Biopharmaceutical Applied Statistics Symposium (BASS) XX

6-Nov-2013, Orlando, FL

Decision Making in Confirmatory Multipopulation Tailoring Trials

Brian A. Millen, Ph.D.

Lilly

Answers That Matter.

Acknowledgments

Alex Dmitrienko

Steve Ruberg

Lei Shen

Guochen Song

Zongjun (John) Zhang

Dominique Williams

Sumithra Mandrekar

Fangyi Zhao

Laura Nisenbaum

Outline

Motivation / Statement of Problem

Analysis Considerations

- Testing Considerations (FWER control)
- Other Inference Considerations
  - Influence Condition (Influence Error rate control)
  - Interaction Condition (Interaction Error rate control)

Design Impact

Summary / Closing Comments

Background

Continuum of Approaches to Clinical Trials

Trials in Overall Population
-- exploratory subgroup analyses

Single population Tailoring trials

Multipopulation Tailoring trials

Clinical context

Marker of response

- Identifies association
- Usually not causative
- Imperfect predictor (but valuable)
  - Drug effect present in marker-positive and marker-negative subgroups, with reduced magnitude of effect in marker-negative subgroup
  - FDA draft guidance on enrichment strategies: “experience suggests that the selected [markers] often do not precisely dichotomize [the overall population] into subpopulations that will and will not respond.”

Motivation

Traditional Development

- Consider subpopulations only after overall population trials result in failure
- Relies on retrospective analyses of subgroups

Herceptin Example

- HER-2 expression
- Single population confirmatory trials (HER-2 positive)
- Remaining question: efficacy in complementary population?
- Study NSABP B-47 initiated 14 years later

Motivation

Efficiency

- Single study: multiple populations, rather than multiple studies each addressing single population

More informative

- vs. subpopulation-only trials
- FDA guidance on enrichment strategies

Treatment Registration “wants”

- Overall population indication, with enhanced labeling with info on subpopulation effects
- Simple overall population indication or restricted (“tailored”) subpopulation indication, if data warrant

Analysis Considerations:

MULTIPLE TESTING

Multiplicity

Notation

- Overall Population, $\mathcal{O}$
- Pre-defined Subpopulation, $\mathcal{G}+$

  where $\mathcal{O} = \mathcal{G}+ \cup \mathcal{G}-$

FWER control

- $H_{\mathcal{O}}$: no effect in overall population
- $H_{\mathcal{G}+}$: no effect in pre-defined (marker +) subpopulation

Successful outcome if either null hypothesis is rejected.

Choice of Multiple Testing Procedure

Guiding Principles

- Logical Relationships of hypotheses
  - "interchangeable"
  - Importance weights
- Performance of procedure
  - Account for positive correlation of test stats for $H_{\mathcal{O}}$ and $H_{\mathcal{G}_+}$ (correlation is known: function of overlap of pops)

Dmitrienko et al (2011)

Choice of Multiple Testing Procedure

~~Fixed Sequence~~

Bonferroni-based

- Simple
- Fallback
- Holm (cyclical chain)

"Semiparametric"

- Hochberg/Hommel

Feedback

Parametric Chain

Holm (1977); Bretz et al (2009); Zhao et al (2010); Millen et al (2011)

Chain Procedures

Governed by 2 sets of parameters:

