Statistics for Classification and Predictive Tests in Medicine

By Sephora Zoro ☺

What is a Test???

- Assesses the disease status of a patient
- Concerned with dichotomous conditions
  - Disease status – D
  - Test Result – Y
- Binary, Ordinal, and Continuous
- Dichotomous tests may be constructed from a continuum of biomarkers with threshold selection
- Test is *positive* if Y>= c, *negative* if Y<c
- Gold-Standard Tests will tell you whether a patient is ill or not without uncertainty

Errors and Positive Fractions

Type I – False Negative

Type II – False Positive

- Errors are present in non-ideal tests
  - False Positive Fraction: $P[Y = 1|D = 0]$
    Type I Error
  - True Positive Fraction: $P[Y = 1|D = 1]$
    1 – Type II
- An ideal diagnostic test has FPF = 0, TPF = 1
- A useless test has FPF=TPF

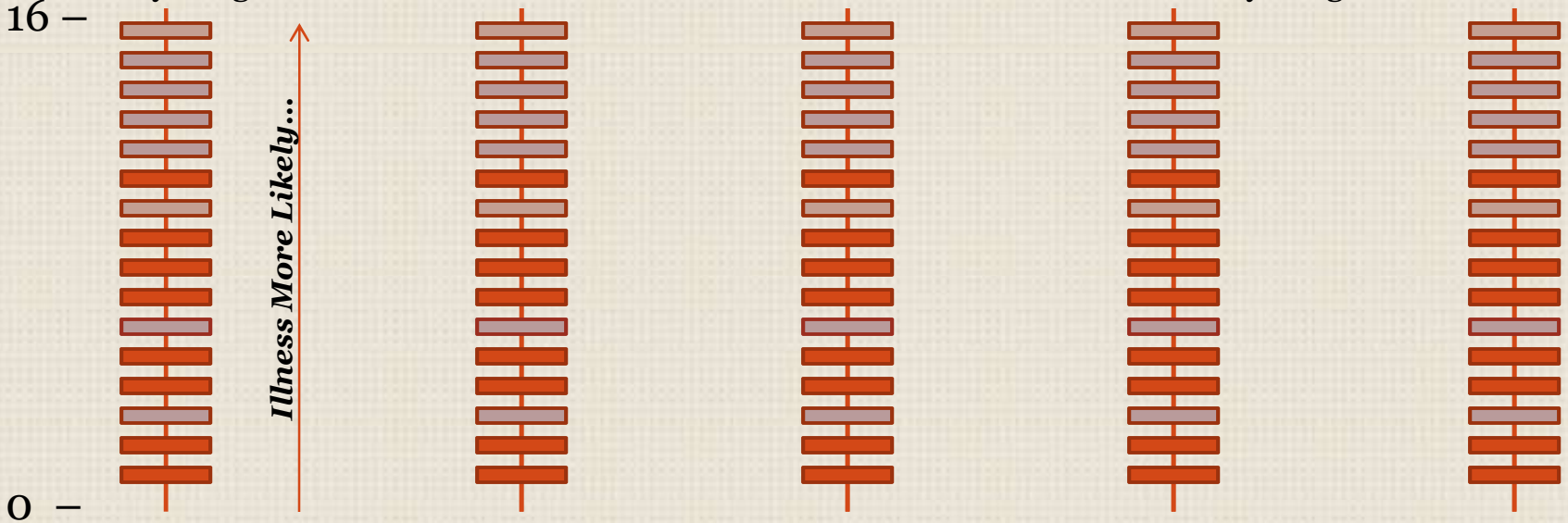
The Receiver Operating Characteristic Curve

The ROC Curve is the set of all possible true and false positive fractions found by transforming a continuous biomarker Y into a dichotomous result, using different thresholds!

The Receiver Operating Characteristic Curve

Biomarker A:

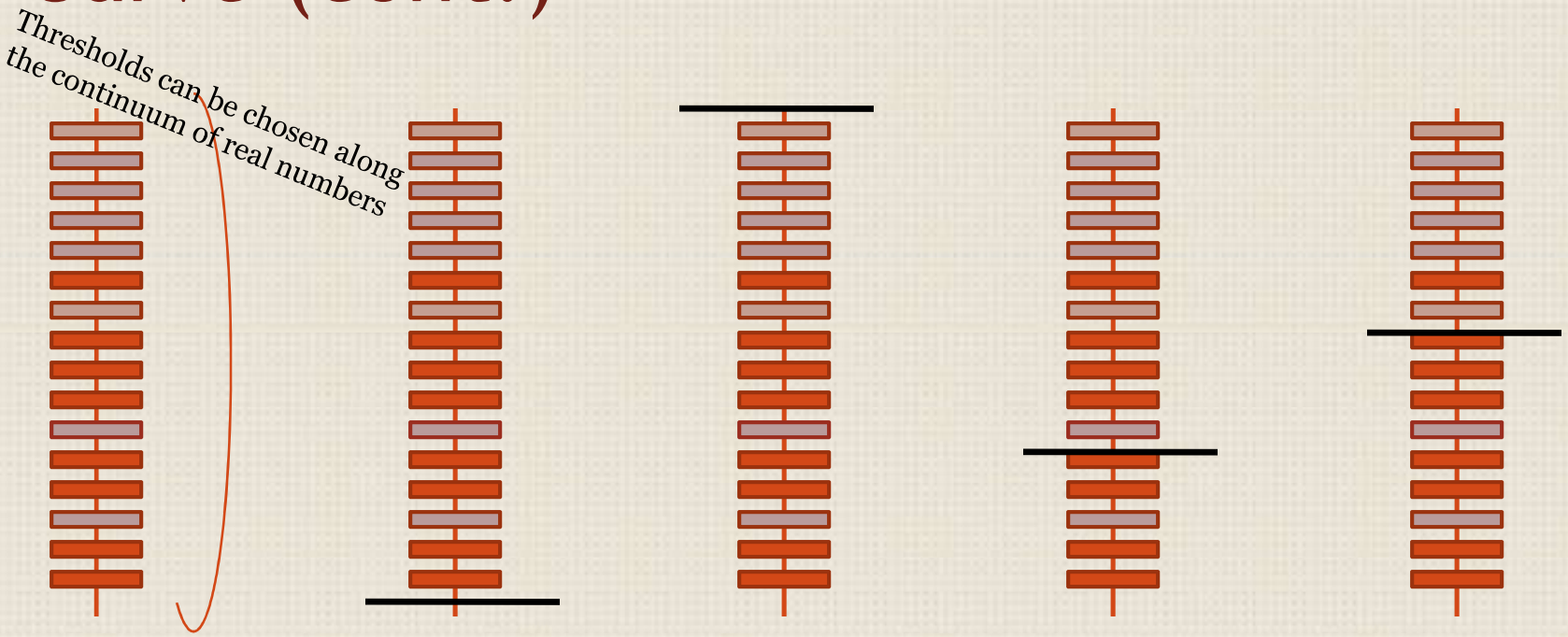
- 0-16
- Continuous
- Purple is ill, red is not ill S1
- Everything below threshold (horizontal line) is **NEGATIVE** test result. Everything at or above is **POSITIVE**



Slide 5

S1 change colors
Sephora, 3/7/2023

The Receiver Operating Characteristic Curve (cont.)



