Chapter 4 Analysis of Multiple Endpoints in Clinical Trials

Wenge Guo

April 15, 2011

Different Types of Endpoints

Clinical trials generally classify the endpoints into *primary, secondary and exploratory* types.

- *Primary endpoints* address primary objectives of the trial.
- They are usually few but are clinically most relevant to the disease and the treatment under study.
- They assess the main clinical benefits of the treatment.
- This is usually done through one or more clinical "win" criteria.

Different Types of Endpoints (II)

- *Secondary endpoints* characterize extra benefits of the treatment under study after it has been demonstrated that the primary endpoints show clinically meaningful benefits of the treatment.
  - O'Neal (1997) supported the idea that "secondary endpoints cannot be validly analyzed if the primary endpoint does not demonstrate clear significance".
- *Exploratory endpoints* are usually not prospectively planned and are generally not rigorously evaluated like primary and secondary endpoints. These endpoints are used in treatment comparisons and also unplanned subgroup analysis with an exploratory (e.g., hypothesis generating) purpose.
  - In certain situations, their results can be useful in designing future new trials. However, they are not useful for confirmatory purpose.

Clinical Win Criteria with Primary Endpoints

- Win criteria are also called "clinical decision rules" for determining clinically meaningful treatment efficacy. They simply define how a positive clinical decision regarding the effectiveness of a test treatment in a trial is going to be reached.
- The criteria are defined relative to one or more relevant clinical primary endpoints in the setting of comparing one or more doses of test and control treatments.
- **Some examples:**
  - **Example 1:** one specified primary endpoint need to be significant.
  - **Example 2:** given $m \geq 2$ specified primary endpoints, all need to be statistically significant.
  - **Example 3:** given three specified primary endpoints, E_1, E_2 and E_3, either (both E_1 and E_2) or (both E_1 and E_3) need to be statistically significant.

Primary Analysis

There is a variety of ways in which the contribution of each primary endpoint can be accounted for in the primary analysis.

- Primary endpoints are treated as independent entities.
- Primary endpoints are treated as manifestations of a single underlying cause.

Example 1. If each endpoint independently provides a proof of efficacy, the trial's outcome is declared positive if at least one endpoint is associated with a significant improvement compared to the control.

In the *VEST trial*,

- mortality
- mortality plus morbidity due to heart failure.

In the *PRAISE-I trial*,

- mortality plus cardiovascular morbidity
- mortality

Primary Analysis (II)

Example 2. The primary objective is based on the development of composite endpoints. A composite endpoint can be based on a sum of multiple scores or combination of multiple events.

In the *LIFE trial*,

- the primary objective was to study the effect of losartan on the composite endpoint of cardiovascular death, myocardial infarction, and stroke.

Primary Analysis (III)

Example 3. When the multiple endpoints are biologically related to each other, the primary effect can be defined in terms of a combination of individual effects across the endpoints.

In the *mitoxantrone trial*, patients with progressive multiple sclerosis are evaluated the overall effect of five clinical measures:

- expanded disability status scale
- ambulation index
- number of treated relapses
- time to first treated relapse
- standardized neurological status

Primary Analysis (IV)

Example 4. In certain cases, a clinically meaningful effect is defined as the simultaneous improvement in multiple measures. In this case, the primary objective of a clinical trial is met if the test drug shows a significant effect with respect to all the endpoints.

Some clinical trials:

- Migraine
- Alzheimer's disease
- Osteoarthritis

Inferential Goals

Consider a clinical trial with two treatment groups. The trial's objective is to assess the effect of the experimental treatment on m endpoints compared to that of the placebo. Let δ_i be a measure of the true treatment effect for the ith endpoint.

At-least-one procedures

- If each multiple endpoint is independently clinical relevant, the multiple endpoint problem can be formulated as a multiple testing problem, and the trial is declared positive if at least one significant effect is detected.
- The global hypothesis testing problem is stated as

$$H_I = \bigcap_{i=1}^{m}(\delta_i \leq 0) \quad vs. \quad K_U = \bigcup_{i=1}^{m}(\delta_i > 0).$$

The global hypothesis is rejected if one or more individual hypotheses of no treatment effect are significant.

