
Costs & Benefits of Multiplicity Adjustments for *Correlated* Endpoints

Andrew C. Leon, Ph.D.

Weill Medical College of Cornell University

Funded, in part, by NIH Grant MH60447

Multiplicity Adjustments

- **Rationale**
- Adjustment Strategies
- Application
- Simulation Study

Multiple Tests of Cognitive Assessments Elevate Risk of Type I Error

- **Multiple primary endpoints:**
 - Multiple domains of cognition and/or functioning
 - Multiple cognitive domains and positive symptoms

FDA / ICH Guidance for Industry (E9):
Statistical Principles for Clinical Trials

“It may sometimes be desirable to use more than one primary variable

... the method of controlling type I error should be given in the protocol.”

http://www.fda.gov/cder/guidance/ICH_E9-fnl.PDF

CONSORT Statement

(Consolidated Standards of Reporting Trials)

- Multiple analyses of the same data create a considerable risk for false-positive findings.
- Analyses that were pre-specified in the trial protocol are much more reliable than those suggested by the data.
- Authors should indicate which analyses were pre-specified.

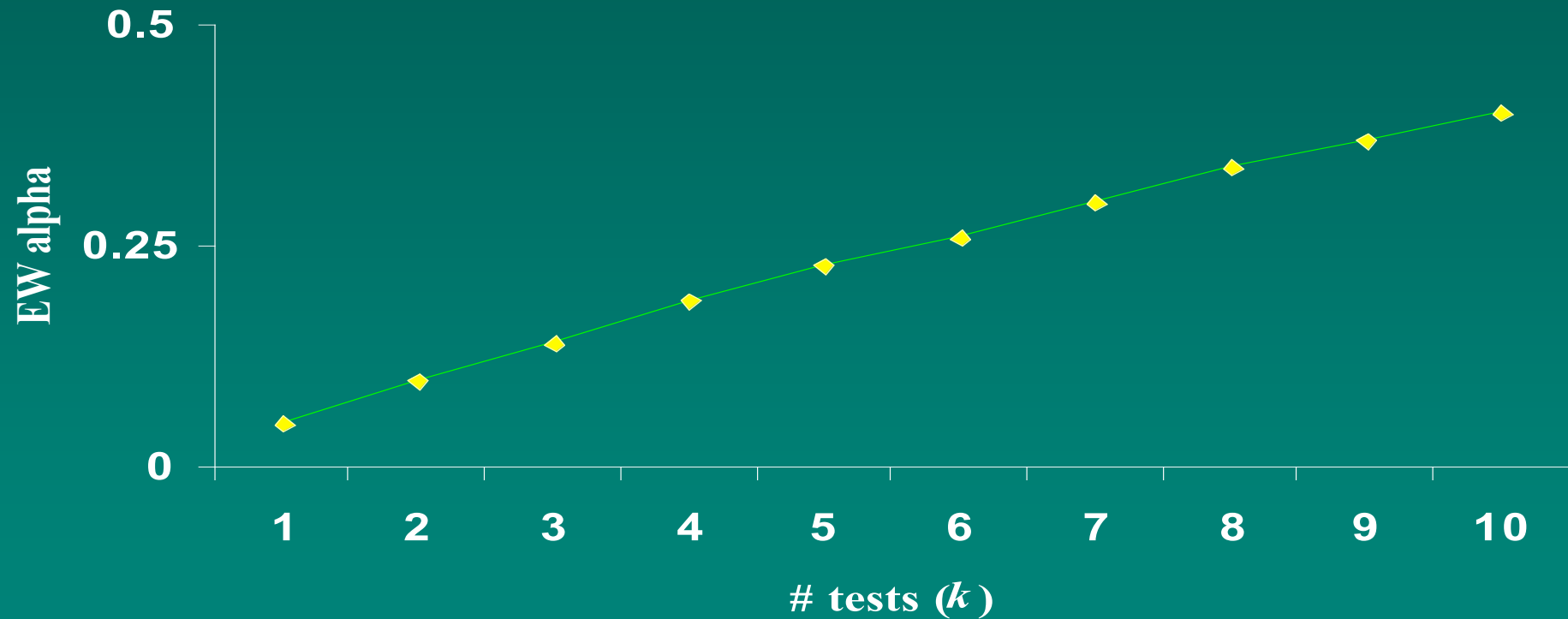
www.Consort-statement.org

Randomized Clinical Trial Design

Tension between

- Falsely concluding that an ineffective agent is efficacious
 - *Type I error*
- Failing to conclude that an effective agent works
 - *Type II error*

