The ROC curve as a measure of diagnostic accuracy

Center for Statistical Sciences
Brown University

Outline

- **Sensitivity and specificity**
- **Positive and negative predictive values**
- **ROC curves**
- **Examples**

The Setup

- **Disease status: present or absent**
- **Test results: Positive or negative**
  - Ordinal categorical or continuous test can be made binary by selecting a threshold
- **The 2x2 table:**

Truth	Test -	Test +	Total
Disease-free	TN	FP	N_-
Disease	FN	TP	N_+
Total	TN+FN	FP+TP	N

Sensitivity and Specificity

- **Assume disease status is known**
- **Sensitivity is the probability that test result will be *positive* given that disease is *present*.**
  - $P(T+|D+) = TP / N_+ = TP / (TP+FN)$
- **Specificity is the probability that test result will be *negative* given that disease is *absent*.**
  - $P(T-|D-) = TN / N_- = TN / (TN+FP)$

Truth	Test -	Test +	Total
Disease-free	TN	FP	N_-
Disease	FN	TP	N_+
Total	T_-	T_+	N

Positive and Negative Predictive value

- **Assume test result is known**
- **Positive predictive value is the probability that disease is *present* given that the test result is *positive*.**
  - $P(D+|T+) = TP / T_+ = TP / (TP+FP)$
- **Negative predictive value is the probability that the disease is *absent* given that the test result is *negative*.**
  - $P(D-|T-) = TN / T_- = TN / (TN+FN)$

Truth	Test -	Test +	Total
Disease-free	TN	FP	N_-
Disease	FN	TP	N_+
Total	T_-	T_+	N

Factoid

- **If you are a patient about to take a diagnostic test, you want to know:**
  - Positive predictive value: $P(D+|T+)$
  - Negative predictive value: $P(D-|T-)$

- **If you are a researcher evaluating a new diagnostic test, you want to know:**
 - Sensitivity: $P(T+|D+)$
  - Specificity: $P(T-|D-)$

An Important Relationship

- **If we know the sens, spec and the disease prevalence (say π), then we can always find the PPV and NPV**

- **PPV = sens*π / (sens*π + (1-spec)*(1- π))**
- **NPV = spec*(1- π) / (spec*(1-π) + (1-sens)*π)**

- **So most studies are designed to estimate sensitivity and specificity.**

Positive Predictive Value | Negative Predictive Value

Sensitivity = 0.9 ; Specificity=0.9
Sensitivity = 0.5 ; Specificity=0.5
Sensitivity = 0.25 ; Specificity=0.25
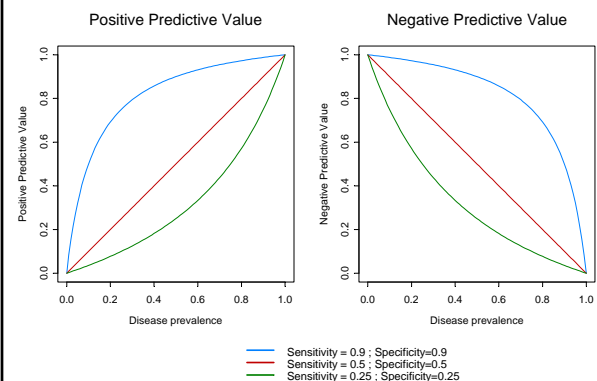
Beyond the Binary Test

- **Most tests are not binary; they are continuous or ordinal categorical**
- **But suppose we test positive only if the test score is above some threshold.**
- **Then, for that threshold, the test is considered binary and we can report sensitivity and specificity.**
- **But as we change the threshold, sensitivity and specificity change as well.**
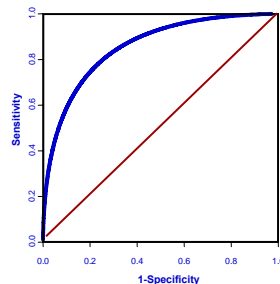
Diseased and Non-Diseased populations
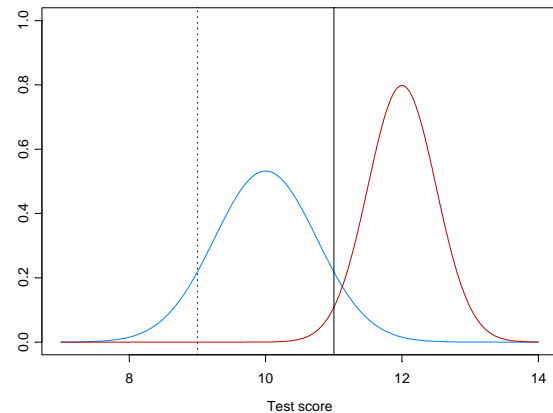
ROC curves

- **ROC curve:**
  - **Plots of all pairs (1-Spec, Sens) as threshold varies**
  - **Anchored at (0,0) and (1,1)**
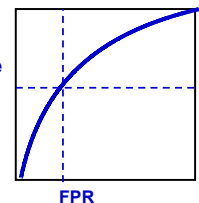 - **Always in top triangle**
  - **Describes the test accuracy at all thresholds**

ROC curves (cont.)

