ENGELBERG CENTER for Health Care Reform at BROOKINGS

Public Meeting: Advancing the Use of Biomarkers and Pharmacogenomics in Drug Development

Engelberg Center for Health Care Reform
The Brookings Institution

Washington Plaza Hotel • Washington, DC
Friday, September 5, 2014

Session Ia Introduction: Critical issues in biomarker development for clinical trial enrichment

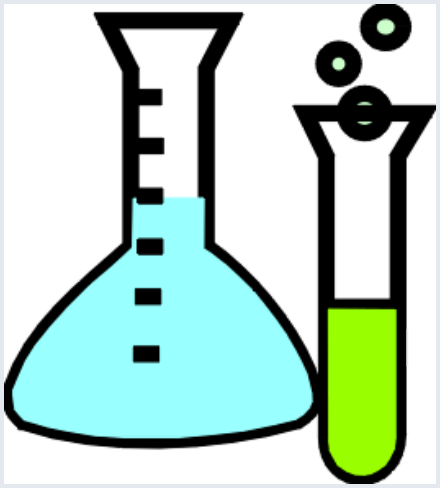
Lisa M McShane, PhD
Biometric Research Branch
National Cancer Institute

Advancing the Use of Biomarkers and Pharmacogenomics in Drug Development Meeting

Washington, DC
September 5, 2014

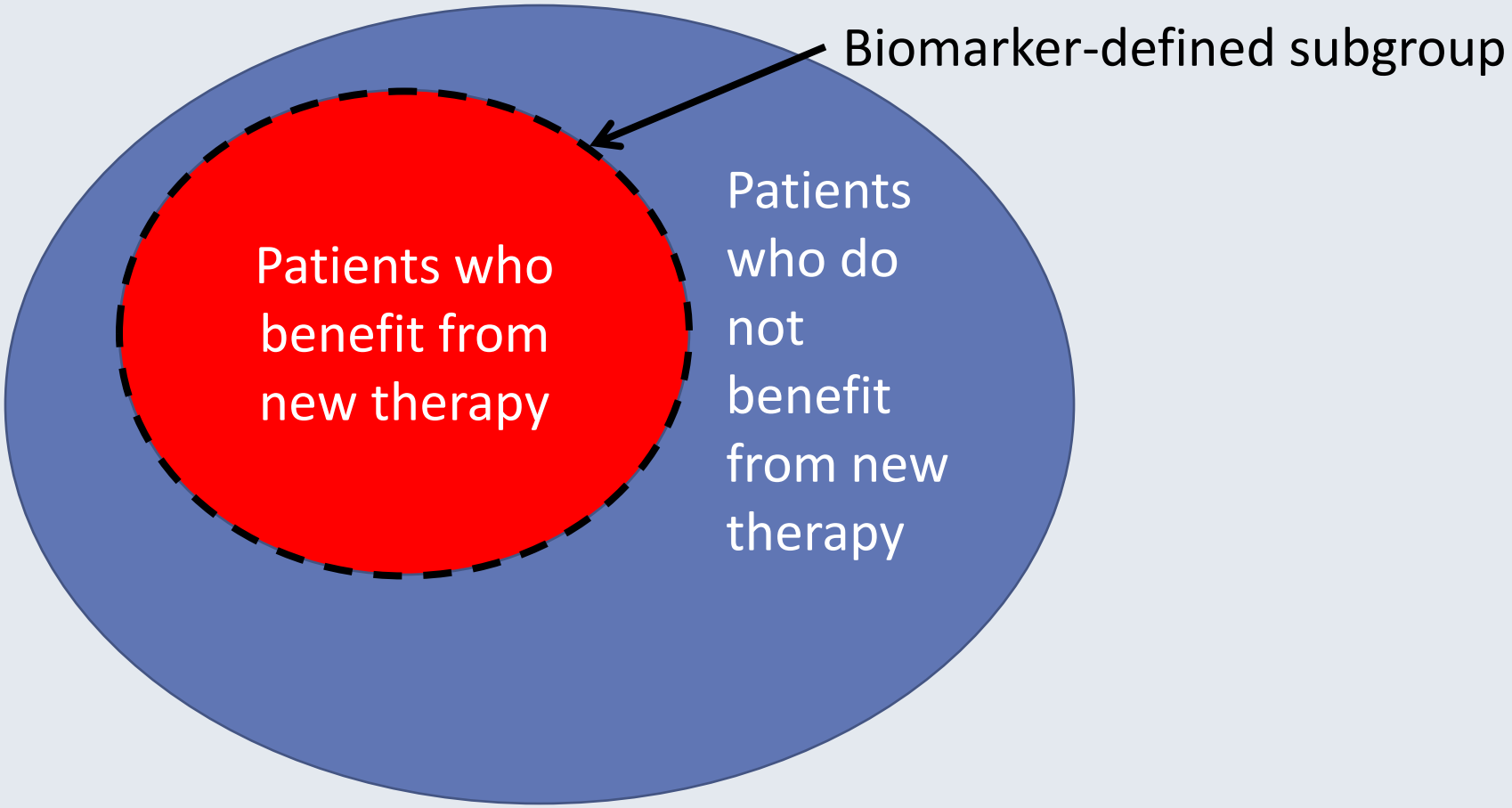
Biomarker and therapy co-development is an iterative process

Identify interesting biomarker

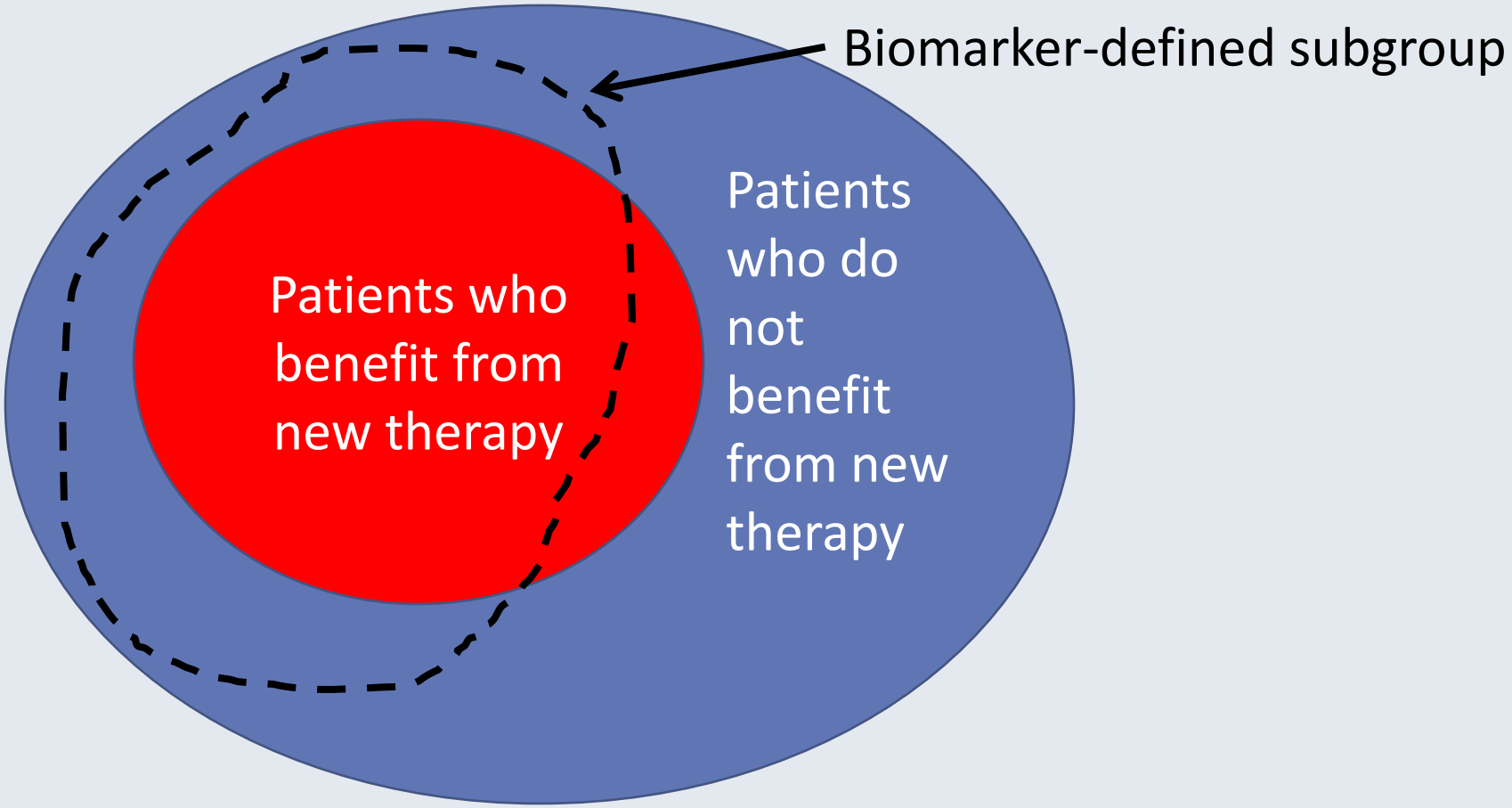


Engineer therapeutic agent to target biomarker

An “ideal” biomarker



A typical biomarker



Initial steps for biomarker assay development

- What molecular format: protein, RNA, or DNA level?
- Preliminary testing of association between biomarker and agent activity
  - Cell lines
  - Animal models/xenografts
  - Phase I trial responses (may be rare)
- Cutpoint determination (if applicable)
- Do results from non-human systems transfer to human clinical setting?

Minimal requirements to move forward to test biomarker in clinical specimens

- Assay analytical performance
  - 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 to see signal
  - Later biomarker refinement often needed

Prospective vs. retrospective evaluation of biomarker

- Retrospective
  - Need availability of adequate number and type of specimens from trials involving relevant treatment(s)
  - Avoid data-dredging to “salvage” failed treatment trial
  - Can be performed rigorously (“prospective-retrospective” study)

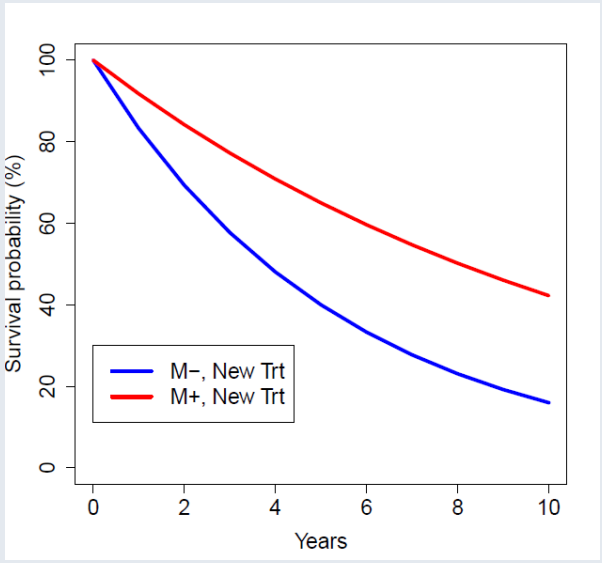
    Simon R et al., *J Natl Cancer Inst* 2009;101:1446–1452

    Polley M et al., *J Natl Cancer Inst* 2013;105:1677-1683
- Prospective
  - Many design options
  - Strive for flexibility to refine biomarker

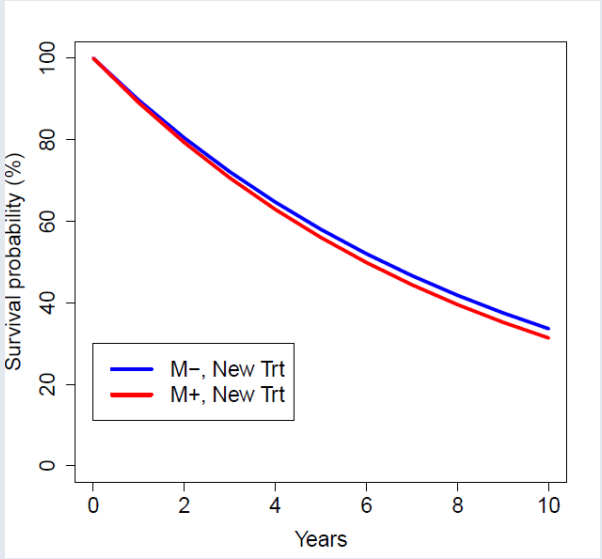
Key issues in evaluation of a biomarker for therapy selection

- Be careful to distinguish prognostic effects of biomarker from treatment effects
- What must be established about treatment effect in the biomarker-negative subgroup?

First instincts . . .



Biomarker ***is useful*** to identify patients who will benefit from new therapy?



Biomarker ***is not useful*** to identify patients who will benefit from new therapy?