Multiple Comparisons and *Experimentwise Type I Error*

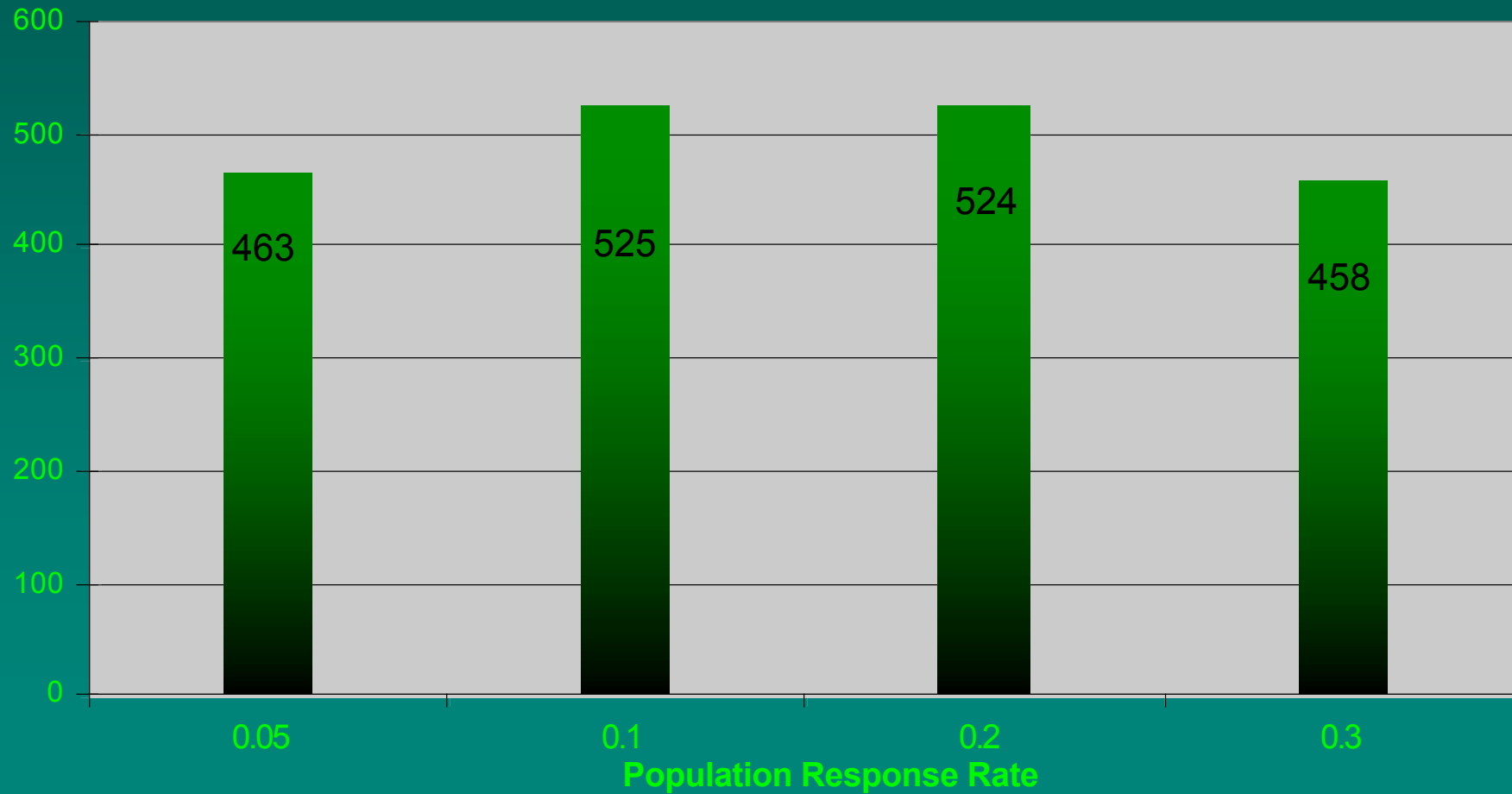


Assumes: $\alpha=0.05$ per test

$$\alpha_{EW} = 1 - (1 - \alpha)^k$$

Simulation Study: Type I Error

Significant χ^2 tests



10,000 data sets with N=100/group per response rate

10,000 χ^2 tests/rate

Multiplicity Adjustments

- **Rationale**
- Adjustment Strategies
- Application
- Simulation Study

Bonferroni Adjustment

- Partitions the $\alpha = 0.05$ among the tests

$$\alpha^* = \alpha / k, \text{ for } k \text{ tests}$$

- Controls for Type I error among all tests
 - Sets an upper limit on *Experimentwise* Type I error (α_{EW}) among multiple tests

Bonferroni Adjusted Alpha-levels for k tests and $\alpha_{EW} = 0.05$

# tests	α^*
1	0.0500
2	0.0250
3	0.0167
4	0.0125
5	0.0100

Experimentwise Type I Error (α_{EW})

With 5 dependent variables:

$$k=5$$

$$\alpha_{EW} = 1 - (1 - \alpha)^k$$

$$\alpha_{EW} = 1 - (1 - 0.05)^5 = 0.23$$

If H_0 is true: 23% chance of false positive result

Bonferroni Correction Reduces α_{EW}

With 5 dependent variables

and an adjusted alpha: $\alpha^* = 0.05/5 = 0.01$

$$\alpha_{EW} = 1 - (1 - \alpha)^k$$

$$\alpha_{EW} = 1 - (1 - 0.01)^5 \cong 0.05$$

Appeal of the Bonferroni Adjustment

- Simplicity of calculations:

$$\alpha^* = \alpha/k, \text{ for } k \text{ tests}$$

- One simple adjustment applies to numerous statistical procedures (and a combination of procedures):

Tests of continuous, categorical, and survival data

