Gatekeeping strategies in Phase III clinical trials with multiple endpoints and doses

Alex Dmitrienko (Quintiles, Inc)
Olga Marchenko (Quintiles, Inc)

EMA workshop on multiplicity issues in clinical trials
Nov 2012
Outline

Multiple “sources” of multiplicity in clinical trials
Multiple endpoints and multiple dose-control comparisons
Gatekeeping procedures
Methods for building gatekeeping procedures
Development of gatekeeping procedures
Multiple “sources” of multiplicity

Multiple families of null hypotheses

- Family 1: Primary hypotheses
- Family 2: Secondary hypotheses
- Family 3: Tertiary hypotheses

... Multiplicity problems with a hierarchical structure
Case study

Latuda (lurasidone) Phase III program in patients with schizophrenia

Multiple doses

Two or three doses versus placebo

Multiple endpoints

Primary endpoint: Positive and Negative Syndrome Scale (PANSS) total score at Week 6

Secondary endpoints: Clinical Global Impression-Severity (CGI-S) score at Week 6 and PANSS total score at Day 4
Case study

Multiple objectives
Multiple doses: Improve success probability
Multiple endpoints: Strengthen lurasidone product label and create differentiating factors

Gatekeeping strategy
Powerful gatekeeping procedures were developed (Brechenmacher, Xu, Dmitrienko, Tamhane, 2011)

Importance of gatekeeping procedures was recognized in clinical publication (Meltzer et al., 2011)
Other examples

Osteoarthritis program
Two dose-placebo comparisons and three endpoints (WOMAC subscale scores, PGA)

Rheumatoid arthritis program
Two dose-placebo comparisons and four endpoints (DAS-28, ACR-20, HAQ, Sharp score)
Lurasidone Phase III trial
Multiple doses
Two doses versus placebo (Dose L, 40 mg/day; Dose H, 120 mg/day)
Multiple endpoints
Primary endpoint E1 (PANSS at Week 6)
Secondary endpoint E2 (CGI-S at Week 6)
Null hypotheses

Endpoint E1

Dose L vs P \( H_1 \)

Dose H vs P \( H_2 \)

Endpoint E2

Dose L vs P \( H_3 \)

Dose H vs P \( H_4 \)

Overall Type I error rate (global familywise error rate) is controlled at two-sided \( \alpha = 0.05 \)
Gatekeeping procedures

Definition
Multiple testing procedures for multiple families of null hypotheses

Type I error rate
Control Type I error rate over multiple families

Power
Optimal distribution of power by accounting for hierarchical structure of multiple families, e.g., more power for more important tests
Gatekeeping procedures

Main classes of gatekeeping procedures

Basic gatekeeping procedures based on Bonferroni test (Bretz et al., 2009; Burman et al., 2009)

Multistage gatekeeping procedures based on Bonferroni and more powerful tests (Dmitrienko, Tamhane and Wiens, 2008)

General mixture/gatekeeping procedures based on Bonferroni and more powerful tests (Dmitrienko and Tamhane, 2011; Kordzakhia and Dmitrienko, 2012)
Principles

A. Incorporate all logical relationships among null hypotheses

B. Utilize available distributional information (joint distribution of hypothesis test statistics)

C. Select an optimal procedure (based on a relevant criterion under trial-specific assumptions)
Clinical information

Establish efficacy based first on Endpoint E1 and then on Endpoint E2

Sufficient to establish efficacy for a single dose but highly desirable to demonstrate efficacy at both dose levels

No evidence of a positive dose-response relationship
Clinical information

Dose L vs P

Endpoint E1

$H_1$

Dose H vs P

Endpoint E2

$H_3$

Take clinical information into account:

- $H_3$ depends on $H_1$
- $H_4$ depends on $H_2$
A: Logical relationships

Serial testing strategy

Endpoint E1

Dose L vs P

$H_1$

Dose H vs P

$H_2$

Endpoint E2

$H_3$

$H_4$

Inflexible strategy which is **not consistent** with clinical objectives: $H_2$ and $H_3$ cannot be tested if $H_3$ is not rejected (Hung and Wang, 2009)
Gatekeeping procedure 1

\[ \alpha/2 = 0.025 \quad \alpha/2 = 0.025 \]

