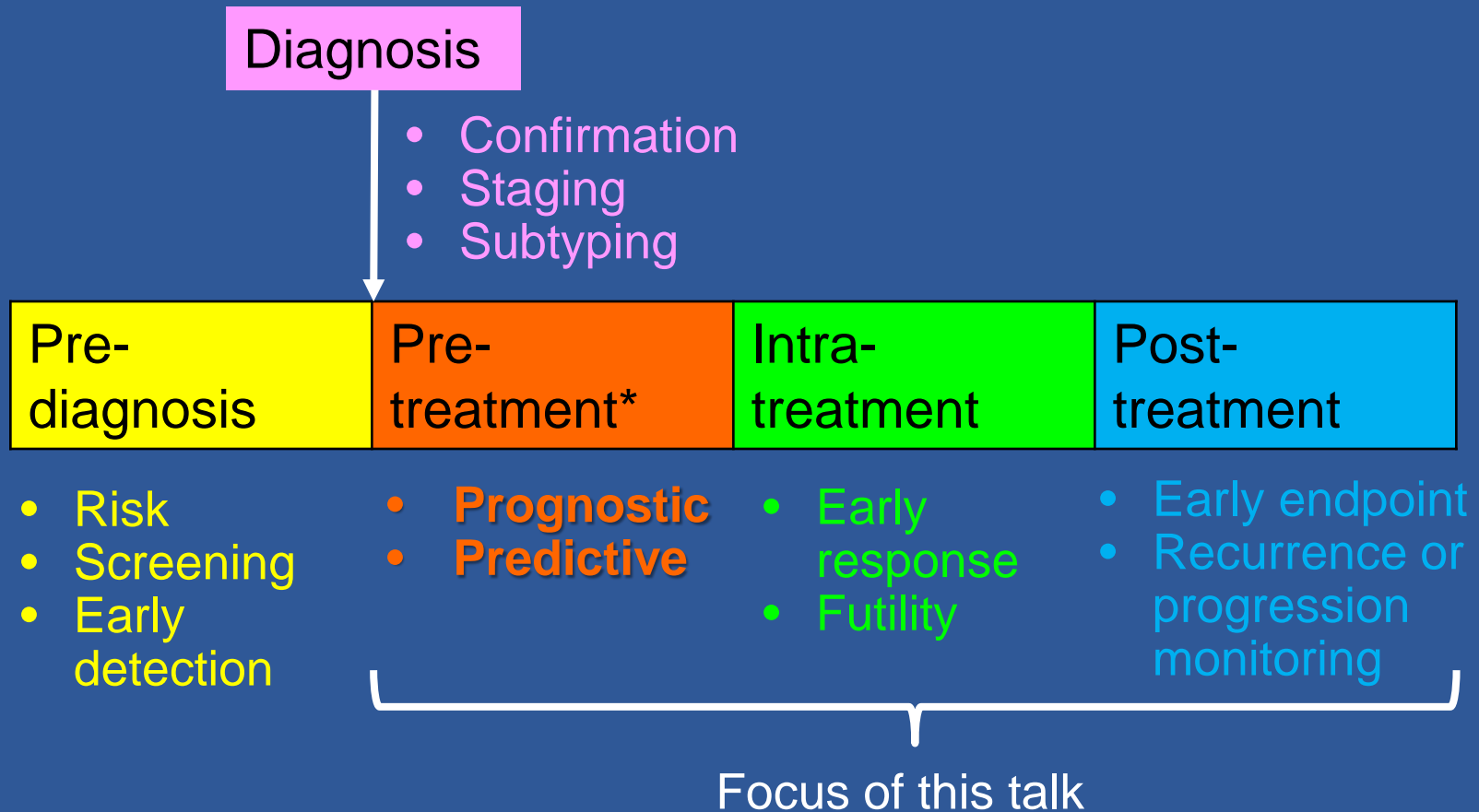


**STATISTICAL CONSIDERATIONS FOR
TRIALS (OR STUDIES) DESIGNED TO
DETERMINE CLINICAL UTILITY OF cfDNA
ASSAYS**

Lisa M McShane, PhD
Biostatistics Branch, Biometric Research Program
Division of Cancer Treatment and Diagnosis
U. S. National Cancer Institute

*Rockville, MD
September 29-30, 2016*

POTENTIAL CLINICAL ROLES FOR cfDNA ASSAYS*



*Statistical principles apply more broadly, and illustrative examples used in this talk are not restricted to cfDNA assays

VALIDATION OF A BIOMARKER TEST

- **Analytical validation** - Establish that the performance characteristics of the locked down test (i.e., completely specified) are acceptable in terms of sensitivity, specificity, accuracy, precision, as applicable
 - Technical performance
 - Says nothing about clinical correlations
 - Poor analytical validation may impede clinical validation
- **Clinical validation** – Demonstrate that the test result is associated with a clinical outcome of interest
 - Establish clinical associations
 - Many statistically significant p-values in published literature
 - Not guaranteed to be useful
- **Clinical utility** – Demonstrate that use of the test results in a favorable benefit to risk ratio for the patient
 - Better clinical outcome, safer, cheaper, easier, less invasive, etc.

BEST RESOURCE

(BIOMARKERS, ENDPOINTS, AND OTHER TOOLS)

Product of the Biomarker Working Group charged by the **FDA-NIH Joint Leadership Council** to develop a glossary of harmonized terminology for biomarkers and endpoints

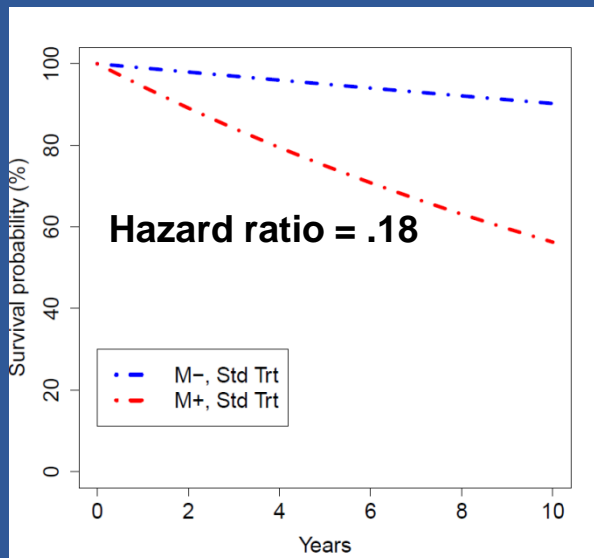
Chairs: Robert Califf (FDA), Pamela McInnes (NIH/NCATS), Michael Pacanowski (FDA)

Project Leads: Nina Hunter (FDA), Melissa Robb (FDA)

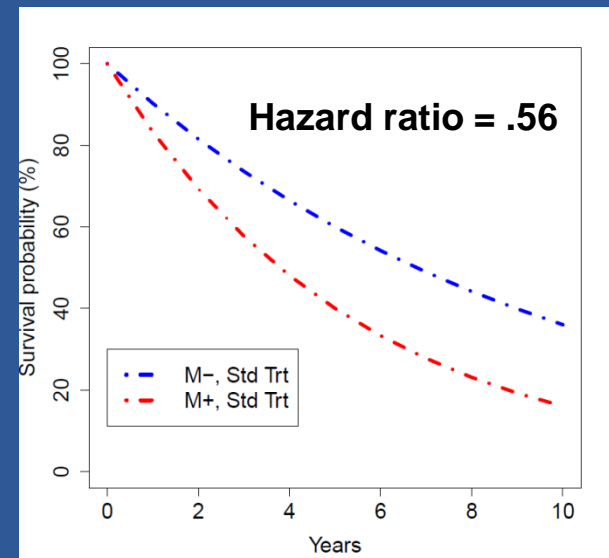
PROGNOSTIC BIOMARKER

- Associated with clinical outcome in absence of therapy (natural course) *or with standard therapy all patients are likely to receive*
- **FDA-NIH:** A biomarker used to identify likelihood of a clinical event, disease recurrence or progression

Clinically useful: Good prognosis group (M-) may forego additional therapy



Possibly (?) helpful in drug development, probably not that useful clinically



(M = biomarker)

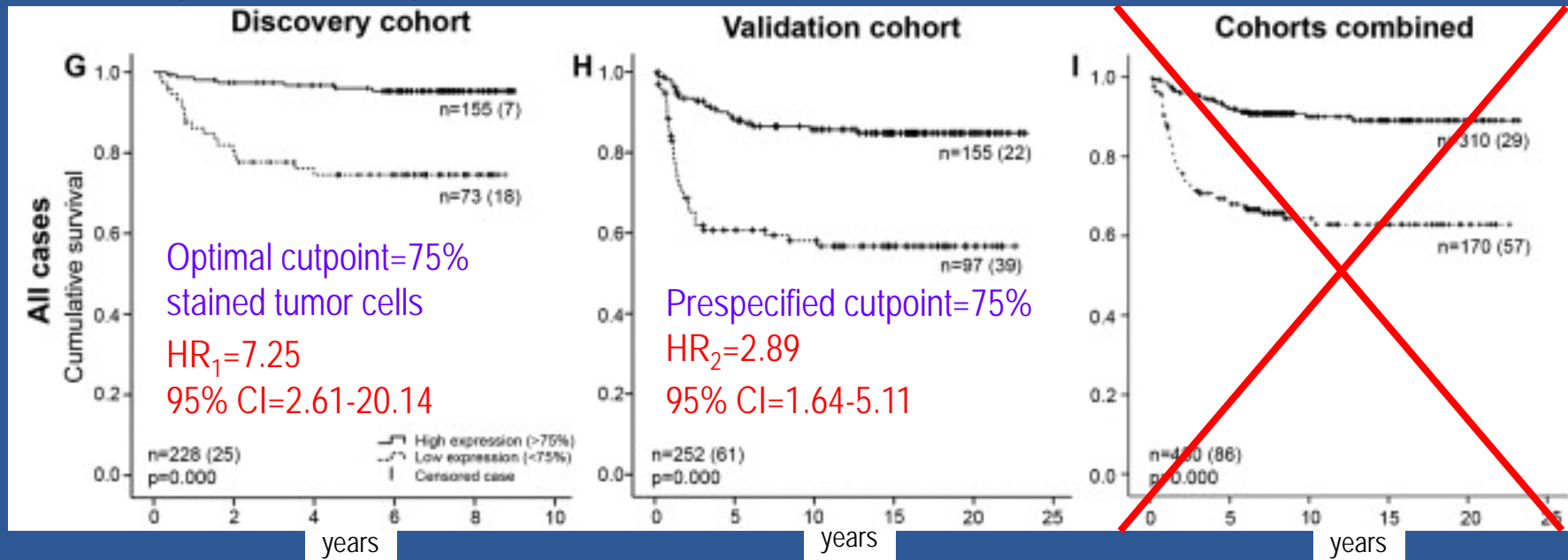
CHALLENGES IN EVALUATION OF PROGNOSTIC BIOMARKERS

- Cutpoint optimization on biomarker can lead to biased effect estimates (e.g., HR=hazard ratio)
- Composition of patient cohort matters
- Absolute risk (e.g., cumulative survival or difference in cumulative survival) may be as important or more than “relative” risk (e.g., hazard ratio) for clinical purposes

