Design of Early Validation Trials of Biomarkers

Daniel Normolle, Ph.D.

Great Lakes-New England Clinical Epidemiological Center
Early Detection Research Network

Department of Radiation Oncology
Comprehensive Cancer Center Biostatistics Unit
University of Michigan

1 Example

We have laboratory results of an interesting marker:

- Twenty normals and twenty cancers of unknown provenance
- Unblinded laboratory analysis
- Two-standard deviation difference between samples

What is the next step?

1 Example (continued)

Weak idea for the design:

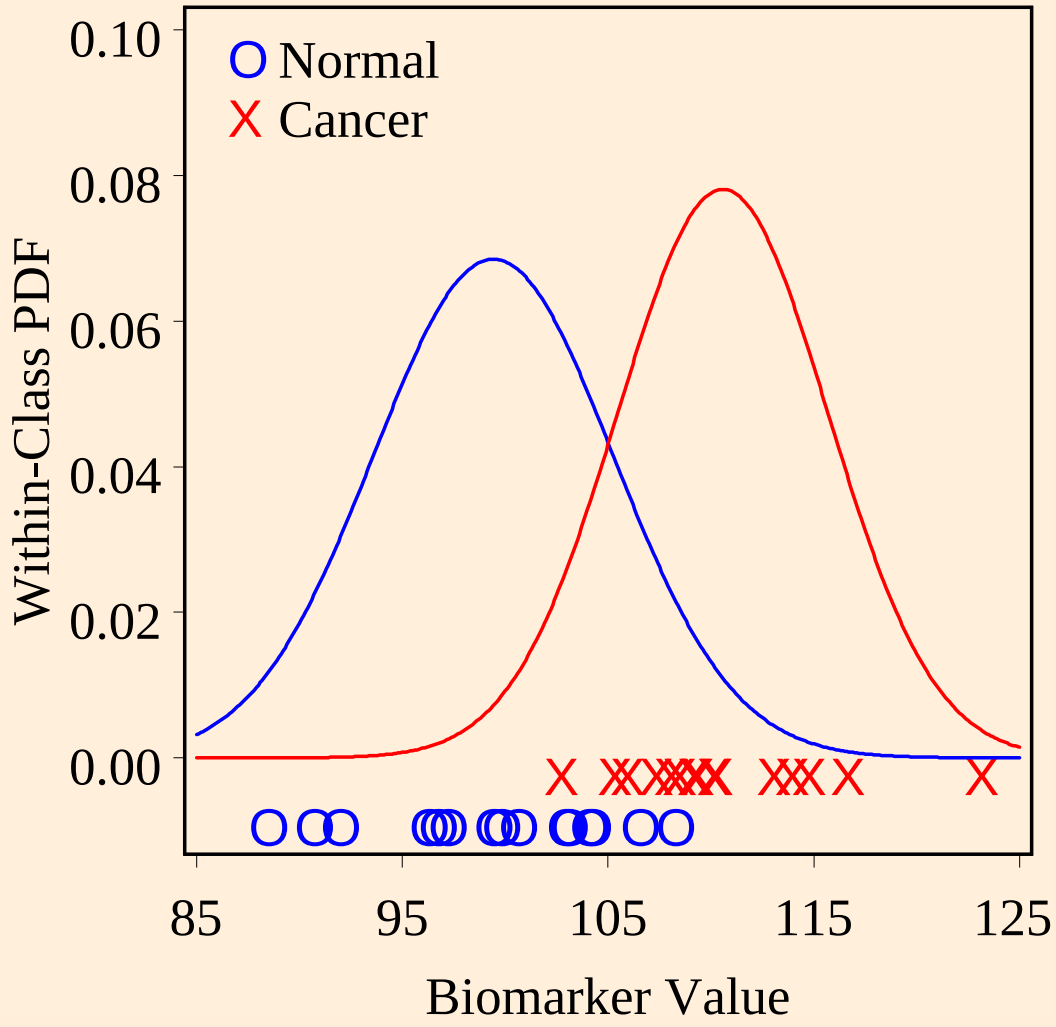
Twenty well-characterized cancers and normals

Two-sample t-test

80% power for two-standard-deviation difference

1 Example (continued)

Results of the trial:



1 Example (continued)

Why is the design weak?

- Is a test that identifies cancers useful?
- Two standard deviations may not be sufficient
- Testing sample = training sample
- Test sample is unblinded
- How does the design relate to the decision?

2 Design Criteria

What are the goals of the marker?

