ROC Analysis

Sahir Bhatnagar, PhD
Assistant Professor
Department of Diagnostic Radiology
Department of Epidemiology, Biostatistics and Occupational Health
sahir.bhatnagar@mcgill.ca

Sensitivity
Specificity
Positive Predictive Value
Negative Predictive Value

Characteristics of a Good Screening Test

- Inexpensive
- Easy to administer
- Minimal discomfort
- Sensitive --> correctly identifies true **disease** cases
- Specific --> correctly identifies true **non-disease** cases

....is a well person

....is a person with a disease

....is a negative test result

....is a positive test result

and therefore....

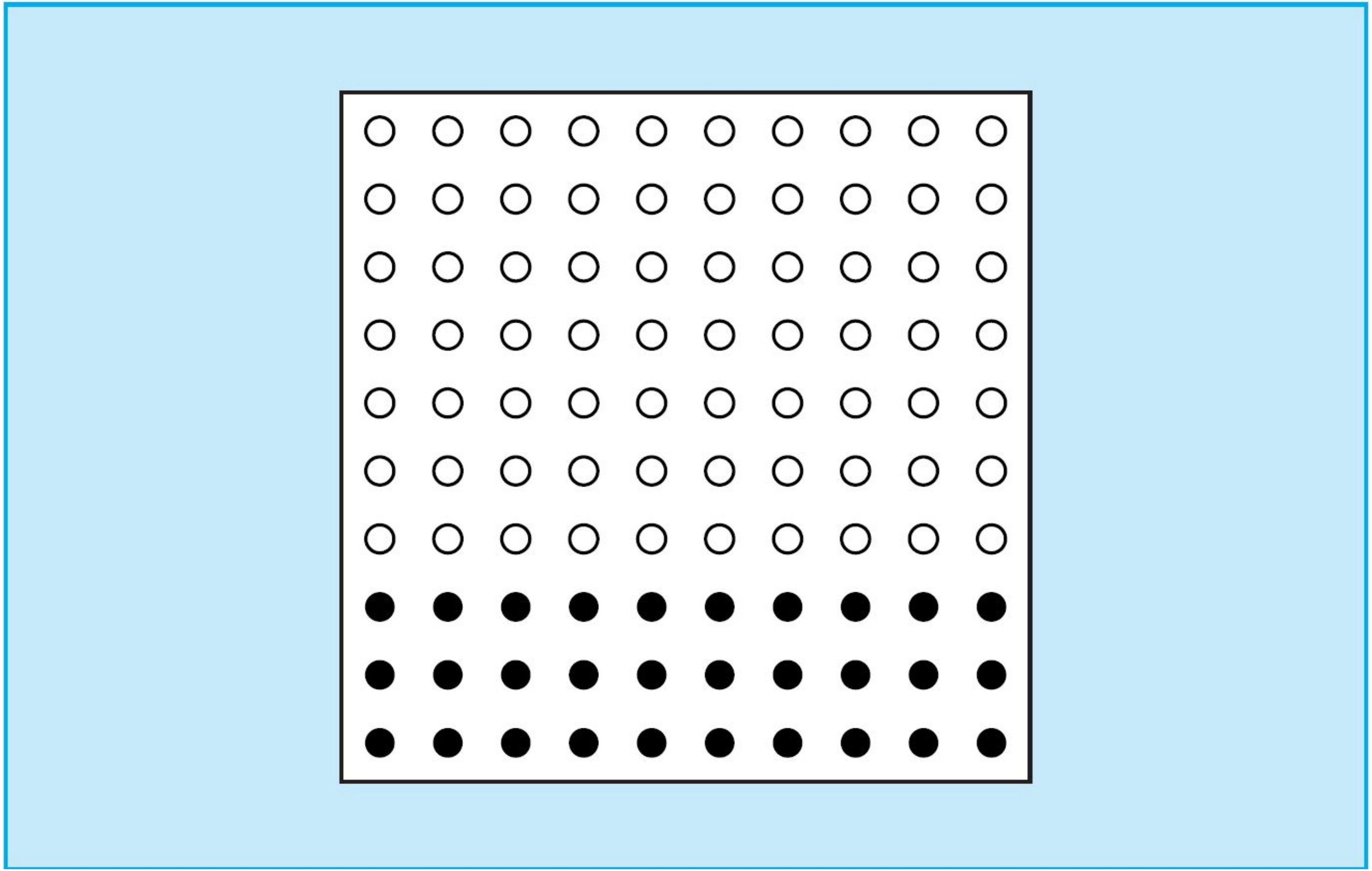
....is a well person who tests negative (a true negative)

....is a person with a disease who tests positive (a true positive)

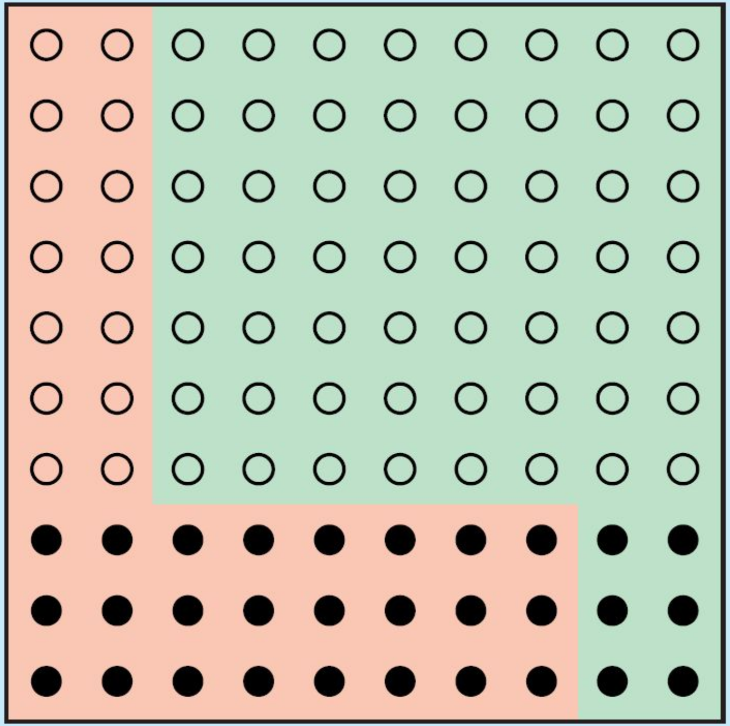
....is a well person who tests positive (a false positive)