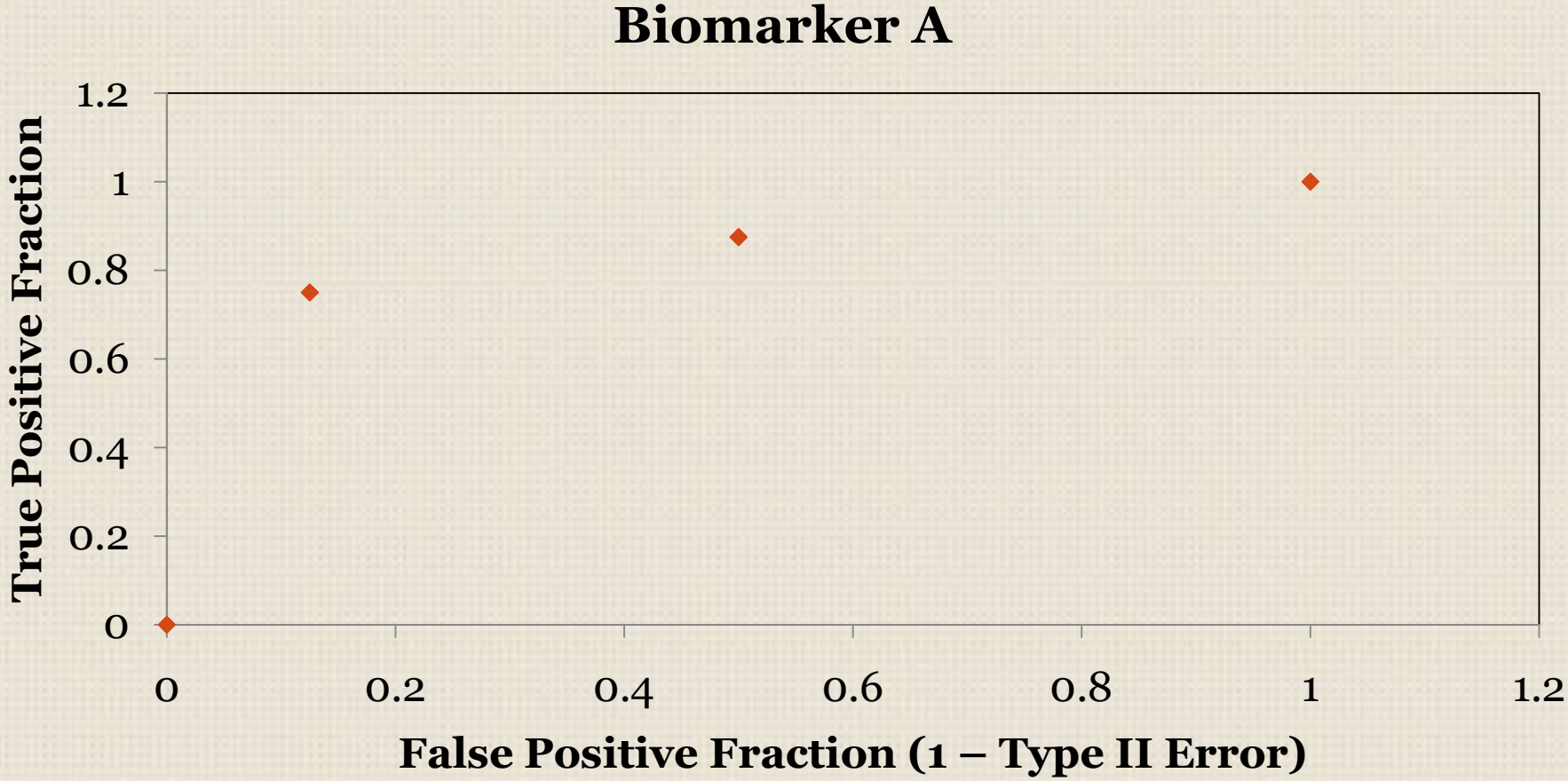
c=0;
TPF = 1.0
FPF = 1.0

c=16;
TPF =0.0
FPF = 0.0

c=5;
TPF = 0.875
FPF = 0.5

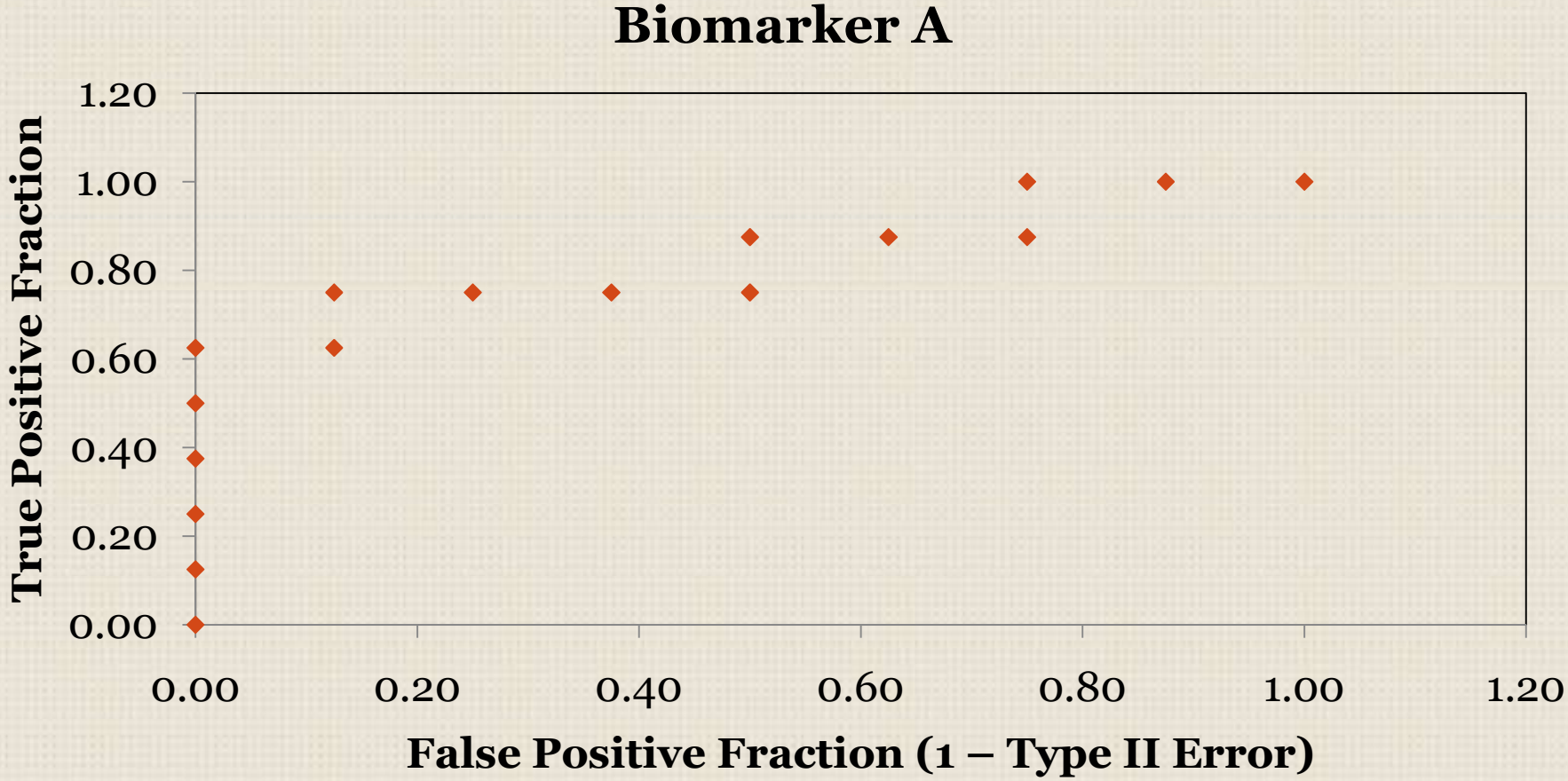
c=9;
TPF = 0.75
FPF = 0.125

The Receiver Operating Characteristic Curve (cont.)



THRESHOLD	TRUE POSITIVE FRACTION	FALSE POSITIVE FRACTION
0	8/8	8/8
1	8/8	7/8
2	7/8	6/8
3	7/8	6/8
4	7/8	5/8
5	7/8	4/8
6	6/8	4/8
7	6/8	3/8
8	6/8	2/8
9	6/8	1/8
10	5/8	1/8
11	5/8	0/8
12	4/8	0/8
13	3/8	0/8
14	2/8	0/8
15	1/8	0/8
16	0/8	0/8

