An Evidence Based Approach for Phase II Proof-of-Concept Trial

Satrajit Roychoudhury
Pfizer Inc.

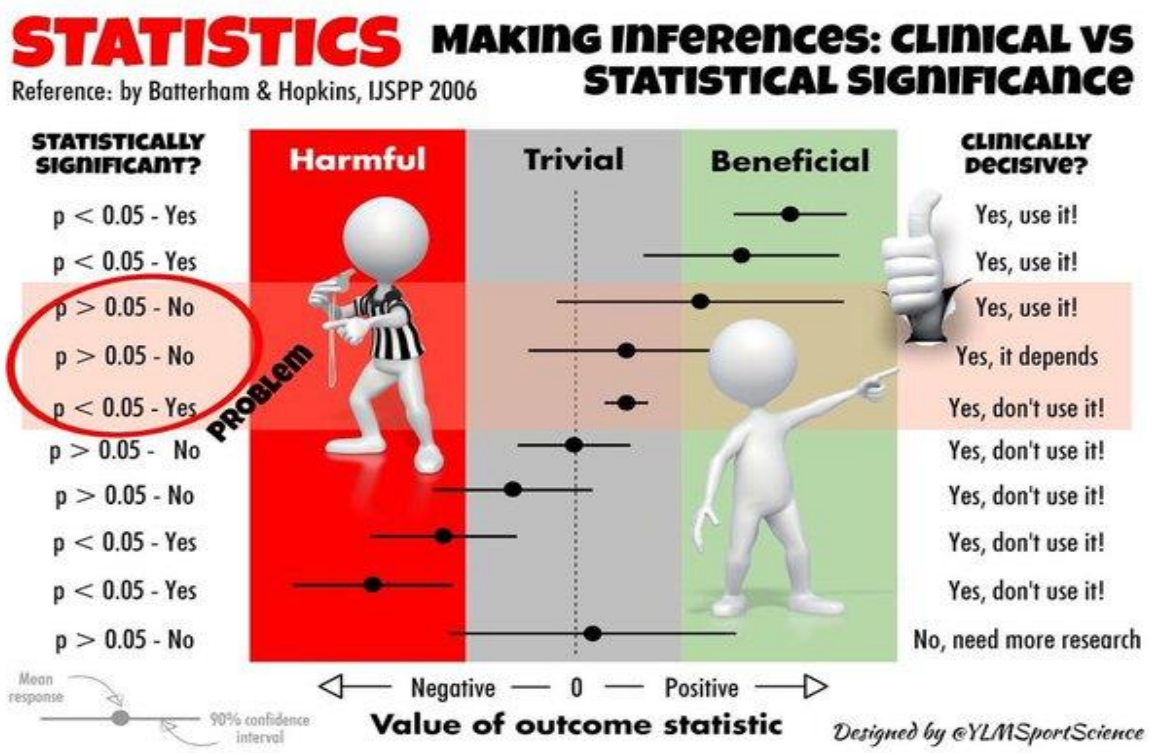
ASA BIOP Regulatory-Industry Statistics Workshop
Washington D.C.
September 25, 2019

