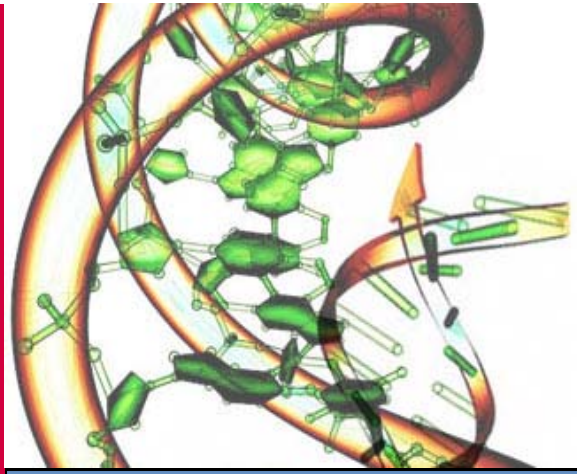
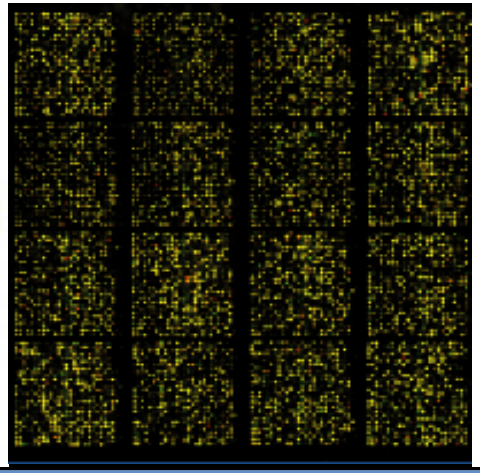
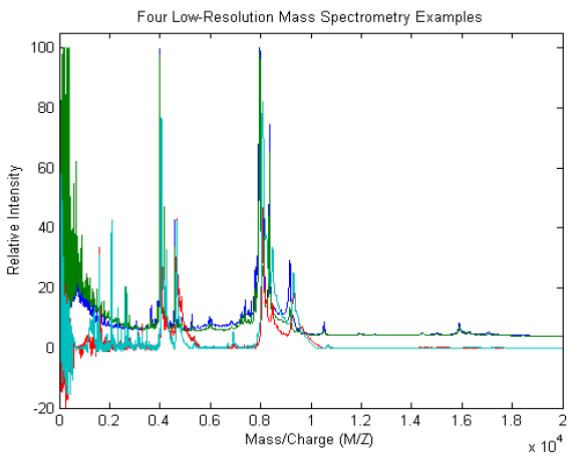
National Cancer Institute







Background for Predictive Assays Breakout Groups

Group A leaders: S. Hamilton, W. Barlow*, D. Carbone

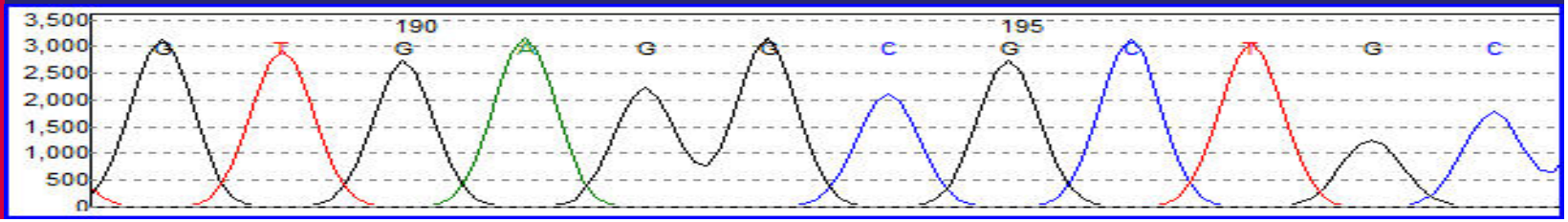
Group B leaders: M. Birrer, L. McShane*, M. Williams

(*Presenting on behalf of groups)

Evidence Needed to Support the Clinical Utility of Molecular Diagnostics

Bethesda, MD

January 10-11, 2012



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

OUTLINE

- Classes of "predictive"
- Formal definitions
- Concepts & principles
- Statistical framework
- Types of evidence
- Examples (interspersed throughout)
- Our task

National Cancer Institute

Classes of "predictive" assays (Our working definitions)

- Class 1: Focused therapy selection
  - Test-pos: Benefit from new therapy
  - Test-neg: No or minimal clinical benefit from new therapy
- Class 2: Therapy activity indicator
  - Identify *at least one* group with therapy activity
- Class 3: Broad indicator of therapy (e.g., adjuvant therapy) benefit
  - Often an "extreme" prognostic indicator
- Class 4: Therapy benefit indicator
  - Identify *at least one* group with therapy benefit

National Cancer Institute

Definition: Prognostic assays

- PROGNOSTIC: Biomarker measurement (test result) associated with clinical outcome in absence of therapy (natural course) *or with standard therapy all patients are likely to receive*
  - Clinical use: Identify patients who have very favorable outcome in absence of *(additional)* therapy or who have poor outcome and <u>might</u> benefit from more aggressive *(additional)* therapy (This makes it a Class 3 test candidate.)
  - Research use: Disease biology, identify drug targets, stratification factor in clinical trials

National Cancer Institute

Definition: Predictive test (Class 1)

- PREDICTIVE: Biomarker measurement (test result) associated with benefit or lack of benefit (potentially even harm) from a particular therapy relative to other available therapy
  - Clinical use: Select one treatment versus another treatment
  - Alternate terms: treatment-selection, treatment-guiding, treatment effect modifier

Class 2 and Class 4 tests are tools for drug development, and are often *assumed* to be Class 1 predictive tests based on other biological rationale.

Some statistical points about prognostic/predictive factors

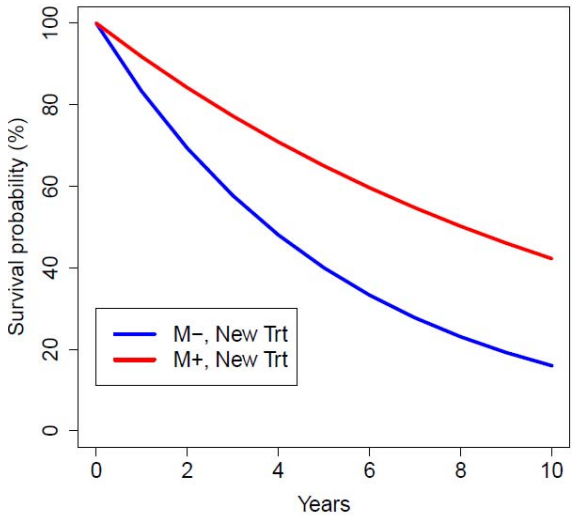
- Prognostic factors are often identified first
  - e.g. HER2 positive breast cancer patients have worse survival

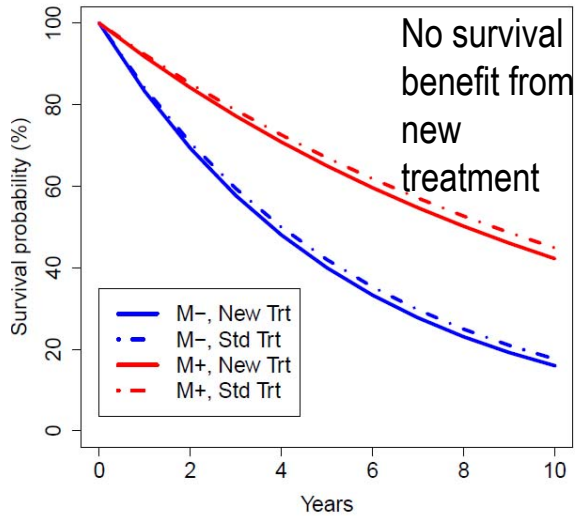
  → leads to the development of anti-HER2 agents

  → HER2 status used as predictive marker
  - Sample sizes to identify prognostic factors do not have to be large if effect sufficiently large
- Predictive factors look for variance in outcomes across treatments
  - Typically requires much larger sample sizes

