A Gatekeeping Procedure to Test a Primary and a Secondary Endpoint in a Group Sequential Design with Multiple Interim Looks

Ajit Tamhane
Northwestern University

Joint work with
Jiangtao Gou, Chris Jennison, Cyrus Mehta & Teresa Curto

To appear in *Biometrics*

MCP2017 (June 21, 2017)

Outline

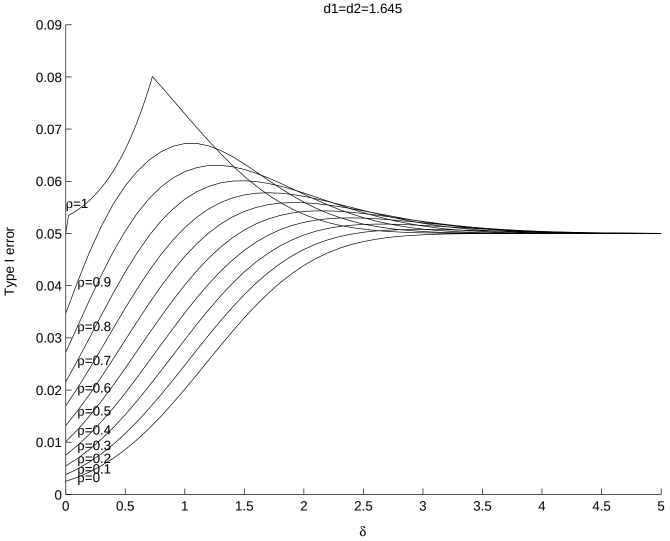
1. Introduction

- A parallel arm trial to compare a treatment with a control on a primary and a secondary endpoint hierarchically using a group sequential design (GSD).
- Gatekeeping condition: The secondary endpoint is tested only if the primary endpoint is significant.
- How to choose group sequential boundaries for the two endpoints to control familywise α?
- The primary endpoint can be tested using any α-level group sequential boundary.
- How to test the secondary endpoint and how to choose the primary-secondary boundary combination to maximize primary and secondary powers?

2. Previous Works

- Hung, Wang & O'Neill (2007), *J. Biopharm. Stats.*: For the two looks case, showed numerically that the common gatekeeping strategy for ordered hypotheses of propagating α from rejected H_1 to H_2 inflates FWER when used in a GSD. Proposed some ad-hoc strategies to fix this, e.g., test H_2 at level $\alpha/2$.
- Tamhane, Mehta & Liu (2010), *Biometrics*: For the two looks case, showed analytically Hung et al.'s (2007) result. Also showed that the secondary boundary can be relaxed to have level $\alpha' > \alpha$.
- Glimm, Maurer & Bretz (2010), *Stats. in Medicine*: Many results similar to Tamhane et al. (2010). Also considered several extensions of the basic procedure.

Plot of FWER for Hung et al.'s (2007) α-Propagation Strategy



3. Problem Formulation

- Bivariate normal responses on the primary and secondary endpoints with means (μ_{t1}, μ_{t2}) for the treatment and (μ_{c1}, μ_{c2}) for the control.
- Assume the variances (σ_1^2, σ_2^2) and the correlation coefficient ρ are common for the two groups.
- This normality setup applies asymptotically to broad types of data including survival and binary data.
- GSD with $K \geq 2$ looks (stages).
- Assume fixed boundaries. Primary boundary: $(c_1, \ldots, c_K)$, Secondary boundary: $(d_1, \ldots, d_K)$.

3. Problem Formulation (Cont'd.)

- Let $\delta_1 = \mu_{t1} - \mu_{c1} =$ primary treatment effect and $\delta_2 = \mu_{t2} - \mu_{c2} =$ secondary treatment effect.
- Test null hypotheses $H_1 : \delta_1 = 0$ and $H_2 : \delta_2 = 0$ against upper one-sided alternatives.
- Strongly control the familywise error rate (FWER):

$$\text{FWER} = P\{\text{Reject at least one true } H_i \ (i = 1, 2)\} \leq \alpha.$$

- Since $R_2 = (\text{Reject} H_2) \subseteq R_1 = (\text{Reject} H_1)$,

$$P\left(R_1 \cup R_2 | H_1\right) = P(R_1 | H_1) \leq \alpha,$$

so to control the FWER under H_1, the primary boundary must be of level α.

