



**Gynecologic Cancer InterGroup**  
**Translational Research Brainstorming**  
**October 2016**                      **Lisbon, Portugal**

# **Considerations in the design of clinical trials to validate predictive biomarkers**

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

# Disclosures

I have no financial relationships to disclose.

- and -

I will not discuss off label use and/or investigational use in my presentation.

- and -

The views expressed represent my own and do not necessarily represent the views or policies of the U.S. National Cancer Institute.

# Enrichment in drug development

- **Enrichment:** Prospective use of any patient characteristic to select a study population in which detection of a drug effect (if one is in fact present) is more likely than it would be in an unselected population<sup>1</sup>
  - Strategies to decrease heterogeneity – reduce inter-patient and intra-patient heterogeneity
  - Prognostic enrichment strategies – choosing patients with a greater likelihood of having a disease-related endpoint event
  - **Predictive enrichment strategies – choosing patients more likely to respond to the drug treatment (i.e., treatment selection using a biomarker)**
    - If successful, may lead to companion diagnostic

<sup>1</sup><http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf>

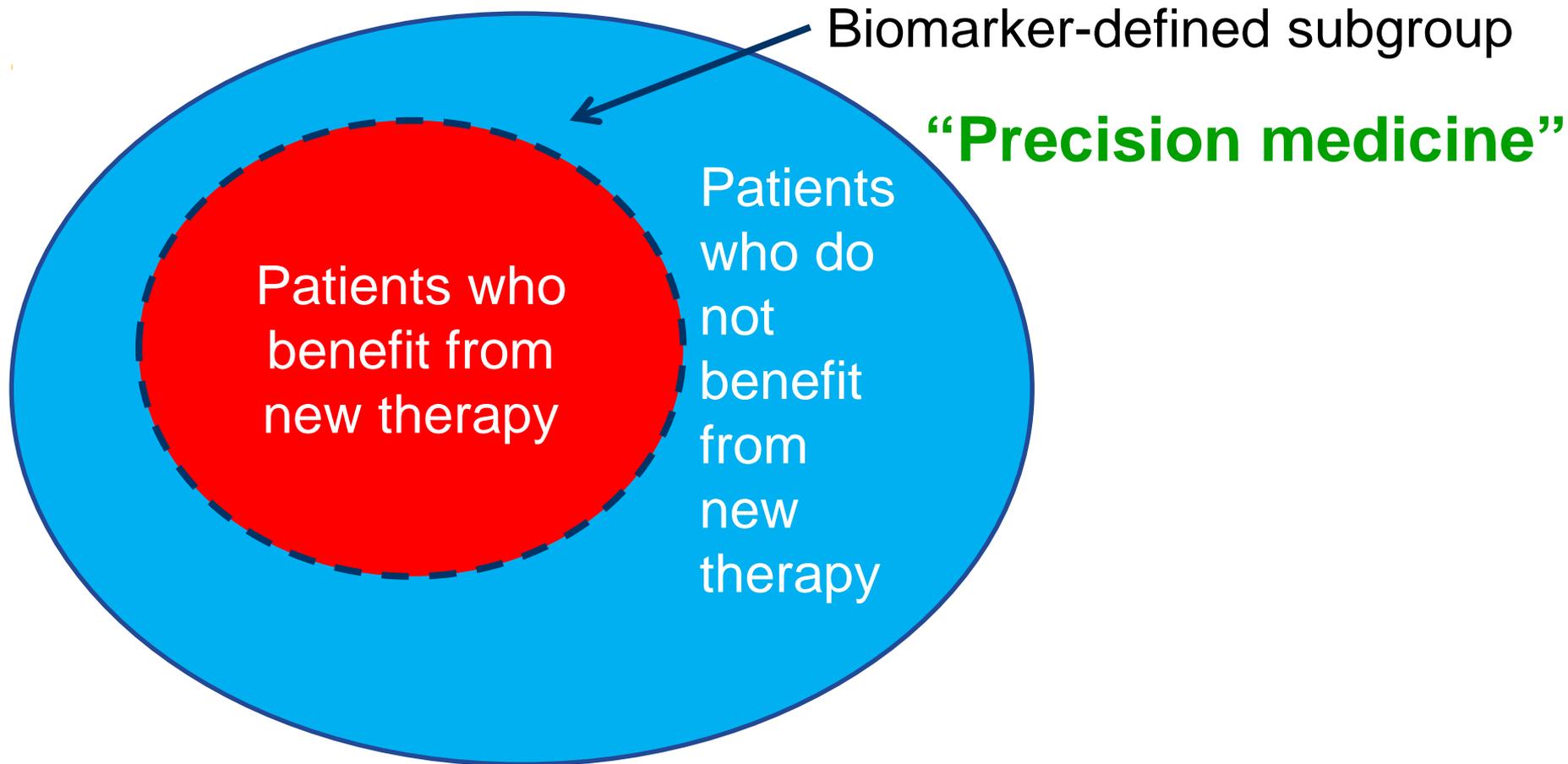
# Predictive biomarker definition

- A biomarker associated with benefit or lack of benefit (potentially even harm) from a particular therapy relative to other available therapy.
- **FDA-NIH “BEST” glossary definition:** 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.<sup>1</sup>

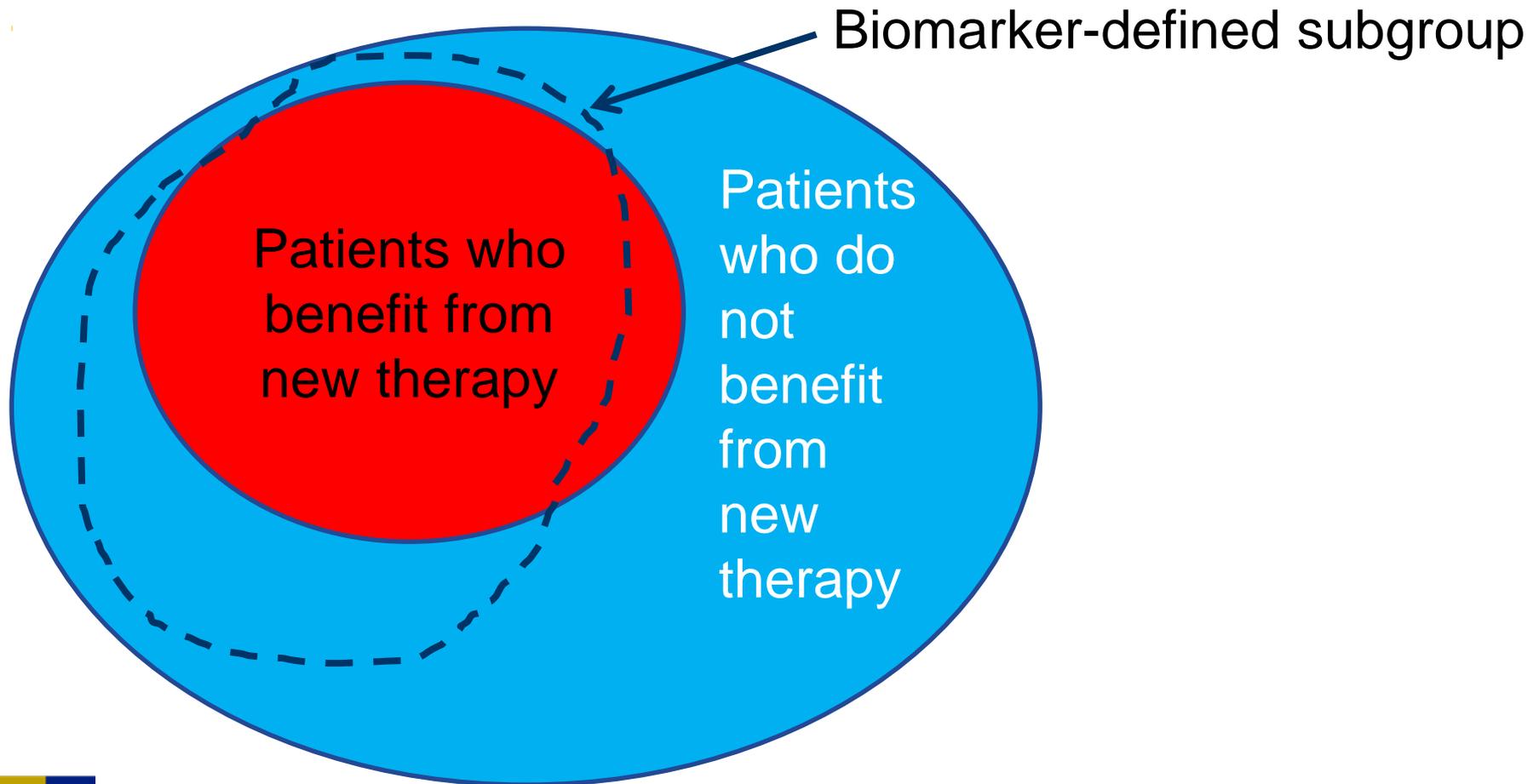
<sup>1</sup>“BEST” Resource glossary:

<http://www.ncbi.nlm.nih.gov/books/NBK326791/>

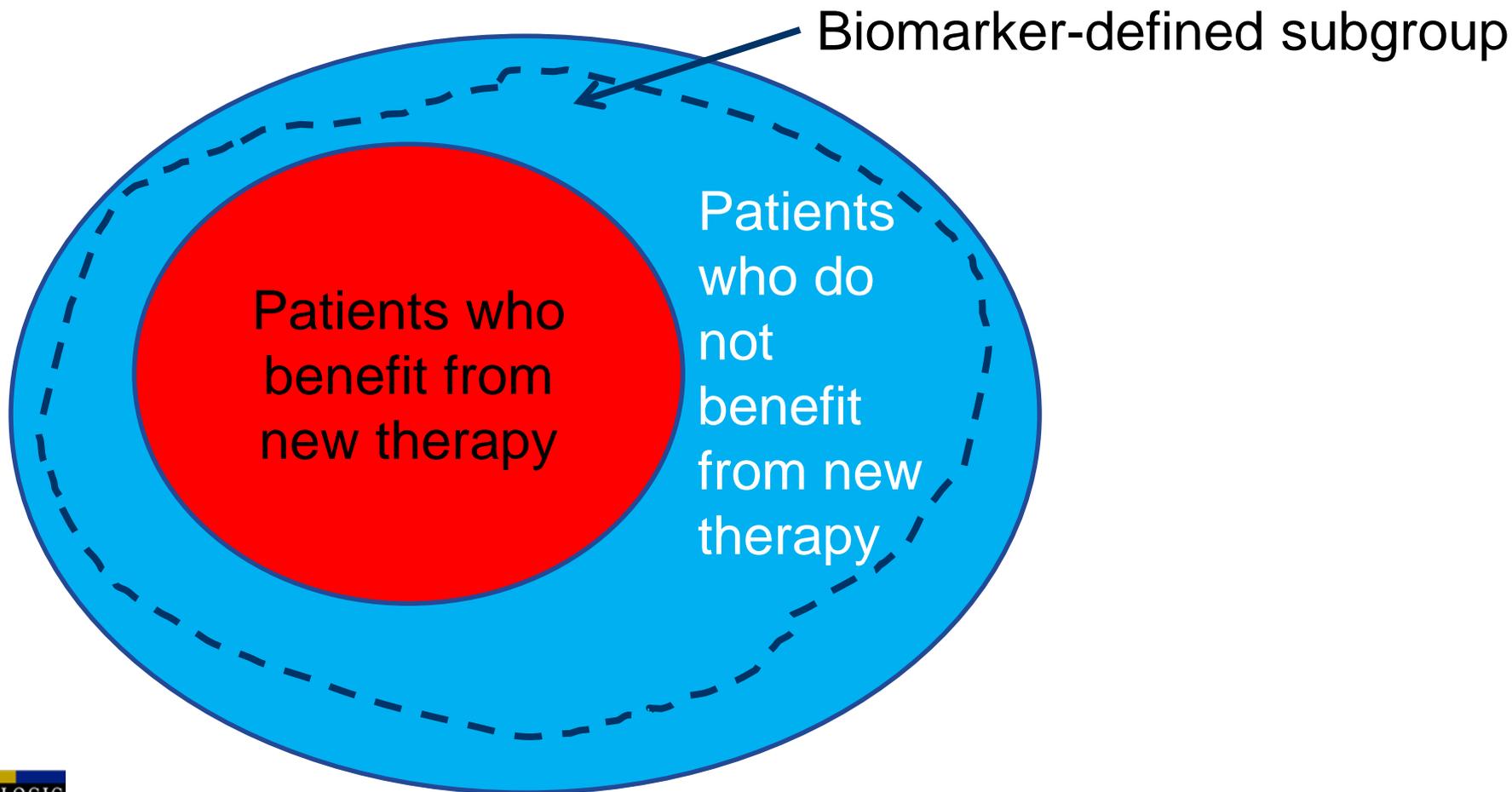
# “Ideal” biomarker for trial enrichment and companion diagnostic (predictive biomarker) development



# Biomarker **useful** for trial enrichment, and likely for companion diagnostic (predictive biomarker) development

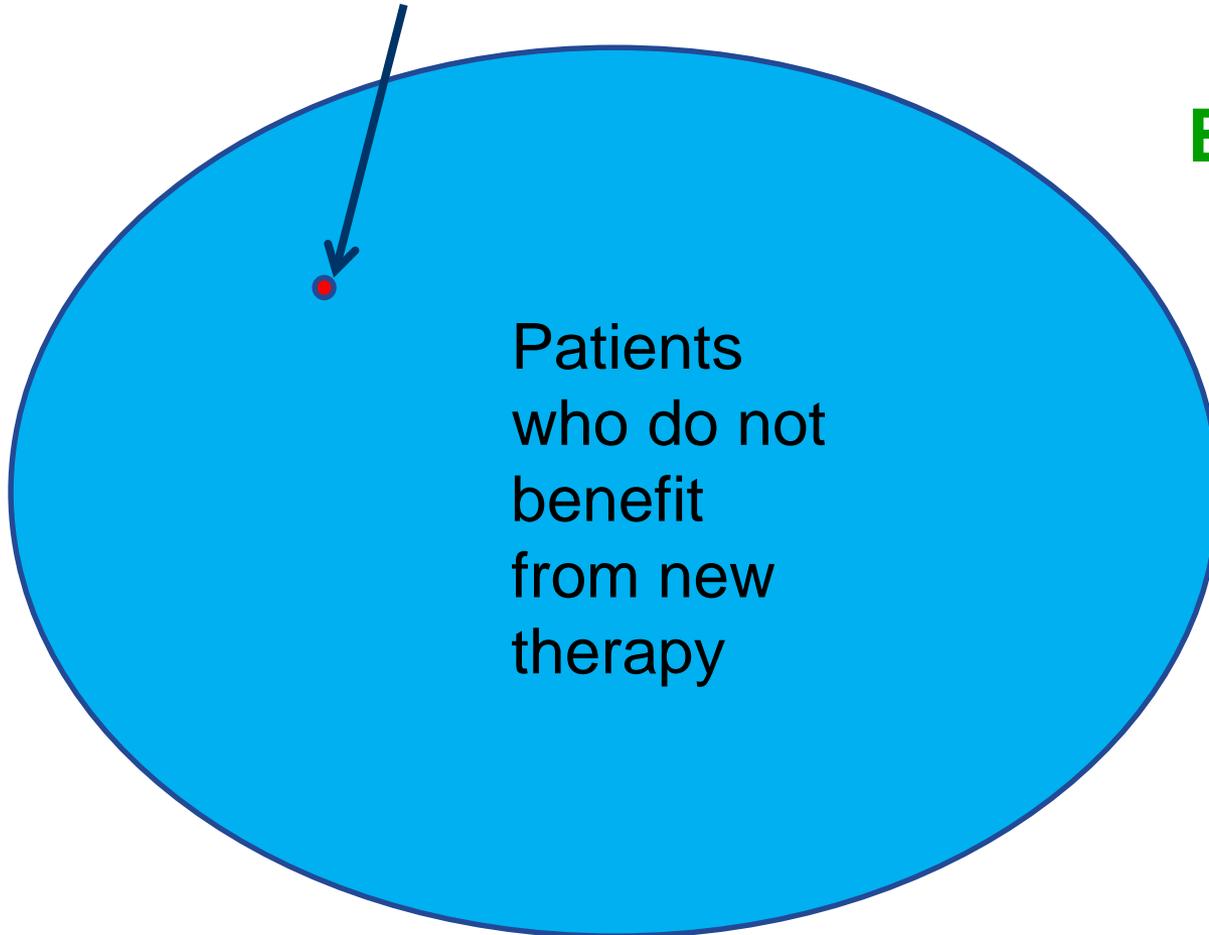


# Biomarker **not cost-effective** to use for trial enrichment or companion diagnostic (predictive biomarker) development



# No drug effect for a biomarker to find

Patients who benefit from new therapy



**Biomarker?**

# A series of questions to answer

- Q1: Does the drug work in any patients?
- Q2: If the drug does not work in all patients, is there a subset in which it does work?
- Q3: If the drug works in only a subset, is there a biomarker that defines that subset?
- Q4: If a biomarker is needed, what is the best way to measure it?

# Tension between assay development and therapeutic development

- **Assay analytical performance** – minimum requirements in early trials
  - Sufficient reproducibility so that study could be repeated
  - Fit for use on anticipated specimen types (specimen format, processing & handling)
- **First priority** is usually to establish that the new agent has promising **activity**
  - Biomarker has to be “good enough” to capture a sufficient portion of the patients who will benefit in order to see signal
  - Later biomarker refinement often needed

# Predictive biomarker development & evaluation

- Must have **biomarker and assay to measure it**
  - *Predictive ability transfers* from pre-clinical models to human
  - Assay requirements: *acceptable reproducibility* and *fit for use on clinical specimens* (may have limited availability)
  - Flexibility for assay evolution, but eventually need *locked assay* with established *analytical performance*
- Typically proceed through **phase II and III trials**
  - Trial design choices depend on *biomarker credentials* and *question(s)* one wishes to answer at each stage
  - *First priority* usually to establish *promising activity* of new agent
  - Biomarker has to be “*good enough*” to capture a sufficient patients who will benefit in order *to see signal of activity*
  - *Sometimes retrospective studies* using banked trial specimens are possible
- **Failure may be due to drug and or biomarker/assay**

# Non-randomized biomarker-guided phase II studies

Can we detect “signal” of activity at least in subgroup defined by “best guess” biomarker?

- **Biomarker enrichment**

- Biomarker positivity required for patient eligibility
- Biomarker-driven is appealing, aids accrual

- **Biomarker stratification**

- Consider results combined and separately within biomarker positive and negative subgroups
- May include biomarker-based adaptive features

McShane L et al., *Clin Cancer Res* 2009;15:1898-1905

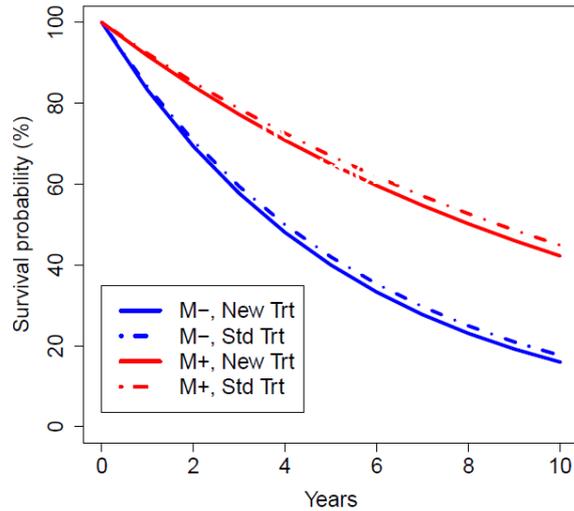
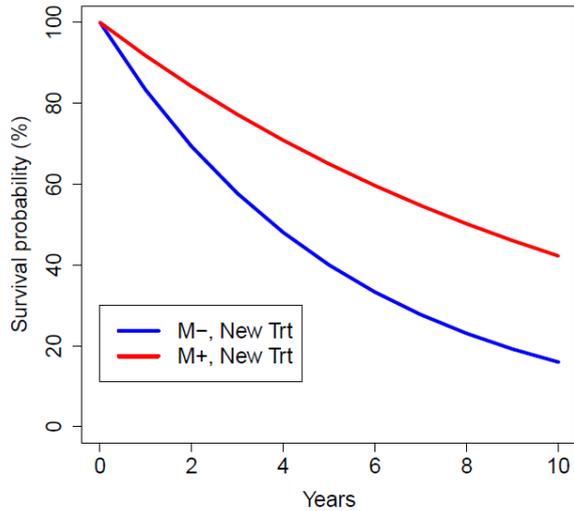
McShane L & Hunsberger S, An overview of phase II clinical trial designs with biomarkers. In *Design and Analysis of Clinical Trials for Predictive Medicine*, Matsui, Buyse, Simon (eds.), Chapman and Hall/CRC, 2015.