Inferential Goals (II)

Global procedures

- In many clinical trial applications, it is desired to show that the treatment has an overall effect across the endpoints without necessarily a large significant effect on any one endpoint.
- To establish an overall treatment effect, usually a point null hypothesis of no difference between the treatment and control is tested against a one-sided alternative:

  $H_I^*: \delta_i = 0$ for all i vs. $K_U^*: \delta_i \geq 0$ for all i and $\delta_i > 0$ for some i.

  One-sided global procedures are needed.

Inferential Goals (III)

All-or-none procedures

- Another formulation of the multiple endpoint problem pertains to the requirement that the treatment be effective on all endpoints.
  - It represents the most stringent inferential goal for multiple endpoints.
- The global hypothesis testing problem is stated as

$$H_U = \bigcup_{i=1}^{m} (\delta_i \leq 0) \quad vs. \quad K_I = \bigcap_{i=1}^{m} (\delta_i > 0).$$

To reject H_U, one needs to show that all individual hypotheses are false (the treatment effects for all endpoints are significant).

Inferential Goals (IV)

Superiority-noninferiority procedures

- It provides a viable alternative to the stringent all-or-none testing approach. In this case, the inferential goal is to demonstrate that the treatment is superior to the control on at least one endpoint and not inferior on all other endpoints.
- Let $\eta_k \geq 0$ denote the superiority threshold (commonly, $\eta_k = 0$) and $\varepsilon_k > 0$ denote the noninferiority threshold for the kth endpoint.
- The treatment is superior to the control on the kth endpoint if $\delta_k > \eta_k$ and is noninferior to the control on the kth endpoint if $\delta_k > -\varepsilon_k, k = 1, \ldots, m$.
- For the kth endpoint, the superiority testing problem is stated as

$$H_k^{(S)} : \delta_k \leq \eta_k \quad vs. \quad K_k^{(S)} : \delta_k > \eta_k.$$

Inferential Goals (V)

- Similarly, the noninferiority testing problem for the kth endpoint is stated as

$$H_k^{(N)} : \delta_k \leq -\varepsilon_k \quad vs. \quad K_k^{(N)} : \delta_k > -\varepsilon_k.$$

- The overall superiority testing problem is given by

$$H_I^{(S)} = \bigcap_{k=1}^{m} H_k^{(S)} \quad vs. \quad K_U^{(S)} = \bigcup_{k=1}^{m} K_k^{(S)}.$$

The global superiority hypothesis $H_I^{(S)}$ is rejected if at least one $H_k^{(S)}$ is rejected.

Inferential Goals (VI)

- Similarly, the overall noninferiority testing problem is given by

$$H_U^{(N)} = \bigcup_{k=1}^{m} H_k^{(N)} \quad vs. \quad K_I^{(N)} = \bigcap_{k=1}^{m} K_k^{(N)}.$$

The global noninferiority hypothesis $H_U^{(N)}$ is rejected if all $H_k^{(N)}$ are rejected.

- Thus, we only need to test the union of the global superiority and global noninferiority hypotheses

$$H_U^{(SN)} = H_I^{(S)} \bigcup H_U^{(N)} \quad vs. \quad K_I^{(SN)} = K_U^{(S)} \bigcap K_I^{(N)}.$$

The trial's objective is met if there is superior efficacy for at least one endpoint and noninferior efficacy for all endpoints.

At-least-one Procedures

The objective is to demonstrate the treatment's superiority on at least one endpoint.

Weighted Bonferroni procedure:

- Let $w_1, \ldots, w_m$ be positive weights representing the importance of the endpoints such that they sum to 1.
- The hypothesis of no treatment effect for the ith endpoint is tested at level α_i, where $\alpha_i = w_i\alpha$ and thus $\sum_{i=1}^{m} \alpha_i = \alpha$.
- The weighted Bonferroni procedure has been proved to strongly control the FWER at level α.

At-least-one Procedures (II)

Prospective alpha allocation scheme (PAAS) method, proposed by Moyé (2000).

