Statistical Issues Related to AVA Vaccine Evaluation

Brian D. Plikaytis

Centers for Disease Control and Prevention

Statistical Issues Related to AVA Vaccine Evaluation

- Correlates of protection studies

  Form statistical models that link the results of animal studies to those of human clinical trials

- Missing Data

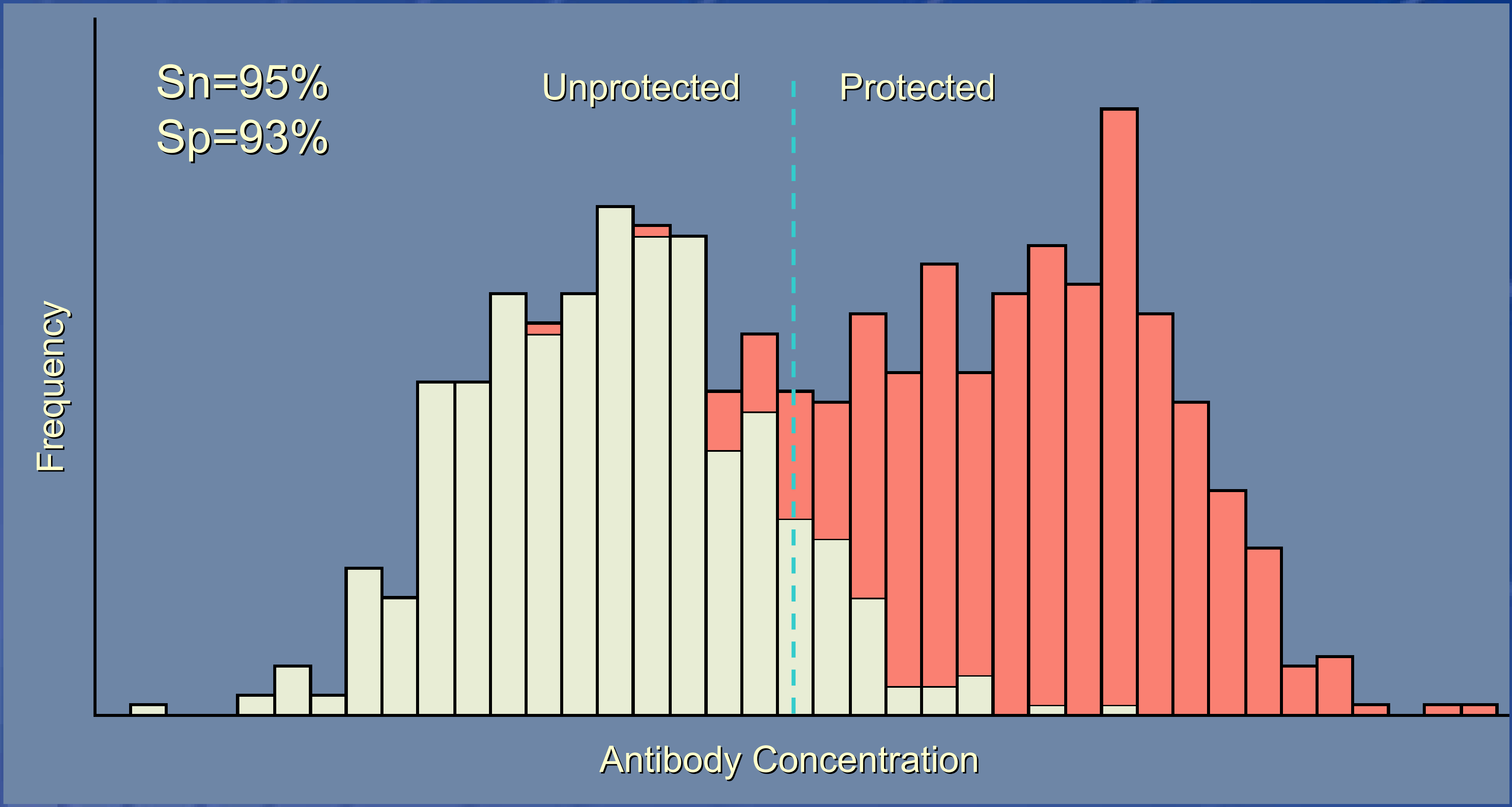
  Techniques which may be applied to resolve missing data

Predicting Protection
(Survival of Challenge)

