Key Multiplicity Issues in Clinical Trials

Alex Dmitrienko (Mediana Inc)

Biopharmaceutical Section's webinar series, June 2017

Outline

Key topics

Overview of multiplicity issues in Phase III trials

Traditional multiplicity problems

Advanced multiplicity problems

Commonly used multiple testing procedures

Power and sample size calculations

Overview of Multiplicity Issues in Clinical Trials

Multiplicity in clinical trials

Drug development challenges

Drug development costs have been increasing steadily

More sophisticated trial designs are used to improve efficiency of drug development programs

Example: Designs with increasingly more complex objectives

Multiplicity issues

Multiple objectives induce multiplicity and increase false-positive rates

Multiplicity issues in clinical trials

Multiplicity adjustment

Multiplicity adjustment methods are required in trials with multiple objectives

Regulatory guidance documents

U.S. Food and Drug Administration (FDA)

European Medicines Agency (EMA)

Multiplicity issues in clinical trials

FDA guidance

Draft guidance on multiplicity issues in clinical trials (January 2017)

EMA guidance

Points to consider on multiplicity issues in clinical trials (September 2002)

Draft guideline on multiplicity issues in clinical trials (April 2017)

Traditional multiplicity problems

Examples

Multiple primary endpoints

Multiple doses and regimens versus common control (e.g., placebo)

Multiple patient populations (overall population and marker-positive subpopulation)

Traditional multiplicity problems

Single family of null hypotheses

$$H_1, \ldots, H_m$$

Trials with a single source of multiplicity

Advanced multiplicity problems

Examples

Multiple endpoints and multiple dose-placebo comparisons

Multiple endpoints and multiple patient populations

Advanced multiplicity problems

Multiple families of null hypotheses

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

$\ldots$

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

Trials with multiple sources of multiplicity

Books

Analysis of Clinical Trials Using SAS

Edited by Alex Dmitrienko (Mediana) and Gary Koch (UNC-Chapel Hill)

Published by SAS Press in 2017

Chapter 5: Multiplicity adjustment methods

Introduction to multiplicity problems arising in clinical trials, popular multiple testing procedures and gatekeeping procedures

Books

Multiple Testing Problems in Pharmaceutical Statistics

Edited by Alex Dmitrienko (Eli Lilly), Ajit Tamhane (Northwestern University), Frank Bretz (Novartis, Hannover Medical School)

Published by Chapman and Hall/CRC Press in 2009

Comprehensive summary of methodological, regulatory and practical issues related to multiplicity problems in pre-clinical research and clinical trials

Review papers

Recent review papers and tutorials

Dmitrienko, D'Agostino and Huque. (2013). Key multiplicity issues in clinical drug development.

Dmitrienko and D'Agostino. (2013). Tutorial in Biostatistics: Traditional multiplicity adjustment methods in clinical trials.

Alosh, Bretz and Huque (2014). Advanced multiplicity adjustment methods in clinical trials.

Online training

Instant Training web site

http://sprmm.com/biostatistical-training/

Available 24 hours a day/7 days a week anywhere in the world

Multiplicity training courses

Traditional multiplicity problems: Key Multiplicity Issues in Clinical Trials (Part I)

Advanced multiplicity problems: Key Multiplicity Issues in Clinical Trials (Part II) [to be released in the summer of 2017]

Overall plan

General approach

Focus on key concepts and fundamental principles

Case study-driven summary of commonly used approaches to multiplicity adjustment

Traditional Multiplicity Problems

Inferential goals

Multiple testing problem

Inferences used in a multiple testing problem depend on the inferential goal

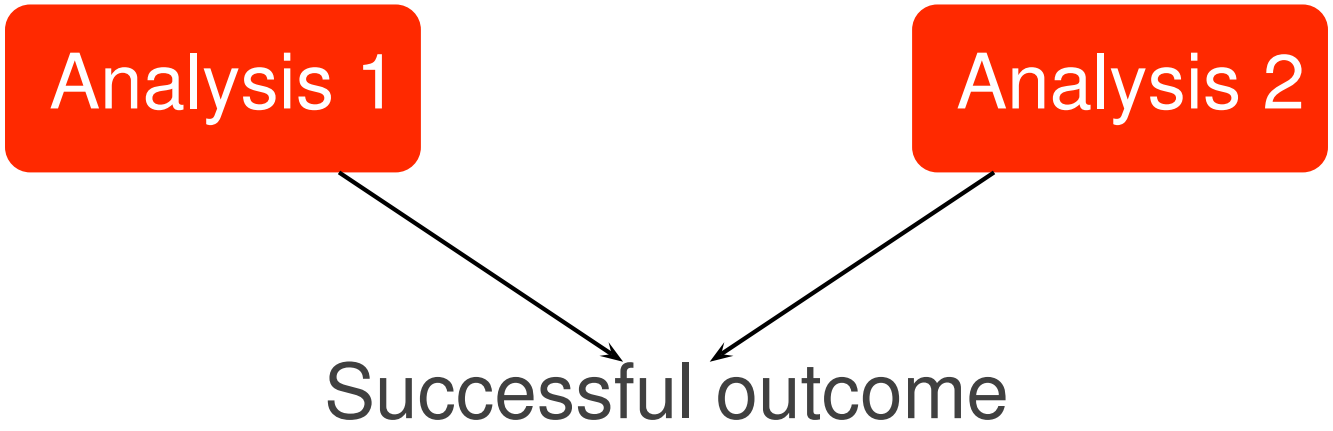
Two inferential goals

Individual analyses separately lead to a successful outcome (at-least-one procedures, also known as multiple testing procedures)

Individual analyses jointly lead to a successful outcome (all-or-none procedures)