- Screen of an untested population
  - No other information about subjects
  - Potentially tens of millions of tests (e.g., PSA)
- Screen of a clinical population
  - Demographic, risk, other marker data
  - Fewer tests, high-risk subjects
- Combine with other markers to make a panel

2 Design Criteria (continued)

What are the goals of the study?

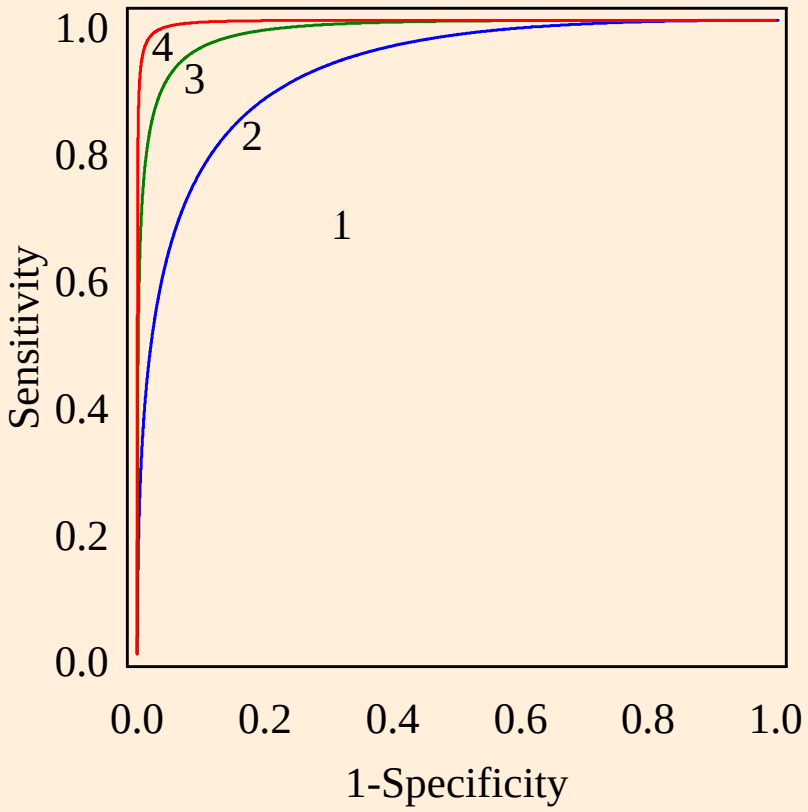
- Feasibility/Technical
  - Further tweaking of assay
  - Characterize features of the assay
  - Identify threats to assay validity (e.g., sample processing)
- Developmental/Pivotal
  - Estimate sensitivity and specificity
  - Verify the validity of the marker
  - Determine quality of reported signal (pathway activation in tissue)

3 Estimate Sensitivty and Specificity

What is sufficient sensitivity and specificity?

Example: ROC curves for normally distributed markers

Number indicates separation in standard deviations



3 Estimate Sensitivty and Specificity (continued)

Minimum cost classification rule:

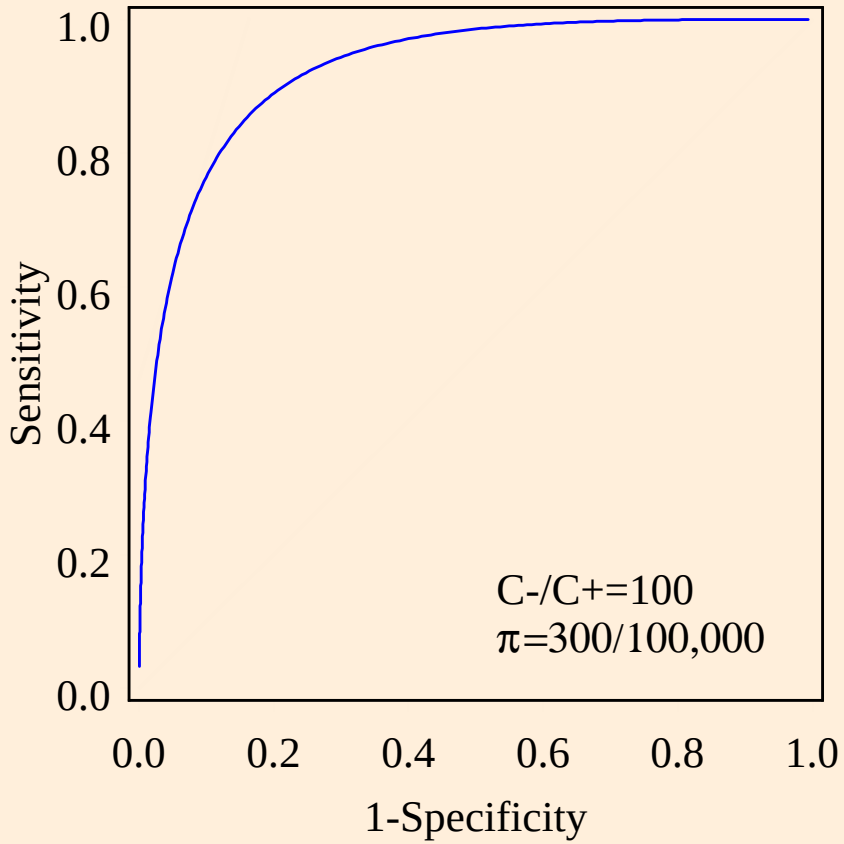
$$\text{If} \quad \frac{\hat{\mathrm{P}}(\boldsymbol{x}|+)}{\hat{\mathrm{P}}(\boldsymbol{x}|-)} > \frac{(1-\pi)}{\pi} \cdot \frac{C_+}{C_-} \quad \text{then} \quad \text{Classify } \boldsymbol{x} \in +$$

