Design and Analysis with Use of Biomarker Classifier for Drug Development

A Regulatory Perspective of Issues

Sue-Jane Wang, Ph.D.

Associate Director, Pharmacogenomics and Adaptive Design
Office of Biostatistics, Office of Translational Sciences,
Center for Drug Evaluation and Research, U.S. FDA

Presented at University of Pennsylvania Annual Conference on
Statistical Issues in Clinical Trials: From Bench to Bedside to Community
U. Penn, PA, April 29, 2009

Outline*

- **Genomic Biomarker: a Baseline Classifier**
- **Biomarker Intended Designs/Objectives in PG A&WC Clinical Trials**
- **Prospective/Retrospective Designs**
- **Biomarker Associated Statistical Analysis Issues**
- **Adaptive Design Consideration**
- **Summary of Regulatory Issues**

* The views presented are those of the authors and not necessarily those of the U.S. FDA

Genomic Biomarker

- A measurable DNA/RNA characteristic that is objectively measured and evaluated
- Recognized as an indicator of
  - Normal biological processes
  - Pathogenic processes
  - **Pharmacologic response to a therapeutic intervention**
- Regulatory impact
  - Genomic surrogate biomarker (endpoint)
  - Genomic biomarker classifier (**single**, composite)

* ICH E-15

Definition of Treatment Effect

Genomic Status*					Scenario C	
					Control	Drug C
g– *g*+					39% 48%	49% 68%

* *g*+ or *g*– is patient's genomic status determined from a diagnostic assay

Predictive
Effect in g+ only; no effect in g-

Prognostic
Biomarker plays a role in disease response only

Prognostic-Predictive
Effect in g+ > in g-; biomarker plays a role both in disease response and in drug response

Qualitative

Quantitative

* Wang, O'Neill, Hung (2007 PS)

Prospective Designs by Study Objectives

- Single subset (well-defined)
- Composite (two types, multiplicity)
- Separate Subsets
  - Depending on the clinical context
- Interaction – a different question
- Biomarker guided
  - power and efficiency issues
- Genomic model based designs
  - describe benefit at patient level

Prospective/Retrospective

Working Definition – In completed or post-interim-analysis trial where genomic samples were collected prior to treatment initiation, whether or not full ascertainment, the genomic hypothesis is 'prospectively specified' because it is prior to diagnostic assay testing. However, the clinical outcome data without genomic information have already been (partially) collected and analyzed. The genomic data analysis might be arguably 'prospectively' performed, which is a retrospective analysis.

Convenience Genomic Sample – The genomic samples may not be obtained on all randomized subjects if some refuse (often optional consent in current practice).

* Wang et al (2006 TPJ)

Statistical Analysis Issues

- Convenience sample issues
  - Bias
- Full ascertainment of subset – design efficiency*
- Prevalence when full ascertainment of ITT
  - Efficacy (effect size vs. diagnostics: PPV, NPV ?)
  - Safety (predictive of group vs. individual - AUC)
- Efficiency with prospective/retrospective designs
  - Intended patient population vs. Intended use
- Replication

* Maitournam & Simon (2005 SIM)

Exploration

Study 1	Primary Endpoint 1		Primary Endpoint 2	
31% of ITT	Low	High	Low	High
effect estimate unadj. p-value	-5.0 0.029	-3.4 0.192	-7.9 0.0141	-6.6 0.135

Study 1	Primary Endpoint 1		Primary Endpoint 2	
ITT	Low	High	Low	High
effect estimate unadj. p-value	-3.1 0.033	-4.5 0.005	-5.1 0.034	-8.1 0.002

Patient Characteristics
(DNA vs. Non-DNA consented sample)

Intent to treat patients	Placebo	Treatment
DNA sample Death rate Dropout rate	887 20% (8%)	892 16% (10%)
Non-consented sample Death rate Dropout rate	829 40% (19%)	839 38% (16%)