Freidlin B et al., *J Clin Oncol* 2012;30:3304-3309

# Reasons to conduct randomized trials (phase II and III designs)

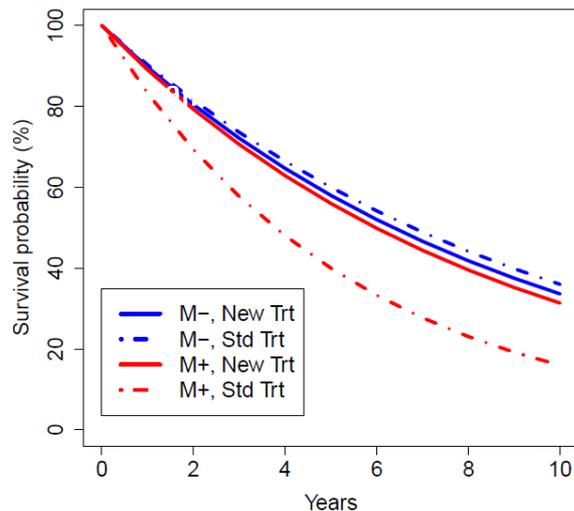
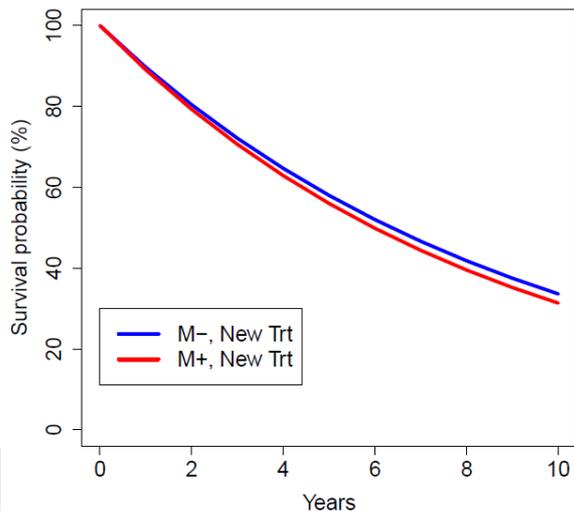
- Desired endpoint is a time-to-event endpoint and prognostic effect of biomarker cannot be ruled out
- If agent not expected to deliver robust tumor shrinkage (e.g., cytostatic), what are the appropriate benchmarks for endpoints such as PFS or SD within biomarker-defined subgroups if no randomization?
- Other effective therapies available
- New (biomarker-directed) agent will be tested *in combination* with a standard therapy (standard therapy  $\pm$  new agent)?

# Prognostic vs. predictive: Importance of control groups



Prognostic  
but not  
predictive

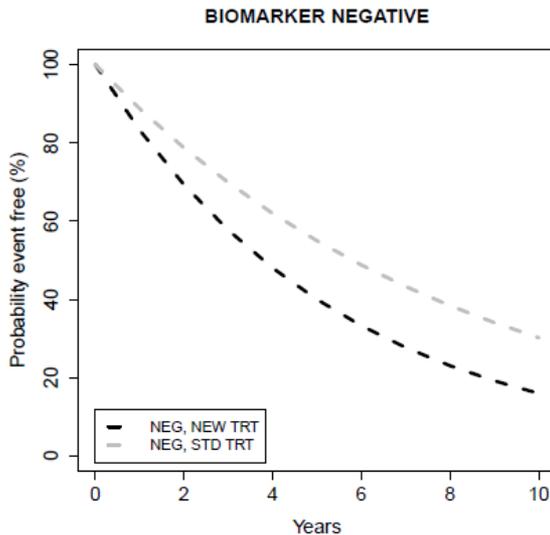
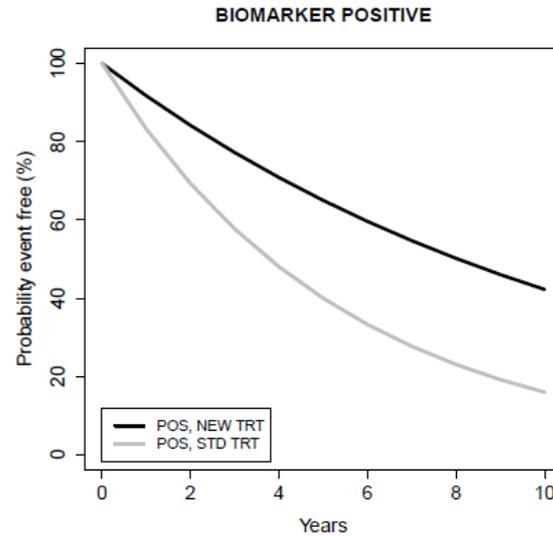
(M =  
biomarker)



Prognostic  
and  
predictive

# CLINICALLY USEFUL predictive biomarker

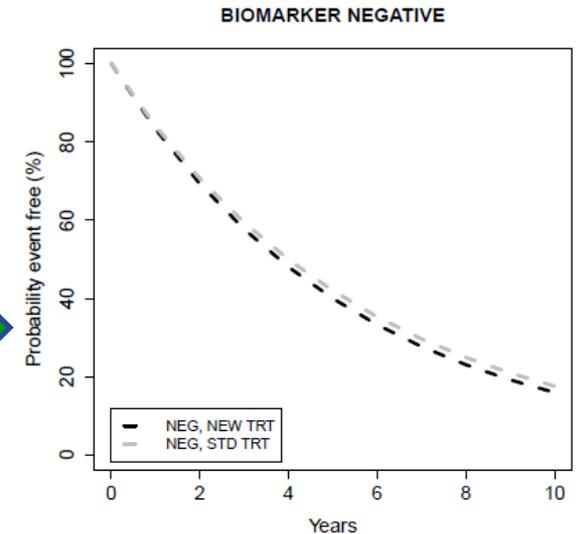
**Qualitative interaction:**  
Patients “positive” for the biomarker benefit from the treatment but others receive no benefit or possibly even harm



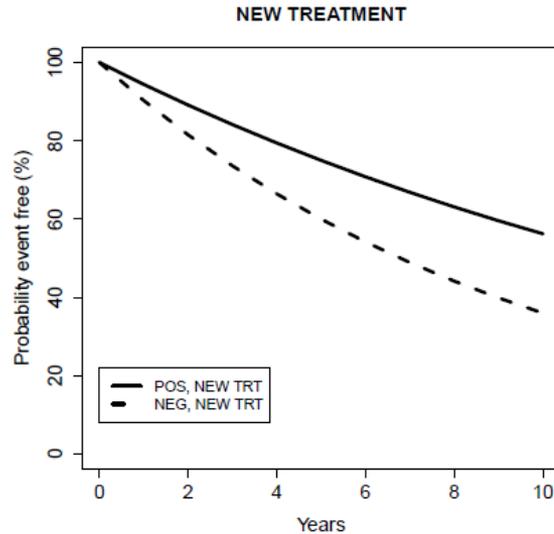
↑

BIOMARKER POS:  
NEW TRT > STD TRT

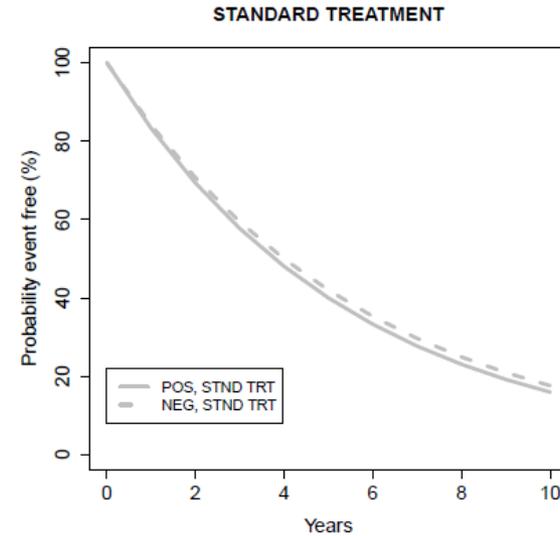
← BIOMARKER NEG:  
NEW TRT ≤ STD TRT →



# How *NOT* to parse evidence for a candidate predictive biomarker



NEW TREATMENT:  
BIOMARKER POS > BIOMARKER NEG



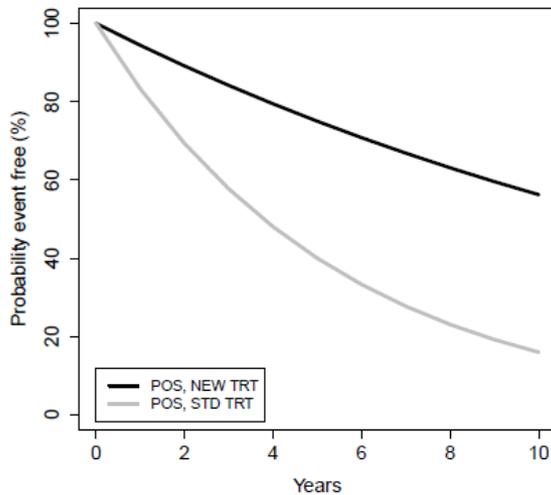
STANDARD TREATMENT:  
BIOMARKER POS = BIOMARKER NEG  
(NOT PROGNOSTIC)

