

Validity of tests

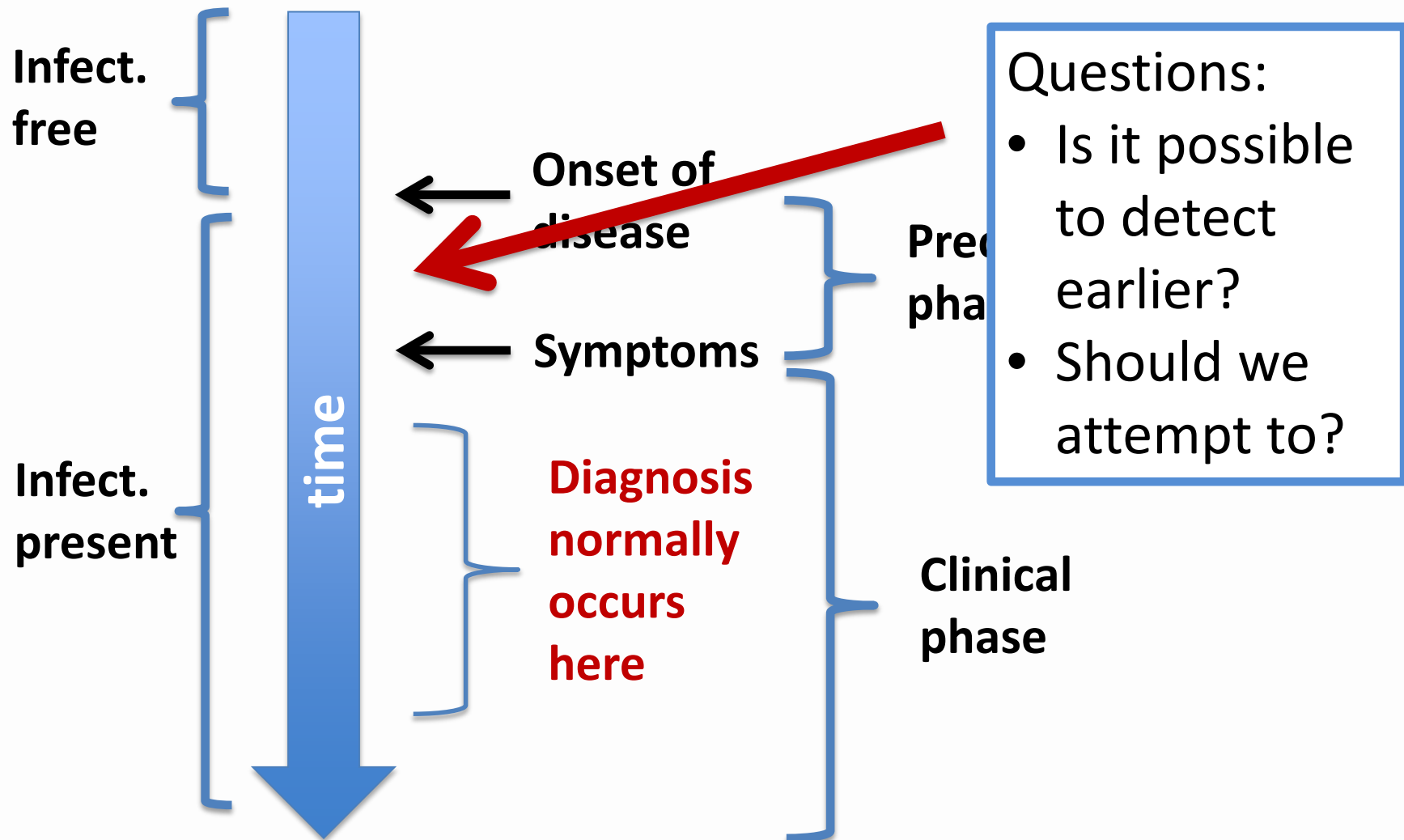
PH101

Spring quarter 2019

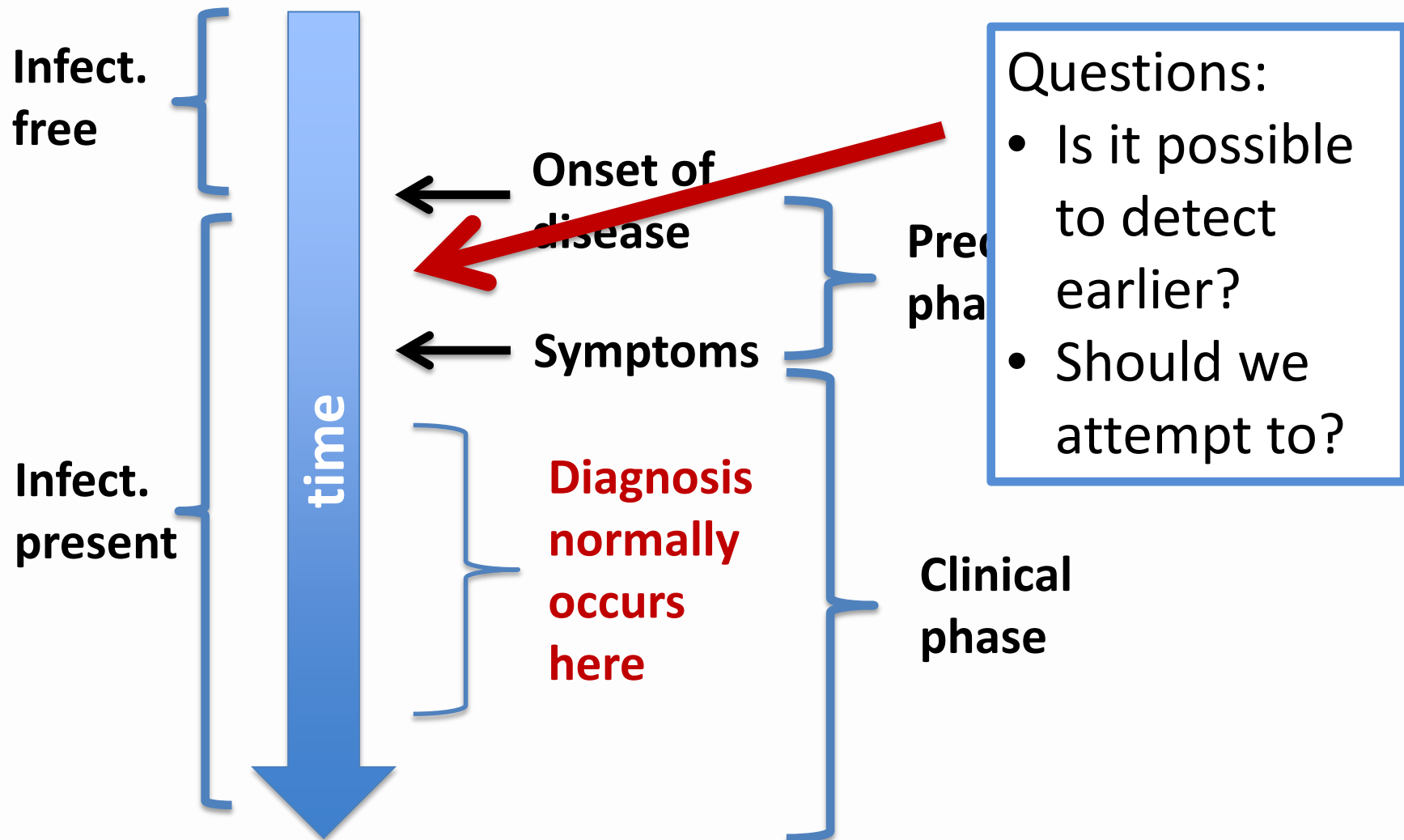
Lecture hall: PCB 1100

Classification problem

- We need to be able to categorize people as either diseased or disease-free
- We want to be able to do this accurately and efficiently



- Diagnostic test: test for absence/presence of disease, given that individual is symptomatic
- Screening test: test for asymptomatic individuals who may have the disease



With regard to tests, we are concerned with:

Accuracy (validity): how well does a test define those with and without a disease?

Reliability: how consistent is a test?

- **Sensitivity:** test ability to correctly identify those who do have a disease
- **Specificity:** test ability to correctly identify those who do not have a disease

- In order to calculate the sensitivity and specificity:
- Must have a **gold standard** test for comparison
- Use a 2 by 2 table to compare the **new test** to the **gold standard** test

people with disease	people without disease
---------------------------	------------------------------

Gold Standard

+	-
people with disease	people without disease

	+	-
+	true positive	
-		true negative

	+	-
+	a (true positive)	b
-	c	d (true negative)

- Sensitivity

$$\textit{sensitivity} = \frac{a}{a + c} = \frac{\textit{true positives}}{\textit{people disease " + "}}$$

- Specificity

$$\text{specificity} = \frac{d}{b + d} = \frac{\text{true negatives}}{\text{people disease " - "}}$$

Example

The study population has a tuberculosis (TB) prevalence of 100 per 1000 people.

	+	-
+	a (true positive)	b
-	c	d (true negative)
	100	900

A test is being used to screen the population. The test has a sensitivity of 80%.

	+	-
+	80	b
-	c	d (true negative)
	100	900

The test has a specificity of 89%

	+	-
+	80	b
-	c	800
	100	900

We can now calculate the remaining parts of the table.

