

Promises and limitations of cancer biomarkers

***Biospecimen Research Network Symposium
March 13-14, 2008***

David F. Ransohoff, MD
Departments of Medicine and Epidemiology
University of North Carolina at Chapel Hill

Promises and limitations of cancer biomarkers

1. 'Disconnect' between claims, products
2. Fundamental problems in study design
 - chance
 - bias
3. How to address (1), (2)
 - role of specimens
 - (other)

'Disconnect' between claims, products

.... for markers of cancer diagnosis, prognosis,
response to therapy

Big claims, little product

Proteomics for diagnosis

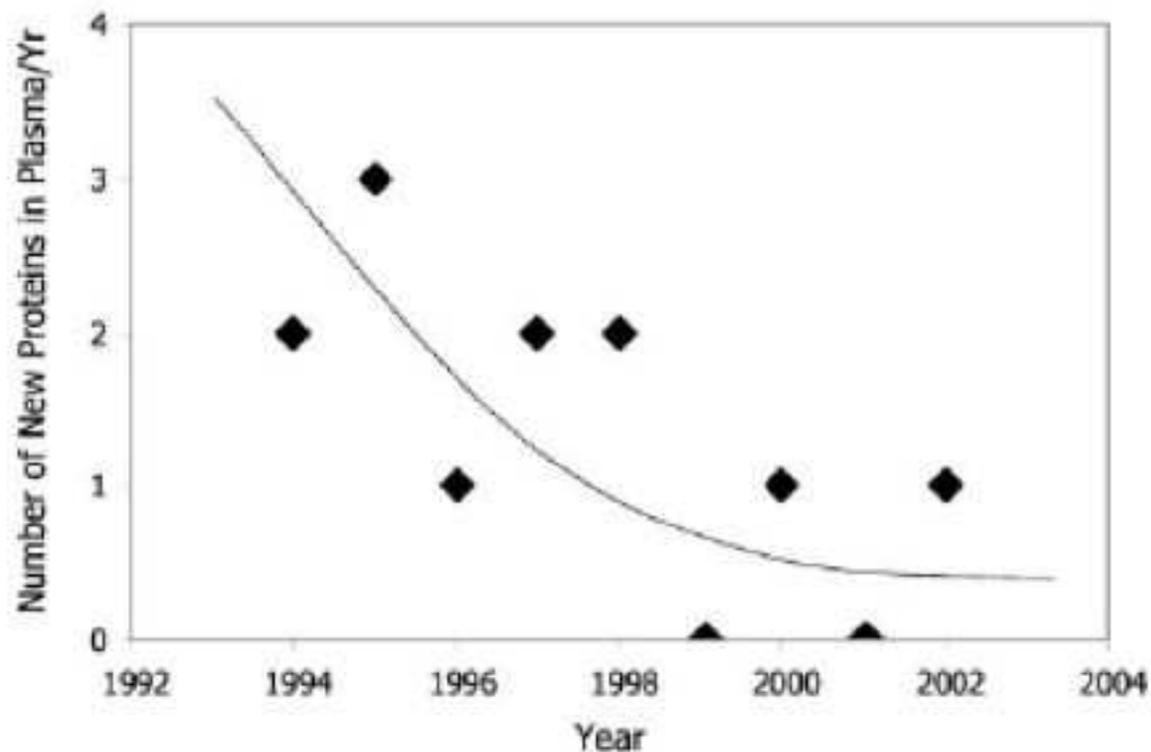


FIG. 5. **The declining rate of introduction of new protein tests.** The data of Fig. 4 are plotted to indicate the rate of introduction of new protein analytes in FDA-approved clinical tests.

Big claims, little product

Proteomics for diagnosis

Asking for the impossible?

I do not understand the surge of activity in the search for biomarkers. The available evidence suggests that proteomics, despite almost a billion dollars investment, has so far failed to deliver any new biomarkers or commensurate returns. Many flagship companies have failed. More worrying, flawed studies, poor business models and exaggerated expectation will take time to reverse.

Walter Blackstock
University of Sheffield
(at Royal Society of Chemistry; London 2006)

Big claims, little product

RNA expression genomics for prognosis

Science, 10.22.04

SPECIAL SECTION

GENES IN ACTION

NEWS

Getting the Noise Out of Gene Arrays

Thousands of papers have reported results obtained using gene arrays, which track the activity of multiple genes simultaneously. But are these results reproducible?

When Margaret Cam began hunting for genes that are turned up or down in stressed-out pancreas cells a couple of years ago, she wasn't looking for a scientific breakthrough. She was shopping. As director of a support lab at the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), she wanted to test-drive manufactured devices called microarrays or gene arrays that measure gene expression; she had her eye on three different brands. These devices are hot, as they provide panoramic views of the genes that are active in a particular cell or tissue at a particular time.

being highly up- or down-regulated.

The disharmony appears in a striking illustration in Cam's 2003 paper in *Nucleic Acids Research*. It shows a Venn diagram of overlapping circles representing the number of genes that were the most or least active on each device. From a set of 185 common genes that Cam selected, only four behaved consistently on all three platforms—"very low concordance," she said at an August forum in Washington, D.C., run by the Cambridge Healthtech Institute, based in Newton Upper Falls, Massachusetts. Using less rigorous criteria, she found about 30% agreement—but never more than 52% between two brands. "It was nowhere near what we would expect if the probes were assaying for the

he gathered on kidney tumor cells, the less significant it seemed.

But those who have persevered with gene expression arrays attribute such problems to early growing pains. They claim that experienced labs are already delivering useful clinical information—such as whether a breast cancer patient is likely to require strong chemotherapy—and that new analytical methods will make it possible to combine results from different experiments and devices. Francis Barany of Cornell University's Weill Medical College in New York City insists that arrays work well—if one digs deeply into the underlying biology.

Imperfections

Digging into the biology is just what Cam did after her experiments produced reams of discordant data. She and colleagues in Marvin Gershengorn's group at NIDDK wanted to pick out a set of key genes active in pancreatic tumor cells undergoing differentiation. From there, they meant to go on to examine how islet cells develop.

