



*Statistical Issues Affecting Biomarker Reliability*

**Power for a Pipeline**

***Steven J. Skates PhD***

***Massachusetts General Hospital and  
Harvard Medical School***

19 April 2013

**Emerging Technologies for 21st Century Clinical Diagnostics: The  
45<sup>th</sup> Annual Oak Ridge Conference**

Perspectives

Possible states of nature

Modeling of pipeline

Review from Statistical Criteria

Power calculations

## Perspective

Quantitative criteria, such as power, provide one source of input into the many criteria for prioritizing candidates for verification

Biology, AIMS, protein biochemistry, etc

Experience as to how to best combine such criteria still being gathered

## Perspective

Statistical modeling of biomarker pipeline is very approximate – first pass approximation

Designed to provide guidance on sample size assuming educated guess at “state of nature”

To answer – “Are we even in the ballpark?”

## Perspective

Why bother?

We have a budget, the budget sets our sample size

What “state of nature” can this sample size detect reliably? (SDs separation, fraction cases producing biomarker)

## Possible states of nature

Bracket by previous experience (e.g. Rx trials)

Separates cases from controls by 1, 2, 3, 4 SDs

Fraction of cases producing biomarker:

10%, 20%, 30%, 50%, 80%

## Possible states of nature

Separates cases from controls by 1, 2, 3, 4 SDs

Fold change if CV between patients = 50%

1 SD = 1.6 fold

2 SD = 2.7 fold

3 SD = 4.5 fold

4 SD = 7.4 fold

4 fold = 2.8 SDs

## Possible states of nature

Separates cases from controls by 1, 2, 3, 4 SDs

Sensitivity at 98% specificity (S98) if all cases produce biomarker, SD cases = 1.5\*SD controls

1 SD = 24%

2 SD = 49%

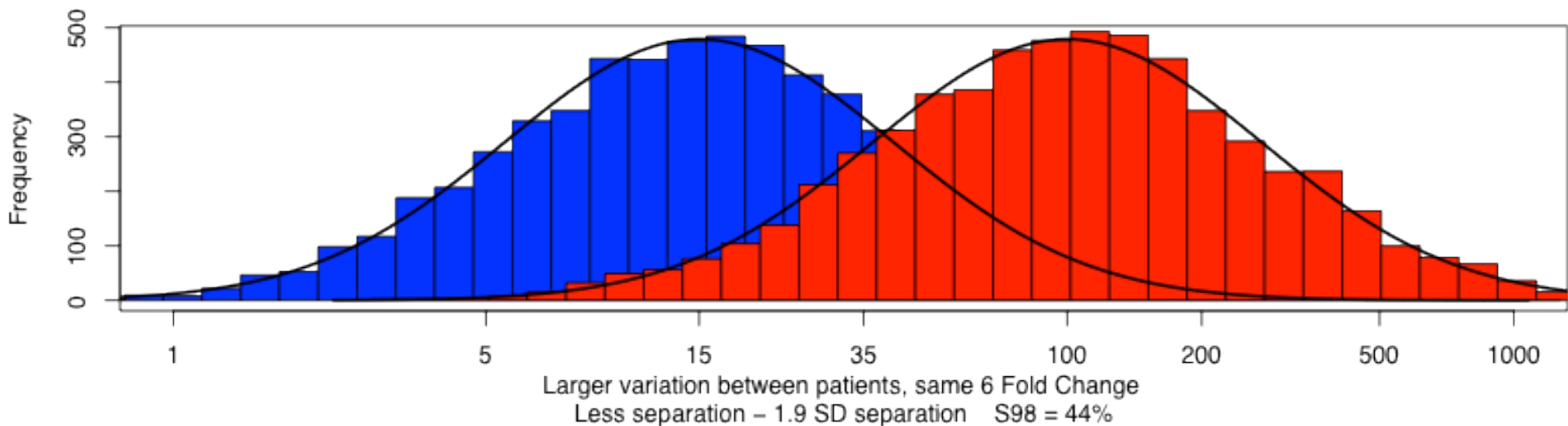
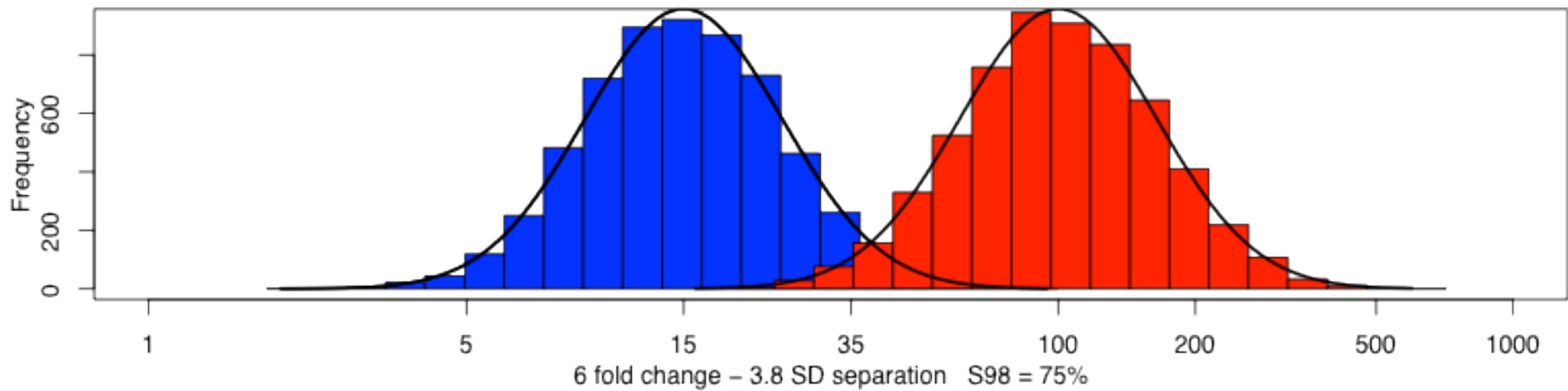
3 SD = 74%