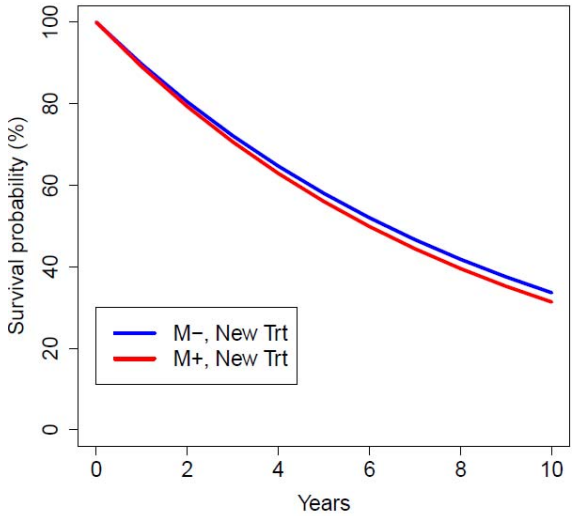
National Cancer Institute

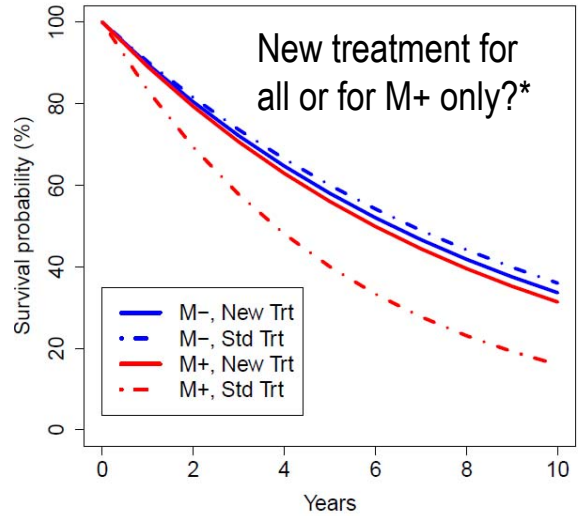
Prognostic versus predictive distinction: Importance of control groups





Prognostic but not predictive





Prognostic and predictive

(*Different considerations might apply for Standard Treatment ± New Treatment)

Statistical language

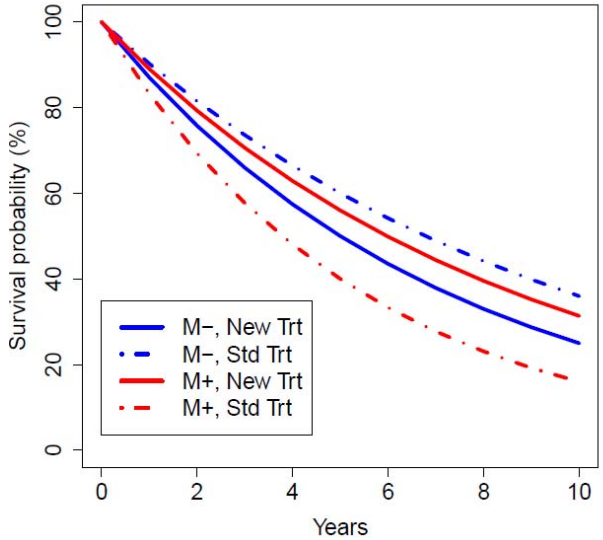
- Treatment by marker interaction: Treatment hazard ratio in marker positive group divided by treatment hazard ratio in marker negative group
  - Quantitative interaction
    - Treatment benefits all patients but may work better for marker positive than marker negative
    - All patients should receive same treatment
  - Qualitative interaction
    - No treatment benefit (none or possibly harm) in the marker negative group
    - Treatment benefit in the marker positive group
    - Preferably would like to show a statistically significant interaction

National Cancer Institute

When is a predictive biomarker-based test clinically useful?

Treatment-by-biomarker interaction: Is it sufficient?

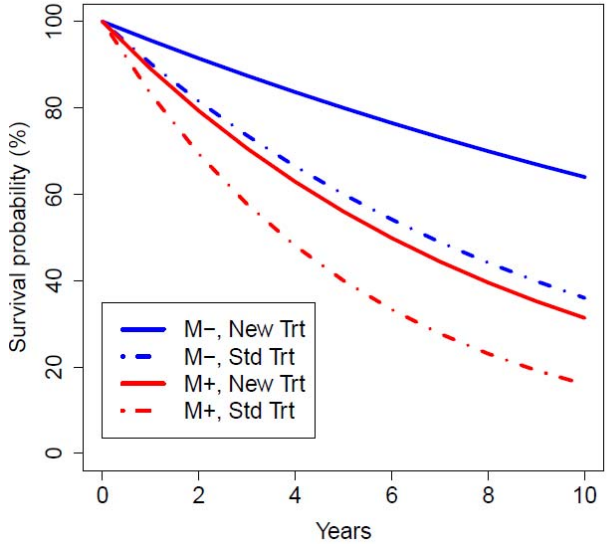
Prognostic and predictive;
New treatment for M+ only



Qualitative interaction

- Std Trt better for M– (HR_= 1.36)
- New Trt better for M+ (HR_+= 0.63)
- Interaction = 0.63/1.36 = 0.47

Prognostic and predictive;
New treatment for all?*



Quantitative interaction

- New Trt better for M– (HR_ = 0.44)
- New Trt better for M+ (HR_+ = 0.63)
- Interaction = 0.63/0.44 = 1.45

Interaction = HR_+/HR_- where $HR=\lambda_{New}/\lambda_{Std}$

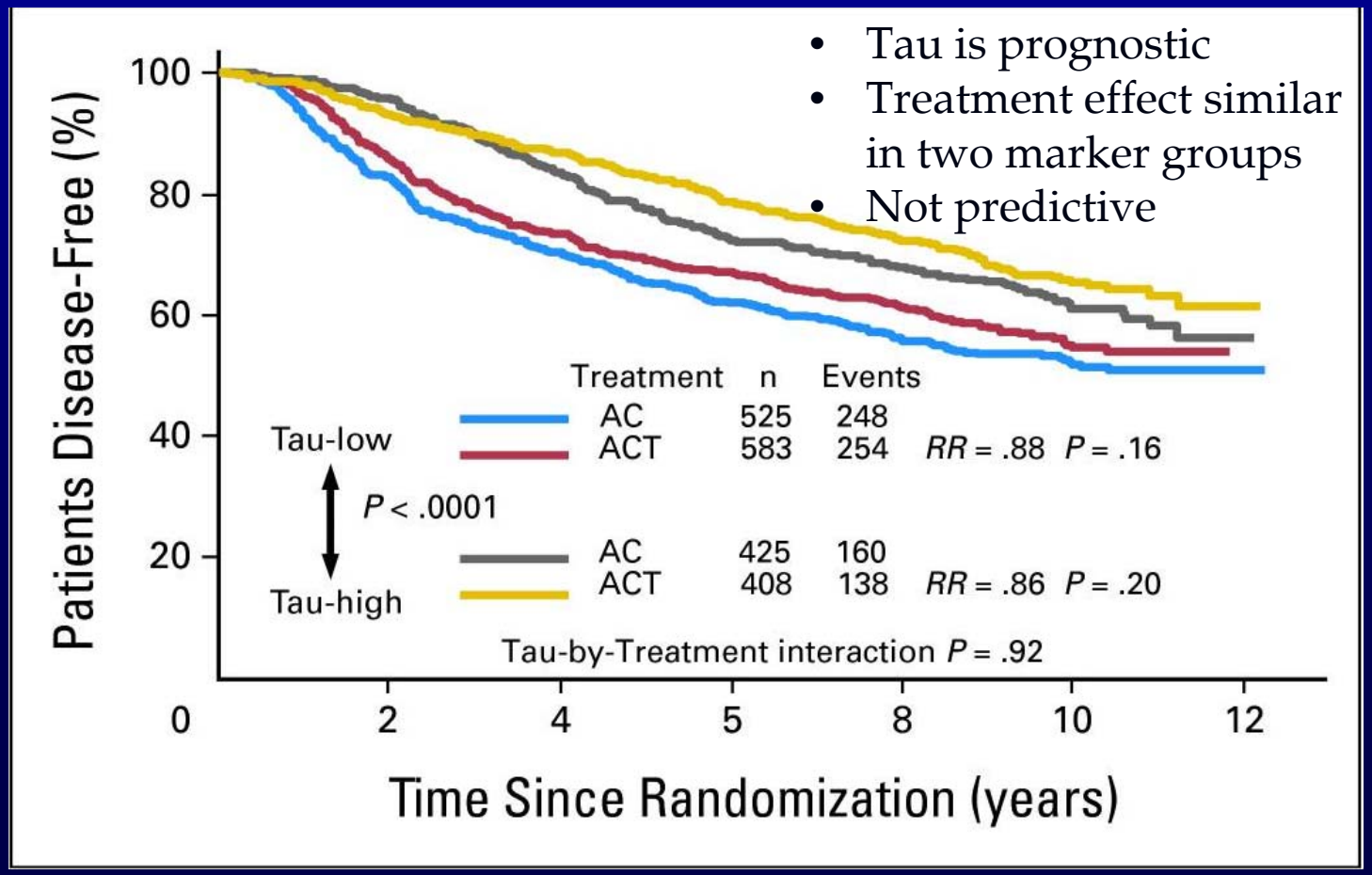
(*Different considerations might apply for Standard Treatment ± New Treatment)

