Statistical Issues in Investigating Prognostic and Predictive Markers for DCIS

James J. Dignam, Ph.D.
Assistant Professor
Department of Health Studies
The University of Chicago

National Cancer Institute Workshop:

Ductal Carcinoma *in Situ*: Strategies for Integrating Tumor Biology and Population Sciences

San Francisco, CA

February 1-2, 2007

Aims

For investigations of ductal carcinoma *in situ* (DCIS), this presentation will address:

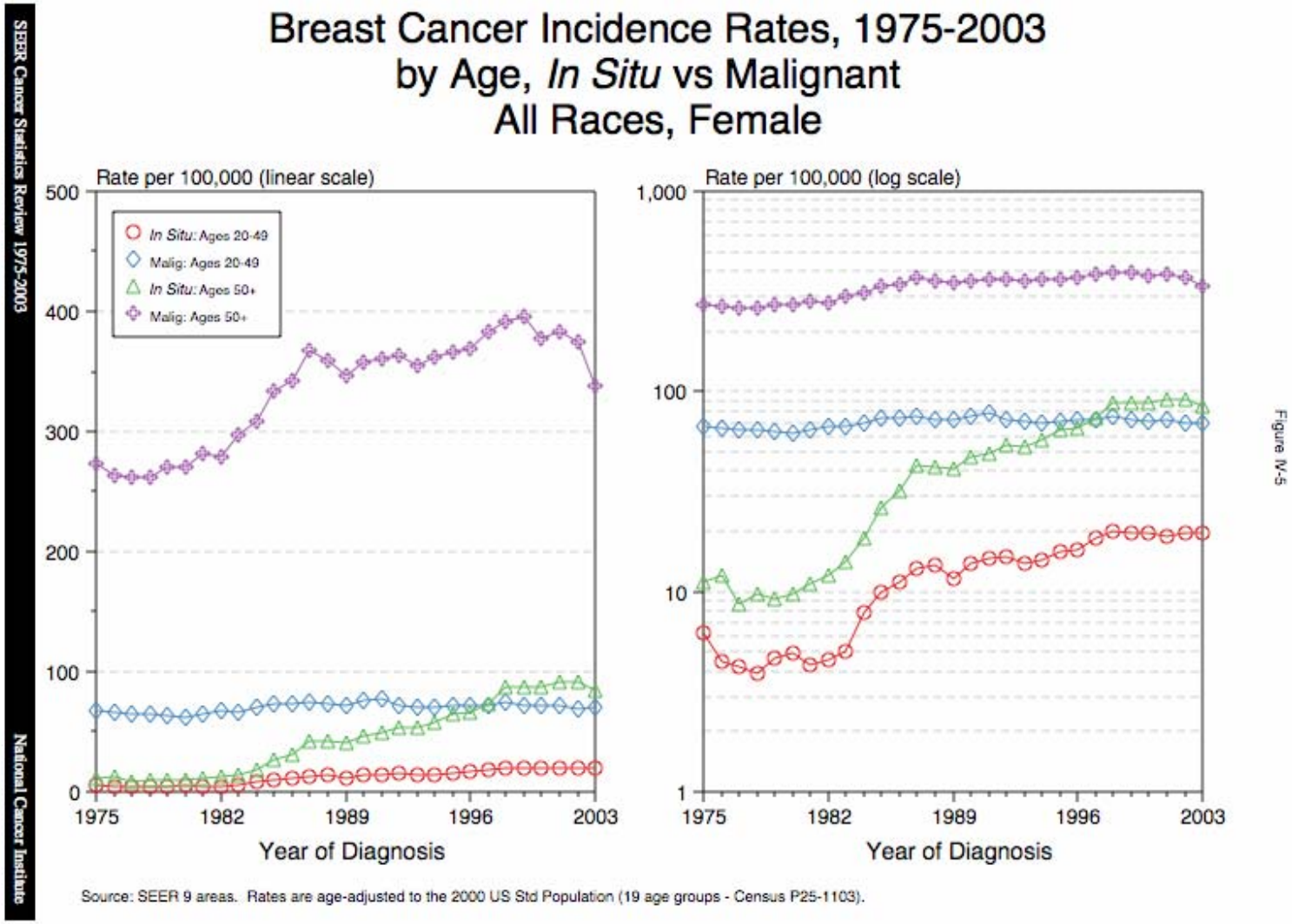
- What are the statistical issues involved in
  - Identifying prognostic markers
  - Identifying treatment-response indicators (i.e., predictive markers)
- Where and how should we develop prognostic/predictive markers?
  - Identification of patient cohorts
  - Criteria for validation of markers
  - Improvements in marker study reporting