- Assuming that the *p*-values for the individual endpoints are independent, we have $\prod_{i=1}^{m}(1-\alpha_i) = 1-\alpha$.
- For the case of two co-primary endpoints, we first set $0 < \alpha_1 < \alpha$, and then calculate $\alpha_2 = 1 - \frac{1-\alpha}{1-\alpha_1}$.
- For example, if $\alpha = 0.05$ and $\alpha_1 = 0.045$, then $\alpha_2 = 0.0052 > 0.005$.

At-least-one Procedures (III)

Adaptive Alpha Allocation Approach (or 4A approach):

- Consider a clinical trial with m endpoints and assume that the endpoints are grouped into two families.
- The first family includes m_1 endpoints that are adequately powered and the second family includes m_2 potentially underpowered endpoints ($m_1 + m_2 = m$).
- The endpoints in the first family are tested using any FWER controlling procedure at level $\alpha_1 = \alpha - \varepsilon$, where $\varepsilon > 0$ is small.

At-least-one Procedures (IV)

Adaptive Alpha Allocation Approach (Cont.):

The endpoints in the second family are tested using any FWER controlling procedure at level α_2, which is adaptively based on $P_{(m_1)}$ as follows:

$$\alpha_2(P_{(m_1)}) = \begin{cases} \alpha & \text{if } P_{(m_1)} \leq \alpha_1, \\ \min(\alpha^*/P^2_{(m_1)}, \alpha_1) & \text{if } P_{(m_1)} > \alpha_1, \end{cases}$$

where

$$\alpha^* = \begin{cases} \alpha_1(1 - \sqrt{2 - \alpha_1/m_1 - \alpha/\alpha_1})^2 & \text{if } \alpha_1 + \alpha_1^2/m_1 - \alpha_1^3/m_1^2 \leq \alpha, \\ \alpha_1(\alpha - \alpha_1)/(m_1 - \alpha_1) & \text{if } \alpha_1 + \alpha_1^2/m_1 - \alpha_1^3/m_1^2 > \alpha. \end{cases}$$

At-least-one Parametric Procedures

Bonferroni-type parametric procedure

- The global null hypothesis of no treatment effect is rejected if at least one test is significant, i.e., if $t_{\max} = \max(t_1, \ldots, t_m) \geq c$, where c is a critical value computed from $P\{t_{\max} < c\} = 1 - \alpha$.

Fallback-type parametric procedure

- Let $t_1, \ldots, t_m$ denote the test statistics for the m endpoints and let $w_1, \ldots, w_m$ denote the weights that represent the importance of the endpoints.
- The test statistics are assumed to follow a standard multivariate normal distribution.

At-least-one Parametric Procedures (II)

Fallback-type parametric procedure (Cont.)

- **Step 1.** Calculate critical values $c_1, \ldots, c_m$ and corresponding significance levels $\gamma_1, \ldots, \gamma_m$ as follows,

  $$P(t_1 \geq c_1) = \alpha w_1 \text{ and } P(t_1 < c_1, \ldots, t_{i-1} < c_{i-1}, t_i \geq c_i) = \alpha w_i,$$

  for $i = 2, \ldots, m$.
  - The probability are computed under the global null hypothesis.
  - The significance levels $\gamma_i = 1 - \Phi(c_i)$, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.
- **Step i.** If s is the index of the last non-significant endpoint, the significance level for the ith endpoint is given by $\max(\alpha w_{s+1} + \ldots + \alpha w_i, \gamma_i)$.

At-least-one Parametric Procedures (III)

Fallback-type parametric procedure (Cont.)

Example. Consider a clinical trial with two unequally weighted endpoints ($w_1 = 0.8$ and $w_2 = 0.2$) tested at the overall two-sided $\alpha = 0.05$.

- The regular fallback procedure with $\alpha_1 = 0.04$ and $\alpha_2 = 0.01$
- The parametric fallback procedure with $\alpha_1 = 0.04$ and $\alpha_2 = 0.0104, 0.0112$ and 0.0146 for $\rho = 0, 0.3$ and 0.6.