4 SD = 90%



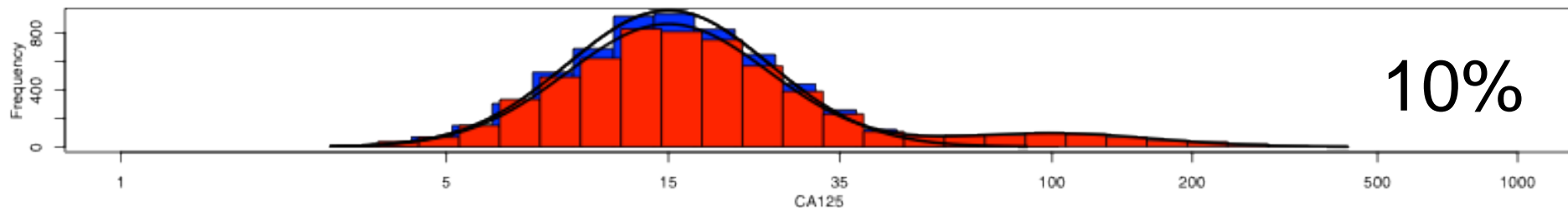
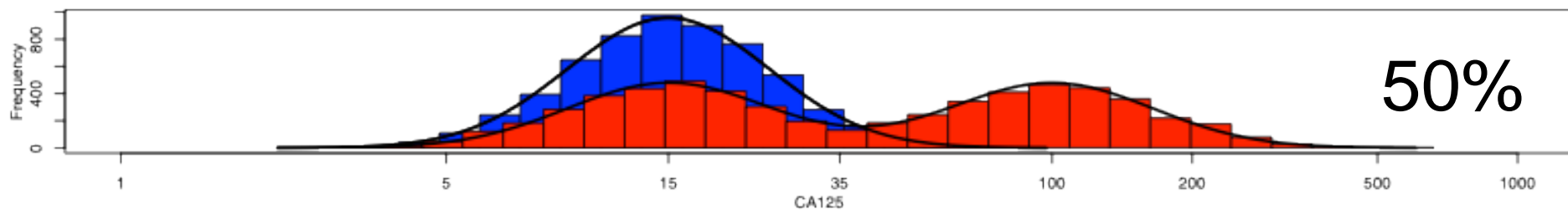
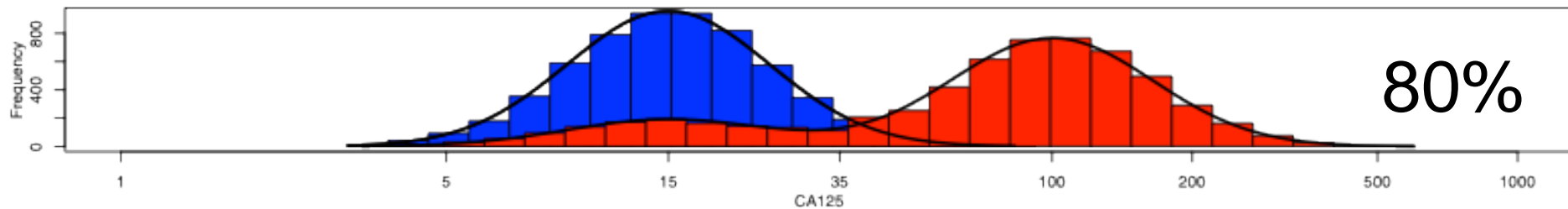
# Early Detection of Ovarian Cancer Biomarkers - Power