3. Problem Formulation (Cont'd.)

- Assume n_i patients on each treatment arm at the ith stage. Let $N_i = n_1 + \ldots + n_i$ denote the cumulative sample sizes and by $t_i = N_i/N_K$ the information times $(1 \le i \le K)$.
- At the ith look, let (X_i, Y_i) denote the standardized sample mean test statistics for the two endpoints.
- Procedure $\mathcal{P}_a$ (Stagewise Hierarchical Rule): Reject H_1 if $X_i > c_i$ for some $i \le K$. Then test H_2 and reject it if $Y_i > d_i$. The trial stops when H_1 is rejected (regardless of whether H_2 is rejected or not) or when the trial ends.
- Note H_2 has only one chance of being tested. Will consider an extension later.

3. Problem Formulation (Cont'd.)

- (X_i, Y_i) are bivariate normal with mean vector $(\Delta_{1i}, \Delta_{2i})$ where

$$\Delta_{1i} = \frac{\delta_1}{\sigma_1}\sqrt{\frac{N_i}{2}}, \quad \Delta_{2i} = \frac{\delta_2}{\sigma_2}\sqrt{\frac{N_i}{2}} \quad (1 \le i \le K)$$

  and correlation structure which depends on ρ and the $\gamma_i = \sqrt{t_i}$.
- Define the standardized treatment effects for the two endpoints by

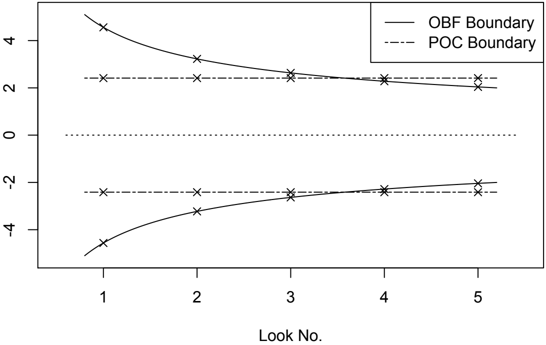
$$\Delta_1 = \Delta_{1K} = \frac{\delta_1}{\sigma_1}\sqrt{\frac{N_K}{2}} \quad \Delta_2 = \Delta_{2K} = \frac{\delta_2}{\sigma_2}\sqrt{\frac{N_K}{2}}.$$

  Then $\Delta_{1i} = \gamma_i \Delta_1$ and $\Delta_{2i} = \gamma_i \Delta_2$.

4. Primary Boundary

- Theorem: Suppose $(X_1, \ldots, X_K)$ has a multivariate normal distribution (more generally an MLR distribution) defined above. Consider two α-level group sequential tests with the same total sample size: Test A with boundary $(a_1, \ldots, a_K)$ and Test B with boundary $(b_1, \ldots, b_K)$ for testing H_1: $\delta_1 = 0$ vs. $\delta_1 > 0$. If for some $k^* \leq K - 1$, $a_i > b_i$ for $i = 1, \ldots, k^*$ and $a_i < b_i$ for $i = k^* + 1, \ldots, K$ then Test A is uniformly more powerful than Test B for all $\delta_1 > 0$.
- Idea of the proof: Test A tends to stop later than Test B and hence tends to take more observations. Hence using the likelihood ratio test, A is more powerful than B.

OBF and POC Two-Sided Boundaries for $\alpha = .05$



- Corollary: The O'Brien-Fleming (OBF) boundary is uniformly more powerful than the Pocock (POC) boundary.