Implication1: bias in the convenience sample ? Different characteristics

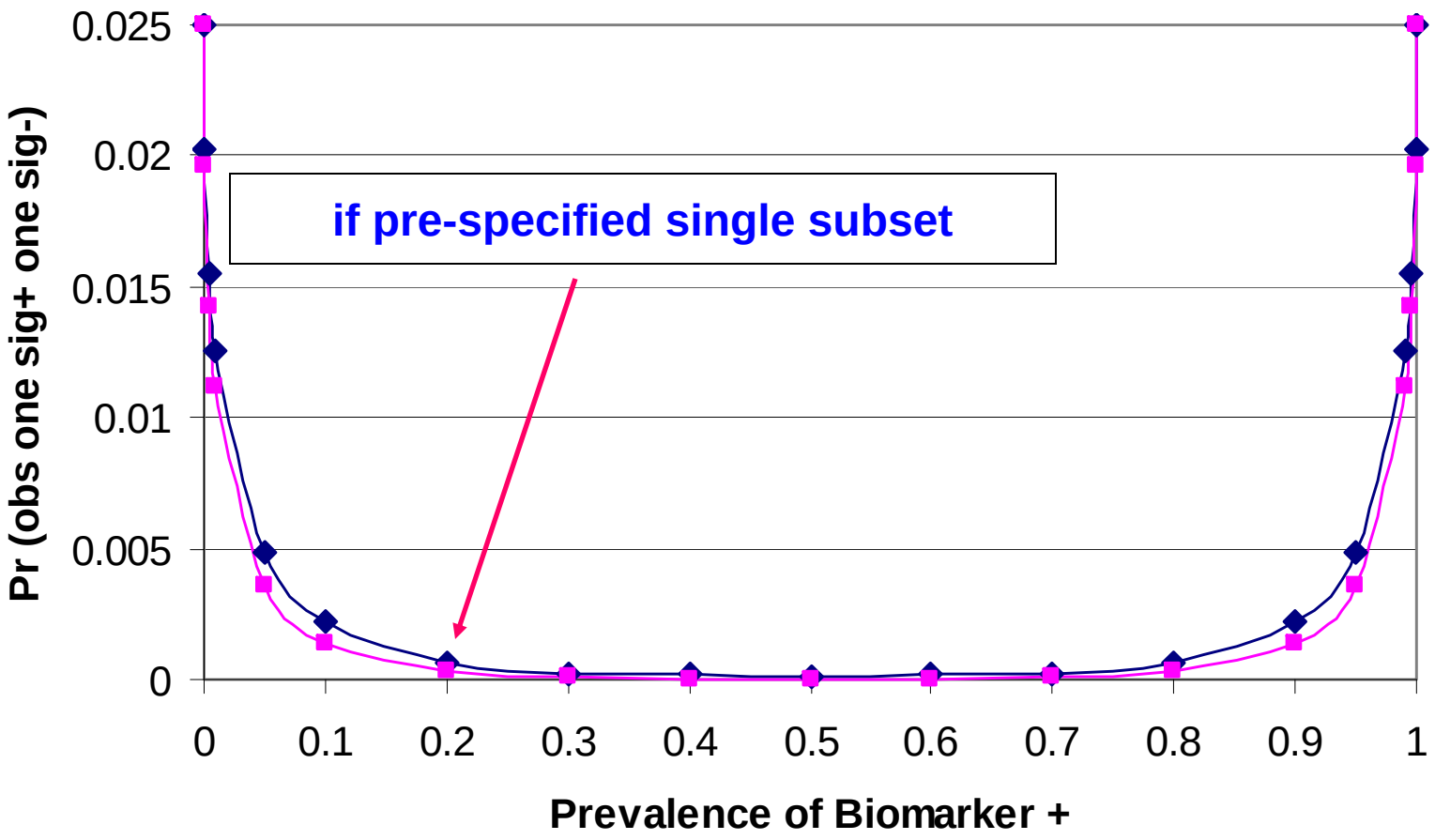
Implication2: bias in the convenience sample ? Differential compliance