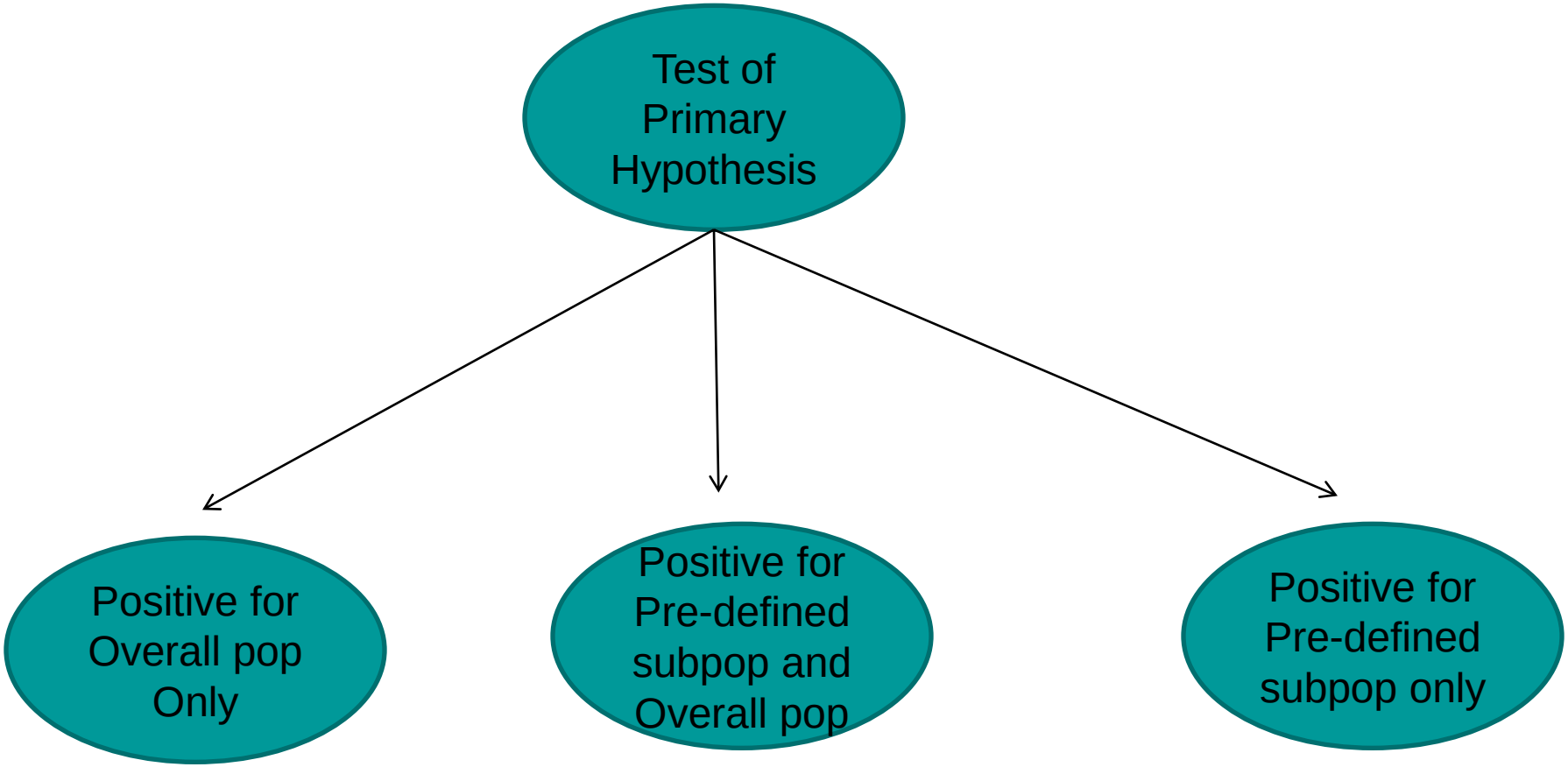
- Hypothesis weights (for initial alpha allocation)
- Transition matrix (for alpha propagation)
- “ordering” set in serial chain procedures; data-driven in cyclical chain procedures.

Parameters may be chosen to optimize performance metrics such as

- probability of rejecting at least one null hypothesis
- probability of rejecting subpopulation null hypothesis, given failure to reject overall population null hypothesis
- Probability of rejecting both null hypotheses

Millen and Dmitrienko 2011; Bretz et al 2009

Multiple Testing Outcomes



Influence Condition

ADDITIONAL CONSIDERATIONS

Overall Population Efficacy

It’s possible to achieve statistical significance in the overall population $\mathcal{O}$ when the effect exists only in the predefined subpopulation $\mathcal{G}+$

Rothmann et al (2012) show that the rate of these errors can be quite high in some scenarios

Appropriate inference should minimize the rate of these “influence errors”

Influence Condition

Application of the influence condition provides control of the influence error rate.

