A Bayesian Adaptive Allocation Method for Clinical Trials with Dual Objectives

Roy T. Sabo[1], Cathy Roberts[2], Amir A. Toor[2]
and John M. McCarty[2]

[1] Department of Biostatistics
[2] Bone Marrow Transplant Division
Virginia Commonwealth University

Monday, November 5, 2012

- Retrospective cost-effectiveness study:
  - Bone-marrow transplant patients treated at VCU Medical Center (2003-2010).
  - Diagnosed with Hodgkin's disease, multiple myeloma or non-Hodgkin's lymphoma.
  - Peripheral blood stem cell mobilization with one of four treatments: two standard therapies and two experimental therapies.
- Two primary outcomes:
  1. Treatment efficacy: $\geq 5 \times 10^6$ CD34+ stem cells collected per kg bodyweight.
  2. Treatment futility: ≥ 5 days needed for stem cell collection.
- A prospective Phase II clinical trial was planned from this study.
  1. Could we use adaptive allocation?
  2. Could we account for both objectives (efficacy and futility)?

- Standard practice in clinical trials: fixed-ratio randomization.
  - Balanced design (e.g. $1:1$ or $1:1:1$ ratio).
  - Unbalanced design (e.g. $2:1$ or $2:2:1$).
  - Can needlessly expose patients to ineffective or harmful treatments.
- Adaptive Allocation or Adaptive Randomization:
  - Allocation proportions can change throughout trial.
  - Patients more likely to receive more efficacious treatments (Berry, 2001 2004).
  - Minimize patients receiving ineffective, inferior or toxic treatments (Berry 2001).
  - "Bandit" methods: Thompson (1933), Bather (1981).
  - "Pick-the-winner" or "play-the-winner" methods: Robbins (1952), Chang (2008).

- What about 'Dual' primary outcomes?
  - Previous methods don't *directly* apply.
- Assume outcomes are dichotomous (e.g. success or failure).
  - Outcomes need not be immediately observable, provided such delays are small (Zelen 1969).
  - Extreme cases may delay changes in weights (Berry and Eick 1995).
- Fix total sample size at n and treatment groups at k.
  - $\theta_j, j = 1, \ldots, k$, represent first outcome.
  - $\lambda_j, j = 1, \ldots, k$, represent second outcome.
  - Dichotomous observations: these are generally proportions.
- Three ways to compare treatments:
  - Inter-treatment comparisons.
  - Hypothesized or historical efficacy / toxicity rates.
  - Hybrid approach.

- Compare "success" rates for both outcomes between treatments:
  - $P^1_{j\ell} = P(\theta_j > \theta_\ell)$ for the first outcome.
  - $P^2_{j\ell} = P(\lambda_j > \lambda_\ell)$ for the second outcome.

$$w_j = \frac{\left(\Pi_{\ell=1}^{k} P^1_{j\ell} P^2_{j\ell}\right)^{c(n)}}{\sum_{i=1}^{k}\left(\Pi_{\ell=1}^{k} P^1_{i\ell} P^2_{i\ell}\right)^{c(n)}}$$

- Compare "success" rates to hypothesized values (p^1_o and p^2_o).
  - $P^1_j = P(\theta_j > p^1_o)$ for the first outcome.
  - $P^2_j = P(\lambda_j > p^2_o)$ for the second outcome.

$$w_j = \frac{\left(P^1_j P^2_j\right)^{c(n)}}{\sum_{i=1}^{k}\left(P^1_i P^2_i\right)^{c(n)}}$$