CHALLENGES IN EVALUATION OF PROGNOSTIC BIOMARKERS

Endometrial cancer prognostic biomarker *ASRGL1* example

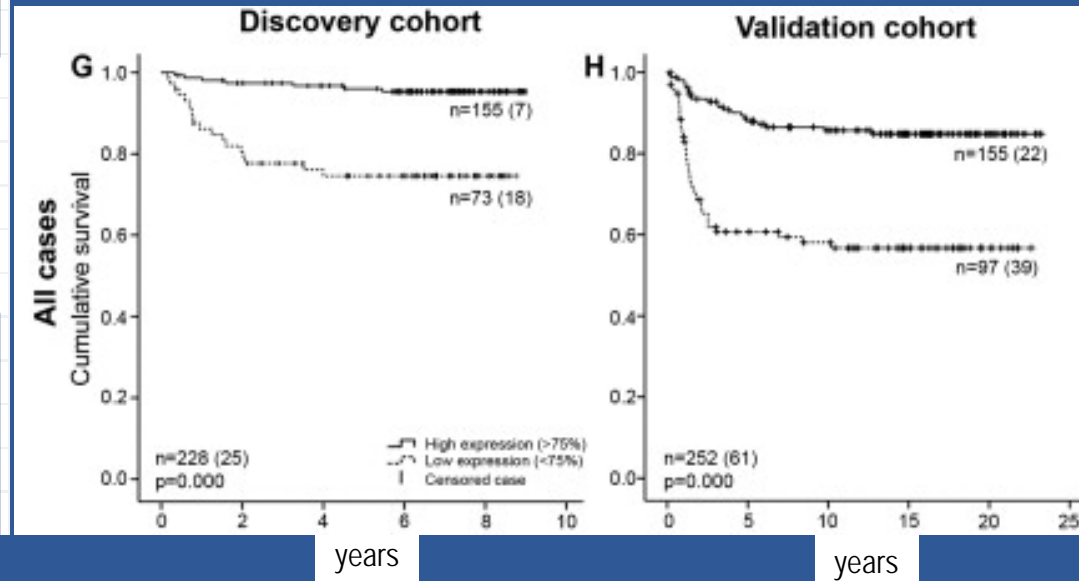
- Cutpoint optimization on biomarker can lead to biased effect estimates
- Absolute risk (e.g., cumulative survival or difference in cumulative survival) may be as important or more than “relative” risk (e.g., hazard ratio) for clinical purposes
- Composition of patient cohort matters



CHALLENGES IN EVALUATION OF PROGNOSTIC BIOMARKERS *(ASRGL1 example cont.)*

Variable	Discovery cohort	Validation cohort
	n (%)	n (%)
Age		
≤ 65	99 (43.4)	127 (50.4)
> 65	129 (56.6)	125 (49.6)
Figo stage (2009)^a		
I–II	190 (83.3)	205 (81.7)
III–IV	38 (16.7)	46 (18.3)
Grade		
Grade 1	117 (51.3)	56 (22.2)
Grade 2	62 (27.2)	137 (54.4)
Grade 3	36 (15.8)	36 (14.3)
Non-endometrioid ^b	13 (5.7)	23 (9.1)
MI^c		
≤ 50%	159 (69.7)	128 (59.8)
< 50%	69 (30.3)	86 (40.2)
LVI		
No	198 (86.8)	n.d.
Yes	30 (13.2)	n.d.
Tumor size		
≤ 2 cm	79 (34.6)	n.d.
> 2 cm	149 (65.4)	n.d.

- Cutpoint optimization can lead to biased effect estimates
- Composition of patient cohort matters
- Absolute risk (e.g., cumulative survival or difference in cumulative survival) may be as important or more than “relative” risk (e.g., hazard ratio) for clinical purposes



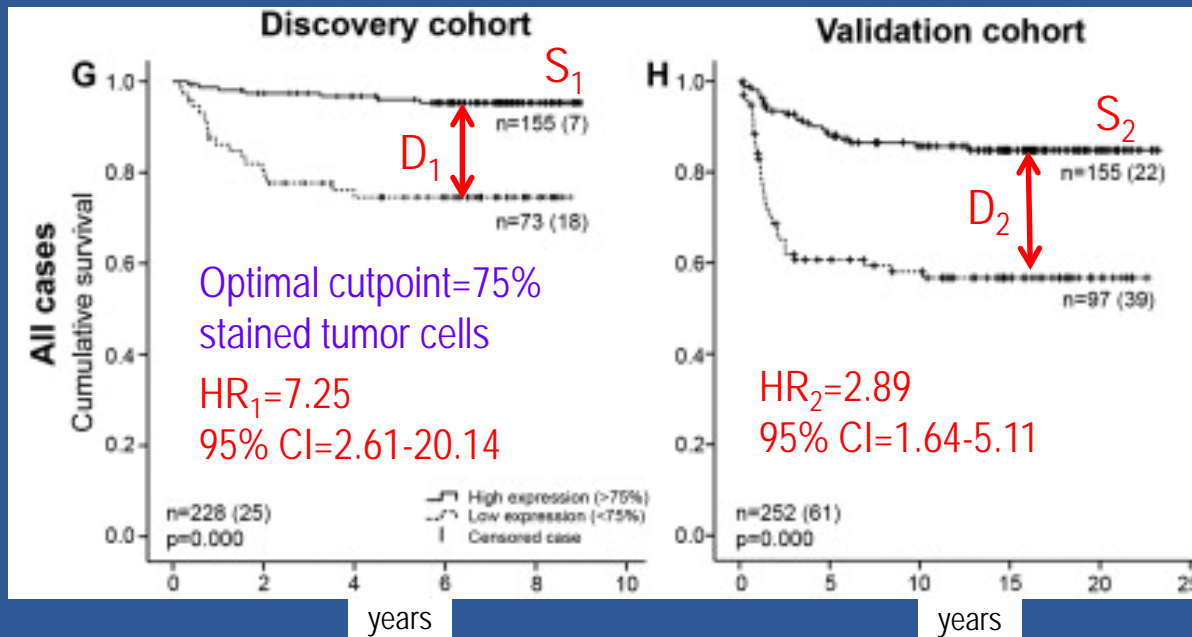
^aData missing from 1 validation cohort case.

^b8 and 8 serous carcinomas; 3 and 15 clear cell carcinomas; 2 and 0 mixed type carcinomas in the discovery and validation cohorts, respectively.

^cData missing from 38 validation cohort cases.

CHALLENGES IN EVALUATION OF PROGNOSTIC BIOMARKERS *(ASRGL1 example cont.)*

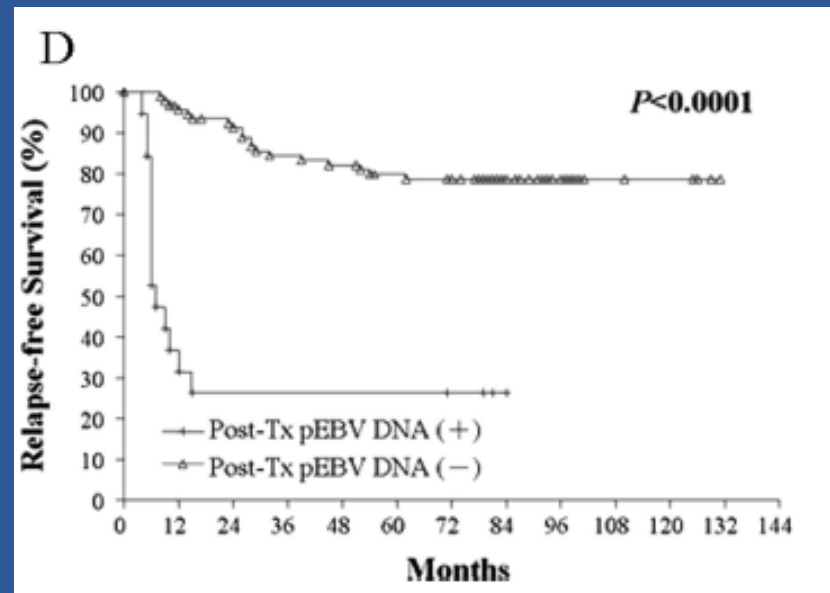
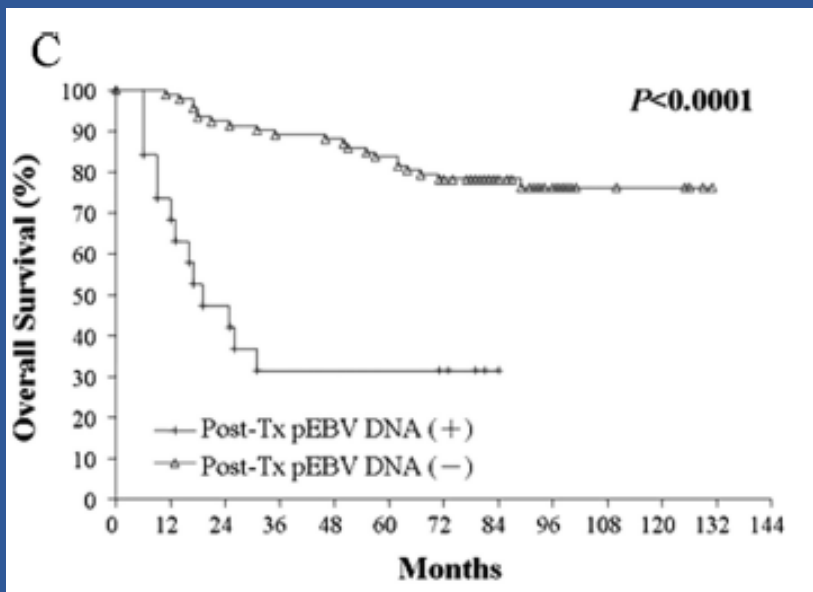
- Cutpoint optimization on biomarker can lead to biased effect estimates
- Composition of patient cohort matters
- Absolute risk (e.g., cumulative survival or difference in cumulative survival) may be as important or more than “relative” risk (e.g., hazard ratio) for clinical purposes



- $HR_1 > HR_2$ but $D_1 < D_2$
- $S_2 < S_1$: Is S_2 sufficiently good that one would base a clinical therapeutic decision on it (e.g., no further treatment for favorable group but more treatment for unfavorable group) ?

