Analysis of clinical trials with multiple objectives

Alex Dmitrienko
Center for Statistics in Drug Development
Quintiles

Outline

Multiplicity problems in clinical trials

Classification of multiplicity problems

Important issues

Selection of secondary endpoints/objectives

Analysis of secondary endpoints/objectives within a development program

Evaluation of multiple testing procedures

Multiplicity problems: Traditional setting

Single family

$$H_1, \ldots, H_m$$

Single source of multiplicity (univariate multiplicity problem)

Single family of null hypotheses

Multiple endpoints

Better characterize efficacy/safety of a new treatment

Multiple dose-placebo comparisons

Evaluate the dose-response relationship

Multiple patient populations

Evaluate efficacy in one or more subsets of overall population and help develop targeted agents

Multiplicity problems: Advanced setting

Multiple families

Family 1 $H_1, \ldots, H_{k_1}$

$\ldots$

Family m $H_{k_{m-1}+1}, \ldots, H_{k_m}$

Multiple sources of multiplicity (multivariate multiplicity problem)

Multiple families of null hypotheses

Multiple primary and secondary endpoints

Evaluate primary endpoints and provide useful supportive information (secondary endpoints)

Multiple endpoints and multiple dose-placebo comparisons

Evaluate efficacy at several dose levels

Multiple patient populations and multiple dose-placebo comparisons

Evaluate efficacy at several dose levels in different groups of patients

Multiple families of null hypotheses

Primary objectives

Directly related to the trial's outcome

Secondary objectives

Provide key supportive evidence

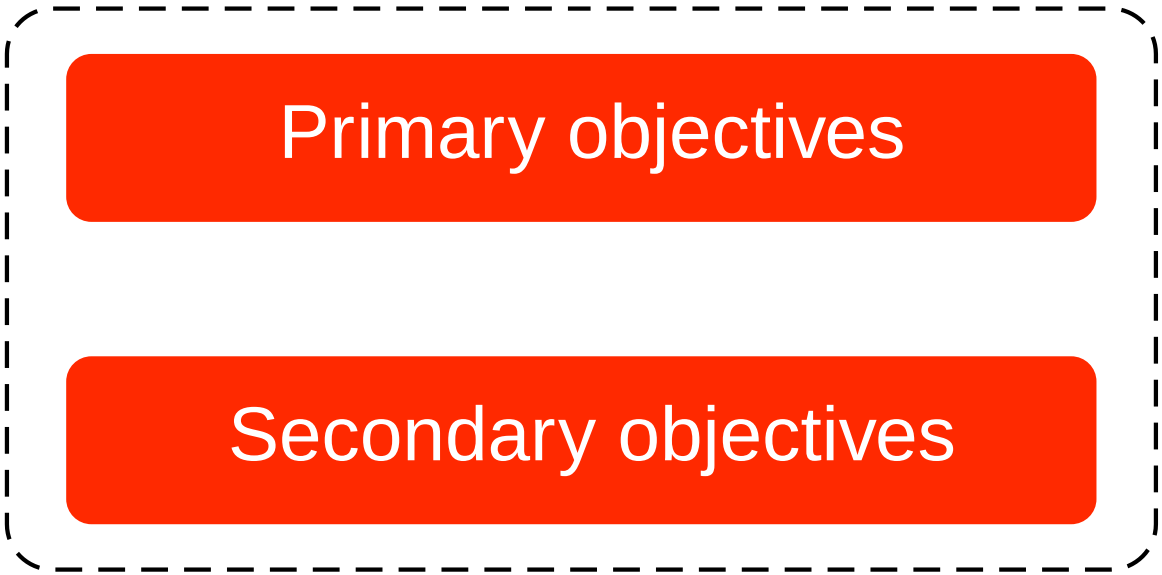
Secondary objectives can be included in the product label to provide important information for prescribing physicians

Tertiary objectives

Play a general supportive role

Multiple families of null hypotheses

Hierarchy of objectives



α control

Tertiary objectives

No α control

Gatekeeping procedures

Definition

Multiple testing procedures for multiple families of null hypotheses

Type I error rate

Control overall Type I error rate

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)

General gatekeeping procedures based on Bonferroni and more powerful tests (Dmitrienko and Tamhane, 2011; Kordzakhia and Dmitrienko, 2013)

Review paper and Tutorial in Biostatistics

Dmitrienko, D'Agostino and Huque (2013) and Dmitrienko and D'Agostino (2013)

Example: Phase III development program

Schizophrenia program

Latuda (lurasidone) Phase III trials

Multiple doses

Three doses versus placebo

Multiple endpoints

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

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

Example

Multiple objectives

Multiple doses: Improve success probability

Multiple endpoints: Strengthen the lurasidone product label and create differentiating factors

Gatekeeping strategy

Powerful Hommel-based gatekeeping procedure (Brechenmacher, Xu, Dmitrienko, Tamhane, 2011)

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

Selection of secondary endpoints/objectives

Product label

Secondary objectives can be included in the product label

Help provide important information for prescribing physicians, patients, payers, etc