....is a person with a disease who tests negative (a false negative)

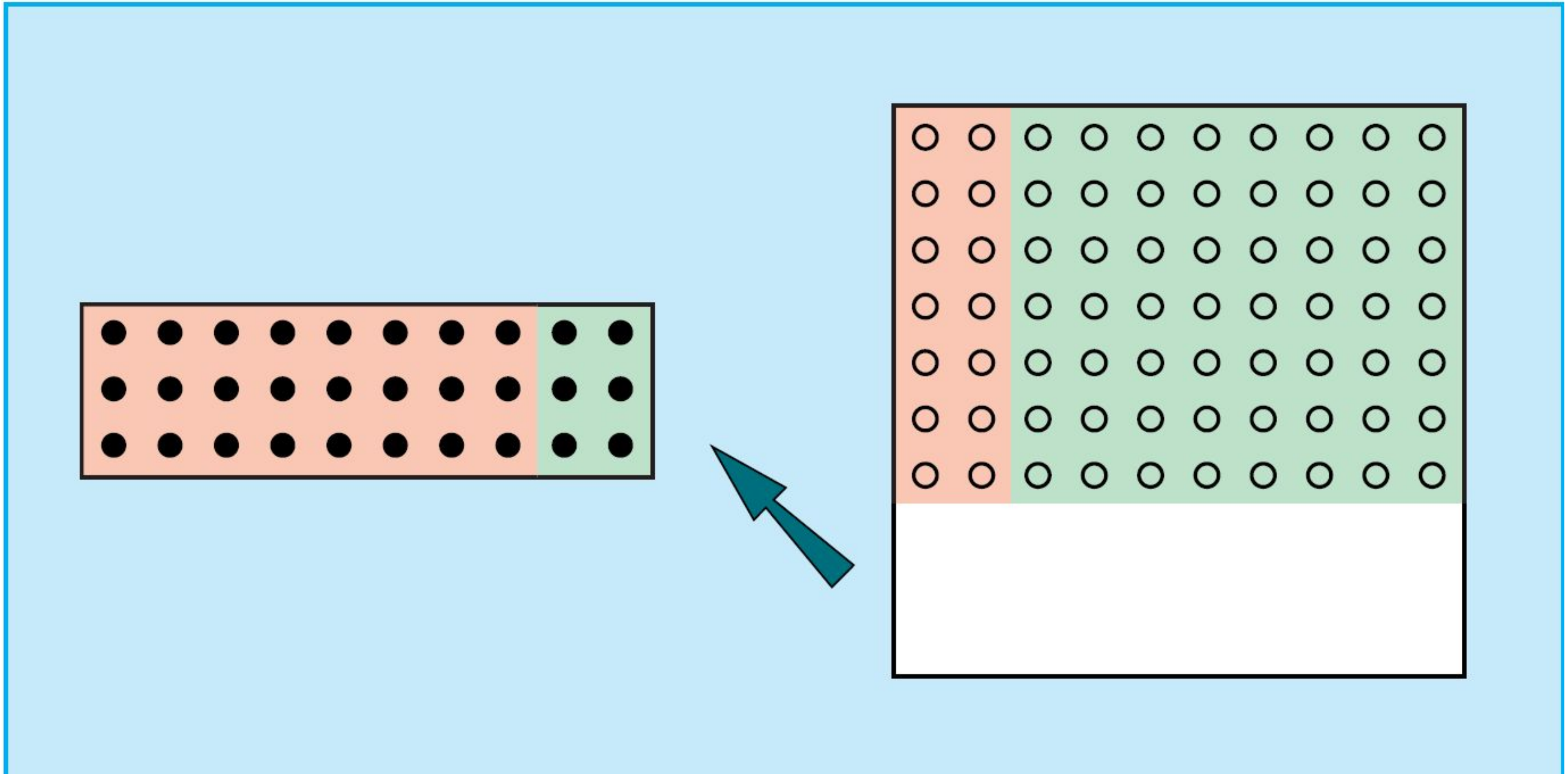