Rapid Increase in Microarray Publications

3000

3000

No common standards are used across platforms, so data are

Promises and limitations of cancer biomarkers

1. 'Disconnect' between claims, products
2. **Fundamental problems in study design**
 - chance
 - bias
3. How to address (1), (2)
 - role of specimens
 - (other)

Principles of science



Cargo Cult Science

by RICHARD P. FEYNMAN

Some remarks on science, pseudoscience,
and learning how to not fool yourself.
Caltech's 1974 commencement address.

Feynman. *Engineering and Science* 1974:10-13.

Principles of science

“Details that could throw doubt on your interpretation must be given, if you know them.... [I]f you know anything at all wrong, or possibly wrong--to explain it.”

Feynman 1974

Ask ***‘what might be wrong.’***

Ask 'what might be wrong'

John Platt said this is the reason “why...some fields advance faster than others.” (Strong inference. Science 1964;146:347)

experimental test.” Or “[o]n any given morning the blackboards of Francis Crick or Sidney Brenner... [will show] the hot new result just up from the laboratory or just in by letter or rumor. On the next line will be two or three alternative explanations, or a little list of ‘what he did wrong.’ Underneath... a series of suggested experiments or controls that can reduce the number of possibilities” [94]. Platt was saying that progress is based on considering alternative explanations and avoiding overinterpretation.

Promises and limitations of cancer biomarkers

1. 'Disconnect' between claims, products
2. **Fundamental problems in study design**
 - chance
 - bias
3. How to address (1), (2)
 - role of specimens
 - (other)

Chance as a threat to validity of clinical research study design

Definition: In multivariable predictive models, overfitting (a problem of 'chance') occurs when large N of predictor variables is fit to a small N of subjects. A model may "fit random variations within the original data that do not represent true relationships that hold for independent data."
(Simon, JNCI 2003)

Consequence: Results not reproducible in independent group.

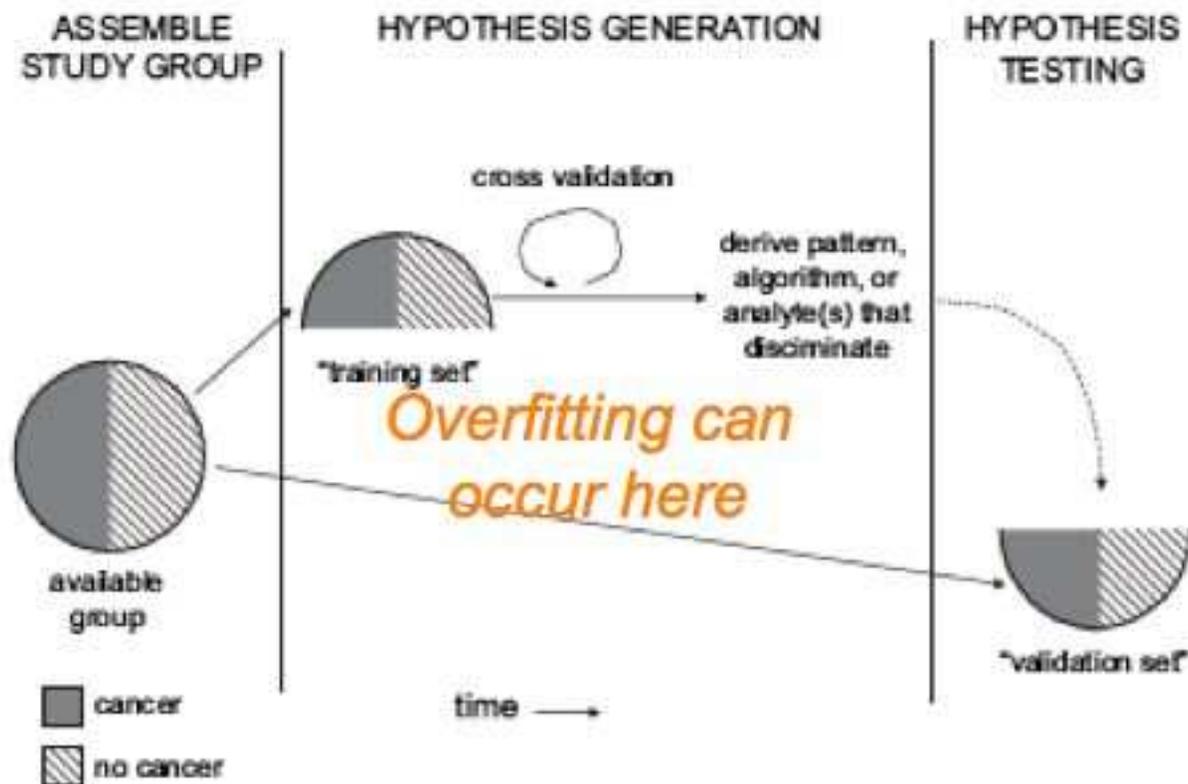
Method to demonstrate that overfitting did not occur:
assess reproducibility in independent group.

Assess reproducibility in totally independent group to show overfitting (chance) did not occur

Ransohoff. Nat Rev Cancer 2004;4:309

Ransohoff. J Clin Epidemiol 2007;60:1205

Clarke. Nat Rev Cancer 2008;8:37



Chance as a threat to validity of clinical research study design

Status in 2008

- 'Overfitting' is *rookie mistake*:
 - easily recognized
 - avoided by showing discrimination in totally-independent specimens
- Situation improving: editors, others become aware

Promises and limitations of cancer biomarkers

1. 'Disconnect' between claims, products
2. **Fundamental problems in study design**
 - chance
 - bias**
3. How to address (1), (2)
 - role of specimens
 - (other)

Threats to validity:

Bias is the main threat in clinical research

Definition

Bias: systematic difference between compared groups; comparison gives distorted answer.
I.e., wrong, misleading

Ransohoff. Nat Rev Cancer 2005;5:142
Ransohoff. J Clin Epidemiol 2007;60:1205

Bias - Example #1

Mechanisms of disease

Lancet 2002; **359**: 572–77

🕒 Use of proteomic patterns in serum to identify ovarian cancer

Emanuel F Petricoin III, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, Charles Simone, David A Fishman, Elise C Kohn, Lance A Liotta

Summary

Background New technologies for the detection of early-stage ovarian cancer are urgently needed. Pathological changes within an organ might be reflected in proteomic patterns in serum. We developed a bioinformatics tool and used it to identify proteomic patterns in serum that distinguish neoplastic from non-neoplastic disease within the ovary.