	+	-
+	80	100
-	20	800
	100	900

	+	-	
+	80	100	180
-	20	800	820
	100	900	

$$\frac{80}{100}$$

sensitivity

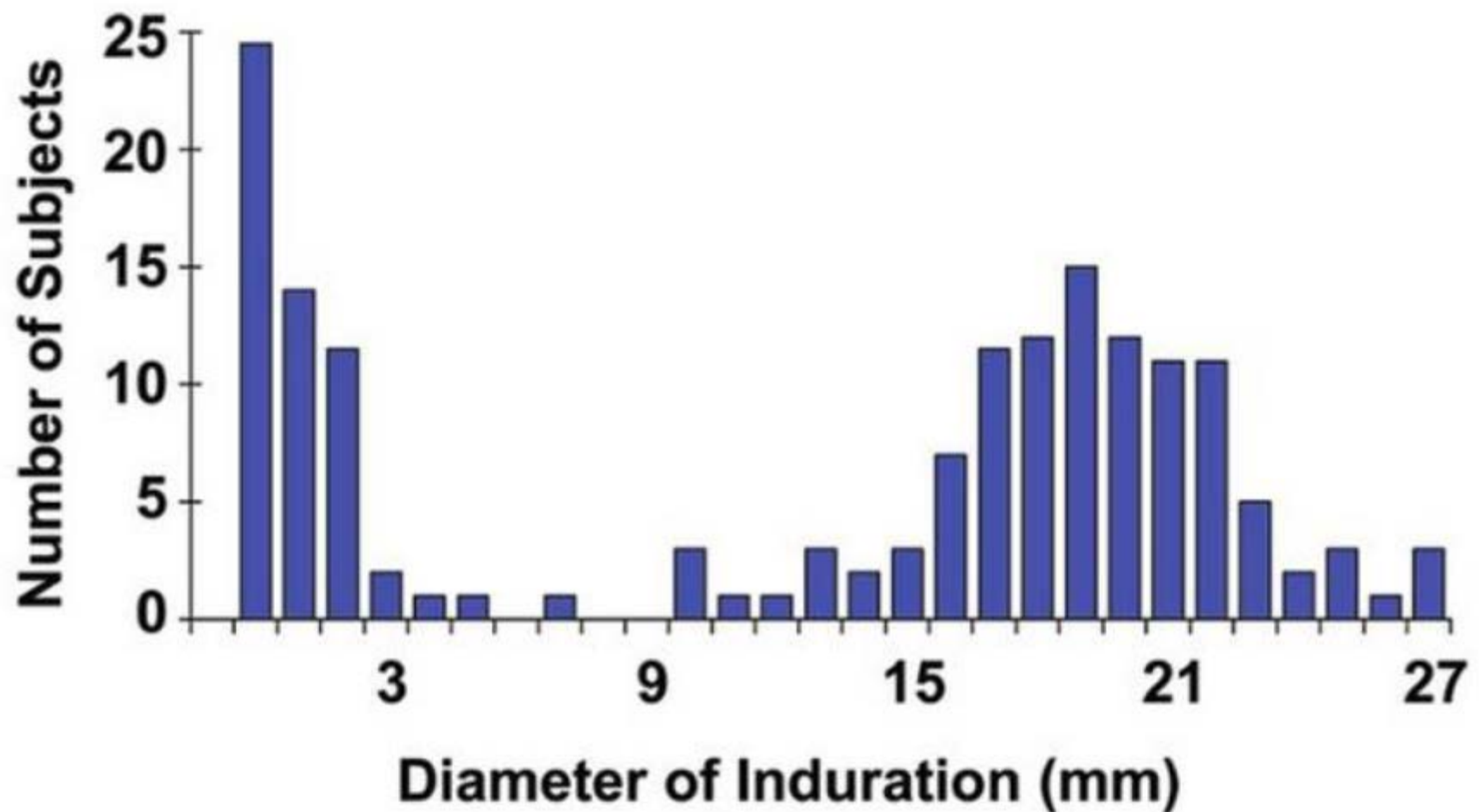
$$\frac{800}{900}$$

specificity

What is actually being measured?
How is it distributed?