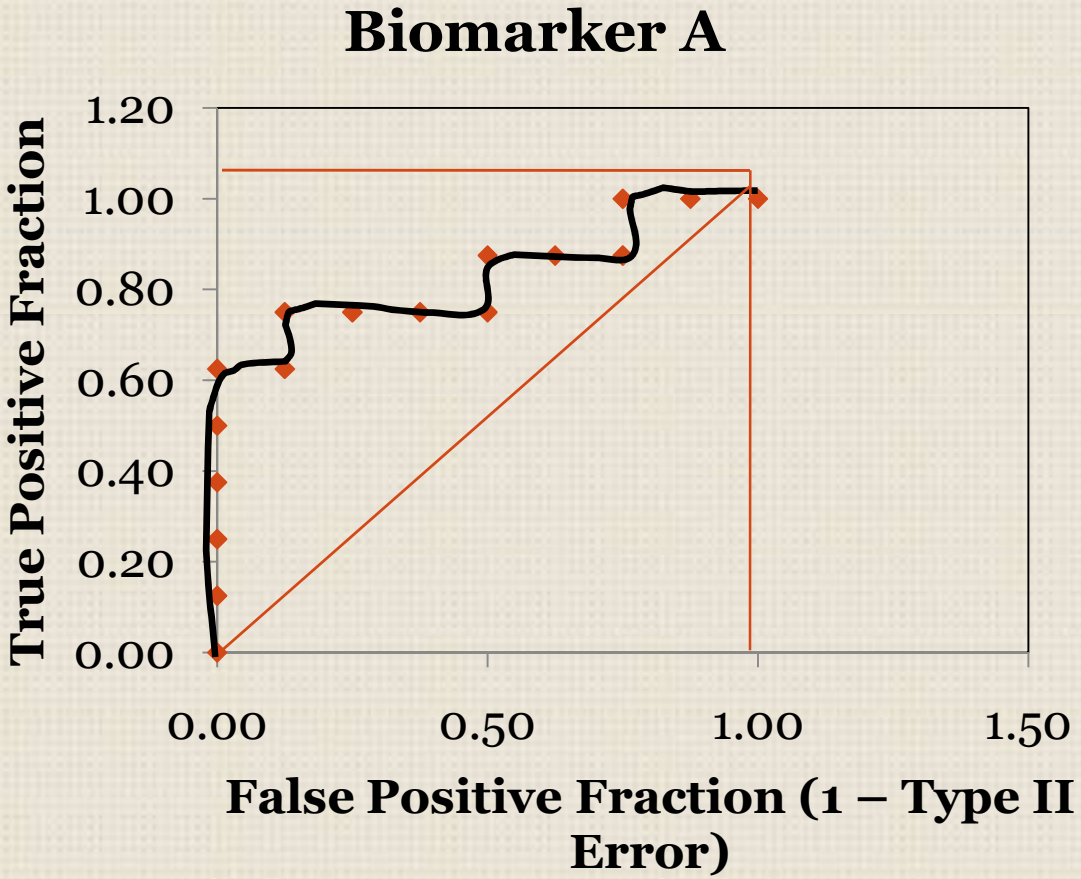
The Receiver Operating Characteristic Curve (cont.)