GLOBAL PRODUCT DEVELOPMENT

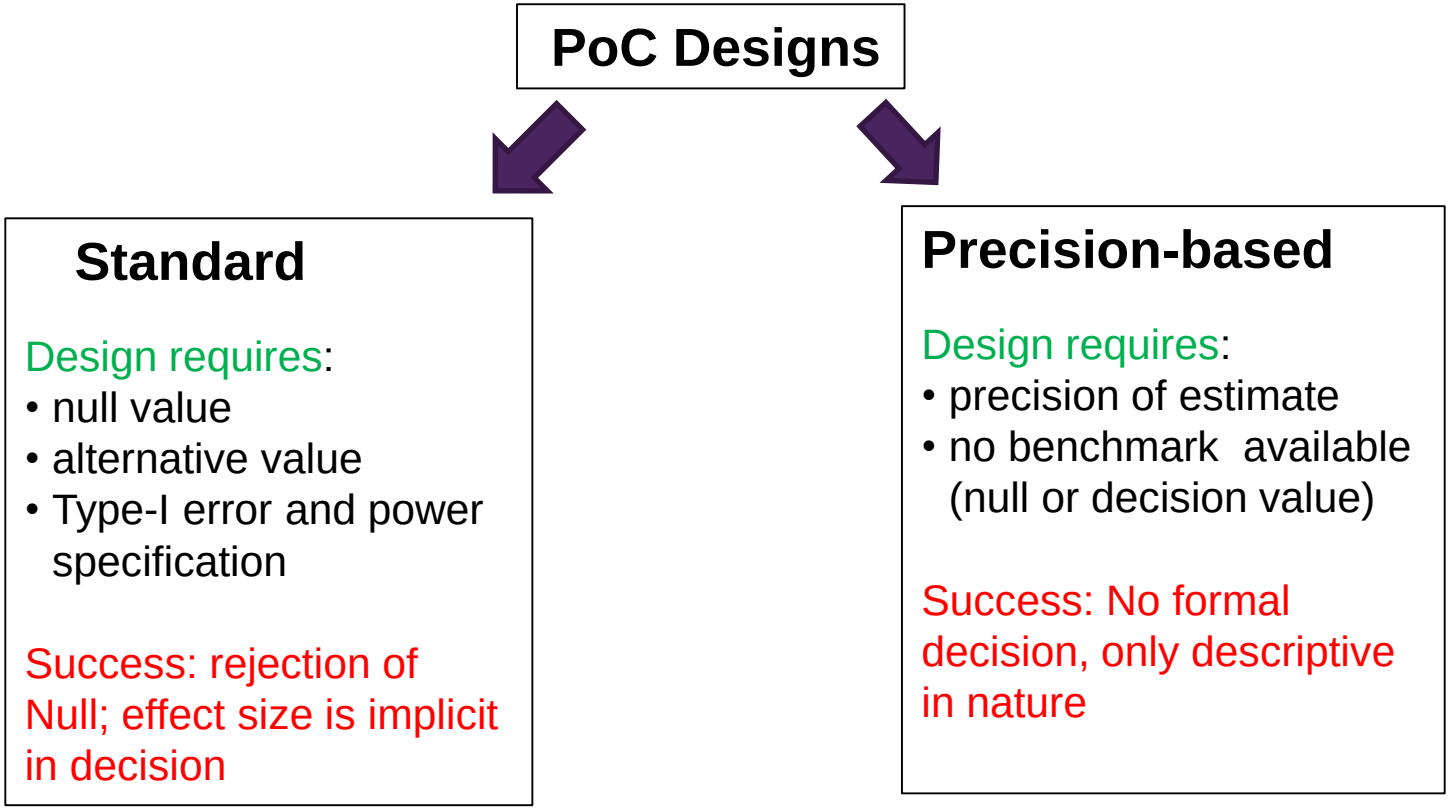
POC Trials: Getting Most Out of It!

- Proof-of-Concept (POC) is a milestone in the clinical development
- Traditionally Go/No-Go based on p-values or upper/lower CI
- A good POC should answer the followings:
  - What do we need to know?: inference regarding the treatment effect
  - How sure do we need to be?: manage the risk of making wrong decision
  - Can we stop early if treatment is not good?: meaningful interim analysis
- Can be formulated using Frequentist and Bayesian framework
  - Quantitative risk assessment for making Go/No-Go decision
  - Bayesian framework may be useful when historical data is available

Statistical vs Clinical Significance: The “Gray Zone”



PoC Study Designs: Commonly Used Design



PoC Study Designs: Evidence based Design

Standard

Design requires:
- null value
- alternative value
- Type-I error and power specification

Success: rejection of Null; effect size is implicit in decision

Dual-criterion

Design requires:
- null value
- decision value (minimum estimated effect size)
- Type-I error