An **influence error** is a conclusion of treatment benefit for the overall population when, in fact, there is no beneficial effect in the complementary subpopulation.

Influence Condition

Principle:

In order to support a claim of effectiveness in the overall population, the beneficial effect must not be limited to only the pre-defined subpopulation.

- otherwise the pre-defined subpopulation exerts undue influence on the overall population effectiveness conclusion

Assessment

- Simple frequentist estimation.
- Bayesian posterior probability calculation.

Application of Influence Condition

Relative size of pre-defined subpopulation	Influence error rate (%)			
	Primary Hypothesis tests only (without influence condition)	Influence condition with λ_{INF}=1.1	Influence condition with λ_{INF}=1.2	Influence condition with λ_{INF}=1.3
Scenario 1: HR=1 (complementary subpopulation) and HR=1.5 (pre-defined subpopulation)				
0.75	80.28	31.34	19.40	11.72
0.5	45.68	22.98	11.64	4.20
0.25	12.92	11.46	6.02	1.88
Scenario 2: HR=1 (complementary subpopulation) and HR=2 (pre-defined subpopulation)				
0.75	99.80	32.04	18.96	10.22
0.5	88.04	25.96	11.14	3.98
0.25	31.58	19.64	6.56	1.84

Comments

Frequentist Approach

- Based on simple point estimation
  - Uncertainty around estimate not factored in directly
    - Operating characteristics for any rule may be evaluated
    - Should be done for study design and at study analysis/inference
  - Reflects estimates which are included in labeling
- Relies on a single parameter
  - Clinical relevance
  - Decision risk tolerance

Bayesian approach

Bayesian formulation

- Directly assesses likelihood (given the data) of positive effect in the marker negative subpop
  - supports decision-making
  - Posterior probabilities not currently included in labeling
- Uses two separate parameters
  - Clinical relevance
  - Decision threshold (risk tolerance)
- Can be more computationally intensive
  - formulations available to make computations simple
  - Availability of software to handle estimating posteriors

Bayesian Approach

Using conjugate priors and modest assumptions, closed form solutions are available for the posterior probabilities.

Example:

- Normally distributed endpoint. Known variance.
- Normal priors on μ_{ij} (i=trmt 1,2; j=pop +,-).
- Then, posterior for effect size $\theta_j == (\mu_{1j}-\mu_{2j})/\sigma_j$ is readily derived.

Result: $\Pr(\theta_- \geq \lambda_1 \mid Y) = \Phi\left([\theta^*_- - \lambda_1]/\sigma^*_- \right)$

Bayesian Approach

Example: Time-to-event endpoint. Assuming normal approximation for effect size distribution (log HR), applying a normal conjugate prior results in a normal posterior with means θ^*_j and variances σ^2_j

and $\Pr(\theta_- \geq \lambda_1 \mid Y) = \Phi((\theta^*_- - \lambda_1)/\sigma^*_-)$

A similar approach is available for binary endpoints, as well.