- The probability of Type I error is tightly controlled

Multiplicity Adjustments

- **Rationale**
- Adjustment Strategies
- Application
- Simulation Study

Application: Gender Differences in Circumstances of NYC Homicides in 1997

- **5 Primary Outcome Measures:**
 - place of injury (victim's home vs. elsewhere)
 - firearm death (yes/no)
 - 3 variables determined by toxicologic analyses:
 - presence of ethanol (y/n)
 - presence of cocaine (y/n)
 - presence of opiates (y/n)
- A χ^2 test was used to test the respective $H_{0j} : \pi_{1j} = \pi_{2j}$.

*PI: K. Tardiff, R01 DA06534

Gender Differences in Circumstances of NYC Homicides in 1997

Variable	Females (N=45)	Males (N=386)	χ^2	df	Unadjusted <i>p</i> -value
	% positive				
Victim's Home	73.3%	28.2%	37.10	1	<.0001*
Firearms	51.1%	75.9%	12.67	1	0.0004*
Opiates	6.7%	5.4%	0.12	1	0.7343
Ethanol	20.0%	37.0%	5.13	1	0.0235*
Cocaine	4.4%	20.2%	6.62	1	0.0101*

Gender Differences in Circumstances of NYC Homicides in 1997

Variable	χ^2	Unadjusted <i>p</i> -value	Unadjusted α	Bonferroni α
Victim's Home	37.10	<.0001	.05*	.01*
Firearms	12.67	0.0004	.05*	.01*
Opiates	0.12	0.7343	.05	.01
Ethanol	5.13	0.0235	.05*	.01
Cocaine	6.62	0.0101	.05*	.01

Problem with Bonferroni Adjustment

- **Does not account for correlations between dependent variables.**
- Reduced statistical power – can lead to false negative findings (with a fixed N).
- Maintain statistical power if *sample size* estimates are based on adjusted alpha level and specified in the protocol.