Attributes of DCIS Potentially Important in Prognostic Marker Studies

- DCIS
  - Has increased in prevalence, changed with respect to presentation, due to screening
  - Is heterogeneous with respect to numerous features
  - Has excellent survival prognosis
  - But, confers considerable excess risk for invasive breast cancer, breast cancer mortality (?)

- These characteristics are shared in common with some other ‘early’ cancer scenarios - for example, stage I/II colon cancer

DCIS Incidence: SEER 1975-2003



Mortality after DCIS - Survival, Epidemiology and End Results (SEER)

Survival %		5-year	10-year
40-49		99.5	97.9
50+		99.0	97.6
Overall		99.1	97.7
Standardized Mortality Ratio		5-year	10-year
1978-83		3.1	3.4
1984-89		1.6	1.9

Ernster et al., *JAMA* 2000

DCIS - A Clinical Opportunity?

DCIS would seem to present an opportunity for early intervention to avoid invasive breast cancer

– Understanding of the heterogeneity of DCIS presentation, characteristics is needed to take advantage of this opportunity - Who is at high risk of invasive recurrence? Who to treat and how much?

Prognostic and Predictive Markers: Definitions

These terms are actually better known and accepted among non-statisticians, popularized in the cancer literature

- Prognostic marker - characteristic associated with prognosis, outcome - usually in terms of relative hazard of failure

- Predictive marker - characteristic that is associated with, ‘predicts’ treatment response - e.g., is involved in an antagonistic or synergistic relationship with treatment - to statisticians, an interaction effect

Prognostic and Predictive Markers: Definitions (cont.)

So, a predictive marker *is* a prognostic marker that. . .

- exerts prognostic influence differentially so, according to treatment
- offers the opportunity for prospective intervention
- may lend insight about biological aspects of the disease

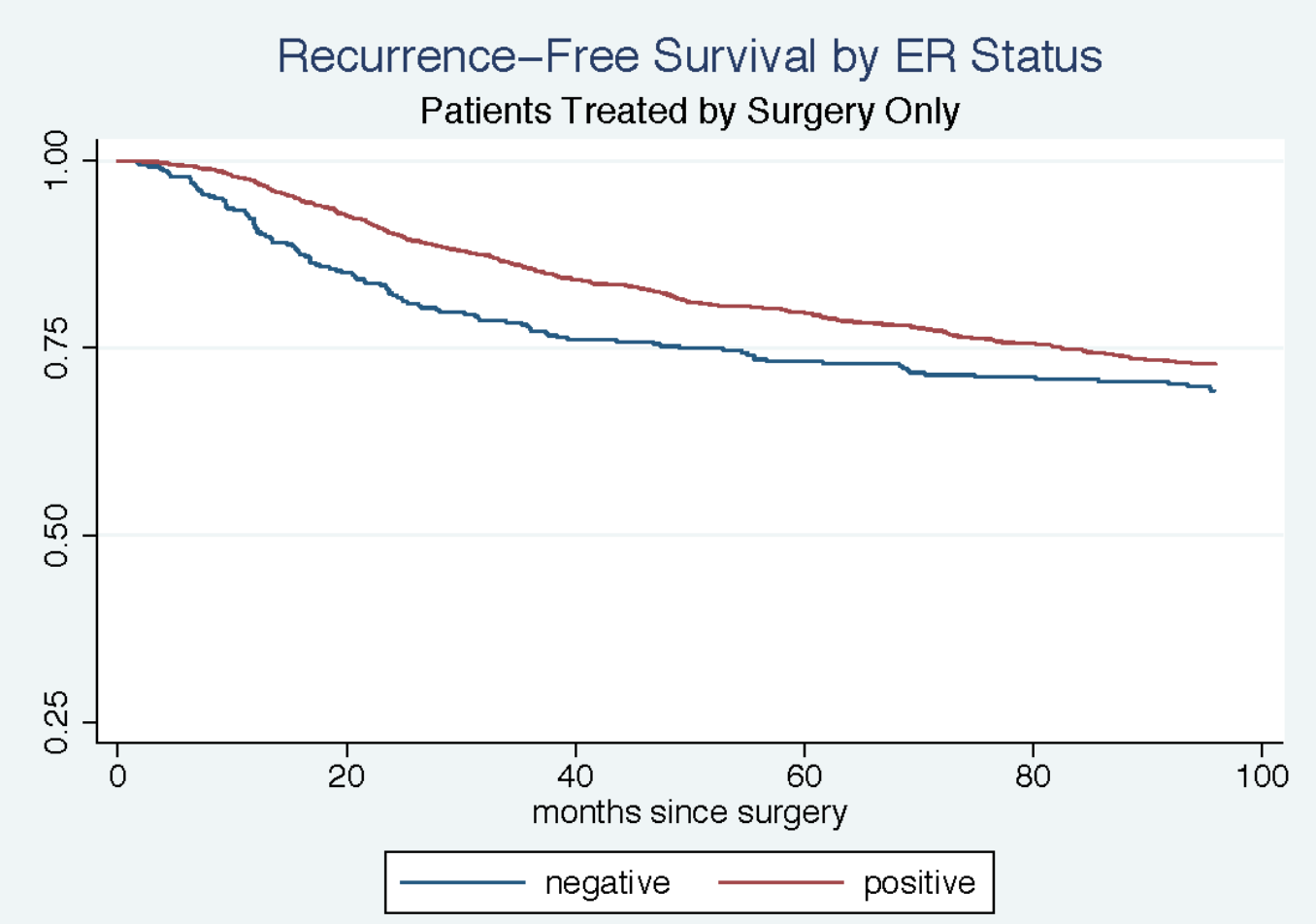
Prognostic and Predictive Markers: Example