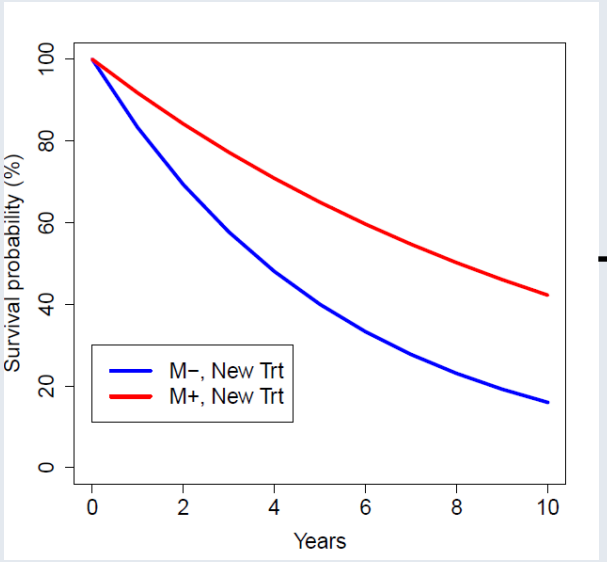
. . . may be wrong in judging value of biomarker for therapy selection

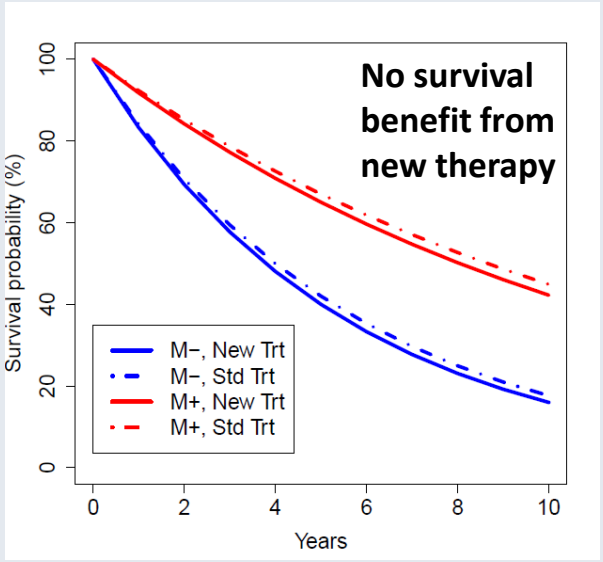
Prognostic and Predictive

- **PROGNOSTIC:** Biomarker-based test producing result associated with clinical outcome in absence of therapy (natural course) *or with standard therapy all patients are likely to receive*
- **PREDICTIVE:** Biomarker-based test producing result associated with benefit or lack of benefit (potentially even harm) from a particular therapy relative to other available therapy
  - **Alternate terms:** treatment-selection, treatment-guiding, treatment effect modifier

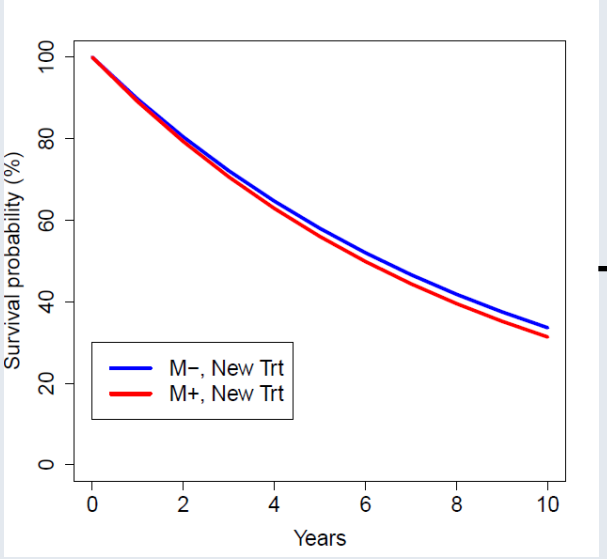
Polley M et al., *J Natl Cancer Inst* 2013;105:1677-1683

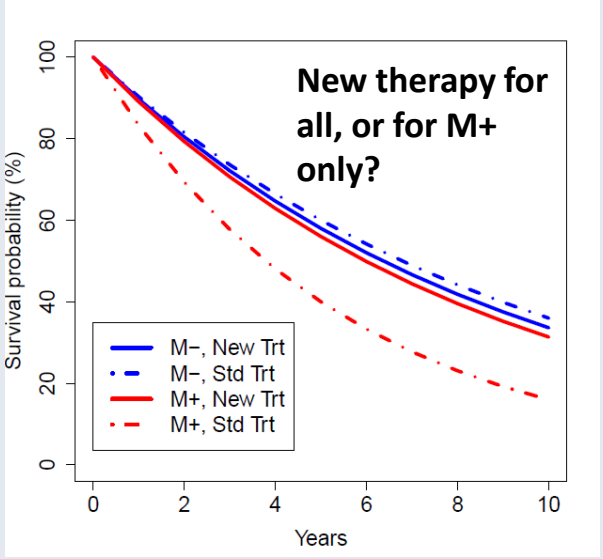
Prognostic vs. predictive: Importance of control groups





Prognostic but not predictive





(M = biomarker)

Prognostic and predictive

Statistical language for examination of predictive markers

- **Treatment by marker interaction:** Treatment hazard ratio in biomarker-positive group divided by treatment hazard ratio in biomarker-negative subgroup
  - Qualitative interaction
    - No benefit of new therapy (none or possibly inferior) in the biomarker-negative group
    - Treatment benefit in the biomarker-positive group
  - Quantitative interaction
    - Treatment benefits all patients but may work better for marker positive than for biomarker-negative
    - In some situations all patients should receive same treatment

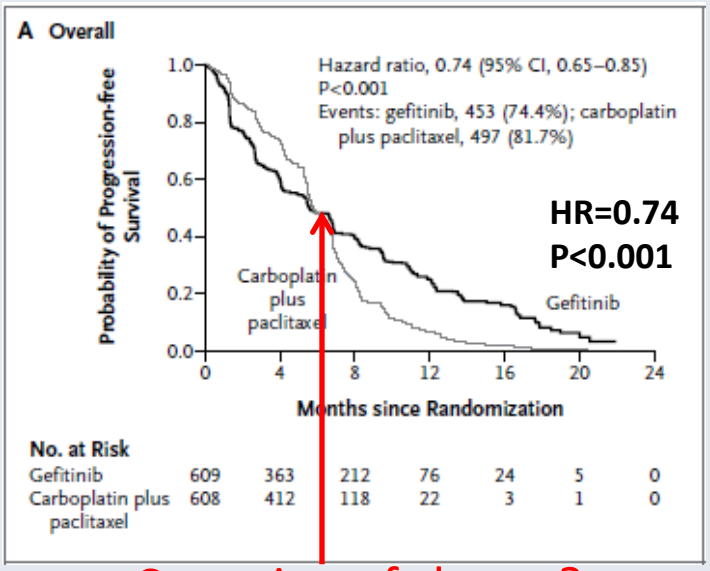
(Preferably would like to show a *statistically significant* interaction, but statistical power is often limited for test of interaction.)

IPASS Trial: EGFR mutation as a predictive biomarker for gefitinib in NSCLC (PFS)

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

EGFR mutation is:

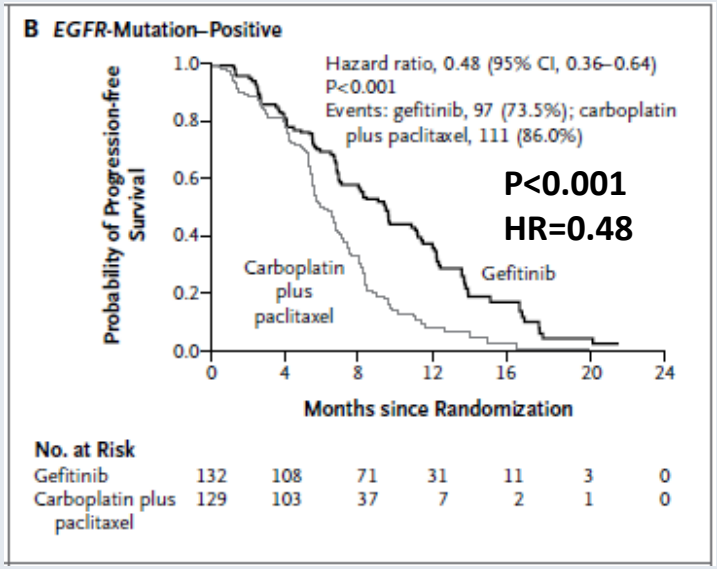
- Positive prognostic factor
- Positive predictive factor for gefitinib benefit (qualitative interaction, p<0.001)

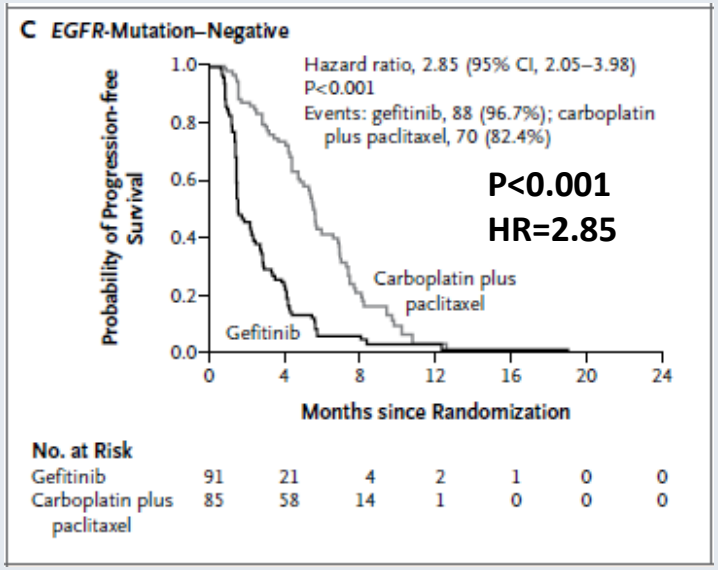


Cessation of chemo?

QUALITATIVE INTERACTION

IPASS: Phase III 1st line advanced adeno NSCLC gefitinib vs. carboplatin+paclitaxel





Plasma IL-6 as a predictive biomarker for pazopanib in metastatic renal-cell cancer?

(Tran H et al., *Lancet Oncol* 2012;13:827-837)

QUANTITATIVE INTERACTION

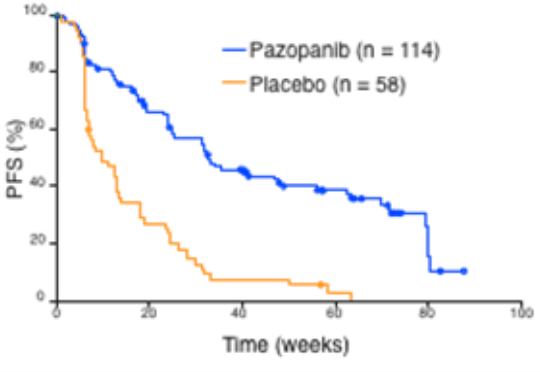
- High plasma IL-6 concentration is prognostic for shorter PFS
- High plasma IL-6 concentration is predictive for improved relative PFS benefit from pazopanib compared to placebo

Is IL-6 helpful for selecting therapy?

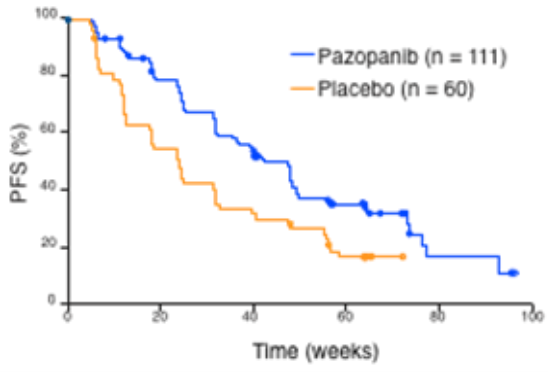
	PFS (weeks)		HR (95% CI)	p value		
	Pazopanib	Placebo		Pazopanib	Placebo	Interaction
Interleukin 6						
Low	42·3	24·0	0·55 (0·38–0·81)	0·445	<0·0001	0·009
High	32·6	9·9	0·31 (0·21–0·44)			

(Adapted from Figure 2 of Tran et al.)

High IL-6



Low IL-6



(Randomized placebo-controlled phase 3 trial)

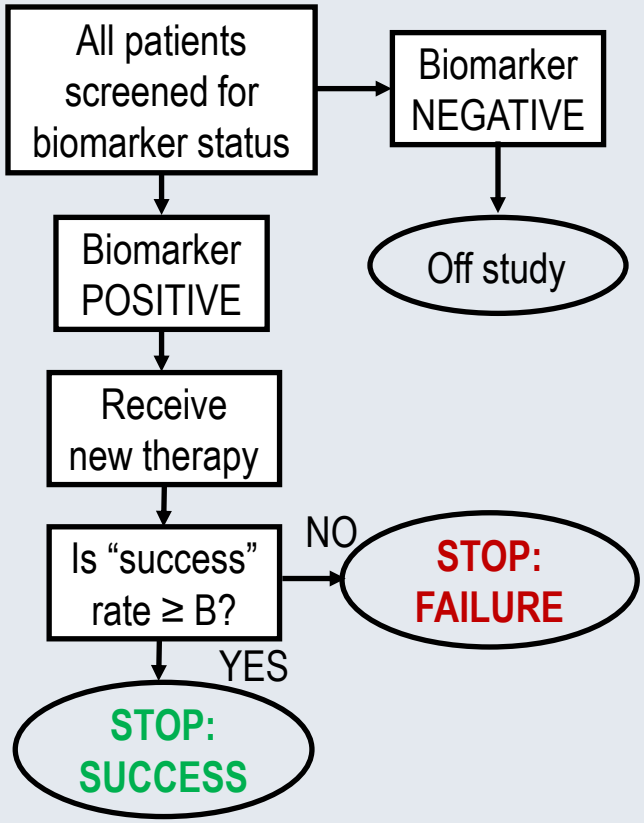
PROSPECTIVE phase II trial design considerations: Role of biomarker

- Biomarker enrichment
  - Biomarker positivity required for trial eligibility
- Biomarker adaptive
  - Trial design features adapted during course of the trial depending on early results within biomarker-positive and -negative subgroups
- All-comers with biomarker stratification
  - Consider results combined and separately within biomarker-positive and -negative subgroups