Sample Size Requirements*

Increase with the Number of Tests

# tests	adjusted ?		ES=0.50
1	0.050		64
2	0.025		78
3	0.0167		86
4	0.0125		91
5	0.010		96

Must increase N by about 20% for 2 tests; 30% for 3 tests.

*Assume: 2-tailed t-test, power=0.80

(Leon, JCP, 2004)

Alternative *alpha*-adjustment strategies

Sequentially-rejective tests

Alpha threshold changes with each successive test.

1) **Holm's** step-down tests

- Holm, *Scandinavian Journal of Statistics*, 1979

2) **Hochberg's** step-up tests

- Hochberg, *Biometrika*, 1988

Hochberg's Sequentially-Rejective Tests

1. Variables are tested in descending order of p -values.
2. Each successively *smaller* p -value has a *more rigorous* alpha threshold.
3. Stop testing after first significant test.
4. All subsequent variables are deemed *significant*.

Hochberg's sequentially-rejective tests

for 5 tests and $\alpha = 0.05$

Each successively smaller *p*-value
has a more rigorous alpha threshold

<i>test #</i>	<i>Hochberg</i>
1	0.0500
2	0.0250
3	0.0167
4	0.0125
5	0.0100

Hochberg's sequentially-rejective tests

Variable	χ^2	Unadjusted <i>p</i>-value	Hochberg α	<i>Decision</i>
Opiates	0.12	0.7343	.05	Do not reject
Ethanol	5.13	0.0235	.0250*	Reject Ho
Cocaine	6.62	0.0101	*	Reject Ho
Firearms	12.67	0.0004	*	Reject Ho
Victim's Home	37.10	<.0001	*	Reject Ho

James p -value Adjustment

- James adjustment for normally distributed endpoints
(*Statistics in Medicine*, 1991, 10, 1123-1135)
- Adjustment procedure incorporates correlations among endpoints
- We have adapted it for correlated binary endpoints and evaluated the approach. (Leon and Heo, *J Biopharm Stat*, 2005)

James p -value Adjustment

Here assume equal pairwise correlation, r , among k variables.

Adjusted p -value for 2-tailed tests is estimated as:

$$p_{\text{adj}} = 1 - D_1(1-r^2) - D_2 r^2 - D_3 r(1-r) - D_4(2 - 2(1-r)^{1/2} - r - r^2)$$

Where:

$$D_1 = (1-p)^k$$

$$D_2 = (1-p)$$

$$D_3 = 0 \quad (\text{by definition for 2-sided case})$$

$$D_4 = k(k-1)\phi(b) \int \Phi(z)^{k-2} \phi(z)^2 dz$$

$$= k(k-1)\phi(b) G(k)$$

p =observed p -value and r = mean correlation

James p -value Adjustment

Cocaine: $p=0.0101$

Observed $r=.1105$

$$\begin{aligned} p_{\text{adj}} &= 1 - D_1(1-r^2) - D_2 r^2 - D_3 r(1-r) - D_4(2 - 2(1-r)^{1/2} - r - r^2) \\ &= \mathbf{.049} \end{aligned}$$

Where: $D_1 = (1-p)^k = (1-.0101)^5 = .9505$

$D_2 = (1-p) = 1-.0101 = .9899$

$D_3 = 0$ (by definition for 2-sided case)

$D_4 = k(k-1)\phi(b) G(5) = 0.01697$

$G(k=5) = .05814$ (James, 1992, Appendix I)

James p -value Adjustment

$$\begin{aligned} D_4 &= k(k-1)\phi(b) \int \Phi(z)^{k-2} \phi(z)^2 dz \\ &= k(k-1)\phi(b) G(k) \end{aligned}$$

$$b = \Phi^{-1}(1-p/2) = \Phi^{-1}(1-.0101/2) = 2.572$$

$$\phi(b) = .0145$$

$$G(k=5) = .05814 \quad (\text{James, 1991, Appendix I})$$

$b = \Phi^{-1}(1-\rho/2)$, ϕ is probability density function of a standard normal variable