Cautions

- Term "predictive" is peculiar to oncology
  - Preferable: treatment effect modifier, treatment selection marker, treatment guiding marker . . .
  - Pubmed search
    - "Prognostic" and "predictive" in the titles; constrained to the last 5 years; human research; abstracts available: 237 publications
    - + "cancer" in title/abstract: 215 publications
    - + "interaction" in title/abstract: 14 publications
- Subgroups identified by "fishing" often spurious
- Difference in marker prognostic effect between two treatment groups is necessary but not sufficient for clinically useful predictive effect

Identifying predictive factors using interaction

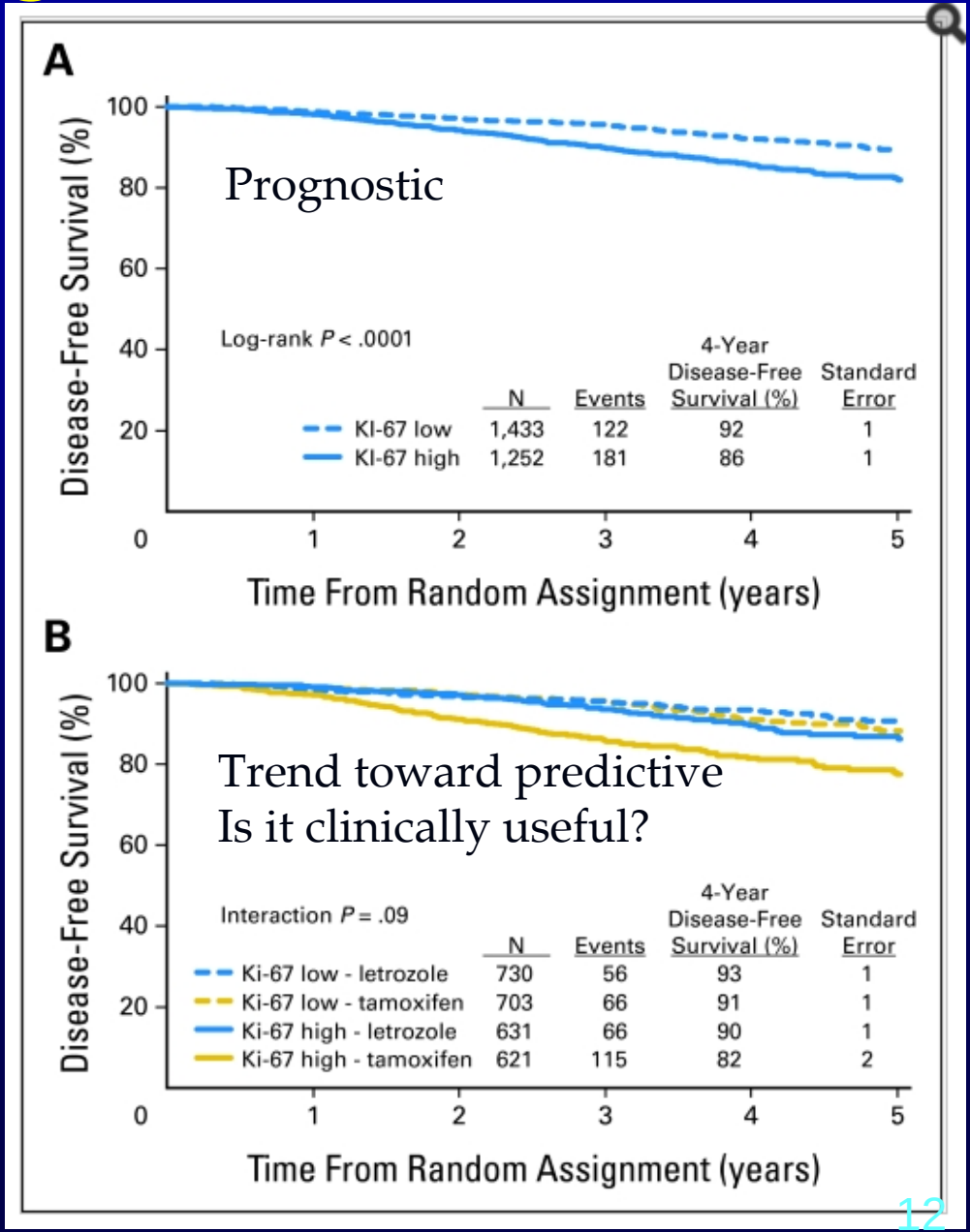
- Pusztai et al. *Evaluation of Microtubule-Associated Protein-Tau Expression As a Prognostic and Predictive Marker in the NSABP-B 28 Randomized Clinical Trial*, **JCO**, 2009



- Tau is prognostic
- Treatment effect similar in two marker groups
- Not predictive

Interaction might not be sufficient

Viale et al. *Prognostic and predictive value of centrally reviewed Ki-67 labeling index in postmenopausal women with endocrine-responsive breast cancer: results from Breast International Group Trial 1-98 comparing adjuvant tamoxifen with letrozole,* **JCO**, 2008



HR for letrozole vs. tamoxifen

High Ki-67 0.53 95% CI 0.39 – 0.72

Low Ki-67 0.81 95% CI 0.57 – 1.15

Give all patients letrozole, or only Ki-67 high?

Types of evidence

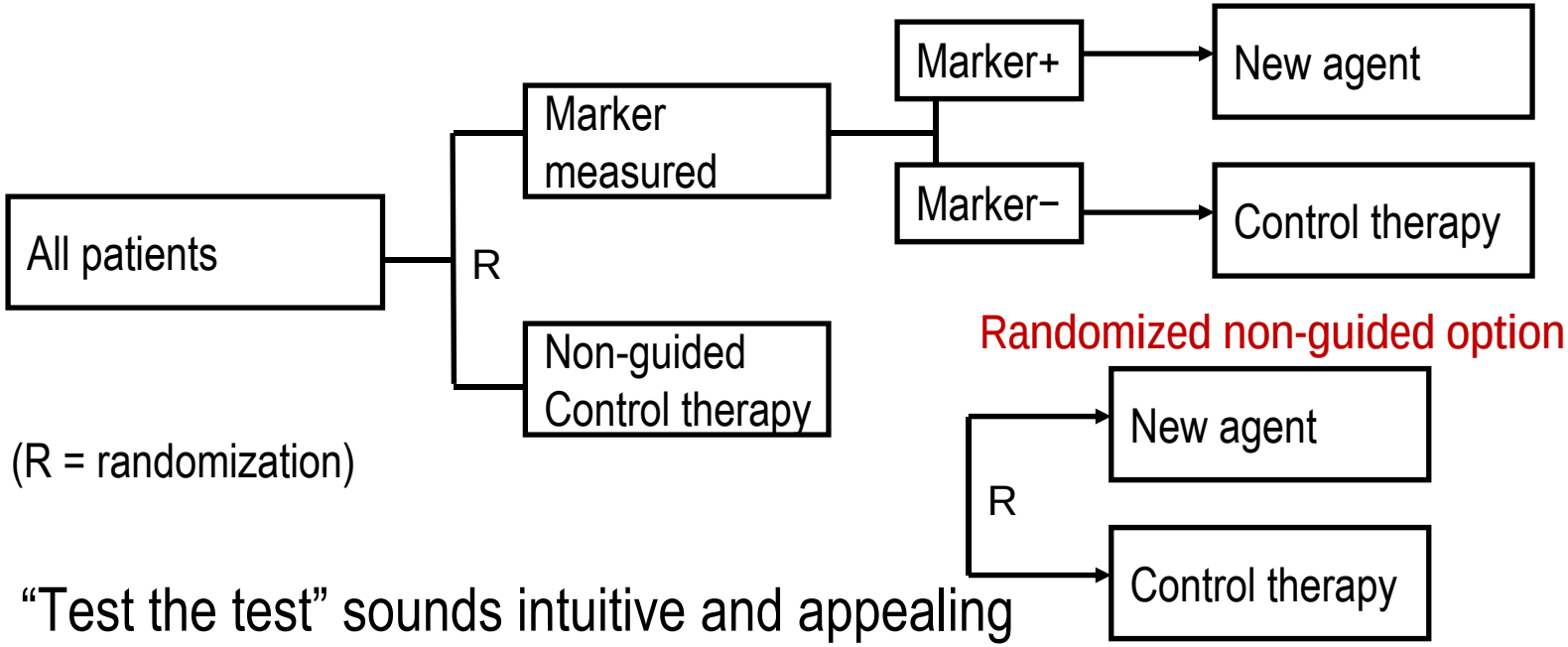
- Anecdotal
- Databases
- Retrospective
- Prospective-retrospective
- Prospective