Success: rejection of Null & sufficiently large effect size

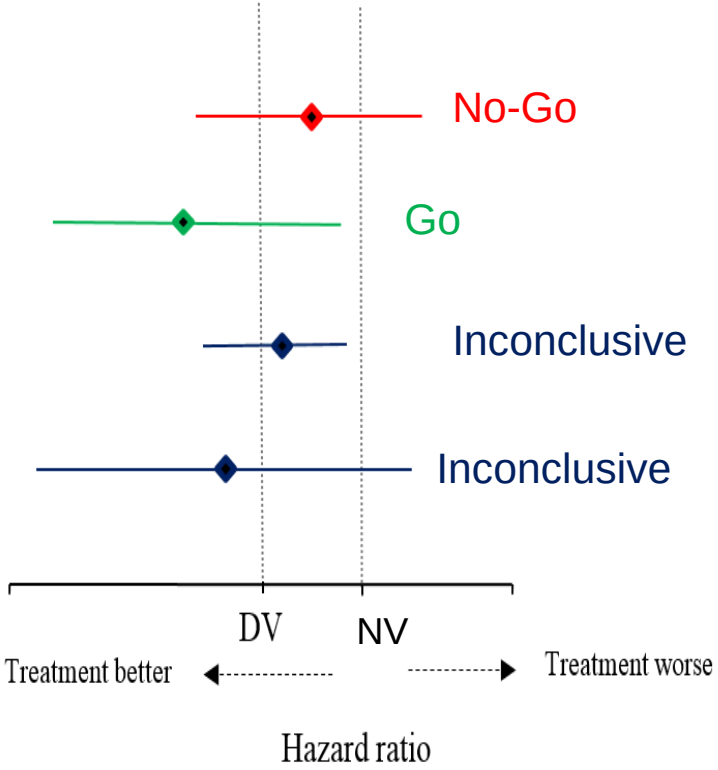
Precision-based

Design requires:
- precision of estimate
- no benchmark available (null or decision value)

Success: No formal decision, only descriptive in nature

POC Trials: Dual Criterion Design

- Formal inclusion of Statistical significance and clinical relevance in design: Decide GO
  - Strong evidence: effect ≥ no effect or null value (NV)
  - Estimated effect ≥ decision value (DV)
- NV and DV are the critical design parameters
  - DV is not classical alternative hypothesis
- Sample size requires consideration of DV
  - Ensure statistical significance when clinical relevance observed
  - Optimal type-I error control/power via simulation
- Applicable single arm and randomized setting



Statistical Significance and Clinical Relevance in POC Design

- Decision value (DV) is also known as
  - *target difference, minimum clinically important difference*, *indifference point*
- The idea to base GO decision on both statistical significance and effect estimate is widely discussed across pharma industry
- **Final decision:** should account for *all relevant information* e.g., secondary endpoints, safety, and subgroups
- Dual-criterion can be formulated in both Frequentist and Bayesian fashion

Design Inputs

Input Parameter	Standard Design	DC Design	Precision Design
Null value	✓	✓	✗
Alternative value	✓	✗	✗
Decision value	Implied	✓	✗
Type-I error	✓	✓	✗
Power	✓	Implied	✗
Estimate	Implied	Implied	✓
Sample Size	Implied	Implied	Implied

Sample Size Calculation

- Sample size calculation of Dual-criterion design needs consideration of both criteria.
- The sample size must ensure statistical significance when clinical relevance is achieved. We define this threshold as *minimum sample size* (n_{min})
  - For normally distributed data, $n_{min} = \sigma^2 \frac{z_\alpha^2}{(NV-DV)^2}$
  - For non-normal data, a grid search over sample sizes may be needed
- A good Dual-Criterion design must have sample size ≥ n_{min}
- The final sample size should be based on desired OC and other considerations (secondary endpoint, safety etc.)

