Key multiplicity concepts and principles addressed in the Draft Guidance: "Multiple Endpoints in Clinical Trials"

Mohammad F. Huque, Ph.D.
Office of Biostatistics
OTS, CDER/FDA

DIA Annual Meeting, Boston, MA
June, 2013

This presentation reflects the views of the presenter and should not be construed to represent FDA's views or policies.

Note

- The Draft Guidance, “Multiple Endpoints in Clinical Trials,” is near completion.
- It is expected to be released soon for public comments.
- Purpose of this presentation is to share with you some key concepts and principles addressed in this document.

About this Draft Guidance

- It is a unique document ever tried at the FDA, written by an FDA committee of statistical and clinical experts.
- It is written in a non-technical language in order to reach a broad audience.
- It includes concepts and methods that were written after much discussions and deliberations for bringing clarity – that is why, it has taken some time to reach to finish it.

The Document has 5 sections

I. Introduction

II. Introductory concepts and principles

III. Multiple endpoints: general principles

IV. Statistical concepts, methods and principles

V. Supportive descriptive statistics and graphs

References

Appendix

Scope

- Multiplicity topics addressed are mostly related to adequate and well-controlled studies.
- Some multiplicity topics are beyond the scope of this Guidance. For example, topics related to
  - Safety
  - Subgroup analyses
  - Sequential/adaptive designs

Illustrative examples

- Includes illustrative examples related to methods that apply to multiple endpoints.
- Emphasizes that these methods also apply to other situations, such as to different doses, time points, and study population subsets

Includes a number of stat methods w. illustrative examples

- Bonferroni Method
- Holm Procedure
- Hochberg Procedure
- Fixed Sequence Method
- Modified Fixed Sequence method
- Gatekeeping Testing Strategies
- Truncated Holm Procedure for Parallel Gatekeeping
- Multi-branched (Tree-structured) Gatekeeping Procedures
- Resampling Based Multiple Testing Procedures
- Graphical method

Addresses a number of multiplicity topics and issues

- Primary and secondary endpoints
- Multiplicity and its extent
- False positive error rate and its control
- Prospectively planned and post-hoc analyses
- Co-primary endpoints and issues
- Composite and multi-component endpoints and issues
- Descriptive statistics and graphs for labeling

Defines and explains endpoint families

- Primary endpoints
  - Endpoint(s) necessary and/or sufficient to establish efficacy (define a successful trial)
- Secondary endpoints
  - Not sufficient to establish efficacy in the absence of an effect on the primary endpoints; not required for establishing efficacy
  - Potentially could lead to additional labeling claims
- Exploratory endpoints
  - Hypothesis generating (clinical utility unknown)

Suggests not using the terms, “key secondary” and “tertiary” endpoints

- Endpoints designated with these terms are not used in the draft Guidance and it recommends that they not be used in protocols.
- Reason: Use of these terms can lead to misunderstandings regarding their intended use, and suitability for such use.

Defines and explains "multiplicity"

- Multiplicity refers to situations in a trial in which multiple statistical tests or analyses create multiple ways to "win" for treatment efficacy or safety.
  - This can cause the **false positive error rate (Type I error rate)** to inflate beyond the desired level, e.g., 0.05, if each test is performed, for example, at the same alpha level of 0.05.
- This inflation in a trial can be substantial and problematic, but

  it can be controlled to a desired level by an appropriate, prospectively planned statistical strategy using the statistical framework of testing multiple hypotheses.

Explains when is it necessary to adjust for multiplicity?

- When there are one or more claims of treatment benefits based on primary and secondary endpoints.
- When the win criteria set up are such that one can win in multiple ways, i.e., there are multiple pathways for winning.

✓ Such situations require multiplicity adjustments when they cause inflation of the Type I error rate.

Examples of multiple pathways for winning are many

- For a trial with a single primary endpoint:
  - Win either for the total patient population or for a targeted subgroup
  - Win either for the high dose or for the medium dose in a trial with 3 doses (low, medium, high) and placebo
- For a trial with two or more primary endpoints:
  - Win in at least on one of the PEs in a 2-arm trial that compares treatment to control
- For trials with multiple objectives:
  - Many more examples involving multiple primary and secondary objectives

Explains what is not multiplicity

- Often there are multiple analyses for the intention-to-treat (ITT) data set for the same PE and by the same method
  - These multiple analyses are done for the same endpoint on varying the assumptions about some data points because of missing data, protocol violations, use of concomitant medications, etc.
- As these analyses are sensitivity analyses for assessing the primary analyses results, there is no multiplicity adjustment for them

