Considerations in Planning and Testing for Multiple Endpoints in Clinical Trials

Mohammad Huque

Division of Biometrics IV/Office of Biostatistics/OTS/CDER/FDA

BASS XIII, Savannah, Georgia, November 8, 2006

Disclaimer

- Views expressed here are of the presenter and not necessarily of the FDA

Sources of multiplicity in confirmatory RCTs

- Multiple endpoints ✓
- Multiple comparisons – more than 2 arms
- Interim analysis
- Subgroup analysis
- Selection of covariates in an analysis model
- Mid-way changes in the trial design
- Others

Outline

- Family-wise error rate and its control in ME testing
- Some general considerations when deciding about MEs
- “Clinical decision rule” concept for efficacy and null hypotheses formulation for ME testing
- A general principle for ME testing
- Co-primary endpoints and the issues of power and type I error (how conservative?)

Outline (Cont'd)

- Examples of some special situations – raising concerns
- Secondary endpoints and their analysis
- Planned subgroup analysis
- Multiplicity - analysis of safety endpoints

Focus of this presentation

- Confirmatory randomized controlled clinical trials (CRCTs)
  - Principle of prospective planning adhere to
- Two arm trial, a test treatment versus a control endpoints: y_1, y_2, …, y_K

  H_{0j}: $\delta_j = 0$, H_{aj} $\delta_j \neq 0$, $j = 1, \ldots, K$

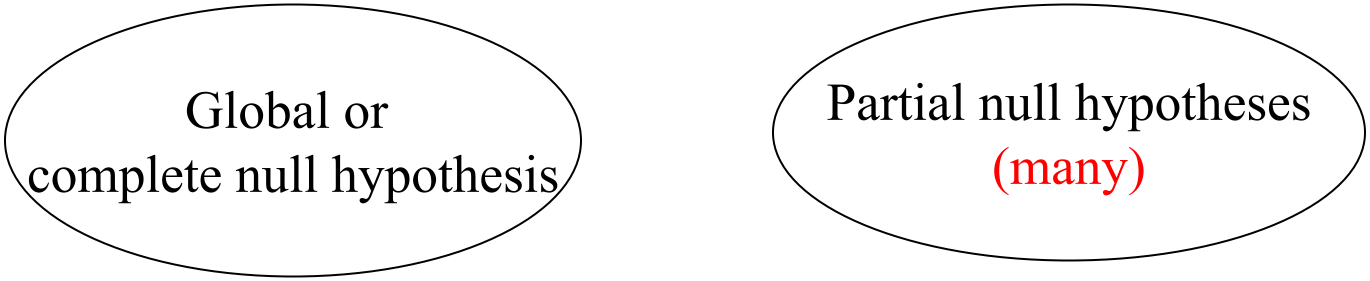
Trial has a single endpoint to test – type I and type II errors

- Conduct a test for claiming that a new treatment is beneficial
- α = Probability of the Type I error
- β = Probability of the Type II error (power = 1- β)

	Concludes Treatment Not beneficial	*Concludes Treatment beneficial*
Truly Not beneficial H_0	Correct Decision	Type I error
Truly beneficial H_a	Type II error	Correct Decision

Familywise type I error rate (FWER)

Family of ME hypotheses: H_{01}, H_{02}, …, H_{0K}. Some may be true and some may be false



Global or complete null hypothesis

Partial null hypotheses (many)

- For a given multiplicity problem there can be many null hypotheses configurations
- Calculate the probability of at least one type error for each null hypothesis configuration
- FWER = Maximum of these probabilities across all null hypotheses configurations

An illustrative example

- A cardiovascular trial: effects of a new therapy on mortality, stroke and MI endpoints

Mortality	stroke	MI	
no effect	no effect	no effect	global null
no effect	no effect	an effect	partial
no effect	an effect	no effect	partial
no effect	an effect	an effect	partial

- 4 null hypotheses configurations for the mortality endpoint
- Total 2^3 -1 = 7 null hypotheses configurations for 3 endpoints
- FWER = maximum of probabilities of at least one type I error across all null hypotheses configurations).

Methods that control FWER only under the global null hypothesis

- Generally, not appropriate for clinical trial applications for efficacy claim
- Can lead to inflated FWER for endpoint specific claims of treatment effect (e.g., in the previous example, mortality benefit for the new therapy)

Examples of Methods (Sankoh et al, DIA Jr.,1999):

Simes test (BMK 1986)

O'Briens OLS/GLS tests (Biometrics 1984)
Hotelling's T^2

Other tests

Two type of FWER control: “weak” and “strong”

- ✓ Weak Control – in reality it does control FWER
  - Control of FWER assured only under the global null hypothesis (the study intervention is not effective in any of the endpoints.) Meant for non-specific claims.
- ✓ Problem
  - The result can be difficult to interpret
  - Type I error rate can remain inflated for endpoint specific claims.

