QUINTILES
Navigating the new health

Janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

Incorporation of Gatekeeping Procedure into Sample Size Re-Estimation Design, a Case Study

Yevgen Tymofyeyev (Janssen RD)

Joint work with Alex Dmitrienko and Jeff Maca (Quintiles)

Boston SCT meeting,
May 21, 2013

Outline

- Gatekeeping procedures
  - Analysis of multiple objectives in Phase III trials
- Adaptive design
  - Gatekeeping procedure integrated into a sample size re-estimation design
- Clinical trial simulations

Multiple Testing Problem

- Phase 3 study in patients with pain has 4 treatment groips
- Compare two doses of the experimental drug (Doses **L** and **H**) to the two control groups (**A** and **B**)

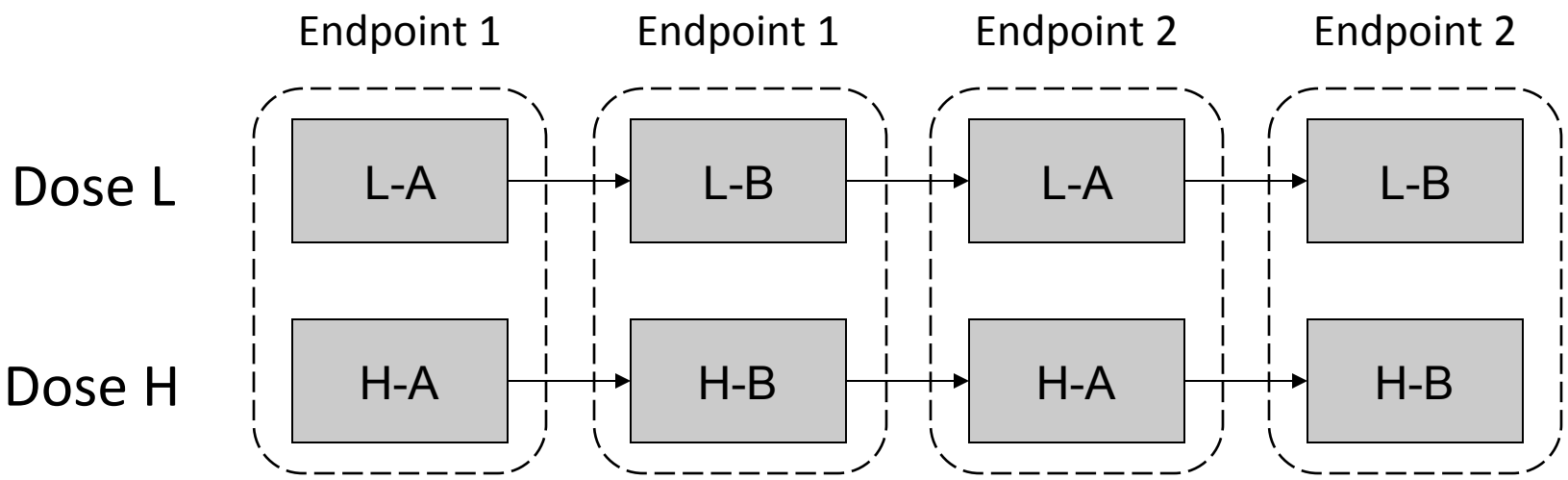
- **Endpoint 1**: (Co-primary endpoint based)
  - WOMAC pain component
  - WOMAC function component
- **Endpoint 2**.
  - Patient global assessment

- To define the multiple hypothesis testing procedure that controls the overall Type I error rate across the eight tests at a two-sided 0.05 level
  - **LA1** (L vs A, Endpoint 1)
  - **LB1** (L vs B, Endpoint 1)
  - **LA2** (L vs A, Endpoint 2)
  - **LB2** (L vs B, Endpoint 2)
  - **HA1** (H vs A, Endpoint 1).
  - **HB1** (H vs B, Endpoint 1).
  - **HA2** (H vs A, Endpoint 2).
  - **HB2** (H vs B, Endpoint 2).

Multiplicity Adjustment

- Gatekeeping strategies (gatekeeping procedures)
  - Methods for addressing multiplicity issues for multiple families of endpoints (primary and secondary)
- Type I error rate
  - Control overall Type I error rate over multiple families of endpoints
- Power
  - Provide optimal distribution of power by accounting for hierarchical structure of multiple endpoints, e.g., more power for more important endpoints
- Gatekeeping procedures
  - General methods for developing gatekeeping procedures based on powerful multiple tests (Dmitrienko and Tamhane, 2011; Kordzakhia and Dmitrienko, 2013; over 15 other papers)