Data Scenarios and Operating Characteristics

- On-study data scenarios showing
  - potential results, along with the Dual-criterion metrics and final study outcome
  - are useful to clarify the decision-making process and ensure they make clinical sense
- To assess the adequacy of the design, the operating characteristics should be measured:
  - Type-I error: under the null value, the probability for a GO decision must be controlled at α
  - Power: for truly promising effects (considerably better than DV), the probability for a GO decision should be large (70-90%)

– On-study scenarios and OCs should both be shown in the protocol

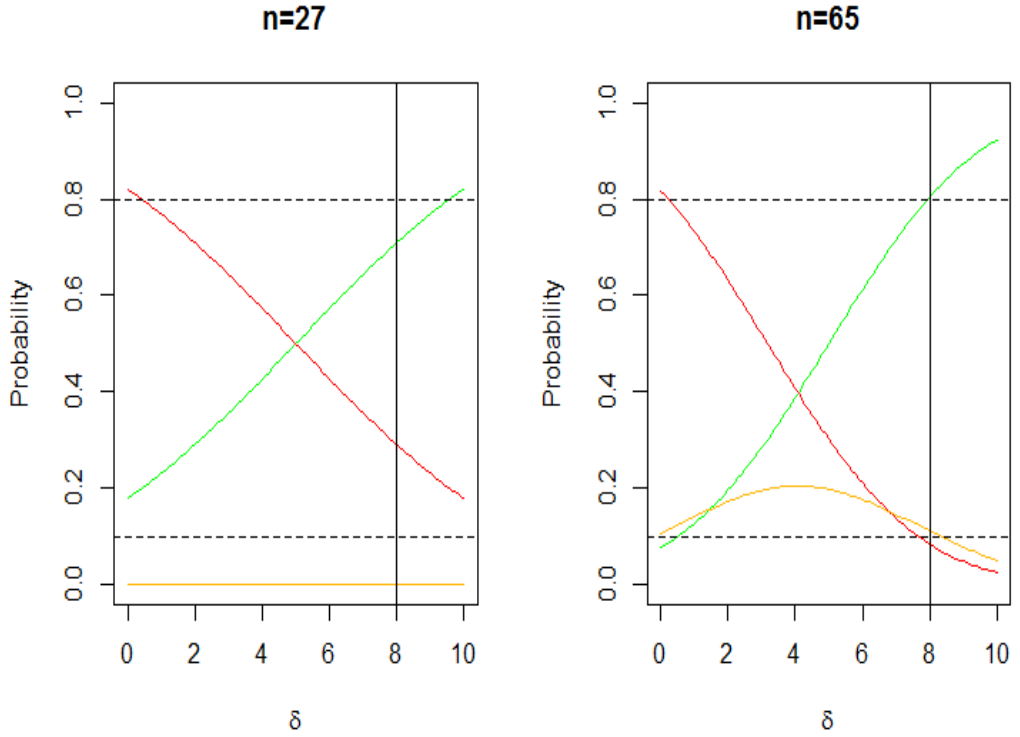
Example 1: PoC Trial in Cystic Fibrosis

- Randomized, double-blind, placebo-controlled, parallel group trial
- **Endpoint:** change from baseline in % of predicted FEV1 at day 28 (δ)
- Specification of NV and DV need literature review and clinical discussion
  - NV = 0 => true drug effect is better than placebo
  - Data from competitor have shown effect size of 8% in a similar population
  - No interest in further development if observed effect **< 5%** in PoC: DV = 5%
- Go criteria is defined as:
  - Statistical significance: one sided p-value < 0.1
  - Estimated change from baseline in % of predicted FEV1 ($\hat{\delta}$) ≥ 5%
- Equivalent Bayesian formulation is also available