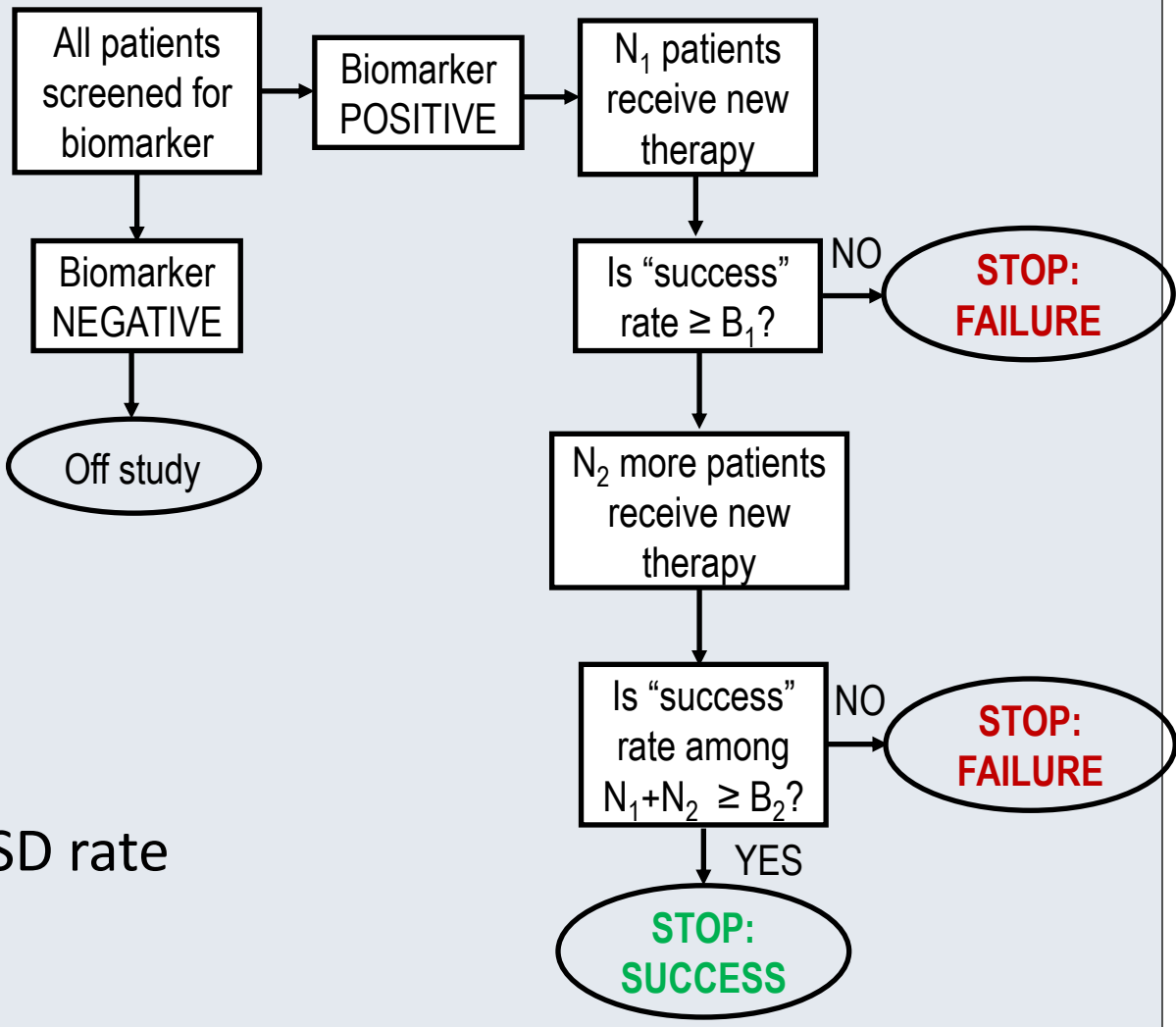
McShane L et al., *Clin Cancer Res* 2009;15:1898-1905
McShane L & Hunsberger S, An overview of phase II clinical trial designs with biomarkers. In *Design and Analysis of Clinical Trials for Predictive Medicine*, in press.

Single-arm biomarker enrichment phase II designs

One-stage design

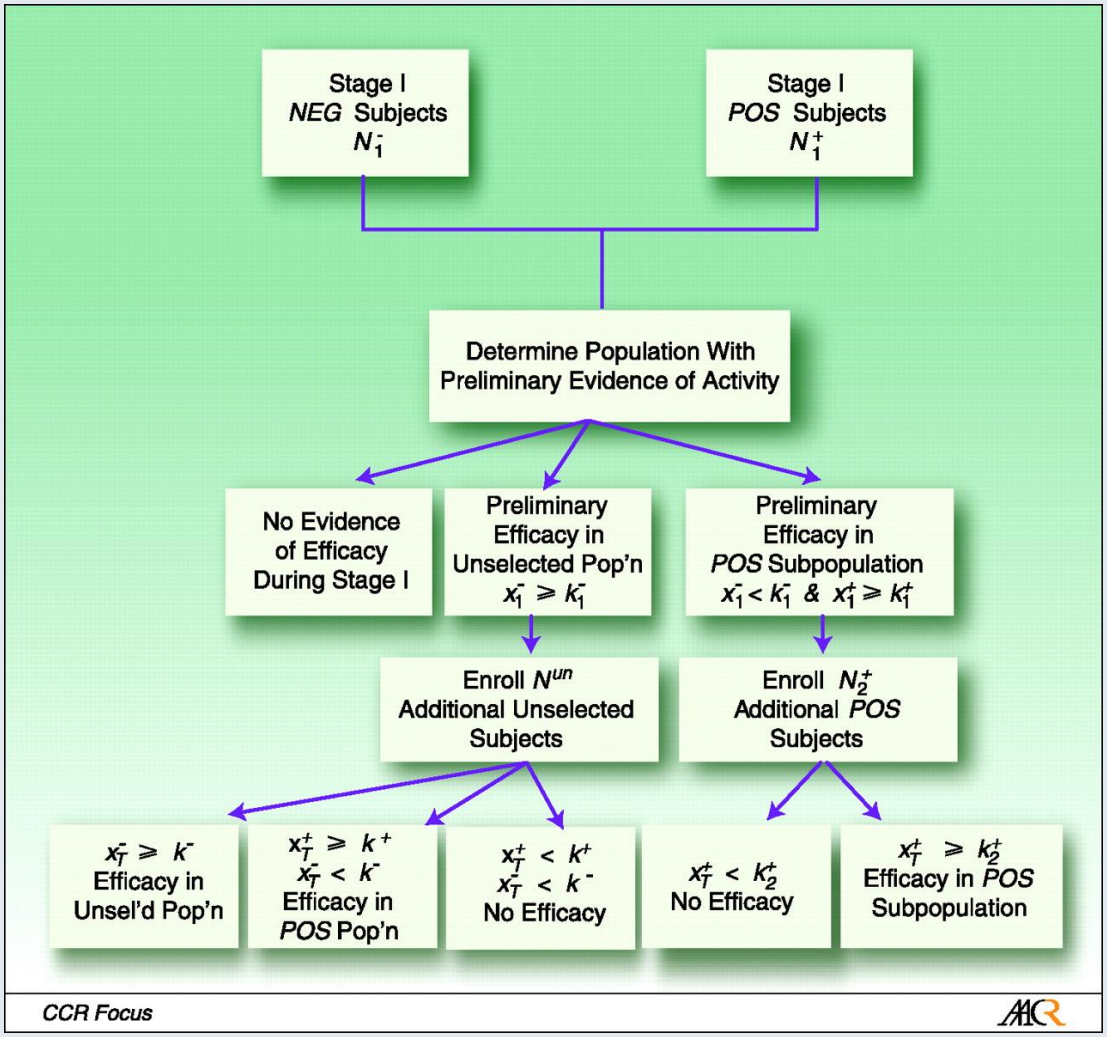


Two-stage design



- Endpoint: ORR, PFS or SD rate
- Typically 30-40 patients
- Limitations:
  - Appropriate benchmark success rate if biomarker is prognostic?
  - Can't assess off-target effects or refine biomarker outside "POSITIVE" group

Schema of the adaptive parallel two-stage design

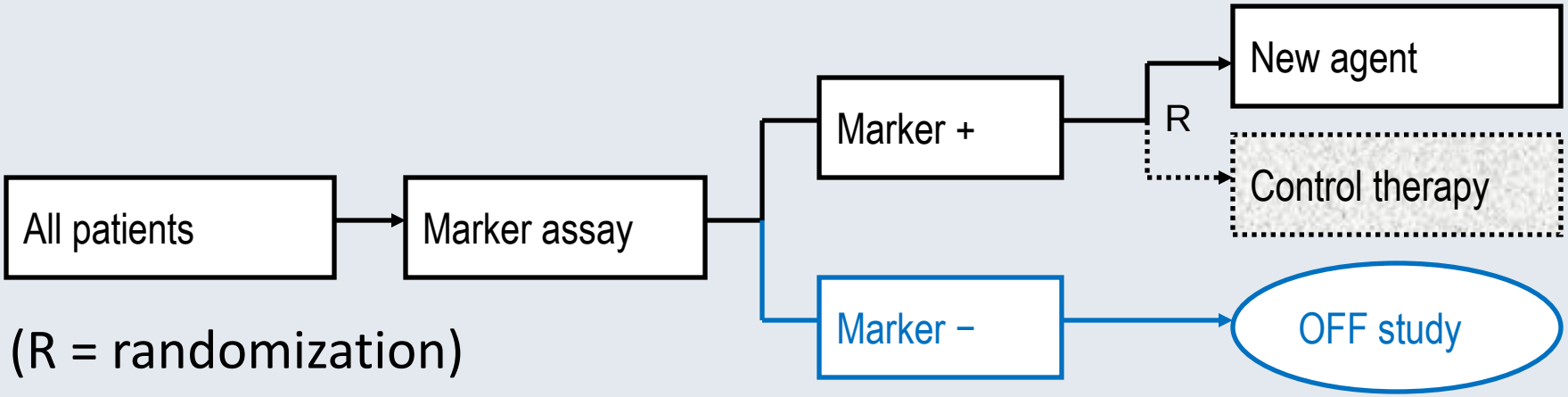


McShane L et al., *Clin Cancer Res* 2009;15:1898-1905, adapted from Jones C & Holmgren E, *Contemp Clin Trials* 2007; 28:654-61

PROSPECTIVE phase II trial design: When is a randomized trial necessary?

- Is the biomarker prognostic?
- Is it possible for a patient's condition to improve and/or resolve with no treatment?
- Are other standard therapies available for the intended patient population?
- Will the new therapy be tested *in combination* with an existing standard therapy (standard therapy $\pm$ new agent)?

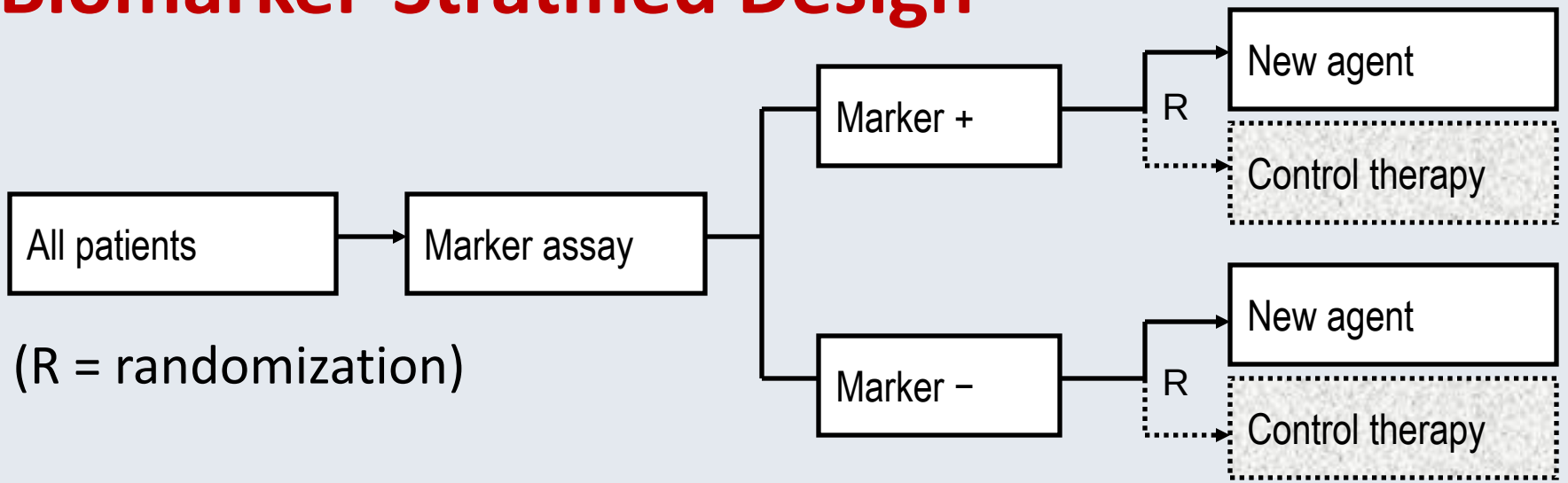
Randomized biomarker-enrichment design



(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-positive subgroup
  - Marker refinement within trial (form of marker or assay) limited to biomarker-positive group

Biomarker-Stratified Design



(R = randomization)

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

Challenges in studying the biomarker negative subgroup

- When are preliminary data sufficiently convincing that biomarker negative patients *should not* be included in trials of the new therapy?
- If a small benefit of new therapy is seen in biomarker-negative patients, is biomarker testing justified?
  - Ratio of benefit (e.g., slightly improved outcome) to harm (e.g., treatment toxicity & cost, risk& cost associated with biomarker testing)?

If additional information about efficacy of new therapy in biomarker-negative subgroup is needed . . .

- Must randomized trial be conducted in biomarker-negative subgroup prior to drug approval for biomarker-positive?
  - Should new therapy for biomarker-positive be “held hostage”?
- Is post-marketing evaluation of therapy in biomarker-negative subgroup feasible?
  - Formal clinical trial
  - Registry – controlled access with data return required for evidence development?

Needs for more rapid and efficient biomarker and targeted therapy development

- Resources for pre-clinical work and assay development (specimens, animal models, reagents)
- Guidance on assay performance requirements and on acceptable post hoc biomarker adjustments
- Broadly accessible trials to accrue sufficient numbers in small biomarker subgroups
  - Nationwide trial accrual system
  - Coordination & comparison of assays among multiple trials
  - Multi-arm trials (“basket”, “umbrella”, “platform” trials) give options for more patients/fewer biomarker-negative

THANK YOU!