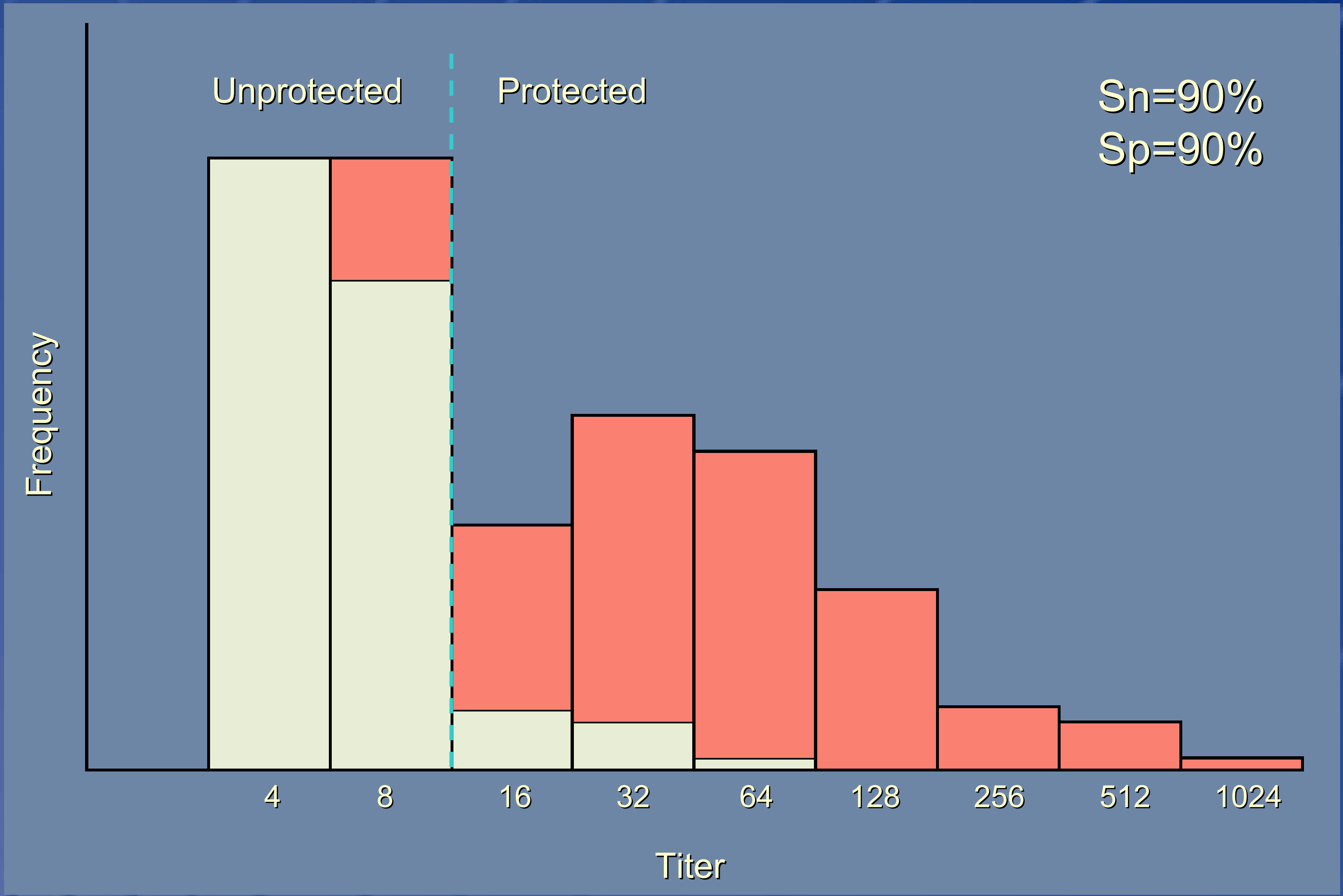
Given that we can measure more than one correlate of protection ….

How do we utilize them to predict protection more effectively than each test individually?