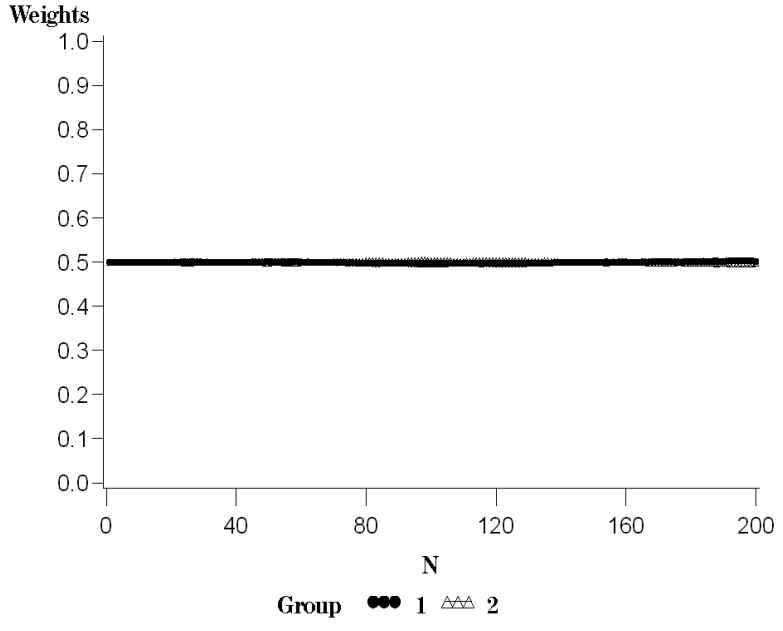
- Hybrid approach: compare one outcome between treatments, the other to hypothesized value.

$$w_j = \frac{\left(P^2_j \Pi_{\ell=1}^{k} P^1_{j\ell}\right)^{c(n)}}{\sum_{i=1}^{k}\left[P^2_i\left(\Pi_{\ell=1}^{k} P^1_{i\ell}\right)\right]^{c(n)}}$$

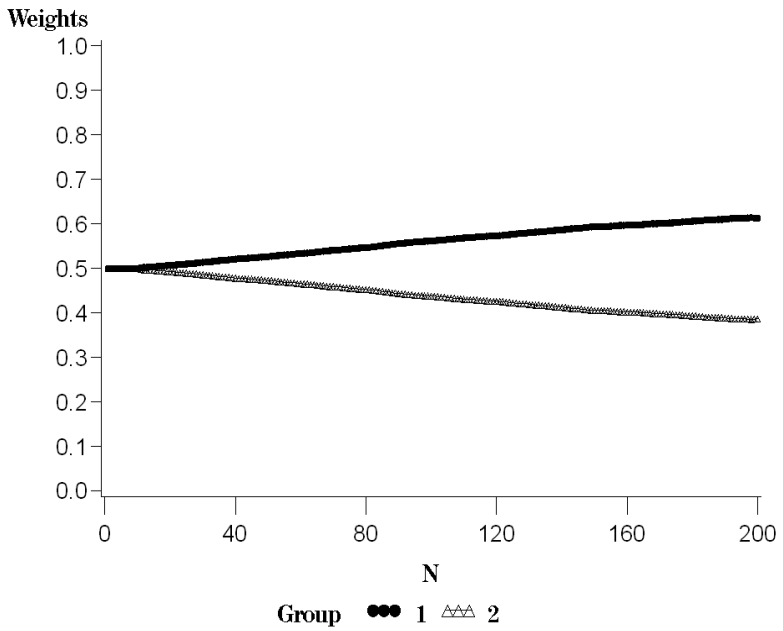
- Each special case is repeated $r = 1,000$ times.
  - Trials consist of $n = 200$ simulated subjects.
  - Lead-in of first 10 subjects.
  - Thereafter, weights are allowed to adapt.
- Calculation of posterior probabilities (for simplicity):
  - **Informative** and **skeptical** *beta* priors on efficacy/toxicity rates.
  - Binomial likelihood for efficacy/toxicity frequencies.
  - Conjugate pair yields *beta* posteriors.
- These choices allow.
  - Comparisons to hypothesized values: probabilities obtained directly.
  - Inter-treatment comparisons: MCMC (or integration) methods used.

Simulation Results from 2-Arm Study.

- Efficacy compared to hypothesized value (30%).
- Toxicity compared to hypothesized value (10%).