National Cancer Institute

Prospective phase III study designs

- Basic phase III designs and hybrids (Sargent et al. 2005, *JCO*; Freidlin et al. 2010, *JNCI*; Clark and McShane 2011, *Stat in Biopharm Res*)
  - Biomarker-strategy design
  - Enrichment design
  - Biomarker-stratified (& completely randomized) design
- Sample size requirement depends on choice of design
- Evidence required vs. produced differs by design
- Start with "easy" case of binary marker with analytically validated test

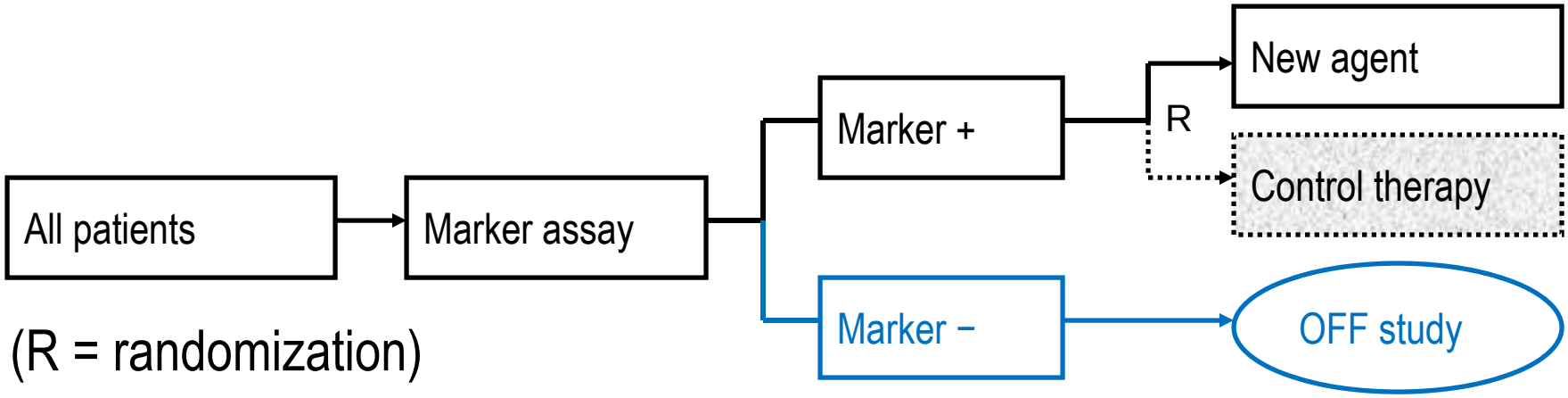
Biomarker-Strategy Design



(R = randomization)

- "Test the test" sounds intuitive and appealing
- Inefficiency: Biomarker-neg patients receive same therapy on both arms
- Real-time assay required
- Must measure biomarker in non-guided arm to distinguish prognostic effect
- Variations
  - Randomize non-guided arm – even more inefficient
  - Clinico-pathologic-guided vs. marker-guided
  - Scales to multi-marker multi-treatment strategies

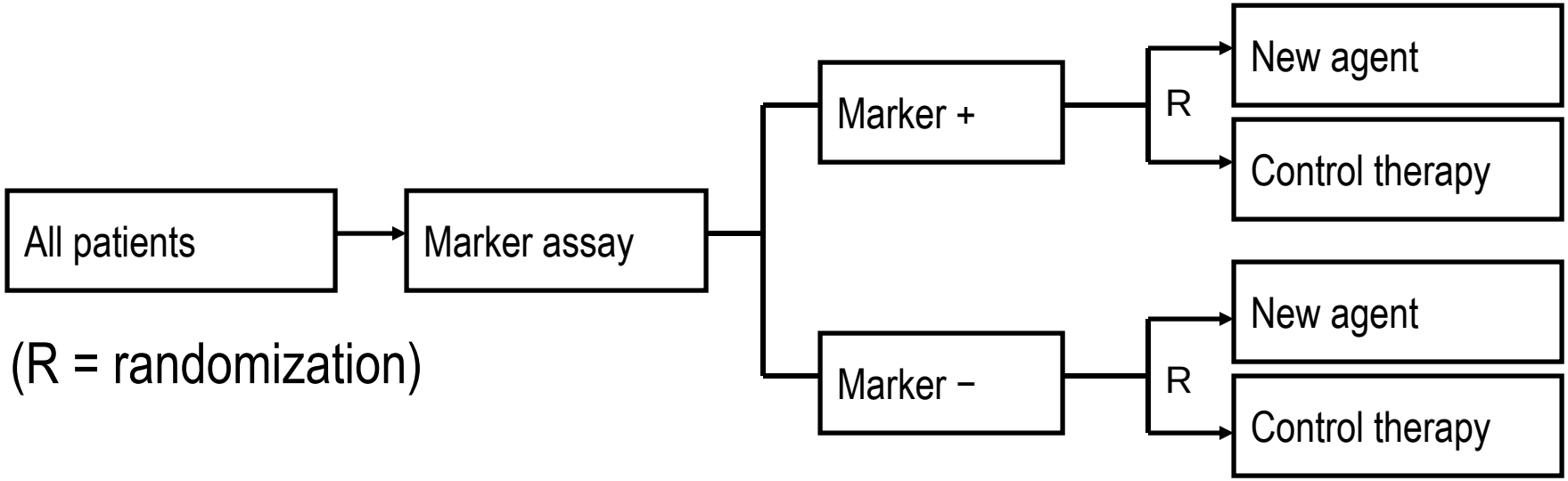
Biomarker-Enrichment Design



(R = randomization)

- Based in knowledge of biology (New agent→ Molecular target)
- Control therapy arm controls for biomarker prognostic effect
- Variation: Standard therapy ± new agent
- Limitations:
  - Off-target effects of new agent not fully evaluated
  - Regulatory indication limited to biomarker+ group
  - Limits future biomarker refinement to within biomarker+ group

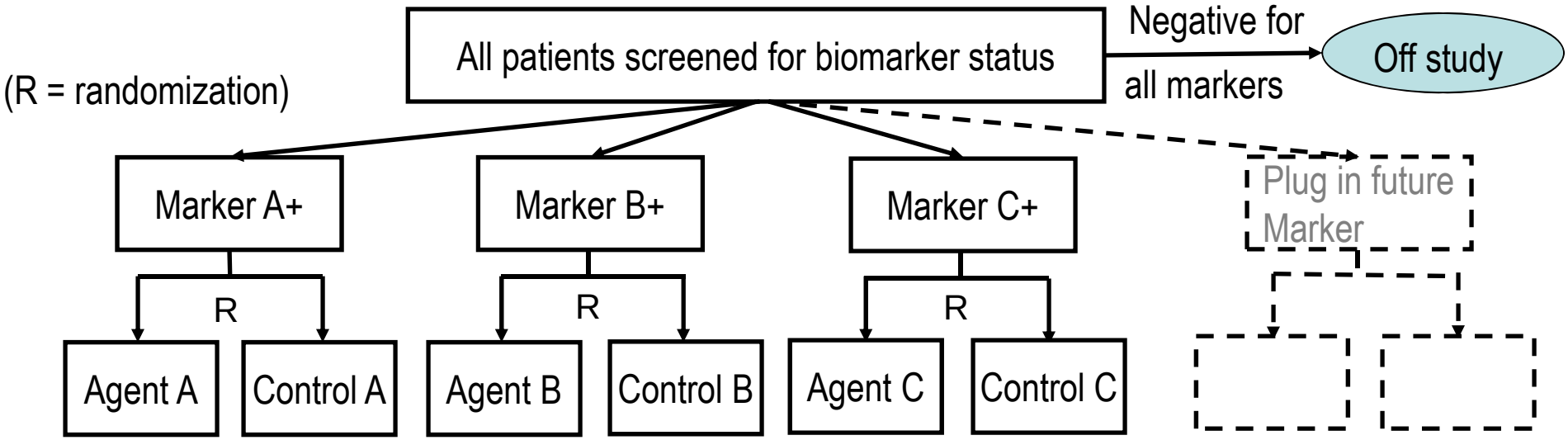
Biomarker-Stratified Design



(R = randomization)