At-least-one setting

Each analysis is independently clinically relevant



Each analysis independently provides a proof of efficacy

The trial's outcome is declared positive if at least one analysis is significant

Multiple endpoints

Prostate cancer trial

Objective

Evaluate the effects of an experimental treatment (enzalutamide) on progression-free and overall survival (Beer at al., 2014)

Design

Experimental treatment versus placebo

Multiple endpoints

Two primary endpoints

Endpoint 1: Radiographic progression-free survival (rPFS)

Endpoint 2: Overall survival (OS)

Overall analysis

At least one endpoint must be significant

Multiple dose-placebo comparisons

Type 2 diabetes trial

Objective

Evaluate the efficacy of an experimental treatment (saxagliptin) in treatment-naive patients with Type 2 diabetes (Rosenstock et al., 2009)

Primary endpoint

HbA1c change from baseline to Week 24

Design

Three dose groups versus placebo and at least one dose must be significant

Multiple patient populations

Non-small-cell lung cancer trial

Objective

Evaluate the effects of a treatment (erlotinib) in advanced non-small-cell lung cancer (SATURN trial, Cappuzzo et al., 2010)

Primary endpoint

Progression-free survival (PFS)

Design

Treatment versus placebo

Multiple patient populations

Two patient populations

Tailored therapy approach is implemented in this trial

General population

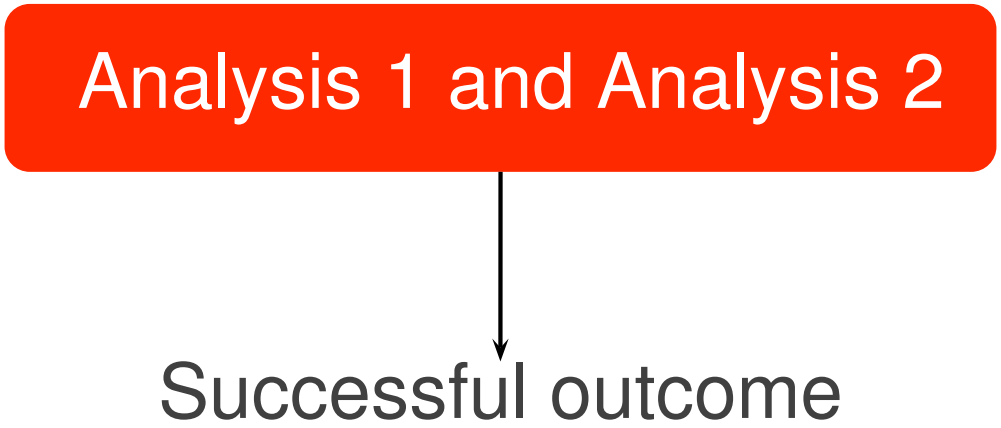
Subpopulation of patients with EGFR (epidermal growth factor receptor) immunohistochemistry-positive tumors

Overall analysis

Treatment effect in at least one population must be significant

All-or-none setting

All analyses must show benefit



The trial's outcome is positive if all analyses produce a significant outcome

Multiple endpoints

Alzheimer's disease trial

Objective

Evaluate the effects of a treatment (rivastigmine) on cognition and global changes in patients with mild to moderate Alzheimer's disease (IDEAL study, Winblad et al., 2007)

Design

Treatment versus placebo

Multiple endpoints

Two co-primary endpoints

Endpoint 1: Cognition endpoint (Alzheimer's Disease Assessment Scale-Cognitive subscale)

Endpoint 2: Clinical global scale (Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change)

Overall analysis

Both endpoints must be significant

All-or-none setting

Intersection-union problem

Problem is known as the intersection-union problem and does not require a multiplicity adjustment

Decision rule

$H_1, \ldots, H_m$, Null hypotheses

$p_1, \ldots, p_m$, p-values

α, Type I error rate, e.g., $\alpha = 0.025$

All null hypotheses are rejected if $p_1 \leq \alpha$, ..., $p_m \leq \alpha$

All-or-none setting

FDA guidance (Section III.C)

"There have been suggestions that the statistical testing criteria for each co-primary endpoint could be relaxed (e.g., testing at an alpha of 0.06 or 0.07)... Relaxation of alpha is generally not acceptable because doing so would undermine the assurance of an effect on each disease aspect considered essential to showing that the drug is effective in support of approval."

Multiplicity issues in clinical trials

At-least-one setting

Analysis of multiple objectives in an at-least-one setting induces multiplicity and increases the false-positive rate (familywise Type I error rate)

Multiplicity adjustments

Multiplicity adjustments (multiple testing procedures) are mandated in Phase III trials with multiple objectives to control the Type I error rate

Importance of addressing multiplicity issues

EMA guidance (Section 5)

"A clinical study that requires no adjustment of the significance level of elementary hypothesis tests (i.e. single statistical tests on one parameter only) is one that consists of two treatment groups, which uses a single primary variable, and has a confirmatory statistical strategy that pre-specifies just one single null hypothesis relating to the primary variable and no interim analysis"

Commonly Used
Multiple Testing Procedures

Multiplicity problem

Notation

$H_1, \ldots, H_m$, Null hypotheses of interest

α, Familywise error rate, e.g., one-sided $\alpha = 0.025$

P-values

$p_1, \ldots, p_m$, Original treatment effect p-values

$p_{(1)} < \ldots < p_{(m)}$, Ordered p-values

Classification schemes

Clinical information

Classification scheme based on clinically relevant logical relationships among the null hypotheses

Single-step and stepwise procedures

Statistical information

Classification scheme based on distributional relationships, i.e., the joint distribution of the hypothesis test statistics