“Strong” control of FWER

- Multiple hypotheses testing leads to multiple null hypotheses configurations
- Strong control: when a method controls FWER for all relevant null hypotheses configurations for a given multiplicity problem

Endpoint specific claims

- It is essential to control FWER “strongly” for endpoint specific claims. Why?
- In seeking a result for a specific endpoint in the presence of other endpoints, there are multiple null hypotheses possibilities.
- Example: CV trial with 3 endpoints, $2^3 - 1 = 7$ null hypotheses scenarios)

Mortality	stroke	MI
no effect	no effect	no effect
no effect	no effect	an effect
no effect	an effect	no effect
no effect	an effect	an effect

Similarly for stroke and MI endpoints

- A specific claim for mortality, stroke or MI benefit requires strong FWER control

Examples of methods appropriate for endpoint specific claims

- Bonferroni, Sidak, Bonferroni-Holms methods
- Hochberg, Hommel methods (with some caveats)
- Hierarchical testing (Wesfall et al, 1999)
- Gatekeeper methods
  - Westfall et al, 1999; Dimetrienko et al, 2003
- Fallback method
  - Wiens, 2003; Wiens and Dmitrienko, 2005
- Closed Testing (Marcus et al, 1976)
- Others

Closed Testing Principle

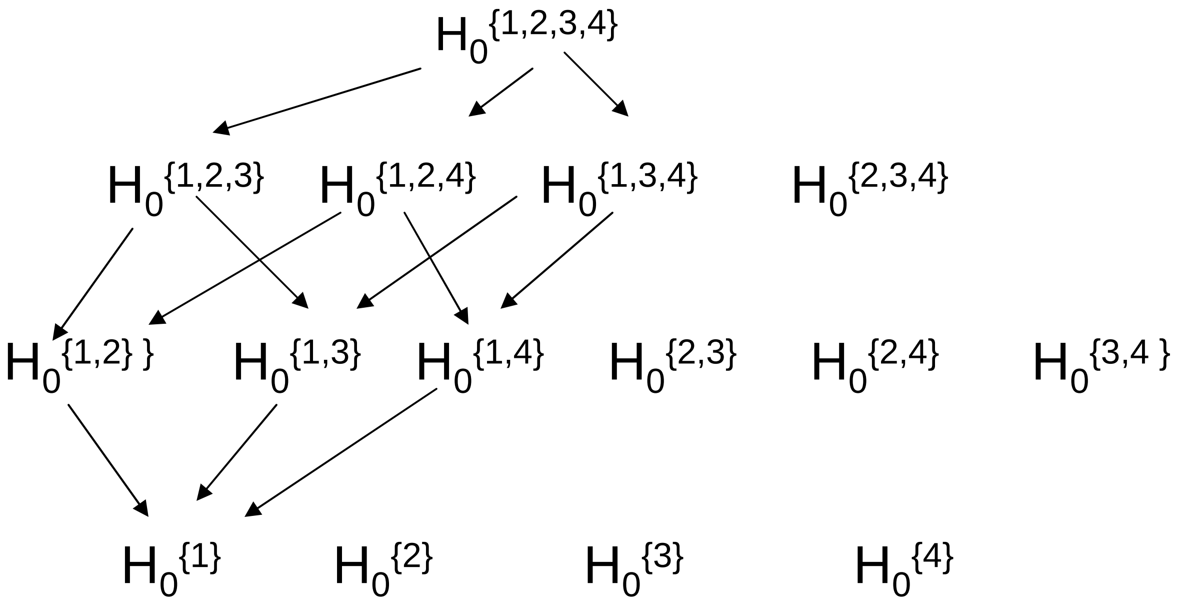
- The family of hypotheses $\boldsymbol{F} = \{ H_0^I \equiv \cap_{k\in I} H_0^k : I \subset (1, 2, \ldots, K) \}$ is closed under intersection. For example, given K= 4,

  $$\boldsymbol{F} = \{ H_0^{\{1,2,3,4\}},$$
  $$H_0^{\{1,2,3\}}, H_0^{\{1,2,4\}}, H_0^{\{1,3,4\}}, H_0^{\{2,3,4\}\ \}},$$
  $$H_0^{\{1,2\}\ \}}, H_0^{\{1,3\}}, H_0^{\{1,4\}}, H_0^{\{2,3\}}, H_0^{\{2,4\}}, H_0^{\{3,4\}},$$
  $$H_0^{\{1\}}, H_0^{\{2\}}, H_0^{\{3\}}, H_0^{\{4\}} \}$$

- In a closed testing procedure all tests are made at the same significance level α, using an appropriate global test statistic (e.g., Hotelling's T^2 test, Simes test). A hypothesis in the family is rejected, if it is rejected and also all the higher dimensional intersection hypotheses containing that hypothesis are also rejected.

Closed Testing Principle

(At Step 1 the test is a Global test)



- Step 1: $H_0^{\{1,2,3,4\}}$
- Step 2: $H_0^{\{1,2,3\}}$ $H_0^{\{1,2,4\}}$ $H_0^{\{1,3,4\}}$ $H_0^{\{2,3,4\}}$
- Step 3: $H_0^{\{1,2\}\ \}}$ $H_0^{\{1,3\}}$ $H_0^{\{1,4\}}$ $H_0^{\{2,3\}}$ $H_0^{\{2,4\}}$ $H_0^{\{3,4\ \}}$
- Step 4: $H_0^{\{1\}}$ $H_0^{\{2\}}$ $H_0^{\{3\}}$ $H_0^{\{4\}}$

Each test is carried out at level α. $H_0^{\{1\}}$ is rejected if it is rejected and all higher dimensional intersection hypotheses containing $H_0^{\{1\}}$are also rejected

Some general considerations about any endpoint

- Prospective definitions and methods of assessments – post hoc definitions can lead to non-solvable multiplicity problem
- Consistency in following the definition and the methods of assessments across all centers of a multi-center trial
- Determination of clinical endpoint values by qualified and trained individuals
- If the endpoint is a lab endpoint – delineation of the assay type and samples

Primary endpoints of a trial

- Addresses the primary objectives of the trial.
- Shows that the study drug has clinically significant beneficial effects
  - A study with null or negative results for the primary endpoints do not support marketing approval of the drug – except perhaps in the most unusual circumstances

Some general questions to ask when deciding about primary endpoints (1)

- Are the endpoints clinically relevant and do they measure meaningful clinical or patient (PRO) benefits?

- Do the endpoints have regulatory and scientific merit?

- What is the history of their use in similar studies of approved products?