Lm5h@nih.gov

ENGELBERG CENTER for Health Care Reform at BROOKINGS

Public Meeting: Advancing the Use of Biomarkers and Pharmacogenomics in Drug Development

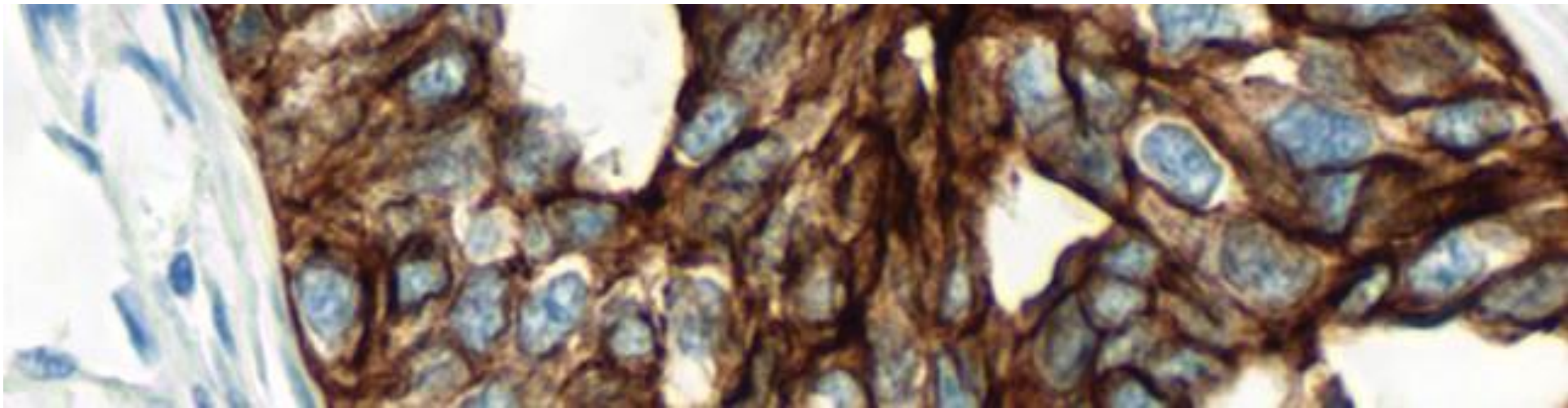
Engelberg Center for Health Care Reform
The Brookings Institution

Washington Plaza Hotel • Washington, DC
Friday, September 5, 2014

Approaches to Collaborative Co-Development of Therapies and Diagnostics

Tracy Bush, PhD
Roche Diagnostics

September 5, 2014



Approaches to Collaborative Co-Development

- Personalized medicines and companion diagnostics can have a huge impact on patients in need.
- Collaboration is necessary for efficient co-development.
- There have been many successes and progress in working with the FDA on co-development.
  - Best practices have been identified in several areas.
  - Recent guidance is especially appreciated.
- Additional clarity is needed on several points.

Diagnostics and therapeutic sponsors must partner with each other and with the Agency to find solutions to the remaining challenges of co-development

Challenges & Best Practices--Use of CDx in Early Phase Therapeutic Trials

- When an investigational assay is used to make a patient management decision during a trial, the device is subject to IDE regulations.
  - Very different from exploratory biomarker research.
  - Regulations necessary to ensure patient safety.
  - IVD manufacturers are familiar with requirements, but Pharma sponsors are not.
- FDA policy is evolving to offer trial sponsors risk-based approaches and options to comply with the requirements.

We urge the FDA to release the draft guidance on Use of Investigational Devices in Clinical Investigations of Therapeutic Products.

Challenges & Best Practices--Communication Between Agency / Manufacturers

- Recent OIR reorganization created a new Division of Molecular Genetics & Pathology.
  - More consistent translation of evolving FDA CDx polices to the project/reviewer level.
- Oncology divisions have led the way in best practices such as inter-center consults and “4-sided” meetings.
- Patients in other disease areas need personalized medicine; and Dx industry is developing CDx based on other technologies besides molecular detection and genetics.

We encourage FDA to ensure that these communication path-ways and best practices extend to other review divisions in both the drug and device review centers.

Challenges & Best Practices--When the CDx is Not Identified Prior to Confirmatory Trials

- FDA has outlined several best practices including bridging studies
- FDA has released draft guidance describing innovative approaches for the late identification and refinement of biomarker thresholds.
- These approaches are at odds with OIR's standard expectation that the assay cutoff must be selected and validated in separate studies; and the assay cutoff should be predefined.
- It is in the patient's best interest to consider the totality of the data in ultimate selection of the most appropriate cutoff.

We ask FDA to clarify that a CDx developed using an adaptive trial design or a refined cutoff should not always be subject to additional validation studies prior to making it available on the market for patient use.

Challenges & Best Practices--When the CDx is Not Identified Prior to Confirmatory Trials

- Final Guidance : markers not “required” in drug labeling are not CDx
- Greater clarity is needed regarding the criteria and requirements for “recommended” vs “required” marker testing.
  - Clear criteria would help industry to make this determination as early as possible in the co-development process.
  - Especially important when the marker is identified late in the drug development.

We ask FDA to clarify that contemporaneous approval should not be required for “recommended” markers.

We urge FDA to strongly consider use of the de novo pathway for co-developed IVDs that are “recommended.”

Challenges & Best Practices--When the CDx is Not Identified Prior to Confirmatory Trials

- Acceleration of drug development (e.g. via Breakthough Therapy Designation) poses a major challenge to CDx co-development.
  - Early phase trials can become pivotal, and CDx may not be ready for submission.
- Expedited Access PMA pathway offers a new pathway for certain Dx to reach patients sooner while still maintaining standards of safety and efficacy.
  - Guidance includes many risk-based approaches developed in collaboration with industry and patient advocacy groups such as Friends of Cancer Research.

We ask FDA to clarify that all CDx should be automatically eligible for the EAP pathway.

We encourage FDA to outline how the EAP program might be leveraged to encourage developments of CDx for orphan indications.

Additional Challenges & Best Practices

- Drug developers are often asked to make investigational therapies available to patient who have no other options. Some of these therapies are targeted drugs that require a CDx.

When FDA determines to provide early access to a drug, the agency needs to align the pathway for providing the companion diagnostic.

Doing now what patients need next

Brookings Institution

Biomarkers as Replacement or Surrogate Endpoints

September 5, 2014

Thomas R. Fleming, Ph.D.
Professor, Dept. of Biostatistics
University of Washington

* IOM, 2010. "*Evaluation of Biomarkers & Surrogate Endpoints in Chronic Disease*:. Washington DC. National Academies Press
* Fleming TR, Powers JH: Biomarkers and Surrogate Endpoints in Clinical Trials *Statistics in Medicine* 2012; 31: 2973-2984.

Some Characteristics for Study Endpoints in Clinical Trials

- Consistently & readily measurable
- Sensitive
- Well defined & reliable
- Clinically meaningful

Invasive Procedures:
E.g., Biopsy, RHC

A "*Clinically Meaningful Endpoint*":

…a direct measure of how a patient "feels, functions or survives"…

… Robert Temple, FDA

Biomarkers & Clinically Meaningful Endpoints

- *Biological Activity:* *…Biomarkers as Surrogates…*
- *Clinical Meaningful Benefit*
  - ~ **Functions**: Ability to conduct normal activities
    - – *Ability to walk, Ability to engage in recreational activities, Ability for self care, Risk of syncope*
    - – *Time in hospital or missing school (overall, or cause specific)*
  - ~ **Feels**:
    - – *Chest pain, breathlessness, fatigue, dizziness*
  - ~ **Survives**

…Physician or Observer administered & PROs…

Biomarkers as Surrogates for Clinical Efficacy Endpoints

"Biomarkers are measurements of biological processes. Biomarkers include physiological measurements, blood tests and other chemical analyses of tissue or bodily fluids, genetic or metabolic data, and measurements from images.

Cholesterol and blood sugar levels are biomarkers, as are blood pressure, enzyme levels, measurements of tumor size from MRI or CT, and the biochemical and genetic variations observed in age-related macular degeneration..."

IOM, 2010. "Evaluation of Biomarkers & Surrogate Endpoints in Chronic Disease". Washington DC. National Academies Press.

Categorization of Nomenclature

Outcome Assessments

Direct Measures of Patient "Functions, Feels, Survives"

- Patient (symptoms: ***chest pain, dyspnea, fatigue, dizziness***)
- Clinician (*PANNS for* schizophrenia *syndrome,* ***Clinician Global Measures***)
- Observer (***seizures, infant behavior, stroke, death***)

Indirect Measures

Biomarkers

e.g. H_bA_{1c}, *CD-4, PSA,* ***PVRI, NT-proBNP, CO HR, Blood Pressure Pulm Arterial Pressure*** *TIMI-III flow HDL, LDL, body temperature, urine GAG, urine KS cardiac rhythm, blood cultures, PCR, quantitative measures from radiology imaging.*

Measures depending on patient motivation or clinician judgment to perform the test

- Patient (*rescue meds for pain, alcohol presentation test*)
- Clinician (*TM bulging, Limb Spasticity,* ***6MWD, 3MSC*** *PFTs, 9-hole peg test*)
- Observer (***rescue meds for pain)***

John Powers, Dave DeMets, Marc Walton, Laurie Burke, Bob Temple...

Biomarkers (as Replacement Endpoints)

…"Post hoc, ergo, Propter hoc"…

Treatment effects on Biomarkers:

- Establish *Biological Activity*
- But not necessarily *overall Clinical Efficacy*
  - ~ How a patient feels
  - ~ Ability to conduct normal activities
  - ~ Overall Survival

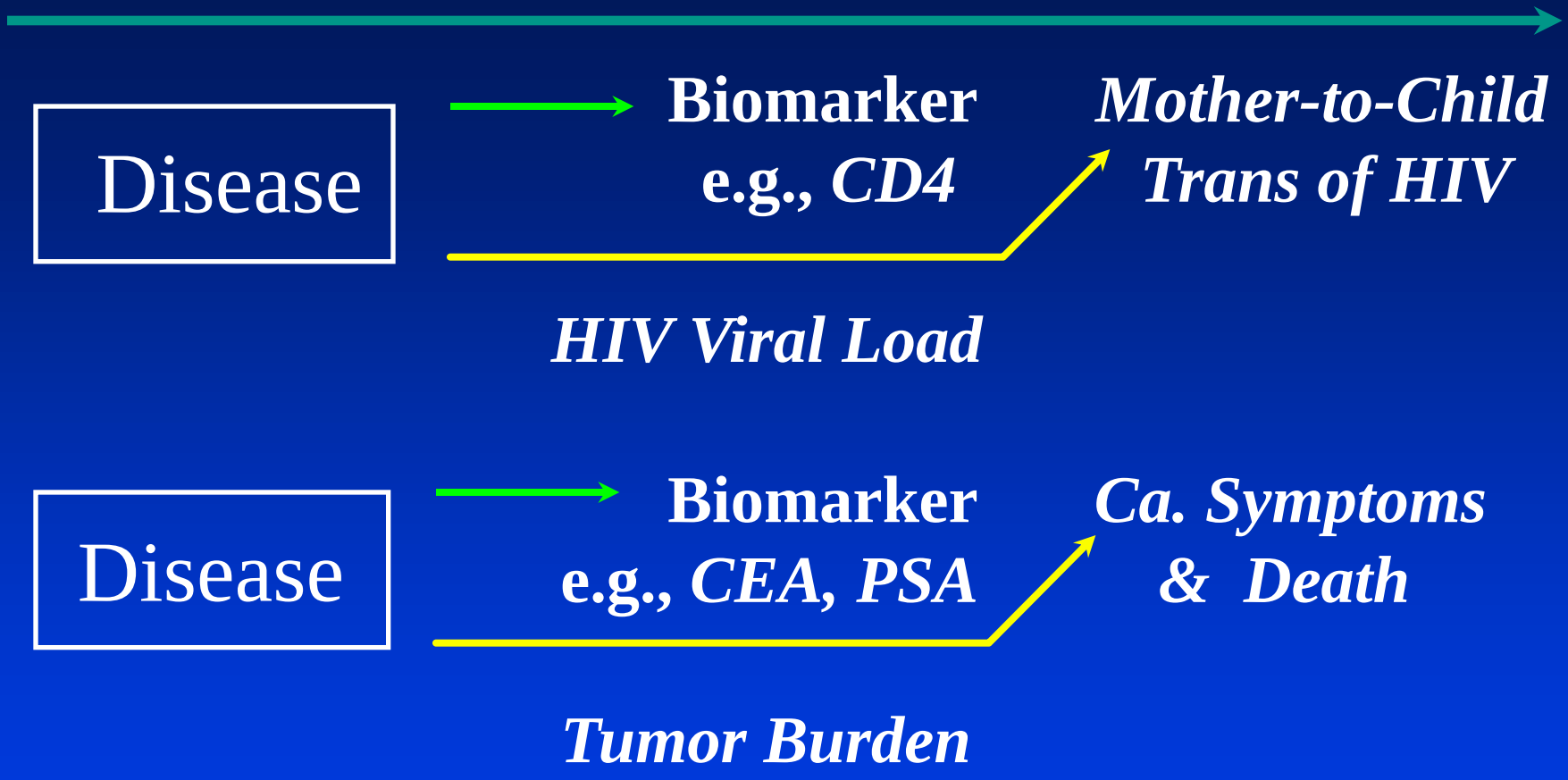
Issues in Surrogate Endpoints

~ Criteria for Choosing Endpoints

~ “**A Correlate does not a Surrogate Make**”