EBV EARLY RESPONSE/PROGNOSTIC BIOMARKER: A LESSON IN ASSAY STANDARDIZATION & REPRODUCIBILITY

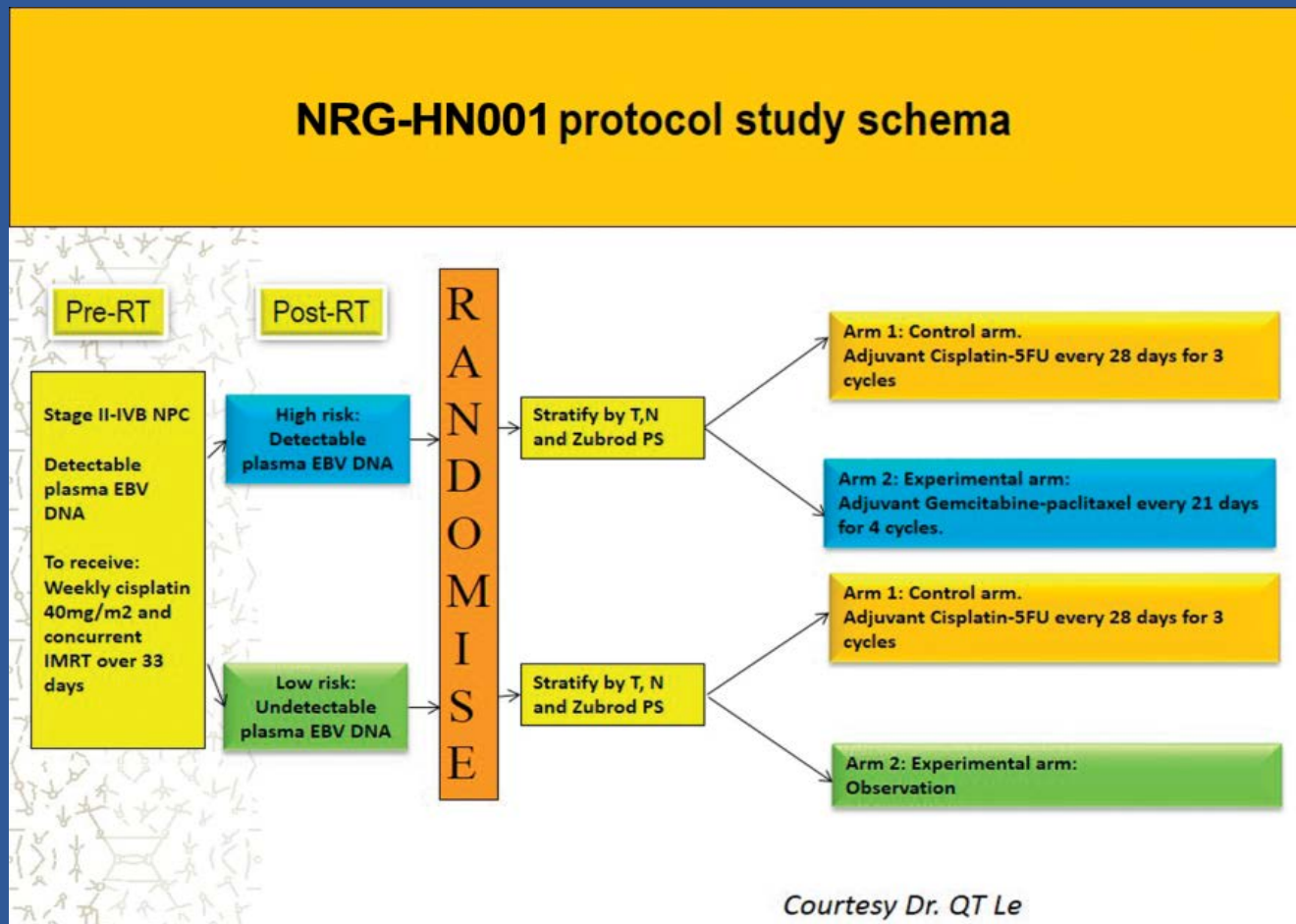
Independent validation of plasma Epstein-Barr Virus (EBV) DNA as an early response/prognostic biomarker (n=111 patients)



Overall survival and relapse-free survival, respectively, according to the plasma EBV DNA “detectable” status one week after the completion of radiotherapy for patients with advanced nasopharyngeal carcinoma (stage III-IV, M0) who received 10 weeks of chemotherapy followed by radiotherapy

PLANNING FOR A CLINICAL TRIAL: IMPORTANCE OF ASSAY STANDARDIZATION AND REPRODUCIBILITY

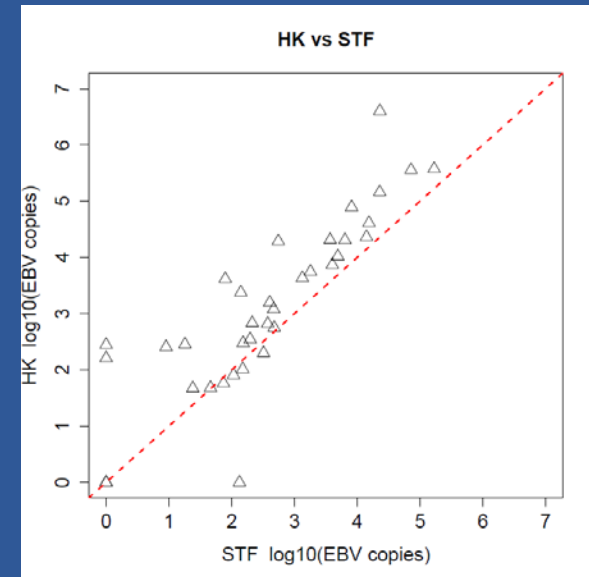
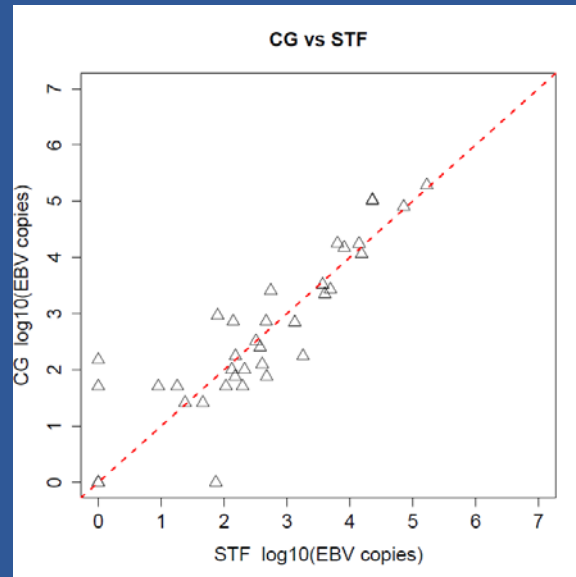
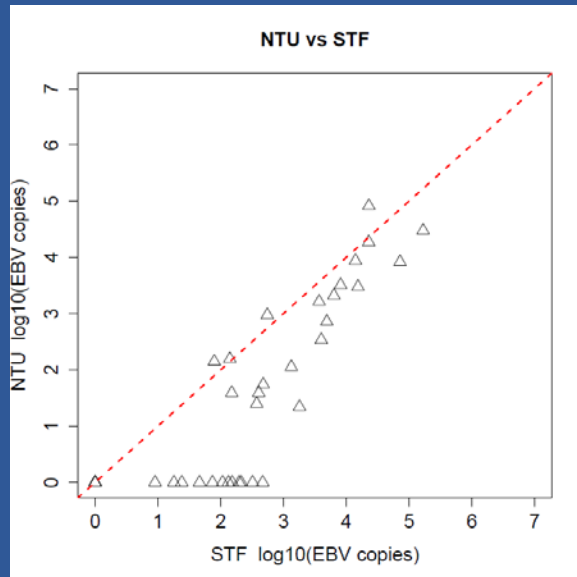
Plasma Epstein-Barr Virus (EBV) DNA as a early response/prognostic stratifier in treatment of nasopharyngeal carcinoma



EBV ASSAY STANDARDIZATION (cont.)

An International Collaboration to Harmonize the Quantitative Plasma Epstein-Barr Virus DNA Assay for Future Biomarker-Guided Trials in Nasopharyngeal Carcinoma (4 sites STF, NTU, CG, HK)

BEFORE HARMONIZATION
(40 samples)



EBV DETECTION RATES: NTU 58%, CG 93%, HK 93%

Le et al, *Clin Cancer Res* 2013;19:2208-2215

EBV ASSAY STANDARDIZATION (cont.)