# How to *CORRECTLY* parse evidence for a candidate predictive biomarker

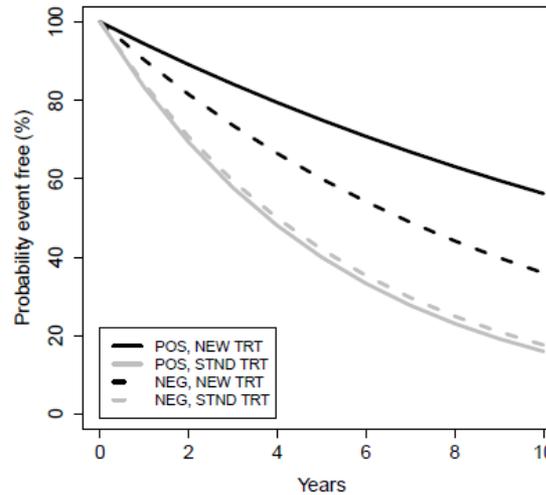
BIOMARKER POS:  
NEW TRT > STD TRT



BIOMARKER POSITIVE



ALL COMBINATIONS OF  
BIOMARKER AND TREATMENT

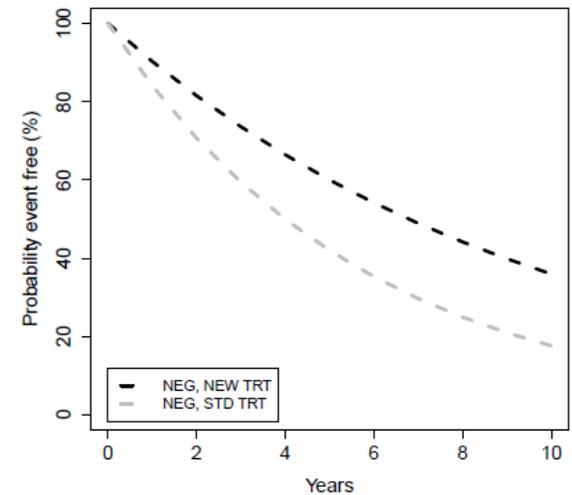


Now we see that the biomarker is not useful for selection of new treatment (because both patient subgroups benefit).

BIOMARKER NEG:  
NEW TRT > STD TRT



BIOMARKER NEGATIVE



**Quantitative interaction:** Treatment benefits all patients but by different amounts

# 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)

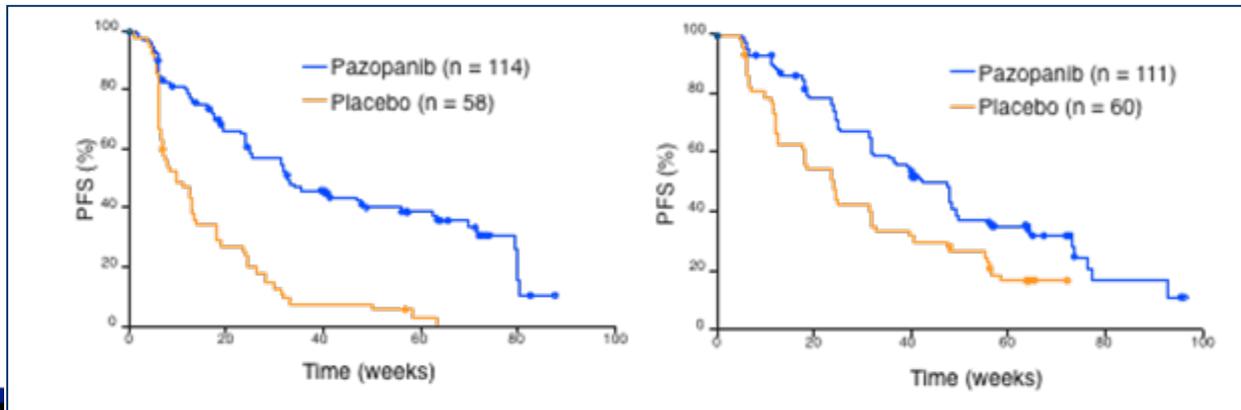
	PFS (weeks)		HR (95% CI)	p value		
	Pazopanib	Placebo		Pazopanib	Placebo	Interaction
<b>Interleukin 6</b>						
Low	42.3	24.0	0.55 (0.38-0.81)	0.445	<0.0001	0.009
High	32.6	9.9	0.31 (0.21-0.44)			

Usefully 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?

# PD-L1 expression as a predictive biomarker in cancer immunotherapy

Therapeutic agent	Detection antibody; membrane staining cutoff (in percent of tumor cells)	Histology	PD-L1 IHC expression (% samples at IHC level)	Response rate for PD-L1-positive versus PD-L1-negative patients
Nivolumab	28-8; 5%	Melanoma ( <i>n</i> = 38)	+ (45%), - (55%)	44% vs. 17% ( <i>P</i> = NR)
	DAKO; 5%	NSCLC ( <i>n</i> = 20)	+ (60%), - (40%)	67% vs. 0% ( <i>P</i> = NR)
	5H1; 5%	Melanoma, RCC, NSCLC, CRC, prostate	+ (60%), - (40%)	36% vs. 0% ( <i>P</i> = NR)
Pembrolizumab	NR; 1%	Melanoma ( <i>n</i> = 71)	+ (77%), - (23%)	51% vs. 6% ( <i>P</i> = 0.0012)
	NR; 50%	NSCLC ( <i>n</i> = 38)	+ (25%), - (75%)	67% vs. 0% (6-month irORR; <i>P</i> < 0.001)
MPDL3280A	Roche/Genentech	NSCLC, RCC, melanoma, CRC, gastric cancer	NR	39% vs. 13% ( <i>P</i> = NR)
	Roche/Genentech	NSCLC ( <i>n</i> = 37)	+ (13%), - (87%)	100% vs. 15% ( <i>P</i> = NR)
	Roche/Genentech	Bladder ( <i>n</i> = 20)	NR	52% vs. 11% ( <i>P</i> = NR)

Abbreviations: CRC, colorectal cancer; NR, not reported

**The story becomes even more complex when looking at PFS and OS endpoints**

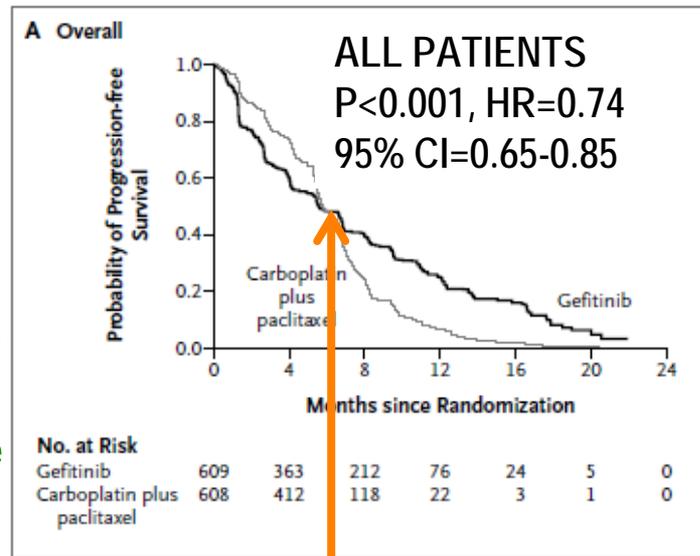
(Adapted from Table 3 in Patel & Kurzrock, *Mol Cancer Ther* 2015;14(4):847-856) <sup>19</sup>

# EGFR mutation predictive for PFS benefit with gefitinib in NSCLC (IPASS trial)

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

**IPASS:** Phase III  
1<sup>st</sup> line advanced  
adeno NSCLC

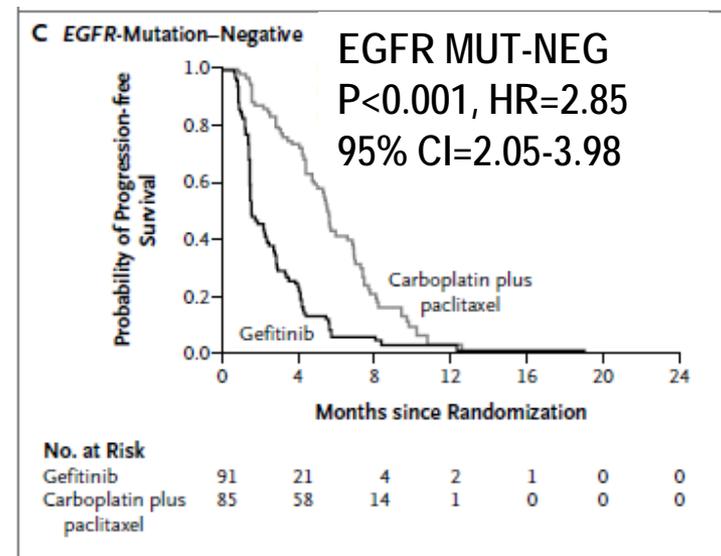
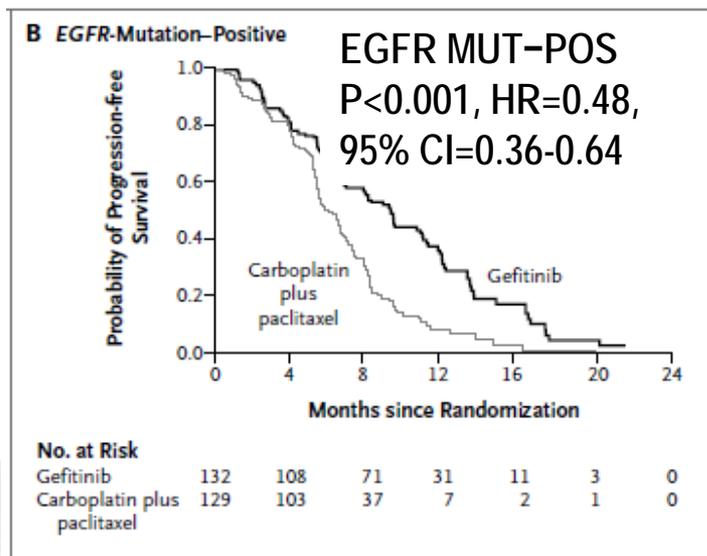
gefitinib  
vs.  
carboplatin+paclitaxel



Cessation of chemo?