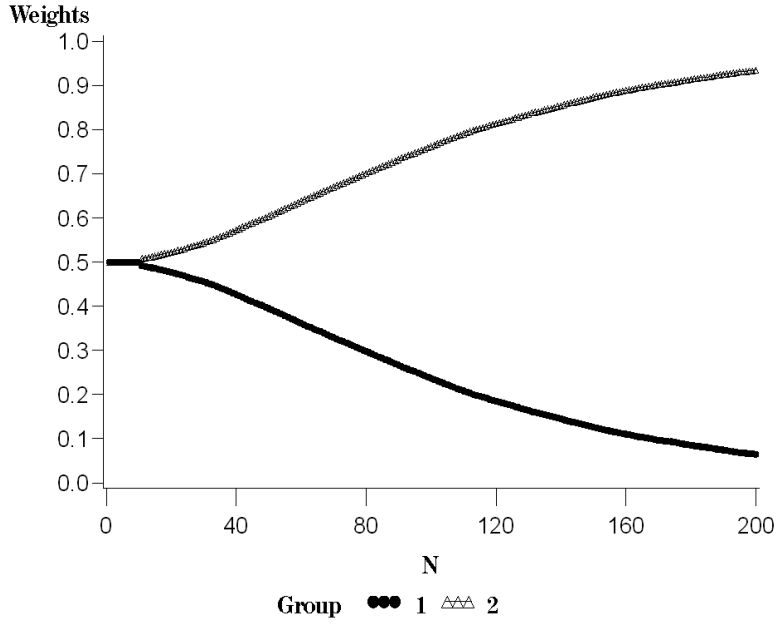
(a) $p_1^e = p_2^e = 0.3; p_1^t = p_2^t = 0.1$



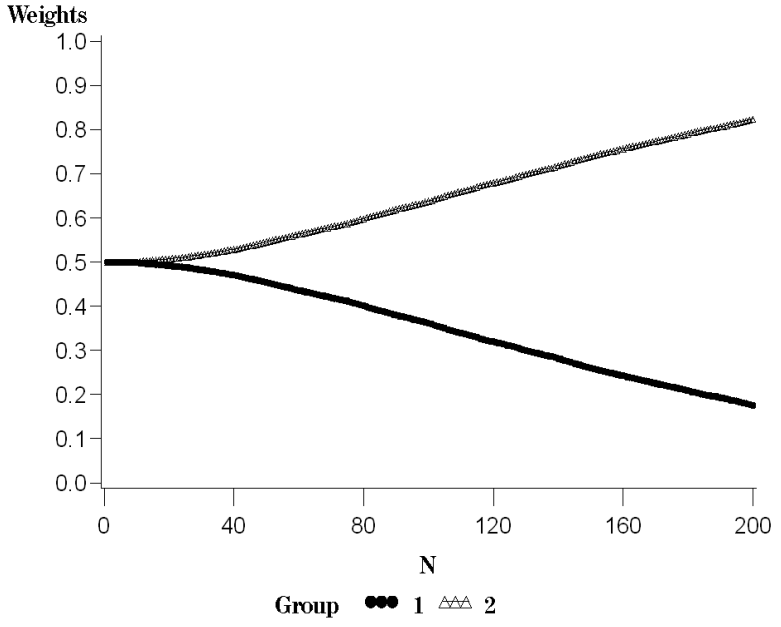
(b) $p_1^e = 0.5, p_2^e = 0.3; p_1^t = p_2^t = 0.1$

Simulation Results from 2-Arm Study.

- Efficacy compared to hypothesized value (30%).
- Toxicity compared to hypothesized value (10%).



(c) $p_1^e = p_2^e = 0.3; p_1^t = 0.25, p_2^t = 0.1$



(d) $p_1^e = 0.5, p_2^e = 0.3; p_1^t = 0.2, p_2^t = 0.1$

Simulation Results from 2-Arm Study.

- Average Sample Size.
- Comparisons made to Hypothesized Values $(p_o^e = 0.3, p_o^t = 0.1)$.