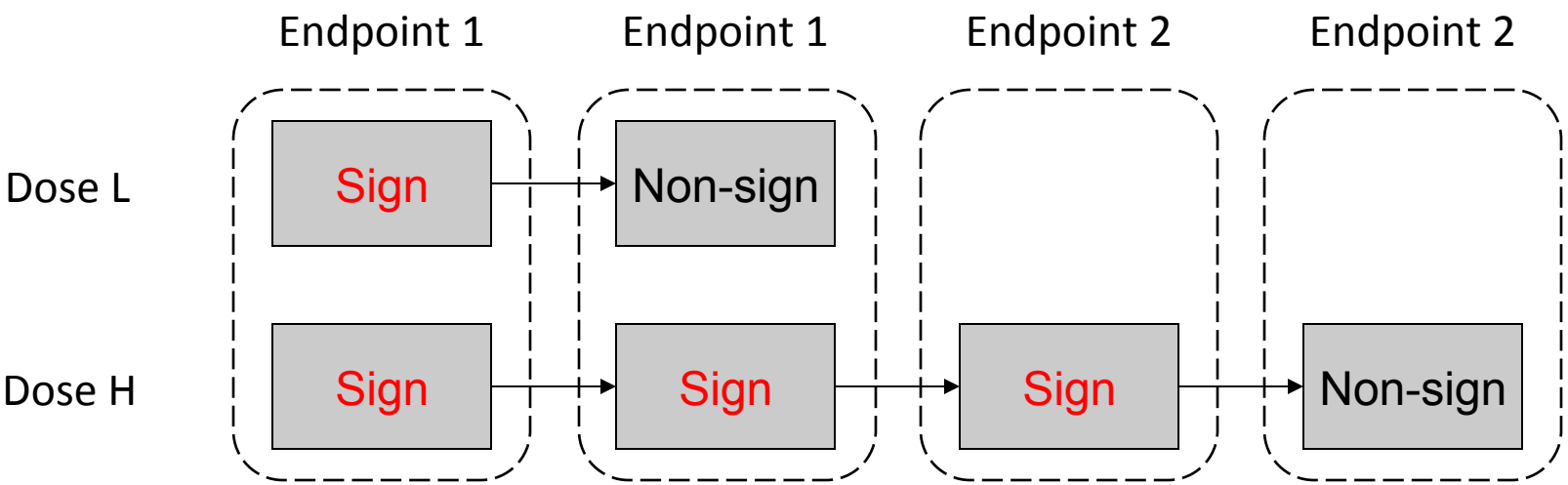
Testing strategy 1



Two independent branches (by dose)