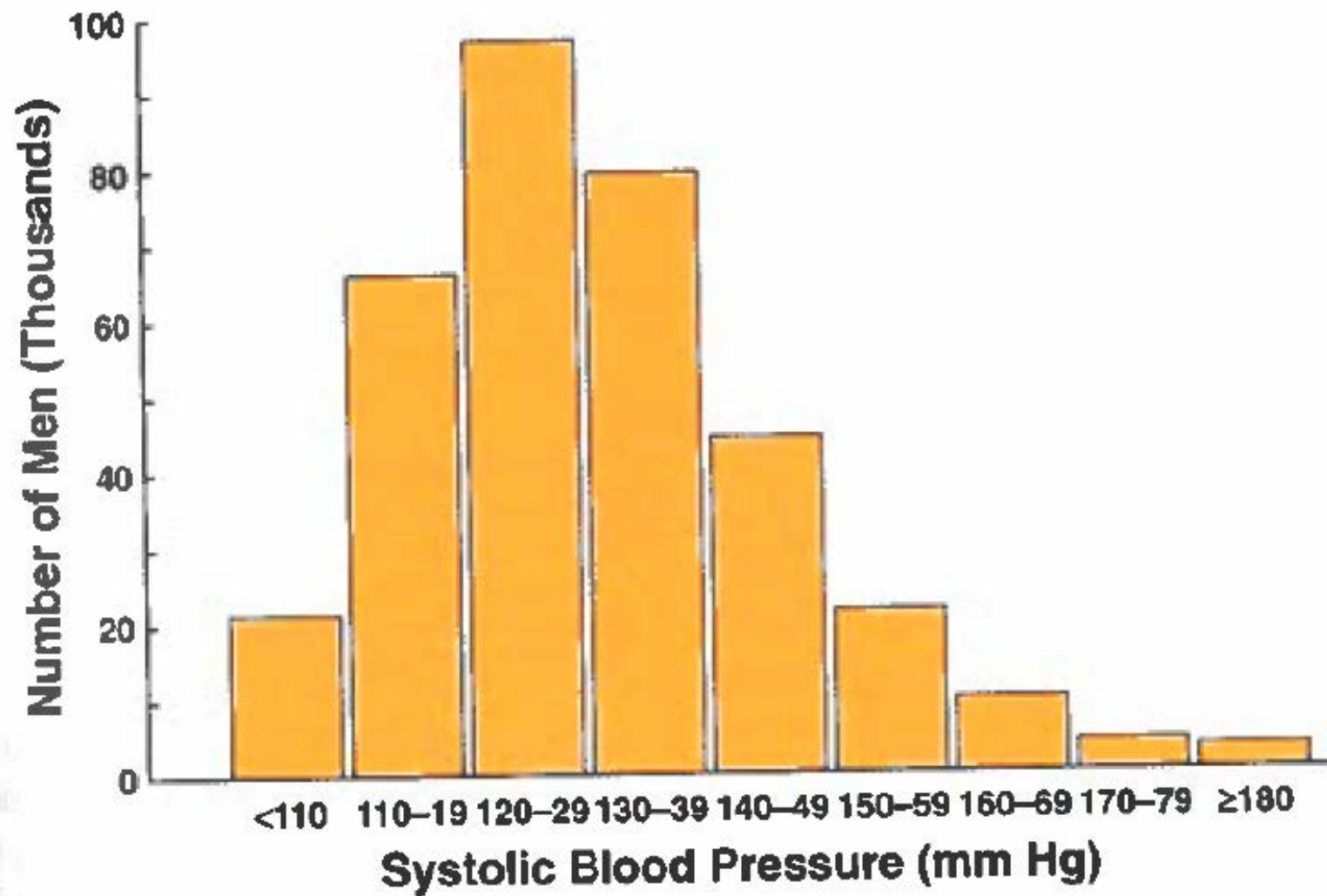
Variation in biological outcomes

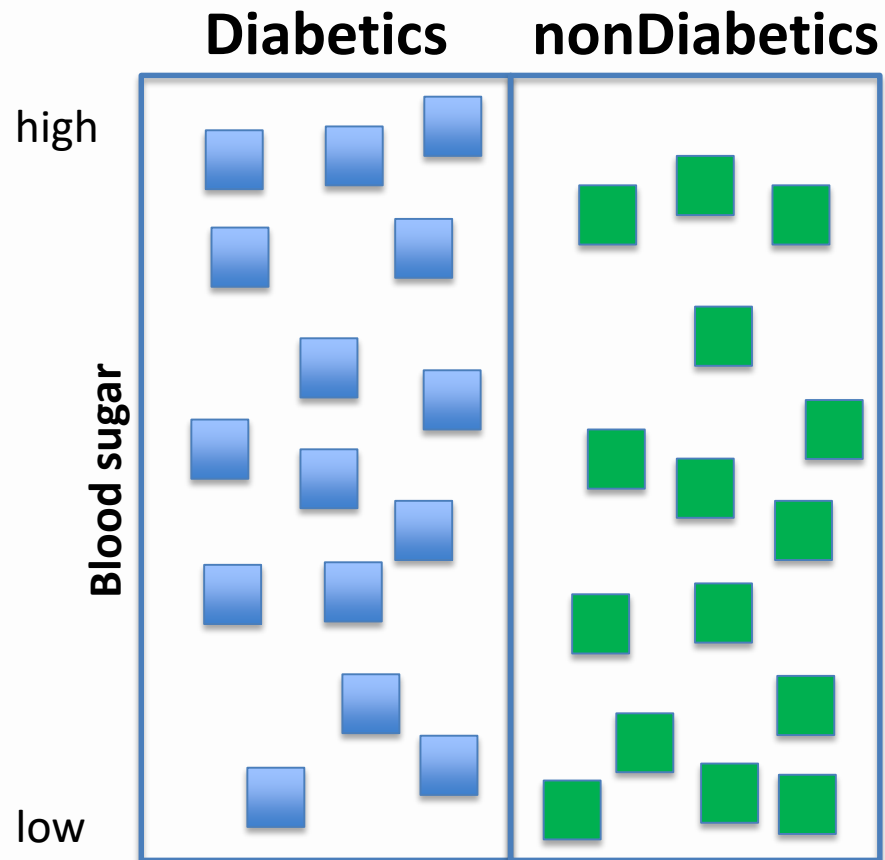
- Binary traits