Large number of secondary objectives

Increased pressure to strengthen product labels/prescribing information and limited scientific advice

Phase III protocols are overloaded with secondary objectives

Selection of secondary endpoints/objectives

Proposal

Develop a list of most relevant indication-specific endpoints or objectives

Example: Treatment of migraine

Primary endpoint: Pain intensity

Key secondary endpoints: Nausea, Photo-sensitivity, Phono-sensitivity

Based on International Headache Society's recommendations

Pseudospecificity

Guidelines for selecting secondary endpoints

Secondary endpoints need to provide additional information on treatment's efficacy

Pseudospecificity

Secondary/supportive endpoints should be clinically independent of the primary endpoint

Supportive endpoints may be excluded if they are perceived to be clinically related to the primary endpoint

Example: Phase III development program

Major depressive disorder program

Primary endpoint: Montgomery-Asberg Depression Rating Scale total score [MADRS]

Key secondary endpoints

(S1) Sheehan Disability Scale Global Functional Impairment score, (S2) Fatigue Association with Depression total score, (S3) MADRS-based remission status at the end of the acute phase

Pseudospecificity

S3 was not accepted by regulators

Secondary endpoints/objectives within a development program

Phase III development program

Primary endpoint: Significant treatment effect is required across all trials with the development program

Secondary endpoints: Requiring consistency across all trials is overly conservative

Alternative approach to analyzing secondary endpoints

Pooled analysis across multiple trials stratified by trial (Wang et al., 2013)

Secondary endpoints/objectives within a development program

Improved power

Pooled analysis improves power of secondary analyses (secondary endpoints tend to be underpowered)

Improved interpretation of results

Pooled analysis removes the possibility of inconsistent results across trials within a program

With multiple secondary endpoints, there is a higher chance of observing inconsistency across trials

Selection of multiple testing procedures

Comprehensive quantitative evaluation

Important to evaluate the performance of several candidate procedures under a broad range of statistical assumptions to arrive at an optimal multiple testing procedure (optimal set of procedure parameters)

Clinical hypothesis evaluation framework

Developed in Benda et al. (2010)

Similar approaches to quantitative evaluation of candidate methods (Millen and Dmitrienko, 2011; Dmitrienko and D'Agostino, 2013)

Clinical hypothesis evaluation framework (CHEF)

Data models

Specify data generation to cover most plausible scenarios

Analysis models

Select multiple testing/gatekeeping procedures that utilize all available clinical and statistical information

Evaluation models

Specify success criteria (power functions) that reflect the overall definition of success in a trial

Selection of multiple testing procedures

Proposal

Clinical trial simulation reports are recommended for adaptive trial designs

Develop a standard simulation report to better understand/characterize the performance of gatekeeping procedures

Additional information

Multiplicity Expert web site

http://www.multxpert.com/

Multiple testing and gatekeeping procedures

Key papers, presentations and short courses on gatekeeping strategies on clinical trials

Software code (SAS macros and R functions)

References

Benda N, Branson M, Maurer W, Friede T. (2010). Aspects of modernizing drug development using clinical scenario planning and evaluation. *Drug Information Journal*. **44**, 299–315.

Brechenmacher T, Xu J, Dmitrienko A, Tamhane AC. (2011). A mixture gatekeeping procedure based on the Hommel test for clinical trial applications. *Journal of Biopharmaceutical Statistics*. **21**, 748–767.

Bretz F, Maurer W, Brannath W, Posch M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine*. **28**, 586–604.

Burman CF, Sonesson C, Guilbaud O. (2009). A recycling framework for the construction of Bonferroni-based multiple tests. *Statistics in Medicine*. **28**, 739–761.

References

Dmitrienko A, Tamhane AC. (2011). Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. *Statistics in Medicine*. **30**, 1473–1488.

Dmitrienko A, D'Agostino RB, Huque M. (2013). Key multiplicity issues in clinical drug development. *Statistics in Medicine*. **32**, 1079–1111.

Dmitrienko A, D'Agostino RB (2013). Tutorial in Biostatistics: Traditional Multiplicity Adjustment Methods in Clinical Trials. *Statistics in Medicine*. **32**, 5172–5218.

Kordzakhia G, Dmitrienko A. (2013). Superchain procedures in clinical trials with multiple objectives. *Statistics in Medicine*. **32**, 486–508.

References

Meltzer et al. (2011). Lurasidone in the treatment of schizophrenia: a randomized, double-blind, placebo- and olanzapine-controlled study. *American Journal of Psychiatry*. **168**, 957–967.

Millen B, Dmitrienko A. (2011). Chain procedures: A class of flexible closed testing procedures with clinical trial applications. *Statistics in Biopharmaceutical Research*. **3**, 14–30.

Wang SJ, Bretz F, Dmitrienko A, Hsu J, Hung J, Huque M, Koch G. (2013). Panel forum on multiple comparison procedures: a commentary from a complex trial design and analysis plan. *Biometrical Journal*. **55**, 275–293.