Nonparametric, semiparametric and fully parametric procedures

Classification based on logical relationships

Basic single-step testing approach

Null hypotheses are tested simultaneously or in a single step

Clinically meaningful relationships among null hypotheses are not taken into account

Examples: Bonferroni and Dunnett procedures

Stepwise testing approach

Null hypotheses are ordered using clinical importance or using significance of test statistics/p-values

Classification based on logical relationships

Pre-specified testing sequence

Null hypotheses are ordered at the design stage to reflect clinical importance or probability of success for associated objectives

Examples: Fixed-sequence, fallback and chain procedures

Multiple dose-placebo comparisons

Strong evidence of a positive dose-response relationship: Doses are tested sequentially beginning with the highest dose

Classification based on logical relationships

Data-driven testing sequence

Null hypotheses are not ordered at the design stage

Examples: Holm, Hommel, Hochberg and step-down Dunnett procedures

Multiple dose-placebo comparisons

Difficult to assume a positive dose-response relationship: Doses are tested in the order determined by significance of test statistics

Classification based on distributional relationships

Nonparametric procedures

Based on univariate p-values and impose no distributional assumptions

Examples: Bonferroni, Holm, fixed-sequence, fallback and chain procedures

Properties

Very popular due to their simplicity

Tend to perform poorly with too many null hypotheses or strongly correlated hypothesis test statistics

Classification based on distributional relationships

Semiparametric procedures

Based on univariate p-values and impose some distributional assumptions (multivariate normal distribution of hypothesis test statistics with non-negative correlations)

Examples: Hochberg and Hommel procedures

Properties

More powerful than nonparametric procedures

Classification based on distributional relationships

Parametric procedures

Based on multivariate p-values computed from a pre-specified joint distribution of test statistics (multivariate normal or t distribution)

Example: Single-step and step-down Dunnett procedures

Properties

More powerful than nonparametric and semiparametric procedures

Classification based on distributional relationships

Resampling-based procedures

Do not make distributional assumptions and approximate true joint distribution of test statistics using bootstrap or permutation methods

Not used in Phase III trials

FDA guidance (Section IV.C)

"Resampling methods are not recommended as primary analysis methods for adequate and well-controlled trials in drug development"

Case Study

Type 2 diabetes trial

Objective

Evaluate the efficacy of three doses of an experimental treatment in patients with Type 2 diabetes

Primary endpoint

HbA1c change from baseline to Week 24

Design

Three dose groups (Dose 1, Dose 2 and Dose 3) versus placebo

Multiplicity problem

Null hypotheses

H_1: No difference between Dose 1 (high dose) and placebo

H_2: No difference between Dose 2 (medium dose) and placebo

H_3: No difference between Dose 3 (low dose) and placebo

Candidate procedures

Parametric procedure

Dunnett procedures (Dunnett, 1955; Dunnett and Tamhane, 1991)

Nonparametric procedures

Holm (Holm, 1979) and chain procedures

Semiparametric procedure

Hochberg procedure (Hochberg, 1988)

Dunnett procedure

Distributional assumptions

Test statistics associated with H_1, H_2 and H_3 follow a multivariate normal distribution with known pairwise correlations, e.g., pairwise correlations are equal to 1/2 in a balanced design

Adjusted significance level

Significance level c is computed under the null distribution from

$$P(p_1 \leq c \text{ or } p_2 \leq c \text{ or } p_3 \leq c) = \alpha$$

Dunnett procedure

Key properties

Powerful procedure that takes into account the joint distribution of the hypothesis test statistics but very sensitive to the assumption that pairwise correlations are known

Example

Can Dunnett procedure be used in a trial with lots of missing observations?

If complex imputation techniques are applied, pairwise correlations are no longer known and Type I error rate control cannot be guaranteed

Holm and Hochberg procedures

Stepwise procedures

Holm is a nonparametric step-down procedure (testing begins with the smallest p-value)

Hochberg is a semiparametric step-up procedure (testing begins with the largest p-value)

Ordered hypotheses

$H_{(1)}$, $H_{(2)}$ and $H_{(3)}$ correspond to $p_{(1)}$, $p_{(2)}$ and $p_{(3)}$

Holm procedure

Testing algorithm

If $p_{(1)} \leq \alpha/3$, reject $H_{(1)}$; Otherwise stop and accept all hypotheses

If $p_{(2)} \leq \alpha/2$, reject $H_{(2)}$; Otherwise stop and accept $H_{(2)}$ and $H_{(3)}$