- Continuous traits
 - Unimodal
 - Multimodal

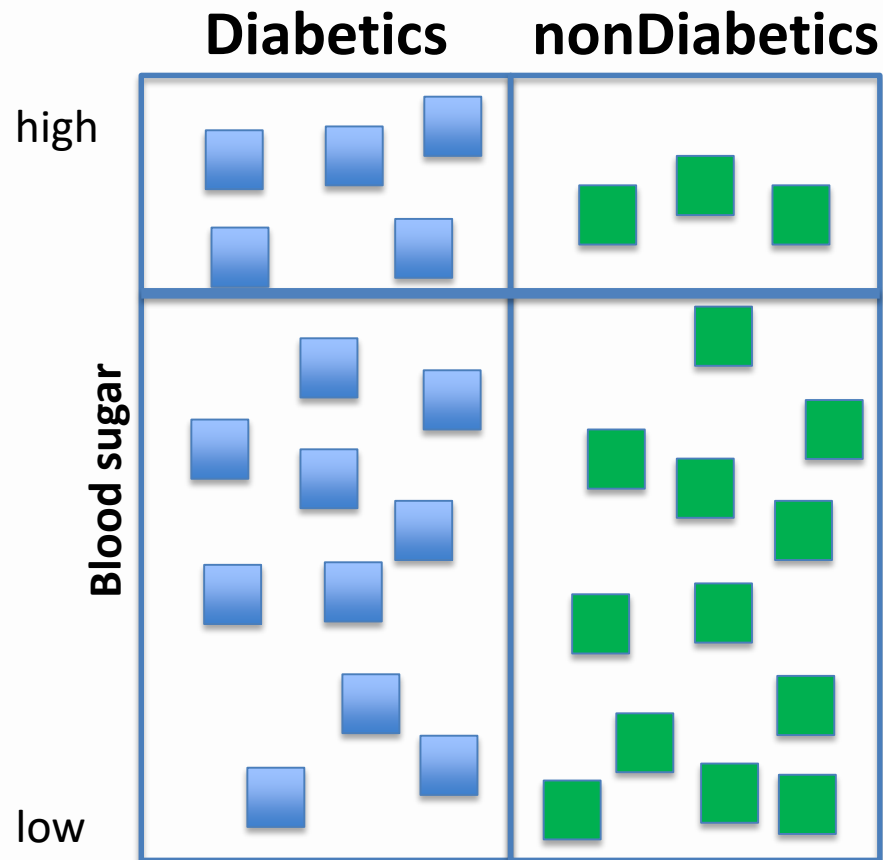


Distribution of tuberculin reactions.

