Decision Making in Multi-population Tailoring Trials

Brian A. Millen, Ph.D.
Outline

Background / Motivation
Testing Considerations
Influence Condition
Interaction Condition
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
Motivation

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

Traditional Development

• Consider subpopulations only after overall population trials result in failure
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
Inferential Outcomes

Test of Primary Hypothesis

- Positive for Pre-defined subpop only
- Positive for Pre-defined subpop and Overall pop
- Positive for Overall pop Only
Considerations for Inference

Multiplicity
• Type I error rate control

Influence Condition
• Influence error rate control

Interaction Condition
• Interaction error rate control
**Multiplicty**

**Notation**
- Overall Population, $O$
- Pre-defined Subpopulation, $G^+$
  
  where $O = G^+ \cup G^-$

**FWER control**
- $H_O$: no effect in overall population
- $H_{G^+}$: no effect in pre-defined (marker +) subpopulation

Successful outcome if either null hypothesis is rejected.
## Choice of Multiple Testing Procedure

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonferroni</td>
<td>Feedback</td>
</tr>
<tr>
<td>Hochberg</td>
<td>Parametric Fallback</td>
</tr>
<tr>
<td>Hommel</td>
<td>Parametric Chain Procedures</td>
</tr>
<tr>
<td>Holm</td>
<td>Bonferroni-based Chain Procedures</td>
</tr>
<tr>
<td>Fallback (Bonferroni-based)</td>
<td>Etc.</td>
</tr>
</tbody>
</table>
Guiding Principles

• Logical Relationships of hypotheses
  – “interchangeable”
  – Importance weights

• Performance of procedure
  – Account for positive correlation of test stats for $H_O$ and $H_{G+}$
    (correlation is known: function of overlap of pops)

• Examples
  – Parametric chain procedures
  – Parametric fallback procedures
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.
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.

Assessment

- Simple frequentist estimation. Effect size estimate in $G_\epsilon > \epsilon$
- Bayesian estimation. $Pr(\text{effect size in } G_\epsilon > \epsilon)$
Influence Condition

Rothmann et al (2012) showed that in some cases, influence error rates can be quite high.

Error rates depend on the relative size of the predefined subpopulation and the magnitude of effect in that subpopulation.
Clinical trial comparing treatment and control. N=300/arm. Applying fallback test with weights of 0.8 to overall pop and 0.2 to pre-defined subpop. Pre-defined subpop e.s. = 0.6; no effect in complementary subpop.

<table>
<thead>
<tr>
<th></th>
<th>Without influence condition</th>
<th>Influence condition ($\varepsilon = 0$)</th>
<th>Influence condition ($\varepsilon = 0.15$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without influence</td>
<td>93.86%</td>
<td>49.6%</td>
<td>9.9%</td>
</tr>
<tr>
<td>condition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influence condition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($\varepsilon = 0$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Influence condition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>($\varepsilon = 0.15$)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Influence Condition

Very simple application of the (frequentist) influence condition allows control of influence errors.

A Bayesian application allows conveyance of likelihood of effect in the complementary subpopulation
   – Readily estimated via MCMC or utilizing closed form solutions based on conjugate priors
Interaction Condition

Principle:

In order to achieve an enhanced claim of 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

• Frequentist: \((\text{est. effect in G+}) / (\text{est. effect in G-}) > \lambda_F\) ?
• Bayesian: \(\Pr((\text{effect in G+}) / (\text{effect in G-}) > \lambda_B) = ?\)
**Decision Framework**

- **Statistical Significance for Overall Population Test?**
  - **Y**
  - **N** ➔ **Statistical Significance for Subpopulation Test?**
    - **Y** ➔ NEUTRAL RESULT
    - **N** ➔ **Tailored Population Label**
  - **N** ➔ **Enhanced Label**

- **Influence Condition Satisfied?**
  - **Y**
  - **N** ➔ **Statistical Significance for Subpopulation Test?**
    - **Y** ➔ NEUTRAL RESULT
    - **N** ➔ **Tailored Population Label**

- **Interaction Condition Satisfied?**
  - **Y**
  - **N** ➔ **Overall Population Label**

- **N** ➔ **Overall Population Label**

Decision labels:
- NEUTRAL RESULT
- NEUTRAL RESULT
- NEUTRAL RESULT
- Enhanced Label
- Tailored Population Label
- Overall Population Label
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
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
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.

Multiple testing procedure and frequentist and/or Bayesian estimation procedures readily facilitate inference from these trials.

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


BACKUP
References