5. Secondary Boundary

- Control the FWER under $H_2 : \delta_2 = 0$ when H_1 is false. We refer to this FWER also as secondary type I error.
- Denote it by $\alpha_2(\Delta_1, \rho)$ and let

$$\Delta_{1i}^0 = (c_i - d_i)/\gamma_i \ (1 \le i \le K).$$

- Theorem: We have

$$\alpha_2 = \max_{\Delta_1, \rho} \alpha_2(\Delta_1, \rho) \le 1 - P_{H_2}\{Y_1 \le d_1, \ldots, Y_K \le d_K\},$$

  and this bound is sharp iff

$$\Delta_{11}^0 = \cdots = \Delta_{1,K-1}^0 \ge \Delta_{1K}^0 \quad \text{and} \quad \rho = 1.$$

- The above condition is satisfied for $K = 2$ if $c_1 > d_1$ and $c_2 < d_2$ (e.g., (c_1, c_2) is OBF and (d_1, d_2) is POC) since $\Delta_{11}^0 > 0$ and $\Delta_{12}^0 < 0$.

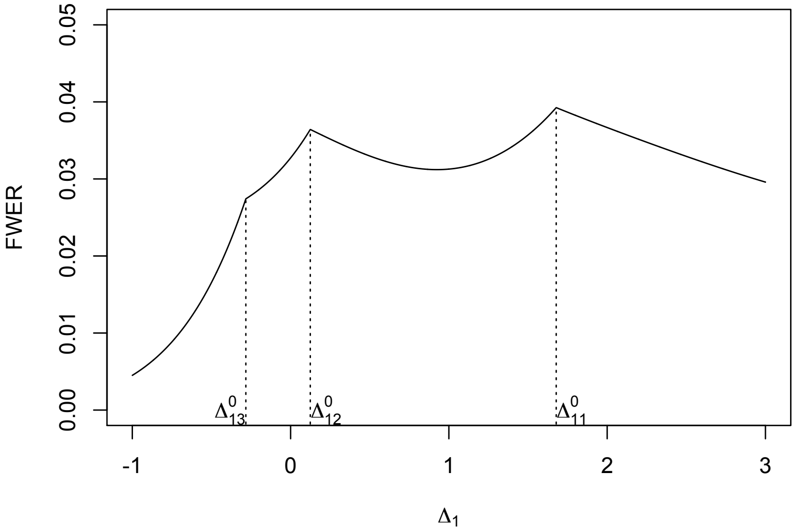
5. Secondary Boundary (Cont'd.)

- For $\rho = 1$, under H_2 we can write $X_i = Y_i + \Delta_{1i}$ $(1 \le i \le K)$ where $Y_i \sim N(0,1)$. So

$$\begin{aligned} & \alpha_2(\Delta_1, \rho = 1) \\ = & \sum_{i=1}^{K} P\{Y_1 \le c_1 - \gamma_1\Delta_1, \ldots, Y_{i-1} \le c_{i-1} - \gamma_{i-1}\Delta_1, \\ & Y_i > \max(c_i - \gamma_i\Delta_1, d_i)\}. \end{aligned}$$

- If $\alpha_2(\Delta_1, \rho = 1)$ is plotted as a function of Δ_1 then the plot has sharp peaks where $\max(c_i - \gamma_i\Delta_1, d_i)$ changes from $c_i - \gamma_i\Delta_1$ to d_i, i.e., where $\Delta_1 = \Delta_{1i}^0 = (c_i - d_i)/\gamma_i \quad (1 \le i \le K)$.

Plot of FWER for $\rho = 1$ as a Function of Δ_1



5. Refined Secondary Boundary

- The upper bound $1 - P_{H_2}\{Y_1 \leq d_1, \ldots, Y_K \leq d_K\}$ on $\max \alpha_2(\Delta_1, \rho)$ is sharp iff

$$\Delta_{11}^0 = \cdots = \Delta_{1,K-1}^0 \geq \Delta_{1K}^0.$$

- If this condition is satisfied then $\{d_1, \ldots, d_K\}$ must be an α-level boundary.
- This condition is satisfied for $K = 2$ if $c_1 \geq d_1$ and $c_2 \leq d_2$ (OBF-POC boundary combination) but not if $c_1 < d_1$ and $c_2 > d_2$ (POC-OBF boundary combination).
- This condition is satisfied for $K > 2$ only if the primary and secondary boundaries are identical.
- Otherwise $\max \alpha_2(\Delta_1, \rho) < \alpha$, so the secondary boundary can be refined to have a level $\alpha' > \alpha$.