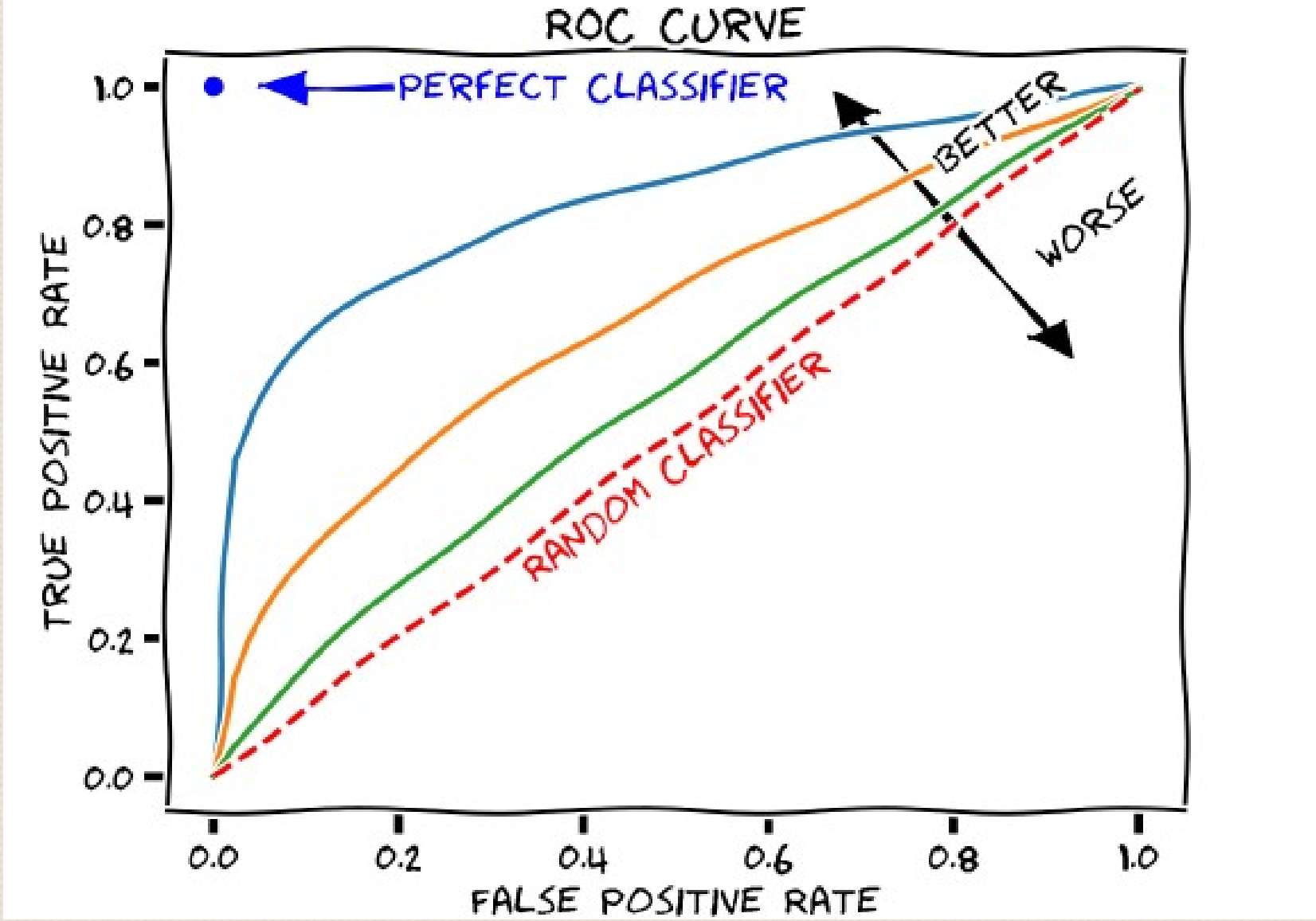


The Receiver Operating Characteristic Curve (cont.)

Threshold Selection:

-Minimizing Type I and Type II error
-Weighting errors
-Minimum TPF/FPF




ROC CURVE
PERFECT CLASSIFIER
BETTER
WORSE
RANDOM CLASSIFIER
TRUE POSITIVE RATE
FALSE POSITIVE RATE
0.0
0.2
0.4
0.6
0.8
1.0

Example: MRI Imaging of Dementia Patients

Cross Sectional Collection

- 416 adults 18-96 yo.
- 100 patients 60+ diagnosed with Alzheimer's
- Gold Standard – CDR classification

MRI Imaging of Dementia Patients

Characteristics of each patient are collected before over study course:

Clinical Dementia Rating (CDR)

Mini Mental State Exam (MMSE)

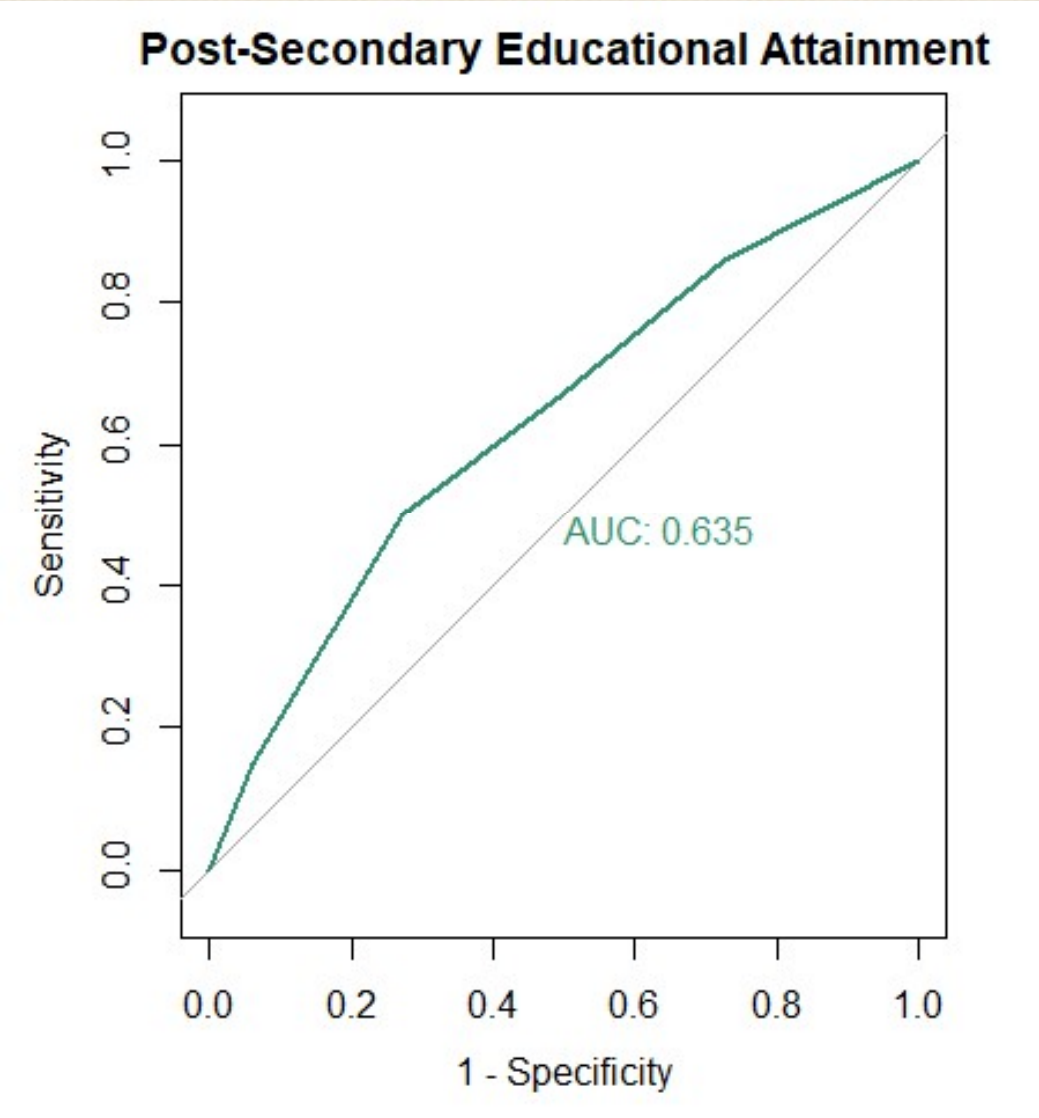
Normalized Whole Brain Volume (nWBV)

