Multiple Endpoints

Weiying Yuan, PhD
Site Head of Biostatistics and Programming, China
Janssen Research and Development
Johnson & Johnson

Outline
- Multiple measurements to characterize disease
- Multiple primary endpoints
- Multiple secondary endpoints
- Repeated measurements
- Composite endpoints and their components
- Methods to control false positive (Type I) error rate
- Criteria to establish efficacy based on multiple endpoints

Multiple Primary Endpoints
- **Definition:** A set of clinical endpoints based on which clinical benefits are assessed.
  - Providing characterization of various aspects of a disease
  - Being used to describe clinical benefits

Examples of Multiple Primary Endpoints & Clinical Decision Rules
1. Primary endpoints - Alzheimer trial
   - Alzheimer Disease Assessment Scale – 11 Cognitive Subscale (ADAS-Cog/11)
   - Clinician Interview Based Impression of Change + Caregiver’s Input (CIBIC-plus)
   - Clinical decision rules:
     - Primary endpoints: Win on both ADAS-Cog/11 and CIBIC-plus

Examples of Multiple Primary Endpoints & Clinical Decision Rules (cont’d)
2. Primary endpoints - Epilepsy trial
   - % reduction in seizure rate
   - % reduction in drop attack rate
   - Parental global evaluation of seizure severity
   - Clinical decision rules:
     - Primary endpoints: Win on seizure rate or win on (drop attack rate and seizure severity)

Examples of Multiple Primary Endpoints & Clinical Decision Rules (cont’d)
3. Primary endpoints - Acne trial
   - Clinical global evaluation
   - Inflammatory lesion counts
   - Non-inflammatory lesion counts
   - Total lesion counts
   - Clinical decision rules:
     - Primary endpoints: Win on Clinical global and (at least 2 out of the 3 lesion counts endpoints)
Multiple Secondary Endpoints

- **Definition:** a set of clinical endpoints intended for possible inclusion in the label after efficacy has been demonstrated
  - Multiplicity adjustment needed to avoid findings by chance
  - Generally not intended for making primary efficacy claim but for labeling perspective

Examples of Multiple Primary & Secondary Endpoints & Clinical Decision Rules

4. Primary endpoints - congestive heart failure (CHF) trial
   - All cause mortality
   - Stroke
   - Myocardial infarction
   - Clinical decision rules:
     - Primary endpoint: All cause mortality
     - Secondary endpoints: stroke, MI

Examples of Multiple Primary Endpoints & Clinical Decision Rules (cont’d)

5. Primary endpoints - Alzheimer trial
   - Alzheimer Disease Assessment Scale – 11 Cognitive Subscale (ADAS-Cog/11)
   - Clinician Interview Based Impression of Change + Caregiver Input (CIBIC-plus)
   - Clinical decision rules:
     - Primary endpoint: Win on both ADAS-Cog 11 and CIBIC-plus
     - Secondary: Win on Neuropsychiatric Inventory (NPI), Activities of Daily living (ADCS-ADL Inventory)

Statistical Issues

- Design issues
  - Win on specific or win on any co-primary endpoints
  - Endpoints correlated vs. independent
  - Impact on Type I & II error rates
- Analysis issues
  - Methods to account for (adjust) multiplicity
  - Interpretation of results
- Controlling the chances of false positive conclusions; multiplicity strategy need to be pre-specified in the protocol

Clinical Decision Rule & Statistical Testing Procedure

Example 1: trial with 3 co-primary endpoints
- **Win on all:** show significance on all 3 co-primary endpoints
- **Win on specific:** a step-down hierarchical closed testing procedure (pre-ordered 3 endpoints);
- **Win on any:** show significance on any 1 of the 3 endpoints. No order.