An International Collaboration to Harmonize the Quantitative Plasma Epstein-Barr Virus DNA Assay for Future Biomarker-Guided Trials in Nasopharyngeal Carcinoma

Intraclass correlation (ICC) for each site when compared to index site (STF) before and after the harmonization of PCR master mixes and calibrators

ICC =
proportion of
total variation
explained by
between-
subject
variation;
 $0 \leq \text{ICC} \leq 1$ with
1 indicating
perfect
reproducibility

Site	Pre-harmonization ICC (95% CI) N = 40	Post harmonization ICC (95% CI) N = 10
NTU vs. STF	0.62 (0.39-0.78)	0.83 (0.50-0.95)
CG vs. STF	0.70 (0.50-0.83)	0.95 (0.83-0.99)
HK vs. STF	0.59 (0.35-0.76)	0.96 (0.86-0.99)

Le et al, *Clin Cancer Res* 2013;19:2208-2215

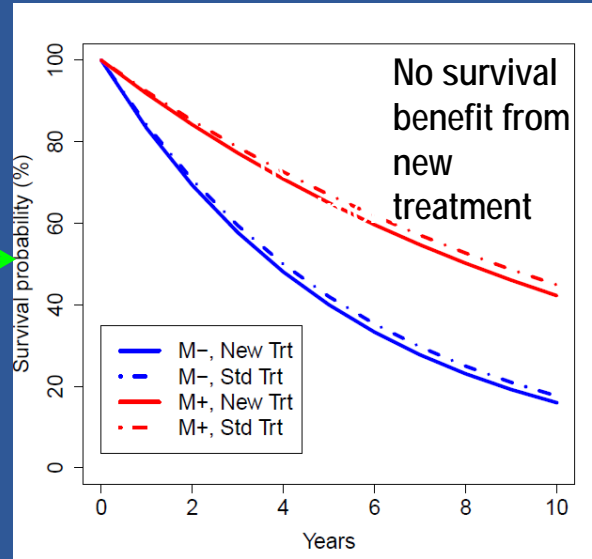
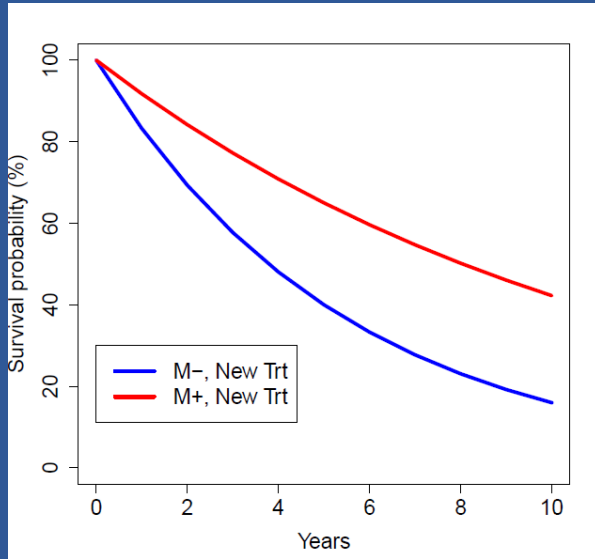
PREDICTIVE BIOMARKER

- Associated with benefit or lack of benefit (potentially even harm) from a particular therapy relative to other available therapy
- **FDA-NIH:** A biomarker used to identify individuals who are more likely than similar patients without the biomarker to experience a favorable or unfavorable effect from a specific intervention or exposure
- 3 sub-classes:
 - Treatment-selection biomarker
 - Enrichment-predictive biomarker
 - Response-predictive biomarker

PREDICTIVE BIOMARKER SUBCLASSES

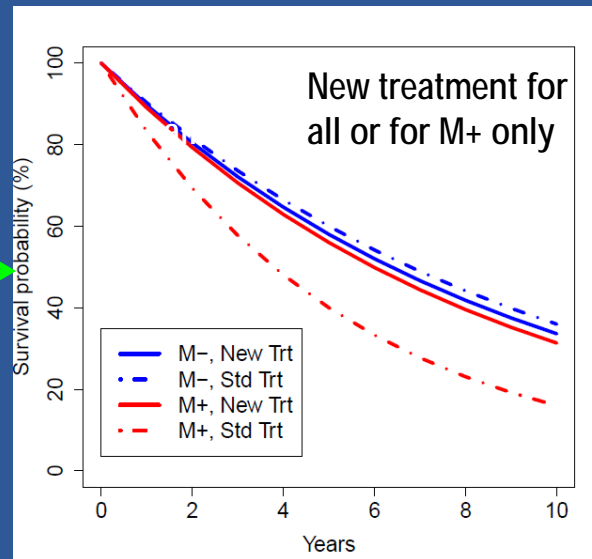
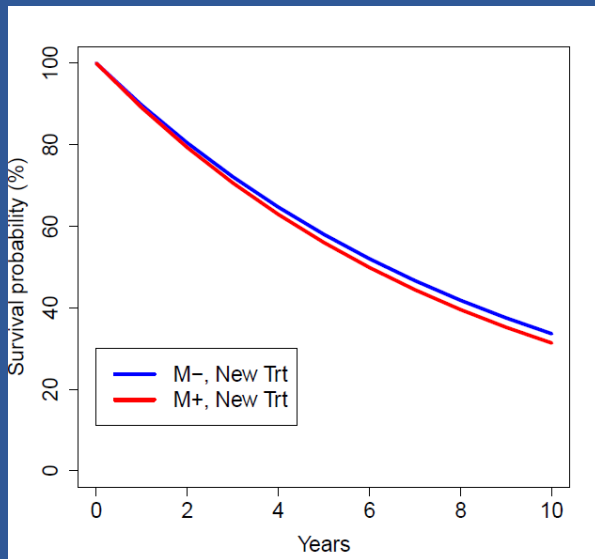
- **Treatment-selection biomarker**
 - Effect of a particular treatment relative to some other treatment (which may be no treatment) varies depending on the value of the biomarker
 - Sometimes called treatment-effect modifier, treatment-guiding
- **Enrichment-predictive biomarker**
 - Used to enrich or select the patient population for clinical trials, particularly for targeted anti-cancer agents
 - Prior evidence suggests biomarker negative cases not likely to benefit
 - Sometimes called selection or efficacy biomarker
- **Response-predictive biomarker**
 - Used to predict tumor response (or possibly prolonged progression-free survival or stable disease), mostly in early phase trials
 - Indicator of likelihood of drug activity (e.g., single arm phase II trial)
 - Improved response not always associated with prolonged survival

PROGNOSTIC VS. PREDICTIVE: IMPORTANCE OF CONTROL GROUPS



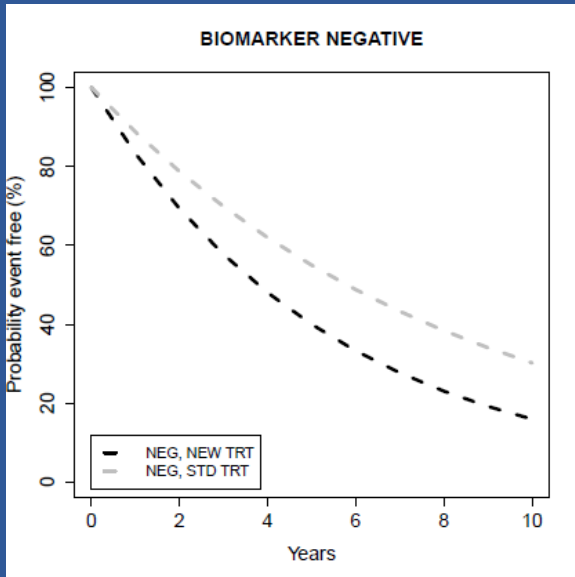
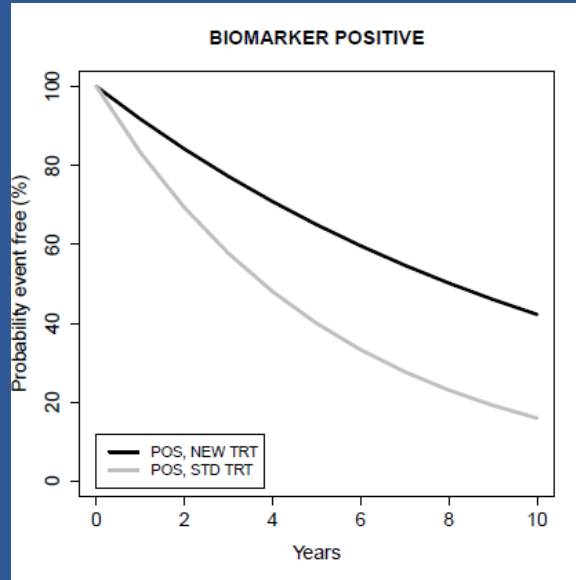
Prognostic
but not
predictive

(M =
biomarker)