Estrogen receptor (ER) content on breast tumors plays two important roles:

- ER is a prognostic marker, in that ER-negative tumors are associated with greater failure hazard
- ER is a predictive marker for response to anti-estrogen drugs. Virtually no response is noted in patients with ER-negative tumors, while those with ER-positive tumors respond

Prognostic/Predictive Markers (cont.)

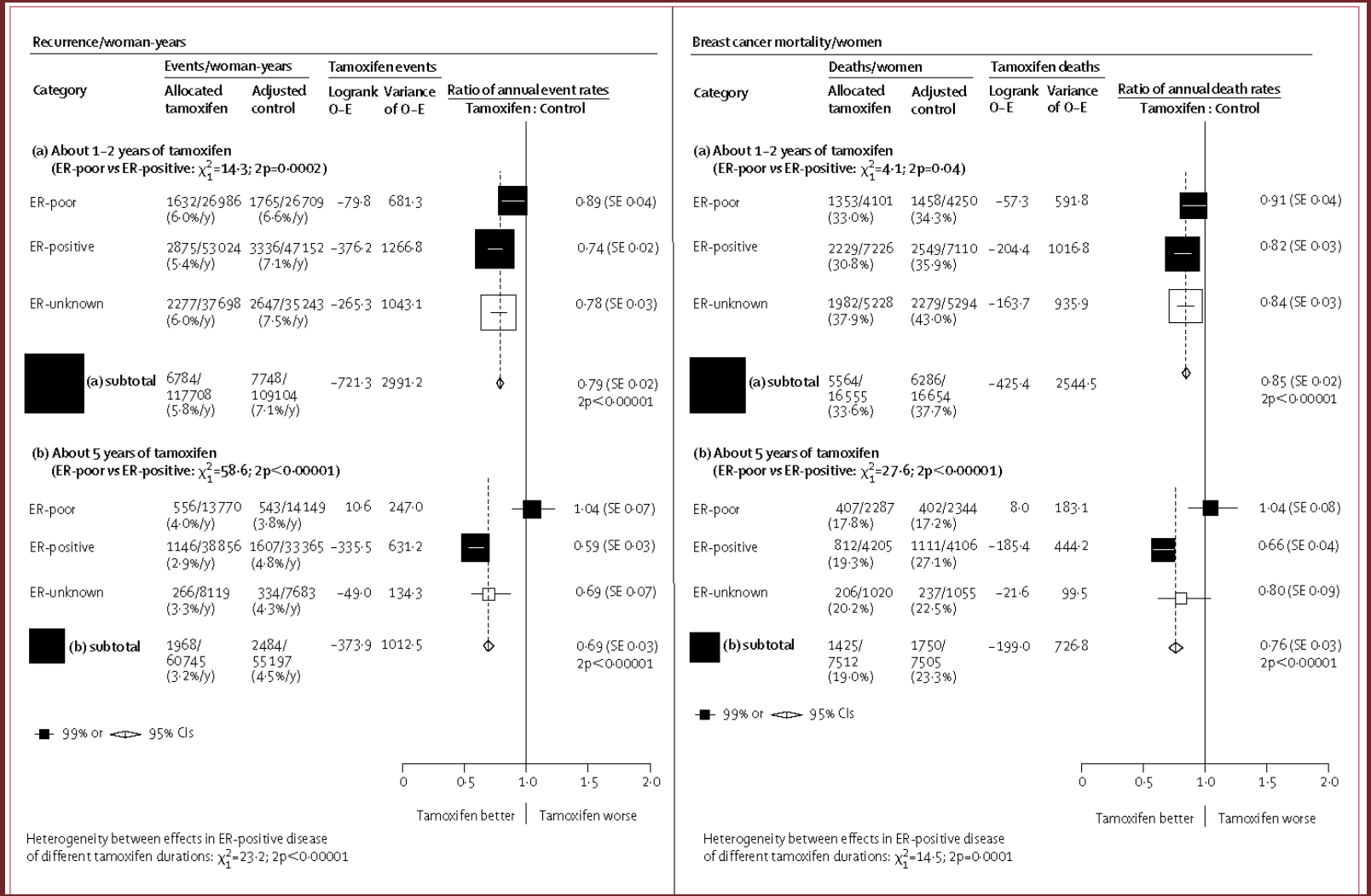
Effect of ER in absence of adjuvant treatment:



NSABP B-14

Prognostic/Predictive Markers (cont.)

Effect of tamoxifen by ER Status:



Early Breast Cancer Trialists Collaborative Group, *Lancet* 2005

Sample Size/Power for Prognostic Markers

Sample size calculations for prognostic markers resemble those in study designs for *treatment effects*, with some important exceptions:

- We cannot manipulate/control the relative frequency of different levels of the marker

- The strength of the relative hazard imparted needs to be larger than for treatments to be of interest/utility

- The marker may be correlated with multiple other known prognostic markers

- Prognostic marker effects may be more prone to change over time (e.g., exhibit non-proportionality)

Sample Size/Power for Prognostic Markers (cont.)

Simple sample size equation (total number of failure events needed) for a single prognostic marker:

$$N = \frac{\left(Z_{1-\alpha/2} + Z_{\beta}\right)^2}{\ln(HR)^2\,\omega(1-\omega)}$$

ω - frequency (proportion) 'positive' for the marker - usually .5 for treatments (equal allocation), usually smaller (say, .10-.20) for prognostic markers

HR - hazard ratio (positive/negative) for marker (proportional hazards assumed here)

Schoenfeld, *Biometrics* 1983

Sample Size/Power for Prognostic Markers (cont.)

When the prognostic marker is correlated with another, sample size is inflated:

$$N' = \frac{\left(Z_{1-\alpha/2} + Z_{\beta}\right)^2}{\ln(HR)^2\,\omega(1-\omega)} \times \frac{1}{1-\rho^2}$$

ρ - correlation between the new marker and the known prognostic marker. So, for ρ = .50, inflation factor is 1.33

Schmoor et al, *Stat Med* 2000

Sample Size/Power for Predictive Markers

For predictive markers, the situation is more challenging. The detection of interaction effects is tantamount to:

- Comparison of treated *vs*. untreated within levels of a prognostic marker

- Comparison of the size of the treatment effect between levels of the prognostic marker - *differential treatment effect*

Sample Size/Power for Predictive Markers (cont.)

	Factor B	
Factor A	Level 1	Level 2
Treatment 1	λ_{11}	λ_{12}
Treatment 2	λ_{21}	λ_{22}

Hazard ratios: $\Delta_1 = \lambda_{11}/\lambda_{21}$, $\Delta_2 = \lambda_{21}/\lambda_{22}$

Sample Size/Power for Predictive Markers (cont.)

