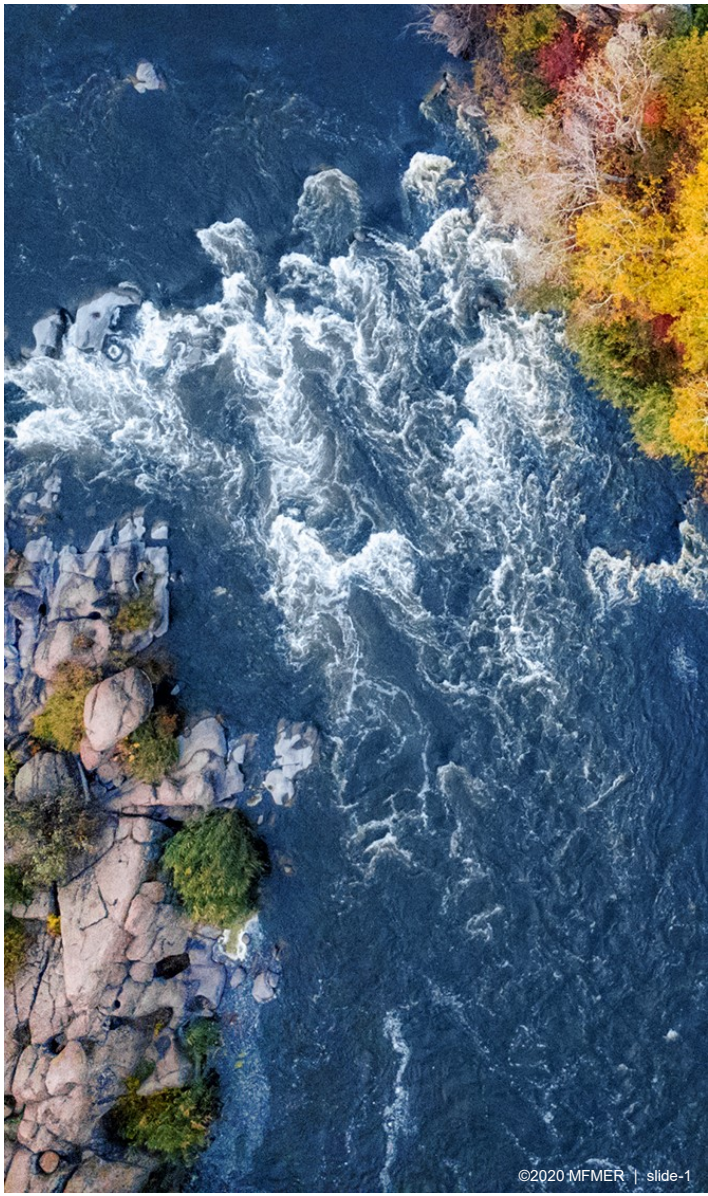
MAYO CLINIC

Surrogate Endpoints in Oncology Clinical Trials: Are We There Yet?

From a Practical Perspective

Qian Shi, PhD., Professor of Biostatistics and Oncology

ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop
September 24, 2020



1

What We Know

Surrogate Endpoints

Make drug development faster, more efficient, safer, and more feasible

Un-validated surrogate endpoints can produce erroneous conclusions regarding the efficacy or safety of therapies

Table 1. Examples of putative surrogate endpoint failures

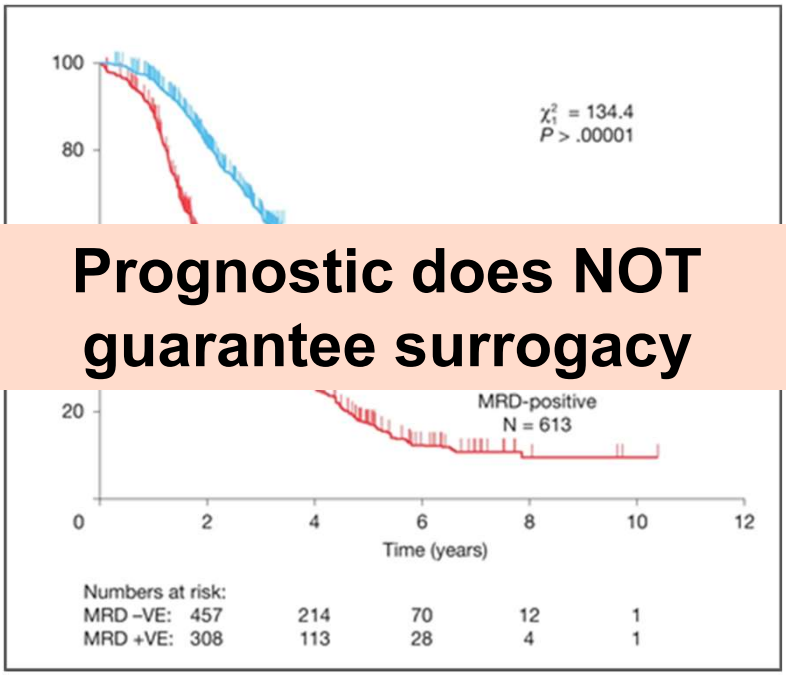
Disease	Treatment	Effects on		Trials or analyses
		Surrogate endpoint	Clinical endpoint	
Postmyocardial infarction	Anti-arrhythmic agents	Reduced ventricular arrhythmia	Increased sudden death	CAST[41]
Atrial fibrillation	Quinidine	Maintained sinus rhythm at 1 year	Increased mortality	Meta-analysis[2]
Congestive heart failure	Milrinone/Flosequinan/ Epoprostenol	Improved cardiac output/ increased exercise tolerance	Increased mortality	PROMISE[87] PROFILE[88] FIRST[89]
Heart disease in postmenopausal women	Hormone replacement therapy	Favorable effect on serum lipoprotein level	Increased coronary heart disease/myocardial infarction	HERS[90] WHIT[91] PEPI[92]
Heart disease	Cholesterol-lowering agents	Lowering cholesterol level	Increased mortality	WHO[93] Gordon meta-analysis[94]
Osteoporosis	Sodium fluoride	Increased bone mineral density	Increased fracture incidences	[95]
HIV	Zidovudine	Lowering CD4+ cell counts	Failed to reduce opportunistic infection	British-French Concord Trial[96]
Normotensive patients	Management of glaucoma	Lowering intraocular pressure	No effect of treatment on long-term visual field loss	[8]