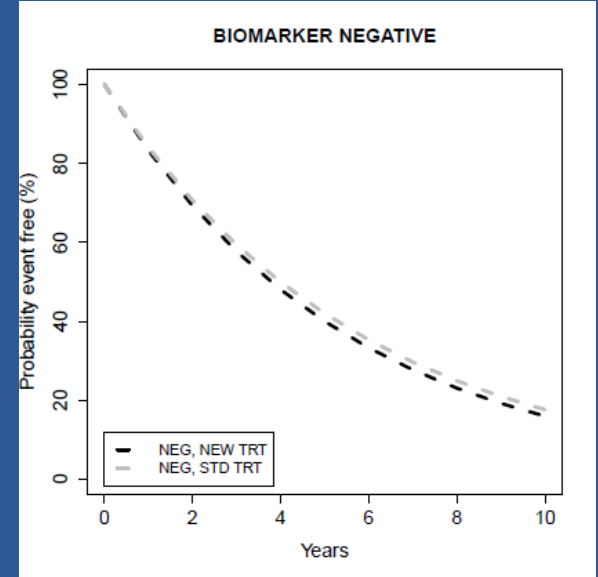
Prognostic
and
predictive

CLINICALLY USEFUL PREDICTIVE BIOMARKER

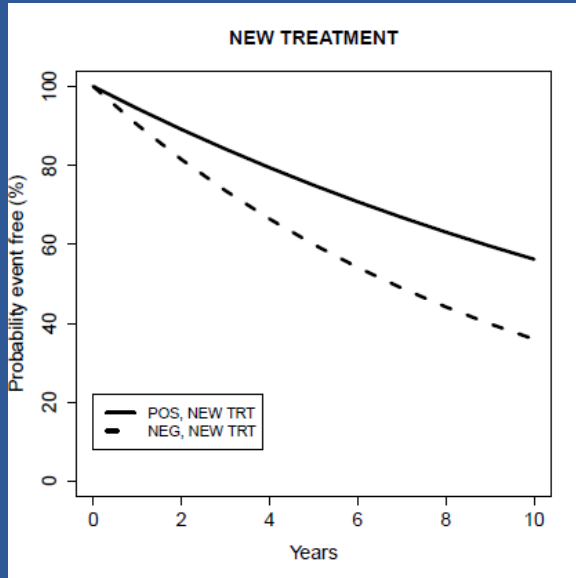


↑
BIOMARKER POS:
NEW TRT > STD TRT

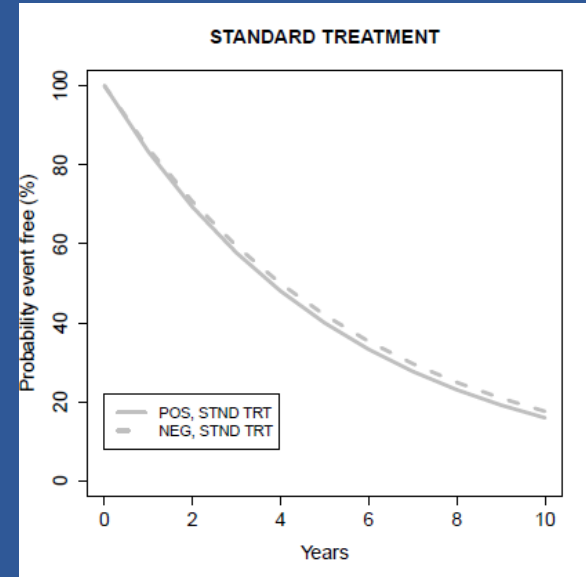
← BIOMARKER NEG:
NEW TRT ≤ STD TRT →



HOW *NOT* TO PARSE EVIDENCE FOR A CANDIDATE PREDICTIVE BIOMARKER



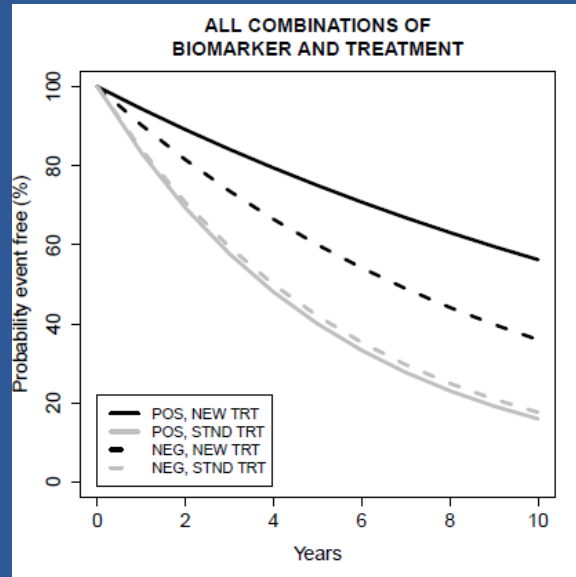
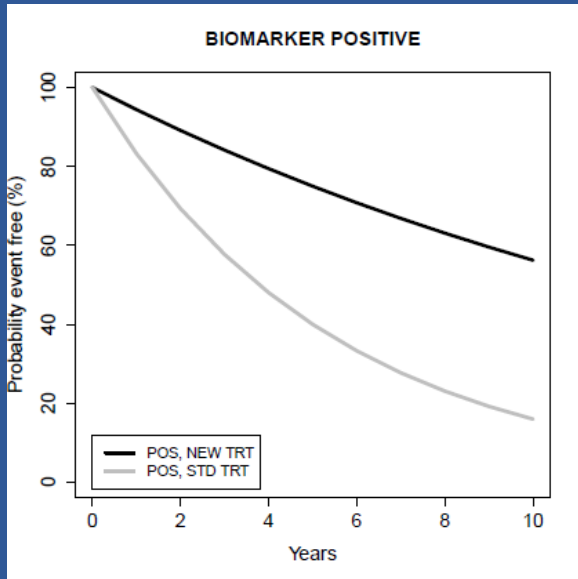
NEW TREATMENT:
BIOMARKER POS > BIOMARKER NEG



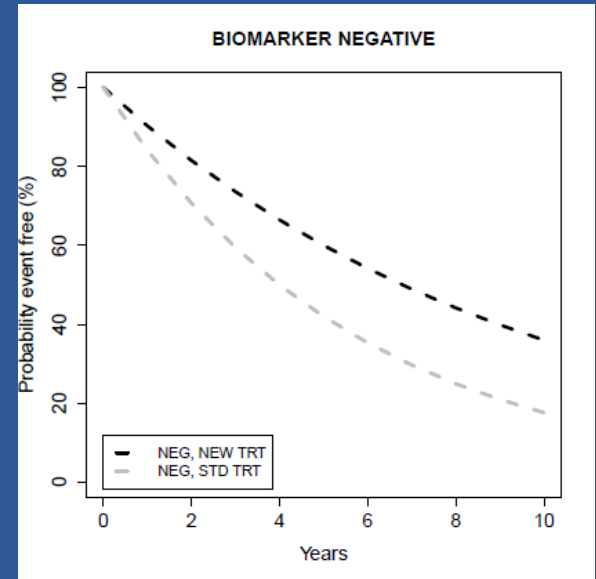
STANDARD TREATMENT:
BIOMARKER POS = BIOMARKER NEG
(NOT PROGNOSTIC)

HOW TO CORRECTLY PARSE EVIDENCE FOR A CANDIDATE PREDICTIVE BIOMARKER

BIOMARKER POS:
NEW TRT > STD TRT



BIOMARKER NEG:
NEW TRT > STD TRT



NOW WE SEE THAT THE BIOMARKER IS NOT USEFUL FOR SELECTION OF NEW TREATMENT (BECAUSE ALL PATIENTS BENEFIT)

STATISTICAL LANGUAGE FOR PREDICTIVE BIOMARKERS: "TREATMENT-BY-BIOMARKER INTERACTION"

- Treatment effect (e.g., hazard ratio) varies by biomarker status
 - **Quantitative** interaction: Treatment benefits all patients but by different amounts
 - **Qualitative** interaction: Patients "positive" for the biomarker benefit from the treatment but others receive no benefit or possibly even harm

Generally strive for qualitative interactions

PLASMA IL-6 AS PREDICTIVE BIOMARKER FOR PAZOPANIB VS. PLACEBO?

Results of randomized placebo-controlled phase III trial in metastatic renal-cell cancer (Tran et al, *Lancet Oncol* 2012;13:827-837)

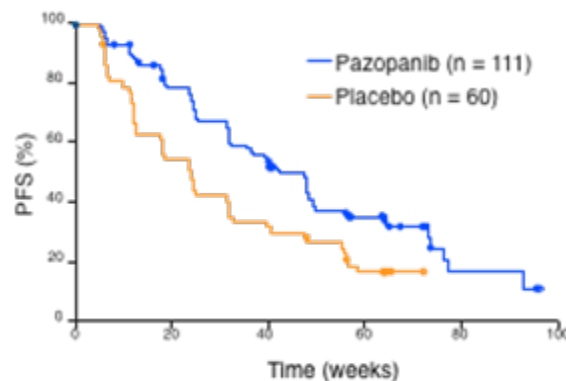
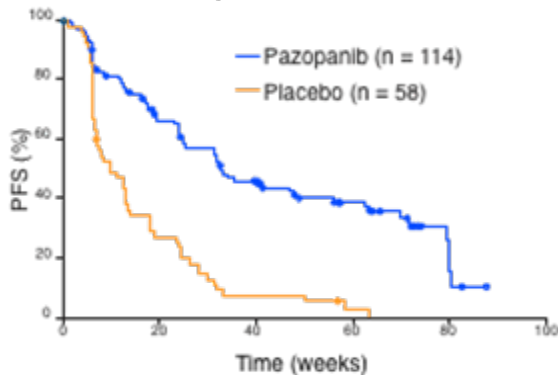
	PFS (weeks)		HR (95% CI)	p value		
	Pazopanib	Placebo		Pazopanib	Placebo	Interaction
Interleukin 6						
Low	42.3	24.0	0.55 (0.38-0.81)	0.445	<0.0001	0.009
High	32.6	9.9	0.31 (0.21-0.44)			

Predictive?
Quantitative
interaction:
P=0.009

Prognostic: P<0.0001

High IL-6

Low IL-6

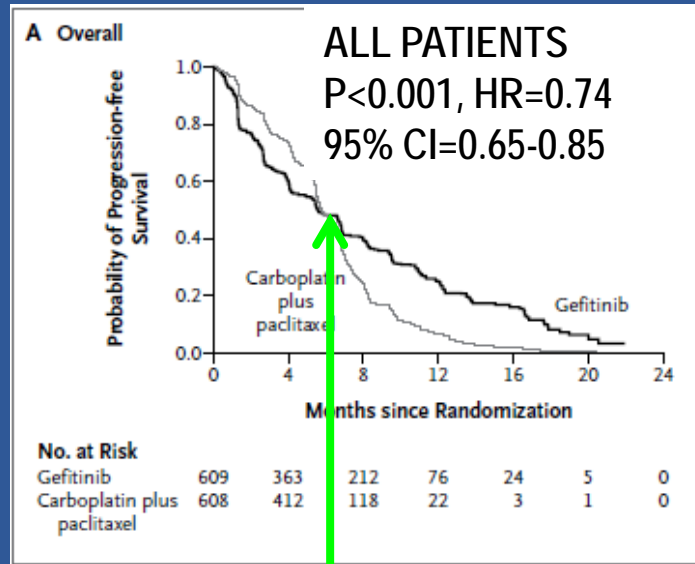


- Does treatment benefit all?
- Is the biomarker cutpoint wrong?