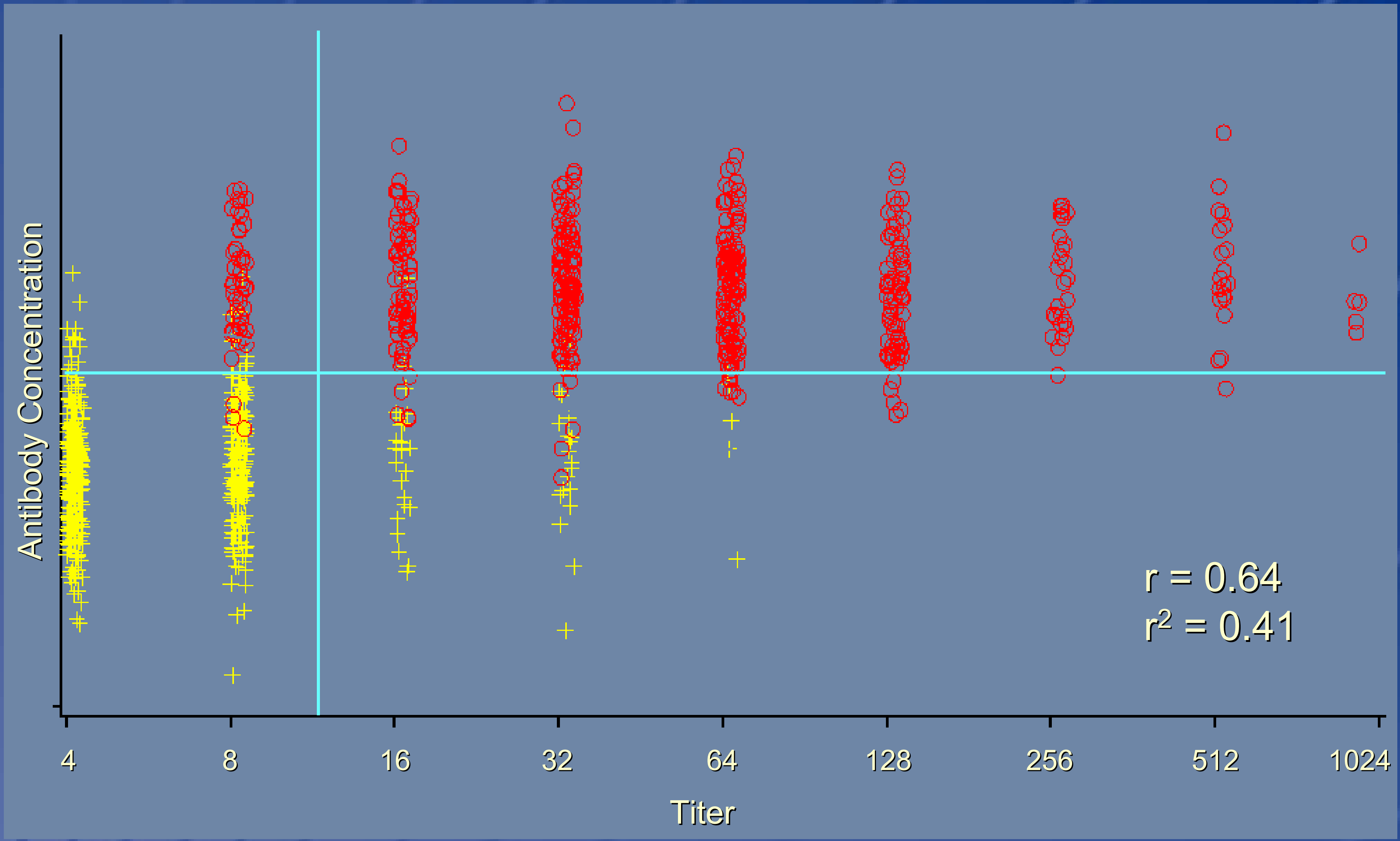
Predicting Protection



Predicting Protection



Predicting Protection



Predictive Model Building

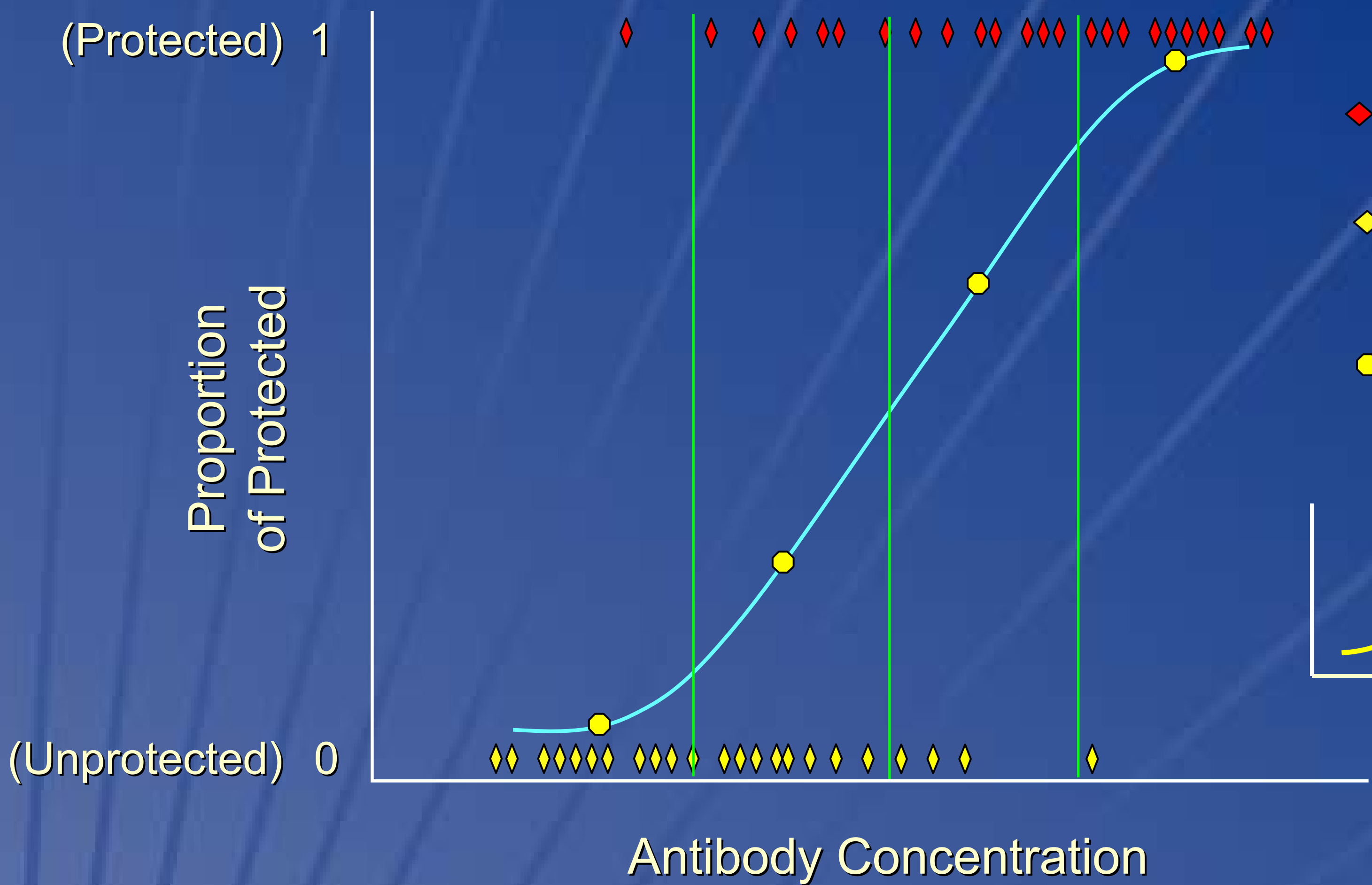
- Multiple approaches
  - logistic discriminant analysis, cluster analysis, other exploratory techniques (CART, etc.)

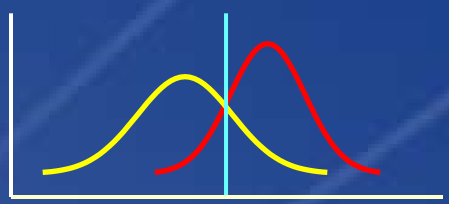
- Logistic Discriminant Analysis
  - Model relates a series of assay endpoints to protection
  - Construct a discriminant function from results:

$$IS = b_1x_1 + b_2x_2 + \ldots + b_px_p$$

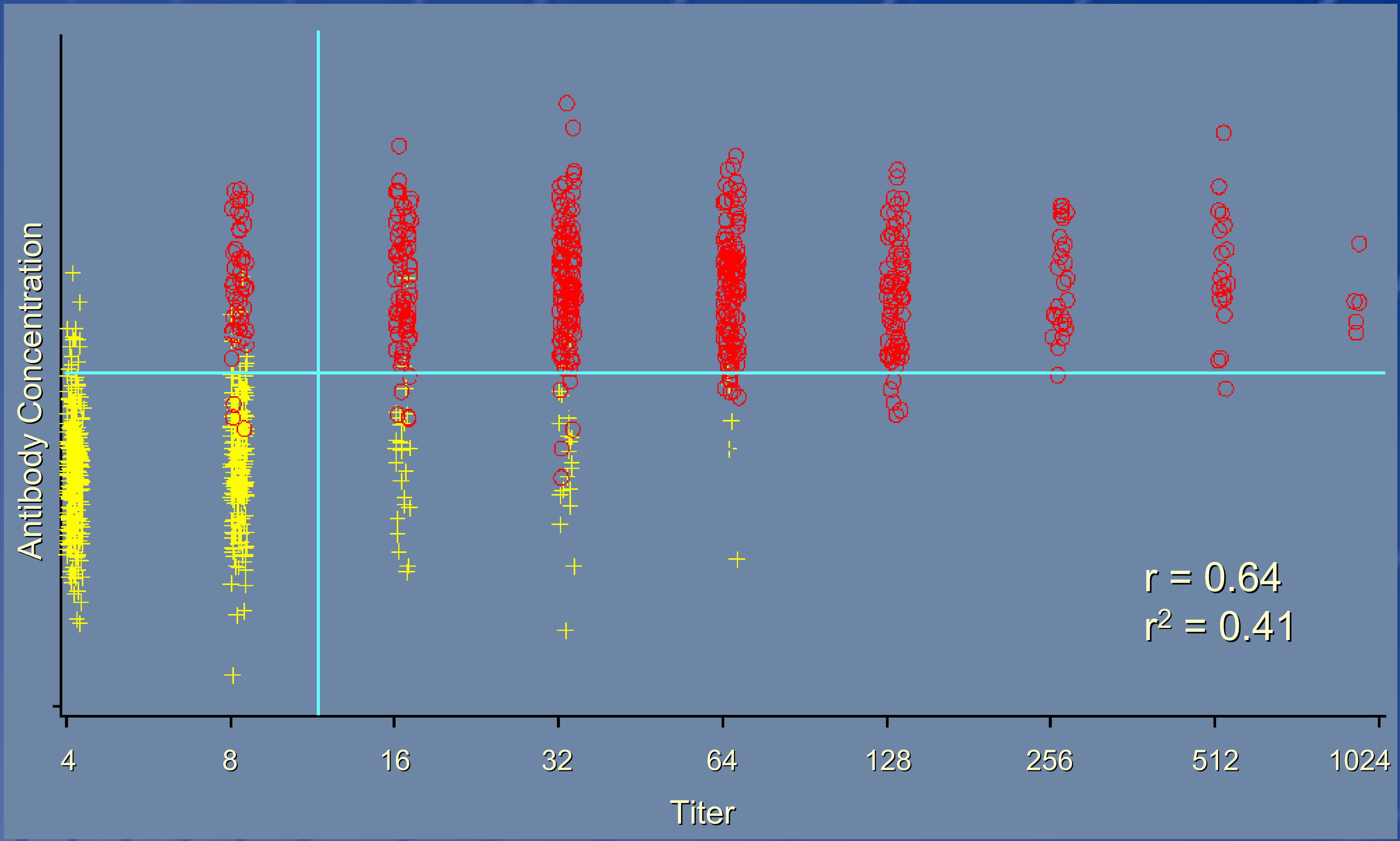
IS = immunologic 'score'
x_i = correlate data
b_i = model coefficients
p = number of variables in the model

