Case Studies of Applying Multiple Testing Procedures in Neuroscience Late-Phase Clinical Trials

Haiyan Xu, Ph.D.
Janssen R&D, Johnson & Johnson

Acknowledgement: Yevgen Tymofyeyev, Ph.D. (Janssen) (major contribution to case studies and simulations)
Problems we face

One experiment

- Multiple dose comparisons
- Multiple endpoints
- Multiple subpopulations
- Non-inferiority + superiority
- Multiple looks, etc

Generate multiple hypotheses: H1, H2, H3, H4, etc

FDA, EMA, other Health authorities

Require controlling study-wise type I error at $\alpha$ level (two-sided 5%, or one-sided 2.5%)
{We often call this familywise error rate (FWER)}

Need to design efficient multiple testing procedures (MTPs)

(6% power difference, ~100 additional subjects, ~200–300 additional subjects in a program, e.g., ~$50k per subject, $10–15MM increase)
My *Naïve* Classification of Multiple Testing Problems

**Class 1:** No prefixed testing sequence of hypotheses

(“Decision paths”)

**Class 2:** Simple (or relative simple) ordering

**Class 3:** More complex hierarchical structure (or “decision paths”)

*E.g., Multiple sources of multiplicity*
  - Multiple doses + multiple endpoints
  - Multiple populations + multiple endpoints
  - Etc

**Class 4:** multiplicities in group sequential/ adaptive design setting

*(will not discuss in detail today)*
Tools for **Class 1 Problem**: No prefixed testing sequence

- Bonferroni (single-step)
  - stepwise
- Holm’s step-down
- Hochberg step-up
- Closed Test, Partition Test *(without restrictions)*

- Distribution assumption
  - Dunnett
  - Sidak, etc
  - Resampling-based (e.g., permutation)
  - Stepwise Dunnett, etc

- Stepwise distribution assumption
Tools for **Class 2 Problem**: Simple (or relatively simple) ordering

- **Fixed sequence**
- **Original Fallback** (Not recommended)
- **Improved fallback**

\[ \alpha_1 + \alpha_2 + \alpha_3 = \alpha \]
Tools for **Class 3 Problem**: More complex hierarchical structure

**Gatekeeping Procedures**
- Serial Gatekeeping
- Parallel Gatekeeping
- Tree Structured Gatekeeping, General Mixture Gatekeeping, etc

**Partitioning Decision Paths Approach** (Liu and Hsu [JASA 2009])

Follow the **Decision Path Principle**: Null hypotheses should be formulated so that decision making naturally follows logical paths.

**Closed Test, Partition Test** *(with restrictions)*
Case Study #1: schizophrenia/bipolar disorder in ph3

- **Primary hypothesis:** Compound X improves symptoms vs placebo as measured by a symptom scale (e.g., PANSS [Positive and Negative Syndrome Scale for Schizophrenia], YMRS for Bipolar disorder)

- **Key secondary hypothesis:** Compound X improves functioning vs placebo based on a functioning scale (e.g., PSP [Personal and Social Performance Scale] for schizophrenia, GAF for Bipolar disorder)

- **4 arms:** placebo, 3 doses of the new treatment
Case Study #1: schizophrenia/bipolar disorder in ph3 (cont’d)

**Dose**

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary</strong></td>
<td>$H_{PE}$</td>
<td>$L_{PE}$</td>
</tr>
<tr>
<td><strong>Secondary</strong></td>
<td>$H_{SE}$</td>
<td>$L_{SE}$</td>
</tr>
</tbody>
</table>

Potential strategy: test primary endpoint first, if both doses are rejected, then test secondary (so-called **serial gatekeeping**).

However, it is of great interest to show effect in one dose, but in both endpoints.
Case Study #1: schizophrenia/bipolar disorder in ph3 (cont’d)

- Initially, Dunnett–based parallel gatekeeping testing procedure was proposed, based on Dmitrienko et al (2006)
- There was concern regarding utilizing sample based correlation between endpoints
- Dunnett–Bonferroni–based parallel gatekeeping procedure was utilized, based on Xu et al (2009)
A closed testing based procedure, with 6 individual hypotheses

- An alternative approach: Partitioning Decision Path (Liu and Hsu [JASA 2009])

<table>
<thead>
<tr>
<th>Intersection hypotheses</th>
<th>Decision rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{111111}, H_{111101}, H_{111011}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^3 &gt; c_1$ or $T_S^1 &gt; c_1$</td>
</tr>
<tr>
<td>$H_{111100}, H_{110111}, H_{111010}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^2 &gt; c_1$</td>
</tr>
<tr>
<td>$H_{110110}, H_{110010}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^2 &gt; c_1$</td>
</tr>
<tr>
<td>$H_{101101}, H_{101011}, H_{101001}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^3 &gt; c_1$ or $T_S^2 &gt; c_2$</td>
</tr>
<tr>
<td>$H_{011101}, H_{011011}, H_{010111}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^2 &gt; c_1$ or $T_S^3 &gt; c_2$</td>
</tr>
<tr>
<td>$H_{101111}, H_{101110}, H_{101101}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^3 &gt; c_1$ or $T_S^3 &gt; c_3$</td>
</tr>
<tr>
<td>$H_{100111}, H_{100110}, H_{100101}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^2 &gt; c_1$ or $T_S^3 &gt; c_3$</td>
</tr>
<tr>
<td>$H_{001111}, H_{001110}, H_{001101}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^3 &gt; c_1$ or $T_S^3 &gt; c_3$</td>
</tr>
<tr>
<td>$H_{000111}, H_{000110}, H_{000101}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^3 &gt; c_1$ or $T_S^3 &gt; c_3$</td>
</tr>
<tr>
<td>$H_{000110}, H_{000101}, H_{000011}$</td>
<td>$T_P^1 &gt; c_1$ or $T_P^3 &gt; c_1$ or $T_S^3 &gt; c_3$</td>
</tr>
</tbody>
</table>
Case Study #2: Neurodegenerative

