Setting Alpha Levels for Testing Key Secondary Endpoints in Trials with Interim Assessments when the Primary and Secondary Endpoints are Correlated

Rebekkah S. Brown, DrPH
Joint research with Greg Cicconetti, PhD

GlaxoSmithKline

BASS XVI

09 November 2009

Outline

- Background
- Issue at hand
- Possible solutions
- Example simulation details
- Comments

Background

- Phase III study
- Large patient population (thousands)
- Long follow-up (2-4 years)
- Purpose: Compare active treatment vs. placebo
- Independent data monitoring committee—reviews ongoing data for safety & efficacy
- Primary endpoint
- Key secondary endpoint(s)—highly correlated with primary endpoint
- Multiple interim assessments & Final assessment

Primary Endpoint

- IDMC can stop trial for efficacy at an interim assessment
- Alpha-spending function specified to preserve overall type I error at two-sided α=0.05
- Examples: Cancer-specific mortality (oncology trial), Major adverse cardiovascular events (cardiovascular trial), ACR20 (arthritis trial), Headache response at 2 hours (migraine trial)

Secondary Endpoint(s)

- Question: What alpha level is used to test the secondary endpoint in order to preserve type I error rate at nominal 0.05 level?
- Secondary endpoint is not tested until the study has ended
- Secondary endpoint is positively correlated with primary endpoint
- Examples: All-cause mortality (oncology, cardiovascular), ACR 50 (arthritis trial), Headache response at 1 hour (migraine trial)

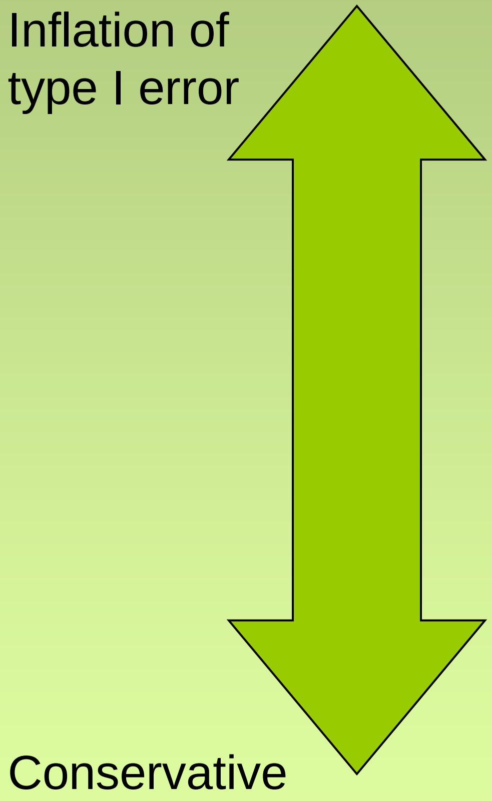
Issue

- Correlated primary & secondary endpoints can result in inflation of type I error rate for secondary endpoint testing when using traditional hierarchical testing
- Reference: Hung, Wang, O'Neill (JBS, 2007)
- Example: Test primary at alpha-spending level α_1 and if significance is achieved at interim 1 → Proceed to step-down to test secondary endpoint at $\alpha=0.05$
- Due to the correlation between primary & secondary, the type I error rate for the secondary endpoint is > 0.05 → inflation

Why does inflation occur?

- Trial stops early at an interim assessment for efficacy → primary endpoint shows significant treatment effect
- Secondary endpoint is also likely to show significant treatment effect because of high correlation with primary endpoint
- Type I error occurs more often for secondary if only rely on hierarchical testing at nominal level

Possible Solutions to Over-inflation



Inflation of type I error

Conservative

Test Secondary at:

- $\alpha=0.05$ → nominal alpha level
- $\alpha_i < \alpha < 0.05$ → optimal alpha level
- $\alpha=\alpha_i$ → alpha level for primary endpoint

How to find optimal alpha-level for testing secondary endpoint?

- Simulations!!!