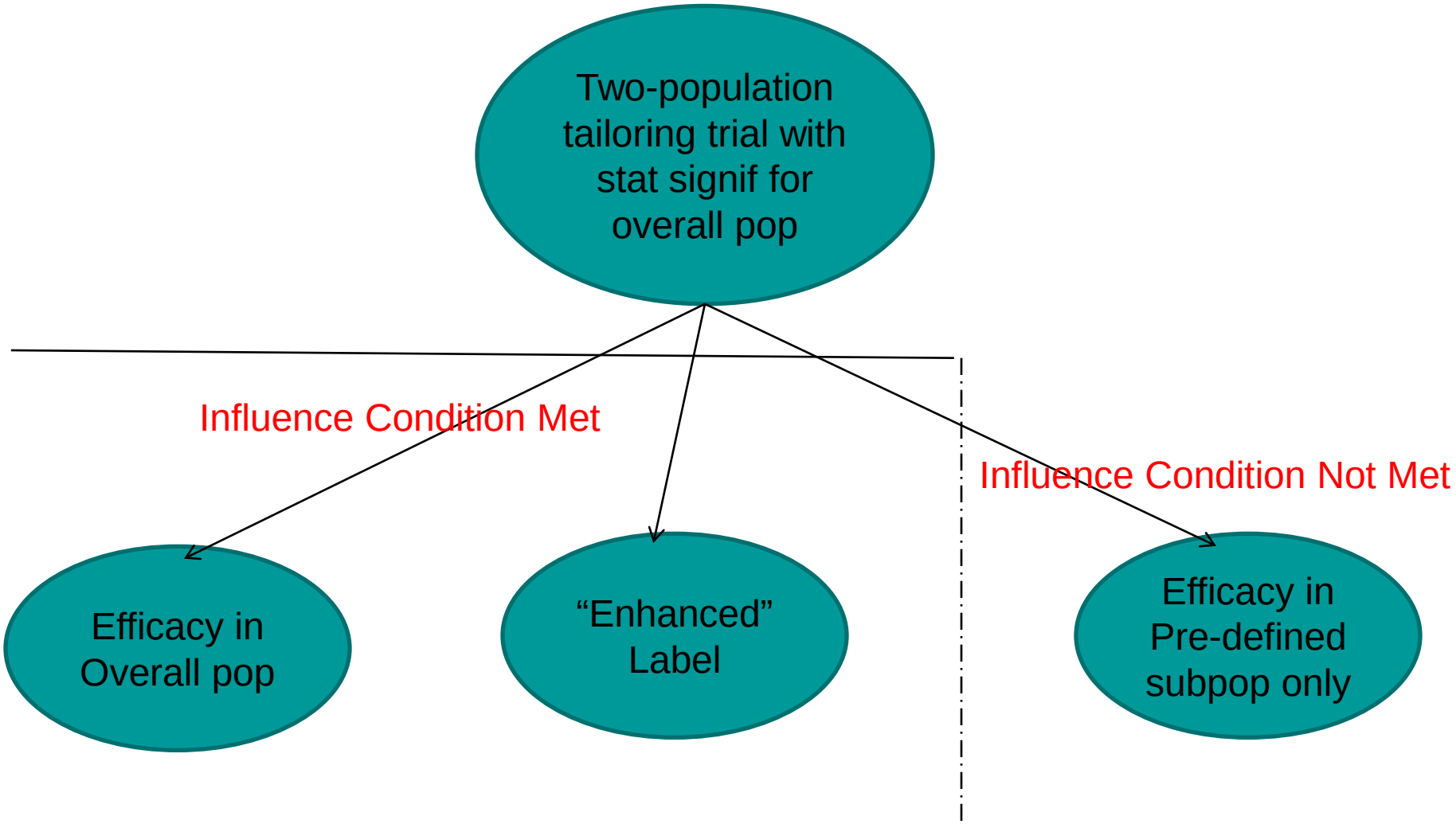
Summary – Influence Condition

For inference of overall population effect, the influence condition must be satisfied.