- What is the recommendation of their use in regulatory guidance documents and by the AC experts?

Some general questions to ask when deciding about primary endpoints (2)

- Are there better endpoints that measure clinical benefits?
  - Less variability, less misclassification error, etc
- Why use certain endpoints when other more acceptable viable alternatives exist?
- Are there practical and ethical issues and how to resolve them?
- Are there some secondary endpoints included in the trial that can complement the results of the primary endpoints?
- Will a composite or responder endpoint be more appropriate?

“Clinical decision rule” or “win” criterion concept for efficacy

- From a regulatory perspective, a clinical decision rule is simply a prospectively defined rule that describes how a positive decision regarding the benefit of a test treatment is going to be reached, i.e., what clinicians usually refer to as, “what defines a win”
- A win criterion usually involves results on multiple primary endpoints on one or more doses of the test drug and a control.

Some Examples of “Clinical Win” criterion CHF trial w. 3 primary endpoints

“win on All-Cause-Mortality” or “win in MI” or “win in Stroke” and specify which endpoints have the effect

- Null hypotheses: $H_{01} \cap H_{02} \cap H_{03}$ (intersection null hypothesis) and partial null hypotheses

- Testing method: use a method that controls the FWER in the strong sense (e.g., method based on closed testing principle) if one seeks endpoint specific claims

Alzheimer trial
(Case of 2 co-primary endpoints)

“ win on ADAS-Cognitive Sub-scale” <u>and</u>

“win on Clinician’s Interview Based Impression of Change”

- Null hypotheses: H_{01} U H_{02} (union null hypothesis)
- Alternative hypothesis: intersection null hypothesis
- Testing method: test $H_{01}$and H_{02} each at the same significance level of α (e.g., $\alpha = 0.05$). No need for multiplicity adjustment. Why? Is the test conservative?
- Impacts the type II error – more co-primary endpoints less power

Epilepsy trial
Example w. 3 primary endpoints

“win on Seizure rate” or
“win on Drop Attack Rate and win on Seizure Severity”

and specify which endpoints have the effect

- Null hypothesis is a complex null hypothesis: $H_{01} \cap (H_{02} \text{ U } H_{03})$ and partial nulls
- Testing: Allocate $\alpha_1 = 0.025$ for testing H_{01}, and $\alpha_2 = 0.025$ for $(H_{02} \text{ U } H_{03})$
- Test H_{02} and H_{03} each at $\alpha_2 = 0.025$

Case of pairwise co-primary endpoints

“win for E1 and E2” or “win for E1 and E3”

- E1 is most relevant endpoint but not sufficient by itself for claim of efficacy
- The decision rule is equivalent to: “win for E1” and (“win for E2 or E3”)
- Hierarchical Testing: Test H_{01} first at $\alpha = 0.05$. If H_{01} is rejected then test for the family $\{H_{02}, H_{03}\}$ using a closed testing procedure, FWER = 0.05

Acne trial example

“Clinical Win” criterion

- 4 primary endpoints:

  Y_0 = physician global

  Y_1= non-inflammatory lesion counts

  Y_2 = inflammatory lesion counts

  Y3 = total lesion counts ($Y_1 + Y_2$)

- Clinical Decision rule:
  - Statistical significance for Y_0
  - In addition, statistical significance in at least 2 of the 3 remaining endpoints
- Possible Rationale: Y_1 and Y_2 lie on different causal pathways, and Y_0 intersects with both.

Hypertension trial example: Case of 2 endpoints and 3 doses

2 endpoints: SBP and DPB (SBP more important);
high doses: D1, D2; low dose D3

Hierarchical Clinical Decision Rule:

(i) Benefit for at least one of the high doses for SBP

(ii) Benefit for at least one of the high doses for DBP

(iii) Benefit for the low dose for the SBP

(iv) Benefit for the low dose for DBP

Statistical methodology: gatekeeper, fallback

Dmitrienko et al, 2003; Wiens, 2003; Wiens & Dmitrienko, 2005; others

Two co-primary endpoints w. relaxed evidence criterion

- Show persuasive evidence of efficacy in at least one of the two endpoints and at least a trend in the other
- Possible “win” criterion:

  $P_1 < \alpha_1 = 0.05$ and $P_2 < \alpha_2 = 0.065$ (2-sided p-values)

  OR

  $P_1 < \alpha_1 = 0.065$ and $P_2 < \alpha_2 = 0.05$

- FWER = 0.05 (strong sense) for endpoints with p-value of < 0.05.