- Optimal alpha-level is dependent on the scenario at hand:
  - Alpha-spending levels for primary endpoint
  - Number & timing of interim assessments
  - Correlation between primary & secondary endpoints
  - Treatment effect size of primary endpoint

Example Simulation Details (1)

- Primary endpoint: MACE—Major adverse cardiovascular events (includes non-fatal MI, non-fatal stroke, CV death)
- Secondary endpoint: All-cause mortality (includes CV death, non-CV death)
- Correlation between primary & secondary endpoint: Not specified as a correlation value, rather primary and secondary endpoints are composites which have a component in common (CV death):
  - Primary (P): X_1 (MI/stroke) + X_2 (CV death) → X_2 is 20% of primary
  - Secondary (S): Y_1 (non-CV death) + X_2 (CV death) → X_2 is 67% of secondary

Example Simulation Details (2)

- Primary & secondary endpoints: time-to-event
- Interim assessments:
  - Interim 1 (after 57% of primary events have accrued)
  - Interim 2 (after 77% of primary events have accrued)
  - Final
- Alpha-spending function for primary (Lan-DeMets, power function):
  - Interim 1: α=0.0005
  - Interim 2: α=0.001
  - Final: α=0.0499

Example Simulation Details (3)

- Treatment effect size of primary endpoint specified as range of possible hazard ratios: 0.6, 0.7, 0.8
- To assess type I error of secondary, set secondary treatment effect = 1 (i.e., no effect)
- Primary and Secondary event times are calculated as composites of X_1 & X_2 and Y_1 & X_2
- Correlation is built into Primary and Secondary due to common component: X_2 (CV death)

Example Simulation Details (4)

- How to set up no effect for all-cause mortality (HR=1) when there is a positive treatment effect on MACE and they have a component in common, CV death?
  - Case I: No treatment effect for Y_1 & X_2 (non-CV death and CV death); positively inflate treatment effect of X_1 (non-fatal MI/stroke) to compensate for no treatment effect for CV death component of MACE
  - Case II: Negatively inflate treatment effect of Y_1 (non-CV death) to compensate for positive treatment effect of X_2 (CV death)

Case I	$\uparrow\uparrow X_1$, $\leftrightarrow X_2$	MACE: trt effect=HR
	$\leftrightarrow Y_1$, $\leftrightarrow X_2$	All-cause mortality: no trt effect
Case II	$\uparrow X_1$, $\uparrow X_2$	MACE: trt effect=HR
	$\downarrow Y_1$, $\uparrow X_2$	All-cause mortality: no trt effect

Example Simulation Details (5)

- Interim assessments performed at 57%, 77%, and 100% of primary events accrued
- For each replication:
  - Interim 1: If primary p-value ≤ 0.0005 then stop study and test secondary at p-value ≤ p_{s1}
  - Interim 2: If primary p-value ≤ 0.001 then stop study and test secondary at p-value ≤ p_{s2}
  - If study is not stopped then proceed to Final: If primary p-value ≤ 0.0499 then test secondary at p-value ≤ p_{s3}

Example Simulation Details (6)

- For each scenario specify values for (p_{S1}, p_{S2}, p_{S3}) for testing of secondary endpoint
- Store p-values from each replication so can run many different scenarios for specification of (p_{S1}, p_{S2}, p_{S3})
- Type I error is calculated as % of replications that result in rejecting null hypothesis for secondary endpoint

Implementation Using an Exponential Distribution (1)

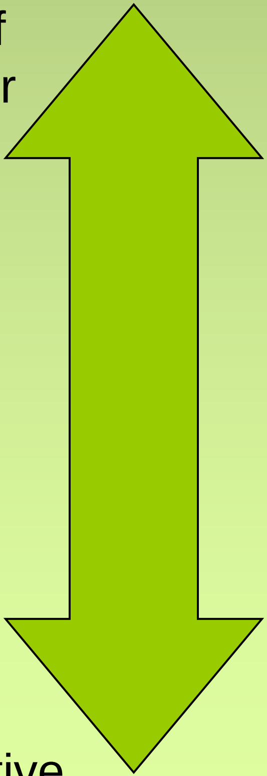
- Event times for X_1, X_2, Y_1 generated from exponential distributions for each subject → ½ subjects in active, ½ in placebo
  - X_1 = Time to non-fatal stroke, non-fatal MI
  - X_2 = Time to CV death
  - Y_1 = Time to non-CV death
  - Primary (MACE) = $\min(X_1, X_2)$
  - Secondary (All-cause mortality) = $\min(Y_1, X_2)$
- Make adjustments to account for enrollment period (subjects enter trial over time, not all at once)

Simulations Using an Exponential Distribution (2)

- Censor event times at 57% of events accrued for Interim 1 & if analysis of primary not significant;
- Censor event times at 77% of events accrued for Interim 2 & if analysis of primary not significant;
- Analyze events at end of trial
- Analysis of primary & secondary endpoint is Log-Rank test
- Calculate Type I error rate (for 10,000 replications)

Exponential Simulations—Type I Error Rate Results: Case I (Inflated Treatment Effect on Non-Fatal MACE)