Clinical Decision Rule & Statistical Testing Procedure – Single Inference (cont’d)

Example 1: trial with 3 co-primary endpoints (cont’d)
- **Win on all:** show significance on all 3 co-primary endpoints
  - Strong control of Type I error rate: all tested at \( \alpha = 0.05 \)
    - 1-sided \( \alpha = 0.025 \)
  - Adjustment of Type II error rate. Most conservative assumption: independence between endpoints
  - \( H_0: \) no effect on both endpoints; \( H_1: \) both are significant. A single inference

Preplanned and described in protocol the pre-specified clinical decision rules and statistical testing procedure
Clinical Decision Rule & Statistical Testing
Procedure (cont’d)

Impact on statistical Power (1 – β)

Power comparison for K=2 endpoints
- Single endpoint power
- Win on both
- versus win on at least one (1-sided test at 0.025)

Power comparison: win in each endpoint at \( \alpha = 0.025 \)
(1-sided test)
- Show decrease in power when K increases,
  - Smaller decrease in power for higher correlation

O’Neill 2004

Clinical Decision Rule & Statistical Testing
Procedure - Multiple Inference (cont’d)

Example 1: trial with 3 co-primary endpoints (cont’d)
- Win on specific: Pre-specify the rank of endpoints in hierarchical order in protocol, step-down closed testing procedures on each endpoint
  - Regardless of final p-values
  - Strong control of Type I error rate

Preplanned and described in protocol the pre-specified clinical decision rules and statistical testing procedure

Clinical Decision Rule & Statistical Testing
Procedure - Multiple Inference (cont’d)

Example 1: trial with 3 co-primary endpoints (cont’d)
- Win on any: show significance on any 1 of the 3
  - Weak control of Type I error rate
  - Statistical testing procedures: eg. Holms, Hochberg, etc.

Clinical Decision Rule & Statistical Testing
Procedure - Multiple Inference (cont’d)

Example 2: Alzheimer trial
- Two co-primary endpoints:
  - ADAS-cog/11 & CIBIC-plus
- Need to win both, ie, show significance on both
  - Each will be tested at \( \alpha = 0.05 \) (1-sided \( \alpha = 0.025 \ )) since not increasing chance
- Statistical power: 1 - β for each at 90%
- Statistical power for the trial: 0.9 \times 0.9 = 81% (assuming two endpoints are independent). If want to maintain 90% power for the trial, each planned at 95% power.

Chi, 2003
Composite Endpoints

**Definition:** An endpoint that is defined based on the responses measured by two or more co-primary clinical endpoints.

- A composite endpoint can be rating scales that yield a total score (e.g., Total PANSS score, the total positive and negative syndrome scales for schizophrenia trials) or index.

**Rationale for using composite endpoints:**
- Disease needs multidimensional characterization based on multiple measurements.
- Low event rates on individual component primary endpoints.
- Increased power by having more events.
- Mortality or ultimate endpoint needs to be accounted for.
- Not certain which component will win.
- Without inflated Type I & II error rates.

**Example 1:**
- Alzheimer's disease trial: definition of response to treatment (no change or improved) if:
  1. Change in ADAS-cog11 from baseline ≤0, or
  2. CIBIC-plus ≤4.

**Example 3:**
- Organ transplant: definition of “failure” if 6 months after:
  1. acute rejection, or
  2. graft loss, or
  3. death.

**Example 4:**
- Rheumatoid Arthritis trial:
  - Primary endpoints:
    - tender joint count
    - swollen joint count
    - patient pain assessment
    - patient global assessment
    - physician global assessment
    - patient self-addressed disability
    - acute-phase reactant (ESR or CRP).
**Composite Endpoints (cont'd)**

**Example 4: Rheumatoid Arthritis trial (cont'd)**

- **Rheumatoid Arthritis (ACR20), a patient is a responder if shows at least a**
  - 20% improvement in tender joint count, and
  - 20% improvement in swollen joint count, and
  - at least a 20% improvement in 3 out of 5 following endpoints:
    - patient pain assessment
    - patient global assessment
    - physician global assessment
    - patient self-addressed disability
    - acute-phase reactant (ESR or CRP)