CAST, the Cardiac Arrhythmia Suppression Trial; FIRST, the Flolan International Randomized Survival Trial; HERS, the Heart and Estrogen/progestin Replacement Study; PEPI, the Postmenopausal Estrogen/Progestin Intervention Trial; PROFILE, the Prospective Flosequinan Longevity Evaluation Trial; PROMISE, the Prospective Milrinone Survival Evaluation Trial; WHIT, the Women's Health Initiative Randomized Controlled Trial

Shi and Sargent, Int J Clin Oncol. 2009

Individual- vs. Trial-level Surrogacy

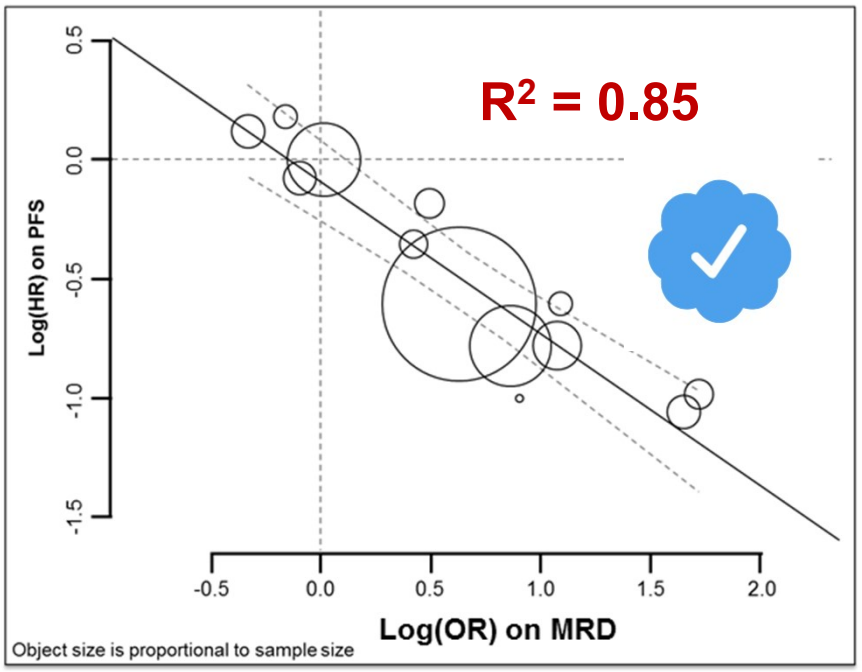
Prognostic association



Prognostic does NOT guarantee surrogacy

Munshi N et al., JAMA Oncol, 2017

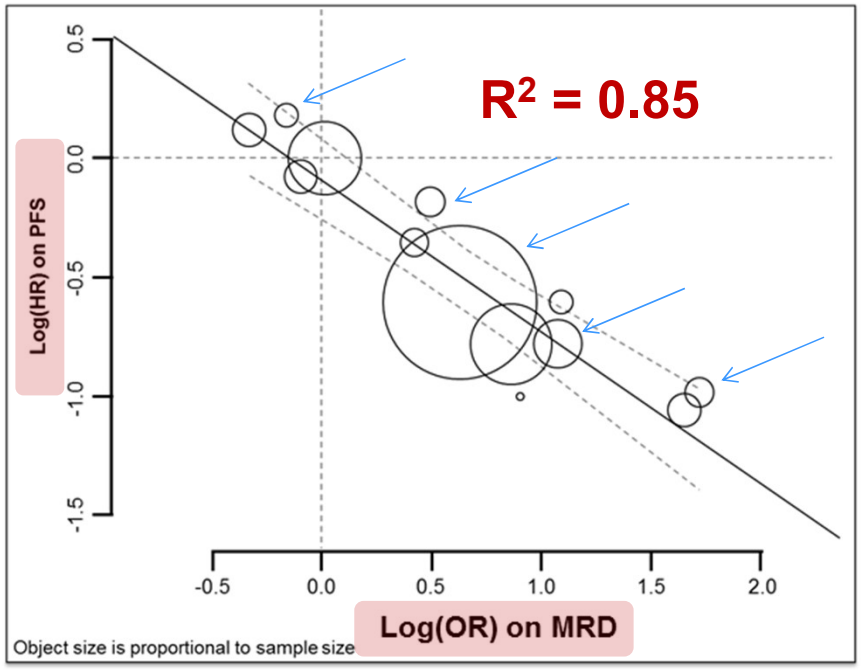
Treatment effect correlation



 |

Trial-level Surrogacy – Primary Validation

Treatment effect correlation



Context of Use:

A substitute for "true" clinical endpoint in future trials

$^{*}R^2_{WLS}$ and $^{**}R^2_{Copula}$

*Sargent et al. JCO, 2005/2007; **Buyse et al. Biostatistics, 2000

2

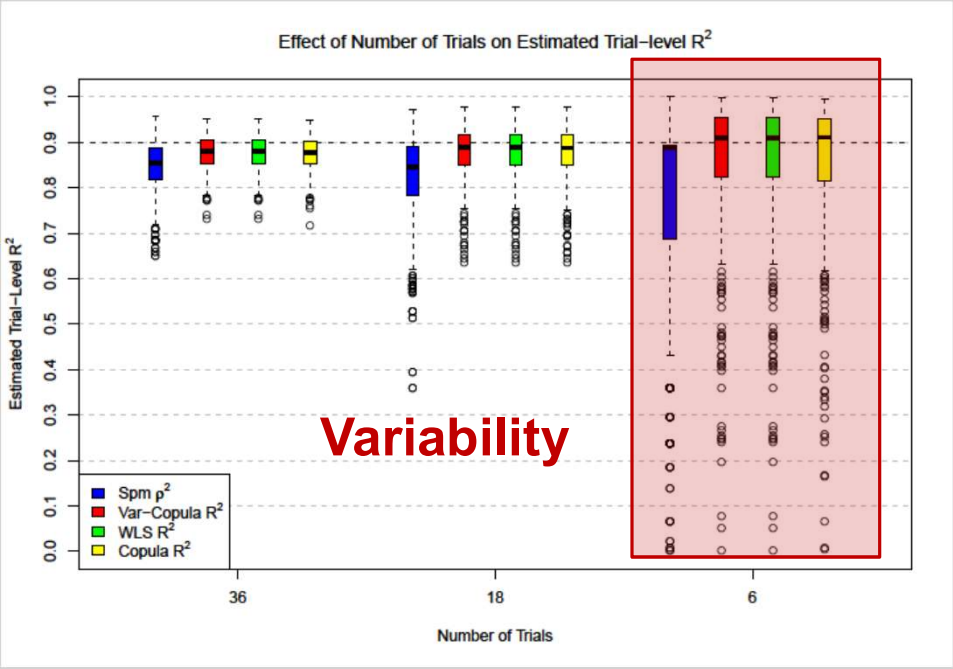
What We Have Learned

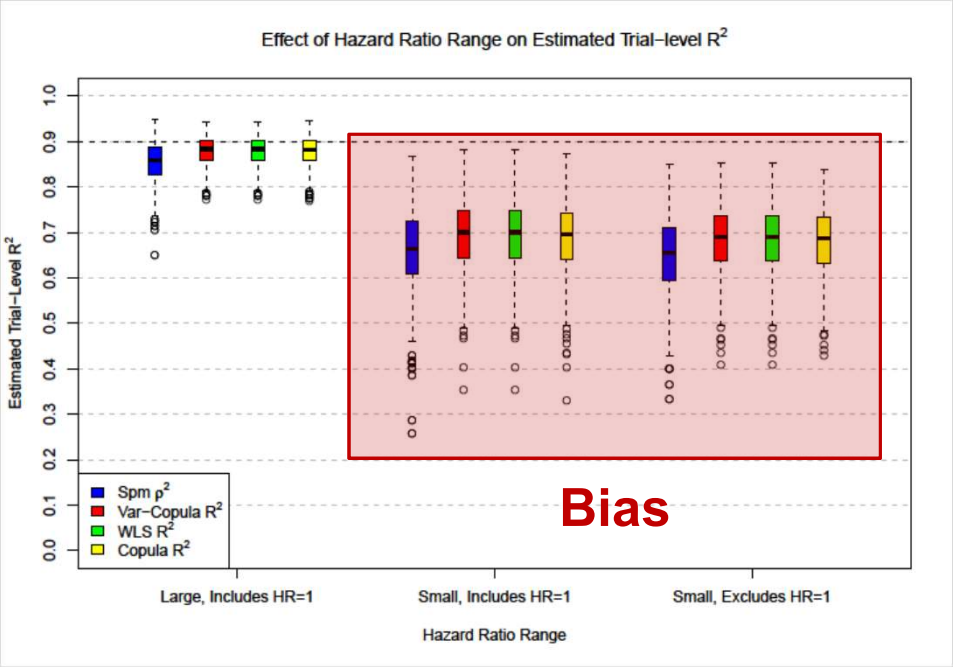
R^2_{WLS} and R^2_{Copula}

most commonly used estimates

2 Critical trial-level factors that impact the estimation accuracy and variability:

- Number of trials
- Range of the treatment effect sizes



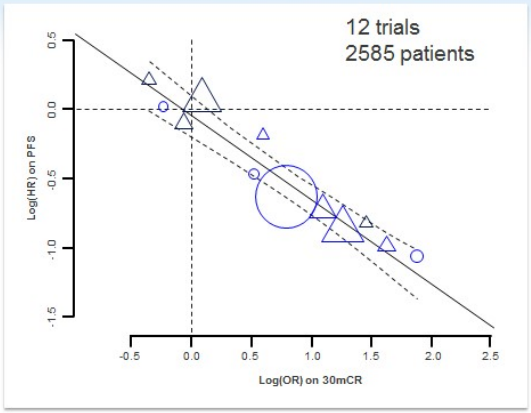


Shi et al. CSDA, 2011

Subgroups in All Comers Trials

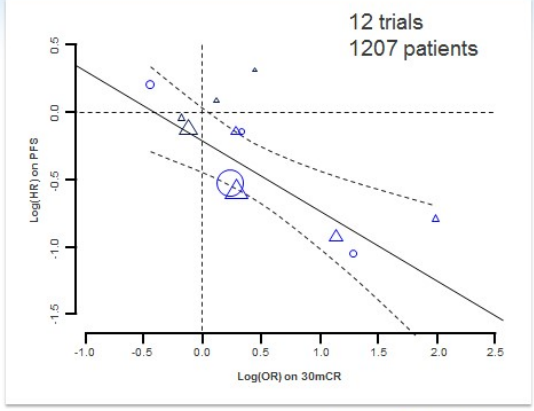
- Previously untreated follicular lymphoma
- Surrogate endpoint: CR30
- True clinical endpoint: PFS

Results: CR30 in Ann Arbor Stage IV Patients



Trial-level Surrogacy	
R^2_{WLS} (95% CI)	R^2_{Copula} (95% CI)
0.92 (0.85, 0.97)	0.94 (0.87, 1.00)
Individual-level GOR (95% CI)	
11.44 (9.30, 13.57)	

Results: CR30 in Ann Arbor Stage I/II/III Patients



Trial-level Surrogacy	
R^2_{WLS} (95% CI)	R^2_{Copula} (95% CI)
0.59 (0.25, 0.87)	0.59 (0.24, 0.95)
Individual-level GOR (95% CI)	
12.57 (9.04, 16.10)	