What is not multiplicity (cont'd)

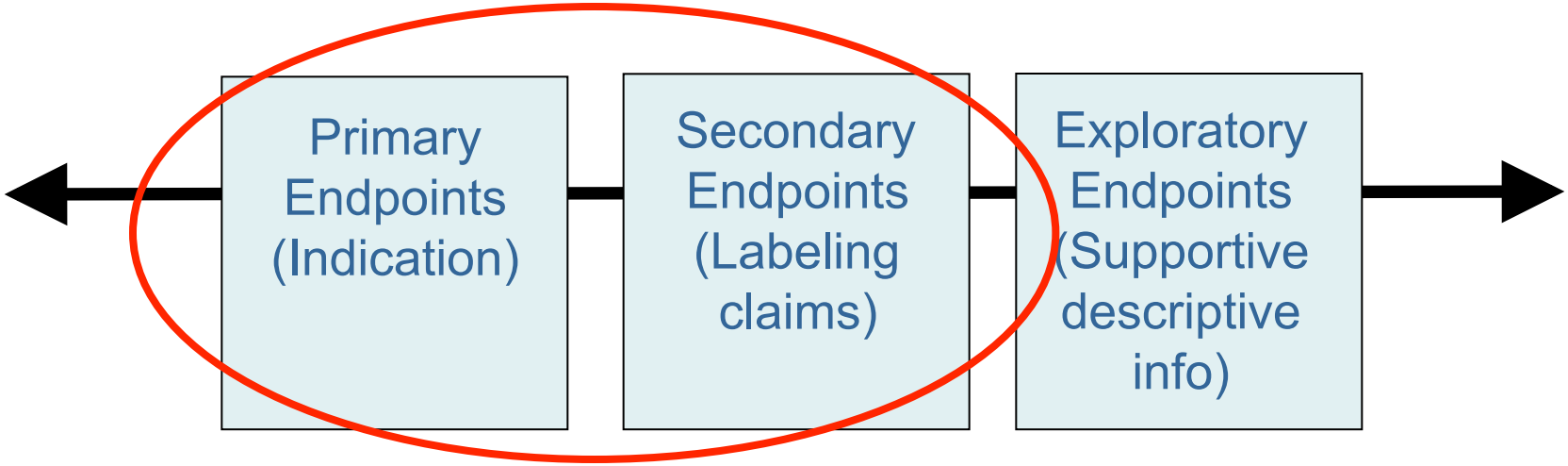
- Often there are analyses of the same endpoint data by alternative methods, e.g.,
  - analysis of the same time-to-event endpoint by log-rank test and by the generalized Wilcoxon test
  - analysis by the parametric and non-parmaetric methods.
- Technically, one can adjust for these multiple analyses if they were pre-specified.
- However, this is rarely done, as the purpose of these analyses is usually to demonstrate that the results found are robust and hold regardless of different methods applied.

What is not multiplicity (cont'd)

- The Draft Guidance considers descriptive analyses and graphs that go into the labeling as being "not multiplicity" – Section V of the document is devoted to this topic.
- These analyses are supposed to be further elaborations of effectiveness that has been established in a statistically rigorous way.
- **Caution: These analyses should be recognized as insufficient to justify additional drug efficacy claims beyond those supported by the prospective analyses**.

Considers error rate control for the primary and secondary families of hypotheses

Continuum for Type I Error Control



→To all primary and secondary endpoints
→Overall error rate should not exceed a pre-specified α

Recommendations on stat methods for controlling the Type I error rate

- Methods generally used for the primary and secondary endpoints should be those that allow finding of significant treatment effects at the individual endpoint level, without inflating the Type I error rate
- These methods permit an individual conclusion about efficacy with respect to each endpoint tested in the primary and the secondary family
- **Some methods (often called global procedures) allow a conclusion of treatment efficacy in the global sense. Such methods generally inflate the Type I error rate for making conclusions on the individual endpoints.**

Emphasizes prospective planning as a key to addressing multiplicity

- An important component in controlling for multiple comparisons is to specify in the protocol all planned study endpoints, time points, subgroups, and analyses in advance.
- Changes in the analysis plan to perform non-prospectively stated analyses can reintroduce a multiplicity problem