✓ Clinical acceptance of this criterion at present not clear

NOTE: The above value of 0.065 (here arbitrary) to be fixed in advance

General principle for solving multiple endpoint problems

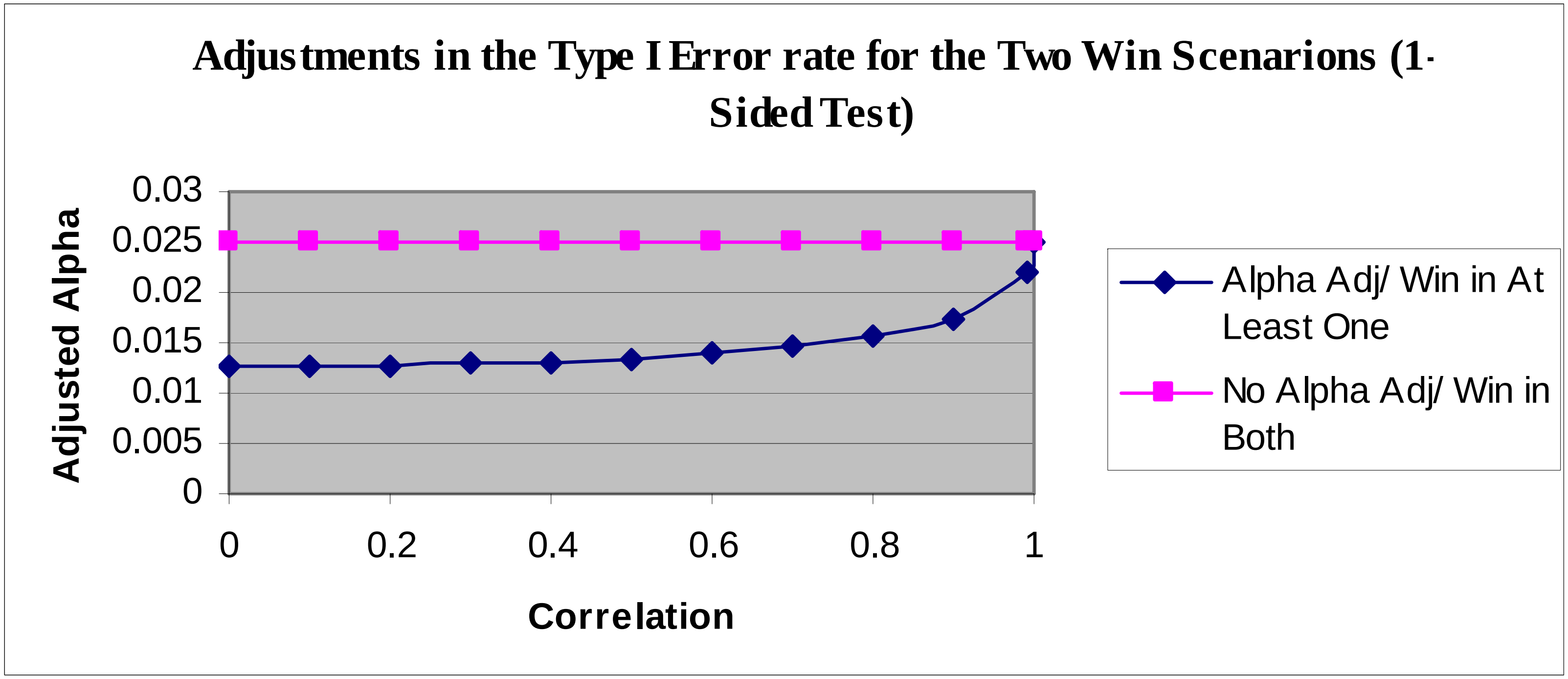
- Specify primary endpoints based on clinical and regulatory considerations
- Specify clinical benefit criterion involving these multiple endpoints?
  - "Alternative Hypothesis" - Win or benefit situation

    ("Null Hypothesis": no clinical benefit scenario(s))
- Select an optimal statistical test strategy for establishing clinical benefit that controls: (1) the familywise type I error rate and (2) has adequate power of the test

Co-primary endpoints and the issues of power

- Case of 2 co-primary endpoints
- Cases of 3 or more co-primary endpoints

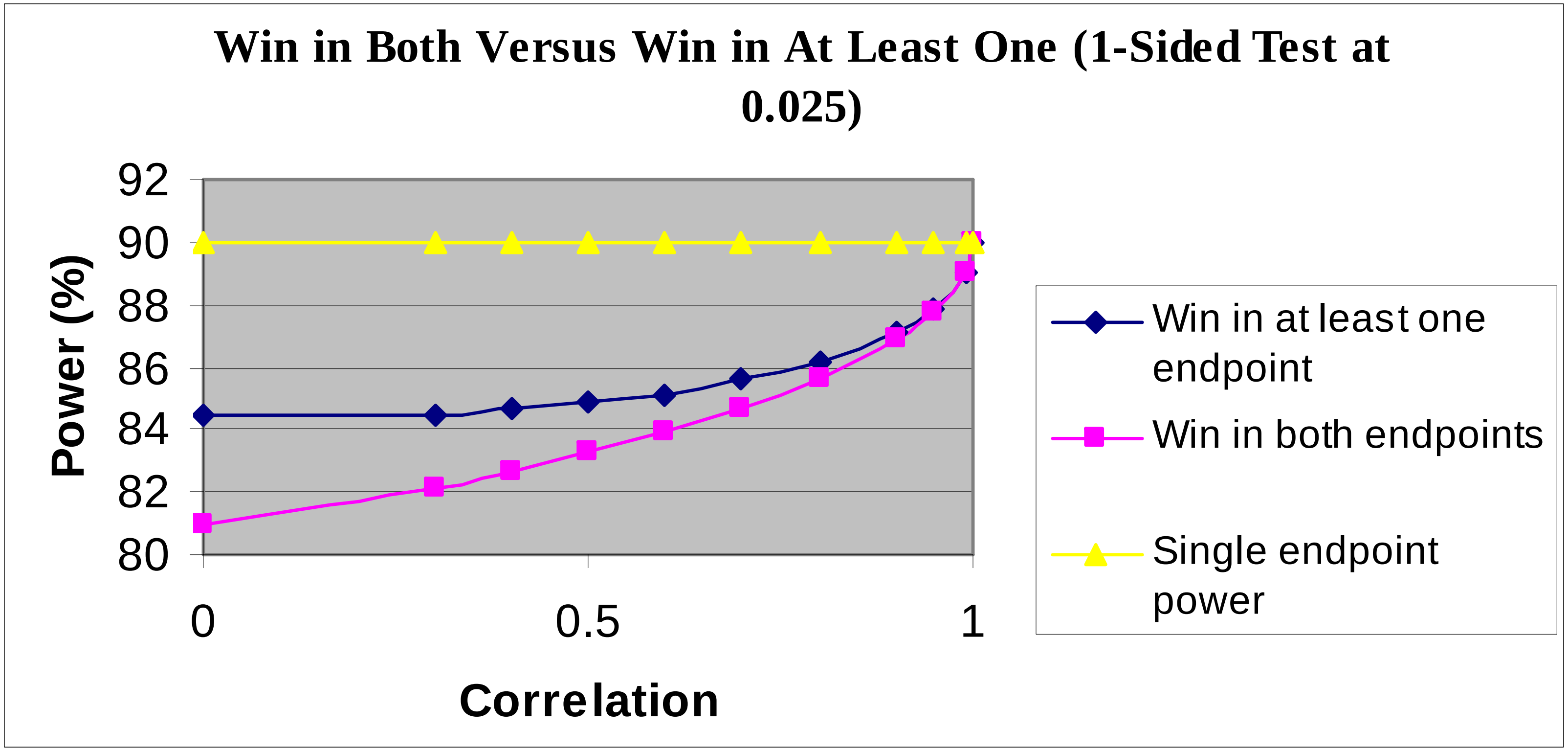
Adjustments in the Type I error rate Case of 2 Endpoints