## EGFR mutation:

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



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

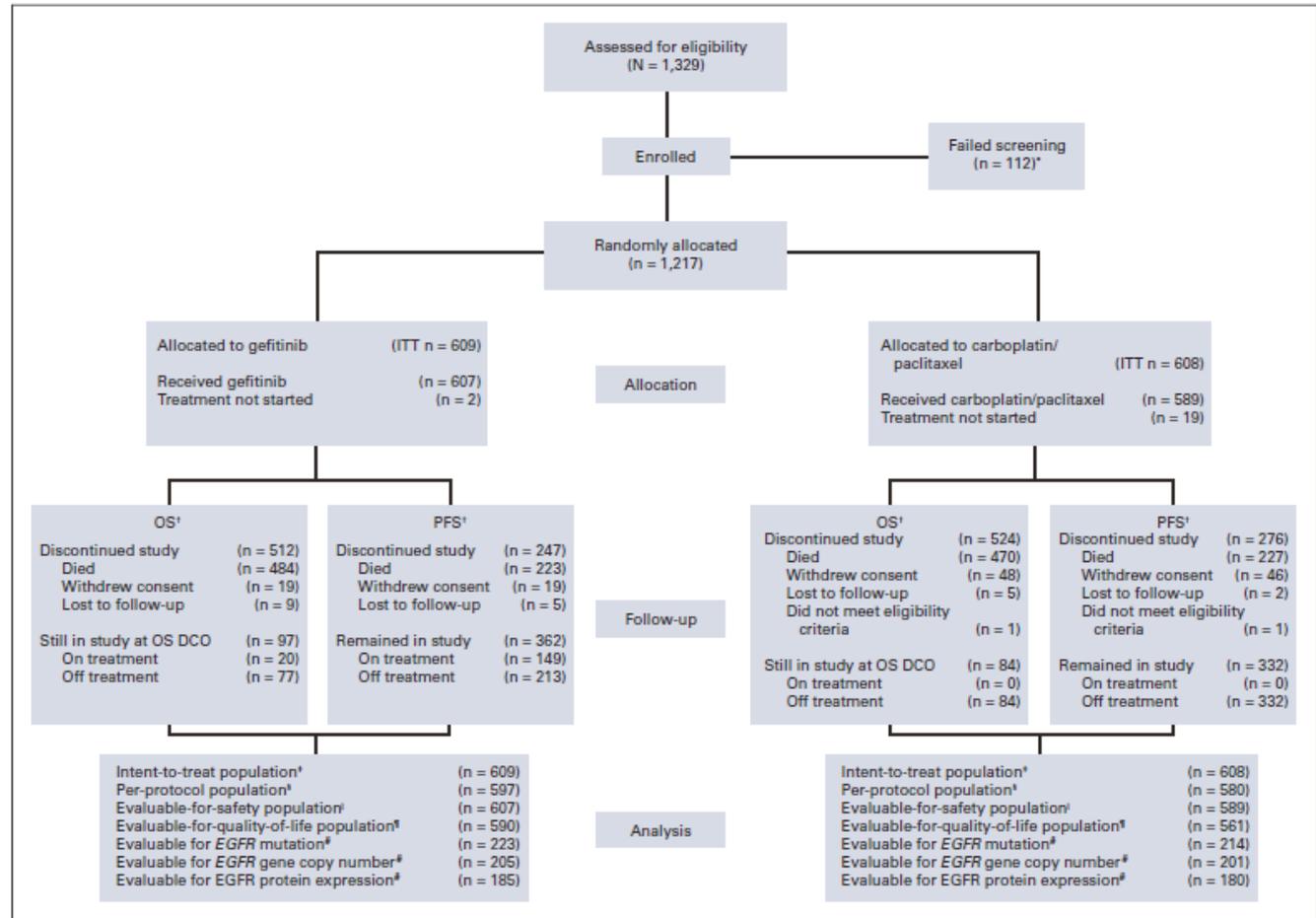
Gefitinib vs. Chemo in NSCLC: Biomarker and Survival Analyses

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

## Marker Availability

IHC 30%  
FISH 33%  
MUT 36%

Marker values  
lacking for  
over half of  
the 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 ( $\leq 60$  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)

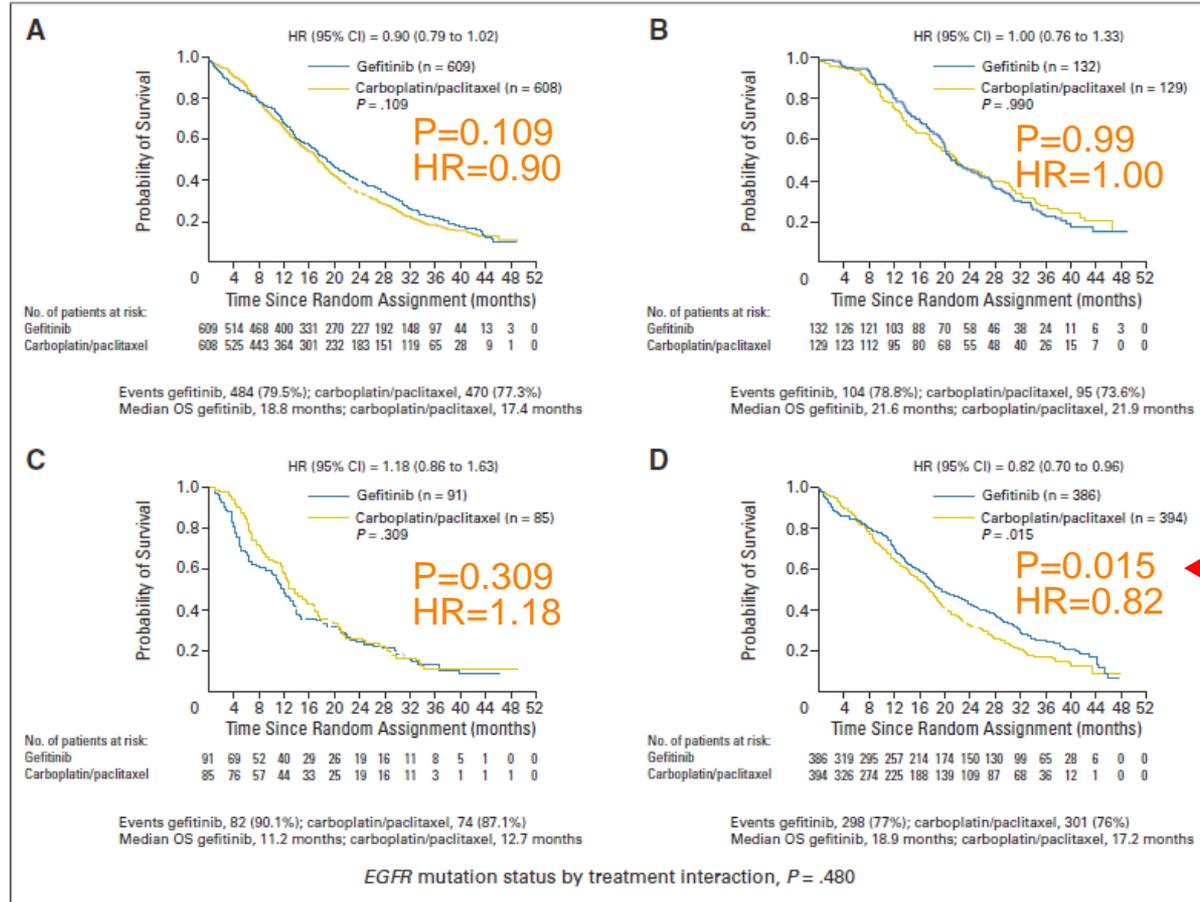
## Gefitinib Versus Chemo in NSCLC: Biomarker and Survival Analyses

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

### Marker Positivity\*

IHC 73%  
FISH 61%  
MUT 60%

\*These rates are high because these were patients in East Asia who were nonsmokers or former light smokers.



**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.

The only statistically significant benefit was in the subgroup with EGFR mutation status unknown.

High rates of crossover; other EGFR-inhibitors showed benefit in unselected patients in second line setting

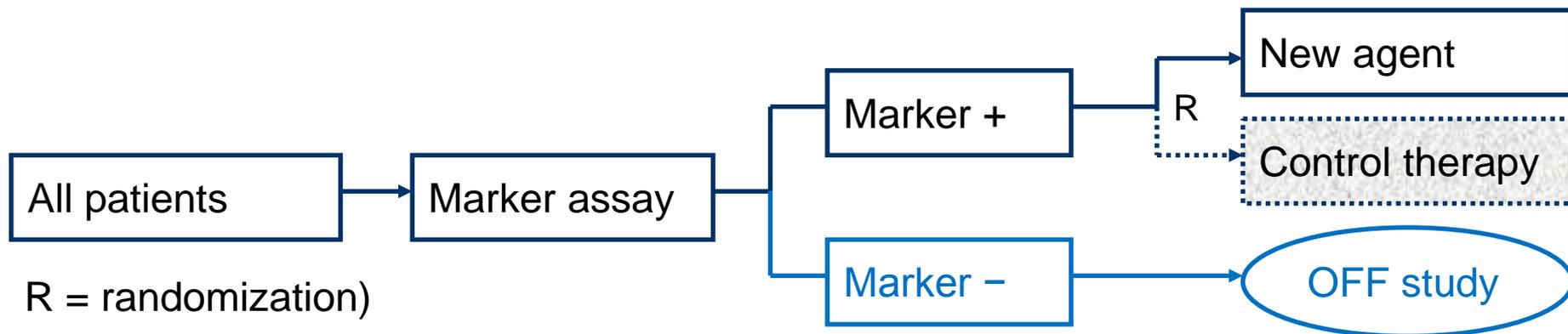
# Randomized phase III biomarker-driven trial designs with time-to-event endpoint

- **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)