Testing is performed sequentially within each branch

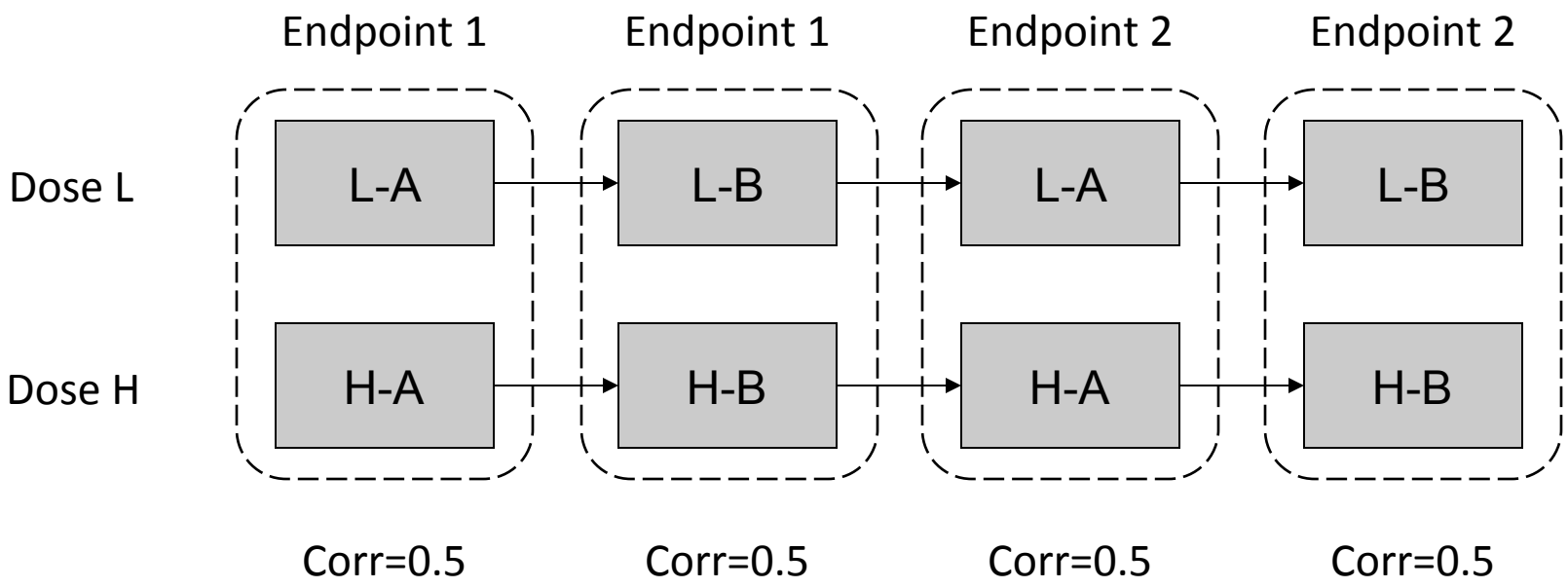
Logical relationships



Sign: Significant outcome

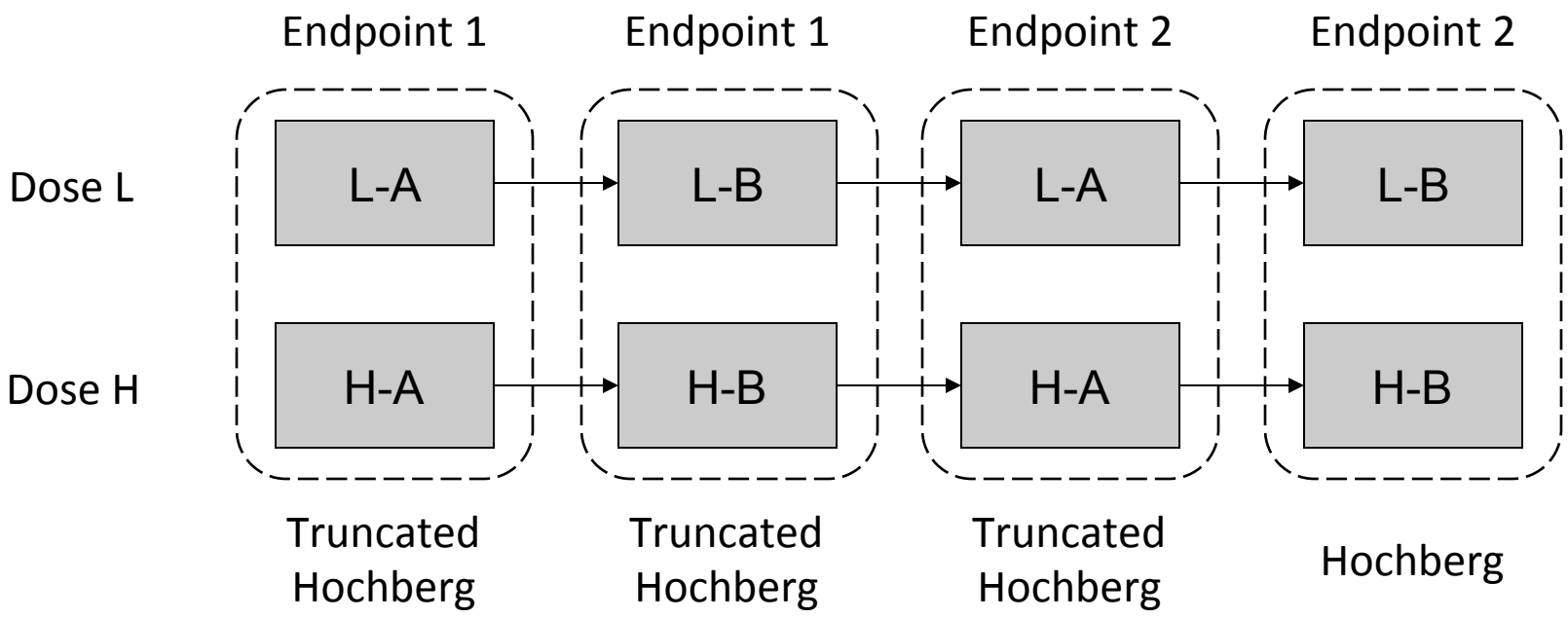
Non-sign: Non-significant outcome

Distributional information



Powerful tests can be applied to each family of tests to take advantage of positive correlation

Distributional information



Hochberg-based gatekeeping procedures help maximize overall probability of success compared to basic Bonferroni-based gatekeeping procedures

Truncated Hochberg Test

- Let $p_{(1)} \leq p_{(2)}$ be ordered p-values and $H_{(1)}$ and $H_{(2)}$ are the associated hypotheses in a family
- Assume α available is 0.05
- If $p_{(1)} \leq \gamma\ 0.05+(1-\gamma)0.025$, where $0 \leq \gamma \leq 1$ *is the* ***truncation parameter***, then reject both $H_{(1)}$ and $H_{(2)}$
- Otherwise,
  - If $p_{(2)} \leq 0.025$, reject $H_{(2)}$
  - If $p_{(2)} > 0.025$, accept both

Available α in the next family

- If both hypotheses rejected, the transferred $\alpha=0.05$
- If only one H rejected, the transferred $\alpha=(1-\gamma)0.025$

Comparison of Bonferroni and Hochberg tests

Power	Low power	More power	High power
	Bonferroni	Truncated Hochberg	Hochberg
Use in gatekeeping procedures	YES: Alpha can be transferred from family to family	YES: Alpha can be transferred from family to family	NO: Alpha cannot be transferred from family to family