Φ the cumulative probability function; Φ^{-1} is the inverse function of Φ

James *p*-value adjustment

Variable	χ^2	Unadjusted <i>p</i>-value	James <i>p</i>	<i>Decision</i>
Victim's Home	37.10	<.0001	<0.001*	Reject Ho
Firearms	12.67	0.0004	0.002*	Reject Ho
Cocaine	6.62	0.0101	0.049*	Reject Ho
Ethanol	5.13	0.0235	0.111	Do not reject
Opiates	0.12	0.7343	0.999	Do not reject

Comparison of Alpha-Adjusted Results

Variable	Unadj. <i>p</i> -value	Bonferroni	Hochberg	James
Victim's Home	<.0001	*	*	*
Firearms	0.0004	*	*	*
Cocaine	0.0101		*	*
Ethanol	0.0235		*	
Opiates	0.7343			

Specify adjustment procedure before the study.

RCT for Subjects with Schizophrenia: Comparison of 2 Active Doses

	Unadj. <i>p</i>-value	Bonferroni _	Hochberg _	James <i>p</i>-value
SST Basal	.0228	.0125	.025*	.0874
Flanker RT Neutral	.0221	.0125	*	.0848
STDT Experiment Time	.0178	.0125	*	.0687
Flanker Correct Incongruent	.0129	.0125	*	.0502

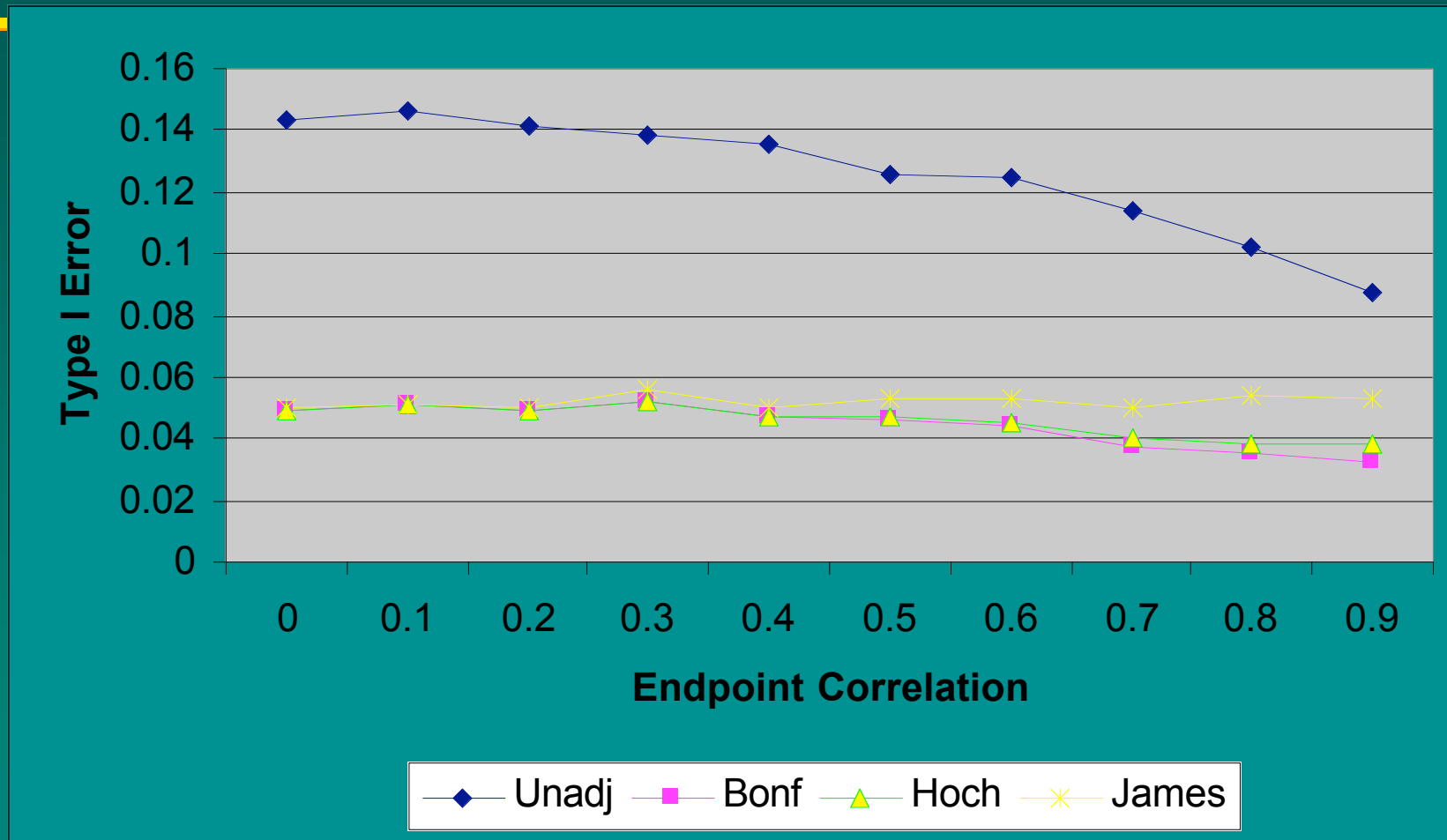
Mean $r=.128$;

