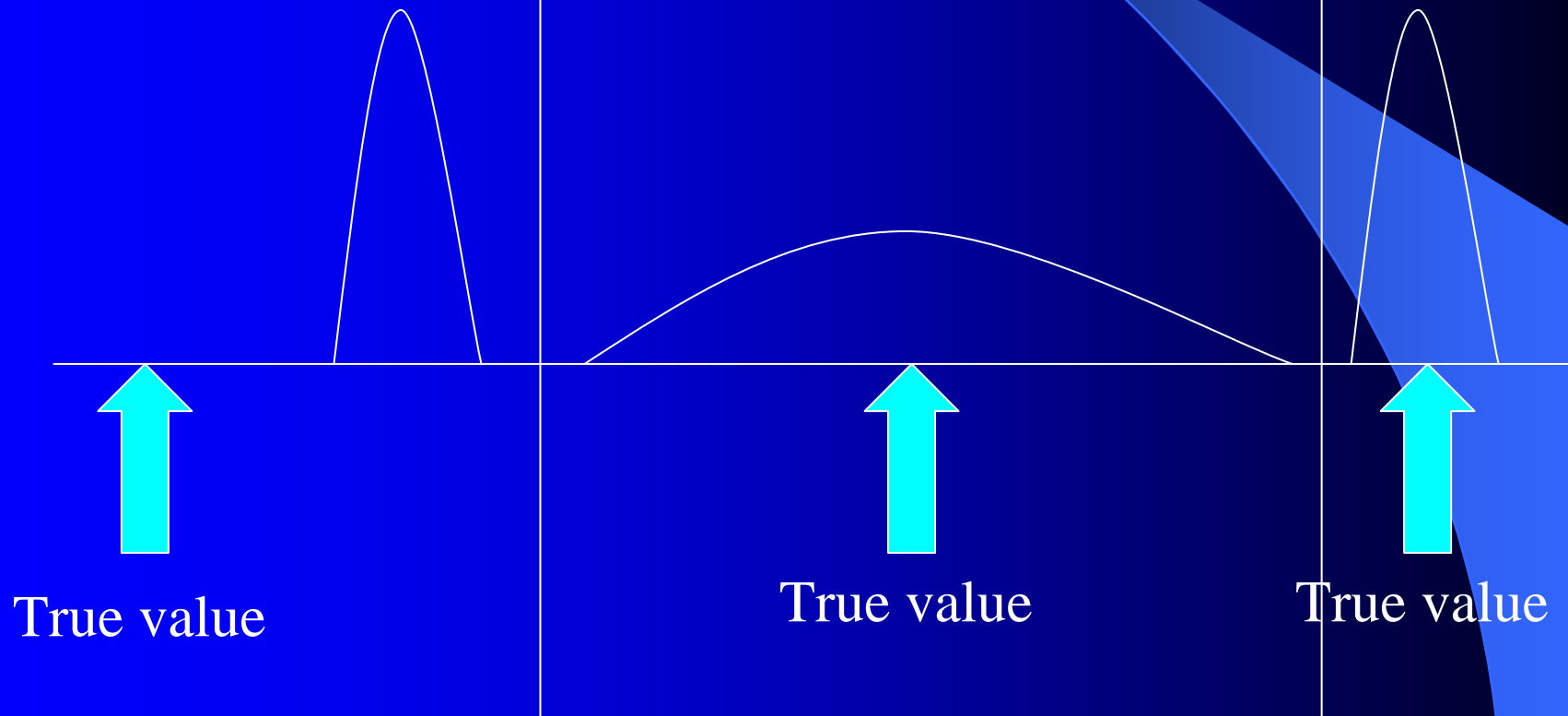
False
Negative
Type 2 Error
beta

True Negative

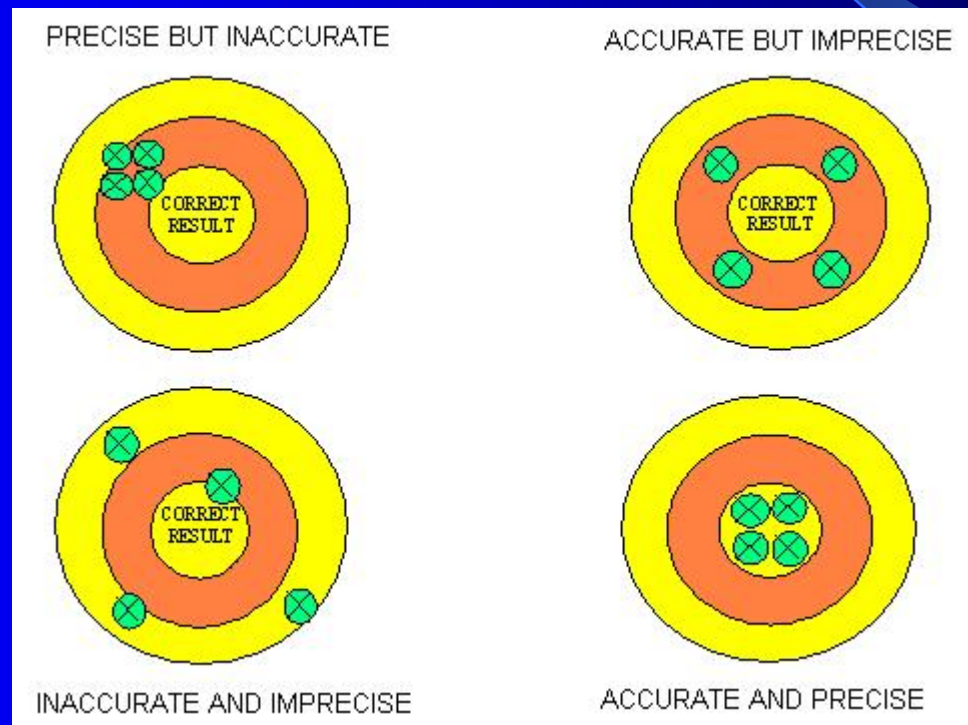
Calculate Kappa

- Percent observed agreement = $a+d / (a+b+c+d) = (41+27)/75 * 100 = 68/75 * 100 = 90.7\%$
- Percent agreement by chance alone = $26.4/75 + 12.4/75 = 38.8/75 = 51.7\%$
- Kappa = $(90.7-51.7) / (100-51.7) = 0.81$

Validity and Reliability



Accuracy and Precision



Examples Sens./Spec. USPSTF data

- Colonoscopy 75-90% sensitivity
- Mammography (<50) 51-83%, PPV = 2-4%
- Pap sens. = 85%, spec. = 60%,
NPV = 99%
- Exercise stress testing Sens=40-62%, PPV 6
- 48%, (- Sx), 40-80% (+Sx)

References

- Scholarly Research Paper Publication/Bibliography Software
- Agrawal A. EndNote 1-2-3 Easy! Reference Management For the Professional. 1st ed. 2005.
- Maran R. Microsoft Office 2000 Simplified. 1999.
- Maran R. Maran Illustrated Office 2003. 1st Ed. 2005.
- Supplementary and Advanced Level References
- Statistics/Biostatistics
- Glantz SA, Slinker BK. Primer of Applied Regression and Analysis of Variance. 2nd Ed. 2000.
- Kleinbaum DG, Kupper LL, Nizam A, Muller KE. Applied Regression Analysis and Multivariable Methods. 4th Ed.. 2007.
- Winer BJ, Brown DR, Michels KM. Statistical Principles in Experimental Design, 3rd Ed.
- Snedecor GW, Cochran WG. Statistical Methods. 8th Ed. 1989.
- Maxwell SE, Delaney HD. Designing Experiments and Analyzing Data: A Model Comparison Approach. 2nd Ed.
- Keppel G, Wickens TD. Design And Analysis. A Researchers Handbook. 4th Ed. 2004.
- McMahon D. Linear Algebra Demystified. 1st Ed. 2005.
- Lay DC. Linear Algebra and its Applications. 3rd Ed. 2005.
- Clark-Carter D. Quantitative Psychological Research: A Students Handbook. 2004.
- Russo R. Statistics for the Behavioral Sciences. 2003.
- Epidemiology/Research Methods
- Evans JS, Evans BT. How To Do Research. 2005.
- Boynton PM. The Research Companion. A Practical Guide for the Social and Health Sciences. 2005.
- Medical Health Informatics
- Englehardt SP. Health Care Informatics: An Interdisciplinary Approach