5. Refined Secondary Boundary (Cont'd.)

- To calculate the refined secondary boundary, set

$$\max_{1\le i\le K} \alpha_2(\Delta_{1i}^0, \rho = 1) = \alpha,$$

where

$$\begin{aligned} & \alpha_2(\Delta_{1i}^0, \rho = 1) \\ = \; & \sum_{j=1}^{K} P\{Y_1 \le c_1 - \gamma_1 \Delta_{1i}^0, \ldots, Y_{j-1} \le c_{j-1} - \gamma_{j-1}\Delta_{1i}^0, \\ & Y_j > \max(c_j - \gamma_i \Delta_{1i}^0, d_j)\}, \end{aligned}$$

where

$$c_k - \gamma_k \Delta_{1i}^0 = c_k - (\gamma_k/\gamma_i)(c_i - d_i) \ (1 \le k \le j).$$

- Parameterize d_i by a single d, e.g., for the POC boundary set $d_i = d$ and for the OBF boundary set $d_i = d/\gamma_i$.
- Solve the above equation for d.

5. Original and Refined Secondary Boundaries (Cont'd.)

Primary Boundary: O'Brien-Fleming, Secondary Boundary: Pocock

	Original		Refined	
K	d	α_2	d	α'
2	1.876	0.050	1.876	0.050
3	1.992	0.039	1.881	0.063
4	2.067	0.033	1.877	0.075

6. Effect of ρ

- The least favorable configuration (LFC) $\rho = 1$ is practically not likely: An example: noninferiority-superiority testing.
- If true $\rho < 1$ how much can the secondary boundary be sharpened? Here are the critical values $d_i = d$ for the OBF-POC combination for different ρ.

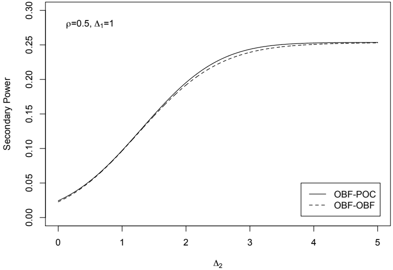
K	ρ					
	0.0	0.2	0.4	0.6	0.8	1.0
3	1.645	1.670	1.698	1.729	1.767	1.881
4	1.645	1.669	1.695	1.726	1.767	1.877

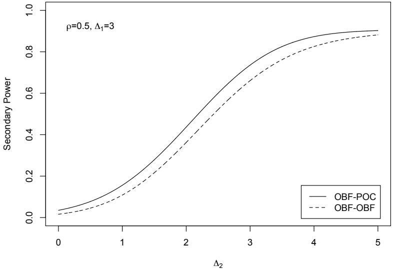
- In practice ρ is unknown. In Tamhane, Wu & Mehta (2012) we showed how to use an upper confidence limit on ρ to sharpen the secondary boundary for $K = 2$. We have not pursued this for $K > 2$.

7. Simulated Power Comparisons

- Compare two boundary combinations:
  1. α-level OBF boundary for H_1 and α'-level POC boundary for H_2.
  2. α-level OBF boundary for both H_1 and H_2.
- Plot power vs. Δ_2 for $\Delta_1 = 1.0, 3.0, \rho = 0.5, K = 3, \alpha = 0.05$.
- Conclusion: OBF-POC combination has higher power (uniformly for large Δ_1).