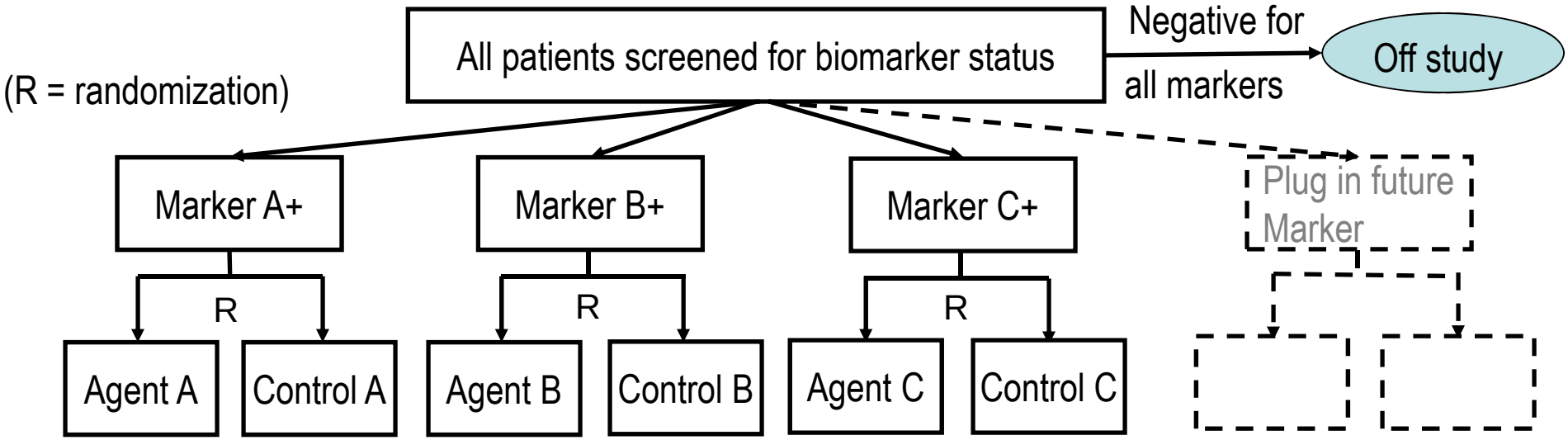
- Reasonable basis for biomarker candidate (target gene or pathway)
- Allows maximum information
  - Controls for prognostic effect of biomarker
  - Directly compares new agent to control therapy in *all* patients
- Allows retrospective evaluation of biomarkers measured by different method (e.g., protein, RNA, DNA) or alternative biomarkers in pathway
- Variation: Standard therapy ± new agent
- Completely randomized design with retrospective biomarker evaluation is an option, but assay results might not be available for 100% of patients

Multiple-Biomarker Enrichment Design



- Based in knowledge of biology (New agent→ Molecular target)
- Control therapy arm controls for biomarker prognostic effect
- Variation: Standard therapy ± new agent
- Limitations:
  - Off-target effects of new agent not fully evaluated
  - Regulatory indication limited to biomarker+ groups
  - Limits future biomarker refinement to within biomarker+ groups
  - Handling overlapping biomarkers
- Efficiencies
  - Common entry for biomarker testing, multiplex biomarker assays possible
  - Master IND for multiple drugs

Multiple-Biomarker Enrichment Design



- Based in knowledge of biology (New agent→ Molecular target)
- Control therapy arm controls for biomarker prognostic effect
- Variation: Standard therapy ± new agent
- Limitations:
  - Off-target effects of new agent not fully evaluated
  - Regulatory indication limited to biomarker+ groups
  - Limits future biomarker refinement to within biomarker+ groups
  - Handling overlapping biomarkers
- Efficiencies
  - Common entry for biomarker testing, multiplex biomarker assays possible
  - Master IND for multiple drugs

Prediction with a continuous or ordinal marker

- Have been considering only dichotomous markers so far (positive/negative)
- Many markers are continuous or ordinal
  - Markers are often dichotomized at some possibly arbitrary cutpoint
- Can more efficiently look for treatment differences that vary by marker when the marker is not dichotomized, but kept in its original form
- Use of different cutpoints across studies hampers ability to draw overall conclusions
- Current SWOG trial S1007 could not have been launched if this efficient method was not used

RxPONDER a clinical trial

Rx for Positive Node, Endocrine Responsive Breast Cancer

Node-positive (1-3) HR-positive and HER2-negative breast cancer

↓

RECURRENCE SCORE evaluated

↓ RS ≤ 25

RANDOMIZATION
Chemo vs. No Chemo

N = 4,000

Stratified by
1. **RS 0-13 vs. 14-25**
2. **Menopausal status**
3. **Axillary node dissection vs. Sentinel node biopsy**

→ **Chemotherapy; appropriate endocrine therapy**
N = 2,000

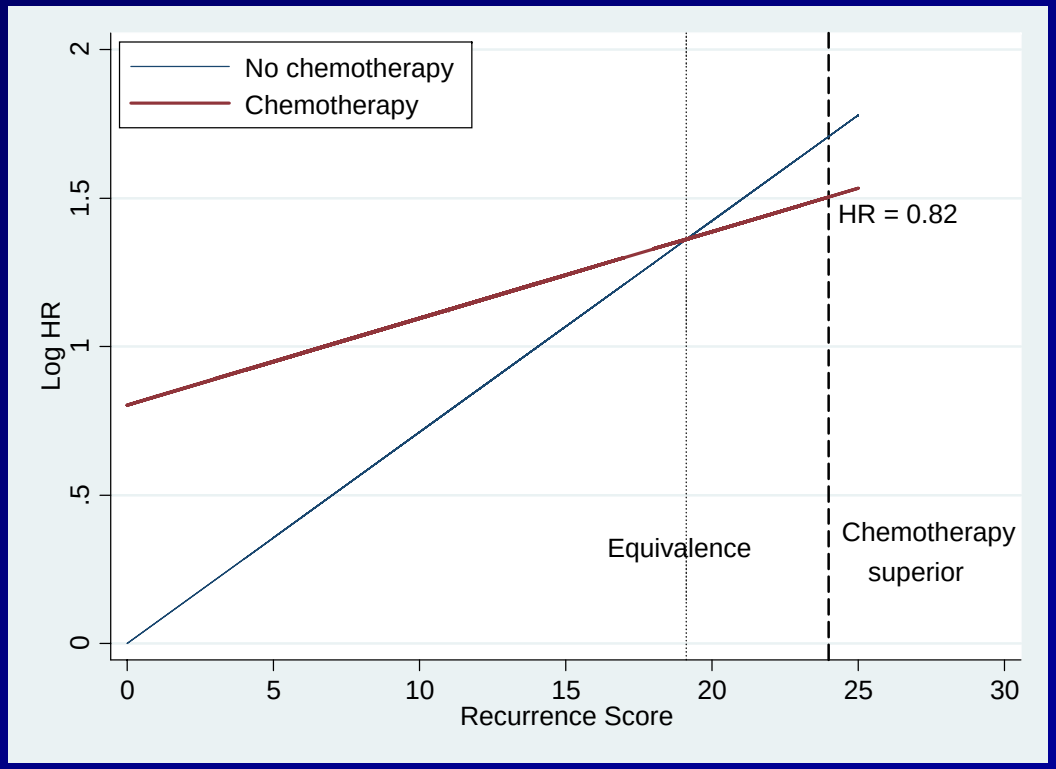
→ **No Chemotherapy; appropriate endocrine therapy**
N = 2,000

SWOG S1007 objectives

- Trial tests prediction of chemotherapy benefit based on OncotypeDX Recurrence Score
  - Survival benefit of chemo (if it exists) should increase with Recurrence Score (i.e. powered to find a significant interaction of randomized treatment assignment and the continuous Recurrence Score)
  - Optimal cutpoint for giving chemo will be estimated at the end of the trial

Trial predicts a significant interaction followed by estimation of the optimal point to treat

- The dashed line represents the computed cutpoint above which chemotherapy is recommended