---

**Composite Endpoints (cont'd)**

**The LIFE Study**

- Losartan (COZAAR) vs. Atenolol in 9193 hypertensive patients
- There were 3 primary endpoints:
  - Cardiovascular death, stroke, myocardial infarction
- Composite endpoint = the time to the first occurrence of any {CV death, stroke, MI}

---

**Composite Endpoints (cont'd)**

**The LIFE Study (cont'd)**

- Trial outcomes
  - Primary composite endpoint (508 (11%) vs. 588 (13%), \( p = 0.021 \))
  - Individual components of primary,
    - Stroke (232 (5%) vs. 309 (7%), \( p = 0.001 \))
    - MI (198 (4%) vs. 188 (4%), \( p = 0.491 \))
    - CV death (204 (4%) vs. 234 (5%), \( p = 0.206 \))

---

**Composite Endpoints (cont'd)**

- Issue of composite endpoints
  - Positive result is not component specific
  - Difficult in claim based on composite endpoint
  - Difficult in determine if and how a component contribute to the composite
  - Single inference of the composite endpoint
  - Careful about interpretation of results.

---

**Clinical Decision Rule & Statistical Testing Procedure - Multiple Inference**

- Single testing problem in which Type I error is uniquely defined as the probability that \( p \)-value is \( \leq \alpha \) given that the null hypothesis of no between treatment group difference is true.
- Multiple testing problem in which no unique definition for Type I error rate. Needs to be pre-specified and properly controlled for Type I error rate
- Cannot test each endpoint at \( \alpha = 0.05 \).
Clinical Decision Rule & Statistical Testing Procedure - Multiple Inference (cont’d)

<table>
<thead>
<tr>
<th>Number of independent endpoints</th>
<th>Probability of making at least one erroneous inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.050</td>
</tr>
<tr>
<td>2</td>
<td>0.098</td>
</tr>
<tr>
<td>3</td>
<td>0.143</td>
</tr>
<tr>
<td>4</td>
<td>0.185</td>
</tr>
<tr>
<td>5</td>
<td>0.226</td>
</tr>
<tr>
<td>10</td>
<td>0.401</td>
</tr>
</tbody>
</table>

Assuming no treatment effect on any of the endpoints

Weak Control of Type I Error Rate for Endpoint Non-specific Inference
- Multiple testing procedure provides only weak control of Type I error rate, if it wins on endpoints not specified in design
  - Does not control all potential sources that may result in erroneous conclusion
  - Does not permit endpoint specific inferences
- Example: Holms, Hochberg, and Simes procedures first orders the p-values obtained from testing individual endpoints. No specification of which endpoint in design

Strong Control of Type I Error Rate for Endpoint Specific Inference
- Multiple testing procedure provides strong control of Type I error rate, if it wins on pre-specified endpoints in design
  - Allow endpoint specific inference
  - Provides information on what specific clinical benefits can be expected of the treatment
- Example: Step-down closed testing procedure with order of the endpoints being tested specified \textit{a priori}.

Conclusions and Recommendations
- Importance of understanding the multiplicity issues in trials with multiple primary, secondary endpoints.
  - Issues associated with inflated Type I error rate
  - Impact on Type II error rate/statistical power and sample size

Conclusions and Recommendations (Cont’d)
- In design of clinical trials.
  - Clearly specify the multiplicity issues due to multiple primary, secondary, and component endpoints, multiple comparisons due to multiple treatment groups
  - Specify clinical decision rules and statistical testing procedures

Conclusions and Recommendations (Cont’d)
- Designs properly specified
  - Primary endpoints (multiple or single) to support the claim/indication in the label
  - Secondary endpoints (step-down after winning the primary endpoint(s) to describe further benefit of the drug in label
  - Adjust for multiplicity to avoid false positive conclusions
References


2. Chi, George Clinical Benefits, Decision Rules, and Multiple Inferences; February 12, 2003, CDER Scientific Roundtable


Thank you!