7. Simulated Power Comparisons (Cont'd.)





8. Example

- Randomized Aldactone Evaluation Study (RALES) (Pitts et al. 1999, Wittes et al. 2001).
- Goal: Evaluate the efficacy of spironolactone for patients who had severe heart failure.
- Multicenter double-blind randomized trial with 811 patients on the treatment and 831 patients on placebo.
- A group sequential design (GSD) with Lan-DeMets (1983) flexible boundary using the O'Brien-Fleming (OBF) error spending function.
- Primary endpoint: All cause deaths, Secondary endpoint: Sudden CV deaths (not used in the trial for formal α control).

8. Example (Cont'd.)

- The trial was monitored semi-annually by the DMC and stopped early at the 5th look due to significant efficacy on the primary ($\alpha = 0.025$).
- The trial was planned assuming a total of 1080 all-cause deaths.
- The looks occurred approximately at equal information times spaced 135 deaths (0.125 units) apart, i.e., total $K = 8$ looks in a fixed GSD trial.
- How to choose the primary and secondary boundaries to control overall α subject to the gatekeeping condition?

8. Example (Primary Endpoint)

Lan-DeMets Boundary Using the OBF Error Spending Function ($\alpha = 0.05$) for the Primary Endpoint with Log-Rank Statistics

Look	Placebo		Treat.		Info.	Rel.	Obs.	Crit.
No.	Enroll.	Deaths	Enroll.	Deaths	Frac.	Risk	X_i	c_i
1	563	81	543	59	0.130	0.755	1.820	6.117
2	830	189	809	139	0.304	0.755	2.719	3.903
3	830	254	810	199	0.419	0.803	2.744	3.278
4	830	327	811	251	0.535	0.786	3.357^*	2.876
5	831	380	811	279	0.610	0.752	4.414^*	2.704

8. Example (Secondary Endpoint)

Lan-DeMets Boundary Using the Refined POC Error Spending Function for the Secondary Endpoint with Associated Log-Rank Statistics

Look	Placebo		Treat.		Info.	Rel.	Obs.	POC'
No.	Enroll.	Deaths	Enroll.	Deaths	Frac.	Risk	Y_i	c_i
1	563	29	543	15	0.130	0.536	2.073	2.345
2	830	57	809	44	0.304	0.792	1.270	2.228
3	830	71	810	59	0.419	0.852	1.113	2.257
4	830	91	811	76	0.535	0.855	1.268	2.236
5	831	109	811	82	0.610	0.771	2.224	2.259

9. Extensions

- Procedure $\mathcal{P}_b$ (Overall Hierarchical Rule): Continue the trial after rejection of H_1 and sequentially test H_2 until it is rejected or the trial stops.
- Theorem: Denote the secondary type I errors of procedures $\mathcal{P}_a$ and $\mathcal{P}_b$ by $\alpha_2^a(\Delta_1, \rho)$ and $\alpha_2^b(\Delta_1, \rho)$. Then we have

$$\alpha_2^b(\Delta_1, \rho) \leq 1 - P_{H_2}\{Y_1 \leq d_1, \ldots, Y_K \leq d_K\}.$$

  This upper bound is sharp iff $\rho = 1$ and $\Delta_1 \geq \max_{1 \leq i \leq K}(\Delta_{1i}^0)$. Therefore the secondary boundary must have level α.
- Other extensions: Multiple primary and secondary endpoints: ordered or unordered.

References

1. Glimm, E., Maurer, W., and Bretz, F. (2010). Hierarchical testing of multiple endpoints in group-sequential trials. *Statistics in Medicine* 29, 219-228.
2. O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* 35, 549-556.
3. Pitt, B. et al. (1999). The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *New England Journal of Medicine* 341, 709-717.
4. Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* 64, 191-199.
5. Tamhane, A. C., Mehta, C. R., and Liu, L. (2010). Testing a primary and a secondary endpoint in a group sequential design. *Biometrics* 66, 1174-1184.
6. Tamhane, A. C., Wu, Y., and Mehta, C. R. (2012). Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (I): unknown correlation between the endpoints. *Statistics in Medicine* 31, 2027-2040.