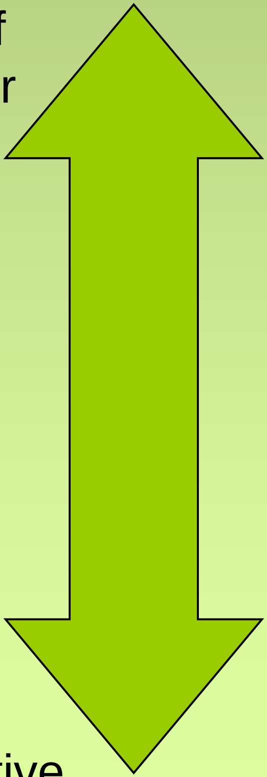
Inflation of type I error



Conservative

α for Secondary Endpoint (p_{S1}, p_{S2}, p_{S3})	Hazard Ratio for Primary Endpoint		
	0.6	**0.7**	**0.8**
(0.05, 0.05, 0.05)	0.1053	0.0957	0.0721
(0.01, 0.01, 0.05)	0.0660	0.0552	0.0450
(0.01, 0.01, 0.045)	0.0613	0.0508	0.0402
(0.01, 0.01, 0.04)	0.0556	0.0457	0.0371
(0.005, 0.01, 0.0499)	0.0558	0.0440	0.0376

Exponential Simulations—Type I Error Rate Results: Case II (Increased Risk of Non-CV Death)