References

- **Epidemiology/Research Methods**

- Gehlbach SH. Interpreting the Medical Literature. Practical Epidemiology For Clinicians. 5th Ed. 2006.
- Gordis L. Epidemiology. 3rd Ed. 2004.
- Hulley SB, Designing Clinical Research. An Epidemiologic Approach. 3rd Ed. 2006.

- **Dataset/Database Management**

- Maran R. Microsoft Office 2000 Simplified. 1999.

- Maran R. Maran Illustrated Office 2003. 1st Ed. 2005.
- Maran R. Maran Illustrated Access 2003. 1st Ed. 2005.
- Maran R. Maran illustrated Excel 2003. 1st Ed. 2005.
- George D, Mallory P. SPSS For Windows: Step-by-Step. 7th Ed. 2006
- Hinton PR, Brownlow C, McMurray I. et. al. SPSS Explained. 1st Ed. 2004.
- Delwiche LD, Slaughter SJ. The Little SAS Book: A Primer. 3rd Ed. 2003
- Acock AC. A Gentle Introduction to Stata. 1st. Ed. 2005.

- **Scholarly Research Paper Publication/Bibliography Software**

- Agrawal A. EndNote 1-2-3 Easy! Reference Management For the Professional. 1st ed. 2005.
- Maran R. Microsoft Office 2000 Simplified. 1999.
- Maran R. Maran Illustrated Office 2003. 1st Ed. 2005.

References

- **Supplementary and Advanced Level References**

- **Statistics/Biostatistics**

- Glantz SA, Slinker BK. Primer of Applied Regression and Analysis of Variance. 2nd Ed. 2000.
- Kleinbaum DG, Kupper LL, Nizam A, Muller KE. Applied Regression Analysis and Multivariable Methods. 4th Ed.. 2007.
- Winer BJ, Brown DR, Michels KM. Statistical Principles in Experimental Design, 3rd Ed.
- Snedecor GW, Cochran WG. Statistical Methods. 8th Ed. 1989.
- Maxwell SE, Delaney HD. Designing Experiments and Analyzing Data: A Model Comparison Approach. 2nd Ed.
- Keppel G, Wickens TD. Design And Analysis. A Researchers Handbook. 4th Ed. 2004.
- McMahon D. Linear Algebra Demystified. 1st Ed. 2005.
- Lay DC. Linear Algebra and its Applications. 3rd Ed. 2005.
- Clark-Carter D. Quantitative Psychological Research: A Students Handbook. 2004.
- Russo R. Statistics for the Behavioral Sciences. 2003.

- **Epidemiology/Research Methods**

- (1)Evans JS, Evans BT. How To Do Research. 2005.
- (2) Boynton PM. The Research Companion. A Practical Guide for the Social and Health Sciences. 2005.

- **Medical Health Informatics**

- Englehardt SP. Health Care Informatics: An Interdisciplinary Approach
- Medical Informatics: Knowledge Management and Data Mining in Biomedicine

References

– Useful WWW Online Resources

- Rice University Virtual Lab in Statistics (online multimedia tutorial and textbook) www.onlinestatbook.com/rvls/
- UCLA Statistical Computing Online Tutorial on SAS, STATA, and SPSS. www.ats.ucla.edu/stat/overview.htm
- Practice Datasets www.vetmed.wsu.edu/appliedregression/

– Video Instruction Resources

- Against All Odds: Inside Statistics www.learner.org/resources/series65.html
- Statistics www.videoaidedinstruction.com/