Methods Proteomic spectra were generated by mass spectroscopy (surface-enhanced laser desorption and ionisation). A preliminary “training” set of spectra derived from analysis of serum from 50 unaffected women and 50 patients with ovarian cancer were analysed by an iterative searching algorithm that identified a proteomic pattern that completely discriminated cancer from non-cancer. The discovered pattern was then used to classify an independent set of 116 masked serum samples: 50 from women with ovarian cancer, and 66 from unaffected women or those with non-malignant disorders.

Findings The algorithm identified a cluster pattern that, in the training set, completely segregated cancer from non-cancer. The discriminatory pattern correctly identified all 50 ovarian cancer cases in the masked set, including all 18 stage I cases. Of the 66 cases of non-malignant disease, 63 were recognised as not cancer. This result yielded a sensitivity of 100% (95% CI 93–100), specificity of 95% (87–99), and positive predictive value of 94% (84–99).

Interpretation These findings justify a prospective population-based assessment of proteomic pattern technology as a screening tool for all stages of ovarian cancer in high-risk and general populations.

Lancet 2002; **359**: 572–77

Mechanisms of disease

🕒 Use of proteomic patterns in serum to identify ovarian cancer

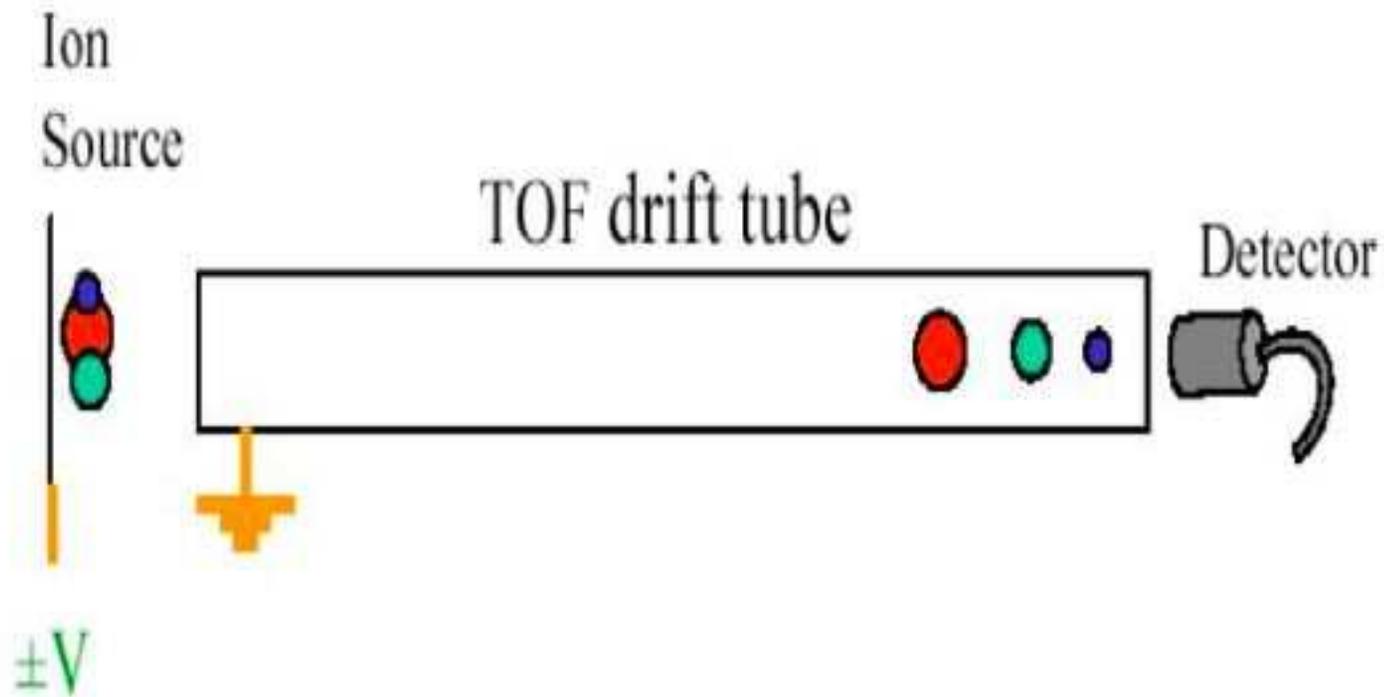
Emanuel F Petricoin III, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, Charles Simone, David A Fishman, Elise C Kohn, Lance A Liotta