Immune Response vs. Protection Model Building

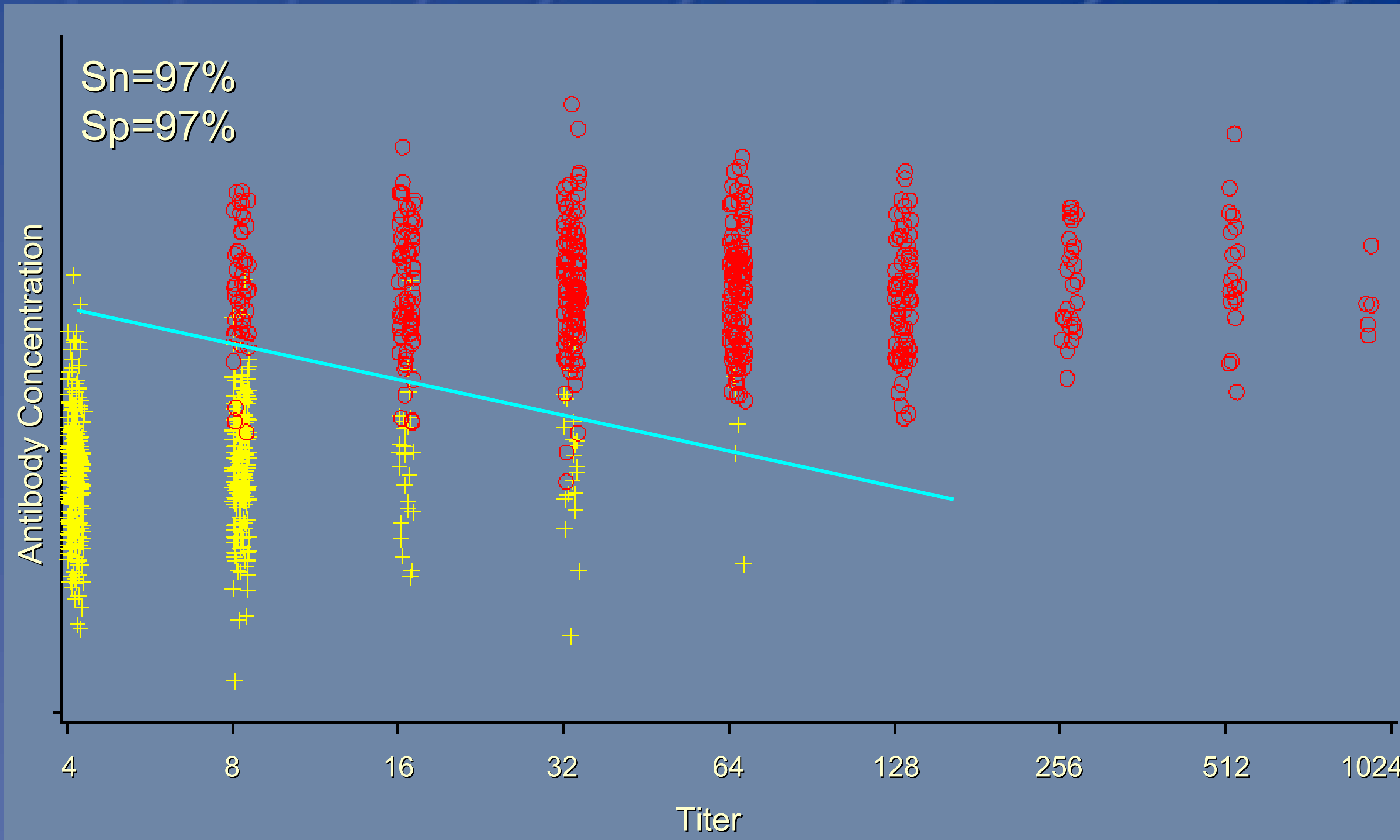




Predicting Protection

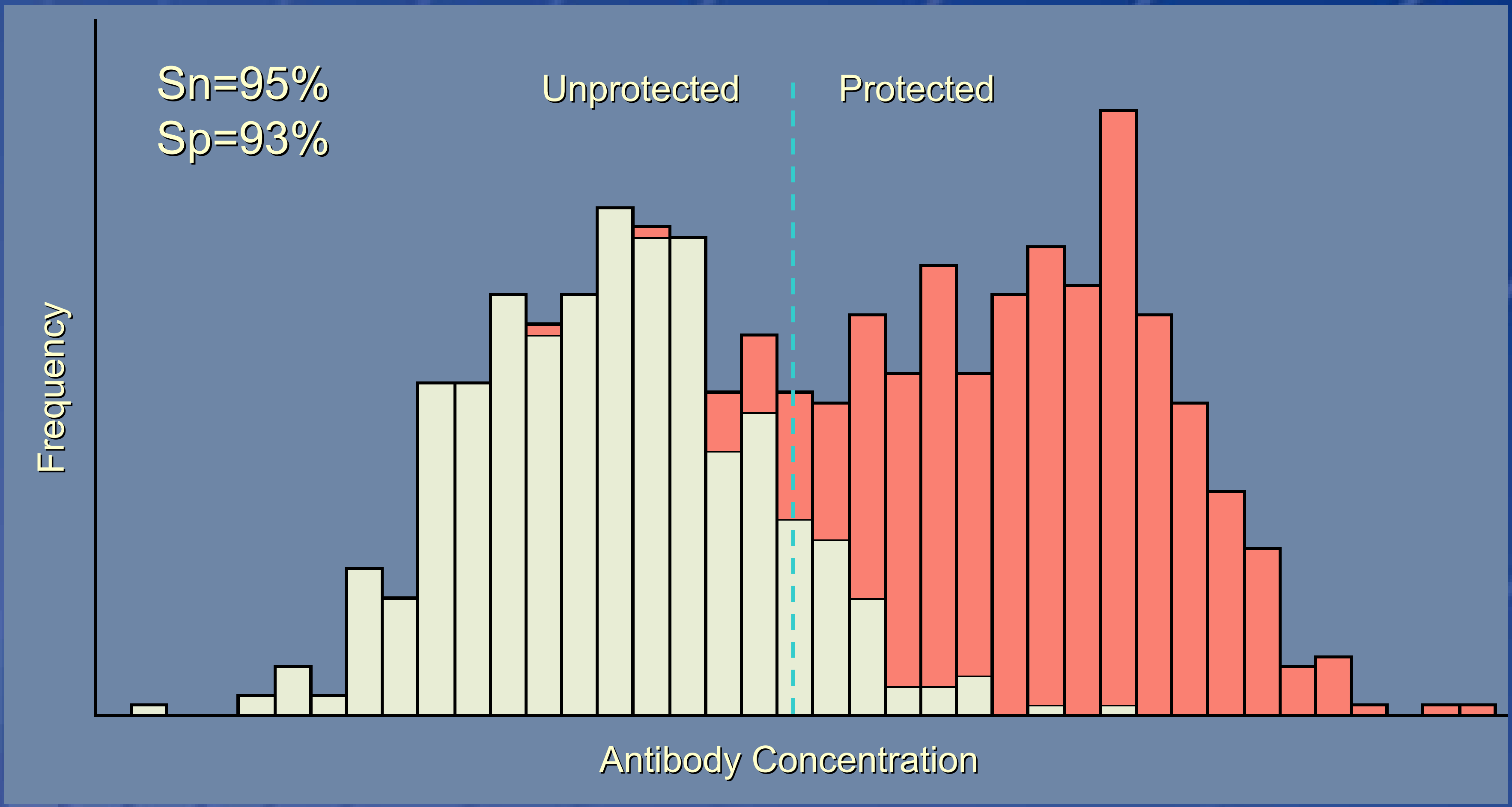


Predicting Protection

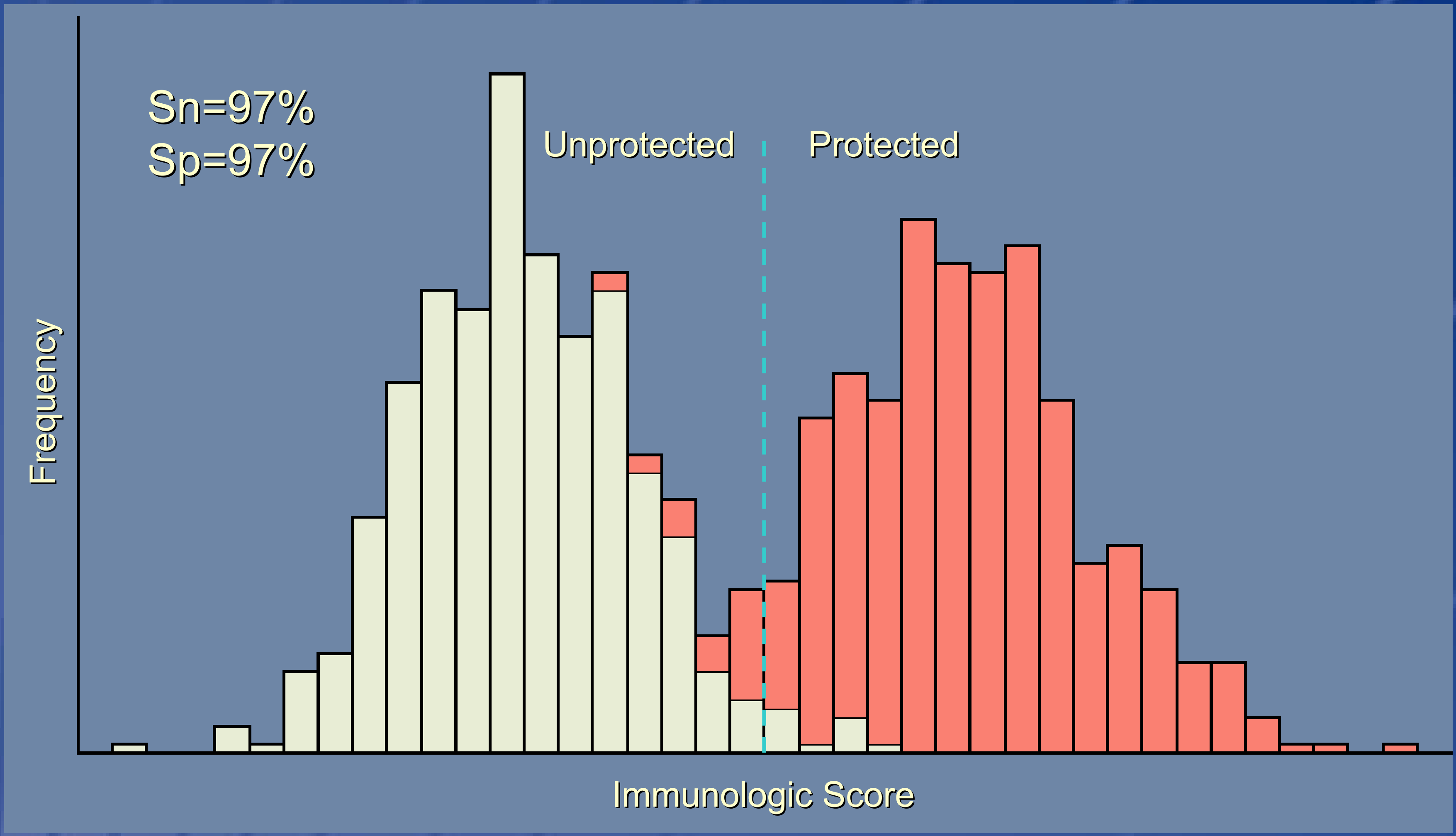


SAFER • HEALTHIER • PEOPLE™

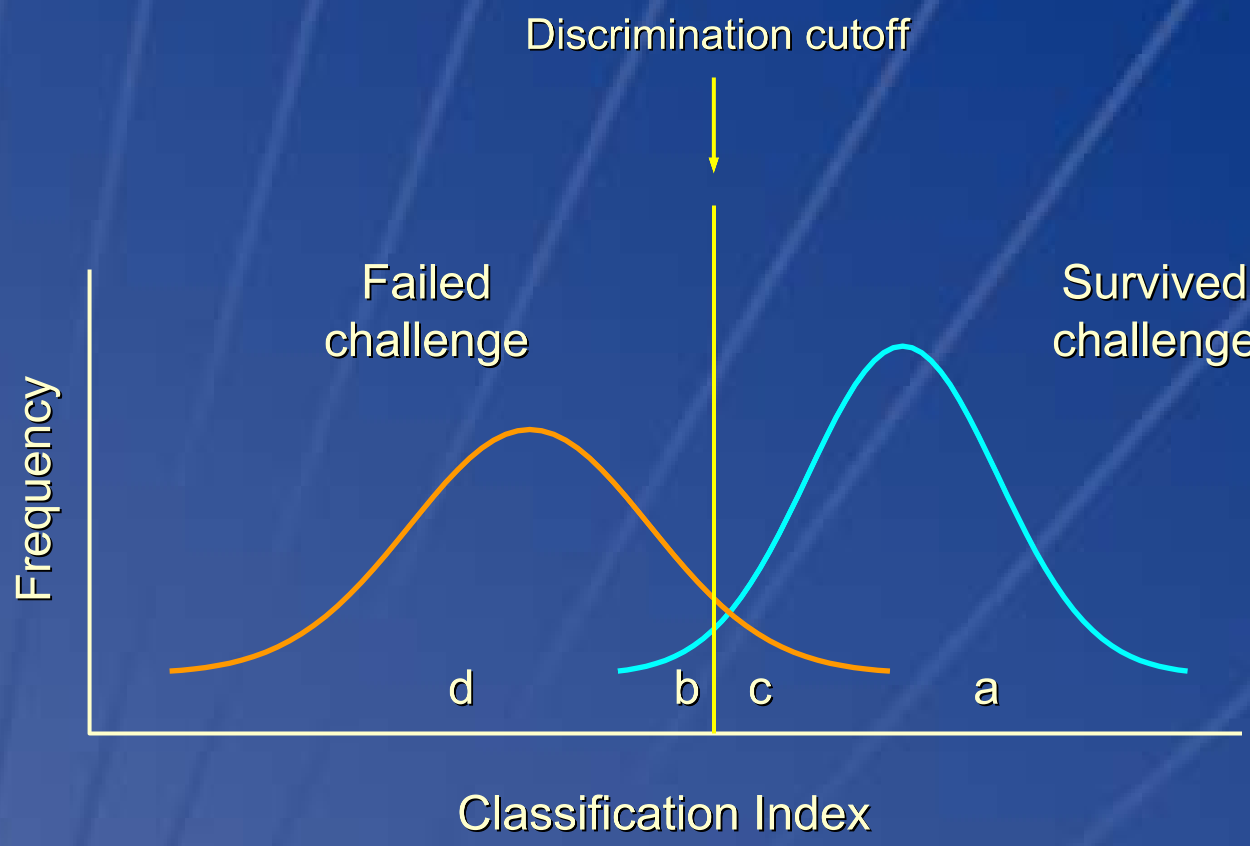
Predicting Protection



Model Application



Model Application



b = survivors incorrectly classified as failed

a = survivors correctly classified as survived

c = failed, incorrectly classified as survived

d = failed, correctly classified as failed

Model Application

		Known Outcome	
		Survived Challenge	Failed Challenge
Discriminant Analysis	Survived Challenge	a	c
	Failed Challenge	b	d

- If survived challenge is defined as a 'positive' event then sensitivity = [a / (a+b)]
- If failed challenge is defined as a 'negative' event then specificity = [d / (c+d)]

Model Application

- Estimate model coefficients
  - Select variables with greatest discriminating power
- Calculate classification indexes for all subjects
- Discrminant cutoff may be placed to maximize sensitivity or specificity

Model Application

- Apply discriminant function to subset of clinical trial participants