Shi et al, JCO 2017

Treatment Mechanisms

- Previously untreated advance colorectal cancer
- Surrogate endpoint: PFS
- True clinical endpoint: OS

Pre-biologics Era

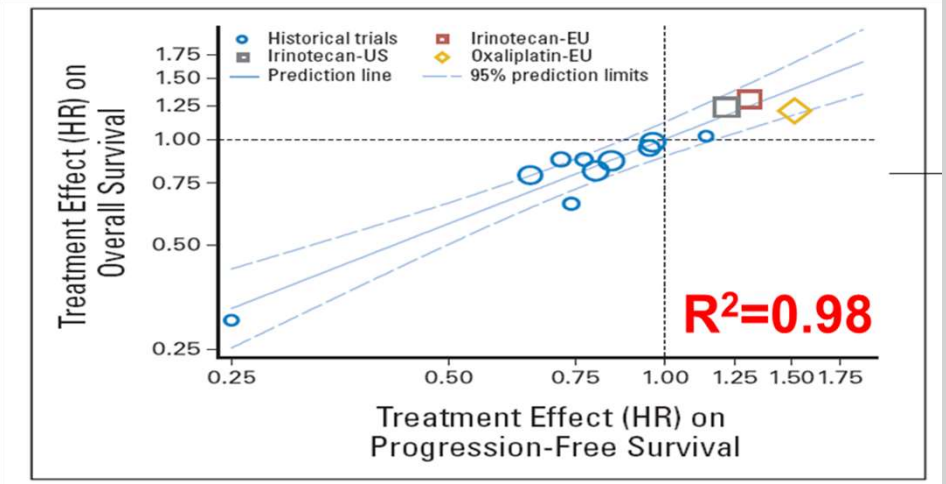
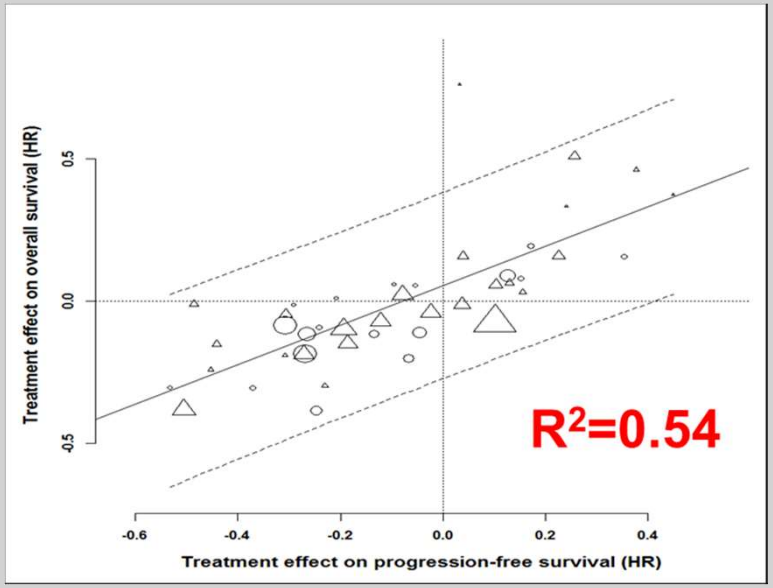


Fig 4. Correlation between treatment effects on progression-free and on overall survival in historical trials (circles), in irinotecan trials (squares), and in oxaliplatin trial (diamond). A logarithmic scale is used for both axes. Symbol size is proportional to the number of patients. HR, hazard ratio; EU, European Union.

