A Subgroup or a Subpopulation

Design and Analysis Issues in Clinical Trials*

Sue-Jane Wang, Ph.D.

Office of Biostatistics, Office of Translational Sciences

Center for Drug Evaluation and Research, US FDA

For presentation at the EMA Subgroup Analysis Workshop, London, United Kingdom, November 18, 2011

* Views expressed are the author's professional views and not necessarily those of U.S. Food and Drug Administration

Acknowledgments

Robert O'Neill (Statistical)

Robert Temple (Clinical)

Badrul Chowdhury (Clinical)

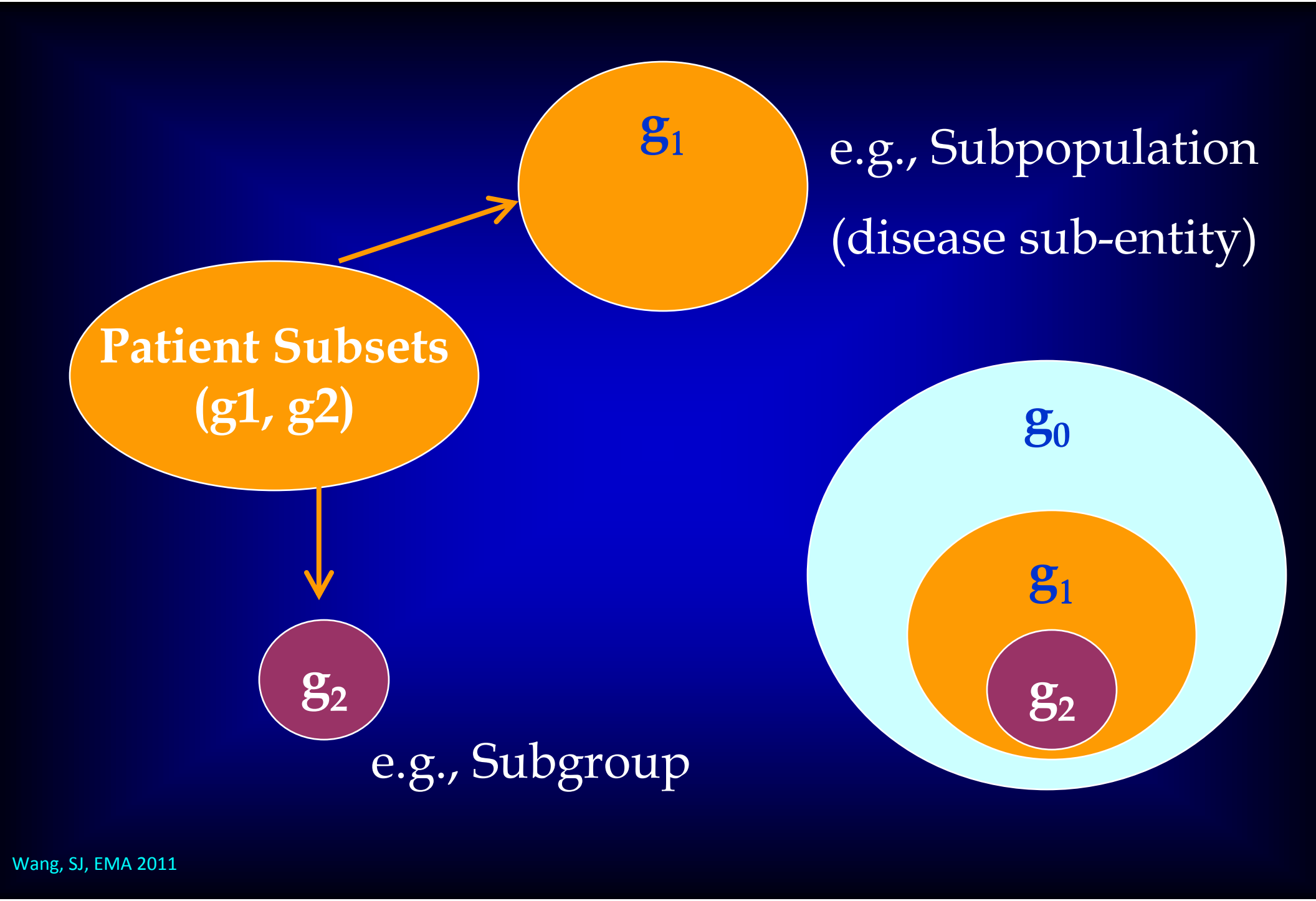
James Hung (statistical)