Parameters	Sample Size	Standard Deviation	Parameters	Sample Size	Standard Deviation
$p_1^e = 0.3$			$p_1^e = 0.3$		
$p_2^e = 0.3$	$\hat{n}_1 = 101.0$	$SD = 22.6$	$p_2^e = 0.3$	$\hat{n}_1 = 52.2$	$SD = 16.9$
$p_1^t = 0.1$	$\hat{n}_2 = 99.0$	$SD = 22.6$	$p_1^t = 0.25$	$\hat{n}_2 = 147.8$	$SD = 16.9$
$p_2^t = 0.1$			$p_2^t = 0.1$		
$p_1^e = 0.5$			$p_1^e = 0.5$		
$p_2^e = 0.3$	$\hat{n}_1 = 111.6$	$SD = 19.9$	$p_2^e = 0.3$	$\hat{n}_1 = 71.2$	$SD = 21.4$
$p_1^t = 0.1$	$\hat{n}_2 = 88.4$	$SD = 19.9$	$p_1^t = 0.2$	$\hat{n}_2 = 128.8$	$SD = 21.4$
$p_2^t = 0.1$			$p_2^t = 0.1$		

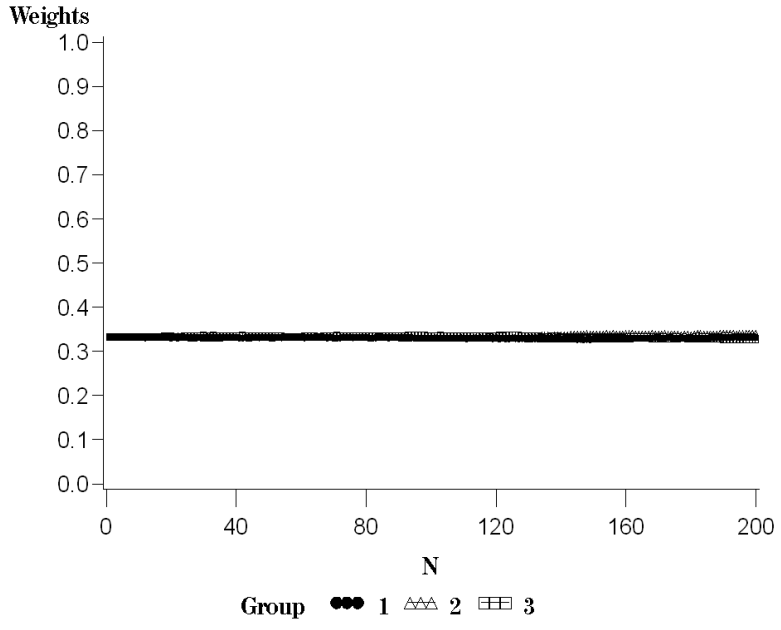
Simulation Results from 2-Arm Study.

- Percentage of Larger Samples and Decisions in Favor.
- Comparisons made to Hypothesized Values ($p_o^e = 0.3, p_o^t = 0.1$).
- Case 1: $p_1^e = p_2^e = 0.3, p_1^t = p_2^t = 0.1$.
- Case 2: $p_1^e = 0.5, p_2^e = 0.3, p_1^t = p_2^t = 0.1$.
- Case 3: $p_1^e = p_2^e = 0.3, p_1^t = 0.25, p_2^t = 0.1$.
- Case 4: $p_1^e = 0.5, p_2^e = 0.3, p_1^t = 0.2, p_2^t = 0.1$.