Hypothetical Population



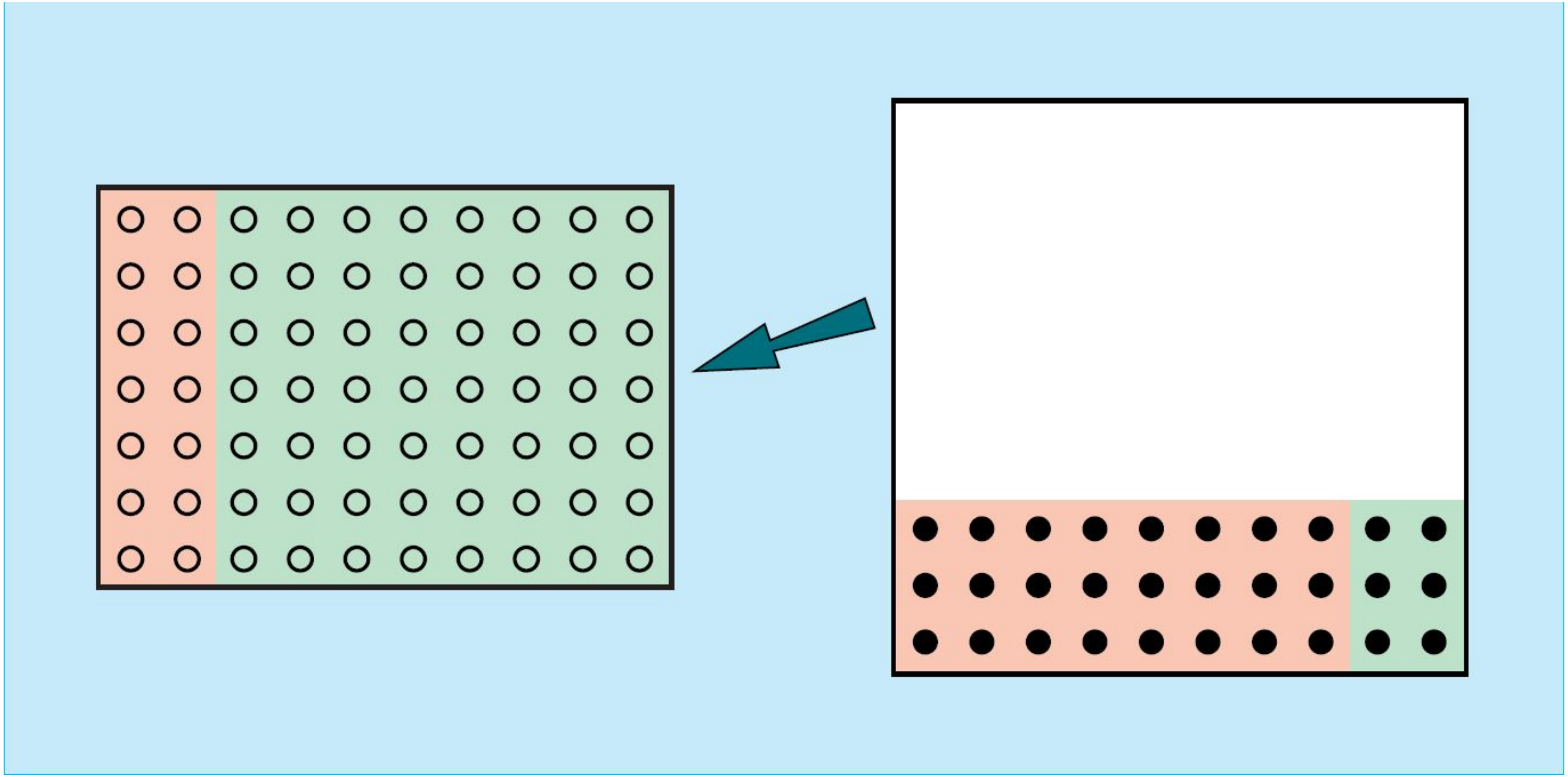
Results of diagnostic test on hypothetical population