- **Primary Hypothesis**: Compound X slows cognitive decline vs placebo as measured by PACC
- **Key Secondary Hypothesis**: Compound X improves cognitive function and performance vs placebo based on CFI
- **3 arms**: placebo, low dose and high dose of new treatment
Fixed Sequence Testing

\[ \alpha = 5\% \text{ used in testing sequence until fail a hypothesis} \]

“Sequential by Dose”
Parallel Gatekeeping

Truncated Hochberg

Truncated Holm’s

Hochberg test for CFI family

Truncation parameter ($\gamma$): ensure that some fraction of $\alpha$ is left for CFI even if only one dose is significant for PACC
Power, at Least One Dose superior to PLB (Primary EP)

- Eff.size H dose = 0.21
- Same effect size for Primary and Secondary EP

- No uniformly most powerful procedure is available.
- Power results depends on the underlying true scenario
  - Need to decide which scenario are more likely and optimize MTP accordingly
Case Study #3: indication not revealed, non-inferiority and superiority + multiple endpoints

- **Primary objective**: establish non-inferiority on primary endpoint, of two doses of Compound X versus active control
- **1st Key secondary objective**: compare the effects of two doses of Compound X versus active control on cognitive function
- **2nd Key secondary objective**: establish superiority on primary endpoint, of two doses of Compound X versus active control
Primary, **non-inferiority on primary endpoint**, with NI margin 5

Key Secondary, **superiority on cognitive function**

**Superiority on primary endpoint**

\[ \gamma \text{ -- controls } \alpha \text{ transferred to primary endpoint superiority} \]
### Power by initial weight allocated to primary endpoint NI high dose

<table>
<thead>
<tr>
<th>Initial Weight for NI on high dose</th>
<th>NI margin/SD</th>
<th>PE true delta high dose/SD</th>
<th>PE true delta low dose/SD</th>
<th>Delta Cognitive high dose, effect size</th>
<th>delta Cognitive low dose, effect size</th>
<th>Power PE NI high dose</th>
<th>Power PE NI low dose</th>
<th>Power Cognitive high dose</th>
<th>Power Cognitive low dose</th>
<th>Power PE superiority high dose</th>
<th>Power PE superiority low dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.864</td>
<td>0.623</td>
<td>0.748</td>
<td>0.609</td>
<td>0.335</td>
<td>0.227</td>
</tr>
<tr>
<td>0.75</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.884</td>
<td>0.607</td>
<td>0.782</td>
<td>0.592</td>
<td>0.338</td>
<td>0.228</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.903</td>
<td>0.546</td>
<td>0.815</td>
<td>0.538</td>
<td>0.343</td>
<td>0.230</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.887</td>
<td>0.878</td>
<td>0.788</td>
<td>0.859</td>
<td>0.487</td>
<td>0.489</td>
</tr>
<tr>
<td>0.75</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.893</td>
<td>0.863</td>
<td>0.798</td>
<td>0.841</td>
<td>0.487</td>
<td>0.489</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.903</td>
<td>0.736</td>
<td>0.815</td>
<td>0.725</td>
<td>0.493</td>
<td>0.491</td>
</tr>
</tbody>
</table>

PE: primary endpoint
Team’s Inputs: Significant Impact on MTP design

- From commercial and regulatory point of view, do we have to have more than two endpoints? And if we do, how do we order their importance?
  - impact on sample size
- Is there a clear dose response to order the arms? How confident are we in terms of betting on a strict order, or a particular arm?
  - impact on simulation and which MTP to choose and also the sample size
- What is the reasonable effect size for each arm and each endpoint?
  - impact on simulation results and hence the decision of which MTP to choose and also the sample size
- How do we define “win”: win on at least one dose, or win on at least two doses?
  - impact on power simulation and hence sample size
- Is there preference between winning on a particular regimen vs. another?
  - impact on allocating weights and redistribution weights, and hence sample size

Objectives, priorities and assumptions imply preferred MT strategies
- Needs to be supported by power simulation under different scenarios
Team Interaction

- Train other functions
  - e.g., Multiple testing workshop – introducing the methods and impact, with real case studies
- Understand Target Product Profile (e.g., certain advantage vs. competitor drug) and get involved in strategic level discussion
- Meta analyses (internal + external vendor) for design assumptions
- Comprehensive simulations with easy-to-understand data display
In Summary

- We encounter many multiple testing problems
- Important to partner with other functions to come up with an efficient MTP, which also aligns with the development strategy (priorities)
- Methods evolving over time:
  - Case 1: Dunnett/Dunnett–Bonferroni gatekeeping: “left alpha on the table”
  - Case 2: Alpha exhaustive, but did not “recycle” alpha back to higher-level families
  - Case 3: Alpha exhaustive, and “recycle” alpha back to higher-level families
- New challenges:
  - Subgroup in confirmatory setting (how to deal with joint distribution for non-continuous outcome)
    - Ding et al (2016); Lin et al (2018), to appear
  - Multiplicity adjustment while searching for subgroups
  - Move towards confidence intervals
References


Rohmeyer, Klinglmueller (2011) gMCP: Graphical approach to multiple comparison procedures.