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

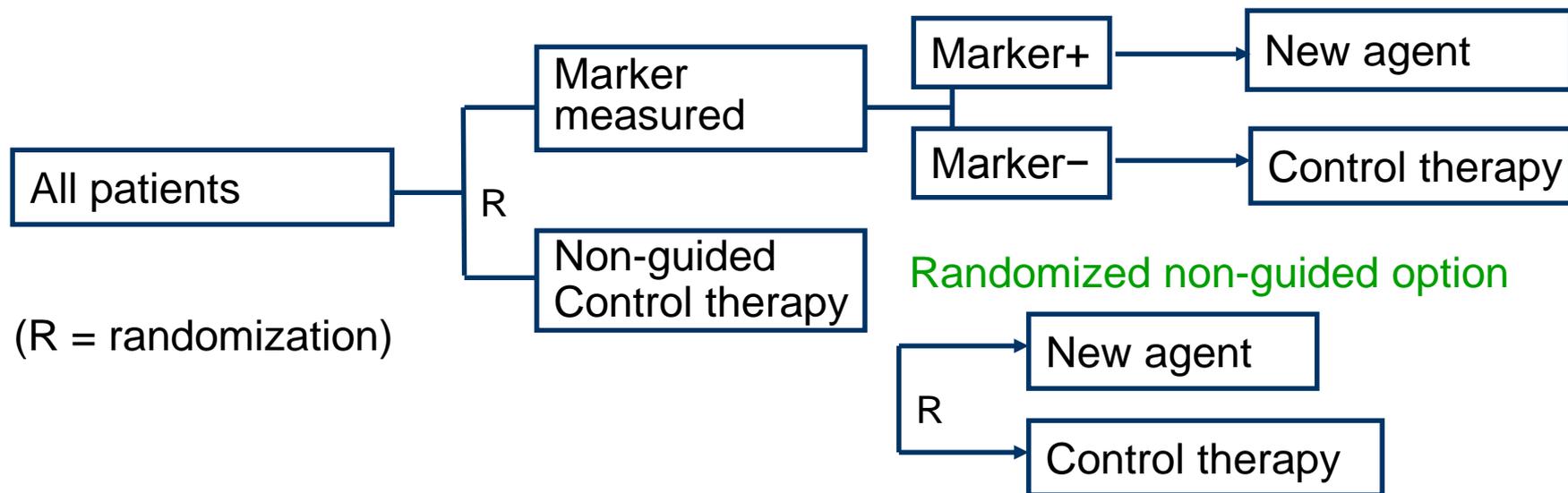
Freidlin B et al. *J Natl Cancer Inst* 2010;102:152-160

# Biomarker-enrichment design



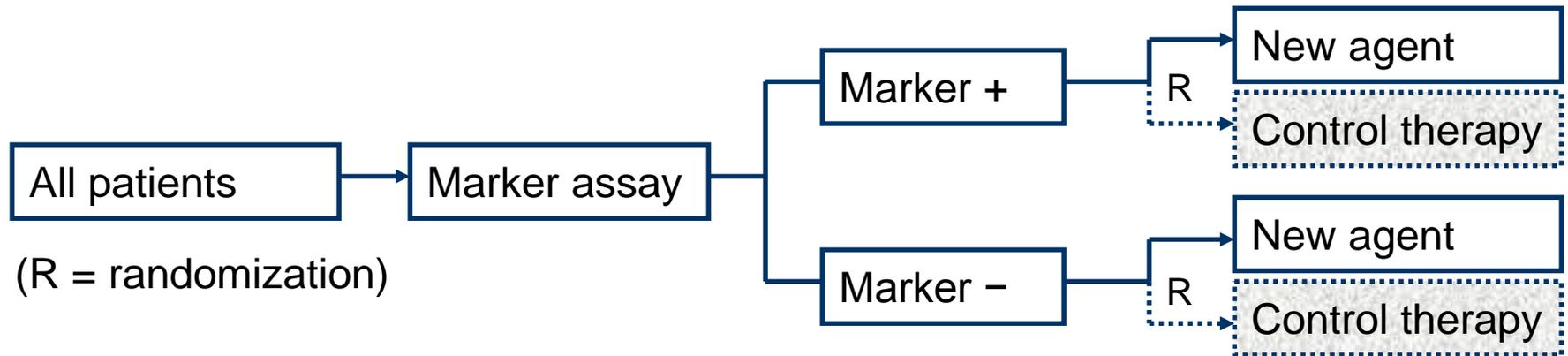
- Based in knowledge of biology (New agent → Molecular target)
- Control therapy arm controls for marker prognostic effect
- Variation: Standard therapy  $\pm$  new agent
- Limitations:
  - Off-target effects of new agent not fully evaluated
  - Regulatory indication limited to marker+ group
  - Marker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

# Biomarker-strategy design



- 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 )

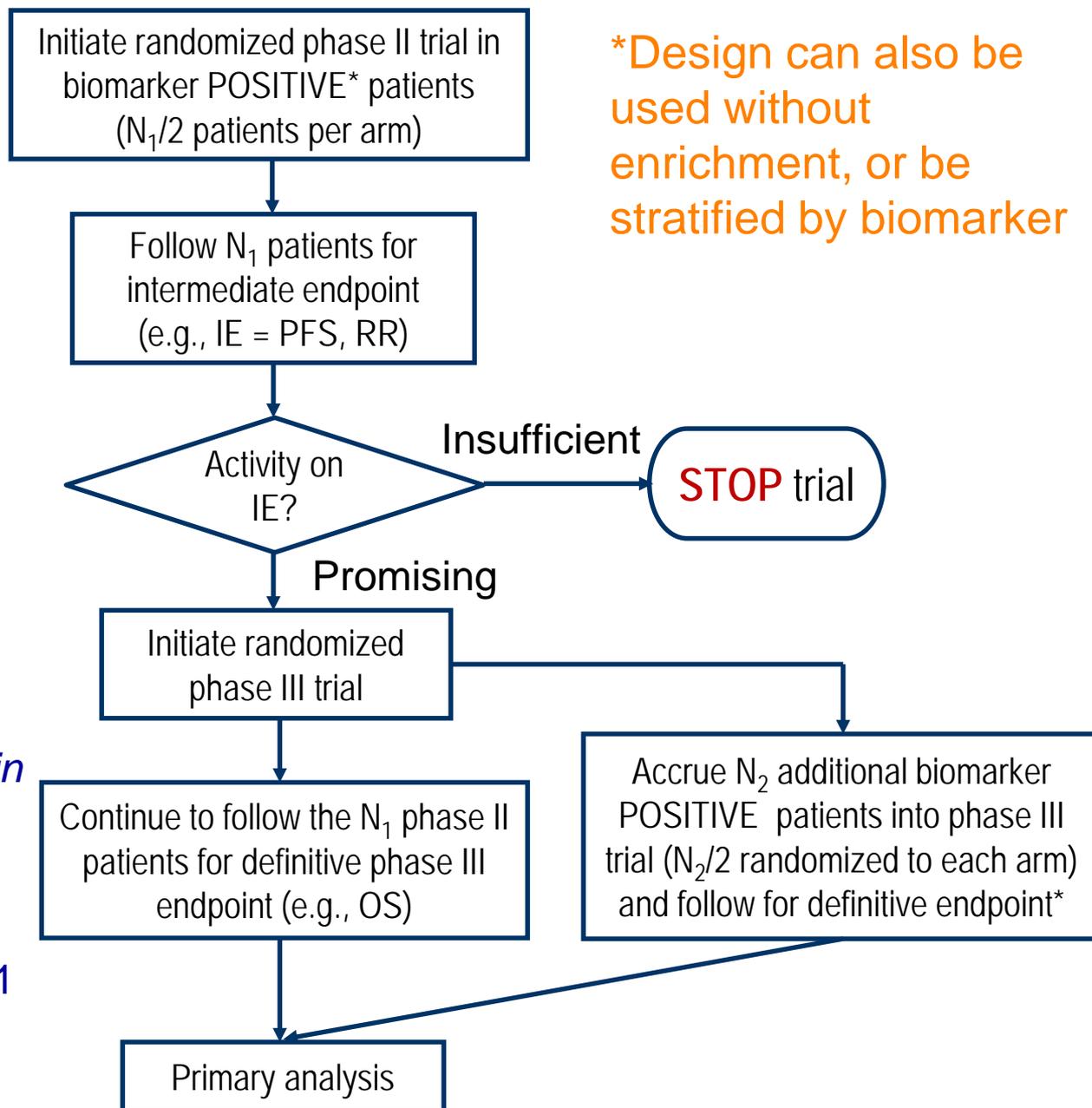
# Randomized phase II/III trial design

## Issues

- Choice of intermediate endpoint (IE)
- Define “promising” activity for Phase II (error rates, timing of analyses)
- Accrual suspension to allow Phase II data to mature

Hunsberger S et al., *Clin Cancer Res* 2009; 15:5950-5955

Korn E et al., *J Clin Oncol* 2012; 30:667-671



# Onartuzumab example

## Phase II trial followed by separate phase III trial

- MET – transmembrane receptor tyrosine kinase (RTK), which binds hepatocyte growth factor (HGF) is associated with poor prognosis and acquired resistance to EGFR-targeted drugs
- Onartuzumab (MetMab) – a recombinant, humanized monovalent monoclonal antibody targeting MET
- Extensive development to optimize MET IHC assay<sup>1</sup> for use in a phase II trial with MET status integral to one of co-primary hypotheses<sup>2</sup>

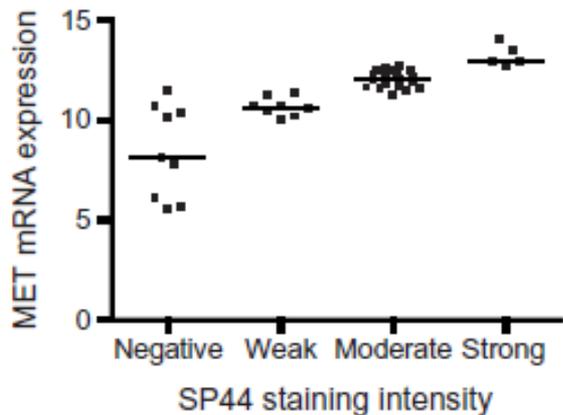
<sup>1</sup>Koeppen et al. *Clin Cancer Res* 2014;20(17):4488-4498

<sup>2</sup>Spigel et al. *J Clin Oncol* 2013;31(32):4105-4114

# Onartuzumab example (cont.)

## Phase II trial followed by separate phase III trial

- Development process for MET IHC assay<sup>1</sup>
  - 16 antibodies tested; SP44 selected
  - SP44 intensities associated with MET protein expression by Western, other anti-MET antibody, and flow cytometry; mRNA expression