Methods

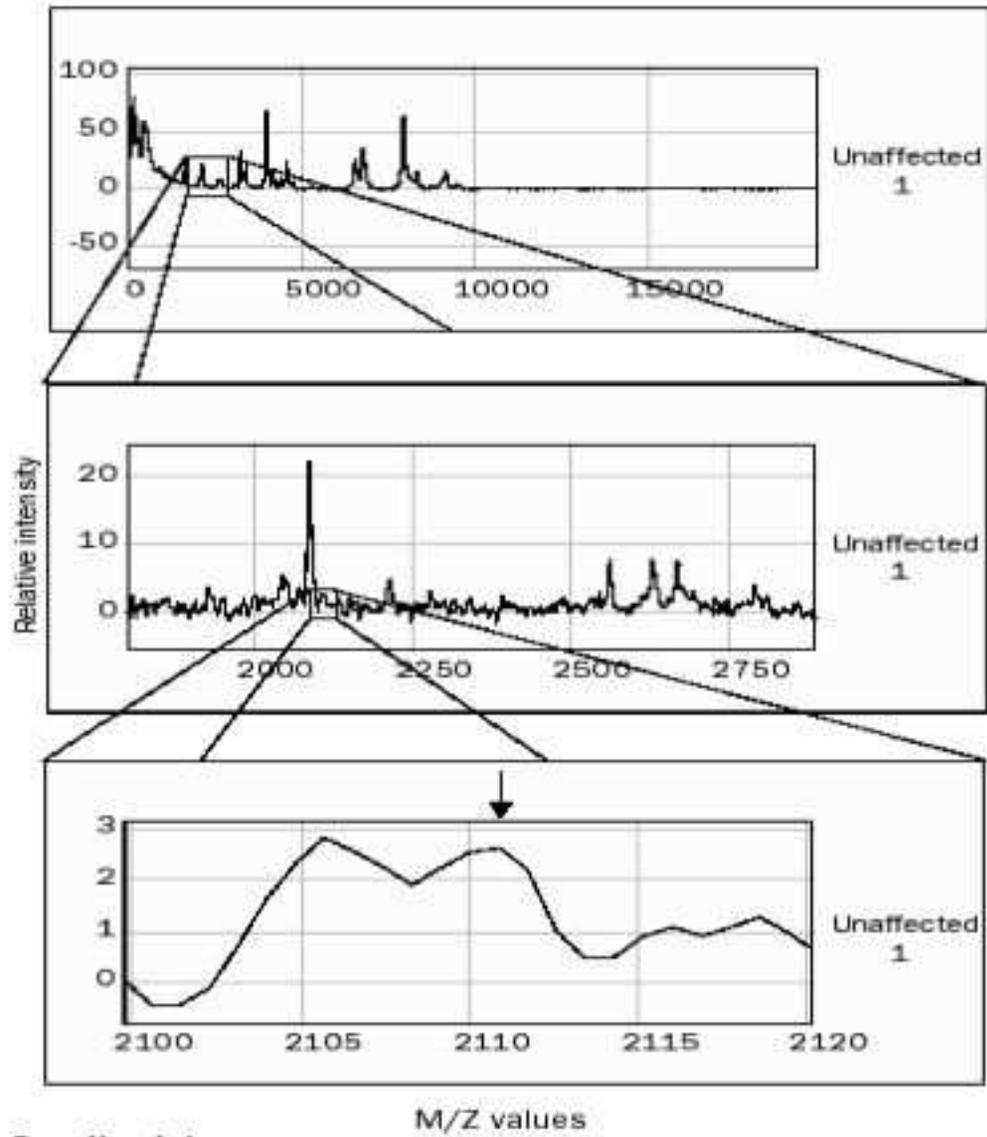
- Subjects: persons with/without ovarian cancer
- Sera were analyzed (training set) to derive pattern
- Pattern was applied to independent validation set

A mass analyzer

(Glish, Nat Rev 2003)



Chromatogram



Mechanisms of disease

Lancet 2002; **359**: 572–77

🕒 Use of proteomic patterns in serum to identify ovarian cancer

Emanuel F Petricoin III, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, Charles Simone, David A Fishman, Elise C Kohn, Lance A Liotta

Results:

“The discriminatory pattern correctly identified all 50 ovarian cancer cases....

[for] a sensitivity of 100%... specificity of 95%...”

Does bias explain some serum proteomics results for ovarian cancer?

(Keith Baggerly's proposal, as reported in Nature news 2004)

Was bias introduced by 'run order' of specimens?

If cancers and non-cancers are run on different days and if the mass spec 'drifts' over time, then ***non-biologic 'signal,' associated with Ca vs no-Ca, is hard wired into results.***

Does bias explain some serum proteomics results for ovarian cancer?

(Keith Baggerly's proposal, as reported in Nature news 2004)

Was bias introduced by 'run order' of specimens?

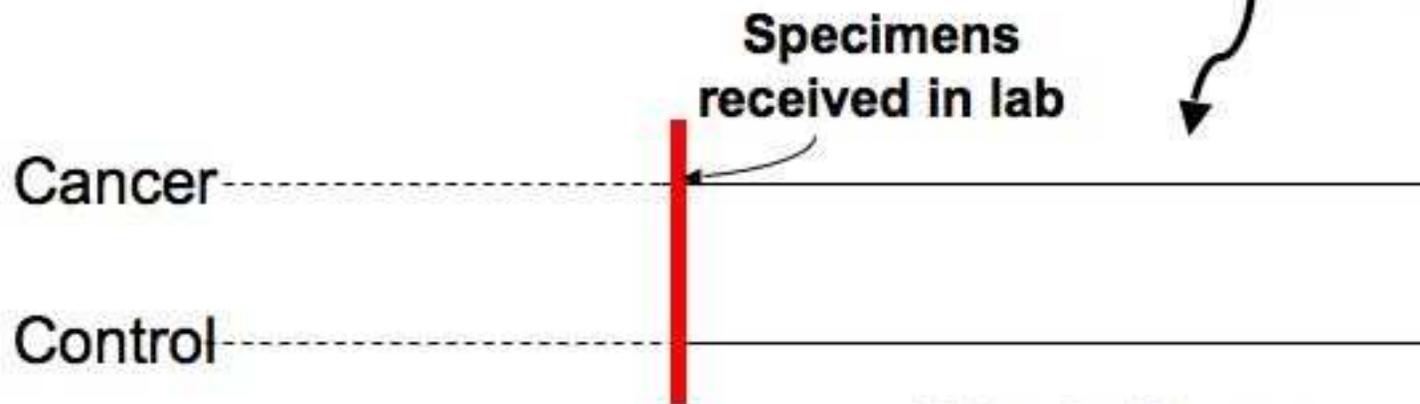
If cancers and non-cancers are run on different days and if the mass spec 'drifts' over time, then ***non-biologic 'signal,' associated with Ca vs no-Ca, is hard wired into results.***

Keith Baggerly has been called 'forensic statistician:' someone who, after study is published, tries to understand what was done.

Many sources of bias

in study of diagnostic test, prognosis, or response to therapy

**After specimens are received,
differences occur in handling:
time, place, etc.**



**Discipline:
laboratory science**

Bias - Example #2

Differential exoprotease activities confer tumor-specific serum peptidome patterns

Josep Villanueva, David R. Shaffer, John Philip, Carlos A. Chaparro, Hediye Erdjument-Bromage, Adam B. Olshen, Martin Fleisher, Hans Lilja, Edi Brogi, Jeff Boyd, Marta Sanchez-Carbayo, Eric C. Holland, Carlos Cordon-Cardo, Howard I. Scher, and Paul Tempst

J Clin Invest 2006;116:271

i.e., Peptide patterns are
~100% sensitive, specific for prostate cancer.