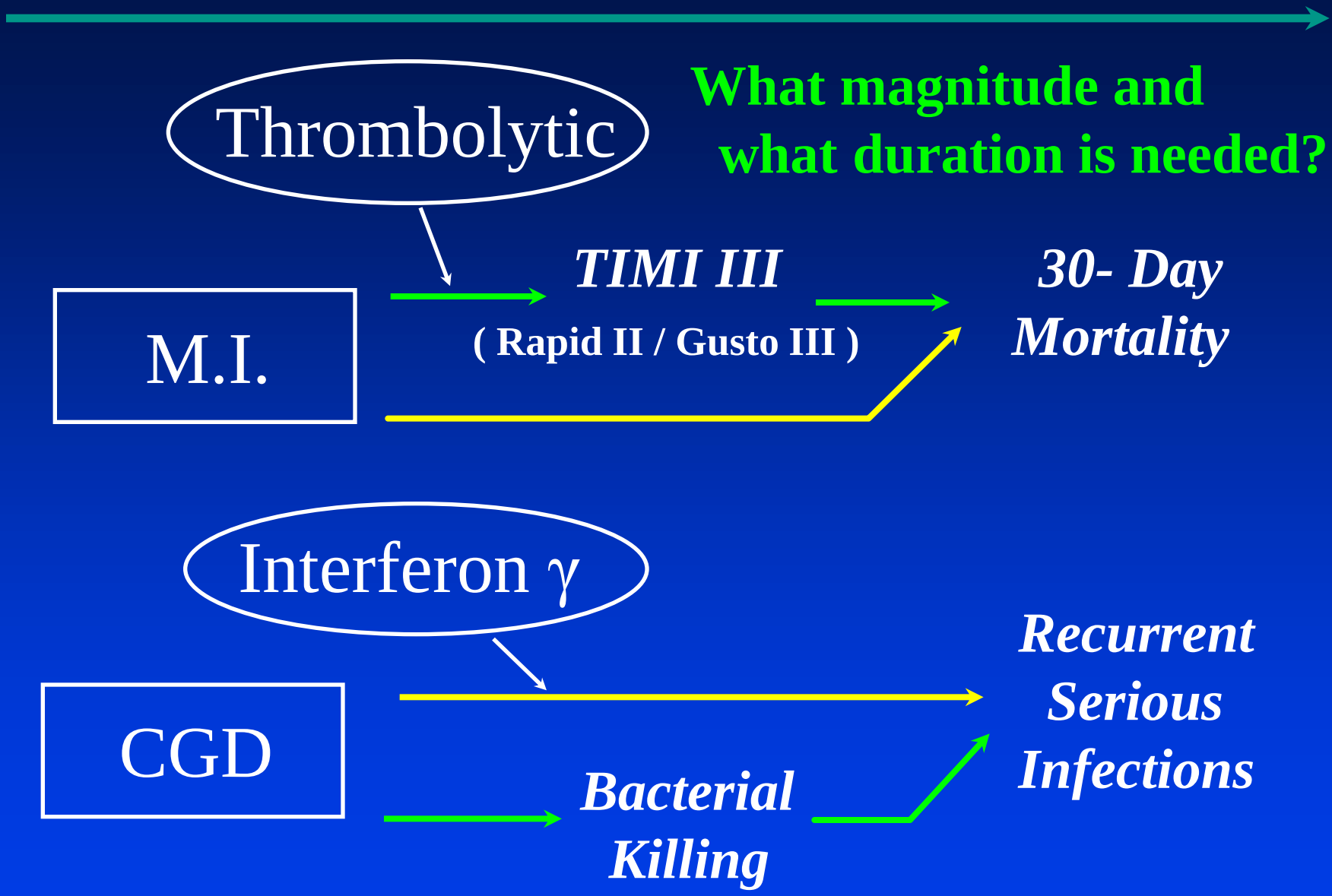
~ Validation of Surrogate Endpoints

The Biomarker Endpoint is not in the Causal Pathway of the Disease Process.



- "Correlates": Useful for Disease Diagnosis, or Assessing Prognosis
- "Valid Surrogates": Replacement Endpoints

Multiple Pathways of the Disease Process



Interventions having Mechanisms of Action Independent of the Disease Process

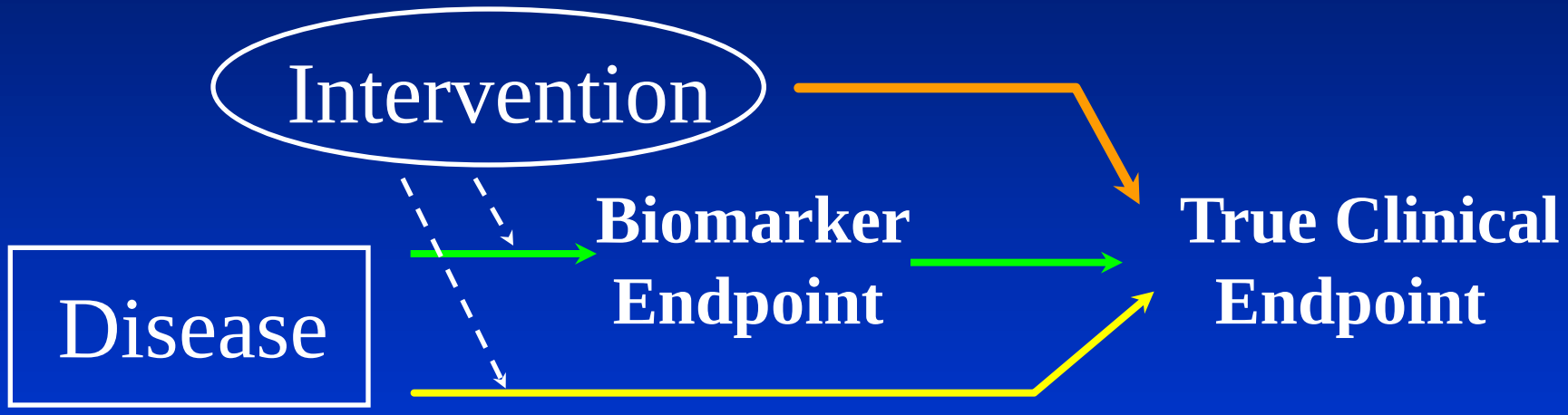


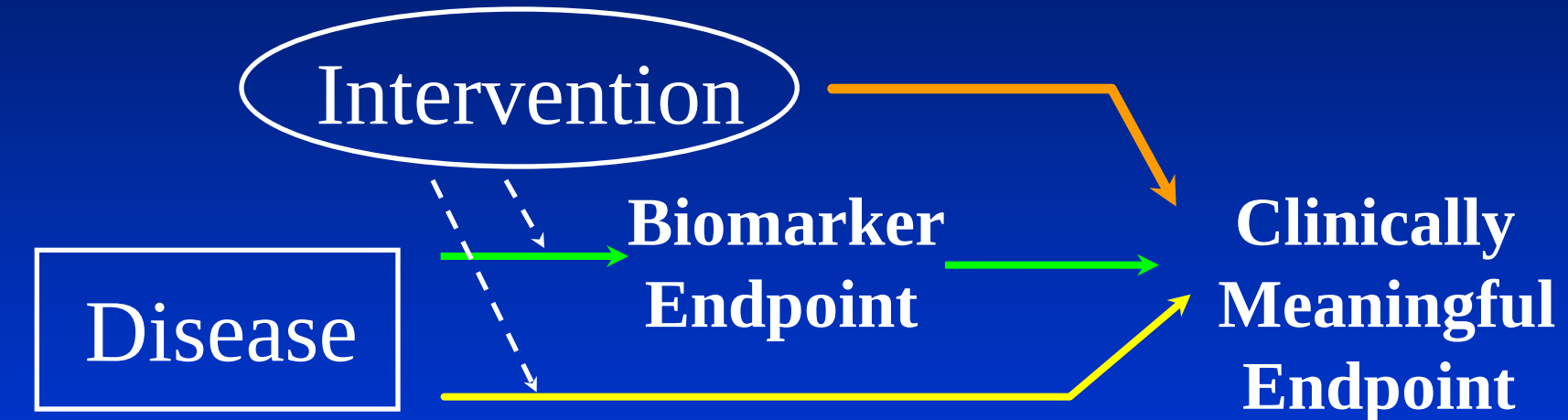
Illustration:
Ventricular Arrhythmia after M.I.

- Arrhythmia:
  - Risk factor for Sudden Death
- Antiarrhythmic Drugs:
  - Class IC antiarrhythmic agents
    …Strong Sodium-Channel Blockade

Cardiac Arrhythmia Suppression Trial:
The drugs, relative to placebo,
TRIPLE the death rate.

Interventions having Mechanisms of Action Independent of the Disease Process

Interventions having Mechanisms of Action Independent of the Disease Process



ESAs: ↑ **Thrombosis** ⇒ ↑ Mortality

Cox-2s, Muraglitazar, Rosiglitazone: ↑ **CV Risk Factors** ⇒ ↑ CV Death/ MI /Stroke

Natalizumab: ↑ **Prog. Multifocal Leukoencephalopathy** ⇒ ↑ Morbidity / Mortality

Torcetrapib: **Activates renin angiotensin system** ⇒ ↑ **BP** ⇒ ↑ Mortality

Troglitazone: ↑ **Serious Hepatic Risks** ⇒ ↑ Morbidity

Long Acting β-Agonists: ↑ Asthma-related deaths

Ezetimibe/Simvastatin: **Block pathways linked to CA prot.** ⇒ ↑ Cancer Mortality?

Issues in Surrogate Endpoints

~ Criteria for Choosing Endpoints

~ A Correlate does not a Surrogate Make

~ **Validation of Surrogate Endpoints**

Validation of Surrogate Endpoints

Property of a Valid Surrogate

- *Net effect of the Intervention on the Surrogate Endpoint*

reliably predicts the

Net effect of the Intervention on the Clinically Meaningful Endpoint

Indirect measures as a replacement for direct assessment of treatment benefit

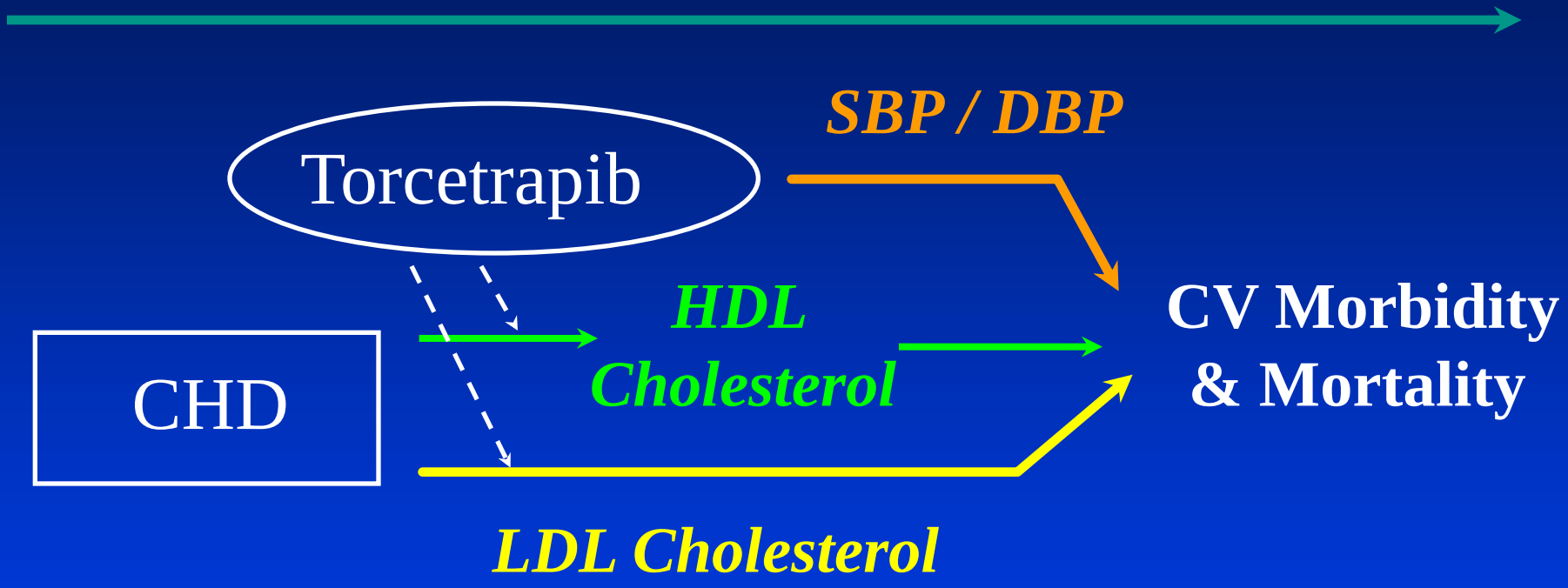
Clinical

- Comprehensive understanding of the
  - ~ Causal pathways of the disease process
  - ~ Intervention's intended and unintended mechanisms of action

Statistical

- Meta-analyses of clinical trials data

Mechanisms of Action of the Intervention & Causal Pathways of the Disease Process



Indirect measures as a replacement for direct assessment of treatment benefit

Clinical

- Comprehensive understanding of the
  - ~ Causal pathways of the disease process
  - ~ Intervention's intended and unintended mechanisms of action