Buyse et al, JCO 2007

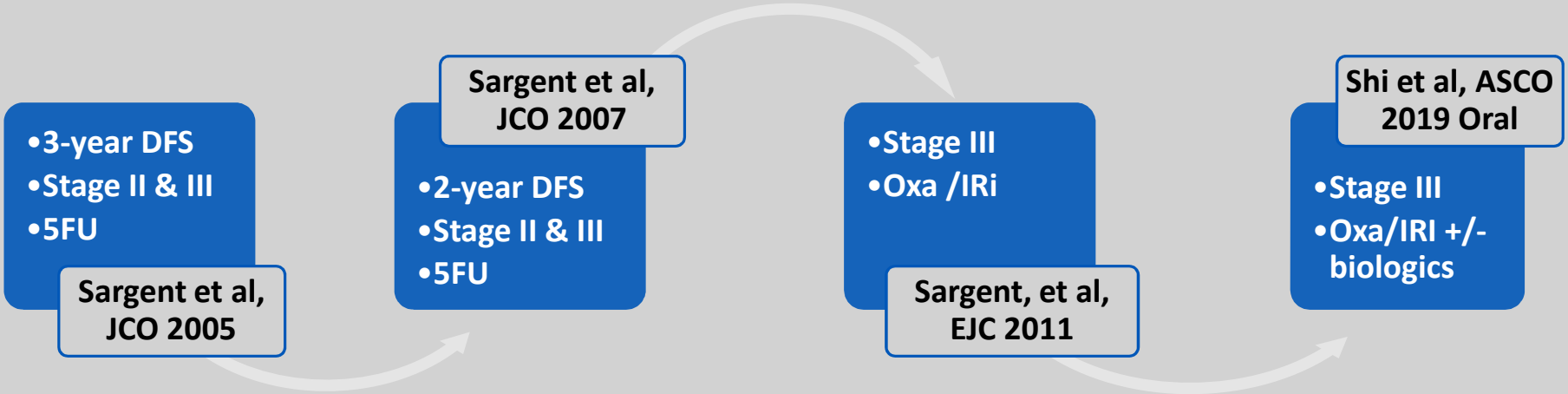
Post-biologics Era



Shi et al, JCO 2014

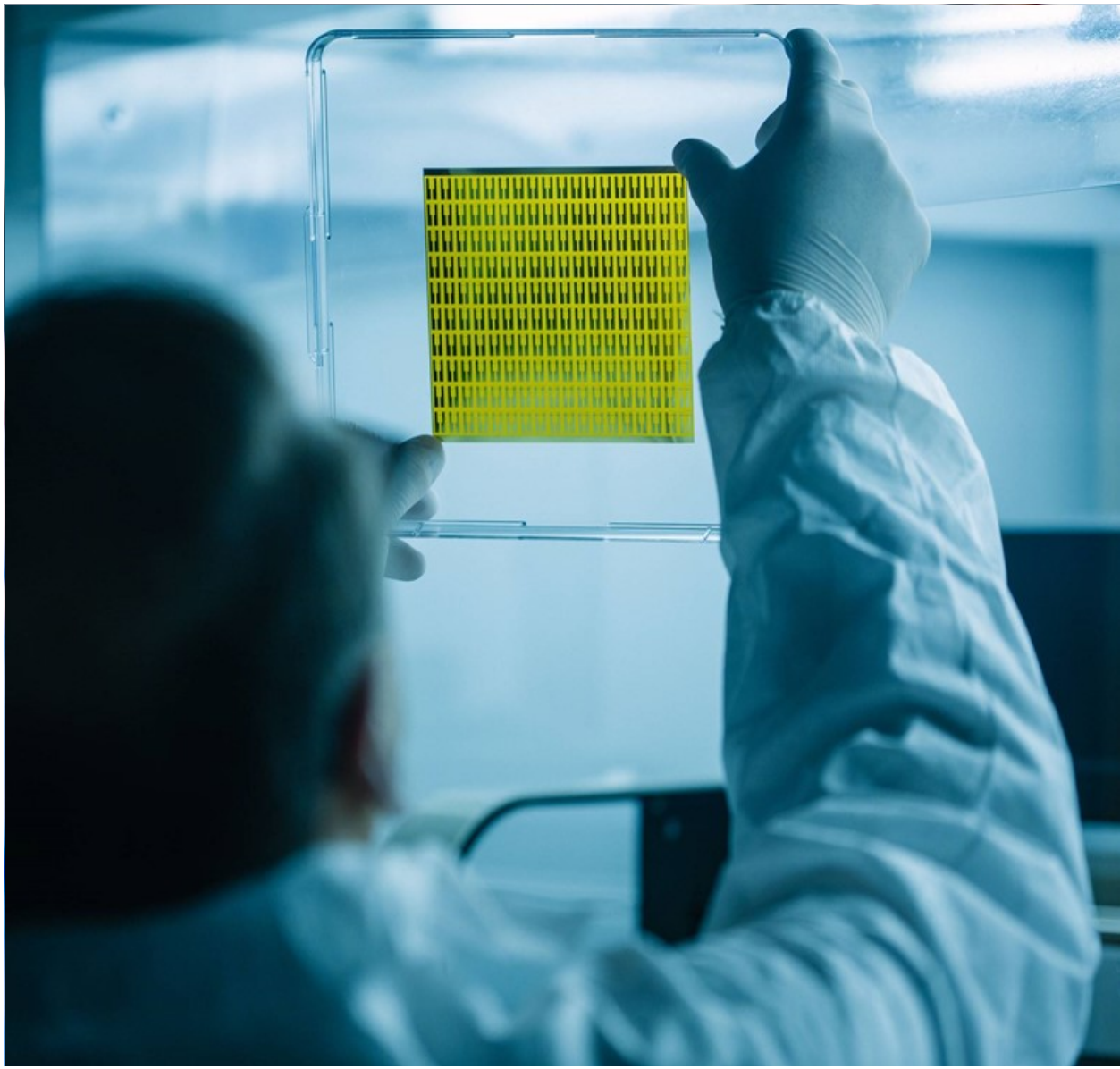
Changing Standard of Care

- Early stage colon caner after curative resection
- Surrogate endpoint: DFS
- True clinical endpoint: OS