Simulations Scenarios to Select Optimal Testing Strategy

- Evaluated different effect size scenarios for 4 groups:
  - "**Base**" scenario (used to power fixed sample size study)
  - "**Conservative**" (sub- efficacious, effect size still of interest)
  - Assume that Control B has better response than Control A
- Accounted for
  - EPs structure (EP1 based on 2 co-EP)
  - Correlation among EPs (high ~ 80%)
- Considered:
  - Gamma parameter scenarios (truncated Hochberg test)
  - Sample sizes scenarios

Recommended Multiple Testing Procedure

- Power characteristics were compared for different testing strategies and parameters
  - Power for each individual hypothesis
  - “Composite” power definitions, e.g.
    - Probability to demonstrate superiority of at least one dose to both control groups with respect to EP1 & EP2
- Strategy (parameter choice) relative performance depends on scenario
- Selected Testing Strategy 1
  - Need to assume most likely group effect scenarios (!)
  - Mostly interested in power to show superiority with respect to **both** control groups simultaneously
  - Verified that N=200 per group will provide adequate power

Interim Analysis (IA) and Sample Size Re-computation Rule

- Given uncertainty in design parameters, sample size adjustment at interim was desired
  - Unblinded look at the treatment effect to be used
  - IA would be performed at about 60% (n_1=120/arm) through the original sample size (N=200/arm)
  - The maximum sample size would be 300 (per arm)
  - Decision at interim will then be to expand to 300, or not.
- **Sample size is increased by same number in each treatment group**
- Decision rule to increase sample size:
  - Compute conditional power (CP) for **HB1** and **LB1** hypothesis (Family 2)
    - Assuming that true effect is equal to the current estimate observed at IA
  - If ***CP.lower*** <CP < **90%** for HB1 or LB1 hypothesis,
    - increase sample size to N=300,
  - Otherwise
    - N=200

Analysis Approach

- For each individual hypothesis, the raw p-value is computed using the combination test approach (Cui et al 1999)

$$Z_f = \sqrt{w} \cdot Z_1 + \sqrt{1-w} \cdot Z_2$$

  - $w = N_{IA} / N_{planned}$
  - Z_1 – Z-score from stage 1
  - Z_2 – Z-score based on data collected after IA only (sample size increase can take place)

- The gatekeeping strategy is applied to the raw p-values to obtain adjusted p-values
- Hypothesis with the adjusted p-value less then 2.5% (one-sided) are rejected

Testing Strategy and SSR

- Gatekeeping procedure requires the assumption (positive dependence condition) within each individual family (not across families)
- Testing at final is based on the combination tests using pooled data. Need to examine the join distribution in the context of SSR design
- The assumption holds, if each treatment group is increased by the same number of patients, Wang (2010)
- **Overall Type I error rate is controlled**

Original vs Re-estimated Sample Size Comparison

- Does the rule designed to increase sample size work?
  - Increase sample size only if results at IA are “promising”
- Evaluate under different true effect scenarios

Original sample size: Outcome	Expanded sample size: Outcome	
	Significant	Non-significant
Significant	Unnecessary SS increase (?)	Prob. of this event is very low
Non-significant	**Gain** due to SS increase	Failure to boost power (?)

Probability of events in the table need to be further considered separately depending on whether the conditions that trigger SS increase are met or not [refer to Mehta and Pocock (2000)]