Statistical Analysis Issues

- Convenience sample issues:
  - Bias
- Full ascertainment of subset – design efficiency*
- Prevalence when full ascertainment of ITT
  - Efficacy (effect size vs. diagnostics: PPV, NPV ?)
  - Safety (predictive of group vs. individual - AUC)
- Efficiency with prospective/retrospective designs
  - Intended patient population vs. Intended use
- Replication

* Maitournam & Simon (2005 SIM)

Figure 1a. Probability of One Negative and One Positive Significances



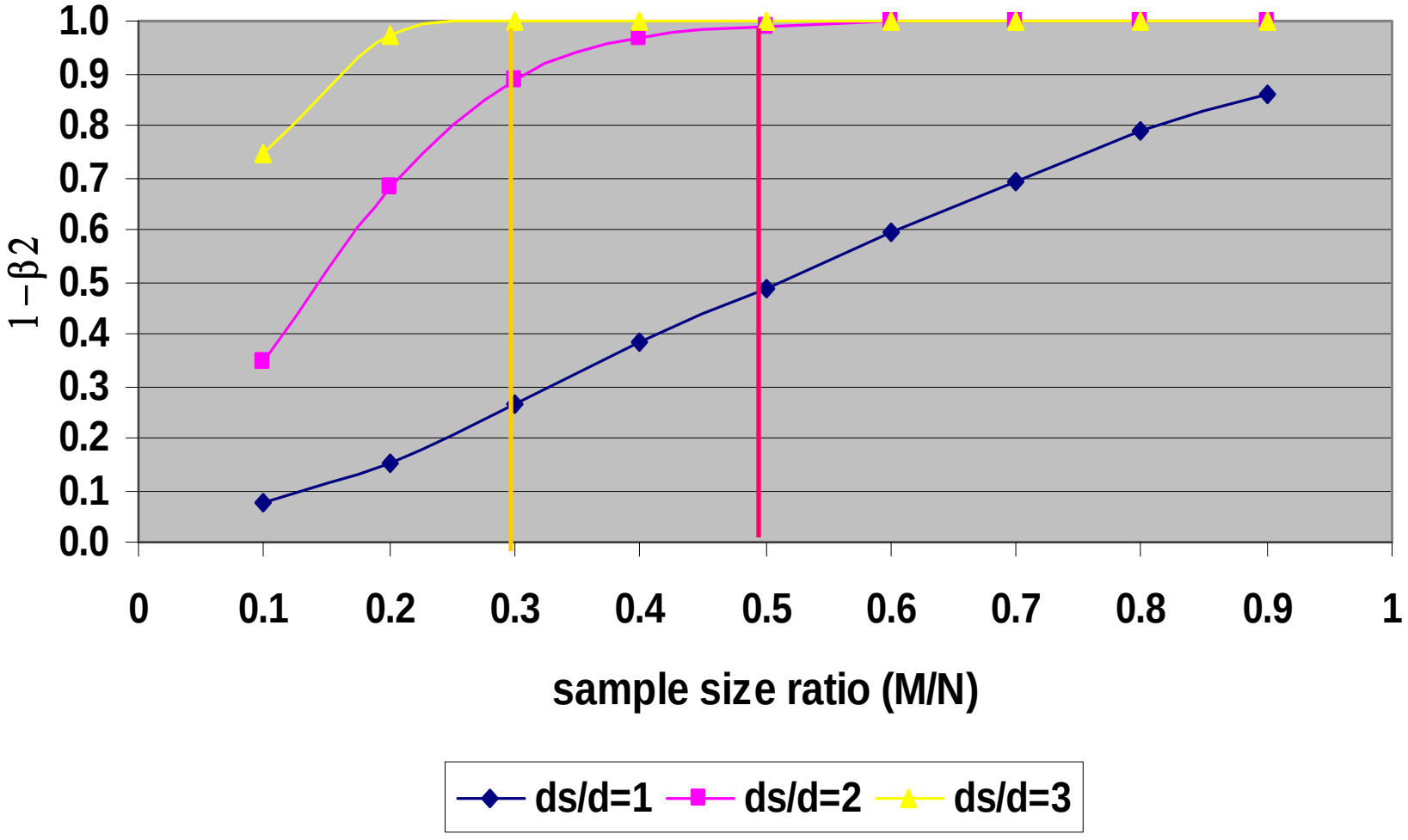
Prasugrel vs. Clopedigrol on composite (CV death, MI, Stroke)
ACS = UA/NSTEMI (primary analysis) + STEMI *