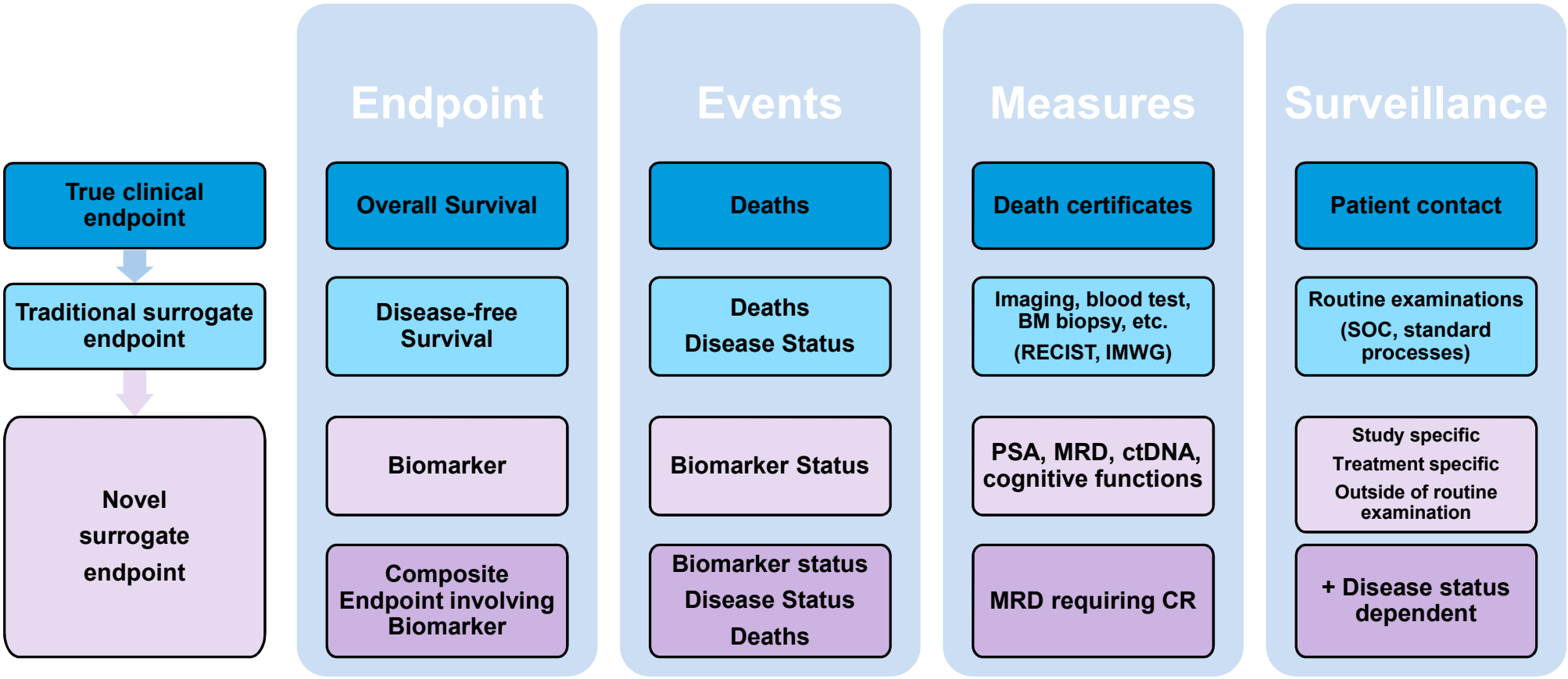


3

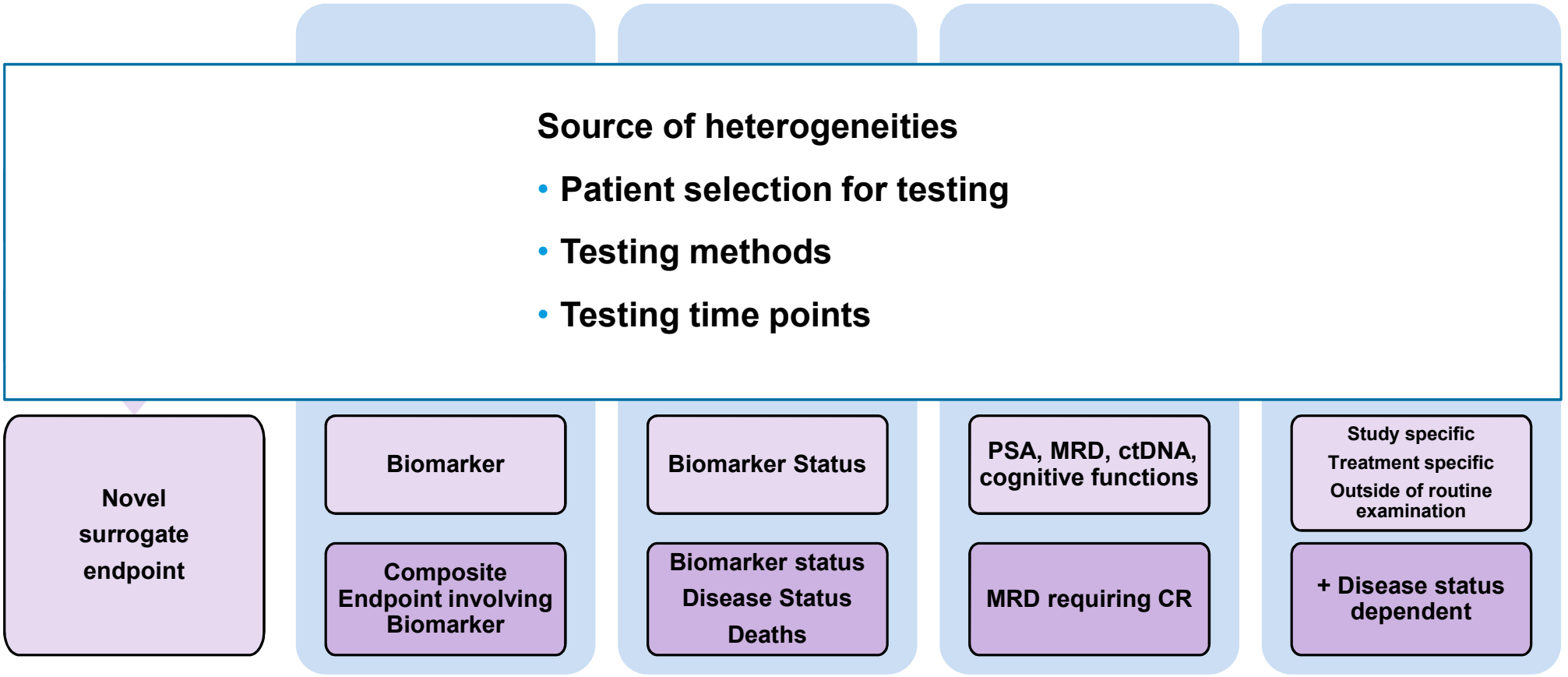
Novel Biomarker – New Challenges



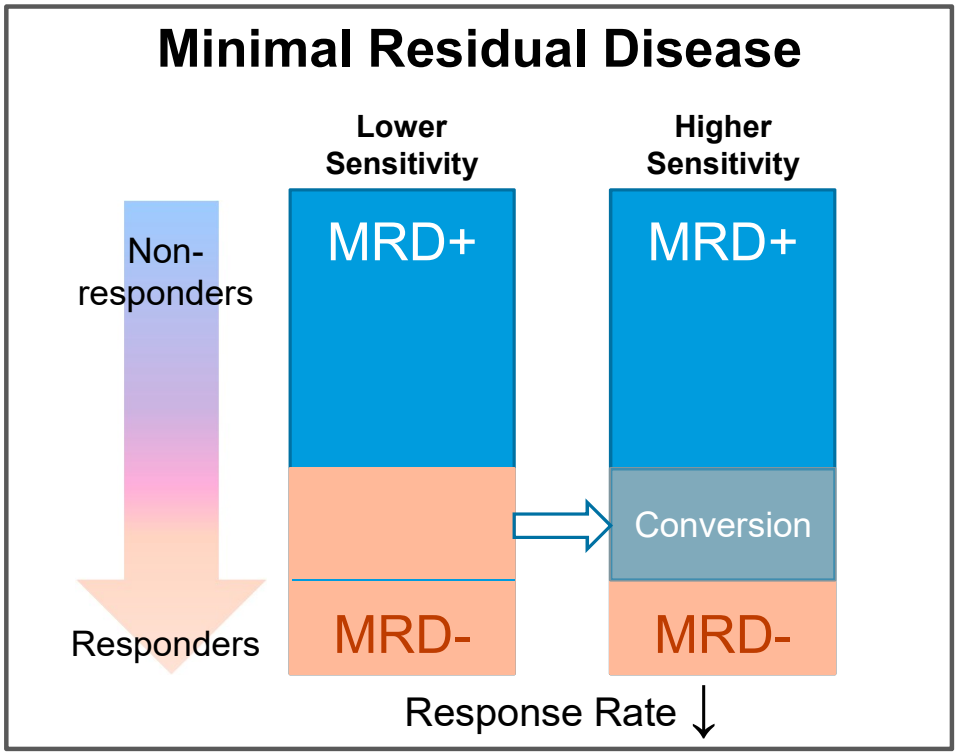
Surrogate Endpoints in Oncology Trials



Surrogate Endpoints in Oncology Trials



Sensitivity Level – Deeper Responses Defined by Biomarkers