National Cancer Institute

Challenges in retrospective assessment of predictive tests

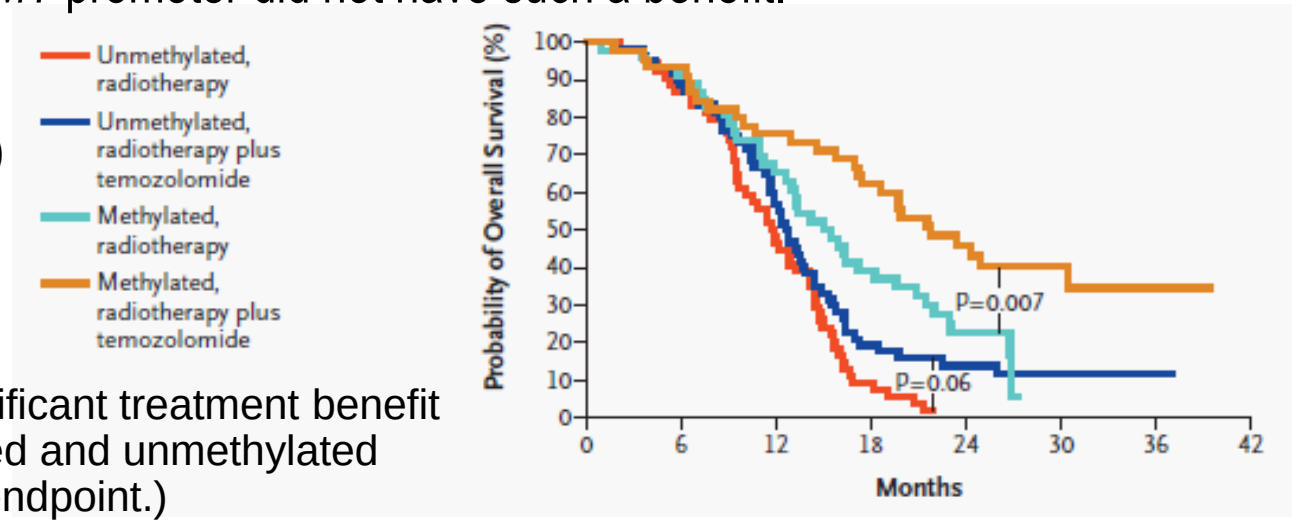
- Prospective-retrospective approach best if feasible
  - Simon, Paik, Hayes 2009, *JNCI*
- Specimens from a relevant trial might not exist
- Randomized clinical trials adequately powered to detect treatment effects are often not sufficiently powered to establish predictive ability
  - Non-significant $\neq$ no effect
- Most relevant endpoint?
- False positives: multiple markers, optimized cutpoint
- "Convenience" retrospective sets: Specimens from a non-randomized or non-comparable control group can lead to uninterpretable results

National Cancer Institute

Assessment of predictive tests: Power pitfalls

CONCLUSION: "Patients with glioblastoma containing a methylated *MGMT* promoter benefited from temozolomide, whereas those who did not have a methylated *MGMT* promoter did not have such a benefit."

(NEJM 2005; 352: 997-1003)



(Statistically significant treatment benefit in both methylated and unmethylated groups for PFS endpoint.)

Overall Survival (OS)	Hazard ratio (95% CI)	Median OS (months)	2-yr OS (%)	P-value
MGMT Methylated				
RT (n=46)	1.00	15.3 (13.0-20.9)	22.7 (10.3-35.1)	
RT+TMZ (n=46)	0.51 (0.31-0.84)	21.7 (17.4-30.4)	46.0 (31.2-60.8)	0.007
MGMT Unmethylated				
RT (n=54)	1.00	11.8 (9.7-14.1)	< 2	
RT+TMZ (n=60)	0.69 (0.47-1.02)	12.7 (11.6-14.4)	13.8 (4.8-22.7)	0.06

National Cancer Institute

Assessment of predictive tests: Power pitfalls

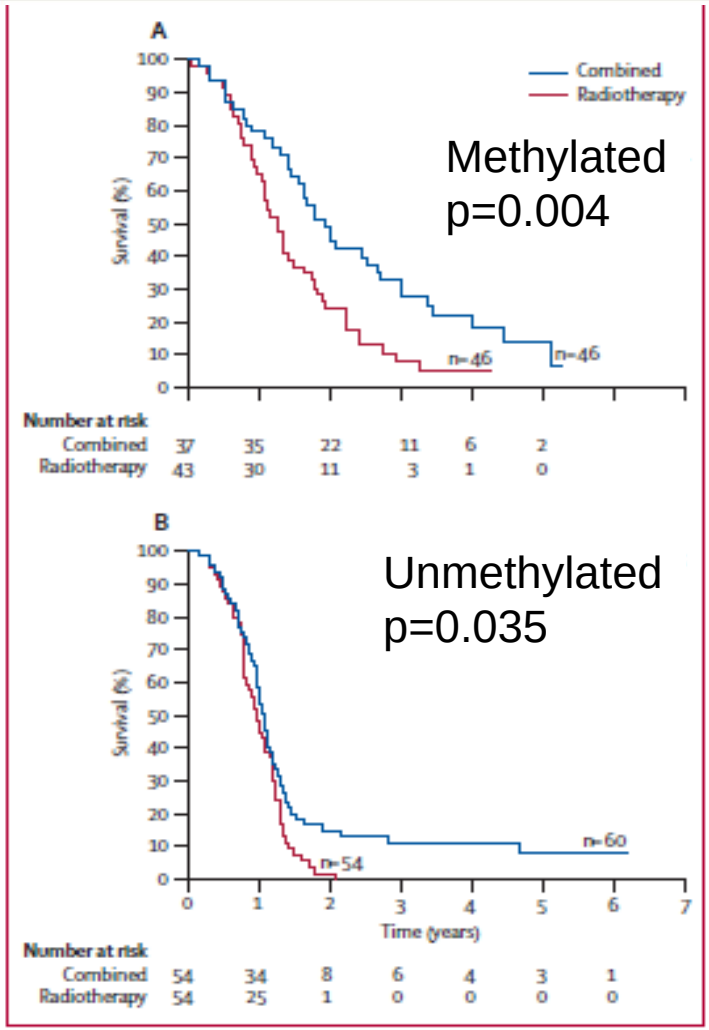


Figure 4: **Kaplan-Meier estimates of overall survival by *MGMT* status**
Patients with methylated *MGMT* (A). Patients with unmethylated *MGMT* (B).

(Salvage therapies, including TMZ, confound OS endpoint.)

With follow-up to 5 years, the OS difference became significant in favor of RT+TMZ even in the unmethylated *MGMT* group (not adjusted for testing in 2 subgroups).
(*Lancet Oncol* 2009; 10: 459-466)

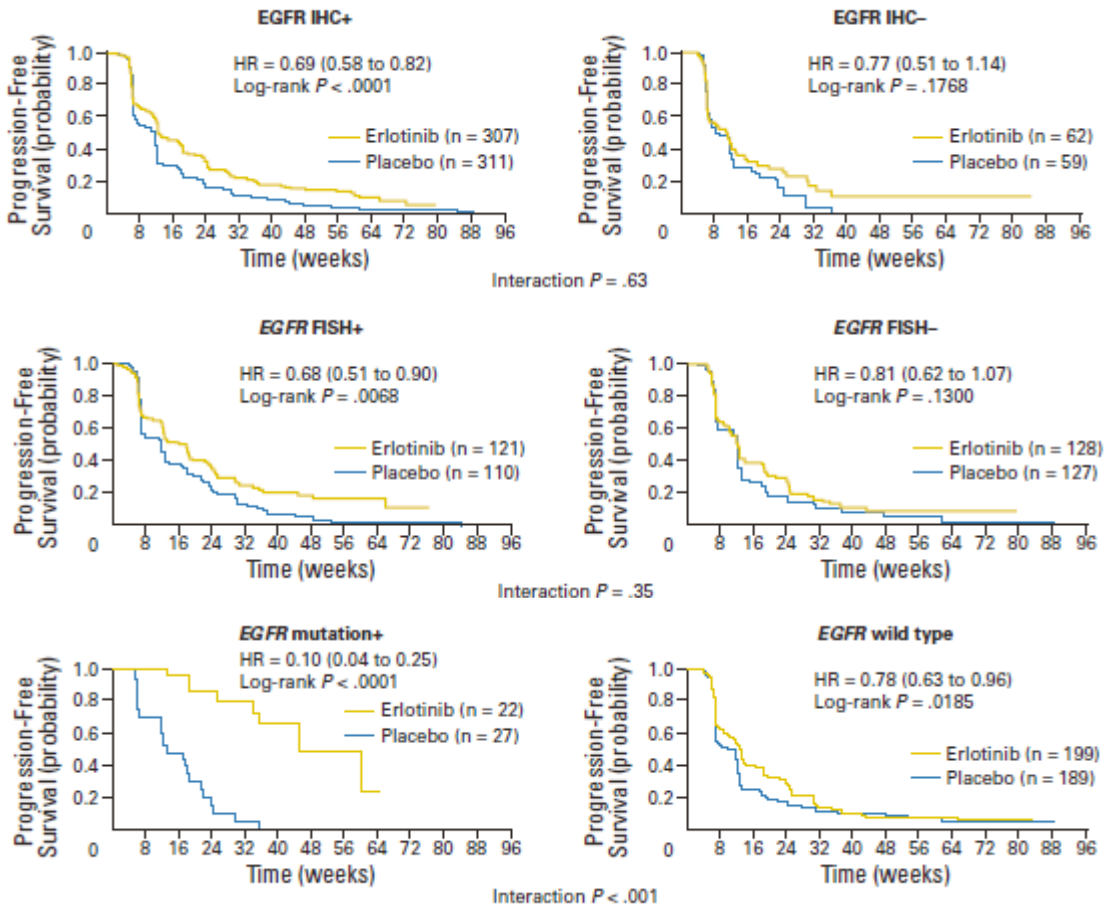
	Hazard ratio (95% CI)	5-yr OS (%)
MGMT Methylated		
RT	1.0	5.2 (1.0-15.0)
RT + TMZ	0.3 (0.2-0.4)	13.8 (4.5-28.2)
MGMT Unmethylated		
RT	1.0	0
RT+TMZ	0.6 (0.4-0.8)	8.3 (2.7-18.0)