All cause mortality	T	C	OR (95%CI)	p-value
ACS n event rate	6813 9.44%	6795 11.49%	0.80 (.72, 0.90)	<0.0001
UA/NSTEMI n event rate	5044 9.30%	5030 11.23%	0.81 (0.71, 0.92)	0.0014
STEMI n event rate	1769 9.84%	1765 12.24%	0.78 (0.63, 0.97)	0.0227

Assume homogeneous effect, then, from the study results, 74% UA/NSTEMI pts, Prob (obs -ve estimate in STEMI & +ve in UA/NSTEMI | D^ sig) ~ 0.0098
➔ if we observe –ve estimate in STEMI, it is likely to be real given these are pre-specified objectives), i.e., effect ~ 0 in STEMI, which was not the case

* Lancet 2009

Figure 5a. Subgroup power for $\delta s/\delta$ = 1, 2 or 3, given α = 0.025, $\alpha 1$ = 0.02, 1-$\beta 1$=0.90



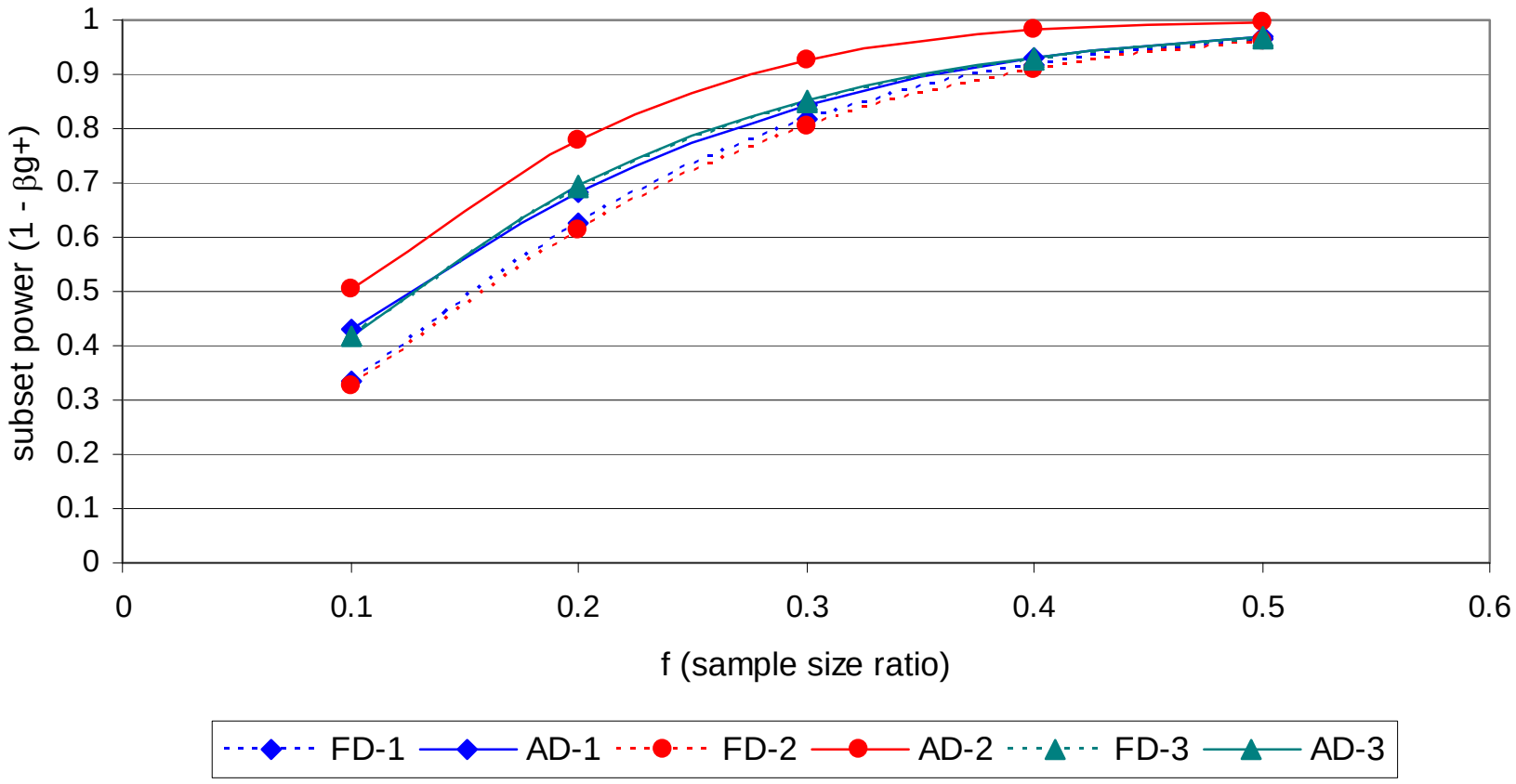
Statistical Analysis Issues

- Convenience sample issues:
  - Bias
- Full ascertainment of subset – design efficiency*
- Prevalence when full ascertainment of ITT
  - Efficacy (effect size vs. diagnostics: PPV, NPV ?)
  - Safety (predictive of group vs. individual - AUC)
- Efficiency with prospective/retrospective designs
  - Intended patient population vs. Intended use
- Replication

* Maitournam & Simon (2005 SIM)

When would Adaptive Design be Useful for Identifying Biomarker-Responsive Patients