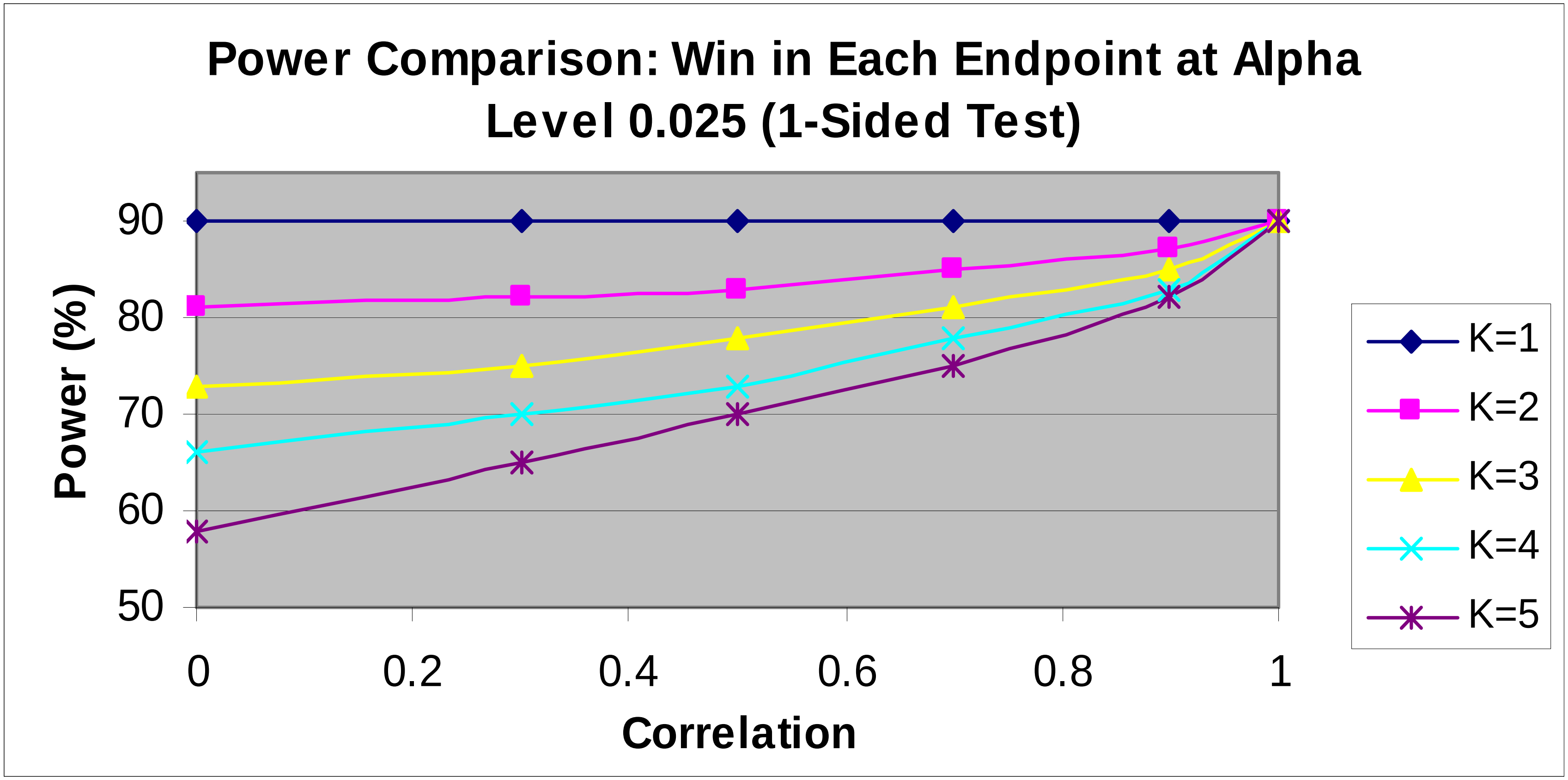
- Ad justment by Sidak's method on accounting for correlation

Power Comparison

Case of K=2 endpoints:



Loss in Power when win in all endpoints, K= # of endpoints



Sample Size Increase: [(1)] When Win in All K Endpoints Compared to Single Endpoint Case

Alpha = 0.025 (1-sided), Power = 0.90

Correlation	K = 2	K=3	K=4
0.0	22.8%	35.9%	45.0%
0.3	21.1	33.1	41.2
0.4	20.2	31.7	39.7
0.5	19.1	29.8	37.3
0.6	17.7	27.5	34.4
0.7	15.9	24.6	30.7
0.8	13.5	20.8	25.8
0.9	10.0	15.3	18.9

[(1)] Calculations using mutivariate normal distribution of the test statistics comparing active treatment versus placebo for a 2-arm trial, assuming same delta/sigma for all K endpoints

Why co-primary endpoints?