At-least-one Parametric Procedures (IV)

Advantages of parametric fallback procedure

- The parametric procedure is robust with respect to the monotonicity assumption and performs well when the first test in the sequence is underpowered.
- When the effect sizes across the tests are comparable, the power of individual tests improves toward the end of the sequence.
- The power of tests later in the sequence declines with increasing correlation.

All-or-none Procedures

- **The all-or-none procedure** has the following form:

  Reject all hypotheses if $t_{\min} = \min_{1 \le i \le m} t_i \ge t_\alpha(\nu)$,

  where $\nu = n_1 + n_2 - 2$ and $t_\alpha(\nu)$ is the $(1-\alpha)$-quantile of the t-distribution.

- Since this procedure does not use a multiplicity adjustment (each hypothesis H_i is tested at level α), it may appear at first that it must be highly powerful as a test of the global hypothesis H_U.

- In reality, the min test is very conservative because of the requirement that all hypotheses must be rejected at level α.

All-or-none Procedures (II)

The conservatism results from the least favorable configuration of the min test which can be shown to be the following form:

- No treatment effect for any one endpoint ($\delta_i = 0$ for some i).
- Infinitely large treatment effects for all other endpoints ($\delta_j \to \infty$ for $j \neq i$).

This configuration leads to marginal α-level t-tests.

Superiority-noninferiority Procedures

Several requirements:

- The treatment is superior to the control on all endpoints — **too strong**
- The treatment is superior to the control on at least one endpoint — **too weak**
- The superiority-noninferiority approach — strengthes the second requirement by augmenting it with the additional requirement that the treatment is not inferior to the control on all other endpoints.

Superiority-noninferiority Procedures (II)

- The null and alternative hypotheses are defined as follows,

$$H_U^{(SN)} = H_I^{(S)} \bigcup H_U^{(N)} \quad vs. \quad K_I^{(SN)} = K_U^{(S)} \bigcap K_I^{(N)}.$$

- The trial's outcome is declared positive if there is evidence of superior efficacy for at least one endpoint ($K_U^{(S)}$) and noninferior efficacy for all endpoints ($K_I^{(N)}$).

Tamhane-Logan Superiority-noninferiority Procedure

Denote the t-statistics for superiority and noninferiority for the kth endpoint by

$$t_k^{(S)} = \frac{\overline{X}_{1\cdot k} - \overline{X}_{2\cdot k} - \eta_k}{s_k\sqrt{1/n_1 + 1/n_2}}, \quad t_k^{(S)} = \frac{\overline{X}_{1\cdot k} - \overline{X}_{2\cdot k} + \varepsilon_k}{s_k\sqrt{1/n_1 + 1/n_2}},$$

where $\overline{X}_{i\cdot k}$ $(i = 1, 2; k = 1, \ldots, m)$ denote the mean response in the ith group on the kth endpoint and s_k^2 denote the pooled sample variance for the kth endpoint.

Tamhane-Logan Superiority-noninferiority Procedure (II)

- Tamhane and Logan (2004) used the UI statistic $t_{\max}^{(S)} = \max(t_1^{(S)}, \ldots, t_m^{(S)})$ for testing the superiority null hypothesis $H_I^{(S)}$ and IU statistic $t_{\min}^{(N)} = \min(t_1^{(N)}, \ldots, t_m^{(N)})$ for testing the noninferiority null hypothesis $H_U^{(N)}$.
- They proposed the following procedure of the global superiority-noninferiority hypothesis:

  Reject $H_U^{(SN)}$ *if* $t_{\max}^{(S)} \geq c^{(S)}$ *and* $t_{\min}^{(N)} \geq c^{(N)}$,

  where the critical values $c^{(S)}$ and $c^{(N)}$ are chosen so that the procedure has level α.

Conclusions

- In this lecture, we introduce several different types of endpoints and discuss clinical win criteria with primary endpoints
- We also introduce several inferential goals in terms of multiple endpoints.
- Finally, we introduce several different testing procedures for the inferential goals such as at-least-one procedures, all-or-none procedures, and superiority-noninferiority procedures.