Estimated Total Intracranial Volume (eTIV)

Educational Attainment (Educ)

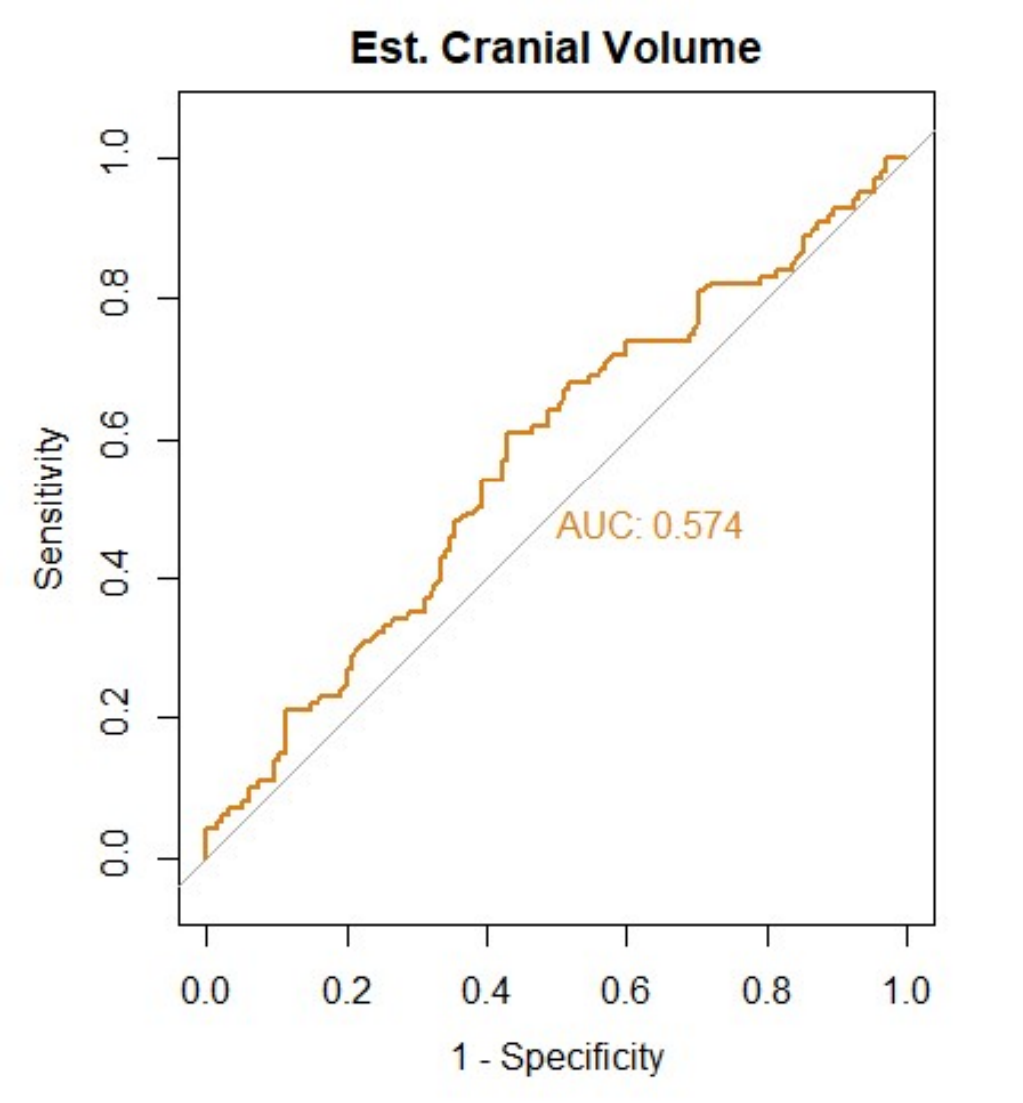
Classification Test

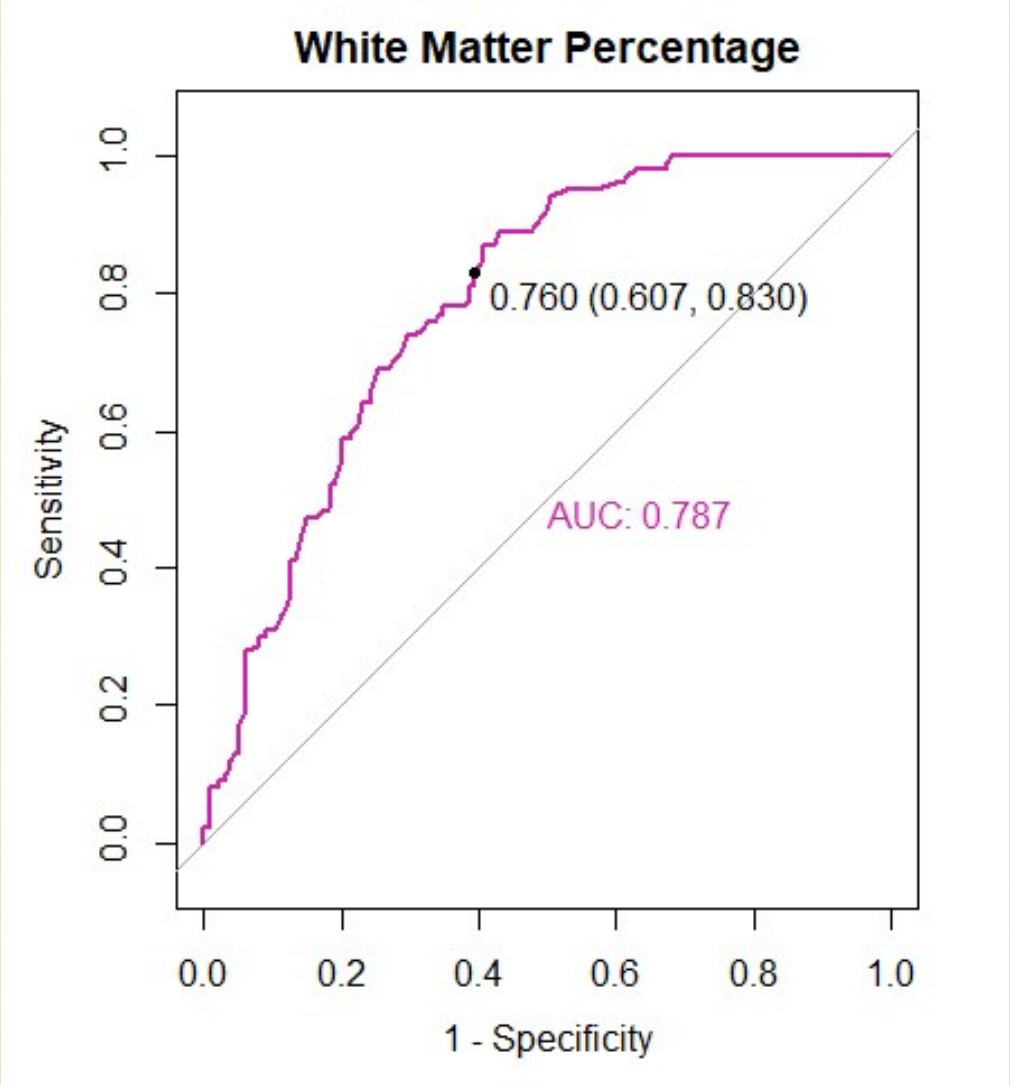
- Educational Attainment