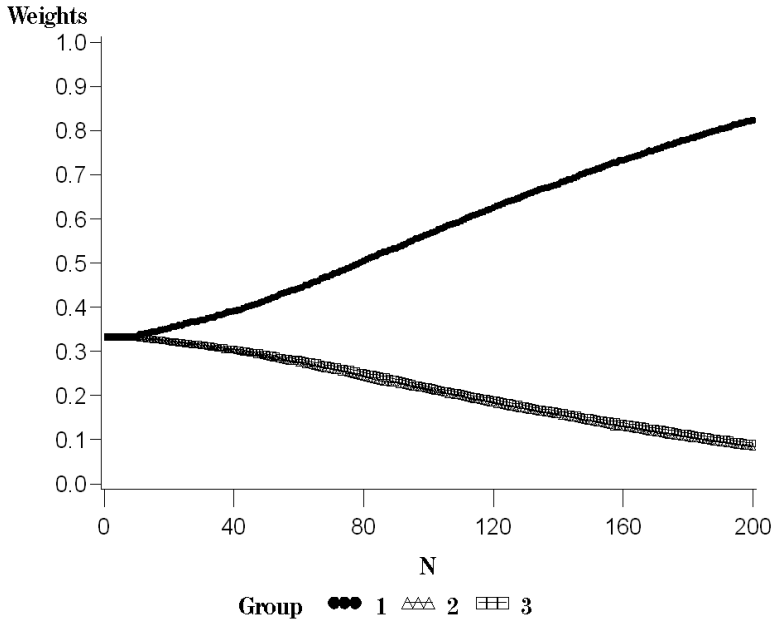
	Reject in Favor of				% of Samples	
	Arm 1 Eff.	Arm 2 Eff.	Arm 1 Tox.	Arm 2 Tox.	$n_1 > n_2$	$n_2 > n_1$
Case 1 (Adapt)	3.3%	3.4%	2.7%	2.4%	50.6%	49.4%
Case 1 (Equal)	3.3%	2.1%	2.6%	2.5%	–	–
Case 2 (Adapt)	83.1%	0.0%	3.4%	3.7%	72.9%	27.1%
Case 2 (Equal)	83.6%	0.0%	2.2%	2.4%	–	–
Case 3 (Adapt)	2.5%	1.9%	80.5%	0.0%	1.3%	98.7%
Case 3 (Equal)	2.0%	2.7%	81.2%	0.0%	–	–
Case 4 (Adapt)	77.0%	0.0%	54.1%	0.1%	8.9%	91.1%
Case 4 (Equal)	81.4%	0.0%	51.4%	0.0%	–	–

Simulation Results from 3-Arm Study.

- Efficacy compared between treatments.
- Toxicity compared to hypothesized value (10%).



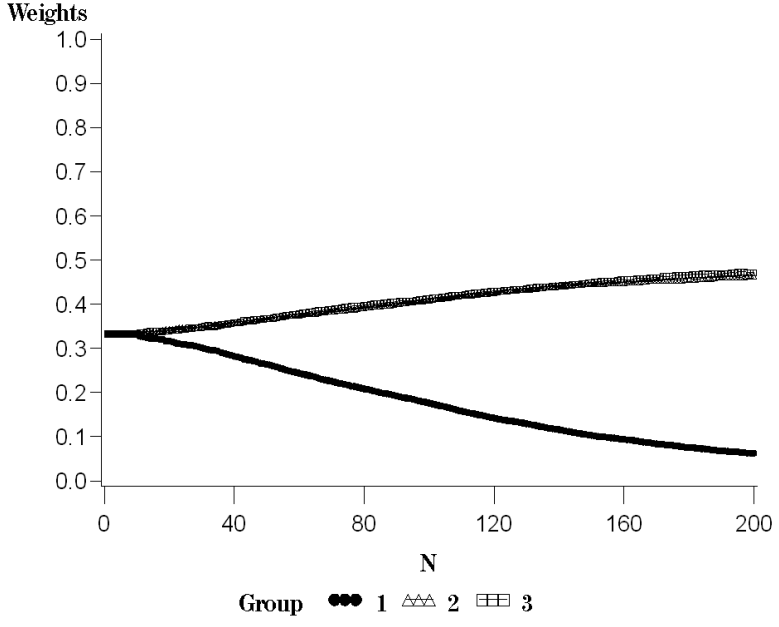
(e) $p_j^e = 0.3, p_j^t = 0.1, j = 1, 2, 3$