- Scientific basis driven by clinical and biological considerations
  - Ophthalmic: Suppose that a benefit of a treatment achieved at a desired time point. Is this benefit sustained at subsequent time points?
  - Migraine: pain-free at 2-hours. How about relief at 2-hours from nausea, photosensitivity and phonosensitivity?
  - Other examples

Considerations for reducing the burden of multiplicity in ME testing

1) Triaging of MEs into Primary, Secondary and exploratory endpoints

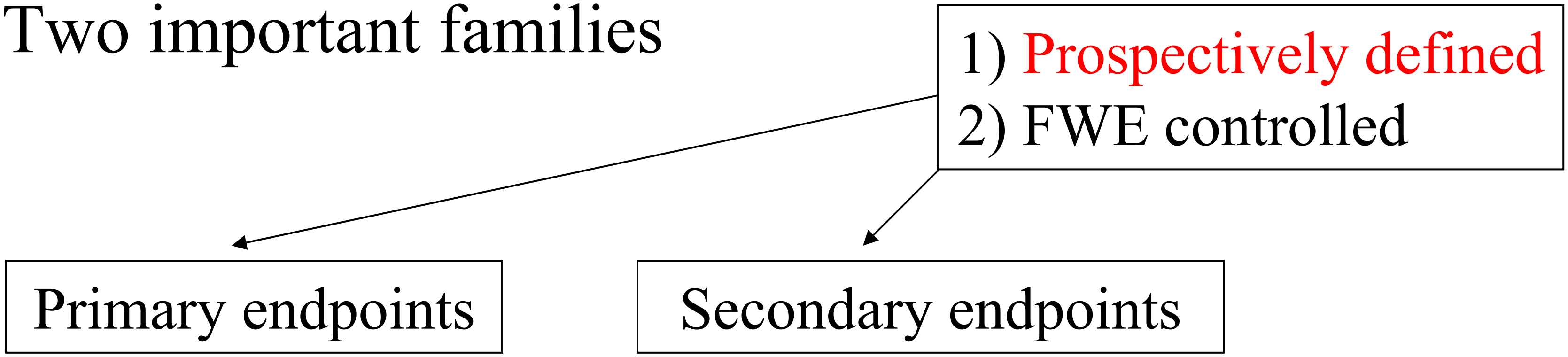
2) Hierarchical ranking of families of endpoints and endpoints within a family

3) Use of dependence/correlation measures in testing of MEs

4) Reducing multiple endpoints to a single composite or a responder endpoint

1) Triaging of multiple endpoints into meaningful families by trial objectives

- Two important families



1) Prospectively defined
2) FWE controlled

Primary endpoints

Secondary endpoints

Exploratory endpoints (usually not prospectively defined)

►Primary endpoints are primary focus of the trial. Their results determine main benefits of he clinical trial's intervention.

►Secondary endpoints by themselves generally not sufficient for characterizing treatment benefit. Generally, tested for statistical significance for extended indication and labeling after the primary objectives of the trial are met.

2) Hierarchical ranking of MEs

Example (CHF- trial):

1. E1 = composite of CHF-related mortality and hospitalizations
2. E2 = CHF-related mortality

(How much of E1 is contributed by E2?)

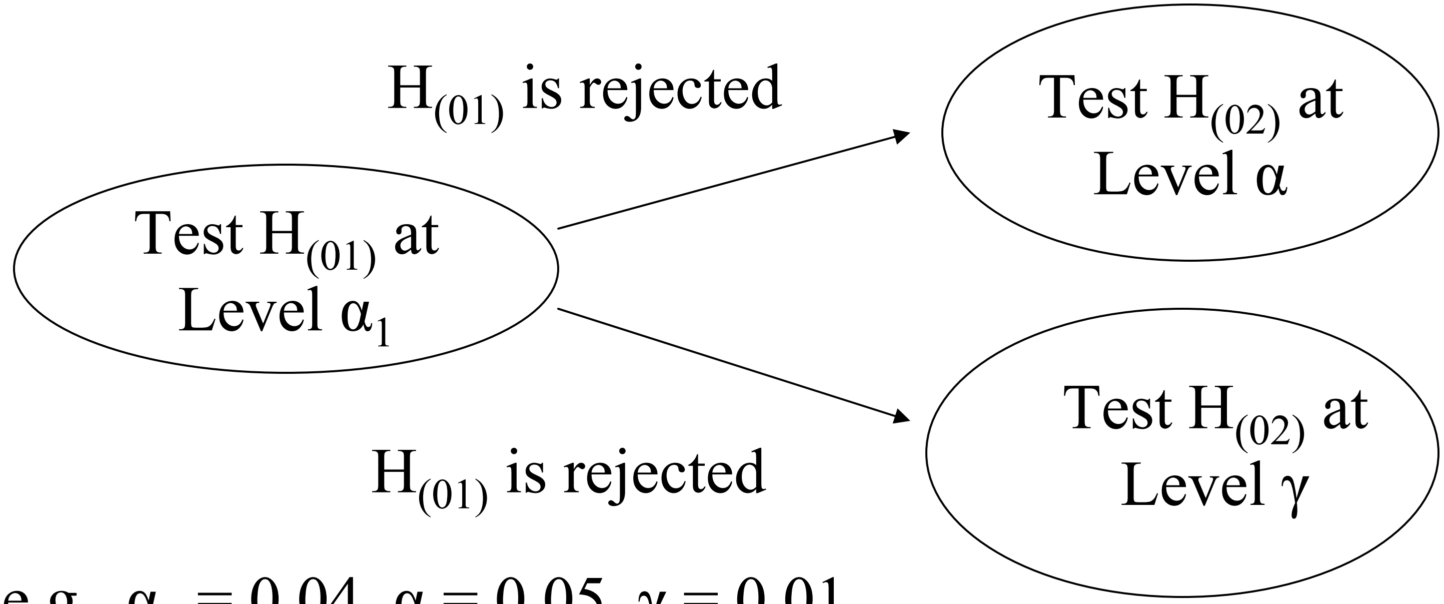
Fixed- sequence testing method:

Test E1 first at $\alpha = 0.05$.