To test

$$H_o: \Delta_1 = \Delta_2 \text{ } vs.$$

$$H_a: \Delta_1 \neq \Delta_2$$

It can be shown that the sample size needed (under some favorable assumptions) is ~ 4 times larger than that needed to detect a main effect of the same magnitude.

Peterson and George, *Controlled Clin Trials* 1993

Sample Size/Power for Predictive Markers (cont.)

Fortunately, some interaction effects *are large*, and so such studies are not infeasible

- So called 'targeted' clinical trials, where pre-selection of patients most likely to benefit, can often be much smaller than trials with more general eligibility*

- Similarly, if a treatment is essentially null in one group and effective in another, then the interaction effect may be detected.

*Simon and Maitournam, *Clin Cancer Res* 2004, *Stat Med* 2004

A Few Comments on Prognostic/Predictive Markers Derived from Modern Biotechnology

The advent of modern biotechnology tools (i.e. microarrays) requires use of statistical methods previously little-used in marker studies, as well as the development of novel methods. However . . .

- These studies tend to be more involved in *discovery* of candidate markers than testing for clinical utility

- Once a candidate marker is developed, regardless of origin, analysis from that point forward largely resembles traditional prognostic/predictive marker problems

Comments on Modern Biotechnology Markers (cont.)

In these situations, marker variable may come in the form of a prognostic score or index that is a function of several other variables

- Statistical principles for model building in general need to be followed to develop the scores.

- Independent validation data is even more critical here - there may be several equally reasonable scoring algorithms. Choice of which to use may be less important than validating one objectively.

- Reproducibility of assay is also critical. Scoring algorithm must be portable to other settings in order to have utility.

Another Statistical Issue: Endpoints

Numerous *time to event endpoints* are of interest in DCIS. Which are most important?

-IBTR-Free interval - Time to ipsilateral tumor - DCIS, invasive, both

-Invasive Breast Cancer-Free interval - Time from study entry to invasive recurrence. Invasive contralateral tumor

-Cancer-Specific Survival - Time to breast cancer death

-Event-free and Overall Survival - Time to any event (DCIS, invasive, 2nd primary, death), time to death from any cause

Competing risks are a major consideration

Data Sources for Studies in DCIS

Data sources may include:

- Health statistics registries (SEER)
- Cohorts from a single institution or health care system (Kaiser Northern CA)
- Randomized clinical trial databases - (NCI Cancer Cooperative Groups)

Data Sources (cont.)

Randomized Trials:

Advantages

- Uniformity of stage at diagnosis, cohort entry criteria
- Randomized treatment assignment, uniform treatment per protocol
- Rigorous follow-up, outcome ascertainment

Drawbacks

- Need for centralized pathology to assure common definitions
- Limited patient diversity, concerns about representativeness of ‘real-world’ patients

Data Sources (cont.)

Observational cohorts:

Advantages

– Diversity of patients, disease presentations
– Ancillary data - comorbidities, SES, etc
– Control over pathology information

Drawbacks

- Treatment selection effects, validity of predictive markers in question
- Institutional or regional pathology definitions
- Differential follow-up possibly related to prognosis, loss to follow-up

Prognostic Marker Guidelines Through the Years

There have been a number of thoughtful commentaries and guideline articles regarding the development and use of prognostic markers in breast cancer. These have been motivated by

- The recognition of the failure of many initially promising markers to ‘pan out’
- The need for high-level evidence (similar to that required for treatments) to recommend markers for clinical use*

*Hayes et al, *J Natl Cancer Inst* 1996

Simon and Altman (1994) - *Statistical Aspects of Prognostic Factor Studies in Oncology*

Outline study progression similar to that used in therapy development:

1. Early exploratory studies of markers, showing association with outcome and development of reproducible assay

2. Studies to determine whether prognostic marker provides improved ability to identify patients at high or low risk of failure

3. Studies to determine which subsets of patients benefit from a given therapy

Simon and Altman, *Br J Cancer* 1994

Statistical Aspects of Prognostic Factor Studies in Oncology (cont.)