EGFR MUTATION PREDICTIVE FOR PFS BENEFIT WITH GEFITINIB IN NSCLC

EGFR mutation:

- 60% mutated
- Positive prognostic factor
- Positive predictive factor for gefitinib benefit (**qualitative interaction**, $p < 0.001$)

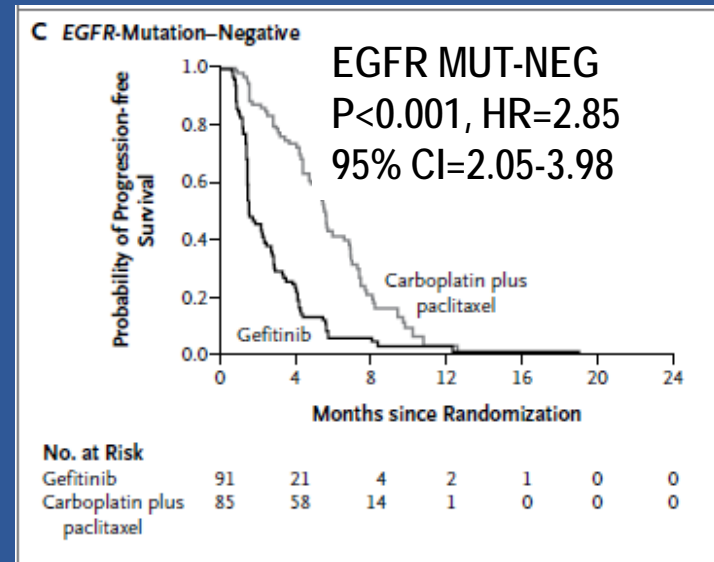
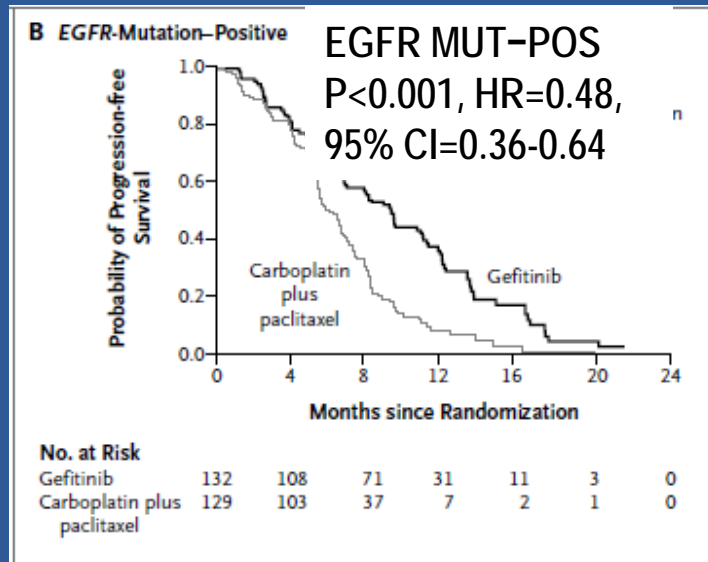


(Mok et al, *N Engl J Med* 2009;361:947-57)

IPASS: Phase III
 1st line advanced
 adeno NSCLC

gefitinib
 VS.
 carboplatin+paclitaxel

Cessation of chemo?



IPASS TRIAL: EVALUATION OF EGFR MUTATION AS A PREDICTIVE MARKER (OS)

Gefitinib Versus Chemo in NSCLC: Biomarker and Survival Analyses

Marker Availability

IHC	30%
FISH	33%
MUT	36%

Marker values lacking for many cases

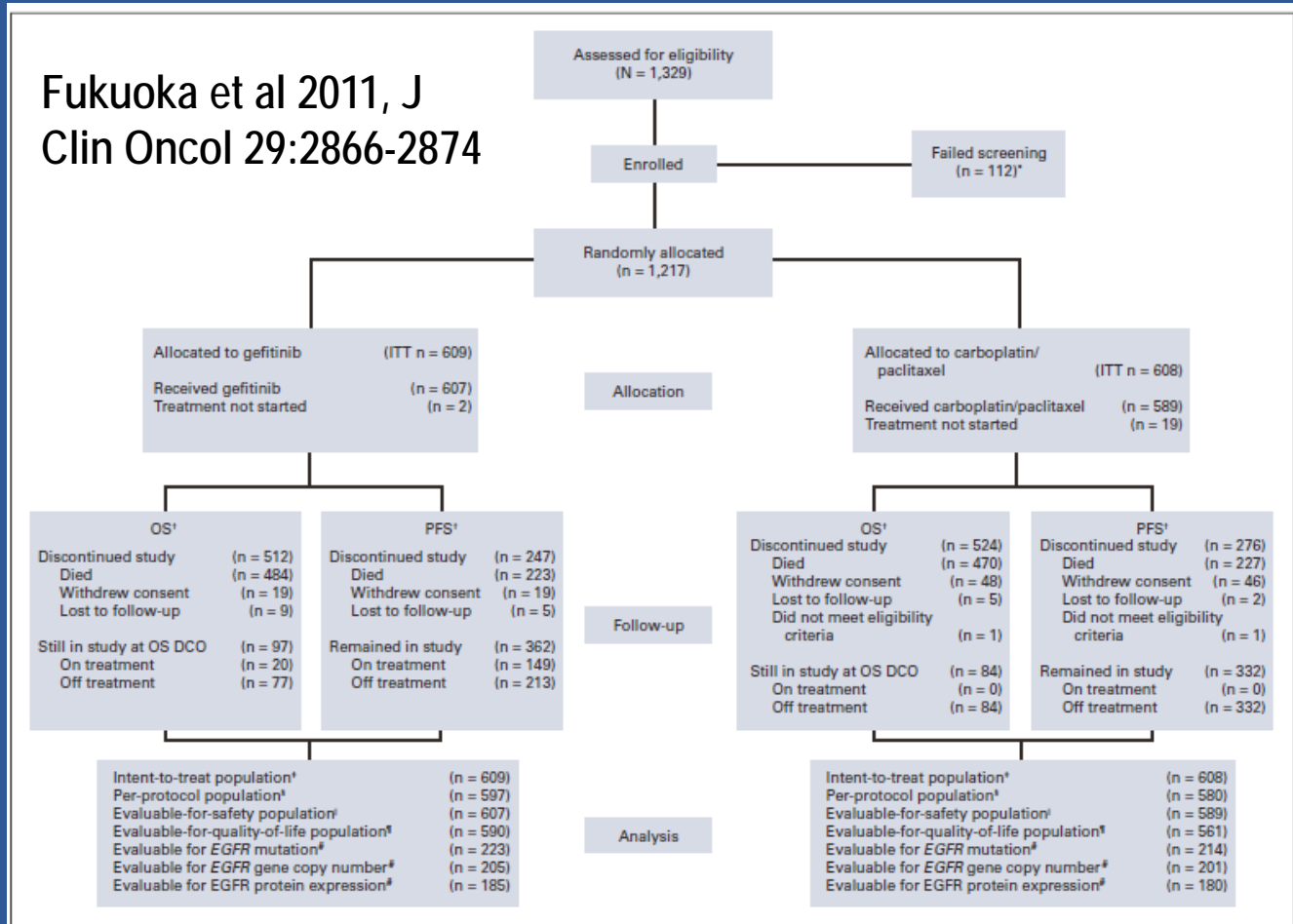


Fig 1. CONSORT diagram. (*) Among the 112 patients who failed screening, the main reasons for exclusion were abnormal serum creatinine ($> 1.5 \times$ upper limit of reference range)/creatinine clearance (≤ 60 mL/min) levels; untreated CNS metastases; or low neutrophil ($< 2.0 \times 10^9/L$), platelet ($< 100 \times 10^9/L$), 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.

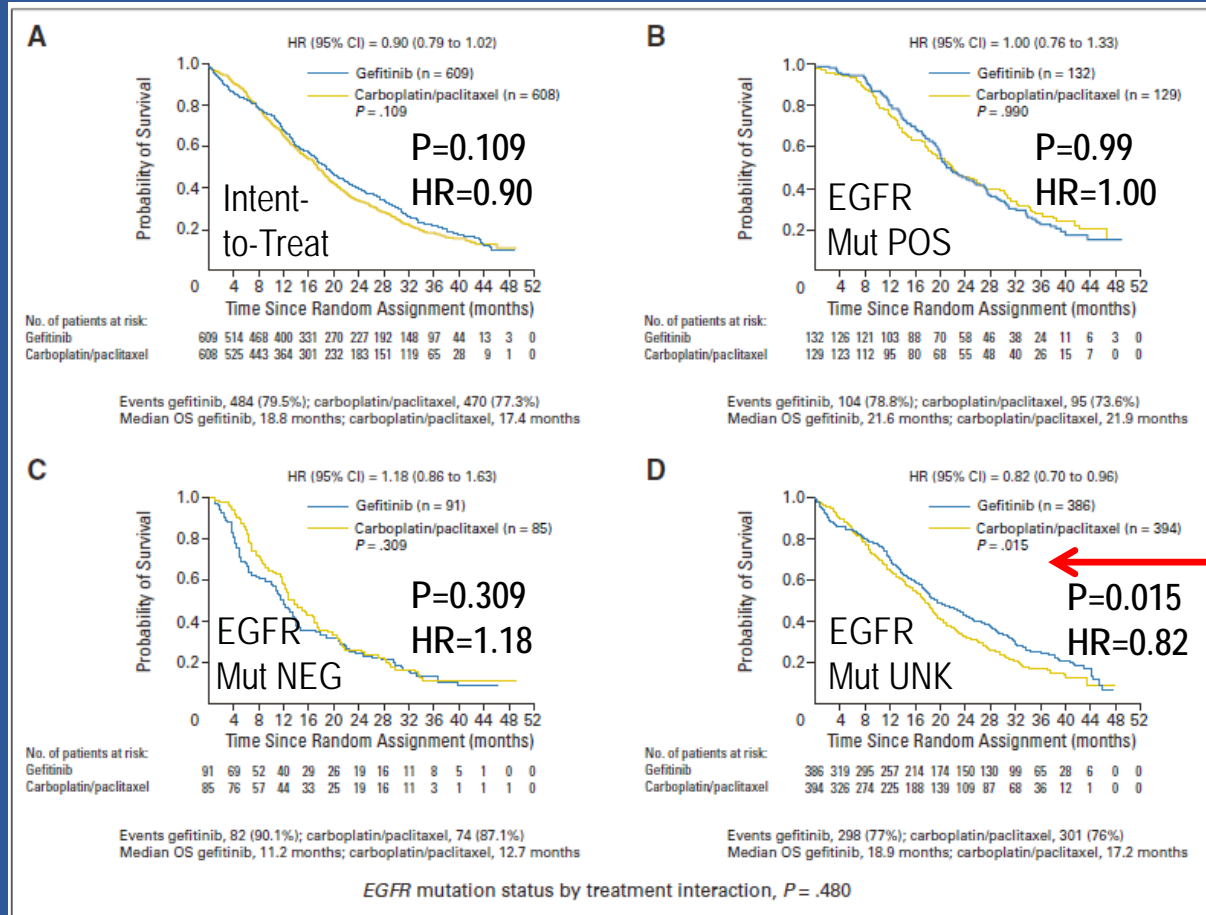
IPASS TRIAL: EVALUATION OF EGFR MUTATION AS A PREDICTIVE MARKER (OS)