Yan Wang (Statistical)

And many clinical and statistical colleagues through review of IND/NDA/BLA submissions across all disease areas in CDER/FDA

Outline

- Patient Subsets
- Traditional Subgroup Analysis
- Prospectively Planned Subset Hypothesis
- Probability of at least one negative subgroup
- Bias, Prob(Imbalance), Sample Size Implication
- Regulatory Considerations of Sub-population
- Concluding remarks



Wang, SJ, EMA 2011

Baseline Covariate(s) that Classifies Patients into Subsets

- There may be subset(s) thought to be, e.g., at greater risk to an event (prognostic) that potentially increase study power
- Some subset(s) who may have a higher probability of response or toxicity to new treatment (predictive of treatment effect)
- In randomized controlled trials, these includes demographic; disease history; medication history; clinical and genomic data

➔ should be consistently collected prior to treatment intervention

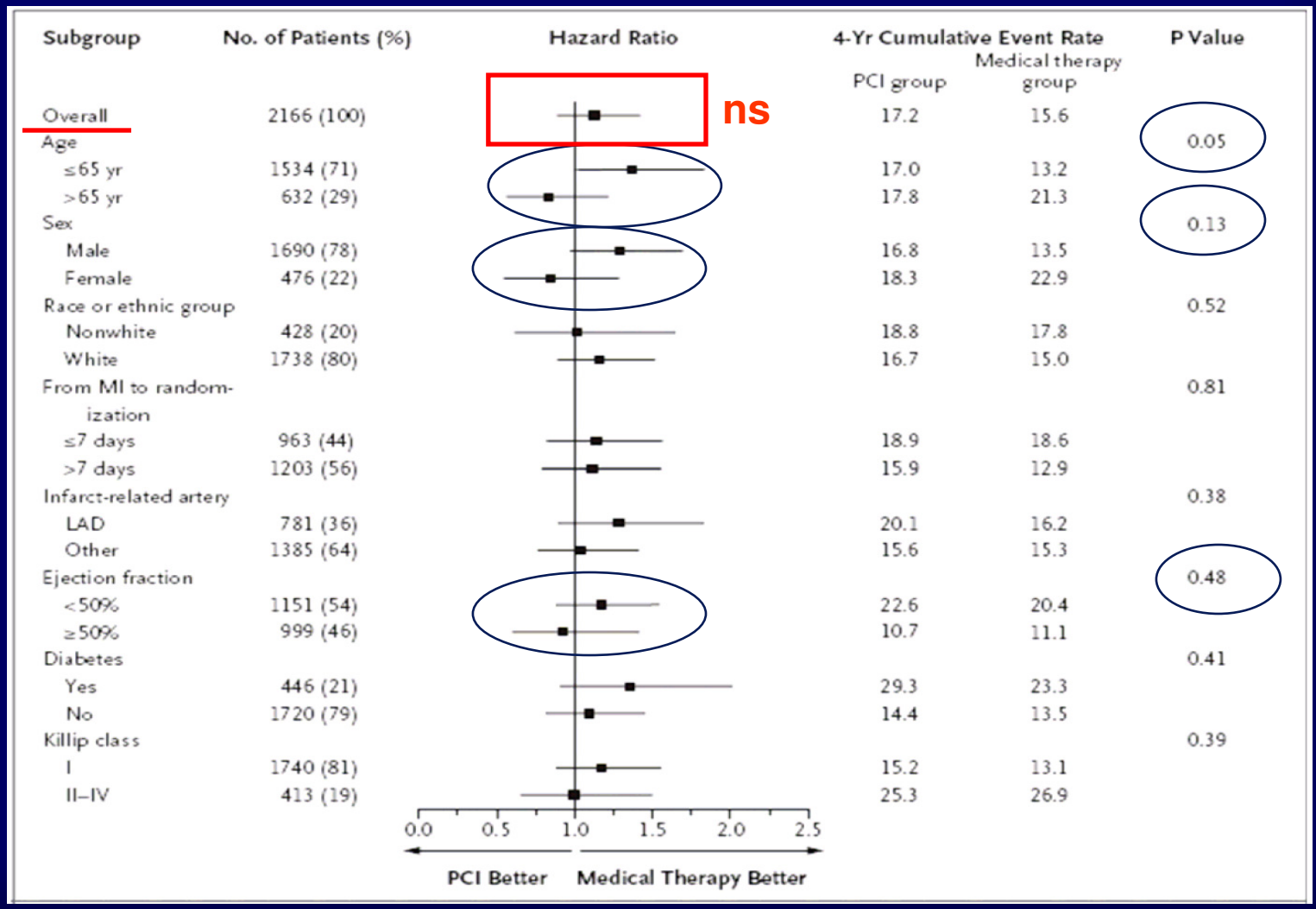
Wang, SJ, EMA 2011

Subgroup (Subset) Analysis

- Traditional: pre-treatment baseline covariates one variable at a time
  - By gender, by race, by age
  - By region (pre-determined or post-hoc grouping)
- Alternatives: aggregate measure of many important baseline attributes, e.g., APACHII score, PANSS total score, multivariate scoring, population risk
  - Potential indicator of disease severity, explanatory variable of heterogeneity of treatment effect, predictive of treatment response or effect
- They are generally considered exploratory (e.g., ICH E-9)

Wang, SJ, EMA 2011

Should One also Prospectively Test T-effect in Younger or Male Patient Subset ?



Hochman et al. (2006, NEJM) Homo p-range: 0.05 – 0.48 (opposite dir)
Homo p-range: 0.38 – 0.52 (pt~0 vs. >0)

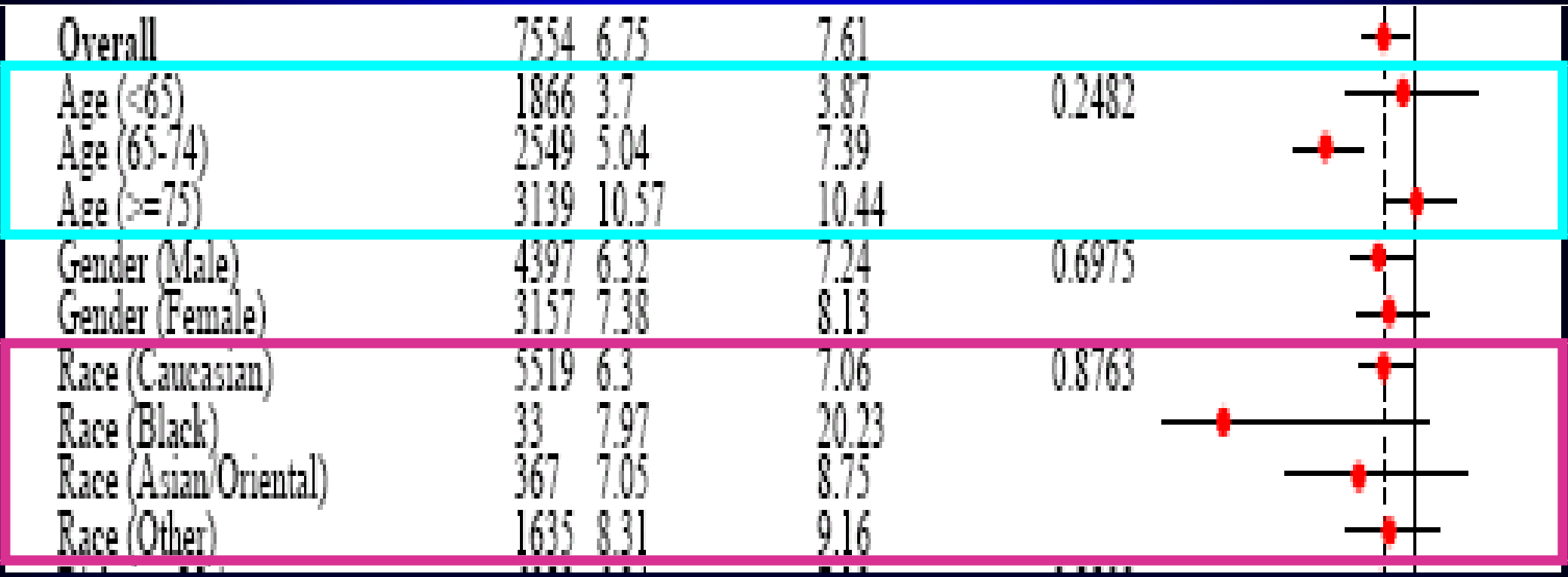
Wang, SJ, EMA 2011

Subset: Prospective vs. Retrospective (Multiplicity?)