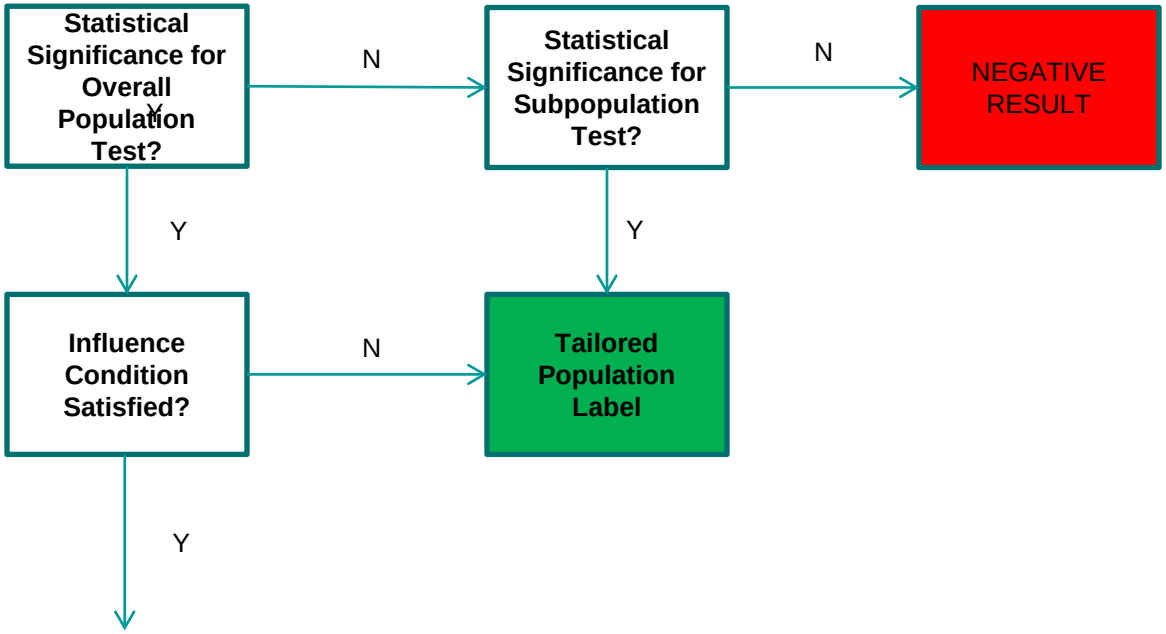
- If influence condition not satisfied, then conclude effect for predefined subpopulation only
- If influence condition is satisfied, then conclude effect for the overall population (Note: this does not mean “equal” effects across subgroups)

Evaluation methods have been proposed based on frequentist estimation of effect (point estimates; confidence intervals) or posterior probabilities of effect.

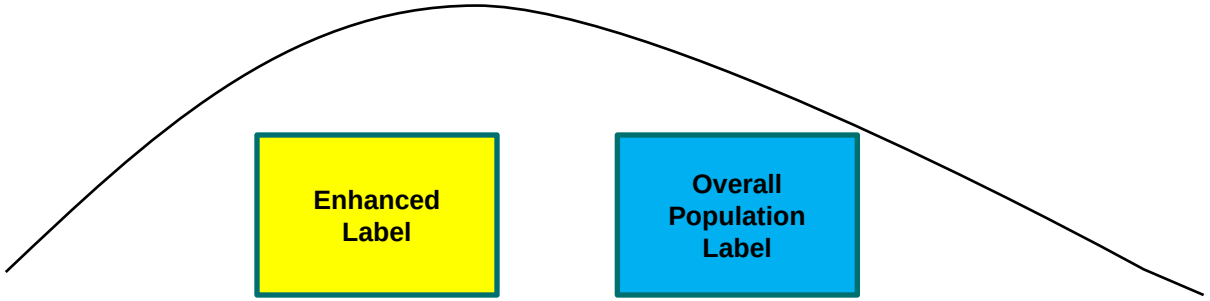
Potential Inferences



Decision Framework



How about addressing these possibilities below?



Interaction Condition

ADDITIONAL CONSIDERATIONS

Interaction Condition

Principle:

In order to achieve a claim of enhanced effect in the predefined subpopulation, along with claim of effect in the overall population, there must be a differential effect between the pre-defined and complementary subpopulations.

- otherwise the broad claim for the overall population is sufficient

Interaction Condition

Application of the interaction condition provides control of the interaction error rate.