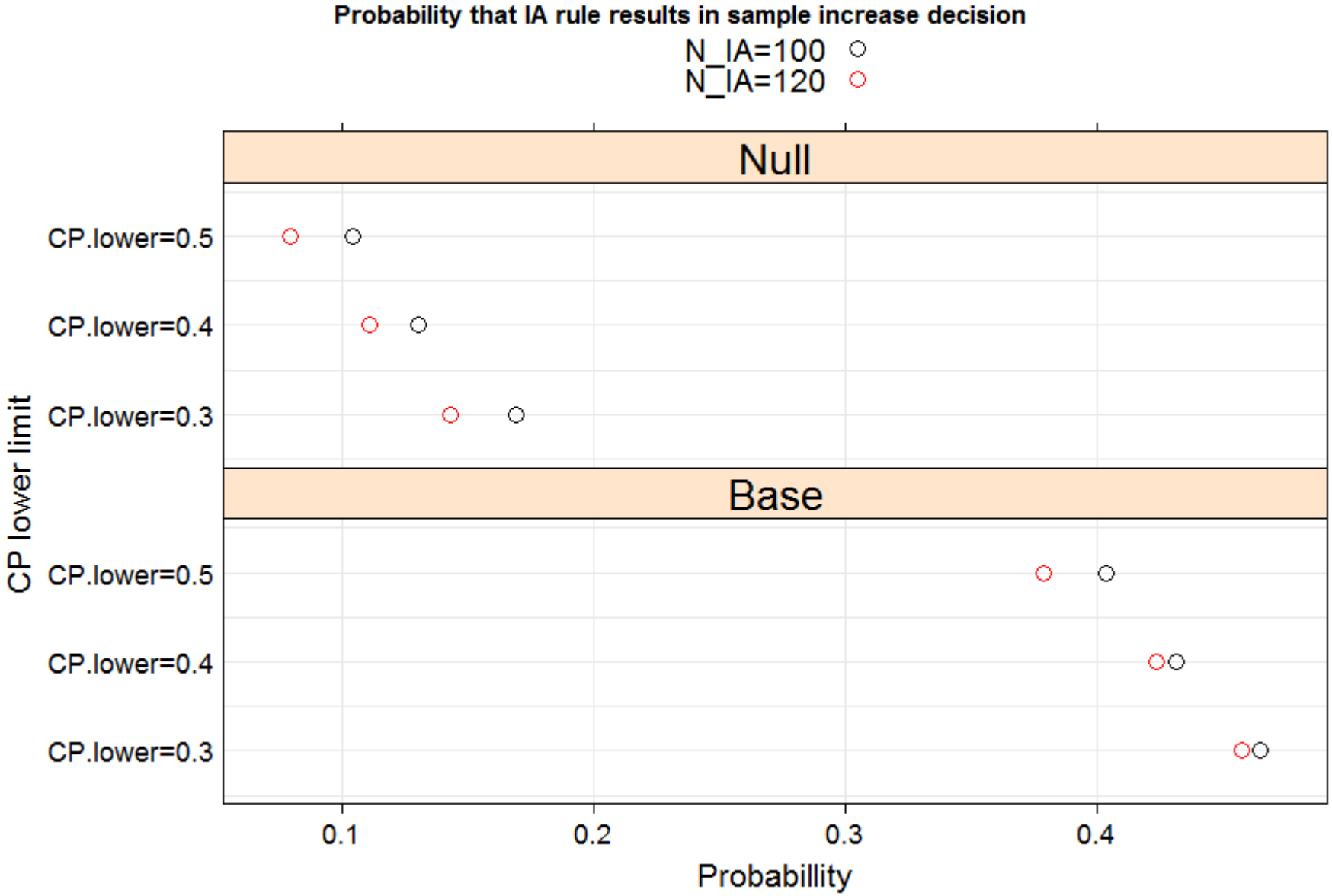
Simulations to Evaluate SSR Design

- Data generated is a such way that the same first 200 patients data (planned for the fixed design) are enclosed in the increase sample
  - N_{IA}=120 (used to compute CP and make SSR decision),
  - followed by data from N=80 subjects (to fixed design N=200)
  - plus extra data, N=100 (for potential SSR)

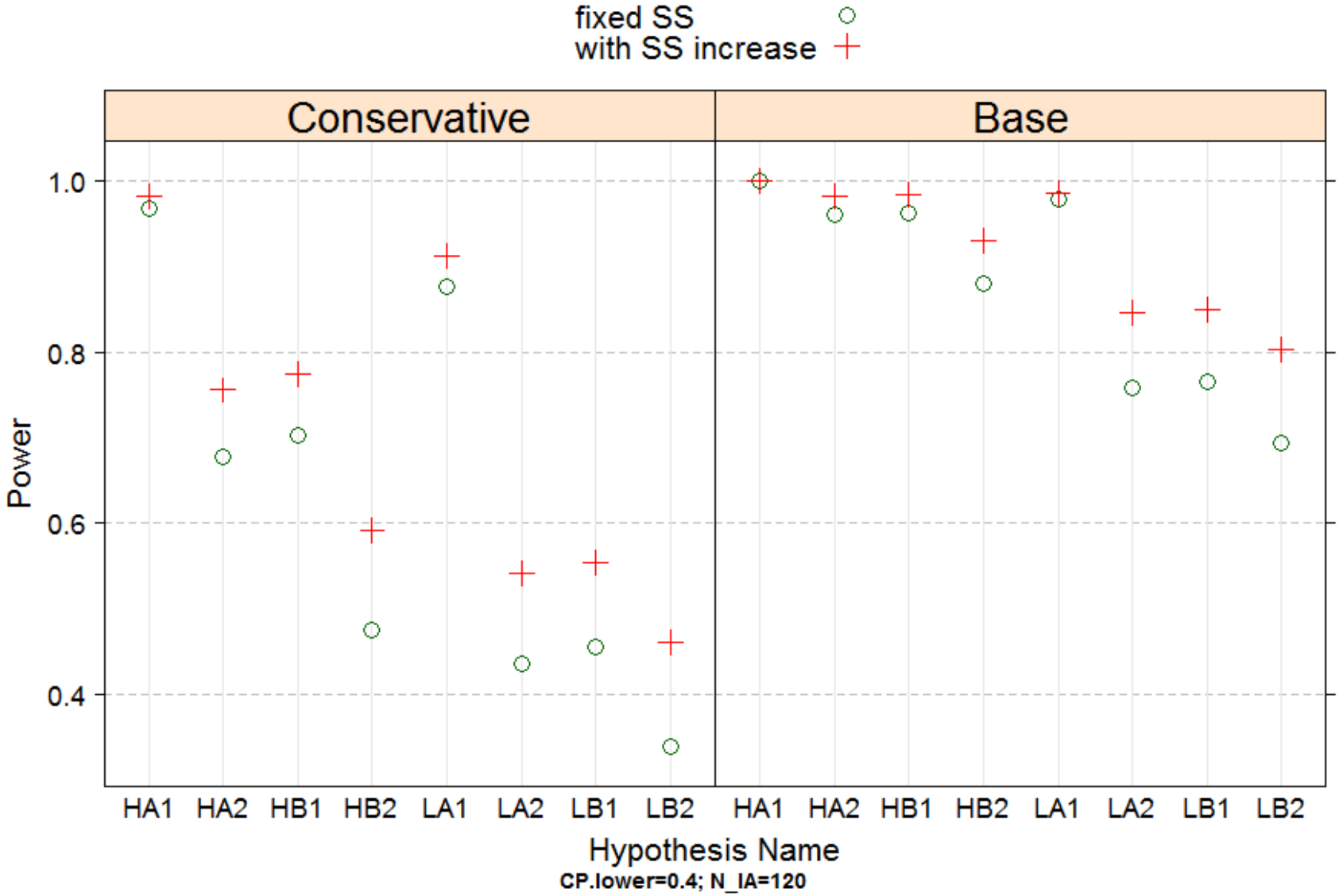
  In R, multi-arm, multi-stage, multi-variate data can be generated efficiently in a single call to `mvrnorm()` from `mvtnorm` package