Bias may explain 'discrimination'

Compared groups are different:

- Cancer: mean age 67 y.o.; 100% men

Bias may explain 'discrimination'

Compared groups are different:

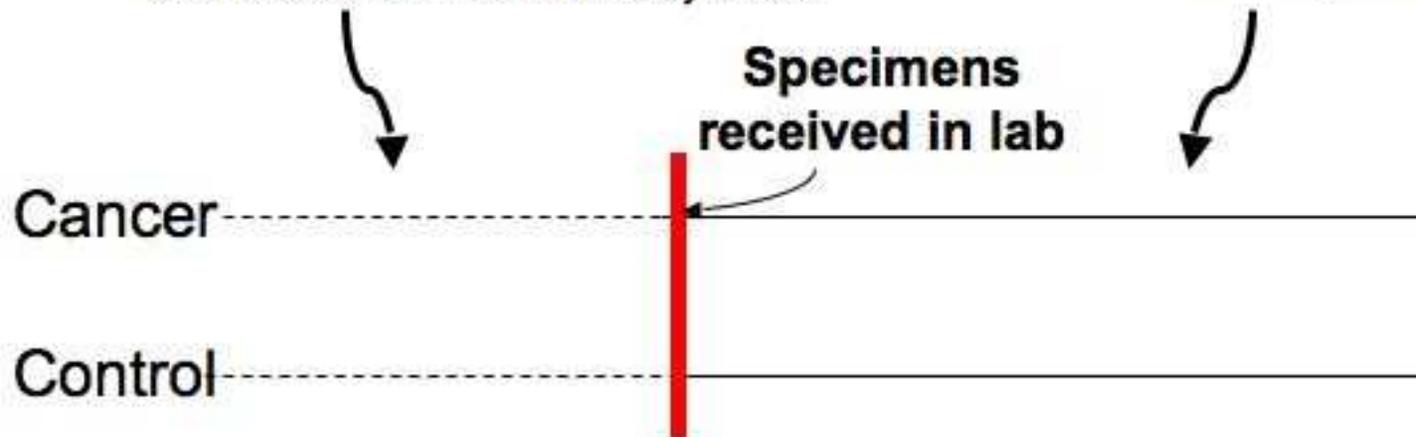
- Cancer: mean age 67 y.o.; 100% men
- Control: mean age 35 y.o.; 58% women

Many sources of bias

in study of diagnostic test, prognosis, or response to therapy

**Before specimens are received,
differences occur in demographics,
collection methods, etc.**

**After specimens are received,
differences occur in handling:
time, place, etc.**



**Discipline:
*(clinical) study design;
clinical epidemiology***

**Discipline:
laboratory science**

Threats to validity:

Bias is the main threat in clinical research

How to address bias (goal):

- a. ***Design (avoid bias)***
- b. **Conduct (learn if bias occurred)**
- c. **Interpretation (was bias important?)**
- d. **Reporting (transparency)**

Ransohoff D. Nat Rev Cancer. 2005;5:142-9

Feynman: “details that could throw doubt on your interpretation... if you know them.... If you know anything at all wrong, or possibly wrong--to explain it.”

Threats to validity:

Bias is the main threat in clinical research

No study can be expected to be perfect.

But every study can - and *must* - be expected to:

- not have ***fatal*** bias (**design**)

- be reported fairly

 - design

 - conduct

 - interpretation

Promises and limitations of cancer biomarkers

1. 'Disconnect' between claims, products
2. 'Fundamental problems in study design'
 - chance
 - bias
3. **How to address (1), (2)**
 - role of specimens**
 - (other)

Role of specimens: *the determining factor in marker research*

- *without* 'right' specimens, cannot 'evaluate the test'

Example: If a plausible blood test for early pancreatic cancer was created tomorrow, it could not be evaluated because no specimens.

Role of specimens: *the determining factor in marker research*

- *without* 'right' specimens, cannot 'evaluate the test'

Example: If a plausible blood test for early pancreatic cancer was created tomorrow, it could not be evaluated because no specimens.

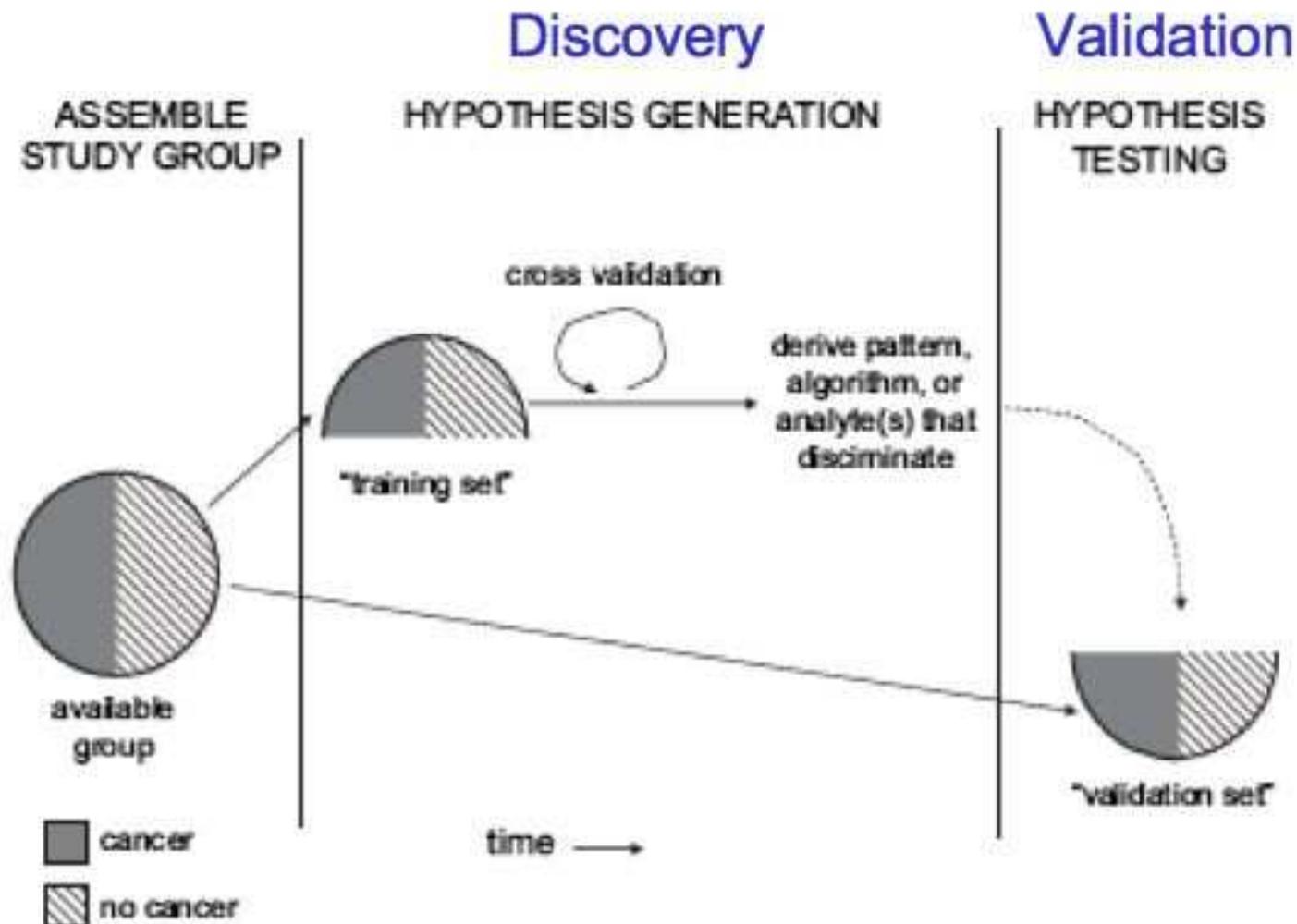
- *with* 'right' specimens, can do amazing unexpected things, in discovery, development, and 'validation.'