If $p_{(3)} \leq \alpha$, reject $H_{(3)}$; Otherwise accept $H_{(3)}$

Hochberg procedure

Testing algorithm

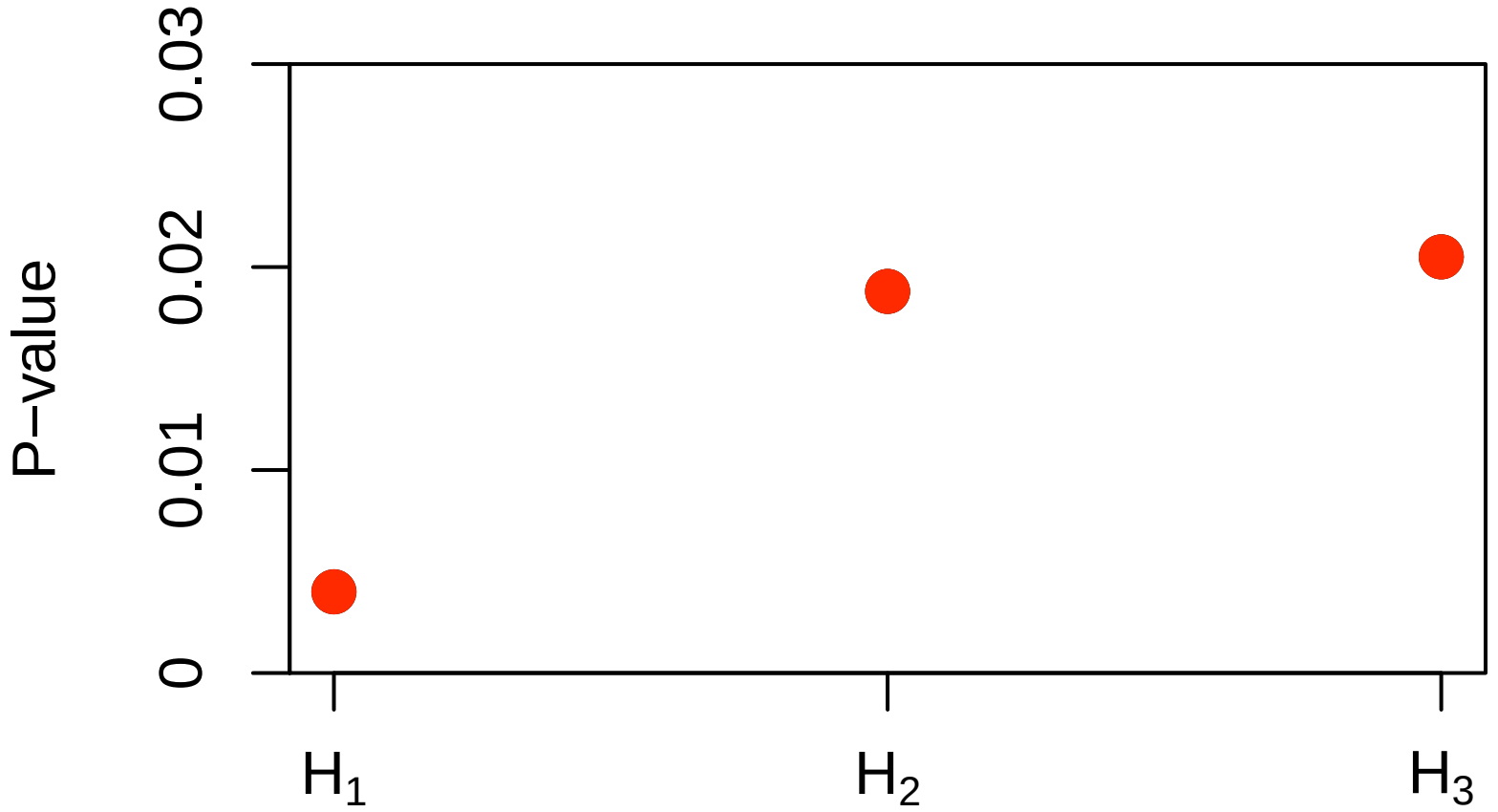
If $p_{(3)} \leq \alpha$, reject all hypotheses; Otherwise continue to $p_{(2)}$

If $p_{(2)} \leq \alpha/2$, reject $H_{(1)}$ and $H_{(2)}$; Otherwise continue to $p_{(1)}$