**Fig S3.** MET mRNA levels vs. MET IHC staining intensity in NSCLC cell lines

Patient ID	Tissue slide	Cells staining at respective intensity (%)				Final call	
		Negative	Weak	Moderate	Strong	IHC score	Dx
1	A	15	70	15	0	1	Negative
1	B	45	50	5	0	1	Negative
2	A	0	0	60	40	2	Positive
2	B	0	0	90	10	2	Positive
2	C	0	0	100	0	2	Positive
3	A	0	20	70	10	2	Positive
3	B	0	25	70	5	2	Positive
4	A	100	0	0	0	0	Negative
4	B	100	0	0	0	0	Negative
5	A	0	0	90	10	2	Positive
5	B	0	0	90	10	2	Positive
5	C	0	0	90	10	2	Positive
5	D	0	0	90	10	2	Positive
5	E	0	0	90	10	2	Positive
6	A	0	0	100	0	2	Positive
6	B	0	0	100	0	2	Positive
7	A	0	70	30	0	1	Negative
7	B	0	75	25	0	1	Negative
8	A	0	20	50	30	2	Positive
8	B	25	0	75	0	2	Positive
9	A	0	0	100	0	2	Positive
9	B	0	0	100	0	2	Positive
10	A	80	0	20	0	0	Negative
10	B	100	0	0	0	0	Negative

Abbreviations: Dx, diagnostic

100% concordance (10/10) within-lab for positive vs. negative calls

**Table S4.** SP44 IHC scores for tissue sections cut from two separate blocks of 10 different NSCLCs

<sup>1</sup>Koeppen et al. *Clin Cancer Res* 2014;20(17):4488-4498

# Onartuzumab phase II NSCLC trial results

- **Successful** randomized double blind **phase II trial** of erlotinib (E) +/- onartuzumab (O) in patients (**n=137**) with recurrent advanced NSCLC (OAM4558g)<sup>1</sup>
  - **Co-primary endpoints: PFS in ITT and MET-POS**
  - **ITT pop'n:** PFS HR=1.09 (p=0.69), OS HR=0.80 (p=0.34)
  - **MET-POS** (54% MET-POS by IHC 2+/3+ on intensity and percent staining):
    - **PFS HR=0.53** (p=0.04), Median 1.5 (E) vs. 2.9 (E+O) mos. (27 vs 20 events)
    - **OS HR=0.37** (p=0.002), Median 3.8 (E) vs. 12.6 (E+O) mos. (26 vs 16 events)
  - **MET-NEG:** **PFS HR=1.82** (p=0.05), **OS HR=1.78** (p=0.16)

# Onartuzumab phase III NSCLC trial results

- MetLung Trial<sup>1</sup>: Randomized double blind phase III trial of erlotinib (E) +/- onartuzumab (O) in patients with recurrent advanced NSCLC who were MET-POS by IHC
  - Primary endpoint Overall Survival (OS)
  - Planned sample size N=490 randomized
- Stopped for futility after 499 patients enrolled (244 events)
  - O+E did not improve survival:  
HR=1.27, p=0.068, median OS 6.8 mos. vs. 9.1 mos.

**Wrong drug, biomarker, or biology; or just bad luck?<sup>2</sup>**

<sup>1</sup>Spigel et al. *J Clin Oncol* 2014;32:5s (suppl; abstr 8000)

<sup>2</sup>Hirsch et al. *Clin Cancer Res* 2014;20:4422-4424

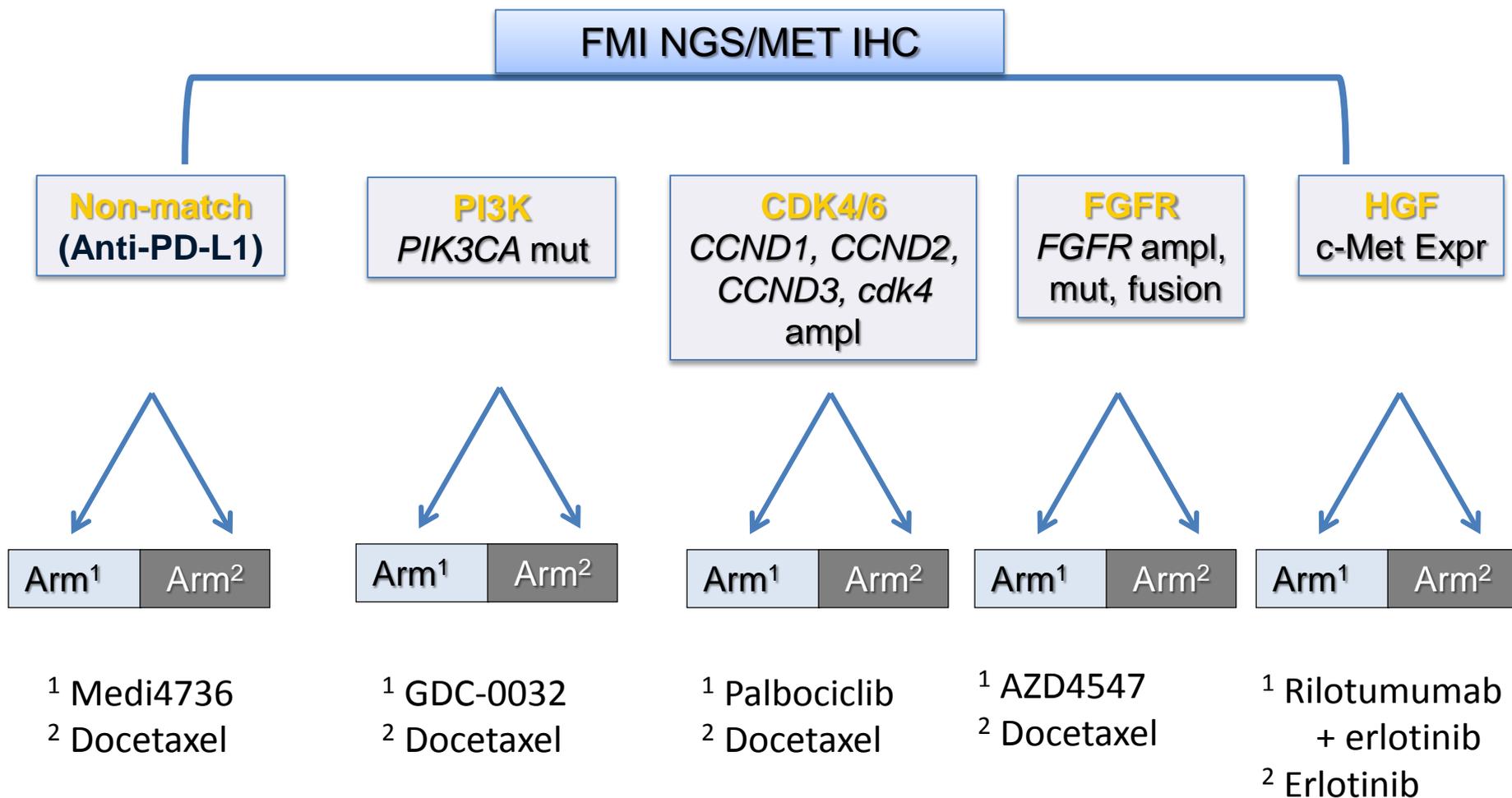
# New generation of oncology clinical trial designs (phase II, III, and II/III)

- 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 or adaptive (I-SPY2 trial, FOCUS trials)
- Combinations of the above (e.g., basket umbrella trial)
  - 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):djv003 (review of molecular profiling trials)

# Lung-MAP: Version 1

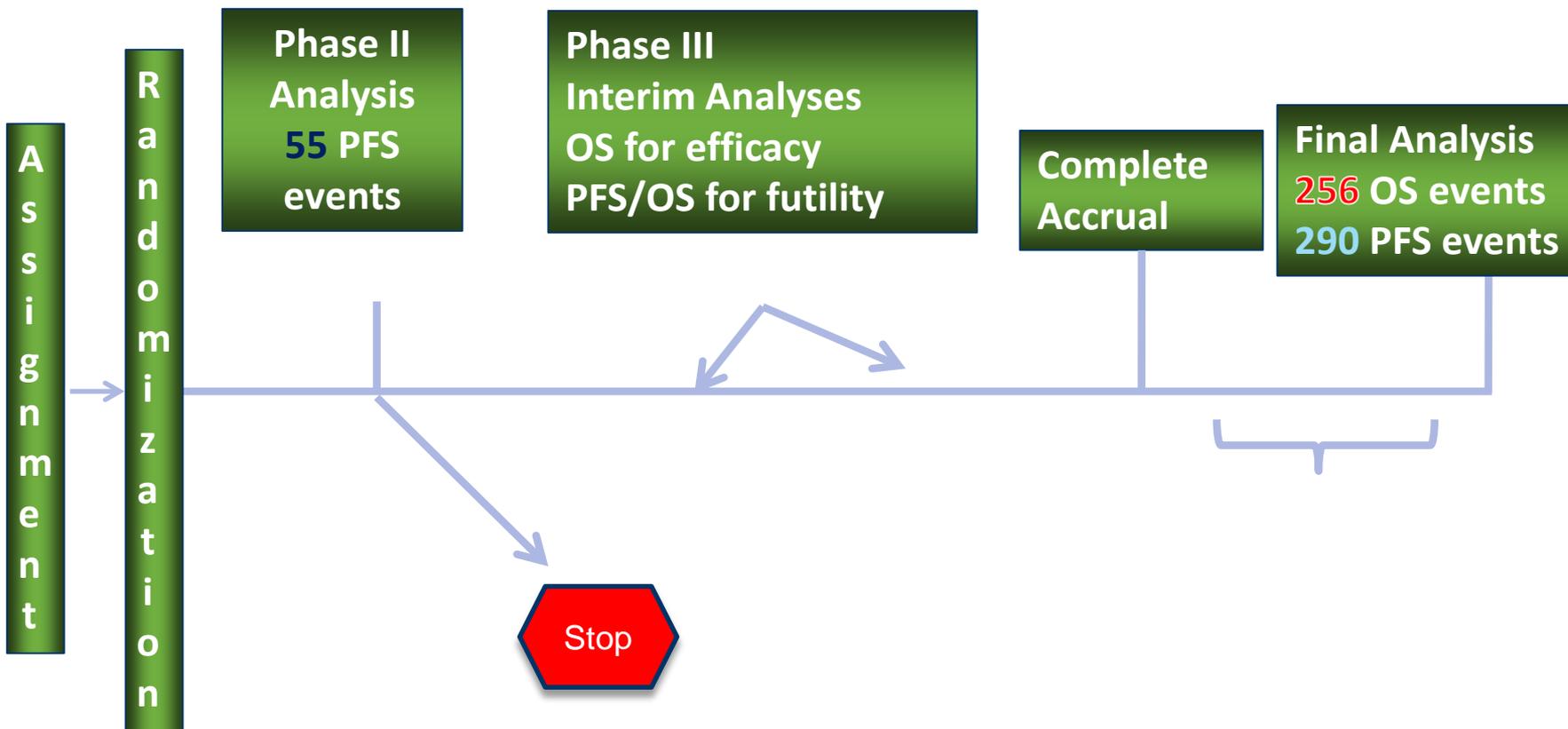
## Randomized phase II/III umbrella basket trial