Endpoint E1

\[ H_1 \quad H_2 \]

Endpoint E2

\[ H_3 \quad H_4 \]

Gatekeeping procedure based on an \( \alpha \)-splitting method:

- \( \alpha \) is split between \( H_1 \) and \( H_2 \)
- \( \alpha \) can be transferred between \( H_3 \) and \( H_4 \)
Gatekeeping procedure 1

Endpoint E1
Bonferroni

Endpoint E2
Holm

Bonferroni and Holm tests do not use available distributional information (test statistics within Families 1 and 2 are strongly positively correlated)
Gatekeeping procedure 2

Family 1
Powerful test

Family 2
Powerful test

Select tests that utilize available distributional information
Specify $\alpha$ propagation rules: how much error rate is transferred from Family 1 to Family 2
Gatekeeping procedure 2

Family 1
Truncated Hochberg

Family 2
Regular Hochberg

Mixture-based gatekeeping procedure:
- Truncated Hochberg test in Family 1 to enable flexible $\alpha$ propagation
- Regular Hochberg test in Family 2
Compare operating characteristics of candidate gatekeeping procedures

**Gatekeeping procedure 1**
Family 1: Bonferroni test
Family 2: Holm test

**Gatekeeping procedure 2**
Family 1: Truncated Hochberg test with truncation parameter of 0.7
Family 2: Hochberg test
C: Performance

Assumptions

<table>
<thead>
<tr>
<th>Family</th>
<th>Dose L</th>
<th>Dose H</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80</td>
<td>80</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>70</td>
<td>56</td>
</tr>
</tbody>
</table>

Dose L: Probability of achieving significant at Dose L
Dose H: Probability of achieving significant at Dose H
Both: Probability of achieving significant at both doses
C: Performance

Multiplicity penalties in Family 1

Gatekeeping procedure 1

Penalty (%)

<table>
<thead>
<tr>
<th>Dose</th>
<th>Penalty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>8.6</td>
</tr>
<tr>
<td>H</td>
<td>8.6</td>
</tr>
<tr>
<td>Both</td>
<td>11.5</td>
</tr>
</tbody>
</table>

Gatekeeping procedure 2

<table>
<thead>
<tr>
<th>Dose</th>
<th>Penalty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>3.8</td>
</tr>
<tr>
<td>H</td>
<td>3.8</td>
</tr>
<tr>
<td>Both</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Multiplicity penalty = Power before multiplicity adjustment – Power after multiplicity adjustment
C: Performance

Multiplicity penalties in Family 2

Gatekeeping procedure 1

<table>
<thead>
<tr>
<th>Penalty (%)</th>
<th>Dose L</th>
<th>Dose H</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penalty (%)</td>
<td>25.2</td>
<td>25.2</td>
<td>16.7</td>
</tr>
</tbody>
</table>

Gatekeeping procedure 2

<table>
<thead>
<tr>
<th>Penalty (%)</th>
<th>Dose L</th>
<th>Dose H</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penalty (%)</td>
<td>19.6</td>
<td>19.6</td>
<td>11.7</td>
</tr>
</tbody>
</table>

Multiplicity penalty = Power before multiplicity adjustment – Power after multiplicity adjustment
General evaluation criteria

Simple disjunctive power (one or more null hypotheses are rejected) or simple conjunctive power (all null hypotheses are rejected)

Subset disjunctive power (one or more null hypotheses are rejected in each family)

Weighted power

See Bretz, Maurer and Hommel (2011), Dmitrienko et al. (2011) for more information
Case study

Hochberg-based gatekeeping procedure

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Dose</th>
<th>Raw $p$</th>
<th>Adjusted $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>L</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>0.011</td>
<td>0.022</td>
</tr>
<tr>
<td>E2</td>
<td>L</td>
<td>0.006</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>0.040</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Both dose-placebo comparisons for Endpoints E1 and E2 are significant at $\alpha = 0.05$
Gatekeeping procedures in confirmatory trials

Type I error rate considerations
Control *global* error rate over multiple families

Power considerations
Based on *powerful multiple tests*

Clinical trial applications
Widely used in clinical trials to enrich product labels and provide *important clinical information* to physicians and patients (lurasidone product label)