If $p_{(1)} \leq \alpha/3$, reject $H_{(1)}$

Numerical example

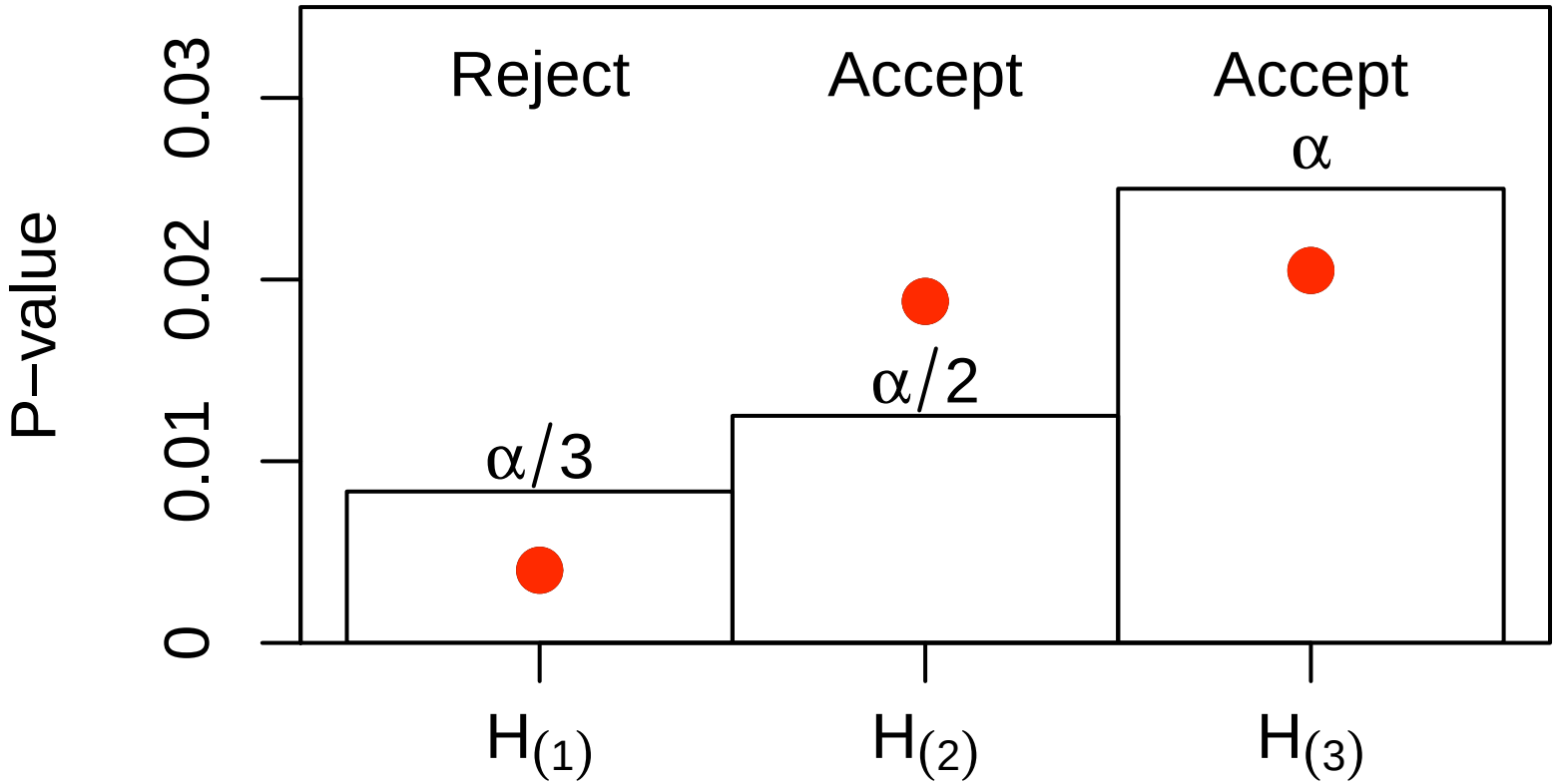
Positive dose-response relationship



One-sided treatment effect p-values: $p_1 = p_{(1)} = 0.0040$, $p_2 = p_{(2)} = 0.0188$, $p_3 = p_{(3)} = 0.0205$

Numerical example

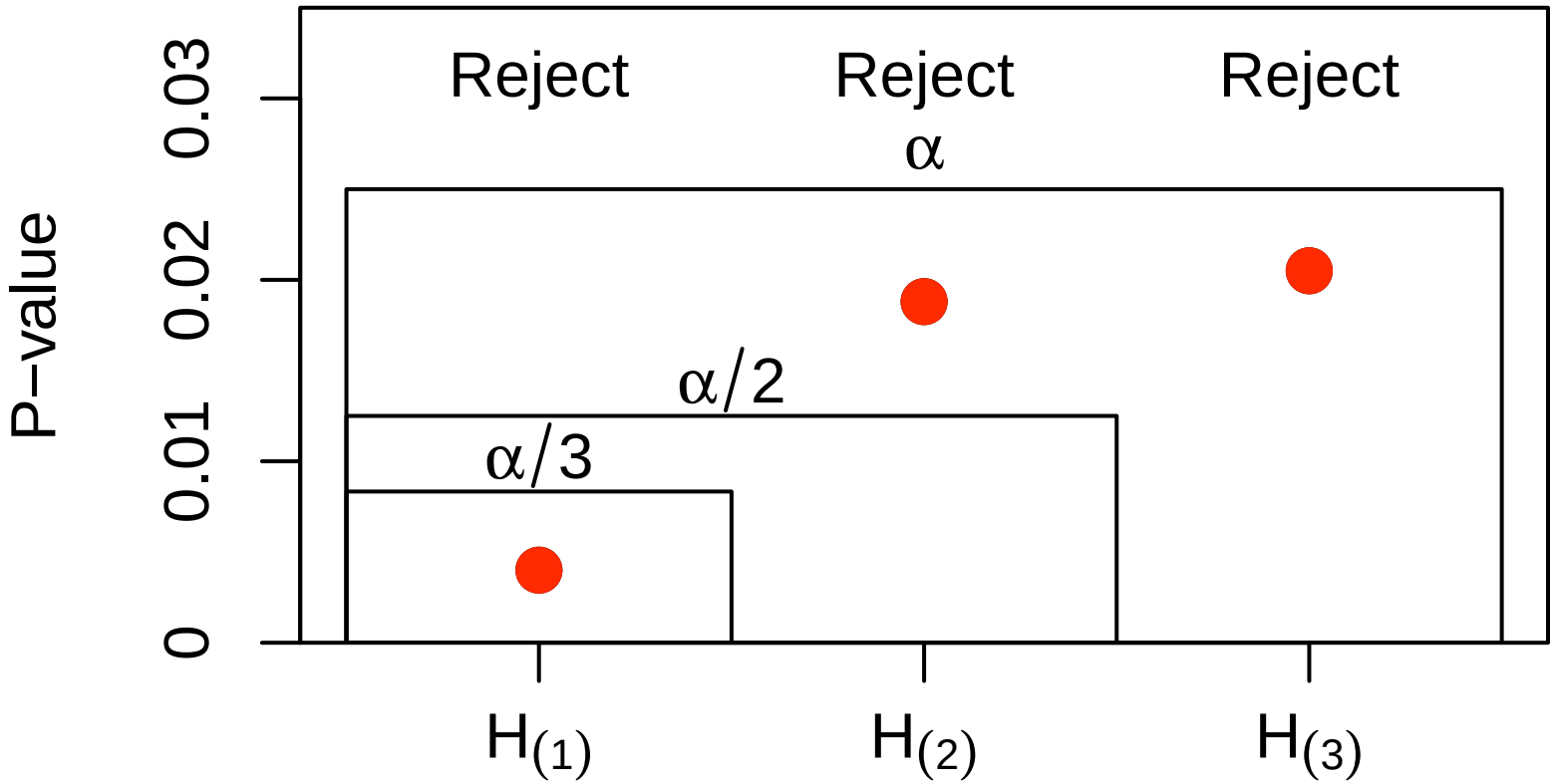
Holm procedure



Holm procedure rejects H_1 (significant effect at the high dose)

Numerical example

Hochberg procedure



Hochberg procedure rejects all null hypotheses (significant effects at all doses)