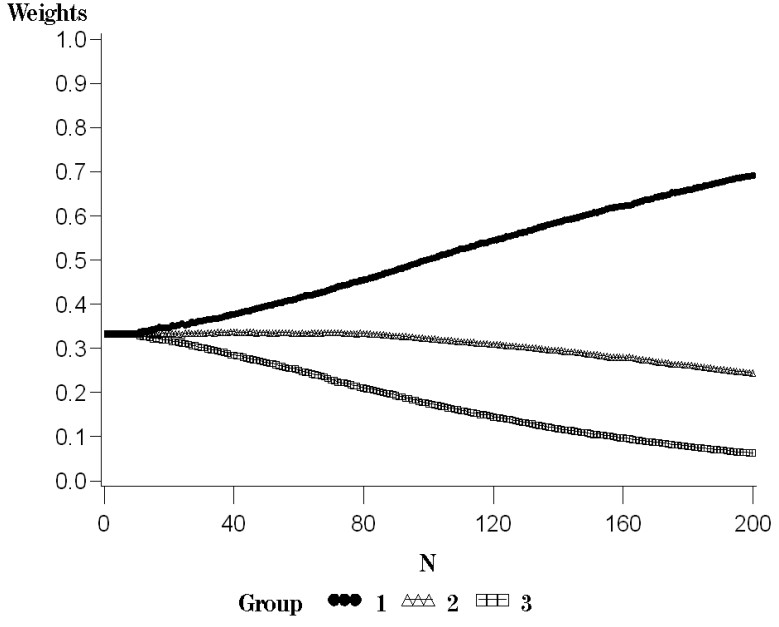
(f) $p_1^e = 0.5, p_2^e = p_3^e = 0.3; p_j^t = 0.1$

Simulation Results from 3-Arm Study.

- Efficacy compared between treatments.
- Toxicity compared to hypothesized value (10%).



(g) $p_j^e = 0.3; p_1^t = 0.25, p_2^t = p_3^t = 0.1$



(h) $p_1^e = 0.4, p_2^e = 0.3, p_3^e = 0.2; p_j^t = 0.1$

Simulation Results from 3-Arm Study.

- Average Sample Size.
- Efficacy Compared between Treatments.
- Toxicity compared to hypothesized values ($p_o^t = 0.1$).

Parameters	Sample Size	Parameters	Sample Size
$p_1^e = 0.3$		$p_1^e = 0.3$	
$p_2^e = 0.3$	$\hat{n}_1 = 66.0, SD = 23.1$	$p_2^e = 0.3$	$\hat{n}_1 = 37.2, SD = 14.7$
$p_3^e = 0.3$	$\hat{n}_2 = 66.2, SD = 23.0$	$p_3^e = 0.3$	$\hat{n}_2 = 81.2, SD = 25.7$
$p_1^t = 0.1$	$\hat{n}_3 = 67.8, SD = 22.5$	$p_1^t = 0.25$	$\hat{n}_3 = 81.6, SD = 26.4$
$p_2^t = 0.1$		$p_2^t = 0.1$	
$p_3^t = 0.1$		$p_3^t = 0.1$	
$p_1^e = 0.5$		$p_1^e = 0.4$	
$p_2^e = 0.3$	$\hat{n}_1 = 113.0, SD = 21.9$	$p_2^e = 0.3$	$\hat{n}_1 = 101.0, SD = 23.4$
$p_3^e = 0.3$	$\hat{n}_2 = 42.9, SD = 17.1$	$p_3^e = 0.2$	$\hat{n}_2 = 61.0, SD = 21.8$
$p_1^t = 0.1$	$\hat{n}_3 = 44.1, SD = 17.3$	$p_1^t = 0.1$	$\hat{n}_3 = 38.0, SD = 15.0$
$p_2^t = 0.1$		$p_2^t = 0.1$	
$p_3^t = 0.1$		$p_3^t = 0.1$	

Simulation Results from 3-Arm Study.

- Percentage of Larger Samples and Decisions in Favor.
- Efficacy Compared between Treatments.
- Toxicity compared to hypothesized values ($p_o^t = 0.1$).
- Case 1: $p_j^e = 0.3, p_j^t = 0.1, j = 1, 2, 3$. Case 2: $p_1^e = 0.5, p_2^e = p_3^e = 0.3, p_j^t = 0.1, j = 1, 2, 3$.
- Case 3: $p_j^e = 0.3, j = 1, 2, 3, p_1^t = 0.25, p_2^t = p_3^t = 0.1$. Case 4: $p_1^e = 0.4, p_2^e = 0.3, p_3^e = 0.2, p_j^t = 0.1$.