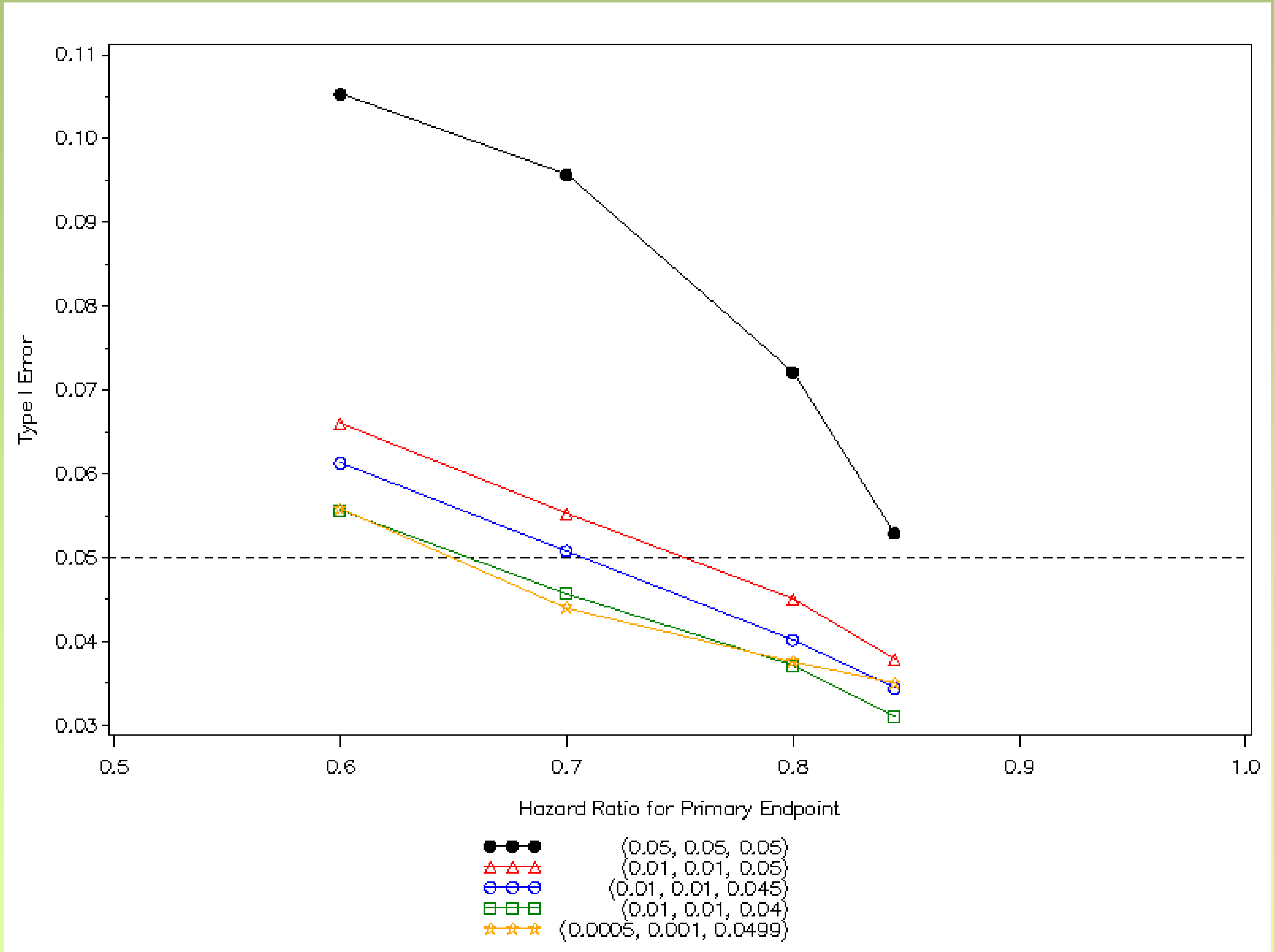


Inflation of type I error

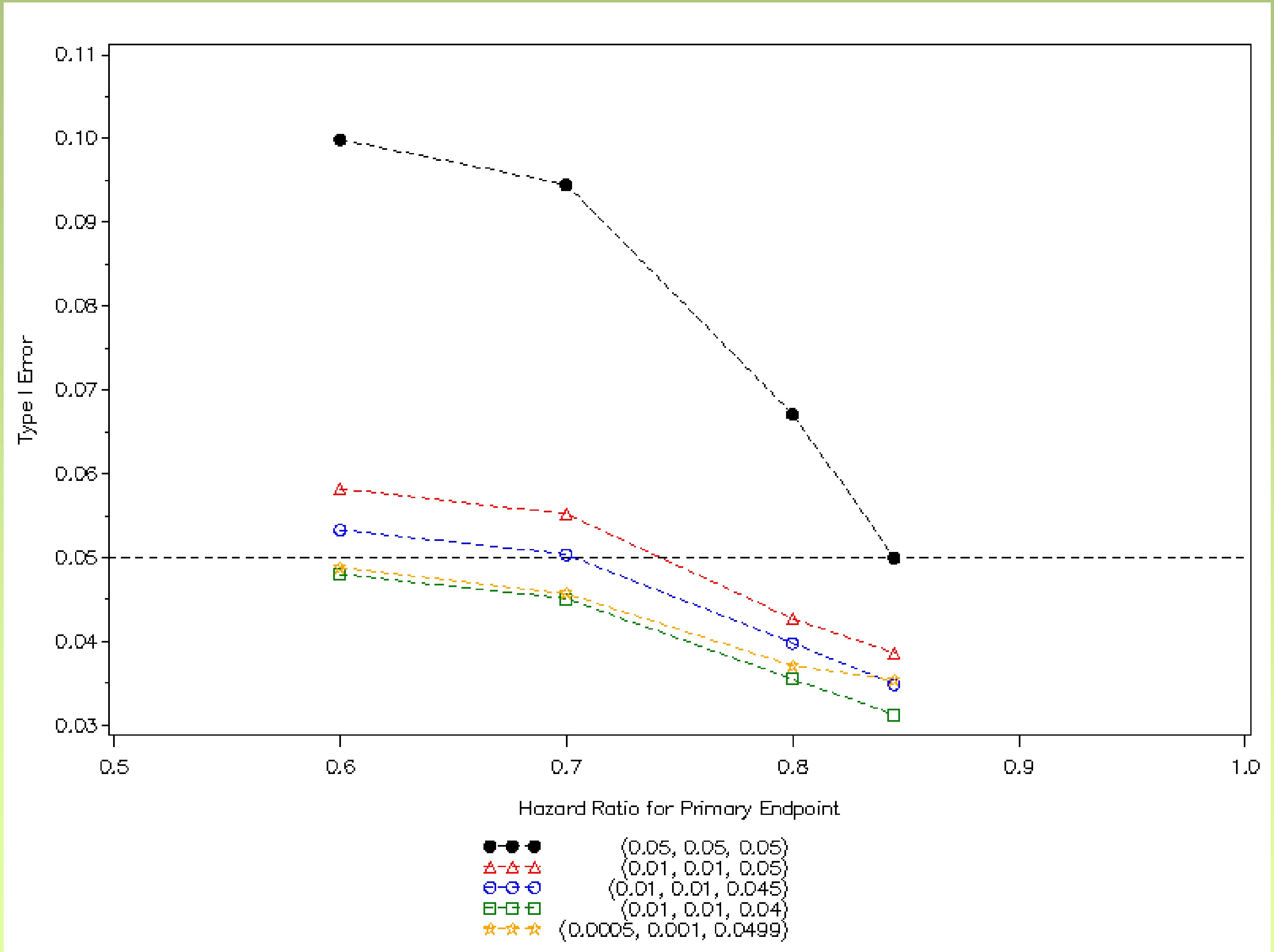
Conservative

α for Secondary Endpoint (p_{S1}, p_{S2}, p_{S3})	**Hazard Ratio for Primary Endpoint**		
	0.6	**0.7**	**0.8**
(0.05, 0.05, 0.05)	0.0999	0.0945	0.0671
(0.01, 0.01, 0.05)	0.0582	0.0552	0.0427
(0.01, 0.01, 0.045)	0.0533	0.0504	0.0398
(0.01, 0.01, 0.04)	0.0480	0.0451	0.0355
(0.005, 0.01, 0.0499)	0.0488	0.0457	0.0371