“Person who controls the specimens controls the field.”

With 'right' specimens, can do
amazing, unexpected things...

1. Discovery and 'validation' from same group of specimens

Discovery and validation from same group of specimens



With 'right' specimens, can do
amazing, unexpected things...

1. Discovery and 'validation' from same group of specimens
2. Discovery and validation of multiple tests simultaneously ('bake-off')

'Bake-off'

Example: Serum tests for CRC

Suppose you collect 10ml sera in 1000 CRC, 1000 controls prior to colonoscopy (i.e., before 'true state' known).

Suppose labs can do 'discovery' on 50 μ l.

Then make 200 identical sets for training (unblinded), and 200 identical sets for validation (blinded) with many CRCs and controls in each set...

Can do 'bake-off' - training and validation - involving 200 labs/tests.

With 'right' specimens, can do amazing, unexpected things...

1. Discovery and 'validation' from same group of specimens
2. Discovery and validation of multiple tests simultaneously ('bake-off')
3. Use banked specimens; e.g.,
 - PLCO - blood specimens: assess markers of diagnosis
 - NSABP (and other RCTs) - FFPE specimens:
can assess markers of prognosis or prediction

With 'right' specimens, can do amazing, unexpected things...

1. Discovery and 'validation' from same group of specimens.
2. Discovery and validation of multiple tests simultaneously ('bake-off')
3. Use banked specimens.

Approaches 1,2,3 not easy logistically... but not impossible..

With 'right' specimens, can do amazing, unexpected things...

1. Discovery and 'validation' from same group of specimens.
2. Discovery and validation of multiple tests simultaneously ('bake-off')
3. Use banked specimens.

Approaches 1,2,3 not easy logistically... but not impossible..

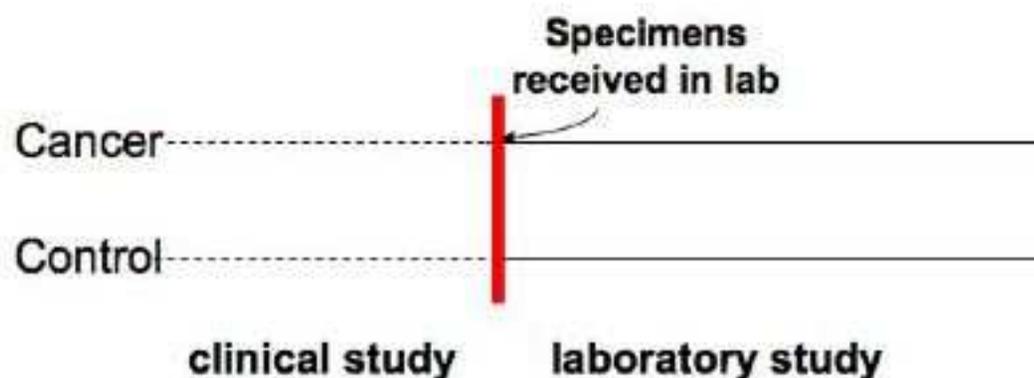
*These kinds of approaches cannot be **thought about** in drug development research.*

J Clin Epidemiol 2007;60:1205-19, 1226-28
(see discussion of 'shortcuts', 'phases')

By the time specimens are collected,
the study was 'designed' and done

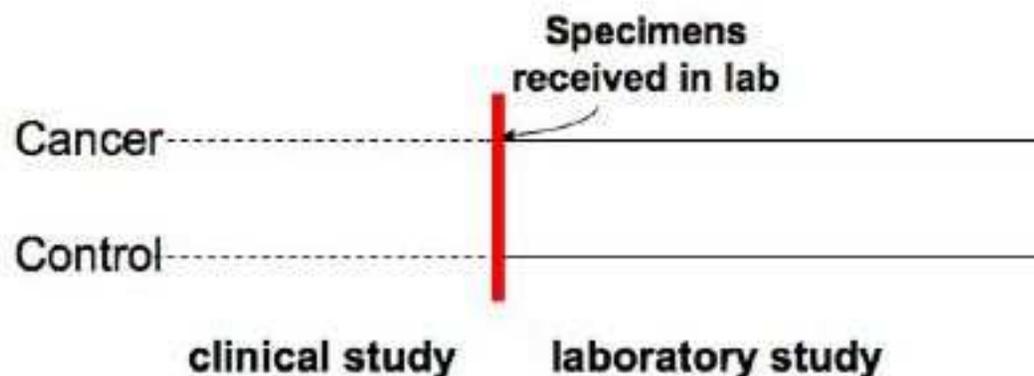
By the time specimens are collected,
the study was 'designed' and done

I.e., 'clinical study, to Left of red line



By the time specimens are collected,
the study was 'designed' and done

I.e., 'clinical study, to Left of red line



So:

1. Do you know where your specimens have been?
2. Will you be able to describe (1) in Methods, and consider impact in Discussion?