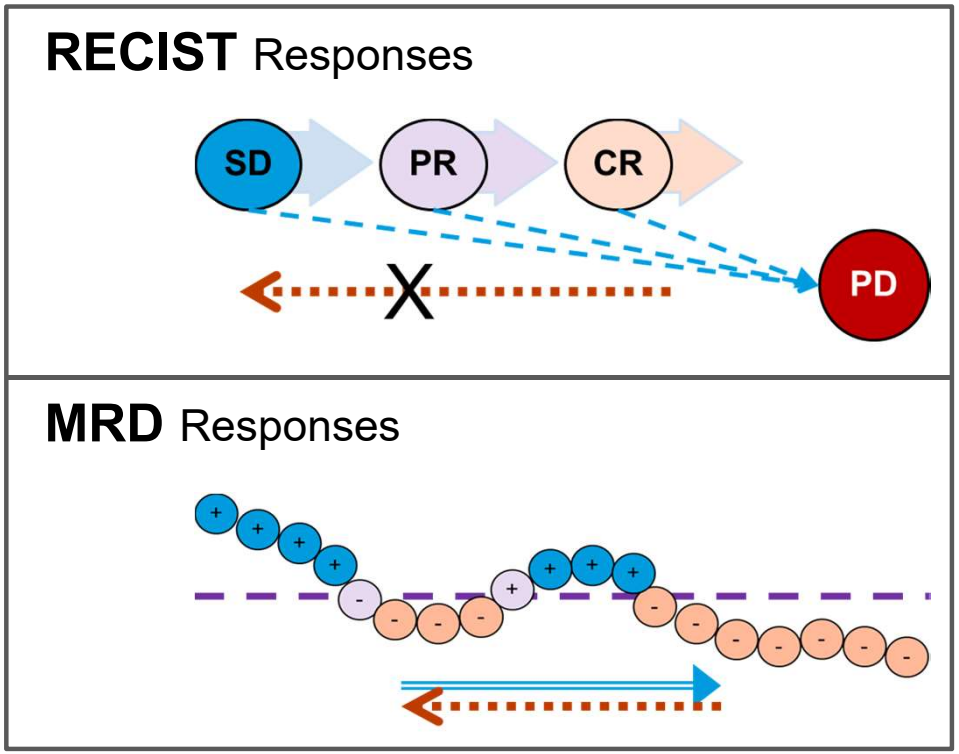
Clinical Utility

- More accurately defining cure

Statistical Challenges: Conversion rate differ per

- Testing
  - Assay
  - Laboratory
  - Specimen
- Disease characteristics
- **Regimen**

Dynamics of Responses Over Time

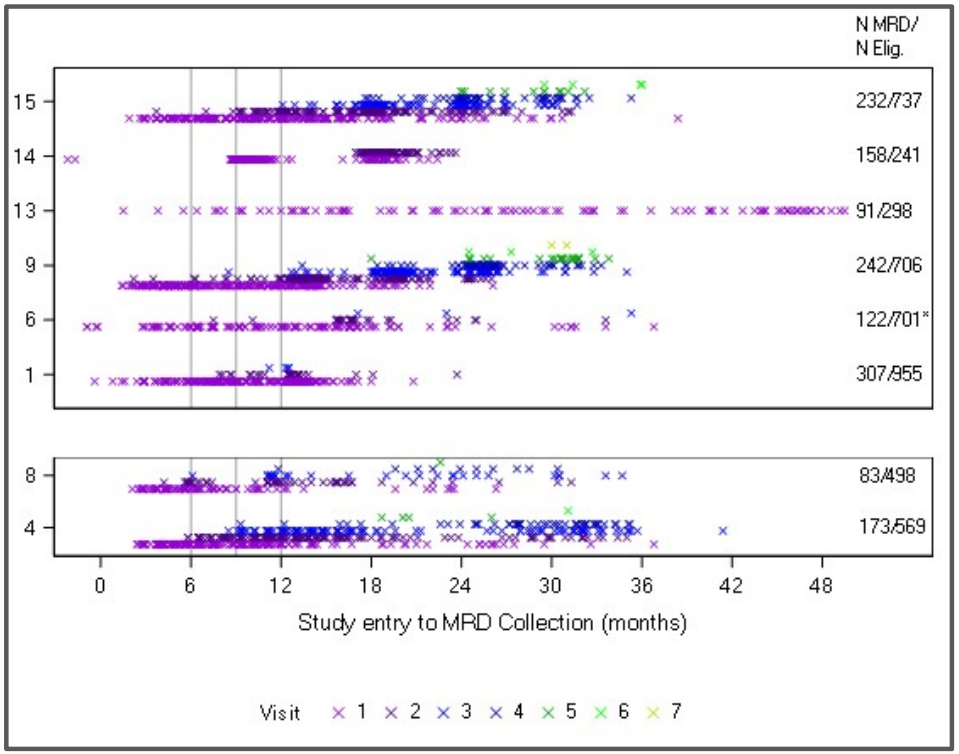


Response Reversing?