Consistency vs Inconsistency → Potential Heterogeneity

Treatment effect estimates: Frequentist vs Bayesian

Subset	N	T-rate	P-rate	p(interact)	HR (95%CI)
Overall	7554	6.75	7.61		
Age (<65)	1866	3.7	3.87	0.2482	
Age (65-74)	2549	5.04	7.39		
Age (>=75)	3139	10.57	10.44		
Gender (Male)	4397	6.32	7.24	0.6975	
Gender (Female)	3157	7.38	8.13		
Race (Caucasian)	5519	6.3	7.06	0.8763	
Race (Black)	33	7.97	20.23		
Race (Asian/Oriental)	367	7.05	8.75		
Race (Other)	1635	8.31	9.16		



Wang, SJ, EMA 2011

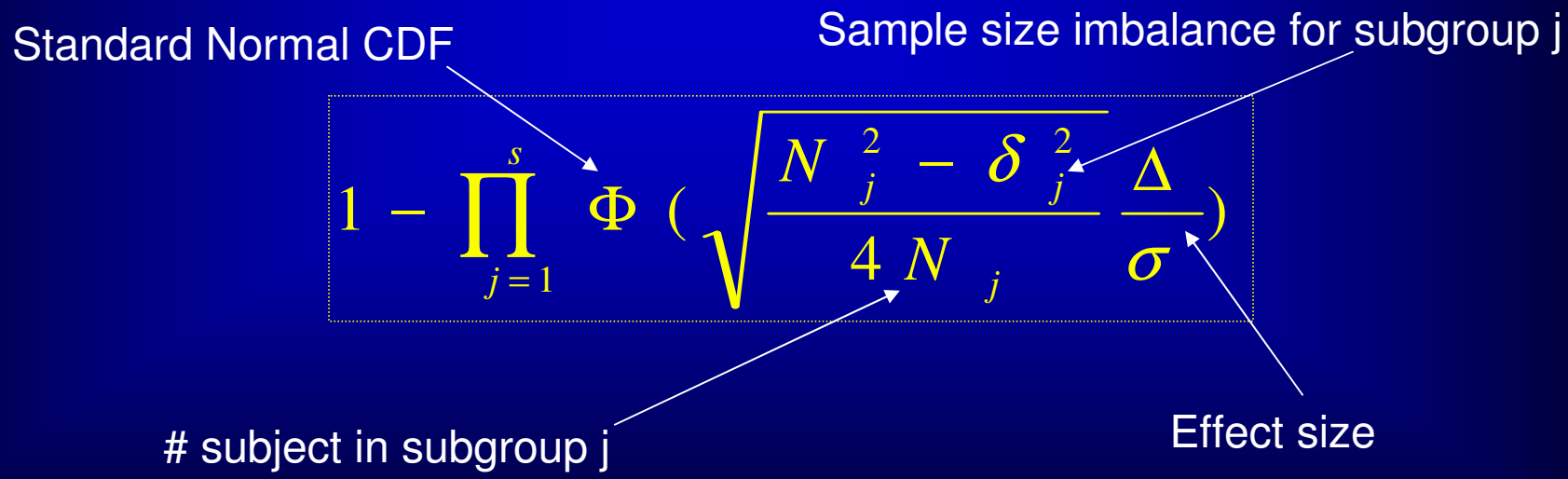
Prospectively Planned Subset Hypothesis

- What is or are the primary study objective(s) ?
- Valid statistical methods including non-adaptive all comer patients and adaptive enrichment via pre-specified multiplicity adjustments are available and not a concern
- Depending on a clear hierarchy of subset hypotheses: in a simplest setting with one subset (ITT and/or subset) – weight and strength of evidence of a statistically significant subset would generally be judged in light of the ITT including safety and efficacy in opposite subset
- Interpretation issues with adaptive enrichment due to patient mixture
- Replication of finding