National Cancer Institute

SATURN Trial: Evaluation of EGFR "status" as a predictive marker (OS)

Erlotinib Versus Placebo Maintenance Therapy in Advanced NSCLC

Brugger et al 2011, *J Clin Oncol 29*: 4113-4120



Statistically significant PFS improvement both in overall ITT population and in EGFR IHC+ subgroup; Absolute benefit small (1 mo. median PFS); EGFR mutation appears to be where the action is, but EGFR WT still benefit

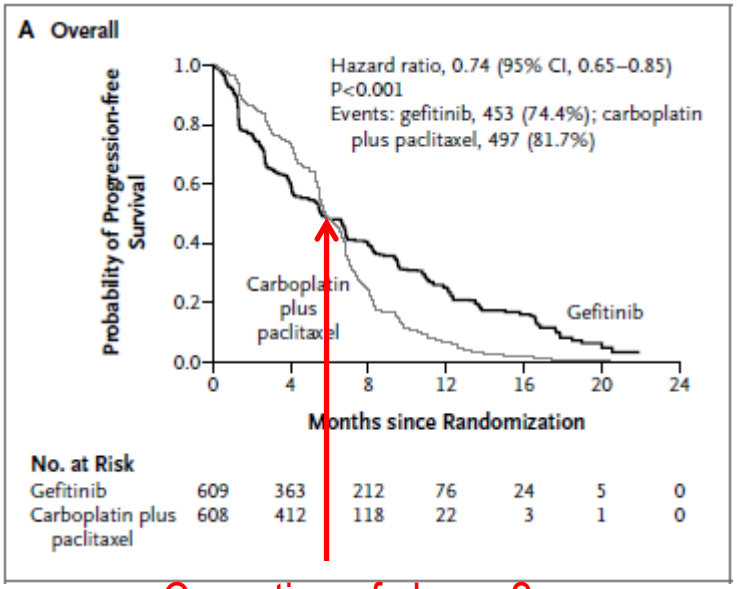
National Cancer Institute

IPASS Trial: Evaluation of EGFR mutation as a predictive marker (PFS)

Mok et al 2009, *N Engl J Med* 361: 947-57

EGFR mutation is:

- Positive prognostic factor (established in untreated groups)
- Positive predictive factor for gefitinib benefit

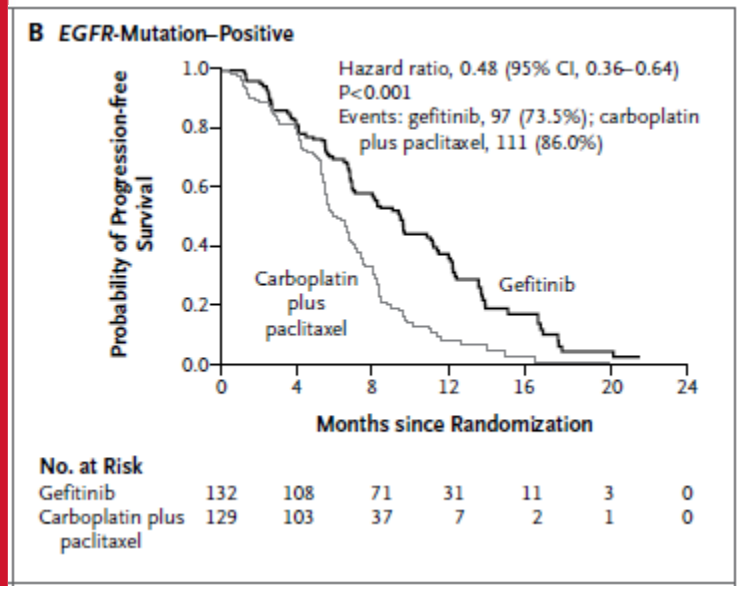


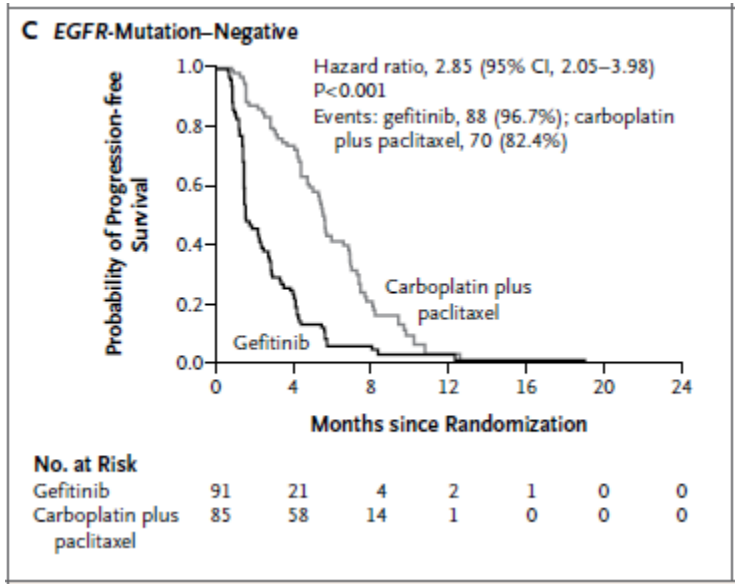
Cessation of chemo?

IPASS: Phase III 1st line advanced adeno NSCLC

Gefitinib (EGFR inhibitor) vs. carboplatin+ paclitaxel (chemo)

EGFR mutation status by treatment interaction, P<0.001)





National Cancer Institute

IPASS Trial: Evaluation of EGFR mutation as a predictive marker (OS)