Summary of Type I Error Results: Case I (Inflated Treatment Effect on Non-Fatal MACE)



Summary of Type I Error Results: Case II (Increased Risk of Non-CV Death)



Comments (1)

- Final testing thresholds for secondary endpoint likely to be based on clinical & statistical arguments
    - How extreme of a treatment effect do we plan for?
    - What is likelihood of trial stopping at an interim?
    - Do we need to control type I error in all situations, even if likelihood of seeing situation is rare?

- What happens if there are multiple secondary endpoints with different correlations w/primary? (Likely not want to set different thresholds for each secondary)

Comments (2)

- Simulations are good starting point for discussion, but may not provide the final answer.

- Reference: Hung, Wang, O’Neill. Statistical Considerations for Testing Multiple Endpoints in Group Sequential or Adaptive Clinical Trials. *J Biopharm Stat* 2007;17:1201-1210.

- Literature search did not uncover any other thoughts on how to deal with this issue in a practical manner (i.e., how to go about specifying thresholds for testing secondary endpoint)…….late breaking…..

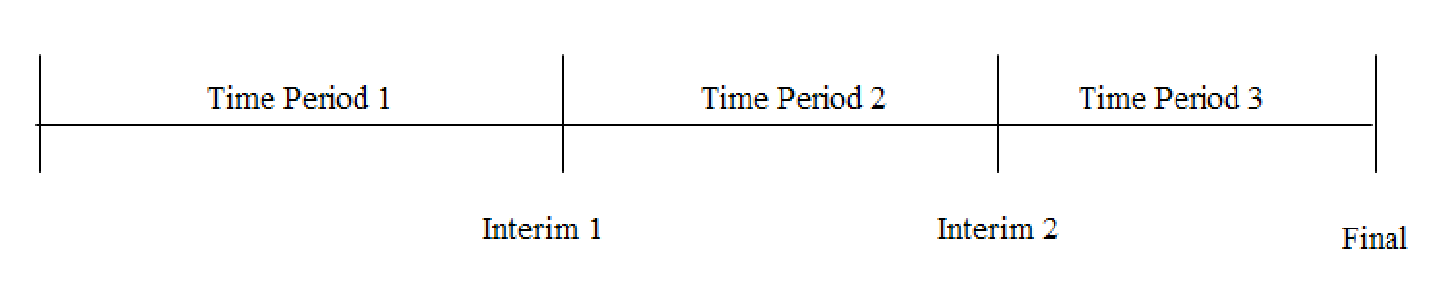
Late-Breaking Publication

- Reference: Glimm, Maurer, Bretz. Hierarchical testing of multiple endpoints in group-sequential trials. *Stats in Medicine;* Accepted 28 August 2009 (pulished online).
- Propose to use alpha spending functions for testing secondary endpoints in group sequential trials to control error rates.
- Suggest that Pocock method is optimal allocation of alpha (for power) for testing secondary endpoints when trial stops early.
- Also discuss other scenarios where interest may be continuing trial until secondary is significant.
- Comments: Need to consider specific trial at hand when deciding on alpha thresholds as the goals of the trial may influence the alpha levels for the secondary endpoints at each interim or final assessment.
- Reference similar work: Tamhane, Mehta, Liu. Testing a primary and a secondary endpoint in a group-sequential design. *Biometrics* 2009; submitted.

Questions/Comments?

Back-ups

‘Quick & Dirty’ Simulations—Using a Binomial Distribution



Time Period 1

Time Period 2

Time Period 3

Interim 1

Interim 2

Final

- Event counts for each time period generated using binomial distributions based on total patient-years of exposure → ½ subjects in active, ½ in placebo
- Analyze events accumulated at Interim 1 & if analysis of primary not significant;
- Analyze events accumulated at Interim 2 & if analysis of primary not significant;
- Analyze events at end of trial
- Analysis of primary & secondary endpoint is Chi-square test
- Calculate Type I error rate