- **ROC curve is generalization of sens/spec, so it is conditional on disease status.**
- **ROC summaries:**
  - **Area under ROC curve**
  - **Partial area under ROC curve**
  - **Sens at given Spec**
  - **Optimal operating point**

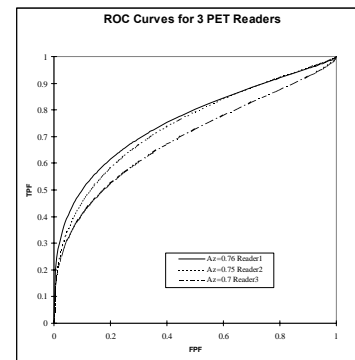## Area under ROC curve (AUC)

- AUC is the most common summary measure
- "Average sensitivity" over all specificity
- "Average specificity" over all sensitivity

- **2AFC (two alternative, forced choice experiment) interpretation**: AUC equals the probability that a pair of a diseased and non-diseased cases are ordered correctly.

- ROC area = 1 for a perfect test
- ROC area = 0.5 for a non-informative test

February 02   Blume and Gatsonis

13

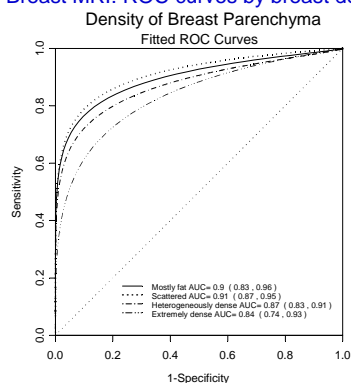PET for axillary node involvement: ROC curves of 3 radiologists

ROC Curves for 3 PET Readers

Az=0.76 Reader1
Az=0.75 Reader2
Az=0.7 Reader3

February 02   Blume and Gatsonis

14

Breast MRI: ROC curves by breast density

Density of Breast Parenchyma

Fitted ROC Curves

Mostly fat AUC= 0.9 (0.83 , 0.96)
Scattered AUC= 0.91 (0.87 , 0.95)
Heterogeneously dense AUC= 0.87 (0.83 , 0.91)
Extremely dense AUC= 0.84 (0.74 , 0.93)

February 02   Blume and Gatsonis

15

CT data- ROC

- Data from Hanley and McNeil, 1982
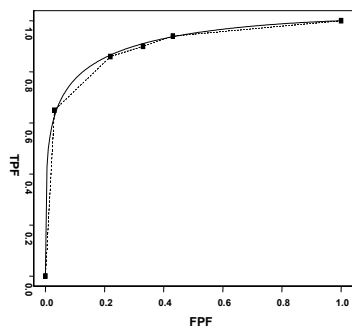- 5-point usual "degree of suspicion" scale, 1="definitely normal",...., 5="definitely abnormal"

Truth	Response category						Total
	1	2	3	4	5		
No disease	33	6	6	11	2		58
FPF	1.00	0.43	0.33	0.22	0.034		
Disease	3	2	2	11	33		51
TPF	1.00	0.94	0.90	0.86	0.65		

February 02   Blume and Gatsonis

16

4

Smooth ROC curve for CT data
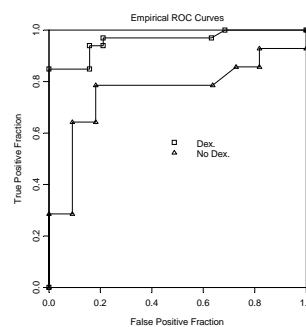
Progesterone data (cont.)

ROC curve for progesterone data
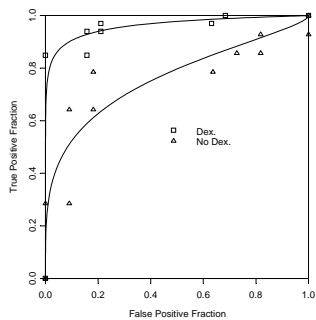
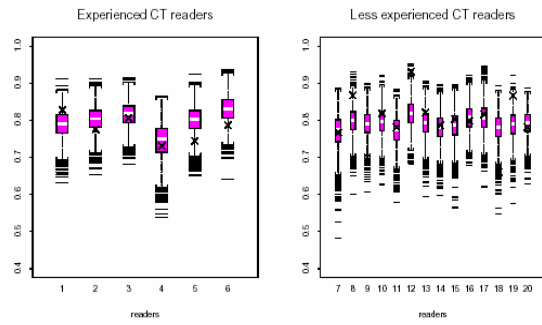## Variability across readers: RDOG data on head and neck cancer

- **38 radiologists interpreted CT and MRI scans on head and neck cancer patients.**
- **Each case was interpreted by 3 readers in each modality. Total of 20 CT readers, 18 MRI readers.**
- **Degree of suspicion about metastasis recorded on 5 point ordinal categorical scale.**
- **Reader AUC presented in next two graphs.**
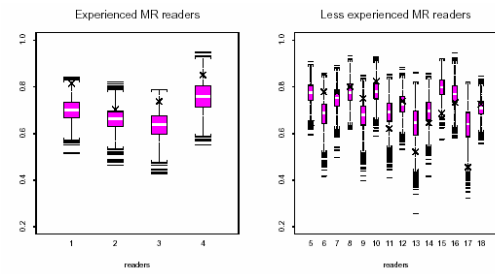
Head and neck cancer data (cont.)

Experienced CT readers | Less experienced CT readers

Head and neck cancer data (cont.)

Experienced MR readers | Less experienced MR readers