- Examine how vaccinated and unvaccinated individuals are classified
- Determine if these variables convey information about vaccine protection

Analytical Model

Outcome	Animal #	Dil.	Assay 1	Assay 2		$Assay_j$
Survived / Died	1	1:40	1, 2, 3, … A_{1k}	1, 2, 3, … A_{2k}	⋯	1, 2, 3, … A_{jk}

j Assays

k Bleeds per animal

$j \times k$ Assay data points per animal

Complete Data

Animal	Assay 1	Assay 2	Assay 3	Assay 4	Assay 5
1	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
2	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
3	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
4	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
5	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
6	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
7	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
8	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
9	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
10	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -

Missing Data

Animal	Assay 1	Assay 2	Assay 3	Assay 4	Assay 5
1	- - - x -	- - - - -	- - - - -	- - - - -	- x - - -
2	- - - - -	- - - - -	- - x - -	- - - - -	- - - - -
3	x - - - -	- - - - -	- - - - -	- - - - -	- - - - -
4	- - - - -	x - - - -	- - - - x	- - - - -	- - - - -
5	- - - - x	- - - - -	- - - - -	- - - x -	- - - - -
6	- - - - -	- - x - -	- - - - -	- - - - -	- - - - -
7	- - - - -	- - - - -	- - - - -	- - - - -	- - x - -
8	- - - - -	- - - - -	- - - - -	- - - x -	- - - - -
9	- x - - -	- - - - x	- - - - -	- - - - -	- - - - -
10	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -

Missing Data

Imputation Strategies

- Recognized by the FDA
- Several methods
  - Last (worst) value carried forward
  - Hot deck
  - Listwise deletion – complete case
  - Mean & regression substitution
  - Pairwise deletion
  - Multiple imputation

Missing Data

Listwise Deletion – Complete Case Analysis

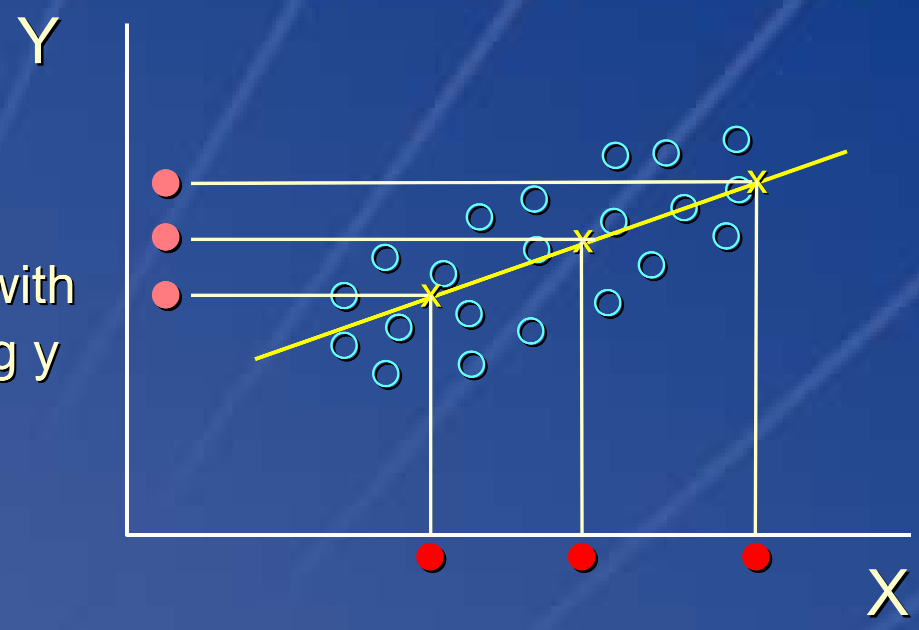
- If data MCAR then probably OK
  - Complete data – random sample
- If data not MCAR then
  - Point estimates biased
  - Inflated variances (smaller n)

Missing Data

Mean & Regression Substitution

Y	X
3	5
4	6
.	2
2	2
7	11
.	6
3	5
1	3
.	4
6	9
$\bar{Y}$	$\bar{X}$

● = obs with missing y



Missing Data

Mean Substitution

- Point estimates may be biased
- Underestimate standard errors

Missing Data

Pairwise Deletion - Available Case Analysis

X	Y
.	2
.	10
.	4
5	6
2	2
7	6
3	2
1	7
6	3
3	.
7	.
5	.