Source: Example is adopted from Fisch et. al. 2015

Example 1: Sample Size of Trial

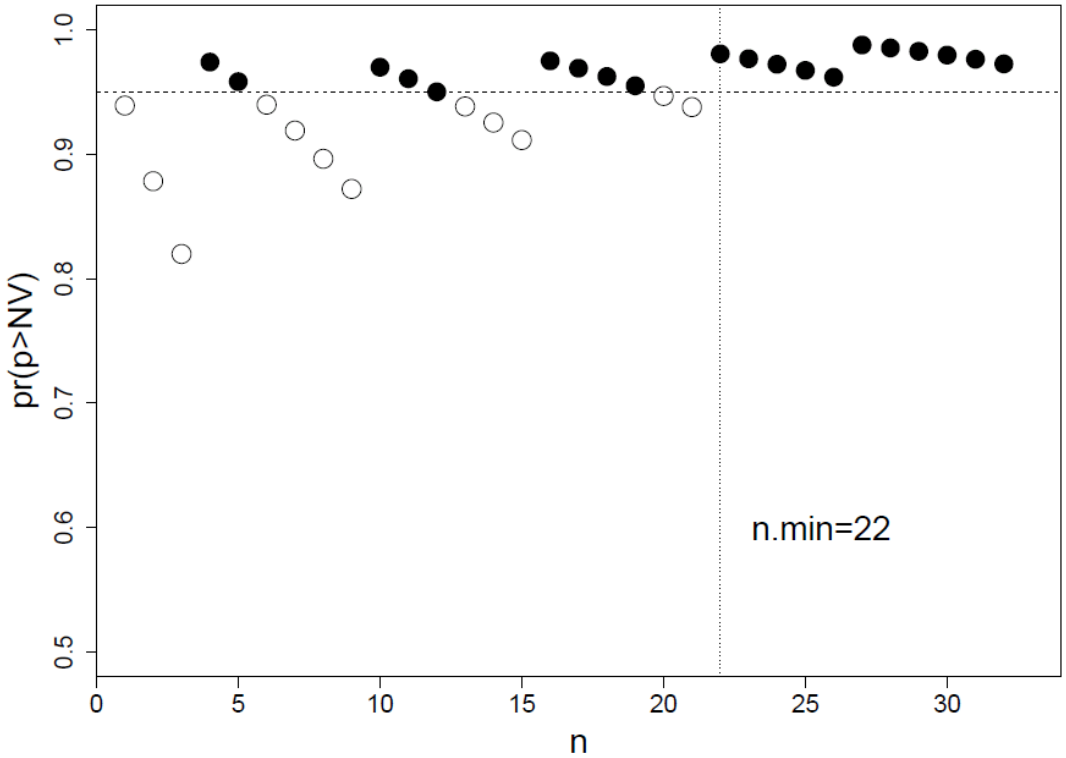
- Standard deviation for $\hat{\delta}$ =20
- Minimum sample size=27
  - n> 27 clinical relevance ensure statistical significance
- However the final sample size depends on frequentist OC
- OC sub-optimal with n=27
  - Desirable OC with n= 65
    - Type I error: 7.7% (δ=0)
    - Power: 80.4% (δ=8)
- Effect size > 8%: *target value*



Example 2: Single-arm PoC Design with Binary Data

- Background: single-arm Ph I expansion PoC trial of an experimental drug in Chinese patients with NSCLC
- Primary objective: Objective Response Rate (ORR)
- Bayesian analysis method: Posterior analysis of ORR based on a binomial sampling model and a Beta prior distribution (a=0.0811, b=1)
  - Prior mean centered at the NV (0.075) and 95% credible interval (0, 0.73)
- Bayesian dual criteria:
  - **Significance**: pr(ORR ≥ 7.5% | data) ≥ 0.95
  - **Relevance**: estimate (posterior median) ≥ 17.5%