Statistical

- Meta-analyses of clinical trials data

Illustration of Validating a Surrogate

- **Anti-Hypertensives**

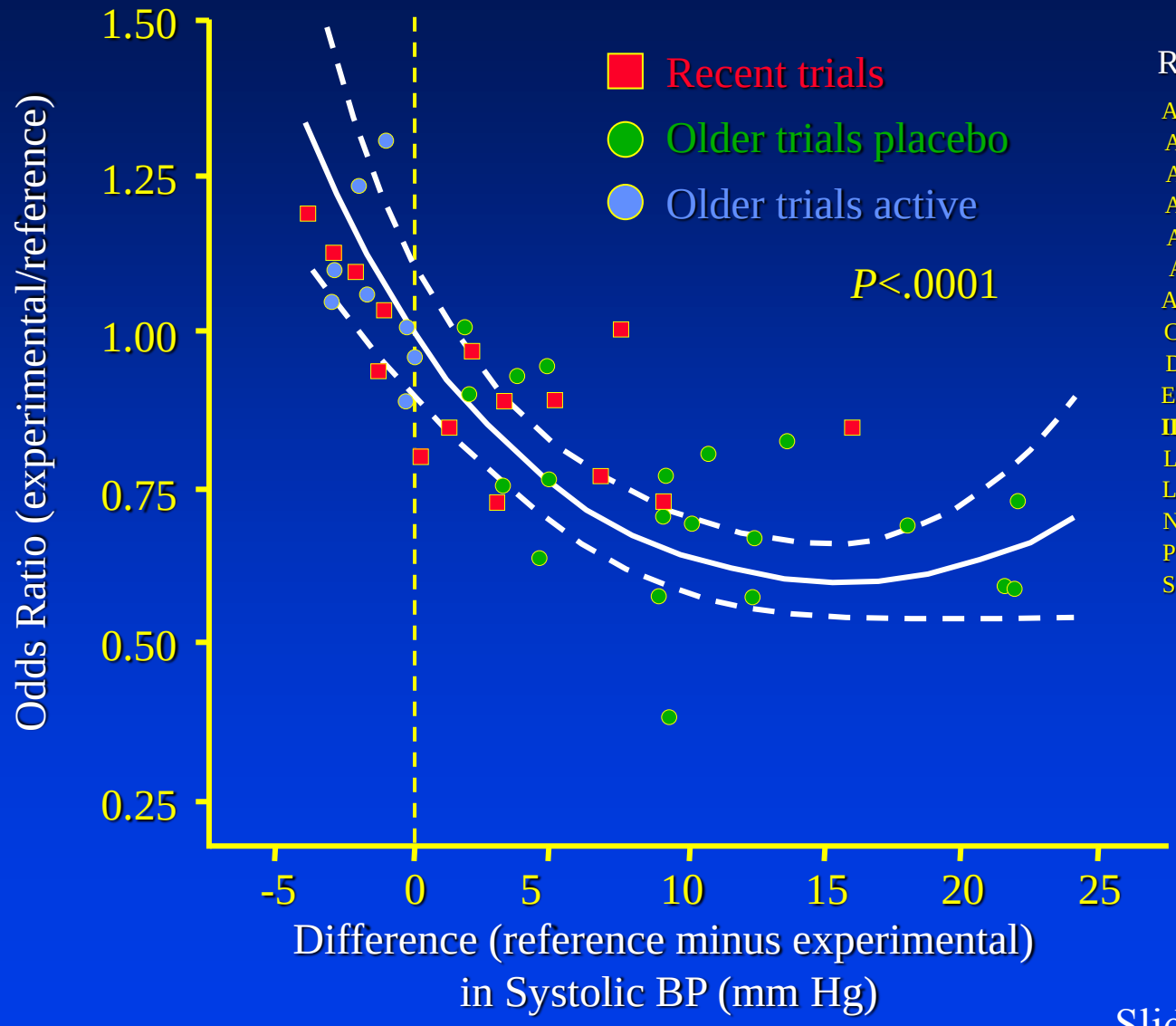
  (>500,000 patients from rand trials)

…β-blockers, low dose diuretics, ACE-I, CCBs, ARBs…
FDA Cardio-Renal Advisory Committee: 6/15/2005

- Effects on *Blood Pressure* predicting effects on each of the following, considered individually:

✓ *Stroke, MI, CVD, Mortality, Heart Failure*

Odds Ratio for CV Events and Systolic BP Difference: Recent and Older Trials



Recent	Older
AASK L vs. H	ALLHAT/Dox
ABCD/NT L vs. H	ATMH
ALLHAT/Aml	EWPHE
ALLHAT/Lis	HEP
ALLHAT/Lis ≥65	HOPE
ALLHAT/Lis Blacks	HOT
ANBP2	HOT M vs. H
CONVINCE	INSIGHT
DIABHYCAR	MIDAS/NICS/VHAS
ELSA	L vs. H
IDNT2	MRC
LIFE/ALL	MRC2
LIFE/DM	PART2/SCAT
NICOLE	PATS
PREVENT	PROGRESS/Per
SCOPE	PROGRESSION/Com
	RCT70-80
	RENAAL
	SHEP
	STONE
	STOP 1
	STOP2/CCBs
	STOP2/ACEIs
	Syst-China
	Syst-Eur
	UKPDS C vs. A
	UKPDS L vs. H

Staessen et al. *J Hypertens*. 2003;21:1055-1076.

Slide: Henry Black's lecture

Illustration of Validating a Surrogate

- **Anti-Hypertensives**

  (>500,000 patients from rand trials)

…β-blockers, low dose diuretics, ACE-I, CCBs, ARBs…
FDA Cardio-Renal Advisory Committee: 6/15/2005

- Effects on *Blood Pressure* predicting effects on each of the following, considered individually:

✓ *Stroke, MI, CVD, Mortality, Heart Failure*

IOM, 2010 "Evaluation of Biomarkers & Surrogate Endpoints in Chronic Disease"

- *Addressing Assay Performance*

  …analysis of analytical performance of an assay…
  e.g., limit of quantitation, across lab reproducibility, etc

- *Evidentiary Assessment*

  …relationship between biomarker & disease state
  …data regarding effects of interventions on both biomarker and clinically meaningful outcomes…

- *Justifying the Proposed Use*

  …determining whether available evidence provides sufficient justification for the context of use proposed…

Replacement Endpoints

- A replacement endpoint cannot be deemed to be a generic surrogate endpoint for a particular disease

Reasons why use needs setting-specific justification:

- Multiple causal pathways of the disease process
- *Magnitude* and *duration* of effect matters
- Intended and *unintended* effects of interventions

- How does evaluating replacement endpoints impact the public?

Response: Need "*reliable*" as well as "*timely*" evaluation …not simply *"a choice"*; rather, "*an informed choice*"

Principles & Insights

"A Correlate does not A Surrogate Make"

* Fleming TR, DeMets DL: Surrogate endpoints in clinical trials: Are we being misled? *Annals of Internal Med* 1996; 125:605-613.

* IOM, 2010. "*Evaluation of Biomarkers & Surrogate Endpoints in Chronic Disease*:. Washington DC. National Academies Press

* Fleming TR, Powers JH: Biomarkers and Surrogate Endpoints in Clinical Trials *Statistics in Medicine* 2012; 31: 2973-2984

ENGELBERG CENTER for Health Care Reform at BROOKINGS

Public Meeting: Advancing the Use of Biomarkers and Pharmacogenomics in Drug Development

Engelberg Center for Health Care Reform
The Brookings Institution

Washington Plaza Hotel • Washington, DC
Friday, September 5, 2014

ENGELBERG CENTER for
Health Care Reform
at BROOKINGS

Advancing the Use of Biomarkers and Pharmacogenomics in Drug Development

Washington Plaza Hotel • Washington, DC
Friday, September 5, 2014

Session IIb
Evidentiary Needs and Implications of Biomarkers as Surrogate Endpoints

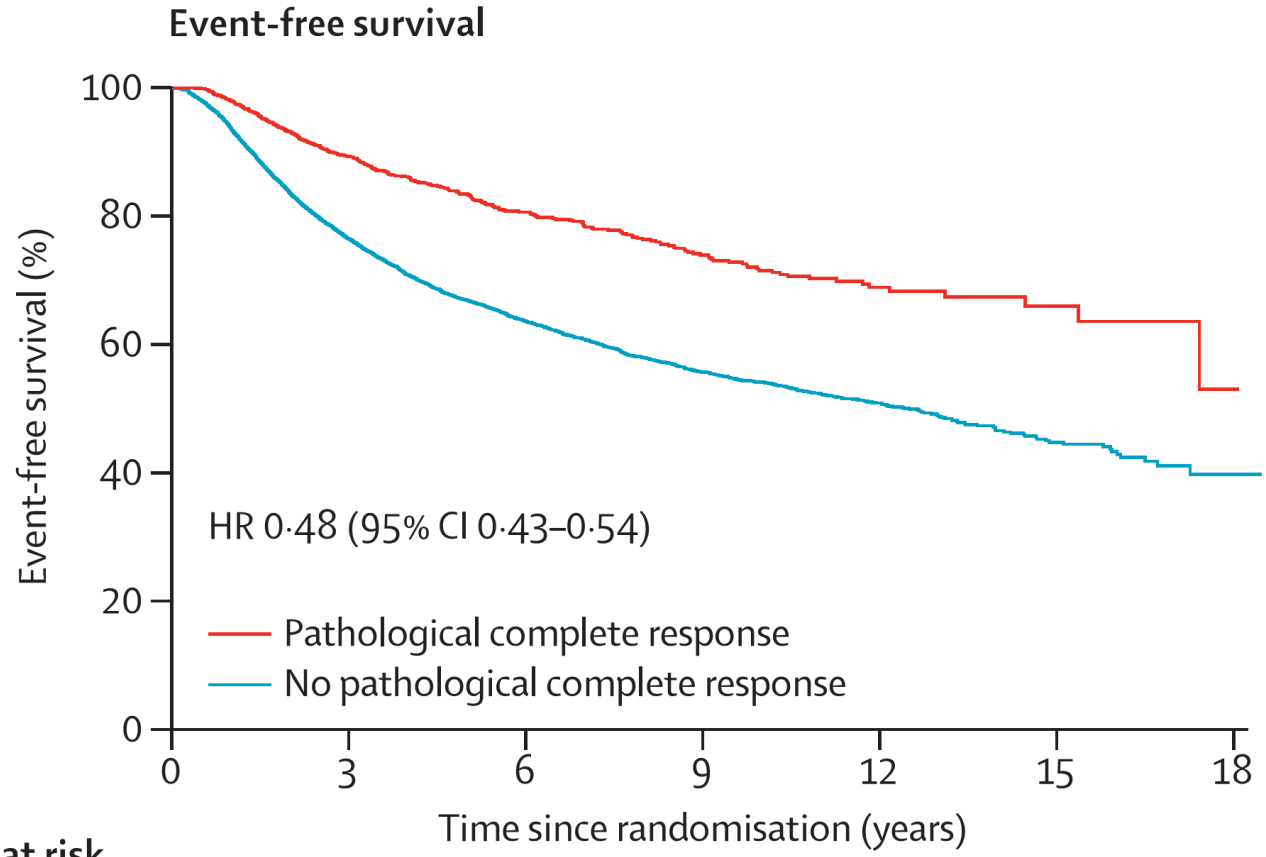
Marc Buyse, ScD
IDDI, San Francisco, CA

Resectable breast cancer: Is pCR a surrogate for EFS?

- Patients with resectable primary breast cancer (any subtype) receiving neo-adjuvant chemotherapies
- Surrogate endpoint: pathological complete response (pCR)
- True endpoint: event-free survival (EFS)
- Meta-analysis of 12 randomized trials including 11,955 patients

Ref: Cortazar et al, Lancet, February 2014.

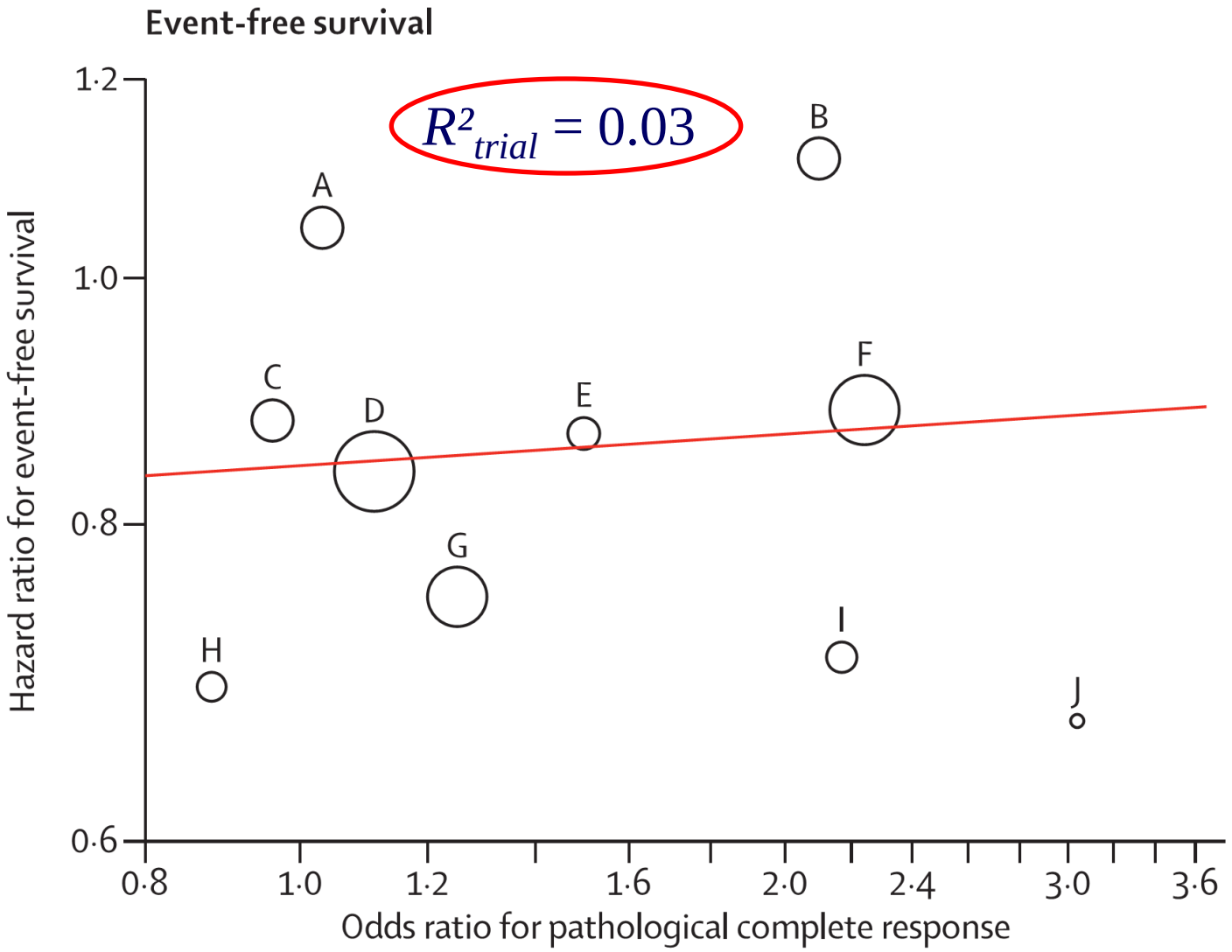
Strong “individual-level association”



Number at risk							
Pathological complete response	2131	1513	583	337	124	35	2
No pathological complete response	9824	6169	2674	1523	525	165	1

Ref: Cortazar et al, Lancet, February 2014.