If significant then test E2 also at $\alpha = 0.05$, else, stop testing for E2

Fallback method – a flexible fixed-sequence approach

Fallback method
(A flexible fixed-sequence approach)



Test $H_{(01)}$ at Level α_1

$H_{(01)}$ is rejected

Test $H_{(02)}$ at Level α

$H_{(01)}$ is rejected

Test $H_{(02)}$ at Level γ

e.g., $\alpha_1 = 0.04$, $\alpha = 0.05$, $\gamma = 0.01$

Other γ can be found using correlation between the test statistics

Fallback testing method: Wiens, 2003; Dmitrienko & Wiens, 2005

Flexible fixed-sequence approach: Huque & Alosh, 2005

3) Use of dependence/correlation measures in testing of MEs

- Dependence: Two response variables are said to be statistically "independent" if the response (or the treatment induced effect) of one occur independently of the other, otherwise, they are said to be "dependent"
- Correlation: Two response variables are said to be correlated if an increase or decrease in one variable is associated with an increase or decrease in the other variable in <u>a linear direction</u>
- A measure of correlation between 2 continuous response variables is the correlation coefficient ρ, where $-1 \leq \rho \leq +1$)

Calculus of FWER for multiplicity adjustments

- This calculus relies on specifying the joint distribution of the test statistics of the MEs (e.g., multivariate Z, multivariate t)
- The joint distribution contains treatment effect parameters of individual endpoints and parameters that define correlation structure among the endpoints (or between the test statistics). This information can be external or internal
- Such a joint distribution, if found and if justified, can lead to a less conservative adjustments for testing of multiple endpoints

Concerns for special situations: Analysis of the same endpoint on varying the patient datasets

- Same endpoint but different analysis data sets:

  (i) Intention-to-treat data set,

  (ii) Modified intention-to-treat data set,

  (iii) Evaluable patient data set

- No statistical penalty as long as primary analysis data set pre-defined and additional analyses are supportive or co-primary

Concerns for special situations: Analysis by alternate statistical methods

- Example: Consider three ANOVA analyses for the same endpoint with models:

Model	Treatment Effect
treatment	$p = 0.10$
center, treatment	$p = 0.06$
region, center, treatment	$p = 0.03$

- Given that the above three analysis of variance models were pre-specified, does the third p-value require any adjustment for multiplicity?

Concerns for special situations: Analysis by alternate statistical methods (cont'd)

- Answer to the previous question is Yes. Why?
- However, because of hyper-correlation between the 3 test statistics the adjustment is likely to be mild. Use of simulation and resampling techniques can assess the extent of this correlation.

Primary versus secondary endpoint roles for confirmatory trials

- Primary endpoints (PEs) relate mainly to primary objectives of the trial.
- E-9 document:
  - PE is a reliable and validated variable with which experience has been gained either in earlier studies or in published literature
  - PE provides a valid and reliable measure of some clinically relevant and important treatment benefit in the patient population described by the inclusion and exclusion criteria.
- Secondary endpoints can not be validly analyzed if the primary endpoint does not demonstrate clear statistical significance (O’Neill, 1997) – exception?
- Secondary endpoints (SEs) has a number of additional functions (D’Agostino, 2000):

Some functions of the secondary endpoints for confirmatory trials

1) SEs supply background and understanding of the primary endpoint result (e.g., efficacy at prior time points in a longitudinal trial)

2) SEs together with PEs can address to broader treatment efficacy and add coherence to the results (e.g., osteoarthritis trials, in addition to pain and physical function endpoints, role of PRO and quality of life endpoints)

3) SE can be a component of a primary composite endpoint

Some functions of the secondary endpoints for confirmatory trials (cont'd)

4) SE can be an important endpoint, e.g., mortality, but because of the expected small size of the treatment effect it is kept as a SE

5) SE can be a safety endpoint, e.g., events of major and minor bleeds in blood coagulation trials

6) SE can address to mechanisms of action of the treatment

7) SE can address to a sub-hypothesis of interest, e.g., treatment cures an infection but increases BP

Analysis of secondary endpoints

- SEs, in general, can not be validly analyzed for confirmatory evidence unless the primary objectives have been successfully met.
- However, if the SE is like a mortality endpoint of importance, and is considered secondary only because of the effect size concern, then it can be validly analyzed. Actually, it is like a PE.
- For example, consider a primary endpoint and the all-cause mortality as a secondary endpoint. Allocate $\alpha_1 = 0.04$ for the primary endpoint, and $\alpha_2 = 0.01$ for the mortality endpoint
- If the primary endpoint is significant at $\alpha_1 = 0.04$, then test the mortality endpoint at the full significance level of 0.05, else test it at $\alpha_2 = 0.01$ (Fallback method; Wiens, 2003)

Analysis of secondary endpoints (cont'd)

- An approach for analyzing secondary endpoints is the hierarchical approach
- Gatekeeper methodology
  - Sequential: If all primary endpoint null hypotheses are rejected at level α (e.g., α = 0.05) then secondary endpoints can be tested on controlling FWER at level α (Westfall & Krishen, 2001).
  - Parallel: If the trial's main objective is to win in at least one of the primary endpoints, then the method allows testing for secondary endpoints. FWER controlled for both the primary and secondary endpoints at level α (Dmitrienko et al, 2003; Chen et al, 2005).