- Compute OC under different true response scenarios:
  - Evaluate proportion of trial replications when the sample size was increased at IA
  - Compare power for fixed and AD
    - **Unconditionally**
      - Depends on how often the sample size increase is triggered by the rule
    - **Conditionally** on the fact that SSR rule was triggered

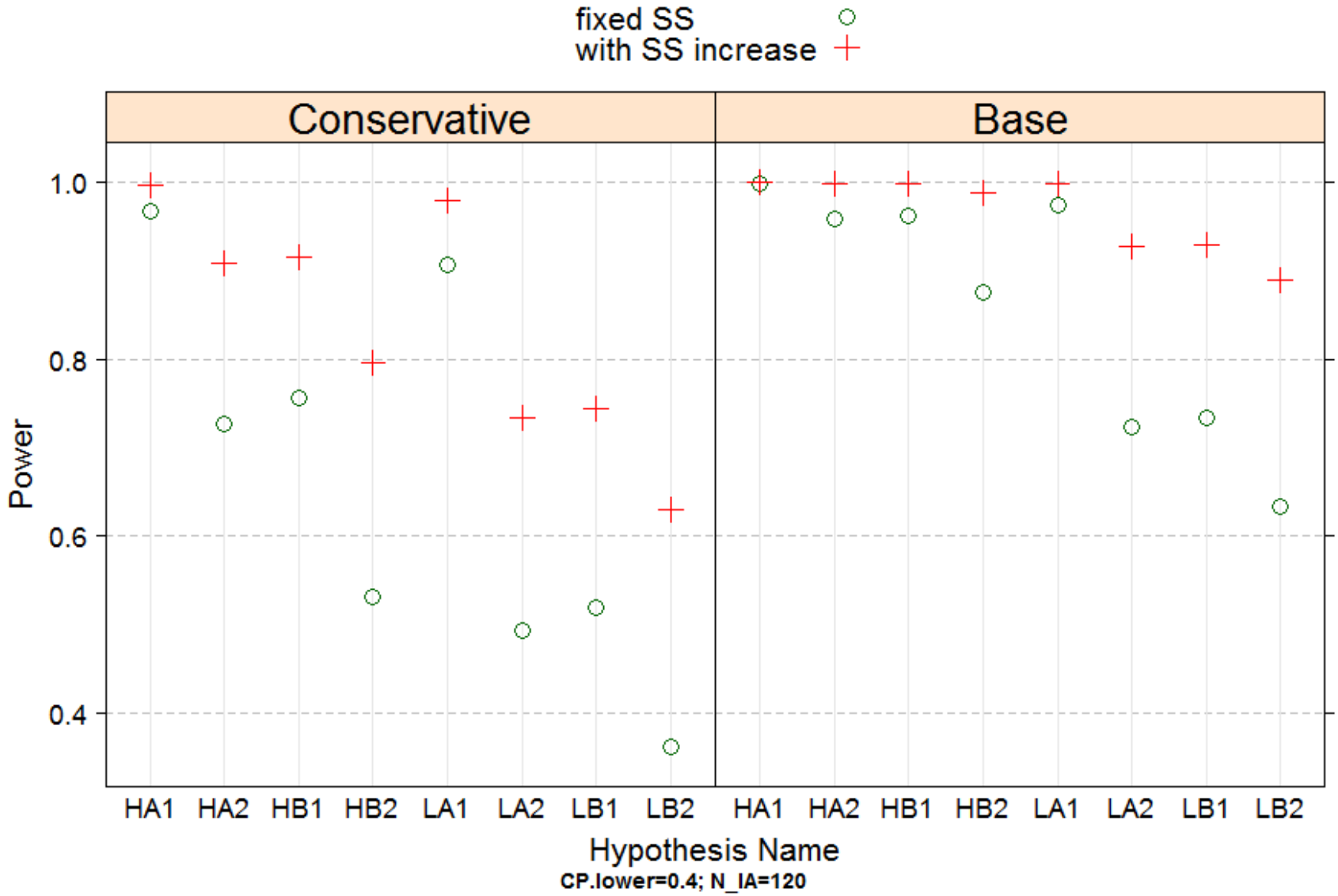
Probability that Sample Size is Increased