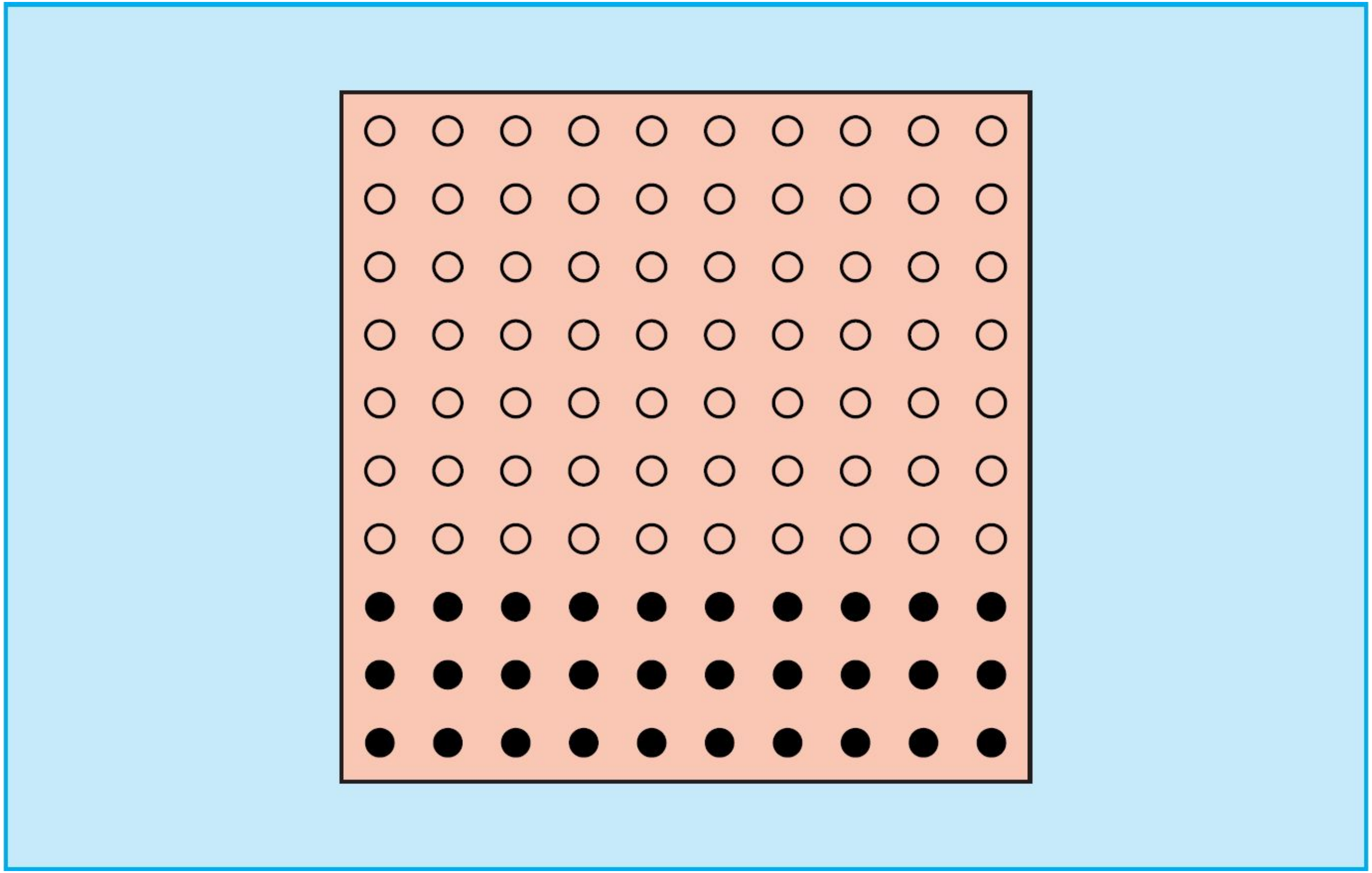
Sensitivity of a Test



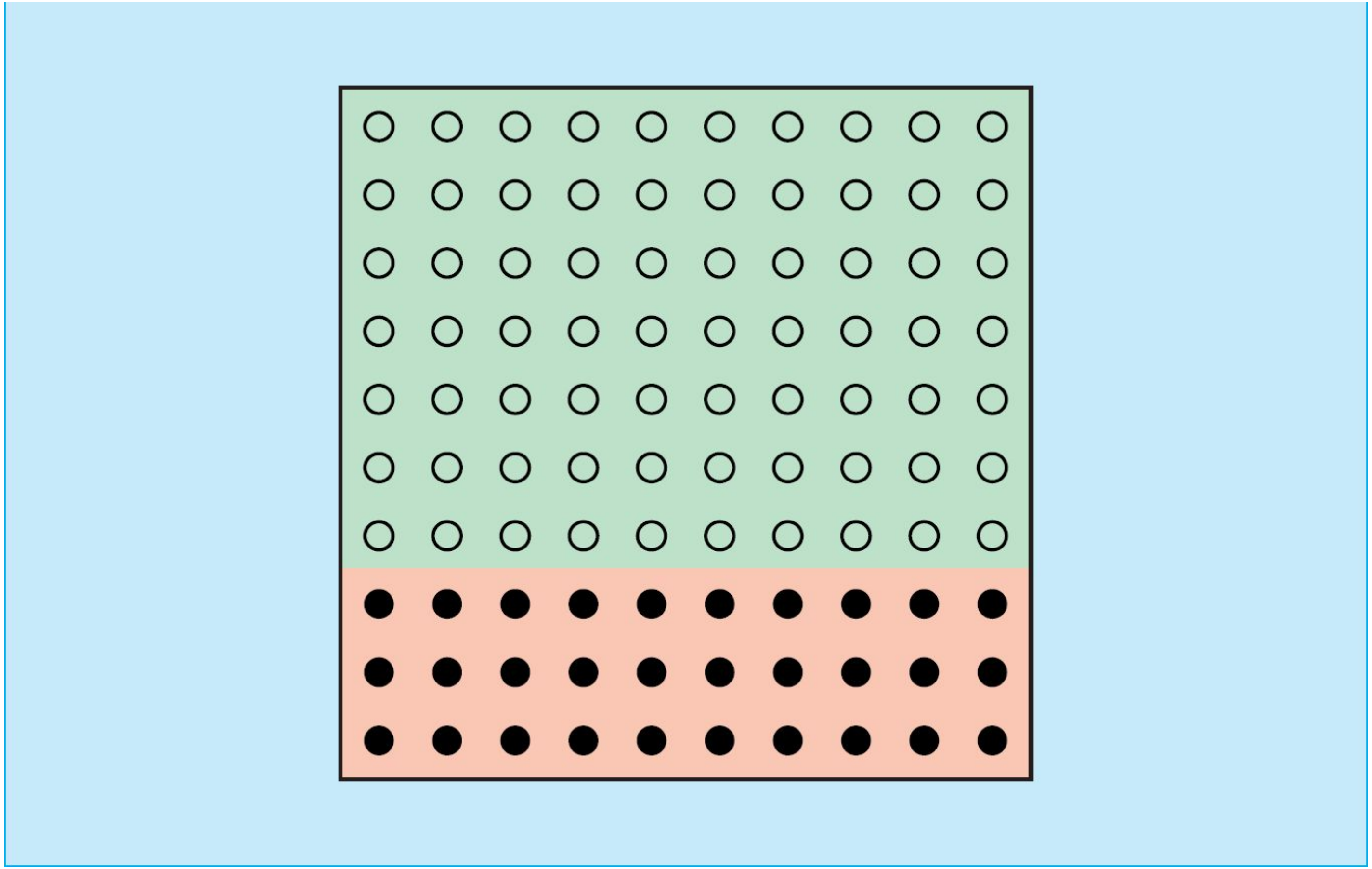
Specificity of a Test



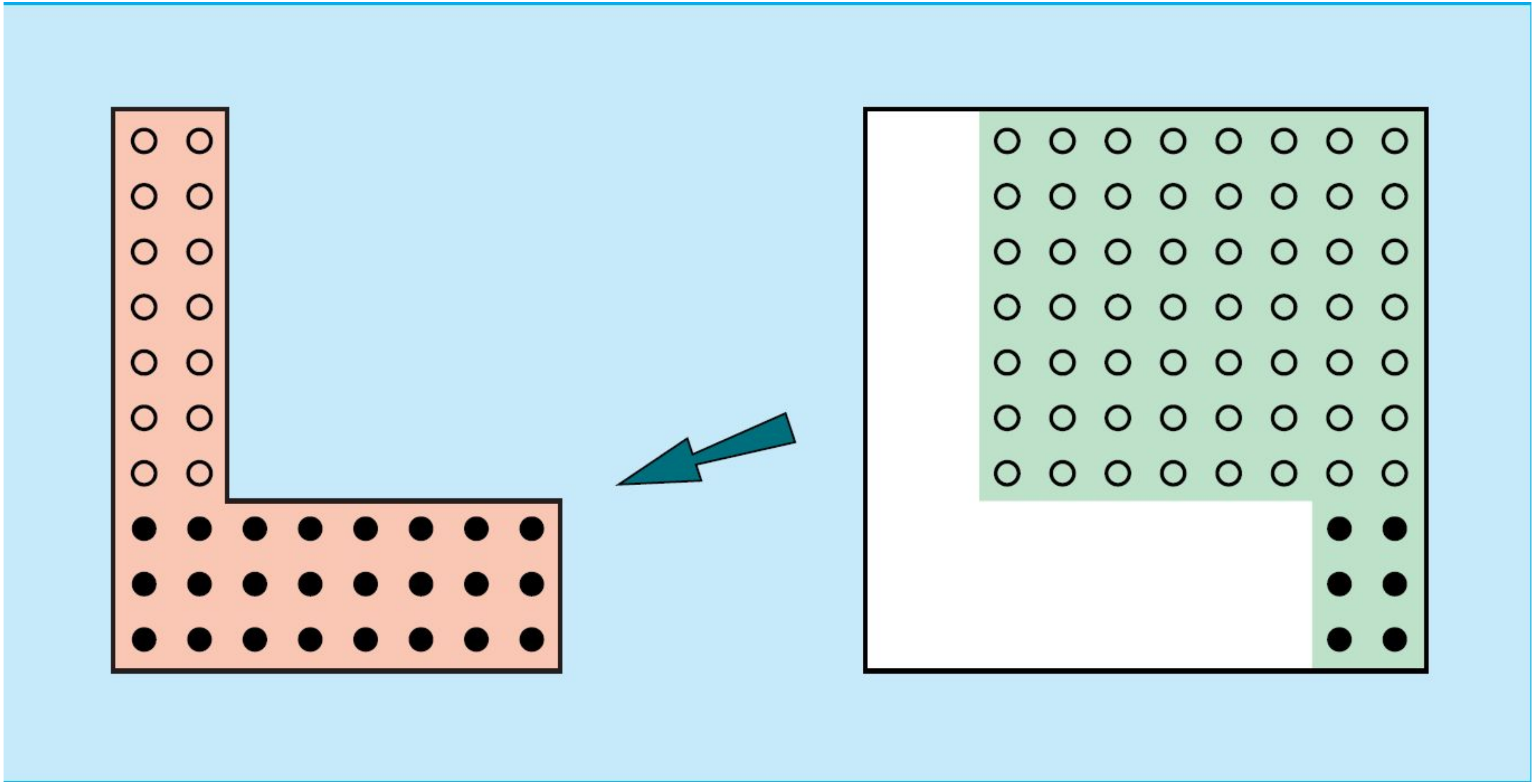
Test with 100% sensitivity



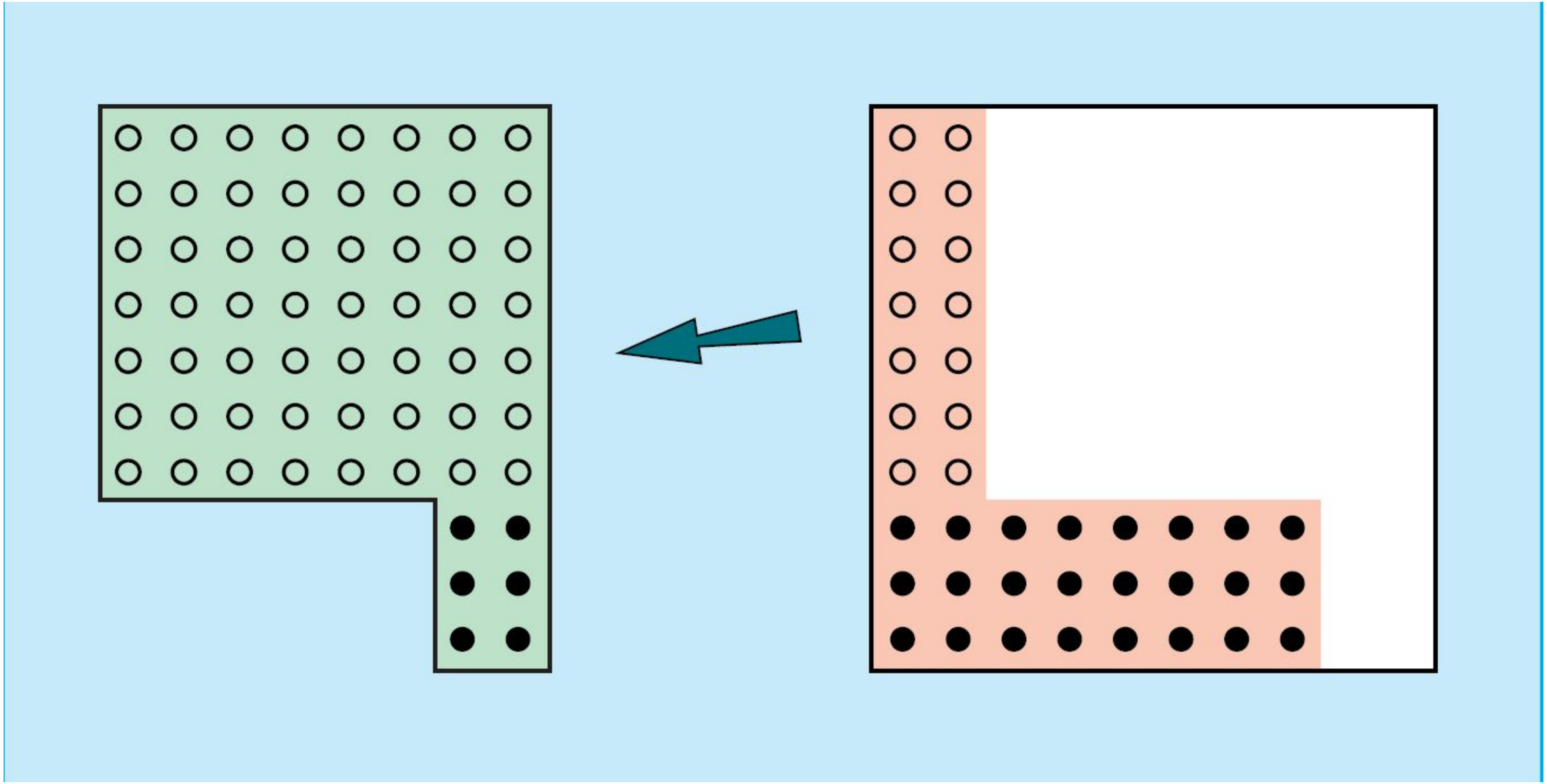
Perfect Test



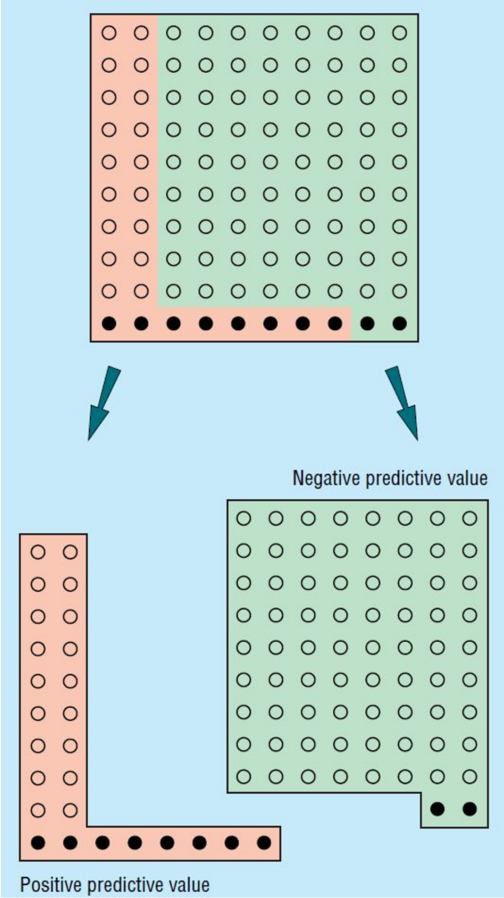
Positive Predicted Value



Negative Predicted Value



Results of testing population with disease prevalence of 10%



2 x 2 table for diagnostic test results

	Disease present (+)	**Disease absent (−)**	**Totals**
Test result positive (+)	a	b	a + b
Test result negative (−)	c	d	c + d
Totals	a + c	b + d	—

Sensitivity (True Positive Rate, Recall):

"I know my patient **has the disease**. What is the chance the test will show that my patient has it?"