Description will be required by Guidelines for *Reporting, Information*

STARD (STARD statement for *reporting* studies of diagnostic accuracy)

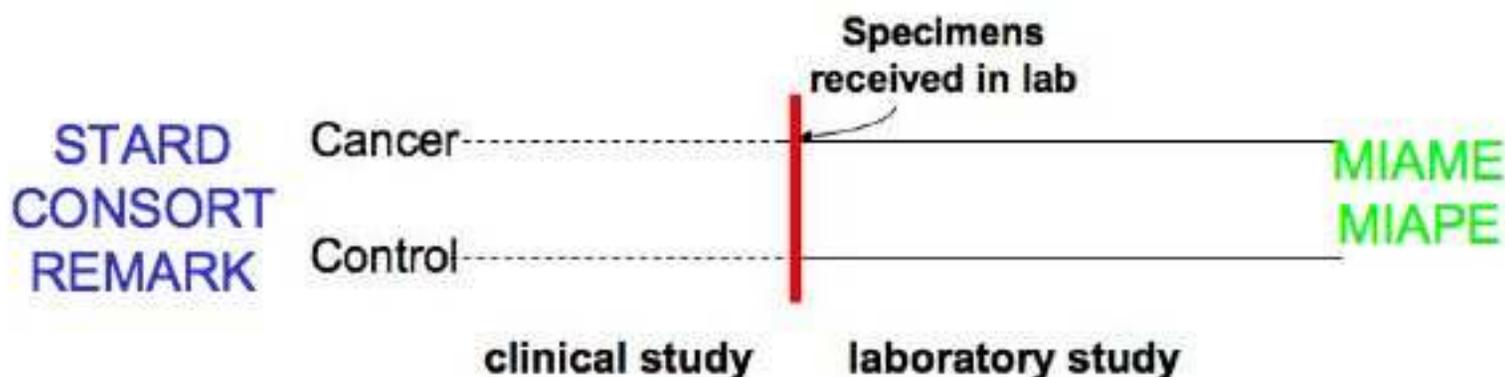
CONSORT (Consolidated standards of *reporting* trials)

REMARK (*Reporting* recommendations for tumor marker prognostic studies)

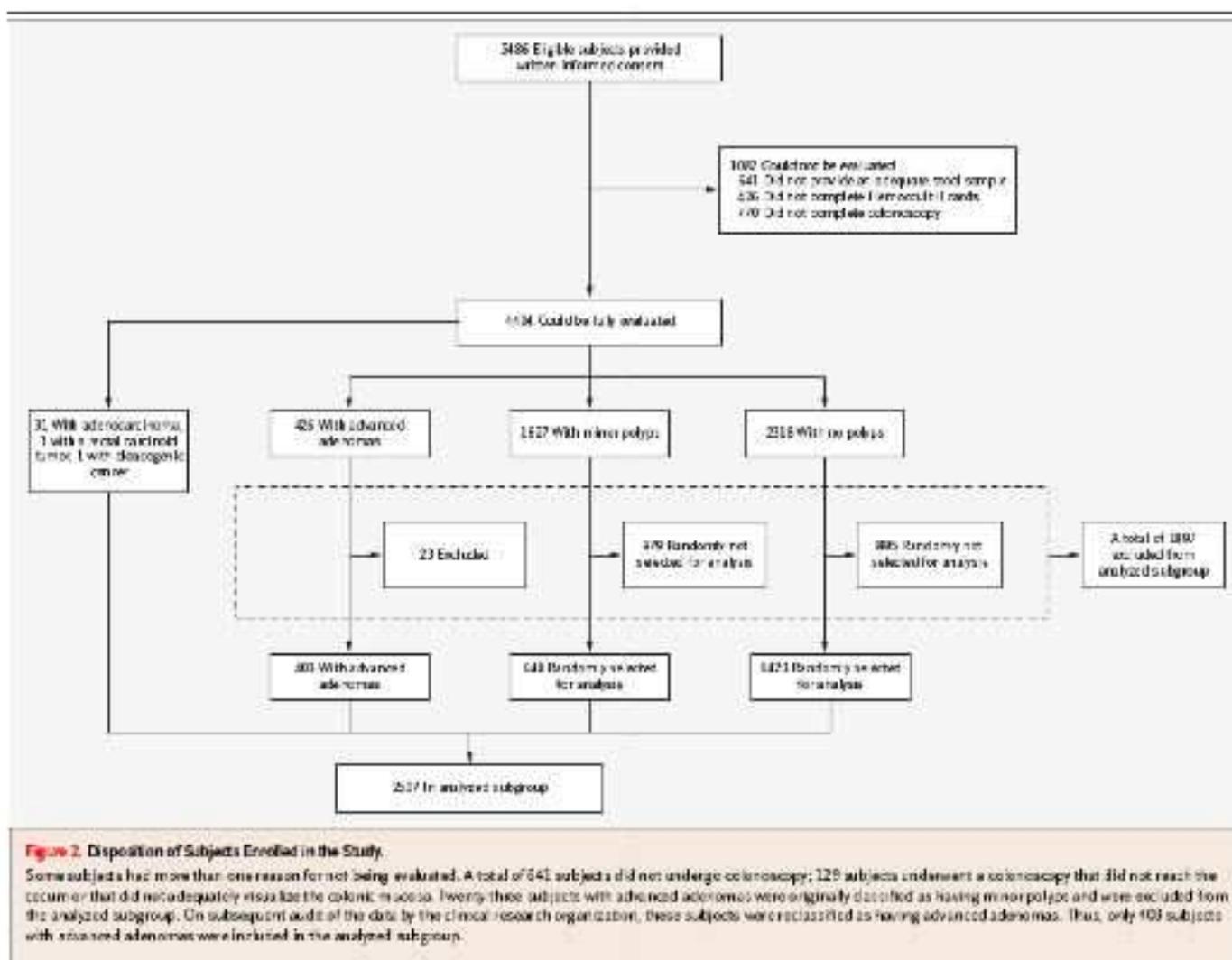
MIAME (Minimum *information* about a microarray experiment)

MAAPE (Minimum *information* about a proteomics experiment)