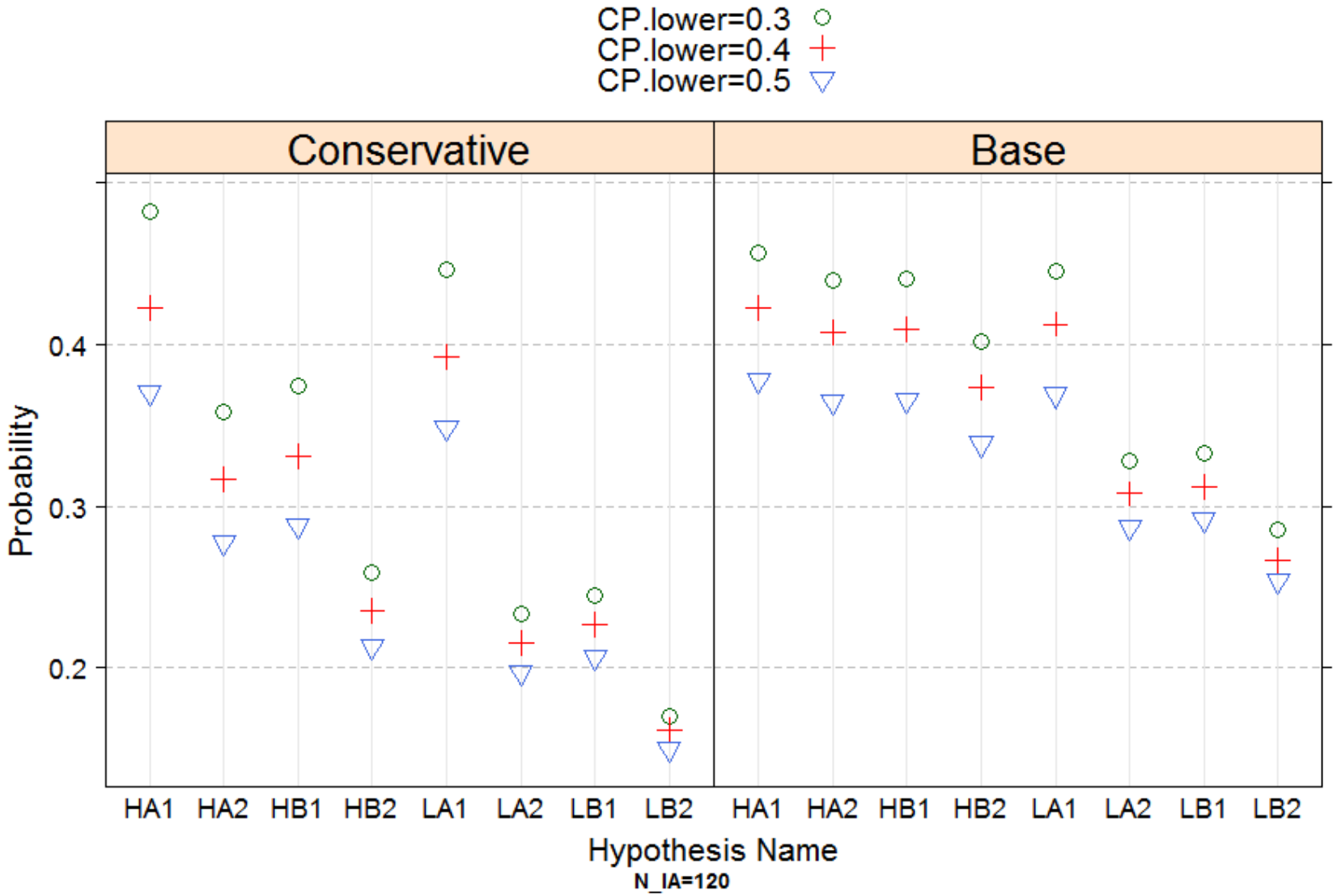
Power Individual Hypotheses for Fixed and SSR Designs (unconditional)