Semiparametric procedures

Power

Hochberg is uniformly more powerful than Holm

Hommel is uniformly more powerful than Hochberg

Type I error rate

Semiparametric procedures provide Type I error rate control under flexible distributional assumptions

Semiparametric procedures can be safely used with any imputation method for missing data

Semiparametric procedures

Positive dependence condition

Simes global test controls Type I error rate when the positive dependence condition is satisfied, i.e., the joint distribution of hypothesis test statistics is multivariate totally positive of order two (MTP2) (Sarkar and Chang, 1997; Sarkar, 1998; see also Huque, 2016)

Positive dependence condition is satisfied for multivariate normal test statistics with non-negative pairwise correlations (Sarkar, 2008)

Positive dependence condition

Multiple endpoints: Prostate cancer trial

Condition is satisfied if the two endpoints are positively correlated

Multiple doses: Type 2 diabetes trial

Condition is satisfied since the doses are compared to a common control

Multiple populations: Non-small-cell lung cancer trial

Condition is satisfied since the subpopulation is a subset of the overall population

Semiparametric procedures

FDA guidance (Section IV.C)

"Beyond the aforementioned cases where the Hochberg procedure is known to be valid, its use is generally not recommended for the primary comparisons of confirmatory clinical trials unless it can be shown that adequate control of Type I error rate is provided."

Recommendation

Semiparametric procedures (Hochberg and Hommel) provide Type I error rate control under very broad assumptions

Chain procedures

Class of chain procedures/graphical procedures

Fixed-sequence procedure: Pre-specified hypothesis ordering

Fallback procedure: Pre-specified hypothesis ordering (Wiens, 2003; Wiens and Dmitrienko, 2005)

General chain procedures: Data-driven hypothesis ordering (Bretz et al., 2009; Burman et al., 2009)

Fixed-sequence procedure

Testing algorithm

$w_1 = 1$ $w_2 = 0$ $w_3 = 0$

α allocation rule: Defines hypothesis weights w_1, w_2 and w_3

α propagation rule: Defines transition parameters g_{12} and g_{13}

Fixed-sequence procedure

Inflexible sequentially rejective algorithm

$w_1 = 1$ $\quad$ $w_2 = 0$ $\quad$ $w_3 = 0$

Reject H_1 if $p_1 \leq \alpha$

Reject H_2 if $p_2 \leq \alpha$ and H_1 is rejected

Reject H_3 if $p_3 \leq \alpha$ and H_1 and H_2 are both rejected

This testing strategy is risky if the effect size at Dose H is small

Fallback procedure

More flexible testing algorithm

$w_1 = w$ $\qquad$ $w_2 = 1 - w$ $\qquad$ $w_3 = 0$

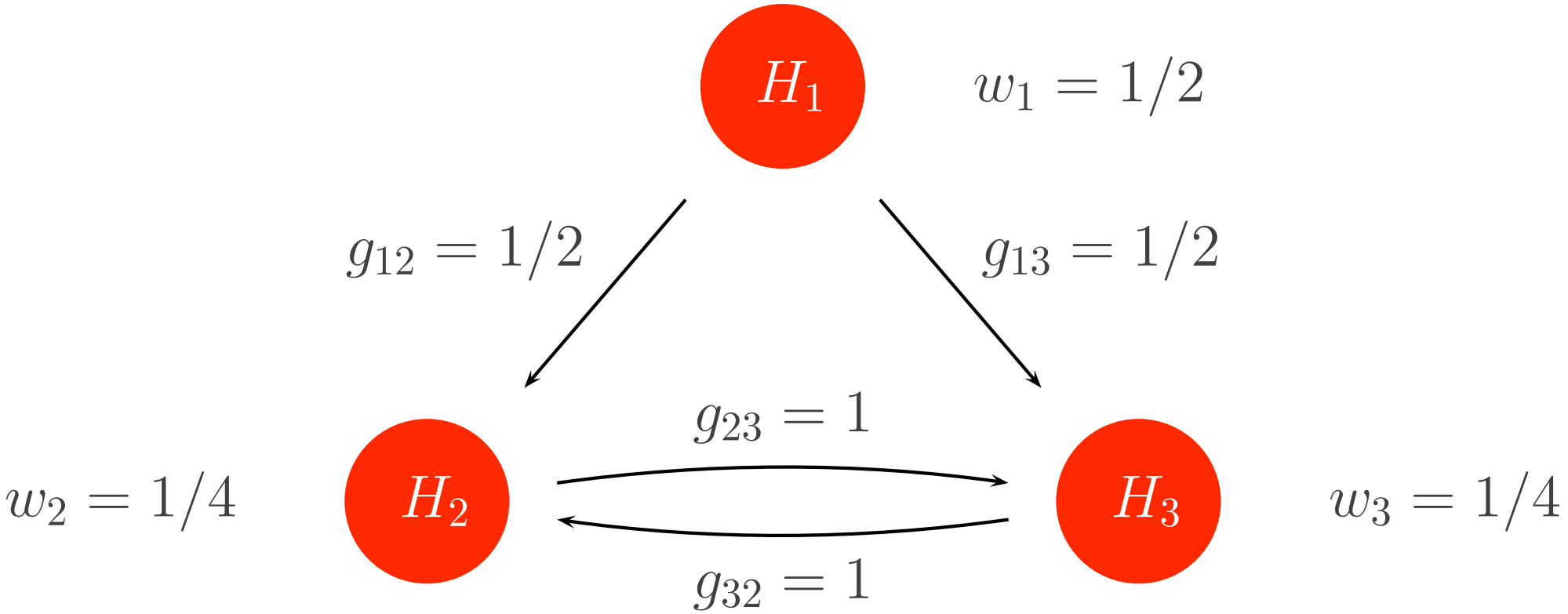
Reject H_1 if $p_1 \leq w\alpha$

Reject H_2 if (1) $p_2 \leq \alpha$ and H_1 is rejected or (2) $p_2 \leq (1 - w)\alpha$ and H_1 is not rejected

