Choice of Multiple Comparison Procedure in Two Pivotal Clinical Trials for Approval of a New Pharmaceutical Product: Power and Aesthetics

Brian L. Wiens
Alcon Laboratories, Inc.
Alex Dmitrienko
Eli Lilly and Co.

Background

• Commonly need (at least) two clinical trials to support a claim
• Complex multiple comparison procedures might be difficult to replicate
• We studied replication
  – Two identical clinical trials
  – Various MCPs
  – Power and consistency
MCPs Considered

• Fixed sequence (Westfall and Krishen, 2001)
• Fallback (Wiens, 2004)
• Weighted Holm (Holm, 1979)
  – Weights of $w_i$ for fallback and weighted Holm

Fixed Sequence

• Test hypotheses in prespecified order
• Test first hypothesis at $\alpha$
  – If fail to reject, stop
  – If reject, test second hypothesis at $\alpha$
    • If fail to reject, stop
    • If reject, test third hypothesis at $\alpha$
• Etc.
fallback

• Allocate $\alpha$ across hypotheses: $\alpha_i'$
• Test hypotheses in prespecified order
• Test first hypothesis at $\alpha_1'=\alpha_1$
  – If fail to reject, test second hypothesis at $\alpha_2'=\alpha_2$
  – If reject, test second hypothesis at $\alpha_2'=\alpha_1'+\alpha_2$
    • If fail to reject, test second hypothesis at $\alpha_3'=\alpha_3$
    • If reject, test third hypothesis at $\alpha_3'=\alpha_2'+\alpha_3$
• Etc.

unweighted holm

• Allocate $\alpha$ across hypotheses: $\alpha_i'$
• Test in adaptive order: smallest p-value first
  – If $p(1) > \alpha/n$, fail to reject associated hypothesis and stop testing
  – If $p(1) < \alpha/n$, reject hypothesis and test next
    • If $p(2) > \alpha/(n-1)$, fail to reject associated hypothesis and stop testing
    • If $p(2) < \alpha/(n-1)$, reject hypothesis and test next
• Etc.
• Weighted version also available
Closed Testing Procedure

- Consider all non-empty subsets of individual hypotheses
  - \(2^n - 1\) subsets for \(n\) hypotheses
  - Each subset tested globally at \(\alpha\) level
- Reject an individual hypothesis if all subsets containing that hypotheses are rejected
  - Controls FWE rate at \(\alpha\) level (Marcus et al, 1976)
- Alpha- exhaustive: every subset tested at full \(\alpha\)

### Closed Testing Procedure - Illustration

<table>
<thead>
<tr>
<th>Intersection Hypothesis</th>
<th>Individual hypotheses</th>
<th>Weighted Holm (\alpha)-exhaustive fallback</th>
<th>Fixed sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(H_1, H_2, H_3)</td>
<td>(p_1 &lt; w_1 \alpha) or (p_2 &lt; w_2 \alpha) or (p_3 &lt; w_3 \alpha)</td>
<td>(p_1 &lt; \alpha)</td>
</tr>
<tr>
<td>2</td>
<td>(H_1, H_2)</td>
<td>(p_1 &lt; w_1/(w_1+w_2) \alpha) or (p_2 &lt; w_2/(w_1+w_2) \alpha)</td>
<td>(p_1 &lt; \alpha)</td>
</tr>
<tr>
<td>3</td>
<td>(H_1, H_3)</td>
<td>(p_1 &lt; w_1/(w_1+w_3) \alpha) or (p_2 &lt; w_2/(w_1+w_3) \alpha)</td>
<td>(p_1 &lt; \alpha)</td>
</tr>
<tr>
<td>4</td>
<td>(H_2, H_3)</td>
<td>(p_2 &lt; w_2/(w_2+w_3) \alpha) or (p_3 &lt; w_3/(w_2+w_3) \alpha)</td>
<td>(p_2 &lt; \alpha)</td>
</tr>
<tr>
<td>5</td>
<td>(H_1)</td>
<td>(p_1 &lt; \alpha)</td>
<td>(p_1 &lt; \alpha)</td>
</tr>
<tr>
<td>6</td>
<td>(H_2)</td>
<td>(p_2 &lt; \alpha)</td>
<td>(p_2 &lt; \alpha)</td>
</tr>
<tr>
<td>7</td>
<td>(H_3)</td>
<td>(p_3 &lt; \alpha)</td>
<td>(p_3 &lt; \alpha)</td>
</tr>
</tbody>
</table>
Which Procedure is Preferred?

• For one trial:
  – Depends on situation
    • Relative importance of hypotheses
    • Relative power to reject hypotheses
    • Correlation among hypotheses
  – Depends on definition of “best”
    • Reject at least one hypothesis
    • Reject all hypotheses
    • Reject first hypothesis
  – What weights to use?
    – Wiens and Dmitrienko (2010) for discussion

Which Procedure is Preferred?

• For two trials:
  – Also consider consistency of results
    • If reject a hypothesis of one trial, should also reject in the other trial
    • Component of “aesthetics” (Hommel and Bretz, 2008)
  – Studied by simulation
    • Two identical trials
    • Three hypotheses
    • Various powers
    • Various decision criteria
First Simulation

• Power of 50%, 90% and 90% (before multiplicity adjustment) for three hypotheses
• Weights of 0.5, 0.3 and 0.2 (Holm and fallback)
• Correlation of 0, 0.2, 0.5 or 0.8 between adjacent hypotheses
Second Simulation

- Power of 90%, 90% and 90% (before multiplicity adjustment) for three hypotheses
- Weights of 0.5, 0.3 and 0.2 (Holm and fallback)
- Correlation of 0, 0.2, 0.5 or 0.8 between adjacent hypotheses
Third Simulation

- Power of 80%, 90% and 90% (before multiplicity adjustment) for three hypotheses
- Weights of 0.5, 0.3 and 0.2 (Holm and fallback)
- Correlation of 0, 0.2, 0.5 or 0.8 between adjacent hypotheses

Probability of Inconsistent Results for Third Hypothesis

![Graph showing the probability of inconsistent results for different correlation values and adjustment methods. The graph includes lines for Fixed Sequence, Fallback, and Holm methods.]
General Conclusions

• Fixed sequence has highest power for
  – Rejecting the first hypothesis
  – Rejecting all hypotheses
• Fixed sequence has lowest power for
  – Rejecting third hypothesis
• Fixed sequence has highest probability of
  – Inconsistent results
• Differences are reduced with high correlations
• Weighted Holm and fallback are nearly indistinguishable

Summary

• Many MCPs are available
• Subtle differences distinguish MCPs
  – Power
  – Consistency of results
• Consistency of results is beneficial
• Simulations will help choose the MCP for a given situation
References