- Green box: For mean X, S_X^2
- Orange box: For mean Y, S_Y^2
- White box: For S_{XY}^2, r_{XY}

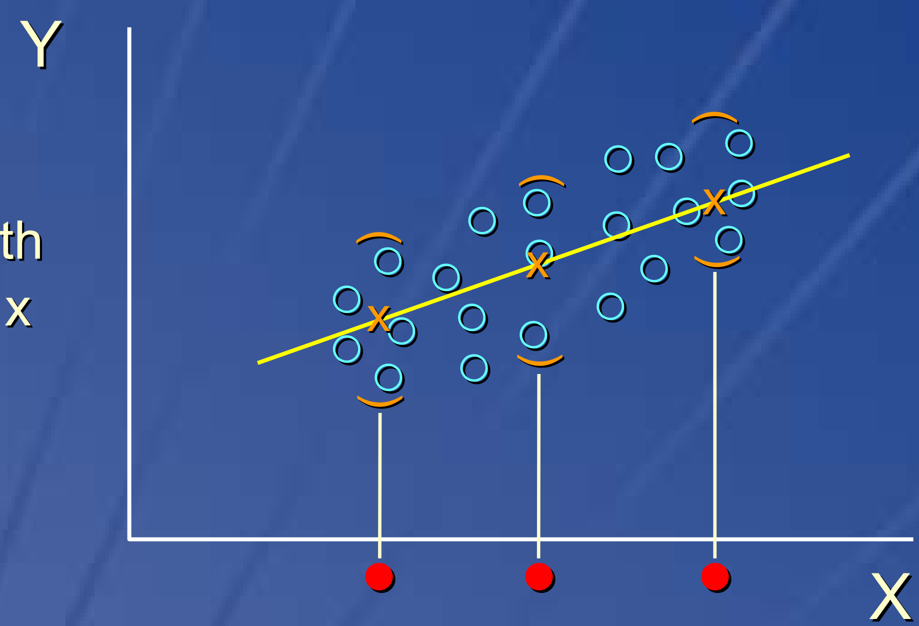
Missing Data

Pairwise Deletion - Available Case Analysis

- Advantages:
  - Easy, but inferences problematic
  - Makes apparent better use of data
- Disadvantages:
  - Correlations may fall outside of (-1,1)
  - Cov. matrix has undesirable properties
  - May be even worse than compl. case

Missing Data

Multiple Imputation



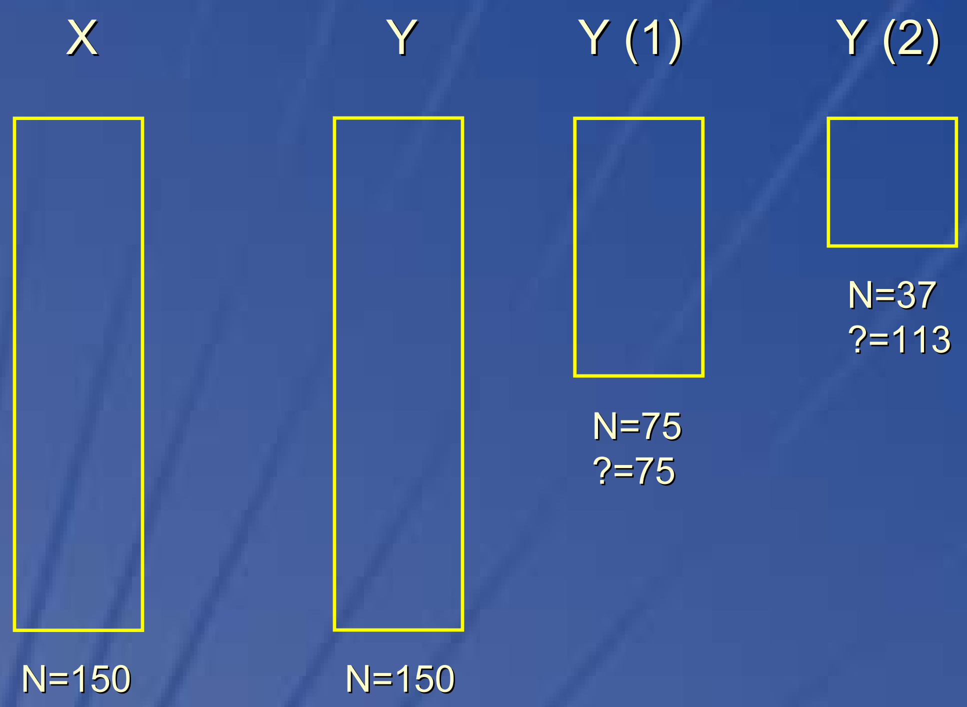
● = obs with missing x

$$y_i = a + bx_i + \varepsilon_i$$

$$y_i = a_j + b_j x_i + \varepsilon_i$$

Missing Data

Multiple Imputation – Example



Y values imputed under 3 correlation structures - low, medium, high

Missing Data

Multiple Imputation – Example

50 imputations (Mean Y ± std. error)

Sample	Correlation: Low	Medium	High
Full (n=150)	29.53 ± 0.47	16.61 ± 0.17	19.48 ± 0.20
1/2 Missing (n=75)	28.61 ± 0.71	16.75 ± 0.23	19.57 ± 0.24
3/4 Missing (n=37)	30.79 ± 0.99	16.44 ± 0.26	19.35 ± 0.24

Missing Data

Multiple Imputation

- Advantages:
  - Estimators consistent, asym. efficient and normal
  - Can be used with any kind of data or model
  - Analysis can be performed in SAS
- Disadvantages:
  - Easy to do in the wrong way
  - May get slightly different answers with each imputation

Missing Data

Multiple Imputation

- Individual data points unimportant
- Primary goal – unbiased estimates
  - Model parameters
  - Standard errors
  - Statistics
- Standard errors reflect degree of 'missingness'