An **interaction error** is a conclusion of differential benefit for the marker negative and positive subpopulations when, in fact, there is no difference in effect

Assessment of Interaction Condition

Frequentist

(est. effect in $\mathcal{G}_+$) / (est. effect in $\mathcal{G}_-$) > λ_F

Bayesian

Pr ((effect in $\mathcal{G}_+$ / effect in $\mathcal{G}_-$) > λ_B | Y, effect in $\mathcal{G}_-$ > λ_1)

Comments

Frequentist Approach

- Based on simple point estimates
- Single parameter to reflect clinical relevance and decision risk tolerance
- Aligns with observed effect information common in labeling

Bayesian Approach

- Conveys likelihood of differential effect to support decision making
- Conditions on the sequential evaluation process (as formulated)
- Two separate parameters reflect clinical relevance and decision risk tolerance
- Can be more computationally intensive, but closed form solutions and availability of software help overcome this limitation

Bayesian Approach

Example:

- Normally distributed endpoint. Known variance.
- Normal priors on μ_{ij} (i=trmt 1,2; j=pop +,-).
- Then, posterior for effect size $\theta_j == (\mu_{1j}-\mu_{2j})/\sigma_j$ is readily derived.

Result:

Pr ((effect in $\mathcal{G}+$ / effect in $\mathcal{G}$-) > λ_B | Y, effect in $\mathcal{G}$- > λ_1)

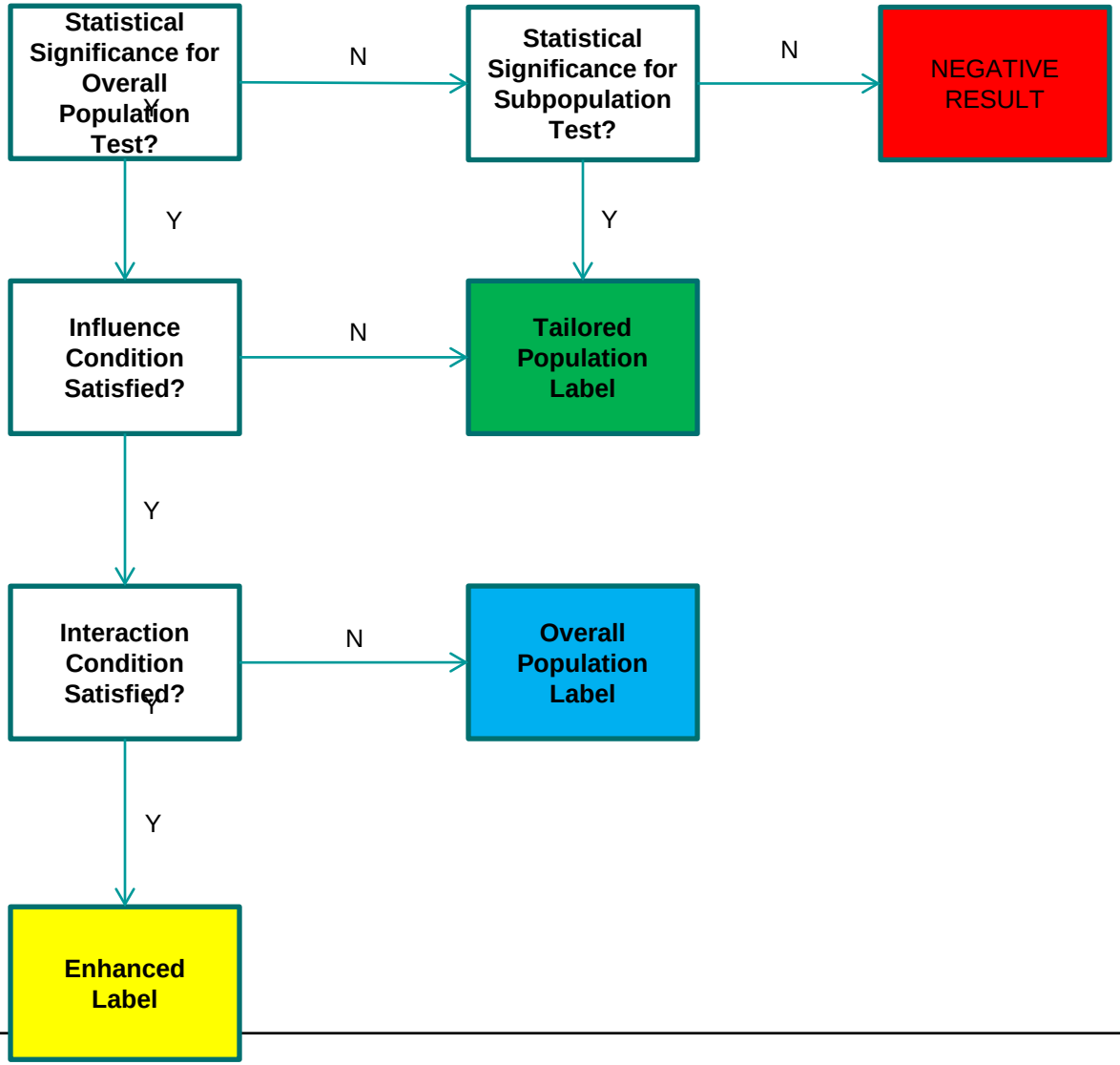
= joint probability / marginal probability