Gefitinib Versus Chemo in NSCLC: Biomarker and Survival Analyses

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

Marker Positivity

IHC 73%
FISH 61%
MUT 60%



Only stat. significant benefit was 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

RANDOMIZED PHASE III BIOMARKER-DRIVEN TRIAL DESIGNS WITH TIME-TO-EVENT ENDPOINT

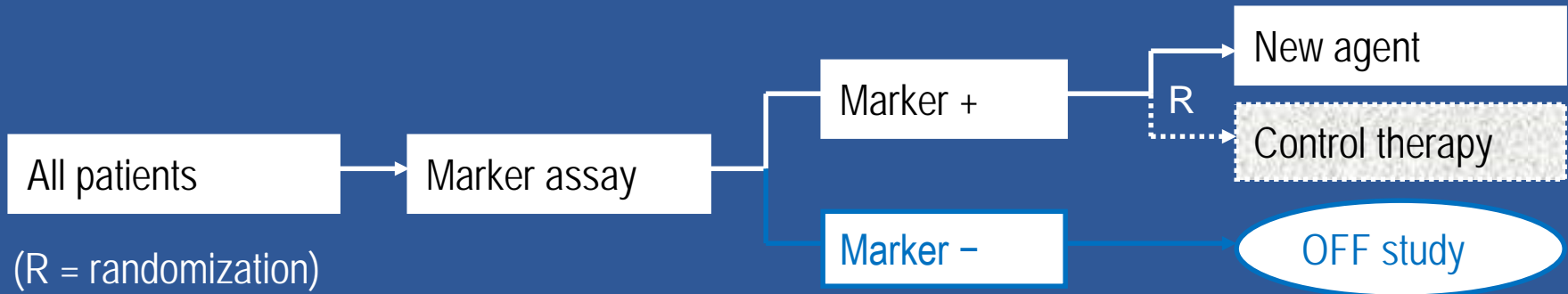
- **Basic designs**
 - Biomarker-Enrichment
 - Biomarker-Strategy
 - Biomarker-Stratified
- **Typical clinical endpoints (depends on context)**
 - Overall survival (OS)
 - Disease-free survival (DFS)
 - Relapse-free survival (RFS)

Note: Assume for purposes of this part of the discussion that the biomarker is binary, assay is analytically validated, and there are 2 treatment arms.

Sargent D et al. *J Clin Oncol* 2005;23:2020-2027

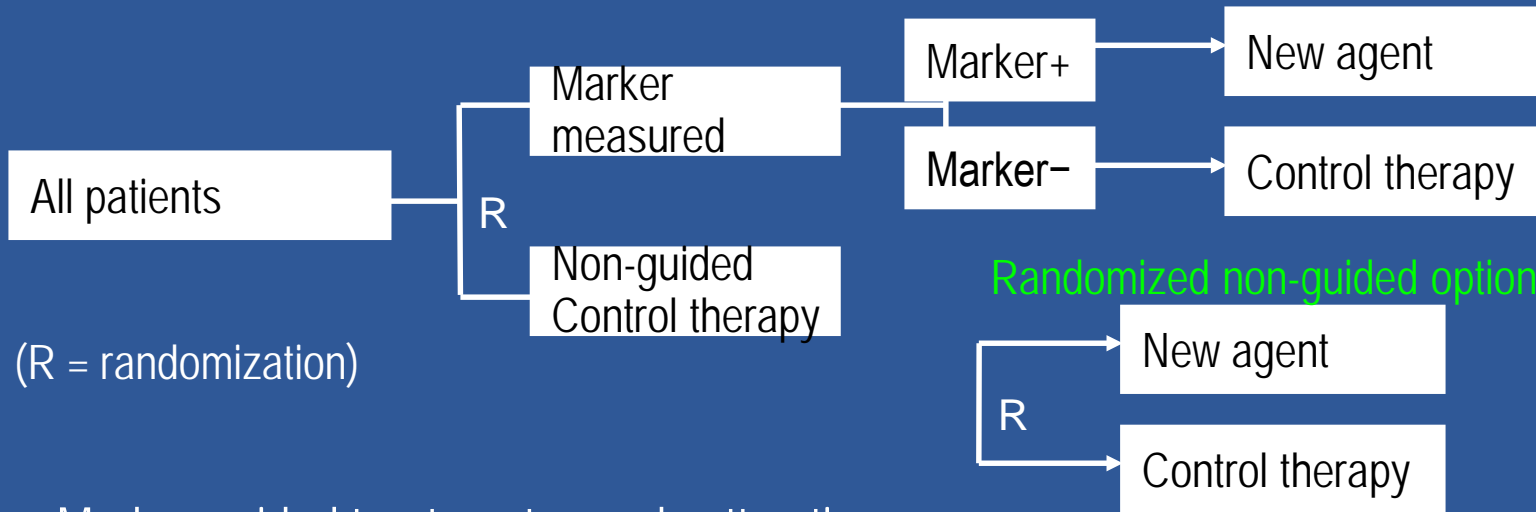
Freidlin B et al. *J Natl Cancer Inst* 2010;102:152-160

BIOMARKER-ENRICHMENT DESIGN



- Based in knowledge of biology (New agent → Molecular target)
- Control therapy arm controls for marker prognostic effect
- Variation: Standard therapy \pm new agent
- Limitations:
 - Off-target effects of new agent not fully evaluated
 - Regulatory indication limited to marker+ group
 - Marker refinement within trial (form of marker or assay) limited to marker+ group

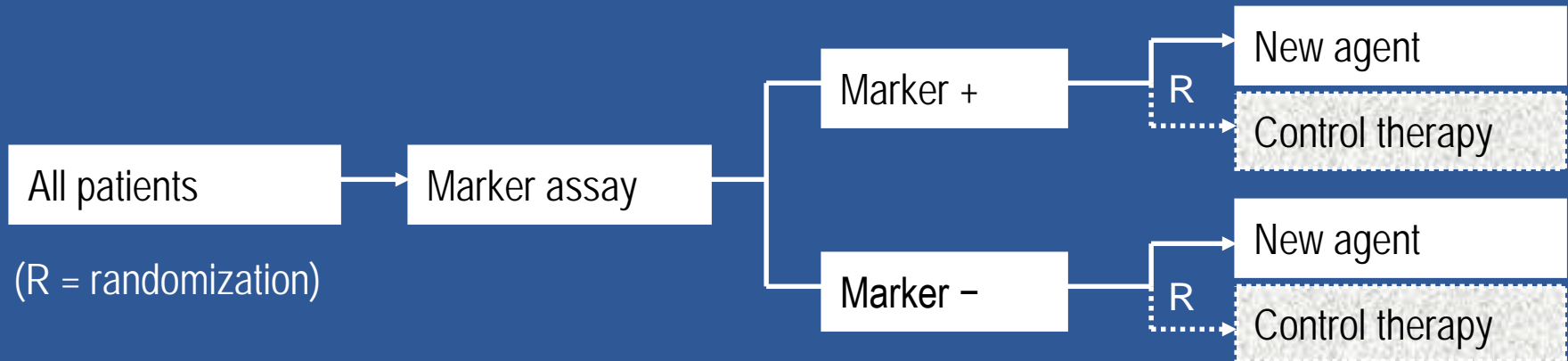
BIOMARKER-STRATEGY DESIGN



(R = randomization)

- Marker-guided treatment sounds attractive
- Might be only realistic option for complex multi-marker guided strategies, but can't separate biomarker and drug effects
- Must measure marker in non-guided control arm to distinguish prognostic effect
- Non-guided randomization allows assessment of new agent effect in marker-
- Statistical inefficiency
 - Marker- patients receive same therapy on both arms in standard strategy design
 - If randomize non-guided group, **even more inefficient**

BIOMARKER-STRATIFIED DESIGN



- Allows maximum information
 - Controls for prognostic effect of marker
 - Directly compares new agent to control therapy in *all* patients
- Allows retrospective evaluation of different markers or assays
- Variation: Standard therapy \pm new agent
- Completely randomized design with *retrospective* marker evaluation is an option, but assay results might not be available for 100% of patients
- Different approaches to testing in biomarker subgroups (Freidlin & Korn, *Nat Rev Clin Oncol* 2014;11: 81–90)

NEW ONCOLOGY TRIAL DESIGNS (PHASE II & III)

- Basket/bucket trials – variety of cancer types; Single drug targeting a single mutation
 - Umbrella trials – multiple biomarker-based cohorts, each matched to a drug; single or multiple histology/cancer types (NCI-MATCH, BATTLE trials, Lung-MAP, ALCHEMIST)
 - Platform trials - standing trial structure, multiple agents enter and exit, single cancer type, possibly biomarker-driven (I-SPY2 trial, FOCUS trials)
 - Combinations of the above
-
- Abrams et al., ASCO Educ Book 2014, pp. 71-76 (NCI-MATCH, Lung-MAP, ALCHEMIST)
 - Barker et al., Clin Pharm & Ther 2009;86:97-100 (I-SPY2)
 - Kaplan et al., J Clin Oncol 2013;31:4562-4568 (FOCUS)
 - Kim et al., Cancer Discovery 2011;1:44-53 (BATTLE)
 - Kummar et al., J Natl Cancer Inst 2015;107(4):dju003 (review of molecular profiling trials)