- Cranium Volume
- White Matter Percentage
- Mini Mental State Exam



* Low AUC– operates close to a random classifierr

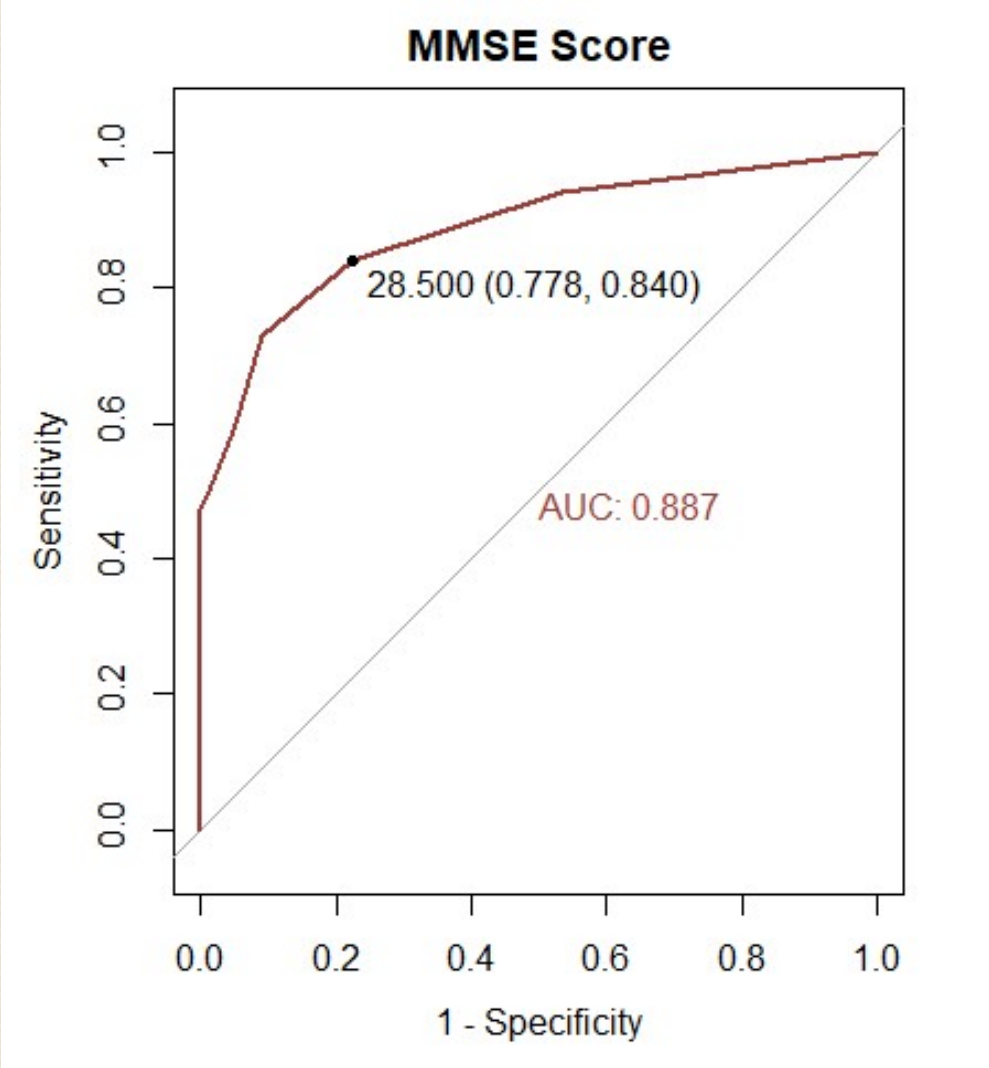
* AUC almost 0.5 – essentially a random classifier