Analysis of secondary endpoints
Some key considerations

1. SEs like the PEs should be prospectively planned for managing multiplicity of tests
2. SE tests should control FWER in the “strong sense” for endpoint specific claims of benefit
3. SE tests together with the PE tests should control FWER uniformly at the same α level – unless a special reason for this (e.g., SE is a safety endpoint).
4. If SE results are direct consequence of PE test – no additional claims of benefit
5. If SEs are analyzed only for supportive evidence w/o any intention of claims of additional benefit, then it should be so clarified in the protocol.

Subgroup analyses

- Post-hoc subgroup analyses
  - using forest plots etc.

    (for data display, crude visual checking of consistency of treatment effect)
- Prospectively planned subgroup analyses
  - With proper design consideration (e.g., appropriate randomization, sample size, minimization of bias, etc)
  - Example: A trial with two subgroups (a biomarker +ve and -ve subgroups) and 2 treatments. It is a 2x2 factorial experiment. Suppose that there is a possibility of different disease progression or manifestation in the two subgroups. What will be the appropriate randomization for this design?

Post-hoc subgroup analyses

- Results usually seriously flawed
  - Non-resolvable multiplicity issues
  - Bias due to various confounding and other factors
  - Can easily produce spurious results
- Results - hypotheses generating
  - If clinically and biologically plausible
  - If adjusted for bias through appropriate statistical modeling

Prospectively planned subgroup analyses - With proper design consideration (1)

Trial Design similar to factorial experiments

1. Clear prospectively defined hypotheses to be tested
2. Randomization and sample size considerations
3. Specification of design variables, e.g.,
   - Center could be a random effect if many and selected from a population of similar centers.
   - Region could be a fixed effect if result sought for U.S. versus non U.S. populations
4. Proper proportional enrichment of patient population if result sought by certain validated biomarker result (+ve versus -ve). Misclassification error?

Planned subgroup analyses - with proper design consideration (2)

5. Method of analysis - prospective
6. Multiplicity adjustment method that assures strong FWER control - prospectively planned
7. Analysis models (e.g. mixed-effect) that give unbiased estimates of treatment effects with right behavior of the residual effects for unbiased statistical testing
8. Confounding issues when testing for main/simple effects

– Other considerations

Analysis of Clinical Safety Data ICH E9 - document

- Safety evaluation of an intervention is a multidimensional problem involving many endpoints
- P-values: appropriate as a flagging device applied to a large number of safety variables to highlight differences that are worthy of further investigation
- Multiplicity adjustments for quantify type I error is appropriate. However, type II error is of major concern for missing a true safety signal.

Analysis of Clinical Safety Data
Three Tier Model (Mehrotra and Heyes, 2004)

- ***Tier 1 AEs***: Adverse experiences with specific hypotheses that are formally tested in the clinical study and both type I and type II errors are formally addressed.

- ***Tier 2 AEs***: Common adverse experiences encountered as a part of the overall patient reporting in the trial. One compares treatment versus control cumulative incidence rates (exposure adjusted) for each AE type

Body System	AE Types	Test Treat	Control	P-Value
Cardiovascular	AE1 AE2	incidence	incidence	p

- ***Tier 3 AEs***: Adverse experiences that are spontaneous AE events, often serious, requiring evaluation by specialty experts. (Generally, no statistical testing for these cases.)

Handling multiplicity for Tier 2 AEs

- Tier 2 AEs in a trial can be many (e.g 40 or more). Need a method that
  - Provides a proper balance between "no adjustment" and "too much adjustment" (e.g., the expected ratio of false rejections to the total number of rejections is controlled)
  - Provides adjusted p-values for flagging purpose
  - Addresses Type II error concerns
- Single FDR method or the DFDR method can be used on taking into consideration the grouping of AEs by body system (Mehrotra and Heyes, 2004)
- Bayesian method: Scott Berry and Donald Berry, 2004)

Type II error concern:
Analysis of Tier 2 AEs

- For a specific application, there is a need to evaluate the operating characteristics of the FDR (α) and DFDR (α_1, α_2) methods - in selecting alphas for these methods

- For adequate type II error control, alphas may not be small unless the trial size is very large.

- For example, if choose $\alpha = 0.05$ when applying the FDR (α) method by each body system, or if choose α_1 and $\alpha_2 = 0.10$ when applying the DFDR (α_1, α_2) method, make sure that type II error is adequately controlled

Concluding Remarks

- Effectiveness of an intervention in CRCTs are usually assessed through a “win” criteria that involve testing of clinically relevant multiple endpoints. This, except for some special cases, causes inflation of the FWER.
- There are 2 types FWER control - weak and strong. CRCTs almost always require FWER control in the strong sense for specific claims of efficacy of an intervention
- In CRCTs, “win” criteria for efficacy of the intervention and the statistical methodology for FWER control are prospectively planned
- Triaging of multiple endpoints into PEs and SEs helps in managing multiplicity

Concluding Remarks (Cont'd)

- FWER for SEs are generally controlled in the strong sense for meaningful labeling of the product - methods, such as, gatekeeper and fallback methods can be applied
- Analyses using different data sets and models raise multiplicity concerns. No concern - if there is a primary analysis and other analyses supportive or co-primary
- Post-hoc subgroup analysis can produce seriously flawed results.
- Planned subgroup analysis can be Ok with proper design considerations and multiplicity adjustments
- Common adverse events: for a flagging method – greater focus on the type II error control than the type I error control.