CAN A cfDNA ASSAY REPLACE A STANDARD TUMOR TISSUE-BASED PREDICTIVE ASSAY?

	cfDNA NEG (D-)	cfDNA POS (D+)
TUMOR NEG (T-)	RR(T-,D-); p(T-,D-)	RR(T-,D+); p(T-,D+)
TUMOR POS (T+)	RR(T+,D-); p(T+,D-)	RR(T+,D+); p(T+,D+)

RR(x,y) = response rate and p(x,y) = prevalence of subgroup (x,y)

Historically, we know $p(T+,.) = p(T+,D-) + p(T+,D+)$ and

$$RR(T+,.) = \frac{p(T+,D-)\times RR(T+,D-)}{p(T+,D-) + p(T+,D+)} + \frac{p(T+,D+)\times RR(T+,D+)}{p(T+,D-) + p(T+,D+)}$$

To justify replacement, ideally we would like to establish that

$$p(.,D+) \geq p(T+,.) \text{ and } RR(.,D+) \geq RR(T+,.)$$

In reality there could be a tradeoff between $p(.,D+)$ and $RR(.,D+)$, i.e., we might dilute $RR(.,D+)$ if $RR(T-,D+)$ is not high

DOES A cfDNA ASSAY HAVE UTILITY WHEN USED IN SEQUENCE WITH A STANDARD TUMOR TISSUE-BASED PREDICTIVE ASSAY?

	cfDNA NEG (D-)	cfDNA POS (D+)
TUMOR NEG (T-)	RR(T-,D-); p(T-,D-)	RR(T-,D+); p(T-,D+)
TUMOR POS (T+)	RR(T+,D-); p(T+,D-)	RR(T+,D+); p(T+,D+)

RR(x,y) = response rate and p(x,y) = prevalence of subgroup (x,y)

Sequential testing

1. Test cfDNA; treat with targeted agent if positive.
2. If cfDNA is negative, test tumor and treat with targeted agent if tumor is positive

Overall response rate with treatment according to this sequential scheme is

$RR(seq) =$

$$\frac{p(T+, D+) \times RR(T+, D+) + p(T-, D+) \times RR(T-, D+) + p(T+, D-) \times RR(T+, D-)}{1 - p(T-, D-)}$$

MONITORING BIOMARKER

FDA-NIH: A biomarker measured serially for assessing status of a disease or medical condition or for evidence of exposure to (or effect of) an environmental agent or medical product

Considerations for clinical utility evaluation:

- Can the biomarker be measured less invasively, less expensively, or more conveniently than currently available clinical indicators?
- Can the biomarker detect the change in disease or toxicity status prior to other standard clinical indicators?
- Are there clinical actions that can be taken on the basis of the biomarker results?
- Does taking clinical action lead to benefit to the patient?

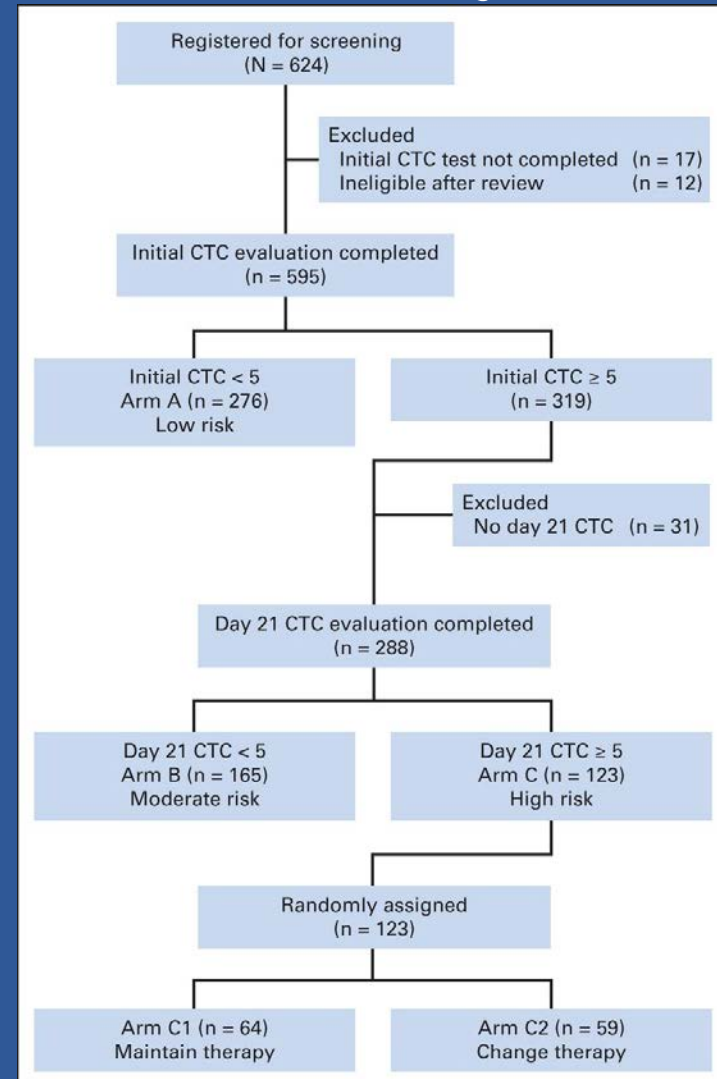
MONITORING BIOMARKERS

CLINICAL VALIDITY VERSUS CLINICAL UTILITY

Example: S0500 clinical trial
in metastatic breast cancer

- Measure circulating tumor cells (CTCs) in whole blood (WB) at baseline and after 21 days of chemotherapy
- Does switching to an alternative chemotherapy improve outcome for those patients who have persistently elevated CTCs (≥ 5 per 7.5 mL WB) after 21 days of therapy?

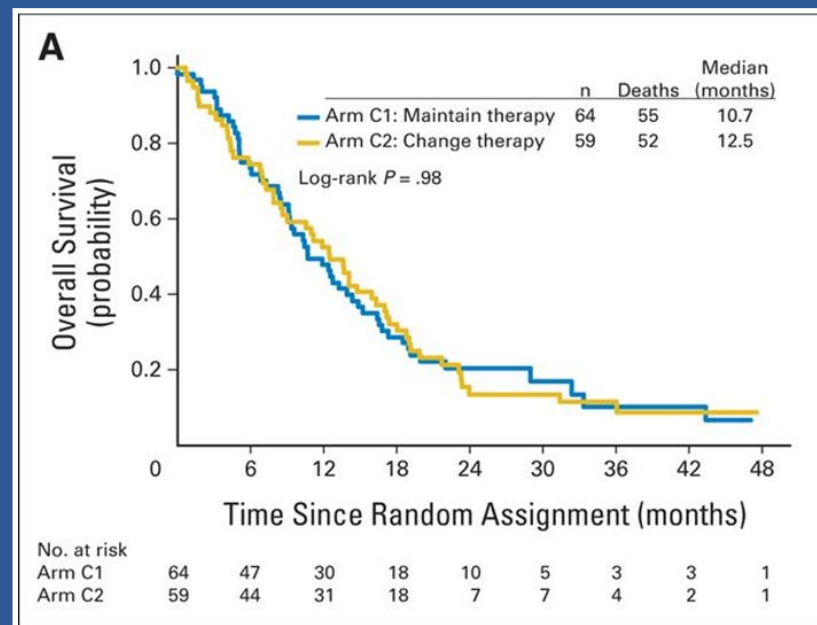
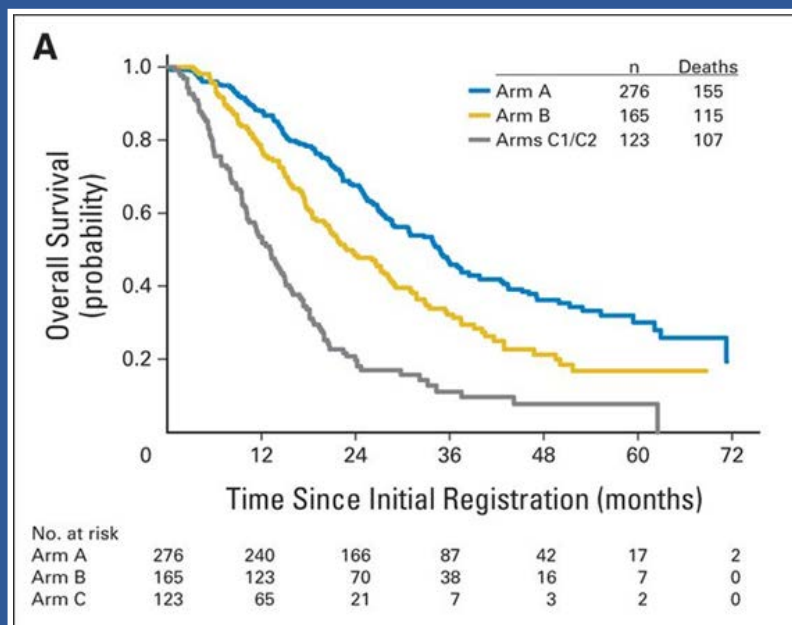
CONSORT Diagram



MONITORING BIOMARKERS

Example (cont.): S0500 clinical trial in metastatic breast cancer

- CTCs confirmed to be prognostic throughout the course of therapy
- Switching to an alternative chemotherapy for those patients who have persistently elevated CTCs (≥ 5 per 7.5 mL WB) after 21 days of therapy **DID NOT IMPROVE** outcome (OS or PFS)



MONITORING BIOMARKERS

Example (cont.): S0500 clinical trial in metastatic breast cancer

Possible reasons why might the study have failed to generate a positive result

- The treatments available were not effective
- CTCs were not measured at the right time or quantified in the right way
- CTCs were not molecularly characterized

For additional helpful discussion of design issues for tumor biomarker monitoring trials see:
Soletormos et al. *Clinical Chemistry* 2013;59(1):52-59

SUMMARY

- Identifying the clinical setting and defining the clinical question is half of the battle
- Clinical utility of a biomarker test will usually depend on the availability of good treatments other medical interventions
- Analytical validation, clinical validation, and clinical utility must all be considered in the translational process
- Careful planning will be required to acquire and make best use of available specimens

THANK YOU!