Gefitinib Versus Chemo in NSCLC: Biomarker and Survival Analyses

Fukuoka et al 2011, *J Clin Oncol* 29: 2866-2874

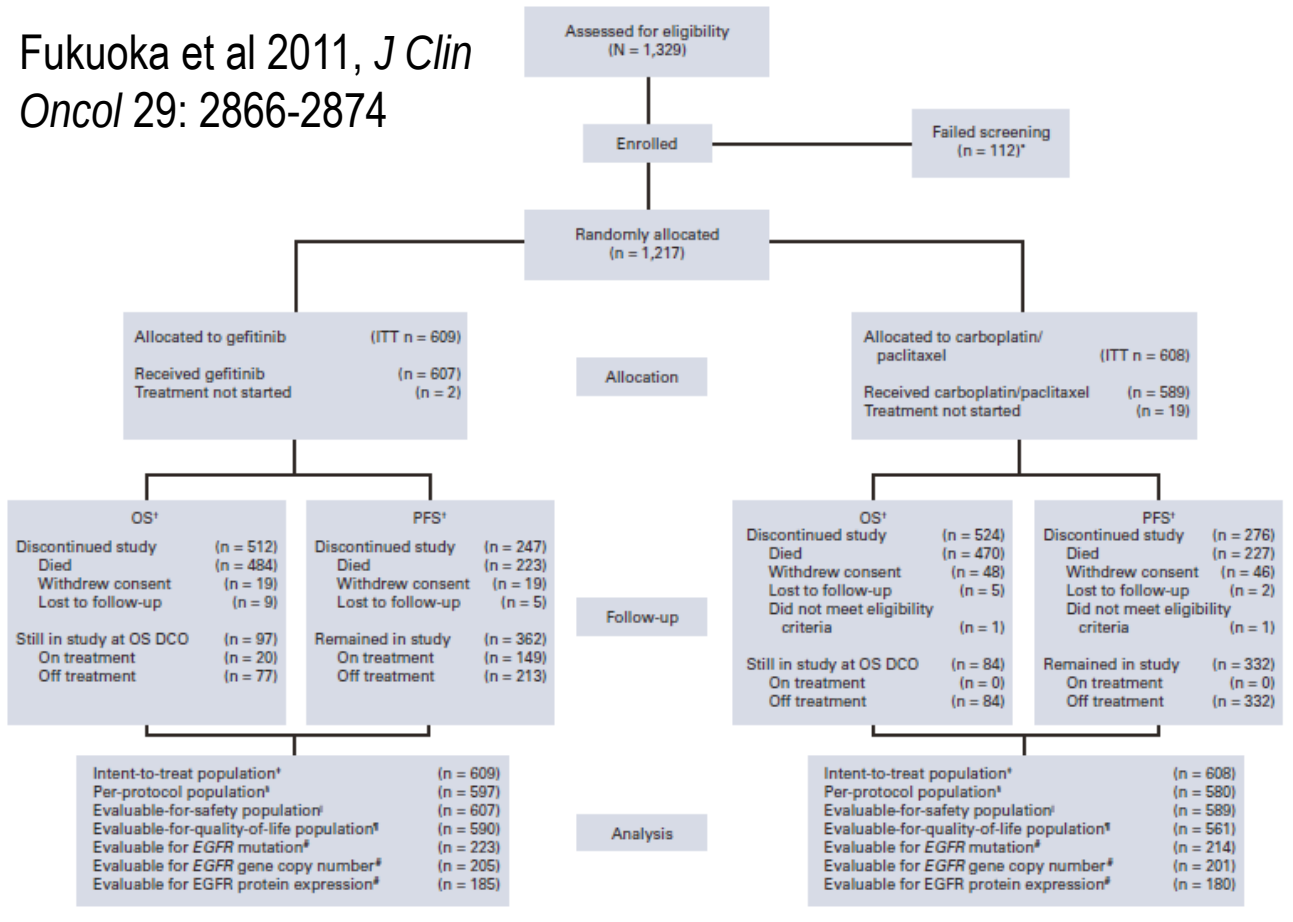
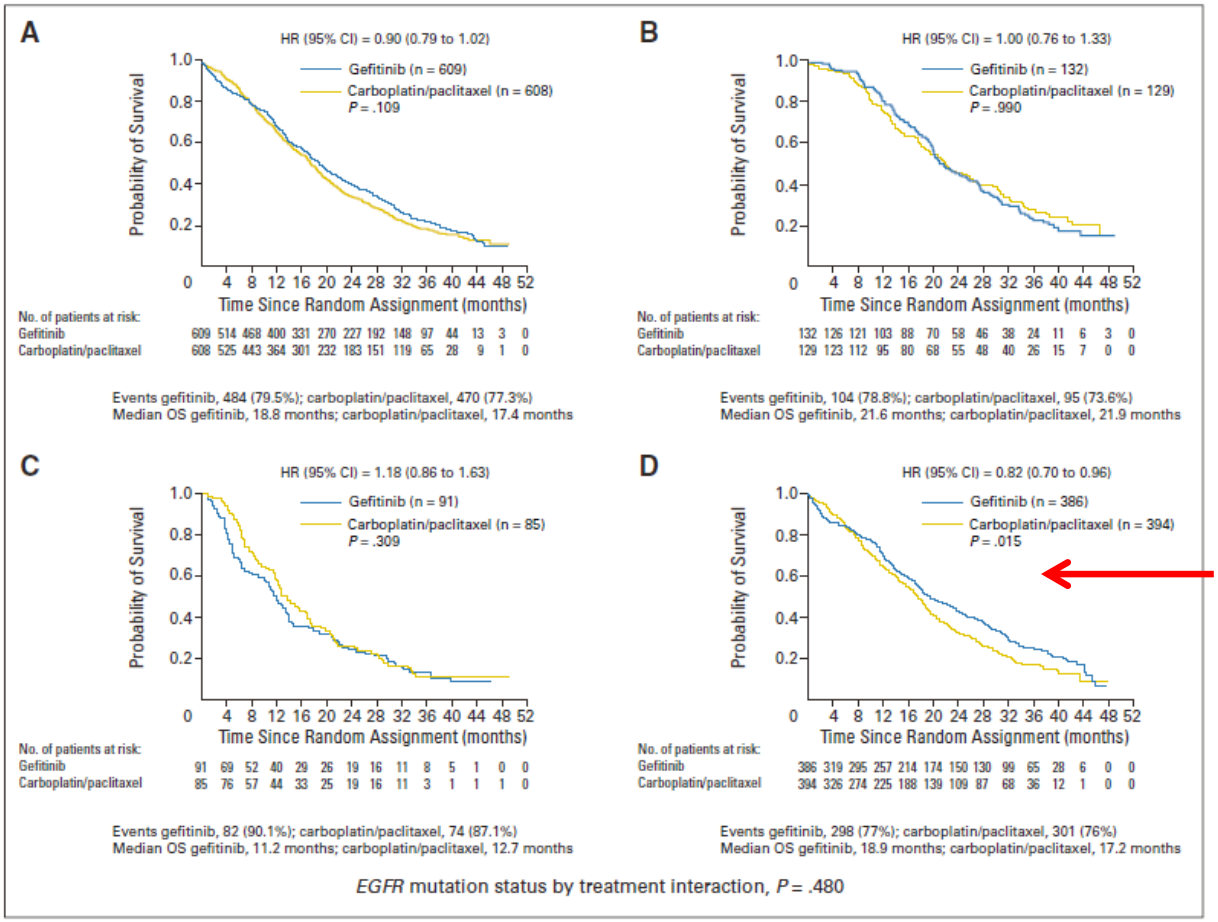


Fig 1. CONSORT diagram. (*) Among the 112 patients who failed screening, the main reasons for exclusion were abnormal serum creatinine (> 1.5 × upper limit of reference range)/creatinine clearance (≤ 60 mL/min) levels; untreated CNS metastases; or low neutrophil (< 2.0 × 10^9/L), platelet (< 100 × 10^9/L), or hemoglobin (< 10 g/dL) counts. (†) Cutoff dates: June 14, 2010, for overall survival (OS) and April 14, 2008, for progression-free survival (PFS). (‡) All patients who were randomly assigned to a study group were included in the intent-to-treat (ITT) analysis. (§) Patients who did not deviate substantially from the inclusion and exclusion criteria at entry or from the protocol were included in the per-protocol analysis. (‖) All patients who received at least one dose of study treatment were included in the safety analysis. (¶) All patients with a baseline and at least one postbaseline quality-of-life assessment that could be evaluated were included in the quality-of-life analysis. (#) All patients in the ITT population with an evaluable tumor sample. Of 683 patients (56%) who provided samples, 118 were cytology samples, and 128 were histologic samples of insufficient quality and were therefore not included in the main analysis. DCO, data cutoff; EGFR, epidermal growth factor receptor.

National Cancer Institute

IPASS Trial: Evaluation of EGFR mutation as a predictive marker (OS)

Gefitinib Versus Chemo in NSCLC: Biomarker and Survival Analyses



Only stat. significant benefit in subgroup with EGFR mutation unknown

Fig 2. Kaplan-Meier curves for overall survival (OS) in the overall population and by epidermal growth factor receptor (*EGFR*) mutation status (intent-to-treat population). Hazard ratio (HR) < 1 implies a lower risk of death for patients treated with gefitinib. Cox analysis with covariates (performance status [0-1, 2], smoking history [never, light ex-smoker], and sex). (A) Overall population. (B) Patients with *EGFR* mutation–positive tumors. (C) Patients with *EGFR* mutation–negative tumors. (D) Patients with *EGFR* mutation status unknown tumors.

High rates of crossover; other EGFR-inhibitors showed benefit in unselected patients in second line setting (Fukuoka et al 2011, *J Clin Oncol* 29: 2866-2874)

National Cancer Institute

It only gets worse

- Alternative data sources
  - Observational databases
    - Confounding with treatment selection
    - Unknown/incomplete covariates
    - Non-comparable or unreliable marker assays
  - Anecdotal evidence
- New assays
  - High-throughput omics, e.g., NGS
  - Complex mathematical predictors

National Cancer Institute

Going Forward

- What criteria need to be satisfied to establish clinical utility?
- How do we obtain reliable evidence?
- What can NCI do to help?

Thank you!