Guidelines apply to both sides of red line

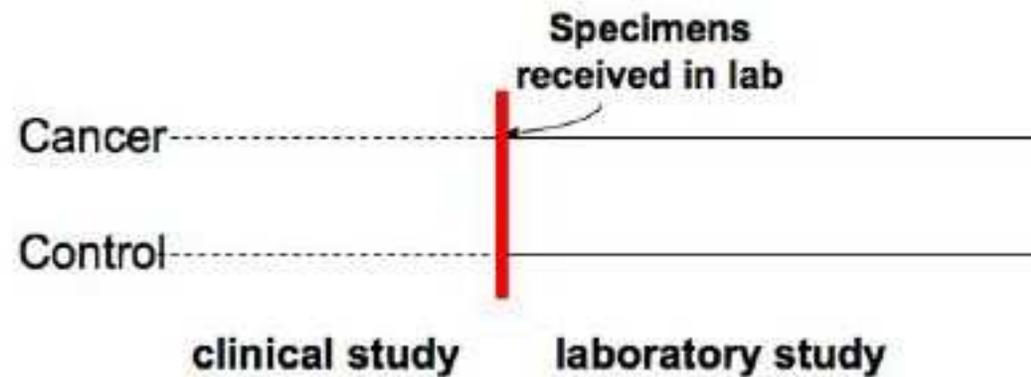


Example of detail to report, on *Left* of red line



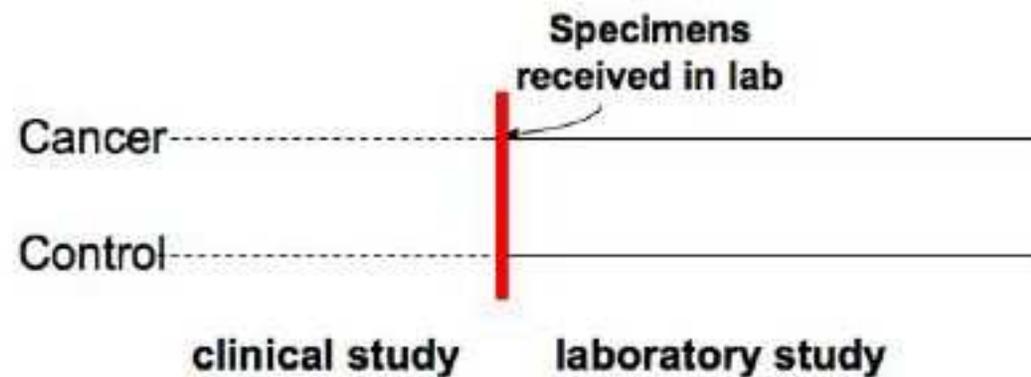
Imperiale, Ransohoff, et al.: Fecal DNA... CRC screening...
 NEJM 2004;351:2704-14

Implications for 'specimen collection'



- On Right of red line,
- collection/handling must be 'biologically sensible'
 - steps must be taken to avoid bias

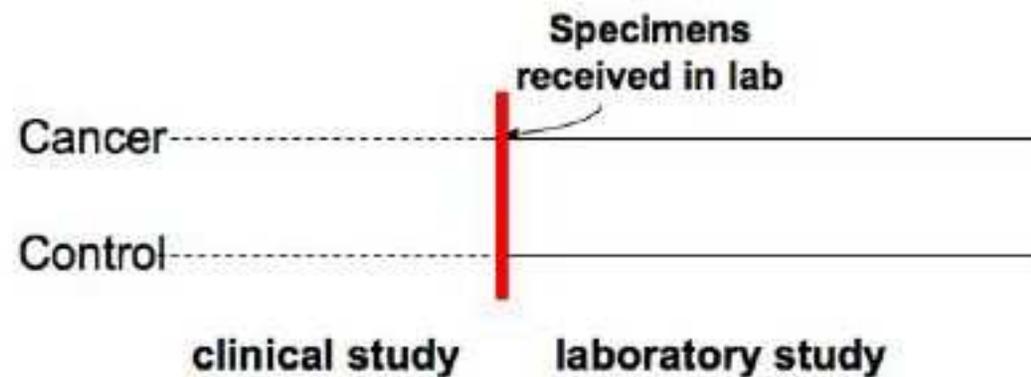
Implications for 'specimen collection'



On Right of red line,
•collection/handling must be 'biologically sensible'
•steps must be taken to avoid bias

On Left of red line,
•specimens/subjects must be 'clinically relevant'
•steps must be taken to avoid bias

Implications for 'specimen collection'



Are you thinking:

“I don't have to worry about what happens on Left of red line; I'm a biologist. Avoiding bias is for 'later phase' research.”

If that's what you're thinking, then think again

Bias is particularly important in 'early' research.

Stan (inventor of medical diagnostic) says:

The worst thing you can do is a weak [biased] early study, because it sets you off in wrong direction... wastes time and money... takes so long to figure out what was wrong.

If that's what you're thinking, then think again

Bias is particularly important in 'early' research.

Stan (inventor of medical diagnostic) says:

The worst thing you can do is a weak [biased] early study, because it sets you off in wrong direction... wastes time and money... takes so long to figure out what was wrong.

We may need, in 2008, to 'turn conventional wisdom on its head' in marker research: use high-quality 'strongly-unbiased' specimens *as early as possible*.

Promises and limitations of cancer biomarkers

1. 'Disconnect' between claims, products
2. Fundamental problems in study design
 - chance
 - bias
3. How to address (1), (2)
 - role of specimens
 - (other)**

Promises and limitations of cancer biomarkers

Other topics

- role of 'guidelines'
- role of 'phases'
- role of journals
- the 'system': Is it 'self-correcting'
- role of 'incentives' funding, publication
- etc etc

Conclusions

1. Promise:
We know so much biology; tools to measure biology are so powerful.
2. Limitations
Large 'disconnect' between claims and products, related to threats to validity from chance and bias.
3. 'Specimens' have a **critical role** to address threats and to improve productivity, efficiency of the process.

Acknowledgements

National Cancer Institute

Division of Cancer Prevention

- Biometry Research Group
- EDRN- Early Detection Research Network
- Early Detection Research Group (PLCO)

CPTAC- Clinical Proteomic Technology Assessment
for Cancer