Specify adjustment before the study.

Multiplicity Adjustments

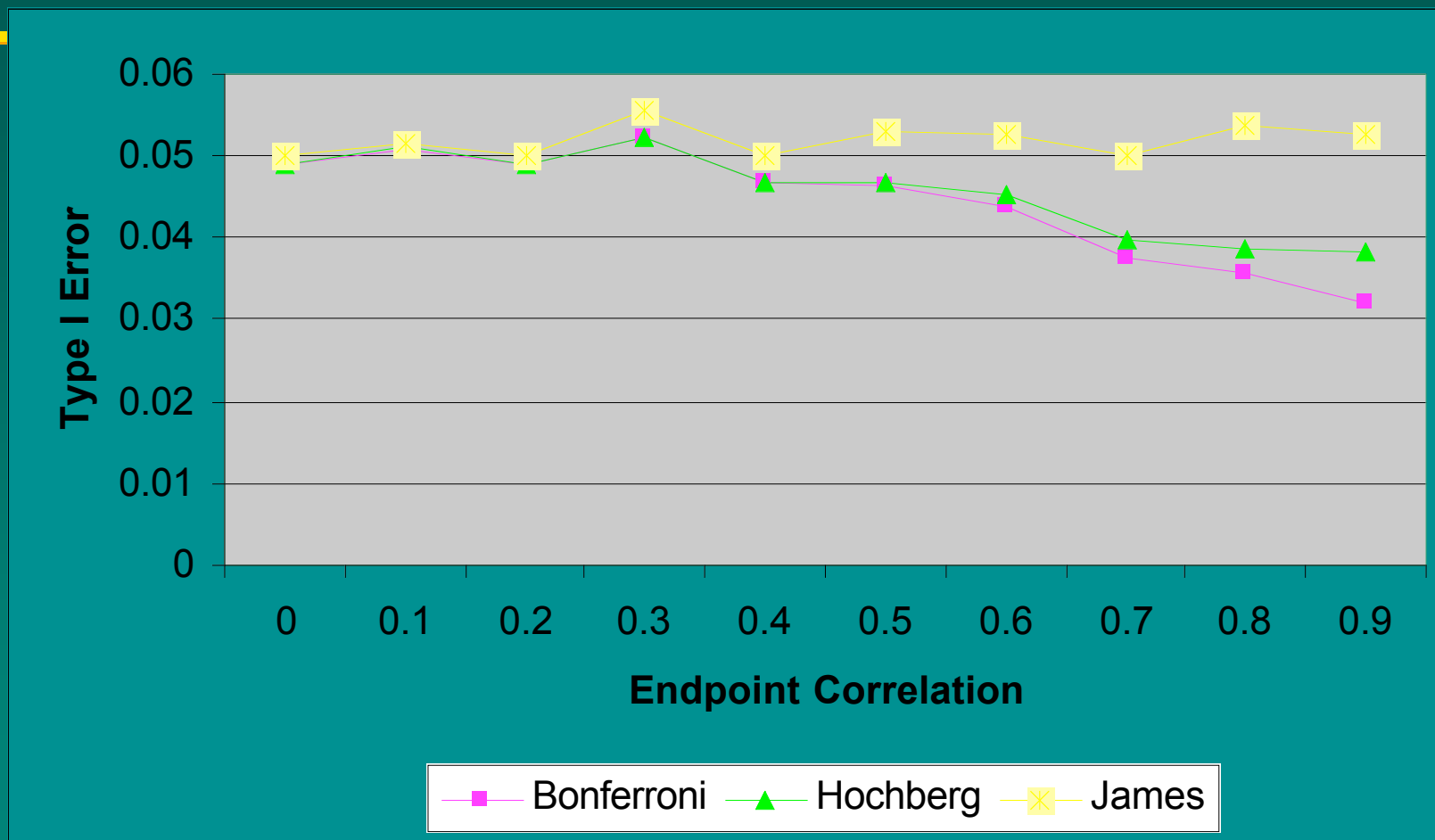
- **Rationale**
- Adjustment Strategies
- Application
- Simulation Study