	Disease present (+)	Disease absent (–)	Totals
Test result positive (+)	a	b	a + b
Test result negative (–)	c	d	c + d
Totals	a + c	b + d	–

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

Specificity (True Negative Rate):

"I know my patient **doesn't have the disease**. What is the chance the test will show that my patient doesn't have it?"

	Disease present (+)	Disease absent (–)	Totals
Test result positive (+)	a	b	a + b
Test result negative (–)	c	d	c + d
Totals	a + c	b + d	–

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

Negative Predicted Value (PPV):

"I just got a **negative test result** back on my patient. What is the chance that my patient actually doesn't have the disease?"

	Disease present (+)	Disease absent (−)	Totals
Test result positive (+)	a	b	a + b
Test result negative (−)	c	d	c + d
Totals	a + c	b + d	—

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

Positive Predicted Value (PPV):

"I just got a **positive test result** back on my patient. What is the chance that my patient actually has the disease?"

	Disease present (+)	Disease absent (−)	Totals
Test result positive (+)	a	b	a + b
Test result negative (−)	c	d	c + d
Totals	a + c	b + d	—

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

Example: Screening test for breast cancer

	Diseased	Not Diseased	Total
Test Positive	132	983	1,115
Test Negative	45	63,650	63,695
Column Totals	177	64,633	64,810