Adapted from Gordis L. Epidemiology. 5th Edition



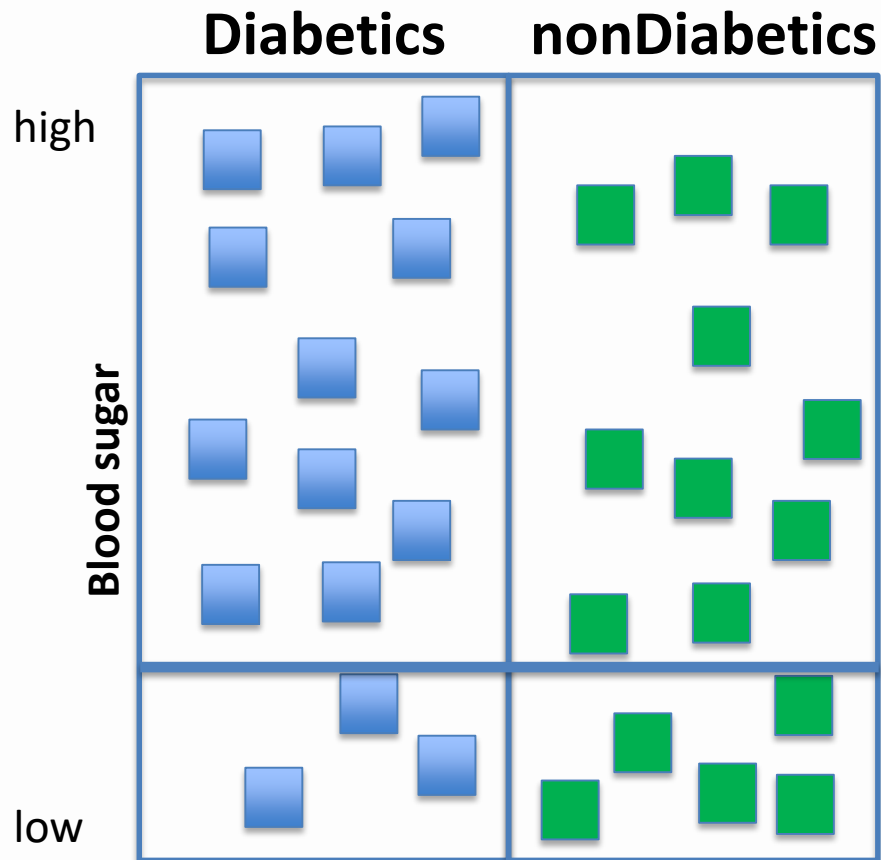




	+	-	
+	5	3	8
-	10	12	22
	15	15	

Sensitivity = 0.33

Specificity = 0.80



Sensitivity = 0.80
Specificity = 0.33

	+	-	
+	12	10	22
-	3	5	8
	15	15	

- Cutoff levels influence who is diagnosed as having a disease
- Cutoff levels influence both sensitivity and specificity
- A cutoff that identifies more true positives will also result in more false positives
- A cutoff that identifies more true negatives will also result in more false negatives

- Finding the right balance
- What is more important?
- Mis-classifying disease + people as being disease -?
- Or classifying people who are disease – as being disease +?

If the prognosis of a disease is poor, or if the disease easily spreads, likely better to err towards high sensitivity (leading to more false positives)

Class exercise

- Calculate the sensitivity and specificity from these data:

	+	-
+	90	90
-	10	810
	100	900

Sensitivity =

A.) 0.09

B.) 9.99

C.) 0.90

D.) 0.79

Specificity =

A.) 0.09

B.) 9.99

C.) 0.90

D.) 0.79

Multiple tests

- Two stage testing (sequential)
- Simultaneous testing

Sequential testing

- An initial screening is conducted
- Individuals who had a “+” screening test are asked to return for a second test
- This should lead to an increase specificity
- Also likely a decrease in sensitivity

Sequential testing

	+	-	
+	350	1900	2250
-	150	7600	7750
	500	9500	10000

Sensitivity = 0.70

Specificity = 0.80

Sequential testing

Sensitivity = 0.90

Specificity = 0.90

	+	-	
+	350	1900	2250
-	150	7600	7750
	500	9500	10000

	+	-	
+	315	190	505
-	35	1710	1745
	350	1900	2250

- The overall specificity has increased $\frac{1710 + 7600}{9500} = 0.98$

Sequential testing

Sensitivity = 0.90

Specificity = 0.90

	+	-	
+	350	1900	2250
-	150	7600	7750
	500	9500	10000

	+	-	
+	315	190	505
-	35	1710	1745
	350	1900	2250

- The overall specificity has increased
- The overall sensitivity had decreased

$$\frac{315}{500} = 0.63$$

Simultaneous testing

- Two or more tests are done concurrently
- If any of the tests result in a positive, then the individual is considered disease “+”
- If all of the tests result in a negative, then the individual is considered to be disease “-”
- This results in an increase in sensitivity
- And a decrease in specificity

Predictive values

- If an individual tests positive for a disease, what is the probability that that person really has the disease?
 - Positive predictive value (PPV)
- If an individual tests negative for a disease, what is the probability that she/he really does not have the disease?
 - Negative predictive value (NPV)