Adjustment Strategies for Multiple χ^2 Tests: Type I Error



Endpoint rates = 30%; and $k = 3$; 10,000 Simulated data sets per correlation.

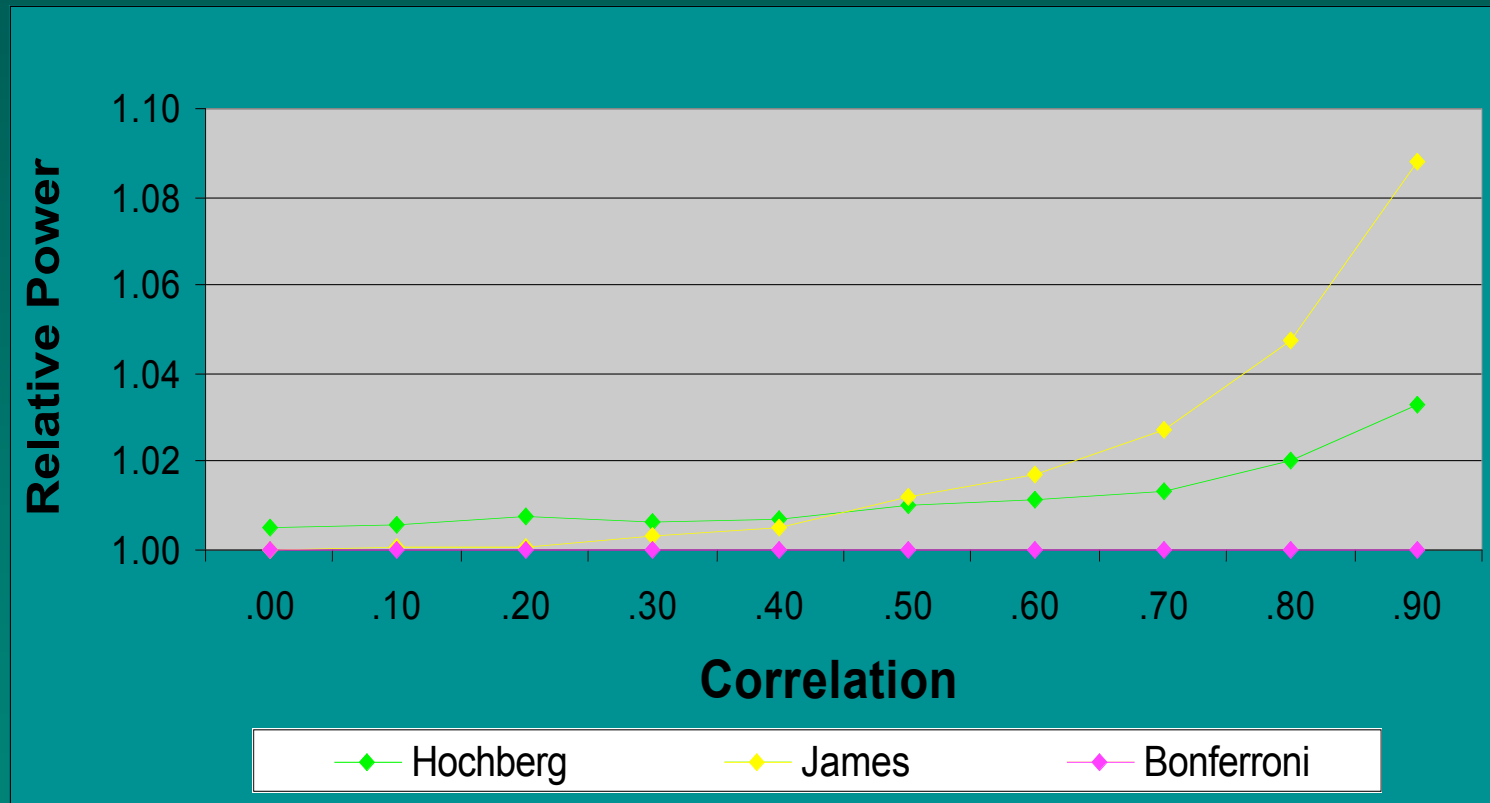
Three Adjustments: Type I Error



Endpoint rates = 30% vs 30%; $k = 3$

Leon & Heo. J Biopharm Stat 2005; 15:839-855

Power Relative to Bonferroni



* Power of 1 or more significant result.
10,000 Simulated data sets/correlation.

Endpoint rates: 25% vs. 40%
 $k = 3$ N per group=152

Summary

Reduce probability of *false positive* results among multiple cognitive endpoints:

- Pre-specify *one primary* endpoint
- If multiple measures are absolutely necessary, pre-specify alpha adjustment strategy and adjust N accordingly.

Summary

There are several alpha-adjustment strategies to reduce the probability of *false positive* results.

- Hochberg's sequentially rejective procedure is less conservative than Bonferroni, but not commonly used.
- For highly correlated endpoints, James' method is less conservative than others discussed here.

“If you half expect an event to occur, you rarely question when it does.”

William Boyd. Blue Afternoon, New York: Vintage Books, 1993.



Pseudospecificity

Indication for Cognitive Impairment in Schizophrenia

- Most indications in psychiatry focus on syndromes. Here we consider an indication for cognitive symptoms.
- How can RCT design demonstrate:
 - One drug treats both positive symptoms and cognitive impairment and comparator only treats positive symptoms
