Key Multiplicity issues. Multistage gatekeeping procedures.

George Kordzakhia,
FDA/CDER/OB/DB I

May 21, 2013

Disclaimer

The views and opinions presented in this talk are solely those of the authors and should not be taken to represent policies or regulations associated with the US Food and Drug Administration in any meaningful way.

Introduction

Multiple hypotheses

Clinical trials with multiple objectives, multiple claims pursued.

Multiple facets of the data result from

- Multiple endpoints.
- Multiple dose levels.
- Multiple timepoints
- Multiple patient subsets (populations)

Introduction

Classification of study endpoints

- Primary endpoint(s): essential basis for establishing drug effectiveness
- Secondary or Key Secondary endpoint(s):
  - selected to demonstrate additional meaningful drug effects that could be included as labeling claims.
  - neither sufficient (without success on the primary endpoint) nor required to demonstrate efficacy.
- Tertiary or Exploratory endpoints: not suited to serve regulatory purpose

Type I error

Overall type I error rate or familywide error rate (FWER):
Probability of incorrectly finding a statistically significant treatment effect for at least one study objective.

Control of type I error rate
required for null hypotheses that could lead to labeling claims.

Multiplicity adjustment

- **not needed** if it is required to win on all claims simultaneously (all-or-none-win criteria), e.g. multiple co-primary endpoints.
- **needed** if each individual hypothesis leads to a successful trial outcome (at-least one win criteria), e.g. one endpoint with multiple doses.

Type I error

Strong control of FWER

Error rate controlled under all possible configurations of true and false null hypotheses. Strong FWER control is required for regulatory purpose.

Weak control of FWER

Error rate controlled for the specific configuration in which all null hypotheses are true.

Type I error

Example. Depression clinical trial

Treatment arms

- Treatment drug
- Placebo

Efficacy endpoints

- Primary endpoint: MADRS
- Secondary endpoints; CGI-I and SDS.

Multiple Testing Approach I

- Test primary endpoint at $\alpha = 0.05$
- If significantly better than placebo, test secondary endpoints at 0.05 each

Type I error

- Weak control: Yes
- Strong control: No

Type I error

Multiple Testing Approach II

- Test primary endpoint at $\alpha = 0.05$
- If significantly better than placebo, test secondary endpoints at 0.025 each

Type I error

- Weak control: Yes
- Strong control: Yes

Multiple testing strategies

Standard multiple test procedures

Bonferroni, Holm, Hochberg, Hommel, Dunnett

Hierarchical and logical restrictions

- Complex relationships among null hypotheses in clinical trials.
- Power of multiple testing may be optimized by taking into account logical restrictions.

Recent developments

- Gatekeeping methods ([6], [9])
- Graphical methods ([5], [4])
- Adaptive alpha allocation, Consistency-adjusted strategies ([1], [2], [3])

Gatekeeping methods

Gatekeeping procedures

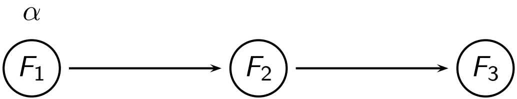
- Designed for multiplicity problems where null hypotheses grouped into families
- Multistage gatekeeping - based on stepwise algorithm.
- Account for complex relationships among the families
- Account for within-family correlations

Two basic types of family relationships

- Linear hierarchical structure
- Symmetric structure - no hierarchy.

Linear Hierarchical structure

- Families are ordered based on the clinical importance.
- Primary family F_1
- Family F_i serves as a gatekeeper for F_{i+1},
  - parallel gatekeeper: family F_{i+1} becomes testable only if at least one null hypothesis is rejected in family F_i.
  - serial gatekeeper: family F_{i+1} becomes testable only if all null hypotheses are rejected in family F_i.
- Families are tested sequentially.



Reference: Dmitrienko, Tamhane and Wiens [7].

Sequential Testing

Algorithm Overview (m families)

- Test primary family F_1.
- If at least one hypothesis is rejected in F_i, then F_{i+1} is tested $(i = 1, \dots, m-1)$. Otherwise stop.
- Test F_m. Algorithm stops.

Alpha assignment and propagation

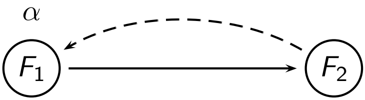
- Initially, full α assigned to F_1.
- Propagation unidirectional: from family F_i to family F_{i+1}.
- The more null hypotheses rejected in $F_1, \dots, F_i$, the larger significance level α_{i+1} for testing F_{i+1}.

Independence property: Primary inference in F_1 is independent of the testing outcomes for secondary families.

Sequential Testing with Retesting

Retesting Option

- Uniform power improvement
- Independence property does not hold. Primary inference depends on the testing outcomes for the secondary family.



Sequential Testing with Retesting

Two family problem: 3 step procedure

1. Test primary family F_1.
2. If at least one hypothesis is rejected in F_1, test F_2.
3. If all hypotheses in F_2 are rejected, F_1 is retested with a more powerful procedure.

Reference: Dmitrienko, Kordzakhia, Tamhane [8].