Reject H_3 if (1) $p_3 \leq \alpha$ and H_1 and H_2 are both rejected or (2) $p_3 \leq (1 - w)\alpha$, H_1 is not rejected but H_2 is rejected

General chain procedures

Very flexible testing algorithms



Summary

Parametric procedure

Dunnett is powerful but not robust (may not control Type I error rate)

Nonparametric procedures

Holm is robust but may lack power

Chain procedures are very flexible procedures but may lack power

Semiparametric procedures

Hochberg and Hommel are more robust than Dunnett and more powerful than Holm

Advanced Multiplicity Problems

Advanced multiplicity problems

Multiple families of null hypotheses

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

$\ldots$

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

Trials with multiple sources of multiplicity

Multiple objectives in clinical trials

Hierarchy of multiple objectives

Primary objectives

Secondary objectives

Exploratory objectives

Primary and secondary objectives

FDA guidance (Section II.A)

"The set of primary endpoints consists of the outcome or outcomes (based on the drug's expected effects) that establish the effectiveness, and/or safety features, of the drug in order to support regulatory action... Secondary endpoints may be selected to demonstrate additional effects after success on the primary endpoint... All other endpoints are referred to as exploratory."

Primary and secondary objectives

EMA guidance (Section 6)

"Secondary endpoints may provide additional clinical characterisation of treatment effects but are, by themselves, not sufficiently convincing to establish the main evidence in an application for a licence or for an additional labelling claim."

Multiple objectives in clinical trials

Primary objectives

Directly related to the trial's outcome and presented in product label using inferential statements

Example: p-values and/or confidence intervals

Secondary objectives

Provide key supportive evidence and presented in product label using inferential statements

Example: p-values and/or confidence intervals

Multiple objectives in clinical trials

Exploratory objectives

Play a general supportive role and presented in product label using descriptive statements

Example: Descriptive statistics or plots (survival curves)

P-values or confidence intervals may not be used

Multiple objectives in clinical trials

Pseudospecificity

FDA's restrictions on secondary objectives in product labels

Key secondary objectives (secondary endpoints) should provide additional information on the treatment's efficacy

Secondary objectives should not be clinically related to the primary objective

Pseudospecificity

Phase III development program in major depressive disorder

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

S3 was not accepted since it was closely related to the primary endpoint

Gatekeeping Procedures

Gatekeeping procedures

Definition

Multiple testing procedures for multiple families of null hypotheses

Global Type I error rate control

Regulatory requirement

Control global familywise error rate over multiple families

Helps provide important information on secondary objectives for prescribing physicians, patients, etc

Gatekeeping procedures

Optimal distribution of power

Trial sponsor's requirement

Maximize power by accounting for hierarchical structure of multiple families

Example: Maximize power in the primary family which may reduce power in the other families

Fundamental principles

Trial information

It is important to account for trial-specific information

Clinical information: Logical restrictions among null hypotheses in different families

Statistical information: Distributional information on hypothesis test statistics

Classification based on logical relationships

Logical restrictions

Based on clinically relevant logical dependencies among the null hypotheses

Different types of gatekeeeers

Serial gatekeepers

Parallel gatekeepers

General gatekeepers

Classification based on distributional information

Nonparametric gatekeeping procedures

Extension of chain procedures (Bretz et al., 2009; Burman et al., 2009)

Semiparametric and parametric gatekeeping procedures

General mixture-based approach to defining gatekeeping procedures (Dmitrienko and Tamhane, 2011, 2013)

Clinical trial application

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

Clinical trial application

Gatekeeping procedure

Mixture-based approach was applied to defining gatekeeping procedures in the lurasidone Phase III trials

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

Importance of gatekeeping procedures was recognized in the clinical trial publications (Meltzer et al., 2011; Nasrallah et al., 2013)

Power and Sample Size Calculations

Power and sample size calculations

Analytical approach

Closed-form expressions used in sample size calculations often rely on simplifying/artificial assumptions

Simulation-based approach

Much more reliable approach to power and sample size calculations in trials with complex clinical objectives

Simulation-based approaches

FDA guidance (Section III.B)

"Determination of an appropriate study sample size to ensure that the study is appropriately powered can be difficult in these cases, and often will be dependent upon computer simulations rather than an analytic formula, which can be used for simpler situations"

Simulation-based approaches

EMA guidance (Section 5)

"Sometimes a series of related objectives is pursued in the same trial, each with its own primary variable... In these situations planning of the sample size becomes more complex due to the different alternative hypotheses related to the different endpoints and due to the assumed correlation between endpoints."

Sample size and power calculations

Success criterion

Important to select a success criterion that is aligned with the trial's goals, e.g., disjunctive power (probability of meeting at least one objective) or weighted power (weighted sum of marginal power functions)

Software

Mediana package is an R package that supports simulation-based power calculations for a broad class of multiple testing procedures

Mediana package

Software implementation

Mediana package also provides software implementation of commonly used multiple testing procedures

Traditional multiplicity problems: Popular nonparametric, semiparametric and parametric procedures

Advanced multiplicity problems: Several classes of gatekeeping procedures, including parallel, multiple-sequence and general mixture-based gatekeeping procedures

Mediana package

Release

First version (Version 1.0.1) was released in July 2015

Latest version (Version 1.0.5) was released in May 2017

CRAN web site

https://cran.r-project.org/web/packages/Mediana

Online manual

http://gpaux.github.io/Mediana/

Clinical scenario evaluation

General framework

Clinical scenario evaluation (CSE) approach was developed in Benda et al. (2010), Friede et al. (2010) and other publications