Sensitivity = 132 / 177 = 75%

Specificity = 63,650 / 64,633 = 98.5%

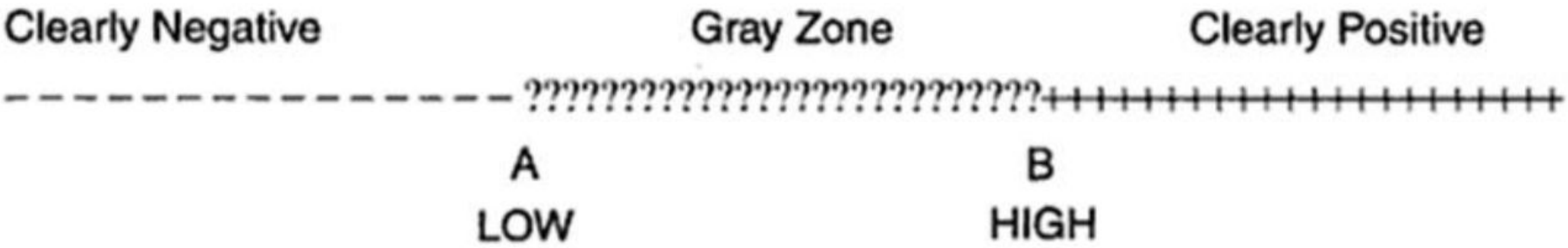
PPV = 132 / 1,115 = 12%

NPV = 63,650 / 63,695 = 99.9%

Questions to consider

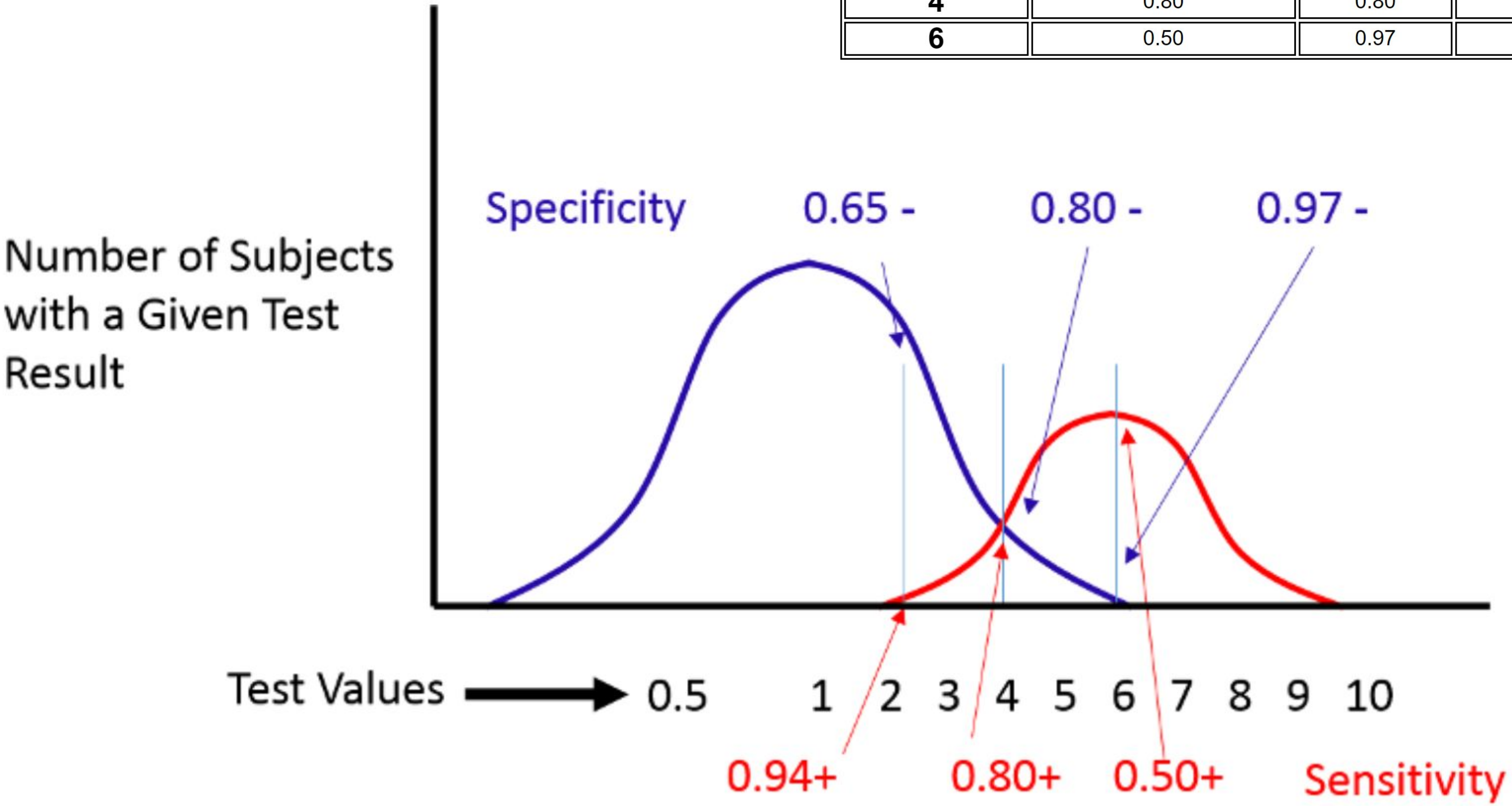
- What is the prevalence of disease?
- Is there a gold-standard test to definitively determine disease status?
- How reliable (stable) is the screening test?
- **How was the criterion for a positive or negative screen determined?**

Test value used to distinguish positive and negative cases



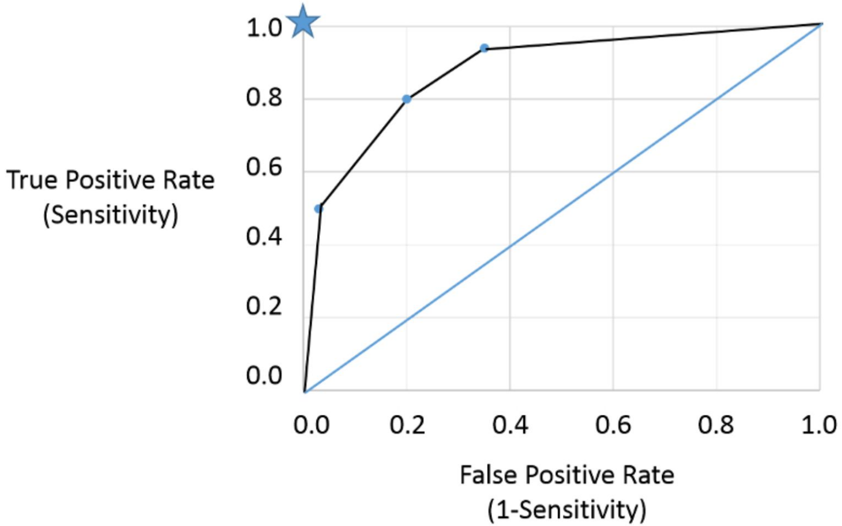
Sensitivity - Specificity Tradeoff

Criterion of Positivity	Sensitivity (True Positive Rate)	Specificity	False Positive Rate (1-Specificity)
2	0.94	0.65	0.35
4	0.80	0.80	0.20
6	0.50	0.97	0.03