## Difference between fold-change and separation



# Early Detection of Ovarian Cancer Biomarkers - Power

Fraction of cases producing biomarker  
e.g. CA125 produced in 80% of ovarian cancers



## Possible states of nature

Given state of nature, and verification/clinical assay throughput, how many samples do I need to analyze to reliably detect such a biomarker?

Reach clinical validation stage with 80% probability

Pass through discovery stage with 90% probability &

Pass through verification stage with 90% probability

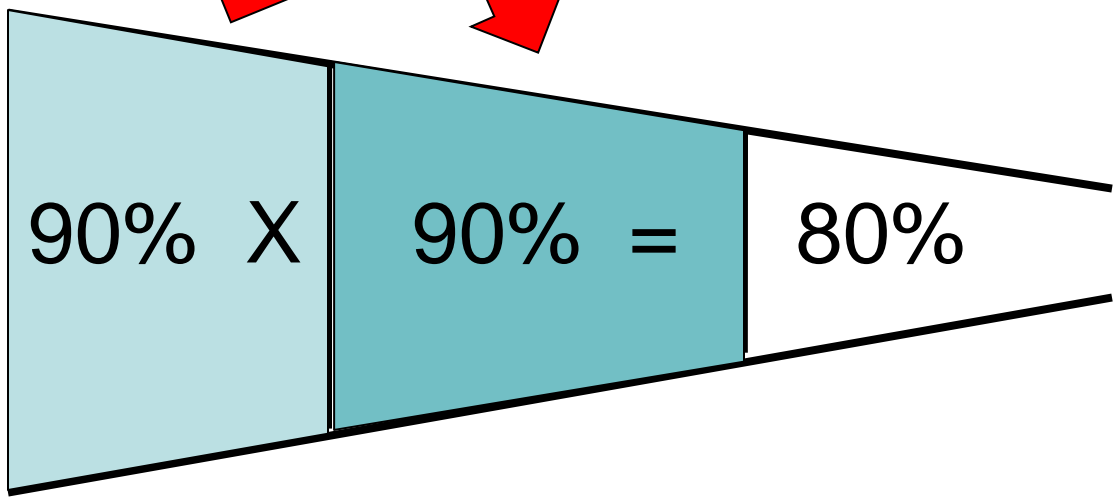
# Statistical Criteria by Stage and Transition for CPTAC Pipeline

## Set power to reach Clinical Validation at 80%

CPTAC Pipeline

CPTAC

Many candidates



A few good biomarkers

Discovery

Verification

Clinical Validation

Low throughput,  
many candidates  
High CV

Moderate throughput,  
moderate # candidates  
Moderate CV

High throughput,  
small # candidates  
Low CV

Power is derived from increased S/N

Signal increased by increasing:

1. Separation between cases and controls - #SDs
2. Proportion of cases producing biomarker

Noise changed by CV of measurement process

3. Throughput of candidates in verification stage

# Statistical Criteria by Stage and Transition for CPTAC Pipeline

## An Approach to Determining if a Biomarker is Clinically Useful

1. Define “Intended Use”
2. Define clinical pathway
3. State minimum benefits to harms ratio
  - a) Benefit of true positive
  - b) Benefit of true negative
  - c) Harm of false positive
  - d) Harm of false negative
4. Derive required specificity (ReqSP) assuming 100% sensitivity to achieve acceptable B/H ratio, within clinical pathway
5. Impact of each biomarker is increase in sensitivity at clinically useful specificity (ReqSP)
6. Evaluate sensitivity of biomarker (panel) at ReqSP

# Statistical Criteria by Stage and Transition for CPTAC Pipeline

## Required Specificity (ReqSP)

Annual incidence of OVCA in postmenopausal women  
1 in 2,500

Desired benefit/harm: 1 OVCA in 5 surgeries

500 fold increase

Ultrasound reduces false positives by 10-fold

Require blood test to reduce by 50-fold

## Biomarker Criterion for early detection of OVCA

Required False Positive Rate: 1 in 50 = 2%

Required Specificity (ReqSP): 98%

Sensitivity at 98% Specificity



# Statistical Criteria by Stage and Transition for CPTAC Pipeline

Estimate Normal distribution - mean, std deviation for intensity (log) for each protein