- Particularly when positive symptoms are a confounding variable in examining the effect of treatment on cognitive impairment.

Pseudospecificity

- **What approaches to design and analysis can provide evidence that the claim is not “artificially narrow”?**
- **Demonstrate that cognition changes independent of positive symptoms.**
- **Post randomization change on positive symptoms confounds the effect of treatment on cognitive impairment.**
- See Laughren T. Biol Psychiatry 2003.

Pseudospecificity: Analysis Stage

- Covariate adjustment to control for positive symptoms
Use of post-baseline covariates - “not advisable” [ICH: E9]
Post-baseline covariates cannot rule out pseudospecificity
- Path analysis (**Wright, 1934**)
 - **Correlational approach**
 - **Series of regression equations (1/dependent variable)**
 - **Covariate adjustment (of positive sx's – post randomization)**
 - **Not for causality in a randomized experiment**
 - **Causal direction is ambiguous for 2 domains assessed contemporaneously. Cannot rule out positive sx's change.**

Pseudospecificity: Design Stage

- **Stratify by a pre-specified confound (e.g., ordinal levels of baseline symptoms).**
 - **If concurrent change in positive symptoms, does provide not evidence of specific cognitive benefit**
- * **RCT Inclusion/Exclusion Criteria**
 - **Exclude patients whose positive symptoms have not stabilized**

-
- [Midsummer2.ppt](#)