Receiver Operating Characteristic (ROC) Curves

Criterion of Positivity	Sensitivity (True Positive Rate)	Specificity	False Positive Rate (1-Specificity)
2	0.94	0.65	0.35
4	0.80	0.80	0.20
6	0.50	0.97	0.03



ROC Plots

1. The ROC plot shows the tradeoff between specificity and sensitivity
2. In ROC plots, classifiers with random performance show a straight diagonal line
3. A ROC curve provides a single performance measure called the Area under the ROC curve (AUC) score
4. AUC is 0.5 for random and 1.0 for perfect classifiers
5. AUC scores are convenient to compare the performances of multiple classifiers
6. It is model-wide because it shows pairs of specificity and sensitivity values calculated at all possible threshold

Limitations of ROC with class imbalance

Balanced

Truth

	Disease Present (+)	Disease Absent (-)
Test Positive (+)	6	4
Test Negative (-)	4	6

Imbalanced

Truth

	Disease Present (+)	Disease Absent (-)
Test Positive (+)	3	6
Test Negative (-)	2	9

Measure	Balanced	Imbalanced
Precision (PPV)	6 / (6+4) = 0.60	3 / (3 + 6) = 0.33
Sensitivity	6 / (6 + 4) = 0.60	3 / (3+2) = 0.60
False Positive Rate	4 / (4 + 6) = 0.40	6 / (6+9) = 0.40
Accuracy	(6+6) / (20) = 0.60	(3 + 9) / 20 = 0.60

With Class Imbalance, Use Precision-Recall (PRC) instead

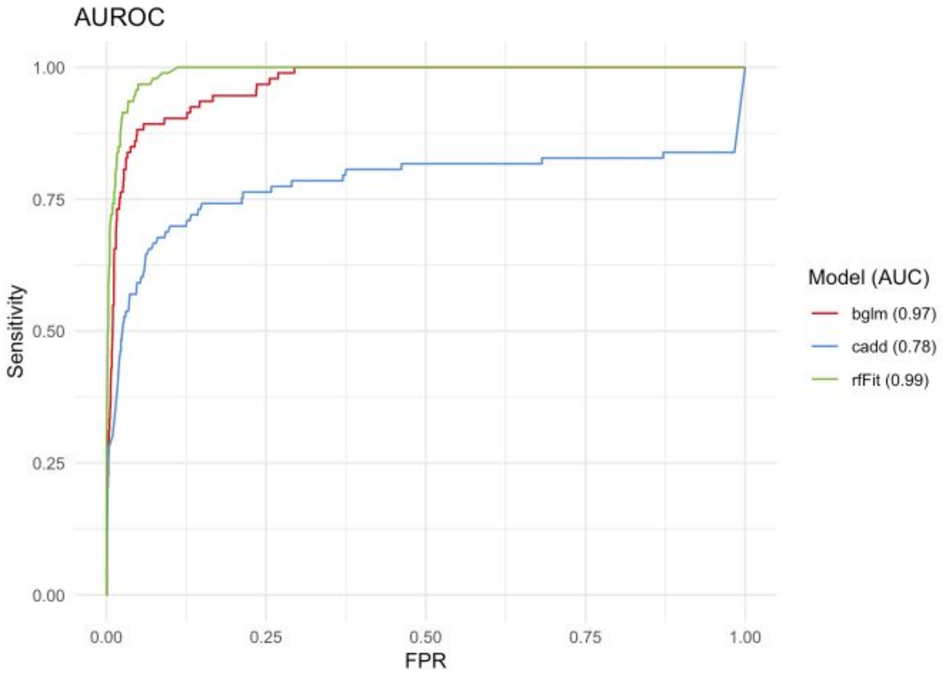
1. While the baseline is fixed with ROC, the baseline of PRC is determined by the ratio of positives (P) and negatives (N) as

$$y = P / (P + N)$$

2. For instance, we have y = 0.5 for a balanced class distribution, but y = 0.09 for an imbalanced class distribution in which the ratio of P:N is 1:10

Example: ClinVar variants for a variety of eye disease

	Disease Present: Pathogenic (+)	Disease Absent: Not Pathogenic (-)
Sample Size	186	8246



Example: ClinVar variants for a variety of eye disease

BGLM (AUC 0.97)

Truth

	Disease Present (+)	Disease Absent (-)
Test Positive (+)	43	35
Test Negative (-)	50	4088

Random Forest (AUC 0.99)

Truth

	Disease Present (+)	Disease Absent (-)
Test Positive (+)	58	18
Test Negative (-)	35	4105

Cadd (AUC 0.78)

Truth

	Disease Present (+)	Disease Absent (-)
Test Positive (+)	10	3
Test Negative (-)	83	4120

- TPR for BGLM = 43 / 93 = 46%
- But the AUC looked great!
- TPR for RF = 62%
- TPR for Cadd = 11%

Area Under Precision Recall Curve (pro: do not take into account negative class)

	Disease Present: Pathogenic (+)	Disease Absent: Not Pathogenic (-)
Sample Size	186	8246