Authors offer a critique of (and suggested solutions to) many statistical design and analysis weaknesses of prognostic factor studies:

- Ill-defined or absent study hypothesis
- Inadequate sample size for the necessary consideration of multiple variables in modeling
- Overuse of stepwise regression
- Inappropriate ‘optimization’ of cutpoint(s) for partitioning continuous variables into low and high risk markers
- Lack of confirmation on independent data

Pajak et al. (2000) - *Statistical Issues in Tumor Marker Studies*

Called for a number of design and analysis improvements to strengthen prognostic marker studies, including

- Consideration of the multiple factors affecting statistical power
- More attention to assay methods
- Better cooperation among clinicians, laboratory scientists, and statisticians to raise the quality of marker studies
- Effort to prioritize markers, perform confirmatory studies

Pajak et al., *Arch Pathol Lab Med* 2000

College of American Pathologists (CAP) (2000) - *Solid Tumor Prognostic Factors - Which, How, and So What?*

Summary document and recommendations for implementation

- Made specific recommendations for several disease sites
- Traced evolution of markers recommended for use since last summary in 1994
- Included concise "General Statistical Recommendations" derived from Pajak et al.

Hammond et al., *Arch Pathol Lab Med* 2000

Solid Tumor Prognostic Factors - Which, How, and So What? (cont.)

General Statistical Recommendations from CAP Conference XXXV

1. Clinical trials should be specifically designed to test whether a factor has prognostic value. This question can be included in a therapeutic trial, but careful attention must be paid to assuring sufficient statistical power to answer both the prognostic and the treatment questions.

2. Prognostic factor question must be prioritized for importance by multidisciplinary groups of investigators working with each cancer type so that the most important factors are quickly evaluated.

Hammond et al., *Arch Pathol Lab Med* 2000

Solid Tumor Prognostic Factors - Which, How, and So What? (cont.)

General Statistical Recommendations from CAP Conference XXXV

3. Journals should adopt publication guidelines for reporting results, including:

- Assessment of possible patient selection bias
- Source of patients for the study
- Difference between patients with and without tumor marker in terms of
  - -Baseline demographic and tumor characteristics
  - -Treatment received
  - -Efficacy outcomes
- Statement about how missing data were handled
- Cut-point selection for method stated
- Adjustments for multiple testing
- Statistical power analysis if conclusion is negative
- Large validation studies should be given publication preference after initial exploratory work for a tumor marker

Solid Tumor Prognostic Factors - Which, How, and So What? (cont.)

General Statistical Recommendations from CAP Conference XXXV

4. Organization addressing prognostic factor categorization should come to consensus about the ranking of factors, or at least harmonize their recommendations relative to each other, so that a clear picture of the relative value of various factors is developed.

5. Continued research into multivariate analysis techniques for incomplete data and for evaluation of multiple factors is needed.

Hammond et al., *Arch Pathol Lab Med* 2000

McShane et al. (2005) - *Reporting Recommendations for Tumor Marker Prognostic Studies*

From 2001-2003, panel of investigators developed guidelines for REporting tumor MARKer studies (REMARK)

- Focuses on standardization and quality improvements in reporting, presentation of marker studies, not on study design, etc. Similar to CONSORT guidelines for reporting clinical trials
- Focuses on prognostic marker studies, suggests that predictive marker study reporting would be more stringent

McShane et al., *J Clin Oncol* 2005

REMARK Guidelines (cont.)

Some highlights- guidelines call for description of:

- Detailed characteristics of study cohort, including source and inclusion/exclusion criteria
- Treatment received and how it was chosen
- Assay methods and material source details
- Basic study design issues: endpoint, rationale for sample size, effect size of interest
- Statistical analysis issues: candidate variables for modeling, variable selection procedure, verification of model assumptions, how missing values were handled

REMARK Guidelines (cont.)

In reporting results:

- Describe distribution of demographic, disease-specific, and marker data, including missing values
- Show relation of marker to other variables
- Present univariate relations to outcome in relative and absolute terms
- Present effect size with confidence intervals in multivariable models for marker and other variables

McShane et al., *J Clin Oncol* 2005

Summary

- Study design requirements for marker studies arise from well-known statistical principles and continue to evolve as do markers

- Different study cohorts have strengths and weaknesses, and most of the latter can be overcome by careful study design

- There is a history of guidelines for quality improvements in designing, conducting, and reporting marker studies. To bring prognostic/predictive markers to full potential, we (researchers, editors, funders) should collectively implement such recommendations.

The End

A bibliography of articles mentioned here and some others can be provided upon request - jdignam@uchicago.edu

Acknowledgements:

This work was partially supported by Public Health Service grants NCI-U10-CA-69651 and NCI-U10-CA-12027 from the National Cancer Institute.