Controls = blue = 0,      Cases = red = 1

Q98 = 98% quantile for controls =  $\text{mean}_0 + 2 * \text{STD}_0$

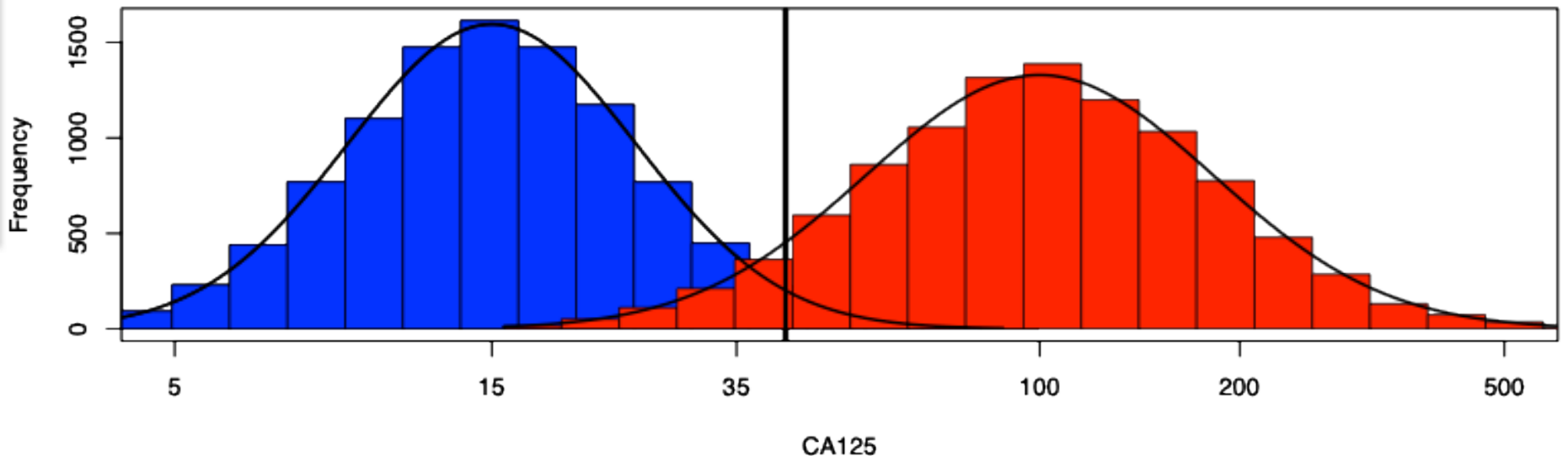
S98 = Sensitivity at Q98 =

$$\text{Prob Normal} > Z = (\text{Q98} - \text{mean}_1) / \text{STD}_1$$

# Statistical Criteria by Stage and Transition for CPTAC Pipeline

98% left of cut-point  
 $X = \text{avg} + 2 \text{ SDs}$

75% right of cut-point  
%ile of Normal Dist ( $X \mid \text{avg}, \text{SD}$ )



Sens = 75%      at 98% Spec      S98 = 75%

6-fold increase – depends on biological variability  
3 standard deviations separates cases from controls

## Reaching Verification

10 cases and 10 controls,  
biomarkers expressed in 80% of cases, and separated by 2 SDs

25 cases and 25 controls,  
biomarker expressed in 50% of cases, and  
separated from controls by 2 SDs,  
expressed in 80% of cases, and separated by 1 SD.

50 cases and 50 controls,  
biomarker expressed in 20% of cases, chance identified at most 53%  
for biomarkers where cases separated by 3 SDs

biomarkers expressed in 10% of cases - unlikely to be detected

## Reaching clinical validation

50 cases and 50 controls

20% of cases producing biomarker, and  
2 SDs apart fail to achieve a high probability

250 cases vs. 250 controls

chance < 90% only if cases and controls are 1.0 SDs apart, **and** proportion of cases producing biomarker is 10% or less

## Conclusions - Power

Statistical modeling produces a first approximation to CPTAC pipeline

Set criteria at each stage based on clinical benefit/harm ratio for intended use

Power of 80% to reach clinical validation stage achieved by attaining 90% power for discovery and verification stages

Factors affecting sample size

1. Separation between cases & controls
2. Fraction of cases producing biomarker
3. Number of assays for verification, clinical validation to be developed