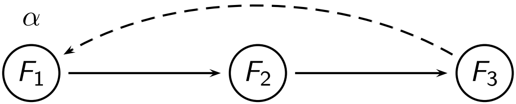
Sequential Testing with Multiple Retesting

Multiple families: Cyclic Stack Procedure

- Think of the families as a stack of objects with the primary family F_1 being on the top, and the family with the lowest rank, F_m, being at the bottom.
- The algorithm goes through the stack from the top to the bottom and returns directly to the top.

Algorithm Overview

- Start testing with the primary family F_1.
- If at least one hypothesis is rejected in the family F_i, the algorithm proceeds to testing family F_{i+1}.
- If at least one hypothesis is rejected in family F_m, the algorithm returns to family F_1.

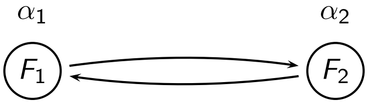


- If a family is fully rejected, it is removed from the stack.
- The method keeps retesting all families in the stack with increasingly more powerful tests as long as additional hypotheses are rejected at each step.

Symmetric Hierarchical structure

Superchain Procedures

- Families correspond to exchangeable sets of objectives, e.g. multiple primary endpoints, subpopulations.
- Exchangeable families can be tested simultaneously rather than sequentially.



Reference: Kordzakhia, Dmitrienko [10].

Simultaneous testing (Superchain procedures)

Algorithm Overview

- Test all families simultaneously.
- If at least one hypothesis is rejected in either family, then at the next step other families are retested with more powerful tests. Otherwise algorithm stops.

Alpha assignment and propagation

- Initially, positive fractions of α allocated to all families.
- Alpha propagation multidirectional: from each family F_i to each family F_j.
- The more null hypotheses rejected in F_i, the larger significance level α_j for testing F_j.

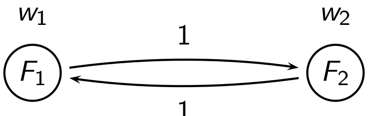
Simultaneous testing

Component Procedures

- Testing procedures $\mathcal{P}_i$ applied within families F_i
- $\mathcal{P}_i$ are based on closed testing principle and control type I error rates locally within F_i
- Allow to account for positive within-family correlations
- At each step of algorithm, families F_i are tested by $\mathcal{P}_i$ at "adaptive" significance levels α_i

Simultaneous testing for two families

- The non-negative weights w_1 and w_2 with $w_1 + w_2 = 1$ are assigned to F_1 and F_2 to quantify relative importance of the two families.
- At Step 1, families F_i are tested at respective significance levels $\alpha_i = w_i \cdot \alpha$.
- If one or more null hypotheses are rejected in a family, a fraction of FWER is "released" and transferred to the other family.
- The sets of rejected hypotheses in each family grow from one step to the next



Simultaneous testing for multiple families

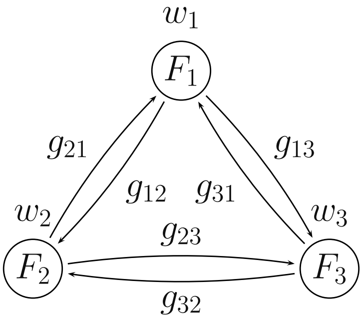
Building Tool

- The multistage procedure (superchain procedure) is defined via a directed graph and components procedures
- Nodes of the graph represent families.
- Directed edges represent connections among families. The families are connected to each other to account for clinically relevant relationship

Key Multiplicity issues. Multistage gatekeeping procedures.

Presentation

Graph for three-family problem



Simultaneous testing for multiple families

Overview

- Node weights w_i define the initial error rate allocation among the families based on clinical importance
- Connection weights g_{ij} define the rules for re-distributing the error rate released from family F_i to other families
- The multistage procedure is set up to control the global FWER across the families at pre-defined level α.

Reference: Kordzakhia, Dmitrienko [10].

Summary

Multiple testing problems

- Multiple sources of multiplicity
- Complex hierarchical objectives.

Criteria for multiple testing procedures

- Strong control of type I error
- Optimize power by accounting for
  - logical relationships
  - correlations
- Simple algorithm for implementation

- Li D, Mehrotra DV. An efficient method for accommodating potentially underpowered primary endpoints. *Statistics in Medicine* 2008; **27**:5377–5391.
- Alosh M, Huque M. A flexible strategy for testing subgroups and overall population. *Statistics in Medicine* 2009; **28**:3–23.
- Alosh M, Huque MF. A consistency-adjusted alpha-adaptive strategy for sequential testing. *Statistics in Medicine* 2010; **29**:1559–1571.
- Burman CF, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. *Statistics in Medicine* 2009; **28**:739–761.

- Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine* 2009; **28**:586–604.
- Dmitrienko A, Tamhane AC. Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. *Statistics in Medicine* 2011; **30**:1473–1488.
- Dmitrienko A, Tamhane AC, Wiens B. General multistage gatekeeping procedures. *Biometrical Journal* 2008; **50**:667–677.
- Dmitrienko A, Kordzakhia G, Tamhane AC. Multistage and mixture gatekeeping procedures in clinical trials. *Journal of Biopharmaceutical Statistics* 2011; **53**:726–747.

- Dmitrienko A, Millen B, Brechenmacher T, Paux G. Development of gatekeeping strategies in confirmatory clinical trials. *Biometrical Journal* 2011; **53**:875–893.
- Kordzakhia G, Dmitrienko A. Superchain procedures in clinical trials with multiple objectives. *Statistics in Medicine* 2012.