Squamous NSCLC; incurable IIIB or IV; failed  $\geq 1$  chemo; measurable Disease; PS  $\leq 2$



# Lung-MAP: Version 1

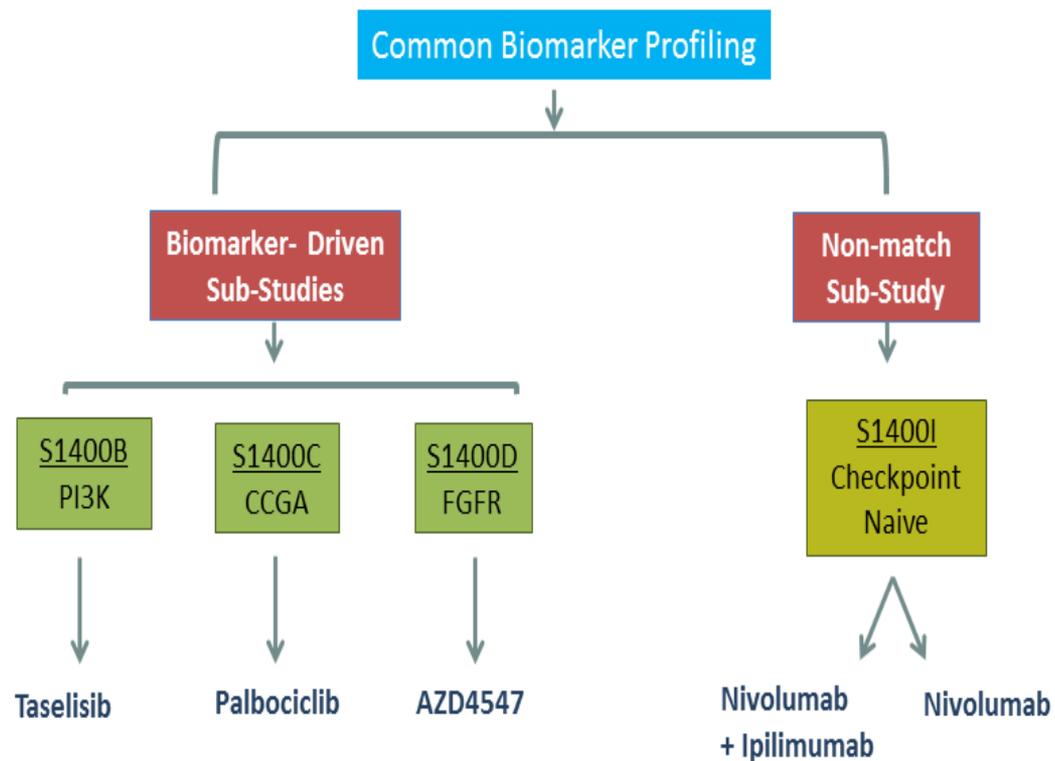
## Phase II/III design for each sub-study



# Lung-MAP TRIAL: Version 2

Lung-MAP (SWOG S1400) is a multi-drug, multi-sub-study, biomarker-driven squamous cell lung cancer clinical trial that uses state-of-the-art genomic profiling to match patients to sub-studies testing investigational treatments that may target the genomic alterations, or mutations, found to be driving the growth of their cancer.

Design change required after approval of nivolumab changed standard of care for advanced squamous NSCLC.



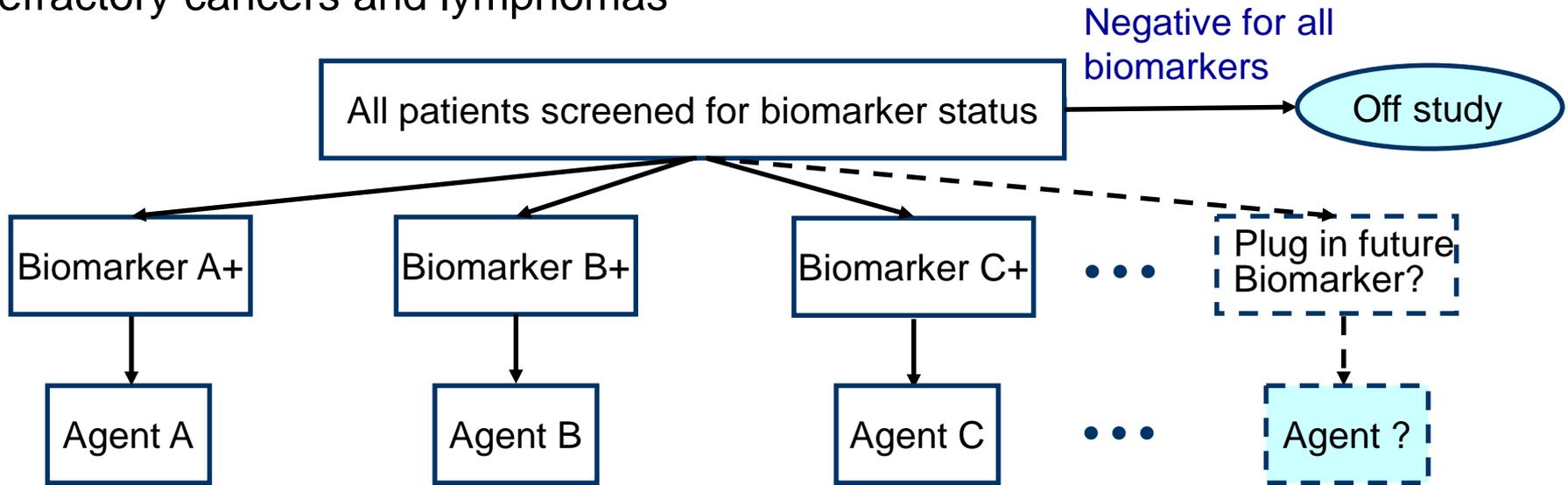
# NCI-MATCH TRIAL

Led by ECOG-ACRIN for NCI National Clinical Trials Network (NCTN)



**Primary objective:** To evaluate the proportion of patients with objective response (OR) to targeted study agent(s) in patients with advanced refractory cancers and lymphomas

**Secondary objectives:** To evaluate the proportion of patients with PFS  $\geq$  6 months of treatment with targeted study agent(s) in patients with advanced refractory cancers and lymphomas



**Signal finding trial:** Patients who have advanced disease that progressed on at least one standard therapy or for which there is no known effective therapy. Master screening protocol directing to multiple biomarker-based mixed histology single arm phase II trial sub-protocols.

<https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/nci-match>  
<http://ecog-acrin.org/nci-match-eay131>

# NCI MATCH trial design\*

- One-stage design (each arm)
  - 31 evaluable patients per arm
  - N = 35 total
- 4066 mutations of interest on targeted NGS panel (plus PTEN IHC) used for treatment assignment
- Primary endpoint: Overall Response Rate
  - H0:  $\leq 5\%$  vs Ha: 25%
  - Reject H0 if  $\geq 5/31$  responses
  - Type I error 1.8% / arm (one-sided)
  - Power 92%
- Secondary endpoints: Progression Free Survival (PFS)
  - 6 months 15% (med PFS 2.2 m) vs 35% (med PFS 4 m)

Arm	Target	Drug(s)
A	EGFR mut	Afatinib
B	HER2 mut	Afatinib
C1	MET amp	Crizotinib
C2	MET ex 14 sk	Crizotinib
E	EGFR T790M	AZD9291
F	ALK transloc	Crizotinib
G	ROS1 transloc	Crizotinib
H	BRAF V600	Dabrafenib+ trametinib
I	PIK3CA mut	Taselisib
N	PTEN mut	GSK2636771
P	PTEN loss	GSK2636771
Q	HER 2 amp	Ado-trastuzumab emtansine
R	BRAF nonV600	Trametinib
S1	NF1 mut	Trametinib
S2	GNAQ/GNA11	Trametinib
T	SMO/PTCH1	Vismodegib
U	NF2 loss	Defactinib
V	cKIT mut	Sunitinib
W	FGFR1/2/3	AZD 4547
X	DDR2 mut	Dasatinib
Y	AKT1 mut	AZD 5363
Z1A	NRAS mut	Binimetinib
Z1B	CCND1,2,3 amp	Palbociclib
Z1D	dMMR	Nivolumab

\*Currently screening > 100 patients per week to find matches for 24 arms

# Challenges: Current and future

- Interdependency between validation of a predictive biomarker and establishing efficacy of matched drug
  - Biomarker enrichment in trials only needs to be “good enough” to find sufficient drug activity; tension with goals for patient care
- Biomarker and its assay may evolve over a series of trials; refinements may occur after trials completed
  - Danger of repeatedly “refined” biomarker assays wandering too far from actual clinical outcome data
- Randomized and “all-comers” trials increasingly challenging due to rapid acceptance of targeted drugs based on robust response rates alone and sometimes premature faith in a biomarker
- As biomarker-defined subgroups continue to shrink, umbrella-type trials for screening on national or international scale will become essential
  - NCI MATCH experience suggests enthusiasm is high