Motivation

Clinical trial researchers have recognized the importance of employing quantitative, comprehensive and disciplined approaches to evaluating the design and analysis of clinical trials to enable better decision making

Clinical trial optimization

Clinical scenario evaluation

An important application of general CSE approach is clinical trial optimization, e.g., optimal selection of multiple testing procedures and their parameters

General goal

Inform decision making in Phase III trials and maximize the overall probability of success

Clinical trial optimization

Publications

Review of general approaches to clinical trial optimization (*Clinical Trial Optimization Using R* edited by Dmitrienko and Pulkstenis, 2017)

Optimal selection of multiplicity adjustments in Phase III trials (Dmitrienko, Paux and Brechenmacher, 2015)

Optimal selection of multiplicity adjustments and adaptive trial designs in Phase II and III trials (Dmitrienko, Paux, Pulkstenis and Zhang, 2016)

Summary

Summary

Multiplicity adjustments

Multiplicity adjustments are required by regulatory agencies in Phase III trials with multiple objectives to control the Type I error rate

Multiple testing procedures

Powerful and flexible multiple testing procedures are available to address multiplicity issues

Thank you!

Alex Dmitrienko (alex.dmitrienko@medianainc.com)

References

References

Alosh, M., Bretz, F., Huque, M. (2014). Advanced multiplicity adjustment methods in clinical trials. Statistics in Medicine. 33, 693-713.

Beer, T.M. et al. (2014). Enzalutamide in metastatic prostate cancer before chemotherapy. The New England Journal of Medicine. 371, 424-433.

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, A.C. (2011). A mixture gatekeeping procedure based on the Hommel test for clinical trial applications. Journal of Biopharmaceutical Statistics. 21, 748-767.

References

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

Cappuzzo, F. et al. (2010). Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: A multicentre, randomized, placebo-controlled phase 3 study. Lancet Oncology. 11, 521-529.

Dmitrienko, A., Tamhane, A.C., Bretz, F. (editors). (2009). Multiple Testing Problems in Pharmaceutical Statistics. Chapman and Hall/CRC Press, New York.

References

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

Dmitrienko, A., D'Agostino, R.B., Huque, M.F. (2013). Key multiplicity issues in clinical drug development. Statistics in Medicine. 32, 10791111.

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

Dmitrienko, A., Tamhane, A.C. (2013). General theory of mixture procedures for gatekeeping. Biometrical Journal. 55, 402-419.

References

Dmitrienko, A., Paux, G., Brechenmacher, T. (2015). Power calculations in clinical trials with complex clinical objectives. Journal of the Japanese Society of Computational Statistics. 28, 15-50.

Dmitrienko, A., Paux, G., Pulkstenis, E., Zhang, J. (2016). Tradeoff-based optimization criteria in clinical trials with multiple objectives and adaptive designs. Journal of Biopharmaceutical Statistics. 26, 120-140.

Dmitrienko, A., Pulkstenis, E. (editors). (2017). Clinical Trial Optimization Using R. Chapman and Hall/CRC Press, New York.

Dmitrienko, A., Koch, G. (editors). (2017). Analysis of Clinical Trials Using SAS (Second Edition). SAS Press, Cary, NC.

References

Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association. 50, 1096-1121.

Dunnett, C.W., Tamhane, A.C. (1991). Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts. Statistics in Medicine. 10, 939-947.

Friede, T., Nicholas, R., Stallard, N., Todd, S., Parsons, N. R., Valdes-Marquez, E., Chataway, J. (2010). Refinement of the clinical scenario evaluation framework for assessment of competing development strategies with an application to multiple sclerosis. Drug Information Journal. 44, 713-718.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple significance testing. Biometrika. 75, 800-802.

References

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 6, 65-70.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika. 75, 383-386.

Huque, M.F., Dmitrienko, A., D'Agostino, R.B. (2013). Multiplicity issues in clinical trials with multiple objectives. Statistics in Biopharmaceutical Research. 5, 321-337.

Meltzer, H.Y. 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.

References

Nasrallah, H.A. et al. (2013). Lurasidone for the treatment of acutely psychotic patients with schizophrenia: A 6-week, randomized, placebo-controlled study. Journal of Psychiatric Research. 47, 670-677.

Rosenstock, J. et al. (2009). Effect of saxagliptin monotherapy in treatment-naive patients with type 2 diabetes. Current Medical Research And Opinion. 25, 2401-2411.

Sarkar, S., Chang, C.K. (1997). Simes' method for multiple hypothesis testing with positively dependent test statistics. Journal of the American Statistical Association. 92, 1601-1608.

References

Sarkar, S.K. (1998). Some probability inequalities for censored MTP2 random variables: A proof of the Simes conjecture. The Annals of Statistics. 26, 494-504.

Sarkar, S.K. (2008). On the Simes inequality and its generalization. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen. Balakrishnan, N., Pena, E.A., Silvapulle, M.J. (editors). Institute of Mathematical Statistics, Beachwood, Ohio, 231-242.

Wiens, B. (2003). A fixed-sequence Bonferroni procedure for testing multiple endpoints. Pharmaceutical Statistics. 2, 211-215.

References

Wiens, B., Dmitrienko, A. (2005). The fallback procedure for evaluating a single family of hypotheses. Journal of Biopharmaceutical Statistics. 15, 929-942.

Winblad, B., Grossberg, G., Frlich, D., Farlow, M., Zechner, S., Nagel, J., Lane, R. (2007). A 6-month, double-blind, placebo-controlled study of the first skin patch for Alzheimer disease. Neurology. 69, S14-S22.