Example 2: Sample Size Calculation

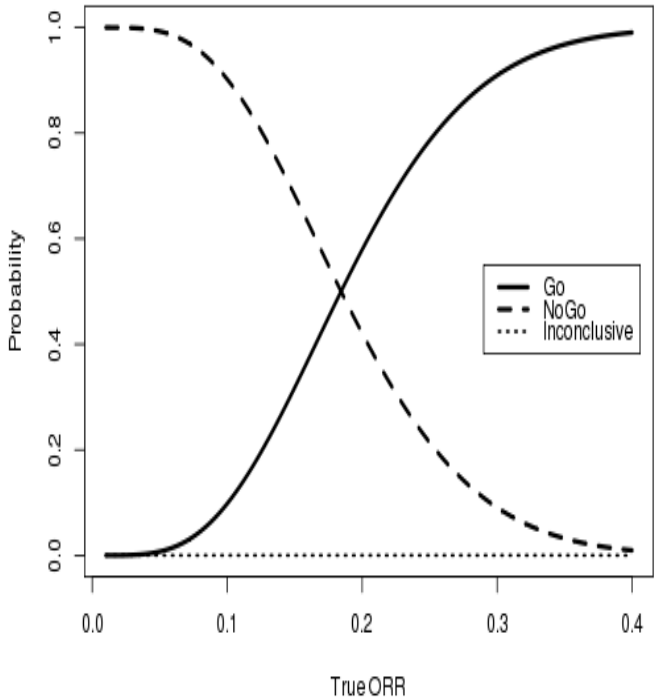


Example 2: Data Scenarios

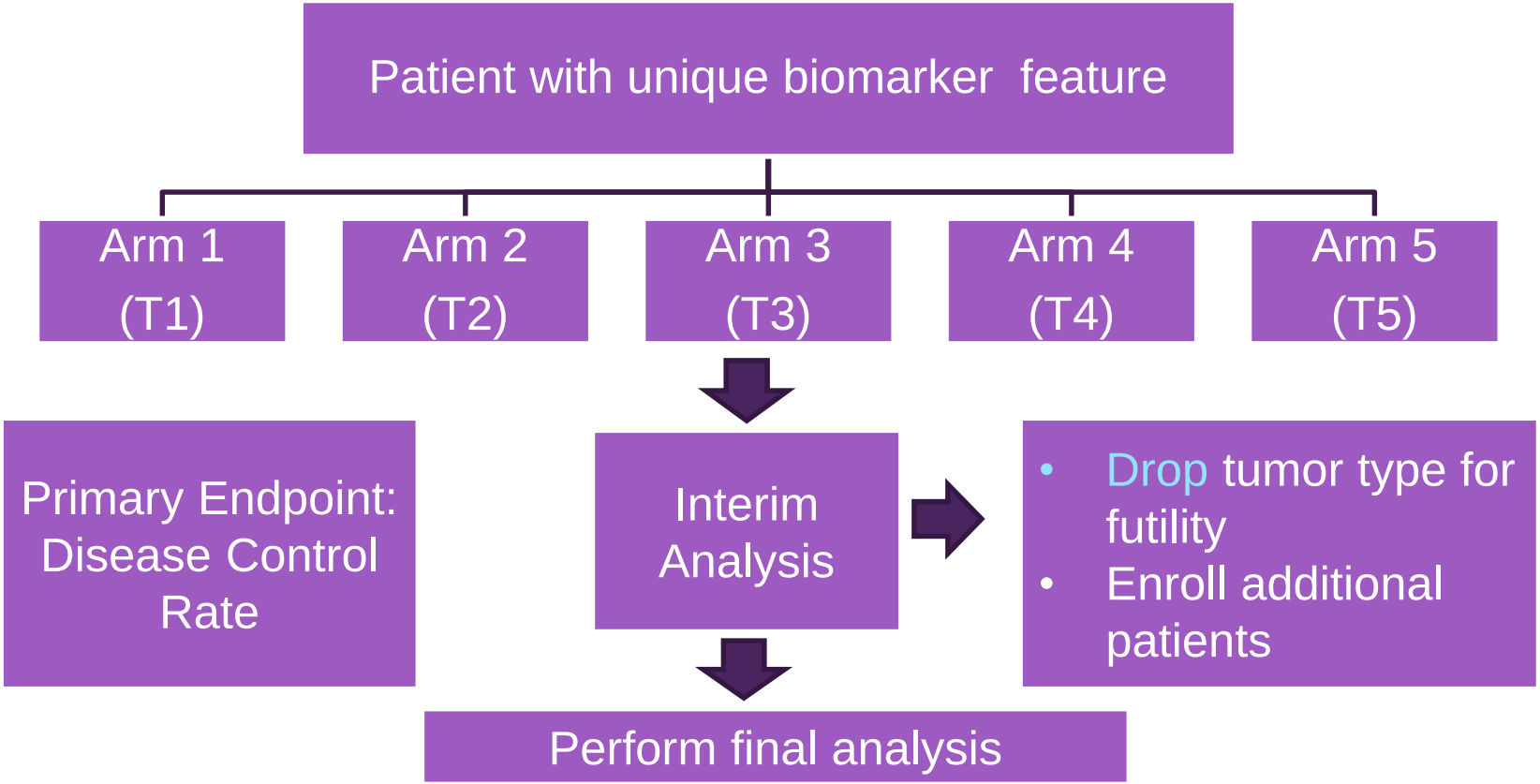
#responders/n	posterior median ORR	posterior pr(ORR>7.5%|data)	PoC decision
1/25	3%	0.161	NO-GO
2/25	6.9%	0.454	NO-GO
3/25	10.8%	0.729	NO-GO
4/25	14.8%	0.895	NO-GO
5/25	**18.7%**	**0.967**	**GO**
6/25	**22.6%**	**0.992**	**GO**