Explains pitfalls of post-hoc analyses

- Although post-hoc analyses of trials that fail on their specified endpoints may be useful for generating hypotheses for future testing, they do not yield definitive results.
- The results of such analyses can be biased, as the choice of analyses surely can be influenced by a desire for success.
- It is difficult to confirm how many different analyses were performed; in this situation, there is no credible way to correct for the multiplicity of multiple analyses and control theType I error rate.
- **Consequently, post hoc analyses generally do not provide evidence of effectiveness.**

Explains when in clinical trials co-primary endpoints are used

- Situation 1: When there are two or more critically important different features of a disorder
  - These features are so critically important to the disease that a drug will not be considered effective without demonstration of a treatment effect on all these disease features.
- Example:
  - Migraine headaches are characterized by the presence of pain, photophobia, phonophobia, and nausea.
  - A treatment is considered effective for migraines if all four aspects of the disorder are shown to be improved by the drug treatment.

Co-primary endpoints (cont'd)

- Situation 2:
  - When there is a single identified critical feature of the disorder, but there is no single patient evaluation that is both specific for the disease feature and clinically interpretable.
  - In these cases, two endpoints are often used.
- Example:
  - Alzheimer's disease trial with endpoints: ADAS-Cog and a global measure of function (e.g., global assessment)
  - One endpoint assures that the effects occurs on the core disease feature, and the other that the effect is clinically meaningful.

Statistical considerations for co-primary endpoints

- **When using co-primary endpoints, testing each individual endpoint at the 0.05 level does not cause inflation of the Type I error rate,**
  - rather the impact of co-primary endpoint testing is on the Type II error rate."
- In general, unless clinically very important, the use of more than two co-primary endpoints should be carefully considered because of the loss of power.
- **Relaxation of alpha is not generally acceptable because doing so will undermine the unequivocal demonstration of an effect on each disease aspect considered essential to showing that the drug has the desired effect.**

Gives an idea of creating a single endpoint from multiple co-primary endpoints

- **Idea:** A successfully treated patient will be that who improves on all the identified necessary endpoints.
- For this, each of the endpoints can be made dichotomous by applying the specified threshold for improvement.
- This can allow classifying patients as responders versus non-responders, and a primary endpoint might be formulated to compare the proportion of responders in each group.

Addresses composite and multi-component endpoint issues in detail (some key points:)

- A common approach in practice has been to combine multiple endpoints (called components) to a single composite (or a single multi-component endpoint) when
  - components individually are expected to yield small treatment effects, but collectively they can show a clinically meaningful benefit.
- Such an approach can effectively reduce the size of the trial if components contribute to the total treatment effect in a meaningful way.
- If individual components were tested simultaneously (e.g., by the Bonferroni test), when expecting only small treatment effects in each, then such an approach would not be practical.

Interpretation of the composite endpoint findings

- The treatment effect on the composite describes the overall clinical effect of the treatment when
  - components all are of reasonably similar clinical importance, and
  - components exhibit some consistency of treatment effects.
- Interpretation difficulties arise when
  - the clinical importance of different components is substantially different, and
  - the treatment effect is mainly on the least important component.

Interpretation of the composite endpoint findings (cont'd)

- If a critical component (e.g., mortality) is adversely affected by the treatment, even if one or more components of less importance are favorably affected, so giving an overall favorable statistical result.
- Then, in that case, while the overall analysis indicates that the treatment is successful, careful examination of the data may call this conclusion into question.

✓ **A key recommendation: For interpretation purposes, component endpoint data are to be fully displayed and carefully examined.**

(Draft Guidance addresses this issue in detail)

Multiplicity issues in composite endpoint trials

- There is no multiplicity issue if the trial has a single composite endpoint as the sole primary endpoint, and there no claim of treatment benefit for its specific components.
  - Component outcomes are analyzed and displayed only in the descriptive sense as an aid to interpreting the result of the composite endpoint.
- Multiplicity issues arise when, for example,
  - claims of treatment benefit are sought for the composite endpoint, as well as for its sub-composites or for its individual components.
- Most of these multiplicity issues can be address by a variety of multiple testing methods (e.g., by gatekeeping and graphical methods)

Concluding Remarks

- The Draft Guidance on “multiple endpoints in clinical trials” is a unique comprehensive document in which both clinical and statistical ideas flow together, and is intended to reach a broader audience.
- This presentation focused on some key multiplicity concepts and principles included in this document.
- Public will have the opportunity to read the full document and provide comments when available for public comments.
- The Agency carefully considers and reviews all comments received, and discusses them extensively for proper final revision of the document.

Thank You