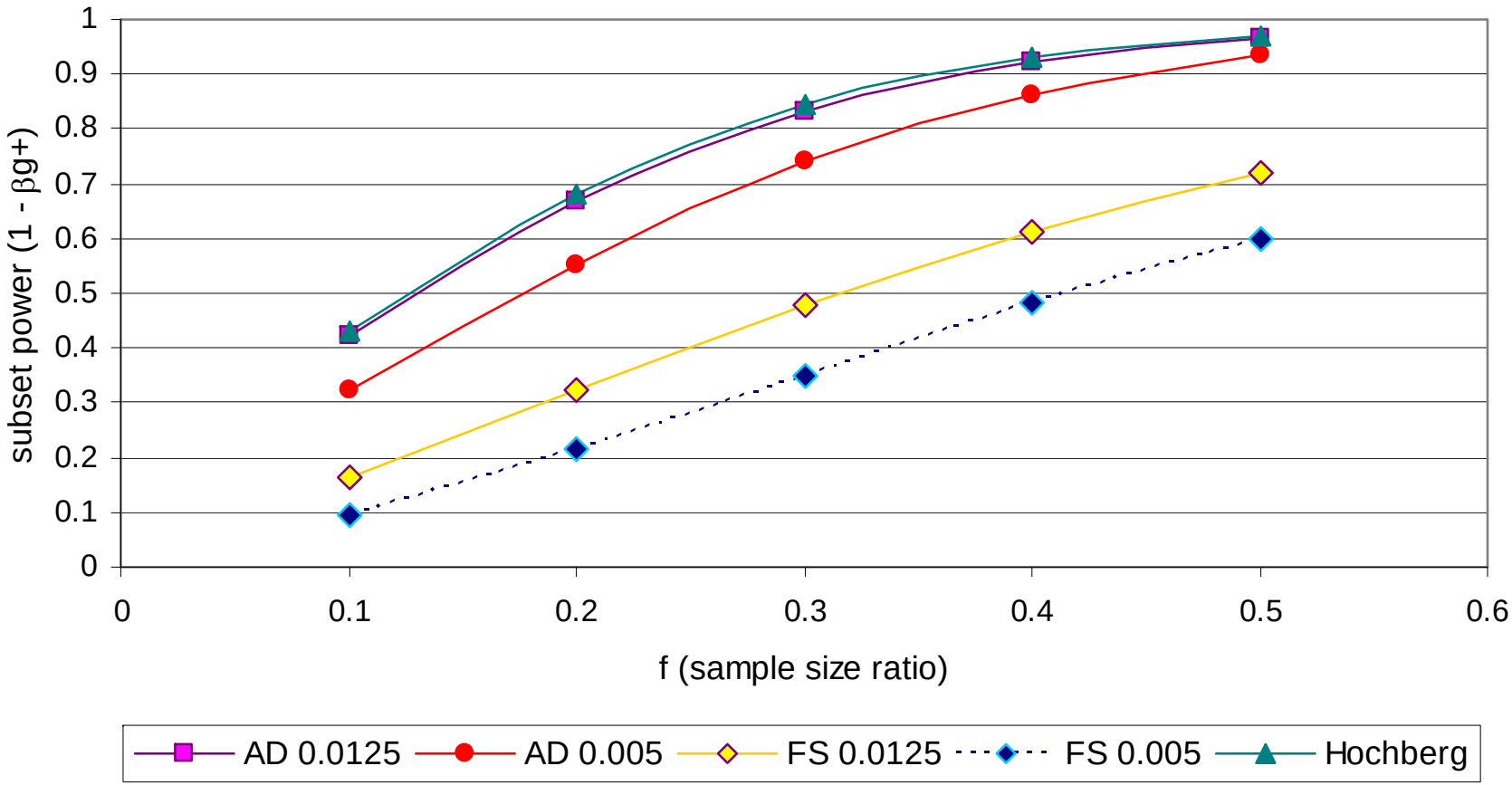
Figure 7. Power Comparison for Δg+ with Hochberg Method
(1=(Δ, 0.4, 0); 2=(0, 0.4, Δg- < 0); 3=(0.2, 0.4, Δg-))



Wang SJ, UPenn 2009

Power Comparison of Adaptive Designs: When Learning is within vs. Outside of Trial

Figure 4. Power Comparison for Δg+ Under Adaptive Design
(Δg+ = 0.4, Δg- = 0)



Level of Evidences for Establishing Biomarker Associated Treatment Effect

- Are there preliminary evidences that genomic biomarker is a predictive factor ?
- In oncology area, clinical research in early trials often entails single arm study, at best describes the prognostic utility
- A biomarker that is predictive in single-arm study does not address the question of biomarker's predictivity of treatment effect
- In 2-arm RCT, simply pre-specifying the hypotheses, it is unclear if predictive of treatment effect is automatically established when hypothesis is rejected

Use of Genomic Models to Predict Treatment Outcome Relies on

- Clear definition of treatment outcome
  - Efficacy vs. safety
- Systematic genomic data collection in clinical studies
  - Case ascertainment
  - Convenience sample
- Quality of genomic measurements for analytical and clinical validation, e.g., undetermined samples
- Seriousness of misclassification of biomarker classifier – performance characteristics of diagnostic
- Prevalence, study power

What and How Much Evidences are Needed for Predictive Biomarker Claim ?

- A study that is not originally designed to address the predictive effect, e.g., prospective/retrospective studies, but is used to interpret treatment effects, how reliable the observed association external to such study is and is it repeatable ?
- Two independent studies, both are prospectively designed to test biomarker-associated treatment effect hypothesis – replication ?
- The strength of statistical evidence relies on
  - Consistent effect size across studies, sufficient sample size, control for false positive conclusions relative to pre-specified multiple hypotheses

Regulatory Issues for including a biomarker classifier in RCTs

- Type of data evidences needed
- Implication of convenience samples for design, analysis, and interpretation
- Conditions of prospective-retrospective study designs that can be considered A&WC investigation for label claim
- Replication