Example 2: Operating Characteristics

- The OCs exhibit a proposed design with satisfactory features



Example 3 : Basket Trial with Rare Tumor Types



Example 3: Statistical Significance and Clinical Relevance

Tumor Type	Minimum SS for Interim	Not Clinically meaningful (C_1)	Clinically meaningful (C_2)	Maximum SS
T1	10	≤40%	≥50%	20
T2	5	≤20%	≥30%	15
T3	10	≤10%	≥20%	20
T4	10	≤10%	≥20%	20
T5	10	≤10%	≥20%	20

Example 3: Criteria to Declare Futility and PoC

- A Bayesian hierarchical model is used for each interim and final analysis
- At each of these interim analyses, a decision will be made based on the calculated posterior probability
  - stop for futility if the calculated probability of being clinically meaningful (response ≥ C_2) is less than 20%
  - otherwise extend recruitment for at least 10 more patients
- For a specific tumor type, a Proof of Concept will be declared if both of the following conditions are met:
  - observed mean response ≥ "clinically meaningful" threshold (C_2).
  - posterior probability of "not being clinically meaningful (response ≤ C_1)" is less than 20%
- For any analysis all accumulated data will be used

Communication, Reporting and Software

- For Dual-criterion designs, the proper specification of the statistical and clinical criteria depends on effective communication with non-statisticians in project team.
- Proper specification of Dual-criterion needs:
  - identification and review of the medical literature with clinical team
  - discussion with key opinion leaders (KOL)
  - input from clinical and regulatory colleagues to understand the competitive landscape and regulatory requirements
- Using graphs, data scenarios and non-statistical language are useful for project team to understand the design components

Summary

- Statistical design
  - Elicit the Dual-criterion in collaboration with the clinical team
- Determine the sample size and share with the team the data scenarios and corresponding OCs
  - include them in the protocol along with the statistical model and the two criteria underlying success
- The dual-criterion can be applied in other settings (e.g. more complex setting)
  - Non-inferiority
  - Bridging studies
  - Dose finding
- Interim analysis (efficacy and futility) can be also introduced

Acknowledgement

- Beat Neuenschwander
- Nicolas Scheuer
- Santosh Sutradhar
- Tingting Yi
- Haige Shen
- Mike Branson

Reference:

Roychoudhury, Scheuer, and Neuenschwander (2018). Beyond p-values: A phase II dual-criterion design with statistical significance and clinical relevance. Clinical trials 15 (5): 452-461 .

Thank You