No “trial-level association”

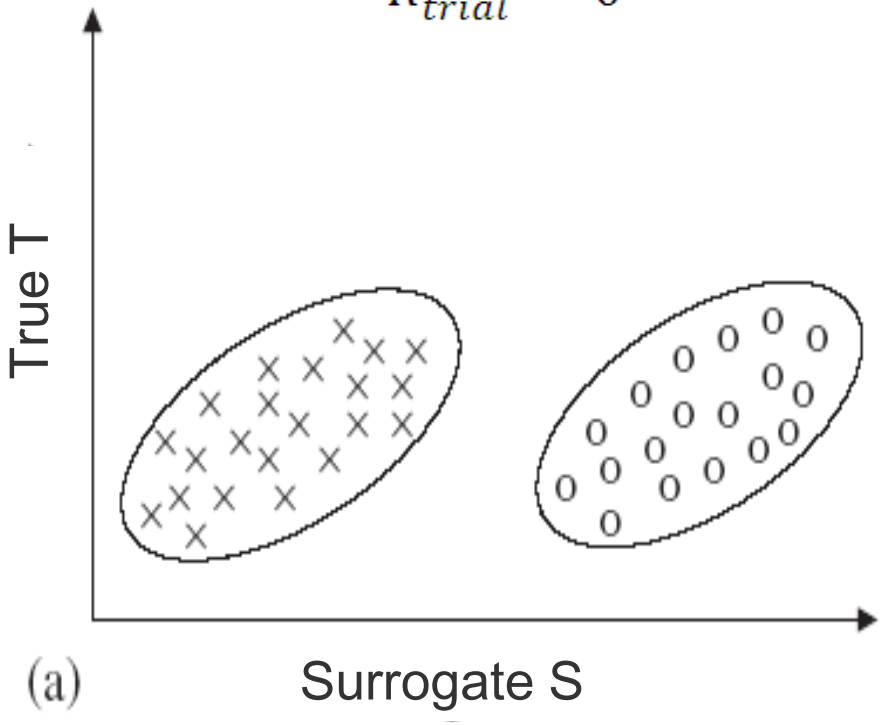


Ref: Cortazar et al, Lancet, February 2014.

Individual-level *vs.* trial-level association

$R^2_{individual} > 0$

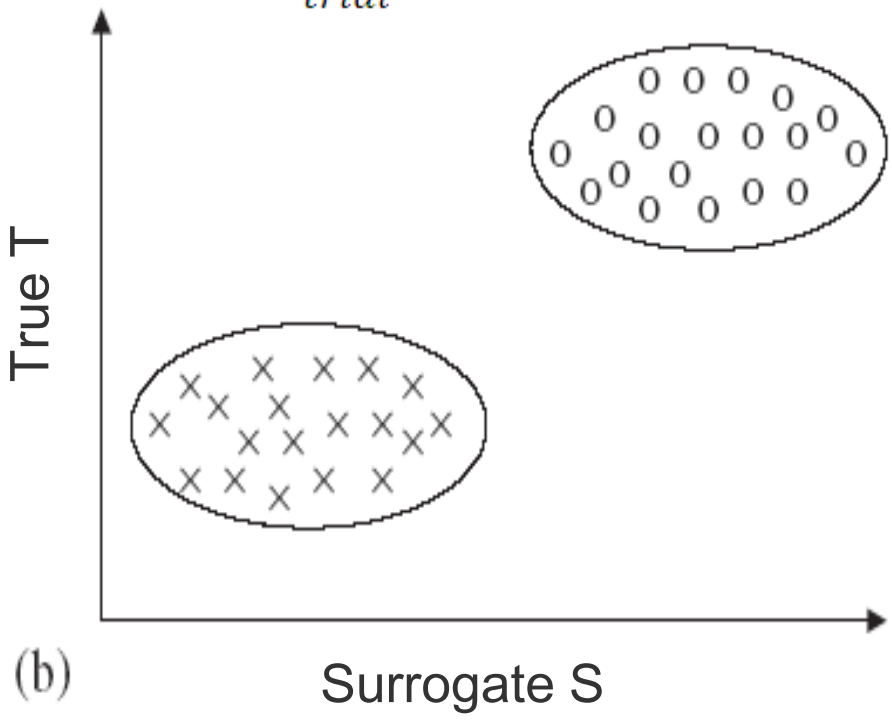
$R^2_{trial} = 0$



S correlates with T
(regardless of treatment)

$R^2_{individual} = 0$

$R^2_{trial} > 0$



Effect on S correlates with
effect on T

Ref: Korn, Albert & McShane, Statist Med 2005;24:163

Individual-level *vs.* trial-level association

"A correlate does not a surrogate make" (Fleming and DeMets 1996)

A change in the surrogate must correlate with a change in the true endpoint

In the context of randomized trials, changes are measured through treatment effects

- in individual patients (requires causal inference)
- in groups of patients (requires meta-analysis)

Ref: Burzykowski , Molenberghs and Buyse, The Evaluation of Surrogate Endpoints, Springer, New York, 2005

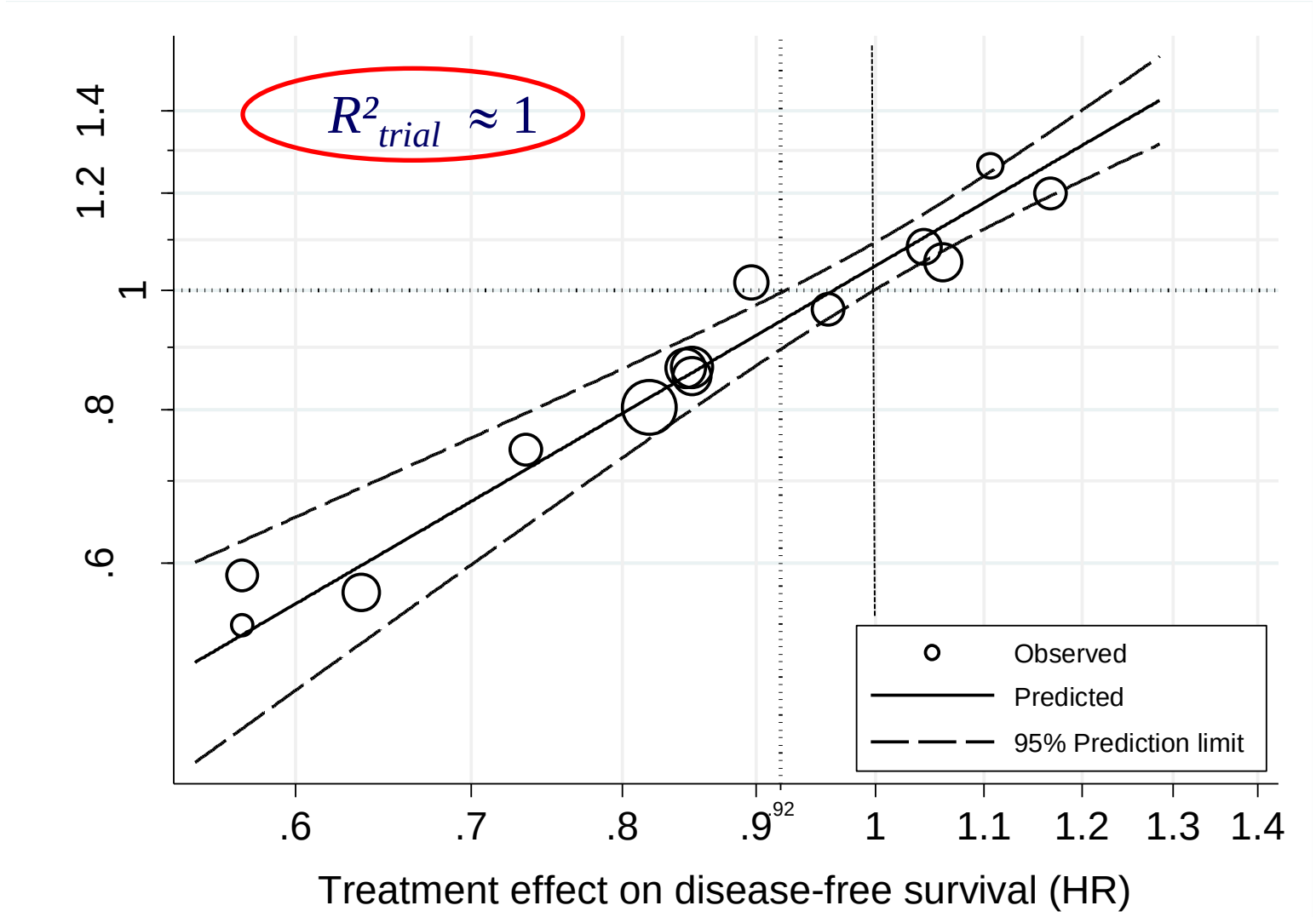
Gastric Cancer (GC): Is DFS a surrogate for OS in localized GC? Is PFS a surrogate for OS in advanced GC?

- Localized gastric cancer:
  - 14 randomized trials
  - Patient-level data (treatment/DFS/OS) on 3,288 pts
  - 5 validation trials (2 with patient-level data)

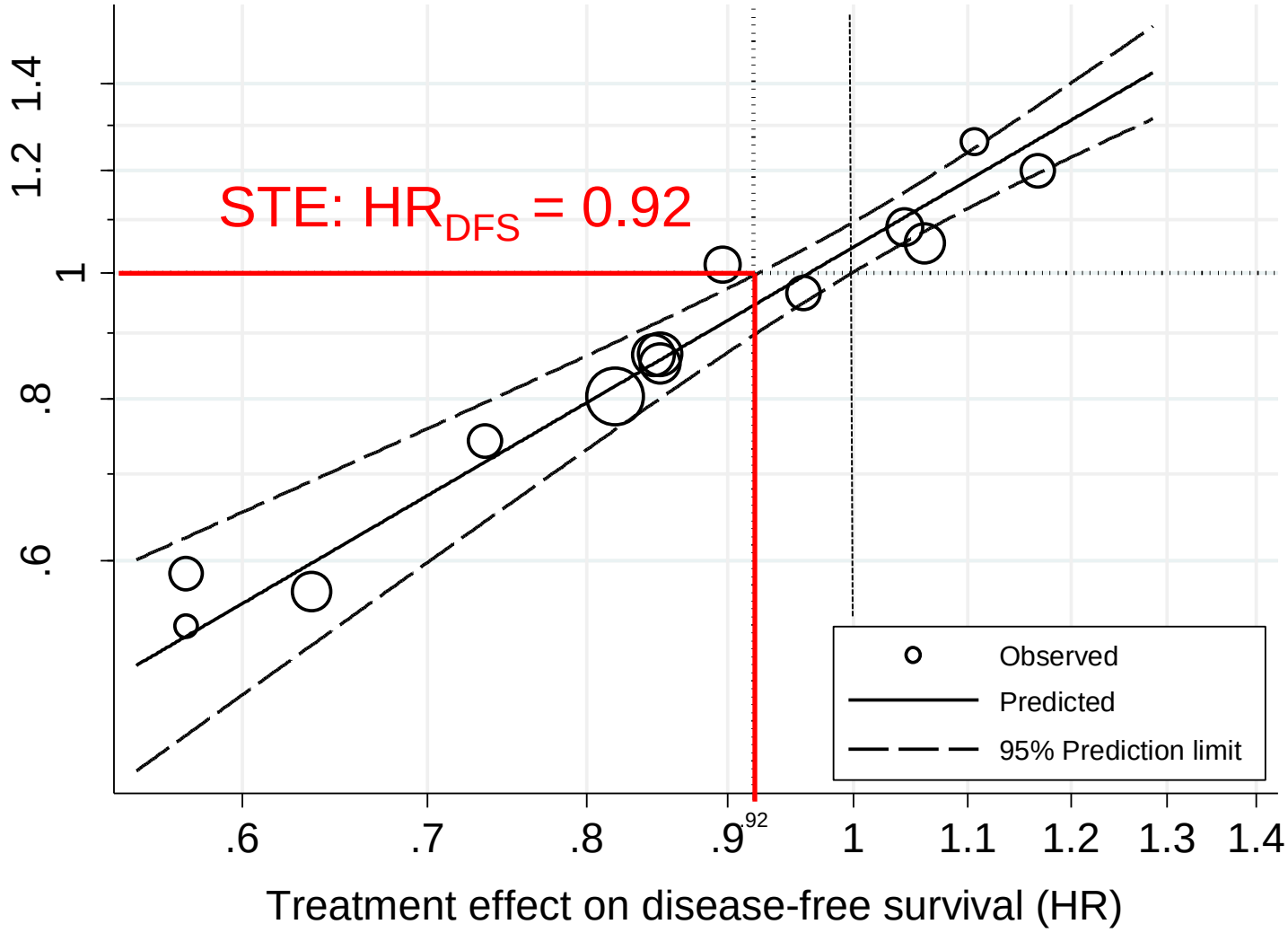
- Advanced gastric cancer:
  - 20 randomized trials
  - Patient-level data (treatment/PFS/OS) on 4,069 pts
  - 12 validation trials with summary data

Ref: Oba et al, JNCI October 2013; Paoletti et al, JNCI October 2013.