Directing Treatment?

Response Lasts?

Biomarker Testing – Protocol Planning to Data Collection



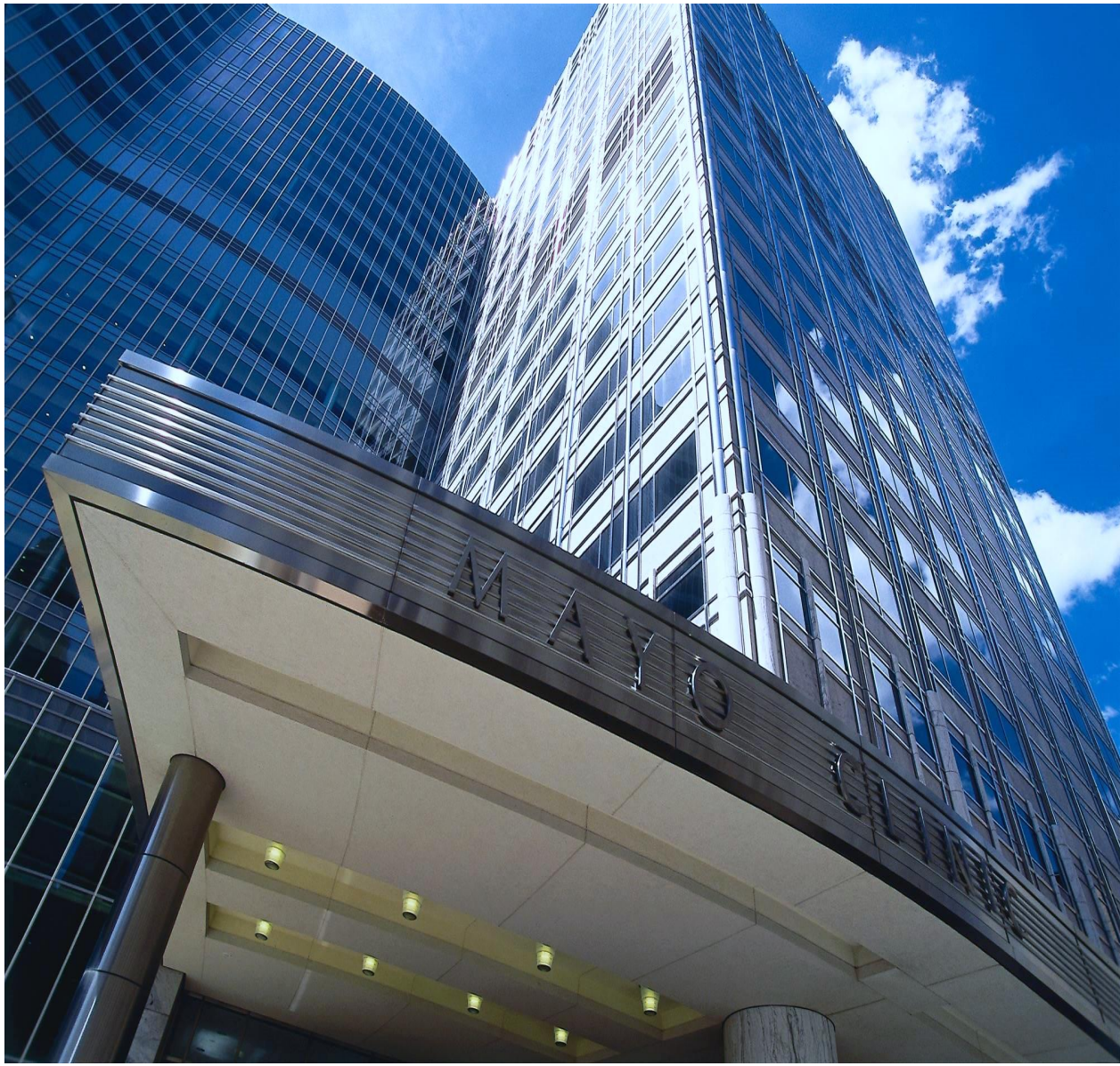
Number of Testing
- Single vs. many
- Varying across patients

Timing of Testing
- Treatment dependence
- Response dependence

Patient Selection
- Response dependence
- Randomization

4

Proposals



General Criteria and Principles of Developing Candidate Surrogate Endpoints

- Clinically meaningful:
  - Capture the targeted treatment effect
  - Correlates the treatment effect on true endpoint
- Statistically meaningful:
  - Programmable derivations based on necessary data elements
  - Consistency across all trials
  - Consistency from evaluation to future use
- Practically useful in further trials:
  - Provides utility advantages compared to the true endpoints
  - Data elements needed can be routinely collected in the future trials

IPD from RCTs are Critical for Surrogacy Evaluations

- Individual patient data (IPD) from RCTs
  - Unbiased estimates of treatment effects within trials
  - Consistent derivations across trials
  - Evaluate (and develop) surrogate endpoints with no published treatment effect data
- RCT selections: Inclusive
  - Prespecified inclusion/exclusion criteria
  - More is better (specificity)
  - Population heterogeneities (universality)
  - Treatment class and treatment sequencing (universality)

Call for Long-term and Large Scale of Collaborations

- Harmonization of testing methods
- Common testing time points
- Common data collection
- Share Data!!

QUESTIONS & ANSWERS