Wang, SJ, EMA 2011

Analytical Approach

- **Assume Δ>0 and is homogeneous across all subgroups, the probability of observing a negative result for at least one of S mutually exclusive subgroups**



Wang, SJ, EMA 2011

(Li et al. 2007)

Pr (observe ≥ 1 negative result)

Subgroup-Inconsistency Not Necessarily Hypothesis Test

- Factor increasing Pr (≥ 1 negative result) ↑
  - # of subgroup is large
  - Severe sample size imbalance b/t treatments
  - Severe sample size imbalance b/t subgroups
  - ➔ Reduced precision of effect estimates

- Factor decreasing Pr (≥ 1 negative result) ↓
  - A profound treatment effect size
  - # of subgroup is small
  - A large sample size

Wang, SJ, EMA 2011

Biased Estimates Under H_0: No Treatment Difference

Treatment group	Biomarker distribution		Response rate		Marginal probability of response
	B+	B-	B+	B-	
T	0.7	0.3	0.6	0.3	(0.7)(0.6)+(0.3)(0.3) = 0.51
P	0.4	0.6	0.6	0.3	(0.4)(0.6)+(0.6)(0.3) = 0.42

Wang, O'Neill, Hung, 2010, Clinical Trials

Wang, SJ, EMA 2011

Probability of Imbalance with Given Sample Size and Subset Size

Prevalence	N=1350/arm d = 5%	N=350/arm d = 10%	N=150/arm d = 15%	N=50/arm d = 20%	N=20/arm d = 20%
10%	0.0000	0.0000	0.0000	0.0017	0.0631
20%	0.0012	0.0011	0.0012	0.0173	0.1636
30%	0.0046	0.0044	0.0045	0.0377	0.2258
40%	0.0080	0.0077	0.0079	0.0519	0.2582
50%	0.0094	0.0091	0.0093	0.0569	0.2682

d: % observed imbalance between the treated group and the comparator group

Wang, O'Neill, Hung, 2010, Clinical Trials

Wang, SJ, EMA 2011

Credibility of Subset Results

Claim Based on Pre-specified "Subgroup"

- Not only pre-specification

  a specific claim of a beneficial effect in a particular subgroup requires pre-specification of the corresponding null hypothesis and an appropriate confirmatory analysis strategy

- Overall treatment effect is the gatekeeper **(?)**

  It is highly unlikely that claims based on subgroup analyses would be accepted in the absence of a significant effect for the overall study population

- Randomization would generally be stratified

Label Restricted to

Subgroup Not Pre-specified → "Subpopulation"

- If a large variety of subpopulations are investigated without proper plans to deal with label consideration
- An **overall positive result** (statistically and clinically) in the whole study population may not lead to valid claims for all subpopulations if there is reason to expect heterogeneity of the treatment effect in the respective subpopulations
- If a meaningful definition of the overall study population is lacking, …… subpopulations can be adequately represented in which statistically significant and clinical relevant results were observed

Label Exclusion of
Subgroup Not Pre-specified ➔ "Subpopulation"

- Strong interaction found …… patients from the respective sub-population may be excluded from the license until additional clinical data are available – often judged on a case-by-case basis

- Results not contradict regulators' belief that a certain sub-population of patients will not benefit due to historical reasons – in regulatory reviews, such scenarios are generally discouraged during IND stage

Concluding Remarks

- Traditional subgroup analysis (pre-specified or post-hoc) assesses internal consistency of treatment effect among subsets
- Usually on a specific endpoint (primary endpoint) and exploratory
- Subgroup analysis can be expanded to more efficacy endpoints other than routine safety for subset benefit/risk assessment to generate hypothesis of treatment effect in potential subpopulation
- Heterogeneity when observed in a post-hoc manner requires additional detective work to sort out if there is subpopulation – sometimes borrowing other completed studies
- Prospective subset hypothesis with clinical/biological plausibility, if done adaptively ensures statistical validity and potential efficiency

Wang, SJ, EMA 2011