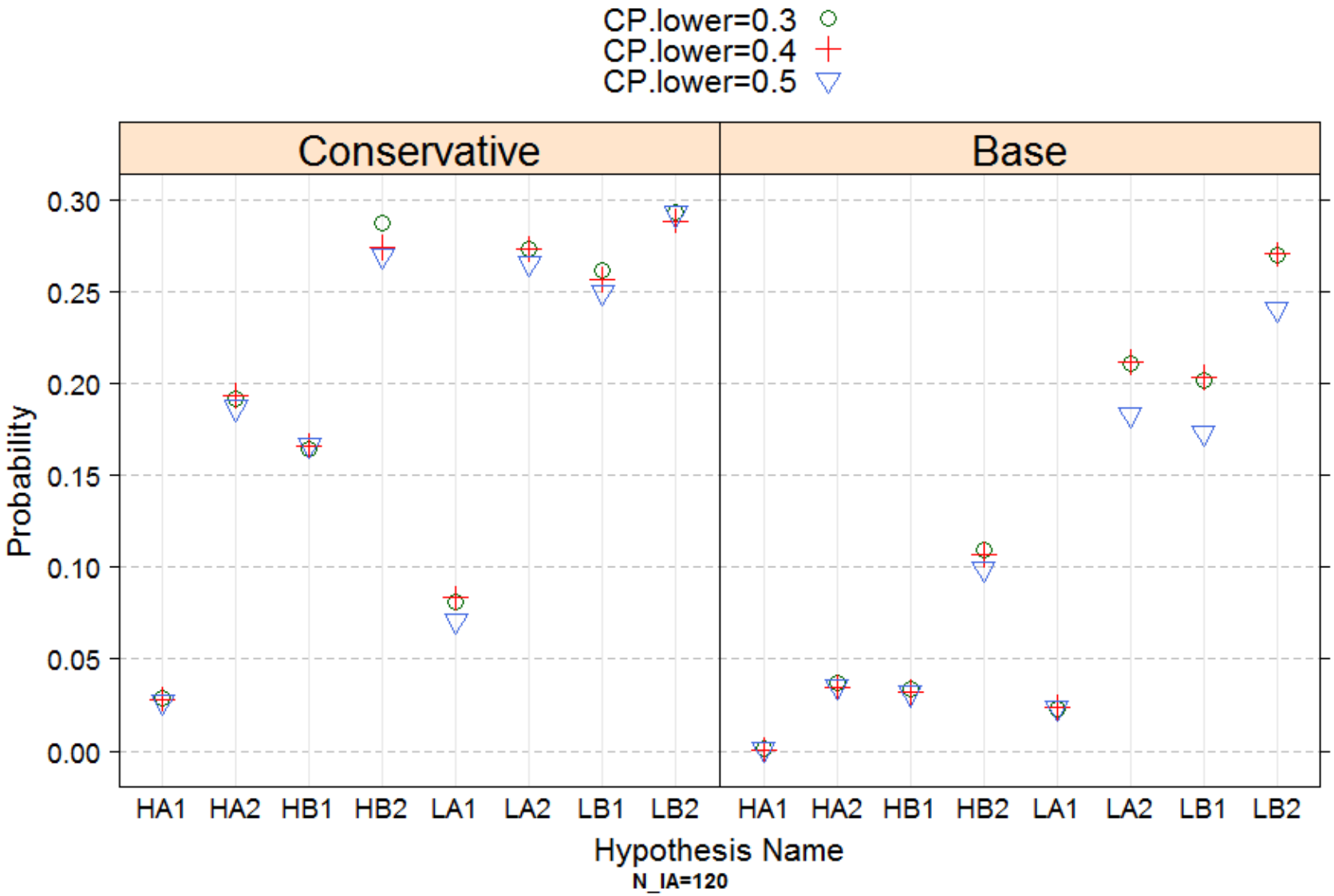
Power Individual Hypotheses for Fixed and SSR Designs (conditional on SS increase Rule at IA)



Probability of "Unnecessary" Sample Size Increase



Probability of Gain (Conditional)



Summary Regarding SSR Aspect

- SSR rule seeks to increase sample in cases when interim results are warranted
  - Minimize chances of SS increase if drug is not efficacious
  - Maximize chances of SS increase if drug is sub-efficacious
- Adjusted parameters accordingly
  - Set CP.lower = 0.4
  - Rule is less sensitive to the timing of the IA. (N_{IA}=100, 120 adequate)
- Application of the SS increase rule results in power improvement compared to the fixed design, especially in cases when SS increase rule is triggered

Conclusions

- Considered case study where gatekeeping procedure integrated into SSR design
  - Theoretically justified strong control of Type I error
  - Verification of assumptions involved in MT procedures is required in a presence of SSR
- Clinical trial simulations to
  - Evaluate operation characteristics
  - Account for special study features
    - Hypotheses, Effect sizes, Correlation among EPs
  - Optimize with respect to
    - Multiplicity strategy, sample size, timing of IA, SSR rules

References

- Dmitrienko A, Tamhane AC. (2011). Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. Statistics in Medicine. 30, 1473-1488.
- Kordzakhia G, Dmitrienko A. (2013). Superchain procedures in clinical trials with multiple objectives. Statistics in Medicine. 32, 486-508.
- Cui L, Hung HM and Wang SJ (1999) Modification of sample size in group sequential clinical trials. Biometrics 55
- Wong J (2010). Many-to-one comparison after sample size reestimation for trials with multiple treatment arms and treatment selection. Journal of Biopharmaceutical Statistics, 20: 927-940
- Mehta CR and Pokock SJ. (2010) Adaptive Increase in Sample Size when Interim Results are Promising: A Practical Guide with Examples. Statistics in Medicine