Predictive values

	+	-	
+	80	100	180
-	20	800	820
	100	900	

PPV

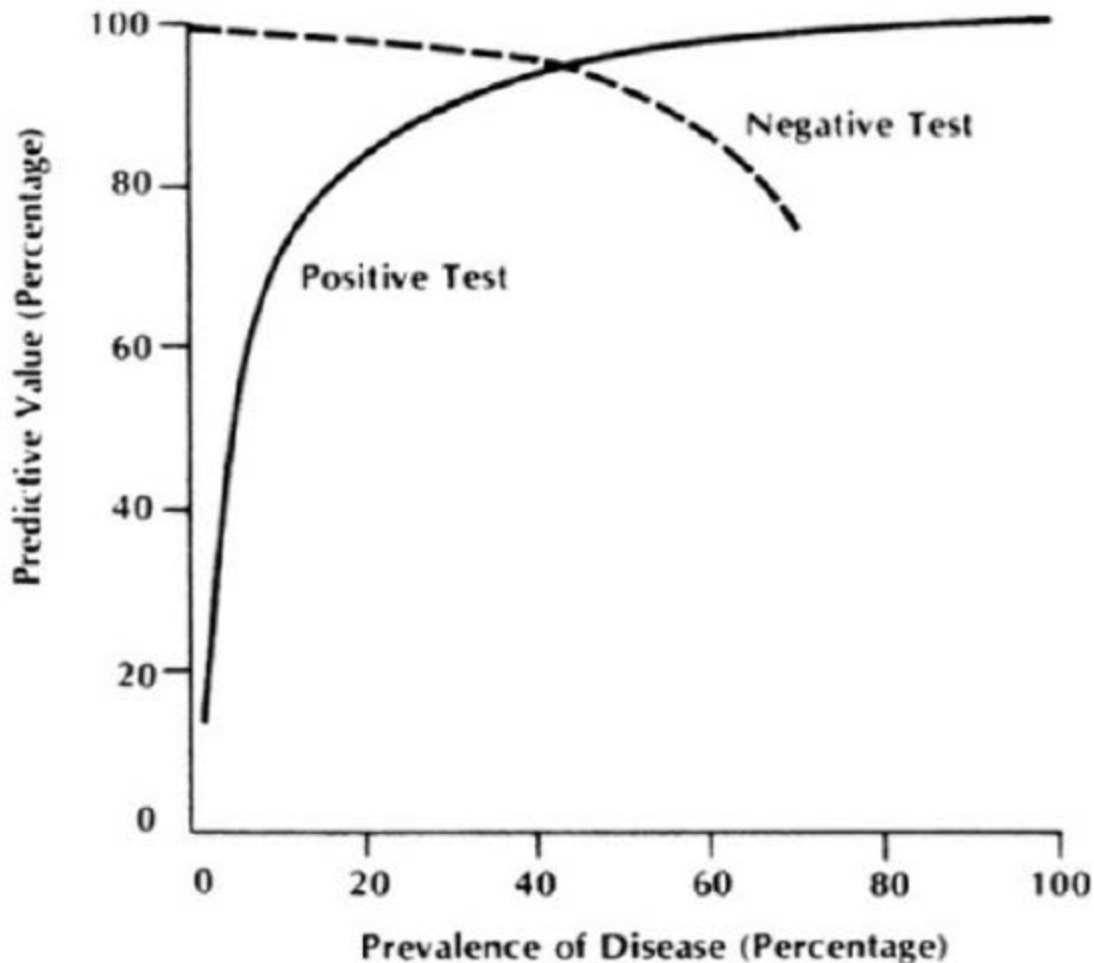
$$\frac{80}{180} = 0.44$$

NPV

$$\frac{800}{820} = 0.98$$

Predictive values

What is the relationship between disease prevalence and predictive value?



Predictive values

- If someone tests +, the probability that they do in fact have the disease depends on:
 - The prevalence of the disease in the population
 - The test characteristics (sensitivity and **specificity**)

Class exercise

- *Calculate PPV and NPV for a given scenario*
- *Recalculate it under a situation of increased prevalence*

- Sensitivity 90%
- Specificity 90%
- Prevalence of 10% (or 100 out of every 1000 people)

- Calculate the PPV:
- Calculate the NPV:

	+	-
+	90	90
-	10	810
	100	900

- The PPV is:
 - A.) 0.50
 - B.) 0.90
 - C.) 0.07
 - D.) 5.00

- The NPV is:
 - A.) 0.50
 - B.) 0.99
 - C.) 0.07
 - D.) 5.00

- Sensitivity 90%
- Specificity 90%
- Prevalence of 30% (or 300 out of every 1000 people)

- Calculate the PPV:
- Calculate the NPV:

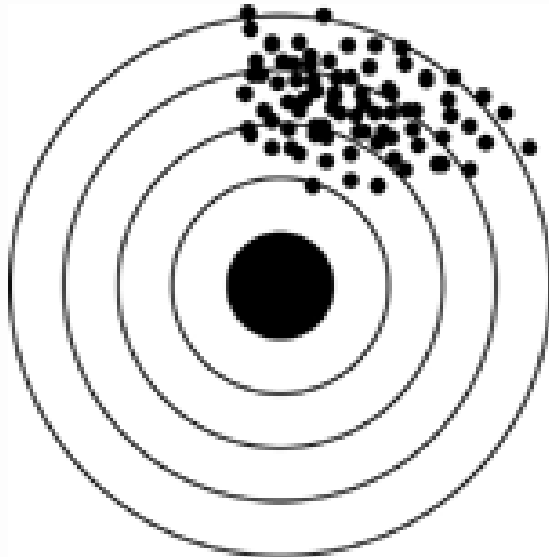
	+	-
+	270	70
-	30	630
	300	700

- The PPV is:
 - A.) 0.56
 - B.) 0.90
 - C.) 0.79
 - D.) 5.67

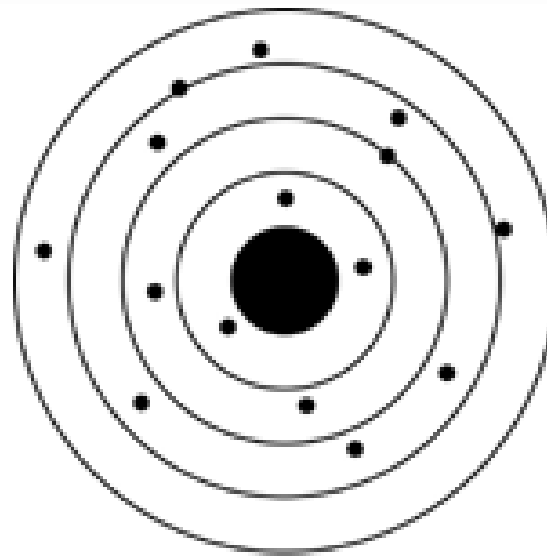
- The NPV is:
 - A.) 0.50
 - B.) 0.95
 - C.) 0.72
 - D.) 5.06

- What is the result of increased prevalence?
- Large increase in PPV
- Slight decrease in NPV

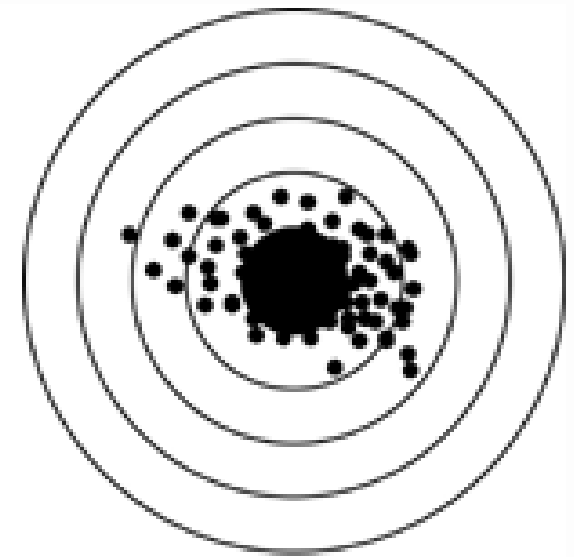
Reliability versus validity



Reliable but Not Valid



Valid but Not Reliable



Valid and Reliable

Reliability and validity

- Intra-subject variation: variation in test outcomes within a single individual
- Intra-observer variation: variations in test observations within a single individual
- Inter-observer variation: variations in test observations across individuals

% Agreement

Overall percent agreement:

$$\frac{a+d}{a+b+c+d} * 100 =$$

		Observer 1	
		Disease +	Disease -
Observer 2	Disease +	a	b
	Disease -	c	d

% Agreement

Perfect agreement:

$$b = 0$$

$$c = 0$$

		Observer 1	
		Disease +	Disease -
Observer 2	Disease +	a	b
	Disease -	c	d

Agreement

Kappa Statistic:

How well in agreement are two different testers, given that there will be some overlap from chance alone

Agreement

Kappa Statistic:

$$\frac{(\% \text{ agreement}) - (\text{random } \% \text{ agreement})}{(100\% \text{ agreement}) - (\text{random } \% \text{ agreement})}$$

Agreement

Kappa Statistic:

$\text{Kappa} < 0.40$ = poor agreement

$\text{Kappa} < 0.75 \ \& \ > 0.40$ = intermediate

$\text{Kappa} > 0.75$ = good agreement

Summary

- screening versus diagnostics
- sensitivity and specificity
- multiple testing
- distributions in biomarkers (binary versus continuous)
- PPV and NPV
- Agreement