- Joint is from Biv Normal ; marginal from std. normal

Decision Framework

SUMMARY COMMENTS

Decision Framework



Design Implications

Trial design reflects the analysis plan

- Employ simulations to ensure adequate “power” to satisfy
  - Multiple testing strategy
  - Evaluation of the influence condition (for given thresholds)
  - Evaluation of the interaction condition (for given thresholds)

Specific design features for consideration

- Sample size in all relevant populations
  - Enrichment strategies
- Stratification by marker status

Summary Comments

Heterogeneity of effects/response within populations exists.

- Understanding heterogeneity is the goal
- When knowledge of predictors of varied effect exists, this should be available for patients/prescribers
- The presence of heterogeneity does/should not mean the absence of treatment availability for a broad population

Summary Comments

Multipopulation Tailoring Trials offer efficient clinical development in the presence of potential markers of efficacy and may accelerate patient access to tailored therapies and “informative” labels.

Inference in these trials is more complex than in single population trials

- Multiple testing procedures and Supportive Analyses are needed to control potential errors.

The decision framework presented supports clinically relevant inference based on these trials and enables transparent discussion across disciplines and between stakeholders.

Limitations / Future Work

Trial level vs. program level inference considerations

Extensions beyond 2-population trials

Recommendations for decision thresholds

References

Bretz F, Maurer W, Brannath W, and Posch M (2009) “A Graphical Approach to Sequentially Rejective Multiple Test Procedures.” Statistics in Medicine.

Dmitrienko A, Millen BA, Brechenmacher T, Paux G (2011) “Development of Gatekeeping Strategies in Confirmatory Clinical Trials.” Biometrical Journal.

Holm procedure

Guidance for Industry: Enrichment Strategies for Clinical Trials to support Approval of Human Drugs and Biological Products. FDA 2012.

Millen BA, Dmitrienko A, Ruberg S, Shen L. A Statistical Framework for Decision Making in Confirmatory Multipopulation Tailoring Clinical Trials. Drug Inf Journal. 2012; 46(6): 657-656.

Millen, B., Dmitrienko, A. (2011). Chain testing procedures: A class of flexible closed testing procedures with clinical trial applications. Statistics in Biopharmaceutical Research.

References

Millen BA, Dmitrienko A, Mandrekar SJ, Zhang Z, Williams D. Multipopulation Tailoring Clinical Trials: Design, Analysis and Inference Considerations. Ther Innov & Reg. Sci. 2013; *to appear*

Millen BA, Dmitrienko A, Song G. Bayesian Assessment of the Influence and Interaction Conditions in Multipopulation Tailoring Trials. J Biopharm Stat. *to appear.*

Rothmann MD, Zhang JJ, Lu L, Fleming TR. Testing in a Prespecified subgroup and the Intent-to-treat Population. Drug Inf Journal. 2012; 46(2): 175-179

Wiens B. A Fixed Sequence Bonferroni Procedure for Testing Multiple Endpoints. Pharm Stat. 2003; 2: 211-215.

Zhao YD, Dmitrienko A, Tamura R. Design and Analysis Considerations in Clinical Trials with a sensitive subpopulation. Stat in Biopharm Res. 2010.; 2: 72-83.