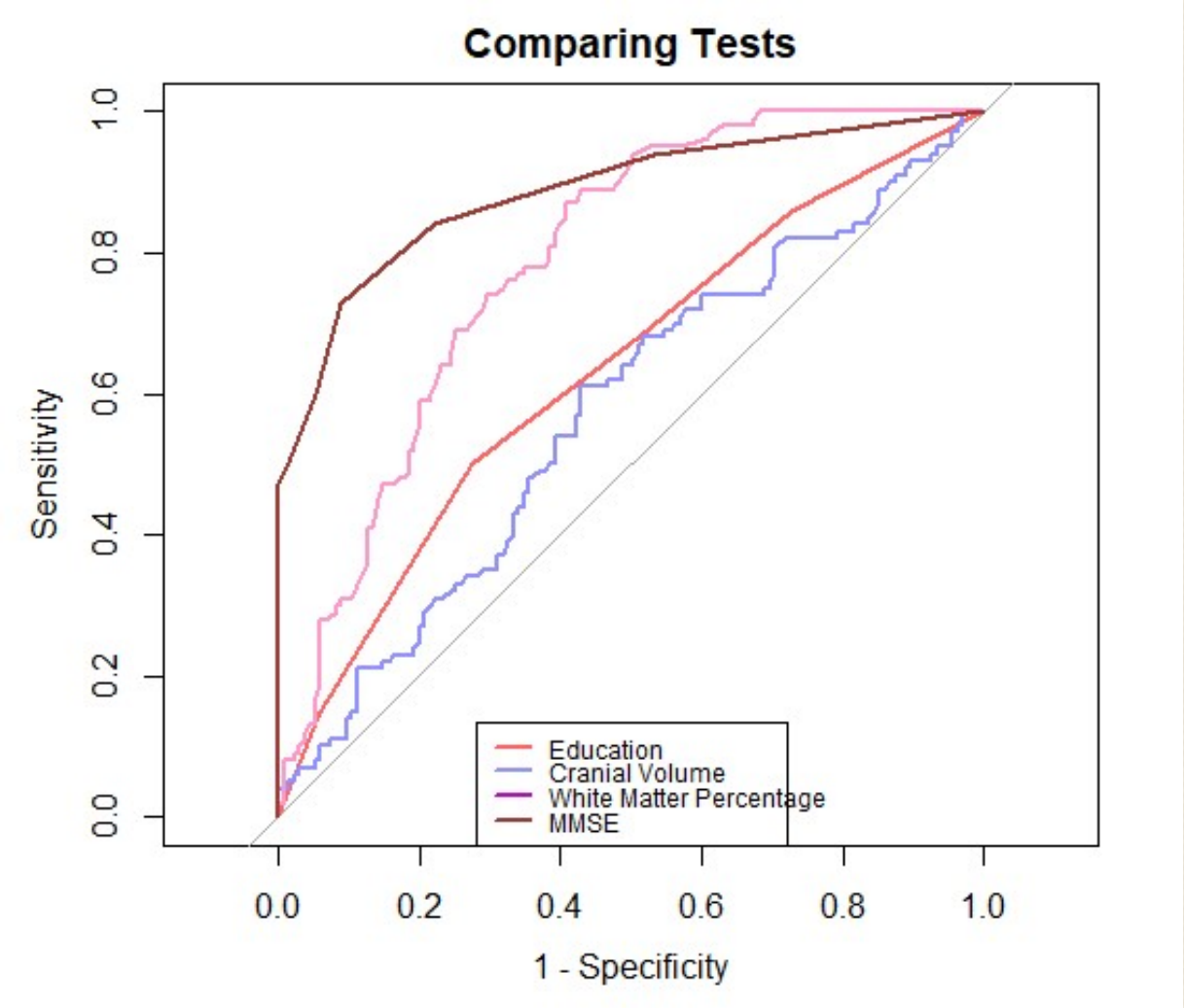


* Higher AUC, closer to perfect classifier than to random classifier
* Sensitivity of 60%, Specificity of 83%

*High AUC;
* TPF and (1-FPF) Higher than previously at selected threshold



Step 1:


Comparing Tests
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity
Education
Cranial Volume
White Matter Percentage
MMSE

Step 2:

- White Matter Percentage (roc3.info, roc1)
- MMSE (roc4.info, roc2)

```
data:  roc3.info and roc4.info
Z = -2.9646, p-value = 0.003031
alternative hypothesis: true difference in AUC is not equal to 0
95 percent confidence interval:
 -0.16475866 -0.03361171
sample estimates:
AUC of roc1 AUC of roc2
  0.7874444   0.8866296
```

Limitations and Summary

- In practice, there is some error in some Gold Standard tests
  - e.g., CDR is *not* gold standard for Alzheimer's diagnosis
- ROC Curves are used to display performance of tests compared to gold standard
- Test usefullness can be quanitified with AUC
  - Close to 1.0, more useful test
- Thresholds for tests can be chosen with different error weights, or specifying a preferred TPF/FPF