Localized Gastric Cancer: Trial-level association



Localized Gastric Cancer: Surrogate threshold effect (STE)



Localized Gastric Cancer: Independent validation trials

STE: $HR_{DFS} = 0.92$

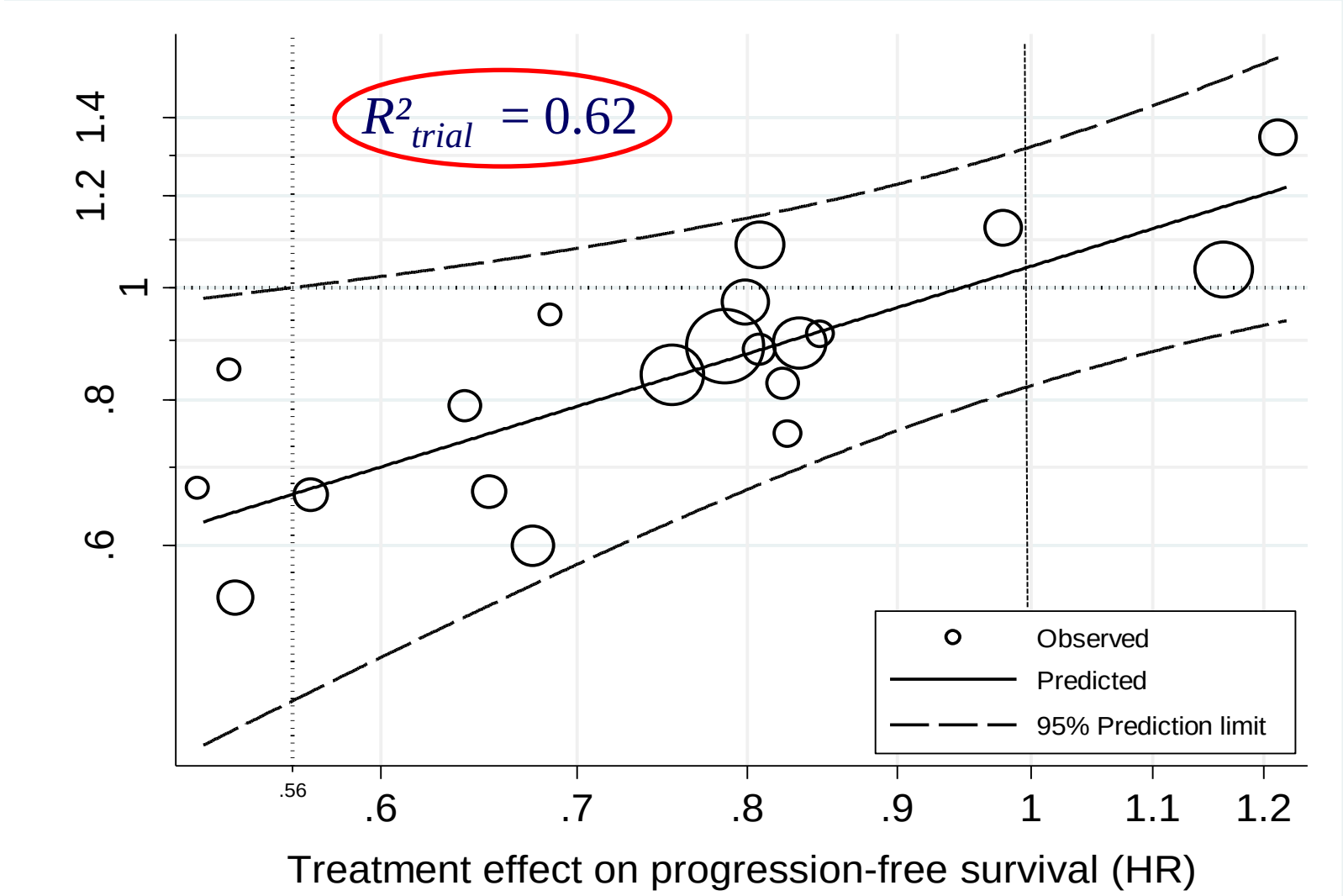
Trial	Type of data	Observed HR_{DFS} (95%CI)	Predicted HR_{OS} (95% limits)
Cirera et al.	Published	0.55 (0.36,0.85)	0.50 (0.28, 0.87)
Sakuramoto et al.	IPD	0.65 (0.54,0.79)	0.61 (0.47, 0.81)
MacDonald et al.	IPD	0.66 (0.53,0.82)	0.63 (0.46, 0.84)
DeVita et al.	Published	0.88 (0.66,1.17)	0.89 (0.62, 1.28)
Di Constanzo et al.	Published	0.92 (0.66,1.27)	0.94 (0.63, 1.42)

Localized Gastric Cancer: Independent validation trials

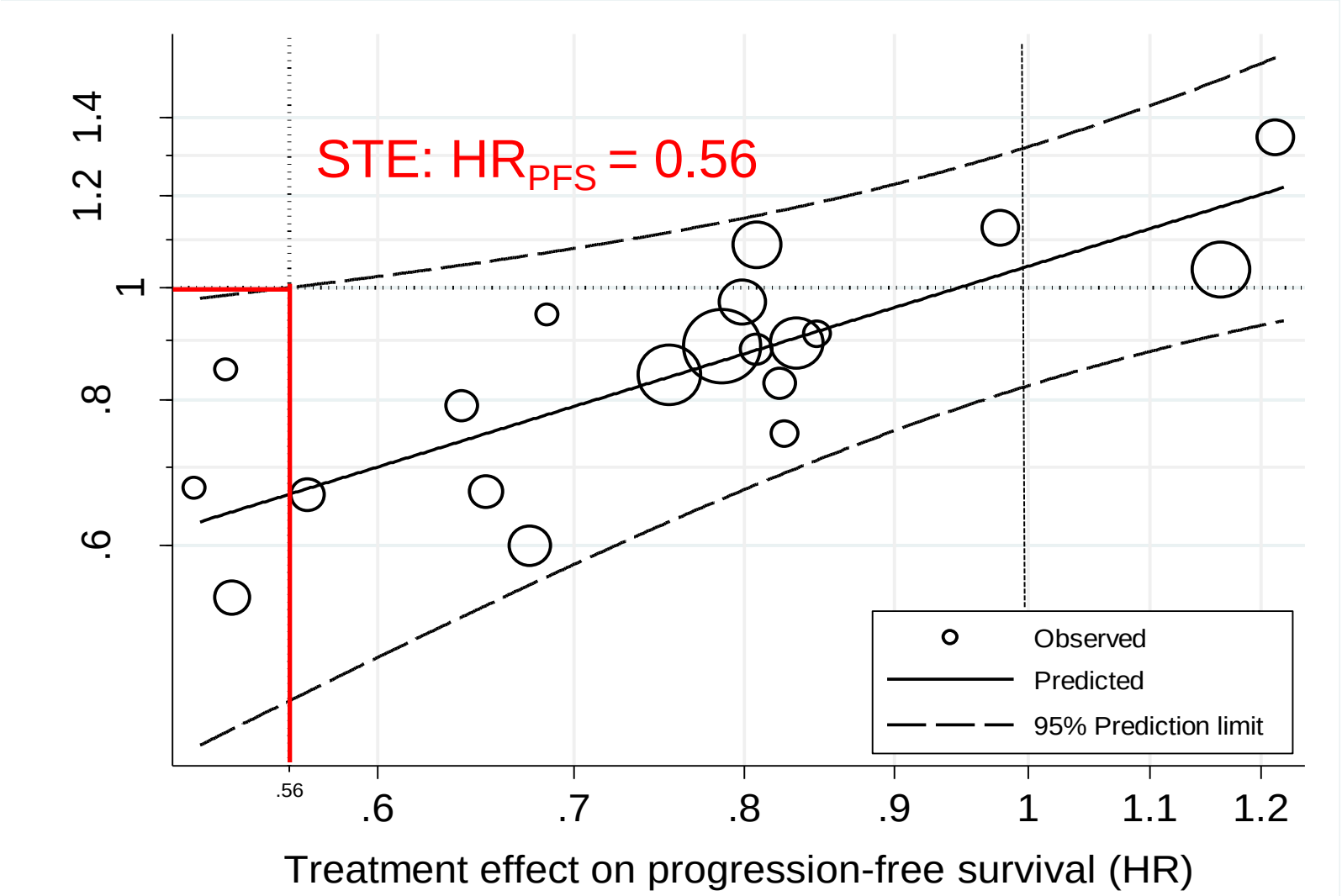
STE: $HR_{DFS} = 0.92$

Trial	Type of data	Observed HR_{DFS} (95%CI)	Predicted HR_{OS} (95% limits)	Observed HR_{OS} (95%CI)
Cirera et al.	Published	0.55 (0.36,0.85)	0.50 (0.28, 0.87)	0.60 (0.39,0.93)
Sakuramoto et al.	IPD	0.65 (0.54,0.79)	0.61 (0.47, 0.81)	0.67 (0.54,0.83)
MacDonald et al.	IPD	0.66 (0.53,0.82)	0.63 (0.46, 0.84)	0.75 (0.61,0.92)
DeVita et al.	Published	0.88 (0.66,1.17)	0.89 (0.62, 1.28)	0.91 (0.69,1.21)
Di Constanzo et al.	Published	0.92 (0.66,1.27)	0.94 (0.63, 1.42)	0.90 (0.64,1.26)

Advanced Gastric Cancer: Trial-level association



Advanced Gastric Cancer: Surrogate threshold effect (STE)



Advanced Gastric Cancer: Independent validation trials

STE: $HR_{PFS} = 0.56$

Trial	Observed HR_{PFS} (95% CI)	Predicted HR_{OS} (95% limits)
Jeung et al.	0.63 (0.38, 1.05)	0.73 (0.46, 1.04)
Albatran et al	0.67 (0.43, 1.04)	0.76 (0.53, 1.07)
Bang et al (TOGA)	0.71 (0.59, 0.85)	0.80 (0.58, 1.09)
Ohtsu et al. (avastin)	0.80 (0.68, 0.93)	0.88 (0.76, 1.14)
Kang et al.	0.80 (0.63, 1.03)	0.88 (0.76, 1.14)
Park et al.	0.86 (0.54, 1.37)	0.93 (0.71, 1.18)
Cunningham et al (a)	0.92 (0.81, 1.05)	0.98 (0.77, 1.22)
Cunningham et al. (b)*	0.92 (0.80, 1.04)	0.98 (0.77, 1.22)
Ross et al.	0.95 (0.80, 1.08)	1.00 (0.79, 1.29)
Ajani et al (FLAG)	0.99 (0.86, 1.14)	1.03 (0.81, 1.31)
Rao et al.	1.13 (0.63, 2.01)	1.14 (0.89, 1.46)
Moehler et al.	1.14 (0.59, 2.21)	1.15 (0.90, 1.48)

Advanced Gastric Cancer: Independent validation trials

STE: $HR_{PFS} = 0.56$

Trial	Observed HR_{PFS} (95% CI)	Predicted HR_{OS} (95% limits)	Observed HR_{OS} (95% CI)
Jeung et al.	0.63 (0.38, 1.05)	0.73 (0.46, 1.04)	0.56 (0.35, 0.88)
Albatran et al	0.67 (0.43, 1.04)	0.76 (0.53, 1.07)	0.82 (0.47 ,1.45)
Bang et al (TOGA)	0.71 (0.59, 0.85)	0.80 (0.58, 1.09)	0.74 (0.60, 0.91)
Ohtsu et al. (avastin)	0.80 (0.68, 0.93)	0.88 (0.76, 1.14)	0.87 (0.73, 1.03)
Kang et al.	0.80 (0.63, 1.03)	0.88 (0.76, 1.14)	0.85 (0.64, 1.13)
Park et al.	0.86 (0.54, 1.37)	0.93 (0.71, 1.18)	0.96 (0.60, 1.52)
Cunningham et al (a)	0.92 (0.81, 1.05)	0.98 (0.77, 1.22)	0.86 (0.80, 0.99)
Cunningham et al. (b)*	0.92 (0.80, 1.04)	0.98 (0.77, 1.22)	0.92 (0.80, 1.10)
Ross et al.	0.95 (0.80, 1.08)	1.00 (0.79, 1.29)	0.91 (0.76, 1.04)
Ajani et al (FLAG)	0.99 (0.86, 1.14)	1.03 (0.81, 1.31)	0.92 (0.80, 1.05)
Rao et al.	1.13 (0.63, 2.01)	1.14 (0.89, 1.46)	1.02 (0.61, 1.70)
Moehler et al.	1.14 (0.59, 2.21)	1.15 (0.90, 1.48)	0.77 (0.51, 1.17)

Tentative conclusions (1 of 3)

- Individual-level association (= “correlation”) is useful for patient management
- Trial-level association is required to replace clinical endpoint by putative surrogate

Tentative conclusions (2 of 3)

- In resectable breast cancer, is pCR “reasonably likely to predict long-term clinical benefit”? Statistical evidence is not compelling, is biological evidence alone compelling?
- In localized gastric cancer, there is convincing statistical evidence that DFS can be used as a surrogate for OS
- In advanced gastric cancer, there is evidence that only major effects on PFS might predict effects on OS – hence, PFS can not be used as a surrogate for OS

Tentative conclusions (3 of 3)

Caveats for meta-analytical approach:

- Large numbers (trials / patients) are needed
- Computational challenges of fitting complex models
- Historical data may be unreliable / inadequate
  - Patient populations may have changed
  - Endpoint assessment may have changed
  - Treatments may have changed
- New treatments may have different mode of action