

## *Paper presentation*

Evaluating markers for the early detection of cancer: overview of study designs and methods

Stuart G. Baker, Barnett S. Kramer, Martin McIntosh, Blossom H Patterson, Yu Shyr, and Steven Skates. *Clinical Trials*; 3: 43-56 (2006)

# Background

- Cancer biomarker development is evolving rapidly
- New developments in biology and statistics are providing increased opportunities for evaluation of markers for early detection or diagnosis

# Purpose

- Review major conceptual and methodological issues in cancer biomarker evaluation
- Emphasis on recent developments
- Make practical recommendations

# Markers

- Marker: measure or indicator of a biological process in **asymptomatic** persons that predicts future clinical cancer
- Examples: Genes, proteins, genetic/proteomic pattern, cell type, tissue abnormality

# The reality

- Most potential markers do not stand the 'test of time'
- The ultimate goal is identification of promising markers for a cancer screening study
- Requires **VARIOUS** phases of study

# Phases of Study for BIOMARKER identification

- Phase I: exploratory
  - Preclinical(?)
- Phase II: clinical assay and validation/preliminary performance
  - Evaluate the performance of the marker for the classification of subjects with and without cancer
- Phase III: retrospective longitudinal/retrospective performance
  - Evaluate the performance of the marker in stored specimens for the classification of asymptomatic subjects into those who later develop cancer and those who do not
- Phase IV: prospective screening/prospective performance
  - Evaluate the performance of the marker for the classification of asymptomatic subjects into those who, depending on the design, later have cancer detected on biopsy or clinical cancer during a follow-up period, or no cancer
- Phase V: cancer control/cancer screening
  - Evaluate the clinical utility of the marker as a trigger of early intervention by estimating the harms and benefits

# Preliminary performance study

- Comparison of clinical cancer versus no cancer
- Usually “convenience” samples: generalizability questionable
- Initial selection of promising markers
- Interpreted with caution: temporality lost

# Preliminary performance study

- Statistics:
  - Classification rules
  - TPR, FPR (sens and spec)
  - Predictive values: caution r.e. prevalence
- Avoid overfitting
  - Same data used to fit model and evaluate performance
  - Training versus test sets
  - Cross-validation
- ROC curves

# Preliminary performance study

- Training sample rules for relatively few markers
  - Baker(2000)
    - “AND/OR” combinations
    - Add cross-classified cells indicating positive test
    - Use rankings of TPR/FPR ratio to determine how to combine
  - Logic regression: minimizes a constant times FPR plus (1-TPR) in successive splits
  - Logistic regression

(a)

Marker	B=1	B=2	B=3
A=1	0.54	0.02	0.10
A=2	0.10	0.03	0.10
A=3	0.05	0.05	0.01

FPR

(b)

Marker	B=1	B=2	B=3
A=1	0.00	0.00	0.15
A=2	0.05	0.15	0.05
A=3	0.20	0.30	0.10

TPR

(c)

Marker	B=1	B=2	B=3
A=1	0 (7)	0 (7)	1.5 (5)
A=2	0.5 (8)	5 (3)	0.5 (6)
A=3	4 (4)	6 (2)	10 (1)

TPR/FPR

# Preliminary performance study

- Training sample classification rules for LARGE number of markers (e.g. microarray)
  - “open question”
  - Individual performance as screen: can miss complementary markers, interactions

# Preliminary performance study

- Study design
  - Types of specimens affects interpretation
  - Better to have early versus late-stage cancer
  - Cancers may not be representative of other factors, as compared to controls (spectrum bias)
  - Can use paired samples from same patient: has drawbacks
  - Repeated marker data in non-cancers.
    - Sense of variability over time
    - Cross-sectional information could be uninformative, but performance over time predictive
  - Sample size: target precision.
    - Suggested 110 without cancer, 70 with cancer with at least same N for training set.

# Retrospective performance study

- Nested case control design within prospective cohort study
- Specimens stored over time
- Include specimens collected BEFORE cancer diagnosis for some subjects
- Precious resource!
- Testing:
  - All those who develop cancer
  - [Matched] controls
- Advantages:
  - More directly related to detection
  - Provides opportunity to look at temporal changes

# Retrospective performance study

- Statistics
  - TPR, FPR (slightly different interpretation now)
- Over fitting still a problem
- Classification rules
  - Use slope of marker values over time with certain slope characteristics
    - Not dependent on sampling
    - Irreversible changes will be detected
    - Need fixed time period defined
  - Growth curve models
    - Markers modeled as functions of age or time
    - Do not require regular sampling
  - Based classification on average value over time
  - Hybrids: combinations of these depending on marker association

# Retrospective performance study

- Study design
  - Screening specimens could cause bias
    - Overdiagnosed
    - Cancers would not have caused medical problems in lifetime
    - Inflate TPR
  - Controls should be chosen with same spectrum of marker confounders
  - Collection, handling, analysis should be same for cancers and controls.
  - Sample size: same as previously stated.

# Prospective performance study

- Marker is tested in asymptomatic people over time
- Large sample size to get enough cases
- Ethical issues: markers are not proven for clinical decisions

# Prospective performance study

- Four design types:
  - All subjects followed to end of study or clinical detection of cancer.
    - Applicable in the early stages of marker development
    - Marker biopsy done at study inception
  - All subjects scheduled to receive biopsy
    - Gold-standard test performed at some point in time.
    - Prone to bias of overdiagnosis
  - Subjects receive biopsy only if positive
    - Cannot make certain inferences because negatives are not tested for cancer
    - Fancy designs deal with ratios of TPR and FPR
  - Randomized trial to compare marker performance
    - Randomized to one of two marker groups.
    - Those testing positive get biopsy, those negative are followed
    - Can estimate differences in detection rates

# Cancer Screening Study

- Trigger for intervention?
- Need to estimate both harms AND benefits (cumulative)
- Harms: unnecessary biopsies, bleeding of perforations, etc.
- Benefit: decrease in cancer mortality rate (potential bias due to ascertainment)
- Need large N

# Cancer Screening Study

- Randomized Trials
  - Randomized to screening or not
  - Typical N: 50,000 to 100,000
  - Efficacy of screening: effect of receiving screening.
  - Need to select optimal time of follow-up
    - Too short: no effect of screening
    - Too long: dilutes mortality

# Cancer Screening Study

- Observational studies
  - Lead-time bias:
    - Comparison of screen versus clinical detection could be strongly affected by self-selection bias
    - Screening preferentially detects slower growing cancers with longer preclinical durations
    - Screening can shift diagnosis to earlier time without affecting time course (lead-time bias)
  - Self-selection bias:
    - Subjects who receive screen are at very high or low risk....see article (page 52)
  - Mathematical models popular: but still often bias is a problem (assumptions of combinability of subgroups)

# Cancer Screening Study

- Novel observational studies less affected by self-selection bias
  - Additional costs of assumptions or limits
  - (1) periodic screening evaluation
    - Uses only data from subjects receiving regularly scheduled screenings
    - Uses older screened subjects as controls for younger screened subjects
    - Provides upper bound on effect of screening
    - Assumptions: age of birth is not predictive; once cancer is detected, it will be detected at later times

# Cancer Screening Study

- (2) Paired availability design
  - Combines before and after data from many areas
  - Estimates the efficacy of screening using a thought experiment: subjects could have been given screen at earlier or later time
  - Key idea: three types of subjects
    - Subjects who would not receive any screening
    - Subjects who would receive screening at either
    - Subjects who would receive screening only in time with more availability of screening
  - Unclear(?)