$$\text{Otherwise} \qquad \text{Classify } \boldsymbol{x} \in -$$

3 Estimate Sensitivty and Specificity (continued)

The Minimum cost point can be identified on the ROC curve:

$$\text{ROC}' = \frac{(1-\pi)}{\pi} \cdot \frac{C_+}{C_-}$$



3 Estimate Sensitivty and Specificity (continued)

Classification criteria (RHS of the minimum cost rule):

C_-/C_+	π 300/100,000	100/100,000	30/100,000	10/100,000
300	1.11	3.3	11.1	33.3
100	3.3	9.99	33.3	99.99
30	11.1	33.3	111.1	333.3
10	33.2	99.9	333.2	999.9

Markers intended for population screens will require very high specificities

3 Estimate Sensitivty and Specificity (continued)

If the marker is intended to be used as a population screen, the minimum cost rule is appropriate, and will tend to recommend very high specificity

In clinical decision-making, risk reduction may be a more relevant criterion, so the bar may be set lower

In either case, a pivotal trial will recommend a marker for further development if it meets minimum criteria for sensitivity and specificity:

- Determine minimum criteria for sensitivity and specificity
- Design the trial as parallel single-sided tests of the null hypothesis of inadequate sensitivity and specificity
- Power the tests appropriately

3 Estimate Sensitivty and Specificity (continued)

Using the training sample as the testing sample tends to overestimate the sensitivity and specificity

- Split the testing and the training sample
- Blind the testing sample
- Cross-validation

3 Estimate Sensitivty and Specificity (continued)

Simulation Example: Generate 1000 data sets with 15 controls and 15 cases, one discriminating variable and five noise variables

		1		1+5 Noise	
		Sens.	Spec.	Sens.	Spec.
Method	Resubstitution	.696	.694	.767	.772
	Holdout	.683	.694	.630	.638
	Crossvalidation	.689	.686	.625	.629

While there are systematic criteria for the size of the testing sample, no such standards exist for the training sample

4 Verify the validity of the marker

What are the biases in testing a potential screen or clincal marker?

- Cases that differ from controls in a way that affects the test
- Case and control samples that have been treated differently
- Wrong type of cases (pre-cancers versus cancers)
- Wrong type of disease
- Wrong marker

5 Determine quality of reported signal

Tumor tissues will not be assayed in real markers

Ideally, markers signalling different pathways will be combined

If there are multiple potential pathways, and the sensitivity of a marker is 30%, is it because:

- The pathway is only activated in 30% of cases, or
- The activation signal in the tumor isn't detected in the tested product?

Estimate the sensitivity and specificity relative to activated tumor, rather than patients

Pivotal trial will accept a sample marker for use in a panel if it is sufficently senstitive and specific for the marker in target tissue

6 Summary

A pivotal study should produce results that directly inform the next decision to be made

Minimal criteria for sensitivity and specificity are the most likely for determining sample size

A marker intended to be used as a population screen must be very sensitive and extremely specific

Sensitivity and specificity should not be directly estimated from the training sample

A blinded testing sample is more convincing

As the dimensionality of the marker increases, the risk of over-fitting increases

A sample marker that indicates a single tissue pathway should be validated against that pathway