Treatment Comparison	Reject in Favor of (Adapt)	(Equal)	% of Samples	Reject in Favor of (Adapt)	(Equal)	% of Samples
	Case 1			Case 3		
Eff: 1v2	6.8%	4.7%	$n_1 > n_2, n_3$	8.9%	5.1%	$n_1 > n_2, n_3$
Eff: 1v3	6.1%	5.1%	31.2%	9.3%	4.5%	1.5%
Eff: 2v3	6.6%	5.4%	$n_2 > n_1, n_3$	5.3%	6.3%	$n_2 > n_1, n_3$
Tox: 1v2	7.1%	4.4%	33.9%	75.3%	78.0%	48.7%
Tox: 1v3	7.2%	5.4%	$n_3 > n_1, n_2$	75.2%	74.3%	$n_3 > n_1, n_2$
Tox: 2v3	7.3%	6.4%	33.8%	4.9%	4.8%	48.8%
	Case 2			Case 4		
Eff 1v2	77.3%	77.8%	$n_1 > n_2, n_3$	32.7%	32.4%	$n_1 > n_2, n_3$
Eff 1v3	75.4%	78.2%	93.2%	74.7%	82.0%	79.7%
Eff 2v3	11.6%	6.3%	$n_2 > n_1, n_3$	36.7%	40.2%	$n_2 > n_1, n_3$
Tox: 1v2	4.3%	5.8%	3.2%	6.0%	6.2%	17.7%
Tox: 1v3	4.4%	6.2%	$n_3 > n_1, n_2$	7.5%	6.7%	$n_3 > n_1, n_2$
Tox: 2v3	11.6%	5.6%	3.3%	12.0%	5.1%	1.9%

- This method assumes objectives are Independent.
- What if dual objectives are Dependent?
  - Motivating Example: Futility is conditional on Efficacy.
  - Is adaptive allocation method affected?
- Simulation Study:
  - Assume $p_1^e = 0.5$ and $p_1^t = 0.1$ in treatment 1, and $p_2^e = 0.3$ and $p_2^t = 0.1$ in treatment 2.
  - Comparisons made to Hypothesized Values for Efficacy $(p_o^e = 0.3)$ and Toxicity $(p_o^t = 0.1)$.
  - Frechet bounds (Chaganty and Joe, 2006): use correlations $-0.2, 0.0, 0.2$ or 0.4.
- As correlation moves away from 0:
  - Slight move toward equal allocation.
  - SDs of weights decrease slightly.
  - Power relatively unaffected.

Simulation Results from 3-Arm Study.

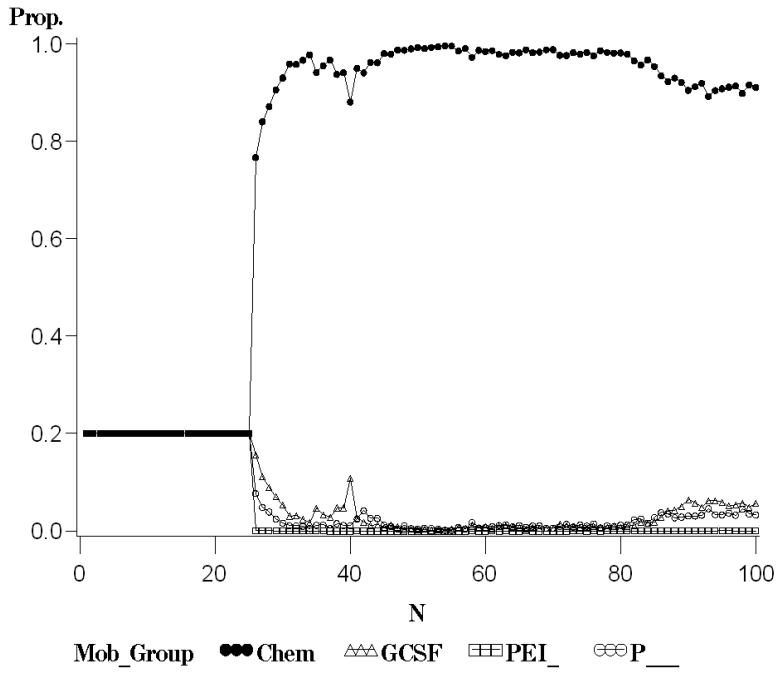
- Average Sample Size (with Standard Deviation) and Decisions in Favor.
- Comparisons made to Hypothesized Values for Efficacy and Toxicity.

	Sample Size	SD	Reject in Favor of Efficacy	Reject in Favor of Toxicity	Sample Size	SD	Reject in Favor of Efficacy	Reject in Favor of Toxicity
			$\rho = -0.2$				$\rho = 0.0$	
Arm 1	110.5	19.3	83.0%	3.2%	111.6	19.8	83.6%	3.3%
Arm 2	89.5	19.3	0.0%	3.0%	88.4	19.8	0.0%	3.9%
			$\rho = 0.2$				$\rho = 0.4$	
Arm 1	111.0	19.4	82.5%	4.2%	107.9	18.7	81.1%	5.5%
Arm 2	89.0	19.4	0.0%	3.4%	92.1	18.7	0.0%	2.4%

- Retrospective analysis of stem cell transplant patients treated at VCU Medical Center $(2003 - 2010)$.
  - Mobilization groups: Plerixafor (AMD), Plerixafor early intervention (PEI), chemotherapy (Chemo) and granulocyte-colony stimulating factor (GCSF).
- Outcome Definitions:
  - Efficacy: patients produce $\geq 5 \times 10^6/kg$ total CD34+ cells.
  - Futility: ≥ 5 days are needed for mobilization.
- Efficacy and futility measurements available in 373 patients:
  - AMD: $n = 19, \hat{p}_e = 0.53, \hat{p}_f = 0.05$
  - PEI: $n = 36, \hat{p}_e = 0.47, \hat{p}_f = 0.11$
  - Chemo: $n = 96, \hat{p}_e = 0.78, \hat{p}_f = 0.10$
  - GCSF: $n = 222, \hat{p}_e = 0.64, \hat{p}_f = 0.33$
- Patient data used as if it were prospectively planned clinical trial.
  - Patients accrued in chronological order.
  - Patient outcomes available in order of mobilization (takes only a few days).

Patterns of adaptive weights for four treatment groups.

- Efficacy compared between treatments.
- Futility compared to hypothesized value (10%).



(i) 4-Group Study

- Adaptive Randomization/Allocation:
  - Can allocate more patients to superior treatment...
  - ... and fewer to inferior treatment.
- Specifically, Bayesian Adaptive Randomization:
  - Relatively simple to implement.
  - Can be designed to have optimal *frequentist* characteristics (e.g. desirable type I and type II error rates).
  - Relatively unaffected by dependent outcomes.
- Note: not all randomized trials should use adaptive allocation.
  - Outcomes not quickly observed (Simon 1977; Berry and Eick 1995).
  - Controlling for covariates can be problematic (Simon 1977).
  - Multi-center trials (Berry and Eick 1995).

- We used $c(n) = n/2N$ in our allocation algorithm (Thall and Wathen, 2007).
  - $c(n) = 0$ at beginning of trial: balanced allocation.
  - $c(n) \rightarrow 1/2$ toward end of trial (Optimal Allocation (Rosenberger *et al.*, 2001)).
  - Automatic "lead-in" at beginning of trial.
  - Keeps allocation weights from changing too quickly.
- We also considered a *decreasingly informative prior* (DIP) method.
  - *Skeptical* prior: *decreasing* function of sample size $\pi(\theta_0, N - n)$.
  - Early trial: close to balanced allocation.
  - As trial progresses: actual evidence outweighs prior skepticism.
- DIP method Behaves similarly to Thall and Wathen (2007) method.
  - Performance can be affected by poor choice of skepticism.

- We used posterior probabilities in our allocation algorithm (Thompson, 1933; Thall and Wathen, 2007).
  - Increases likelihood patients allocated to superior treatment.
  - Benefits (Berry, 2010) and shortcomings (Korn and Freidlin, 2011).
- We could have used predictive probabilities.
  - Account for uncertainty due to unobserved data.
  - Slower adaptation in early portions of trial.
  - Has better power and error rates than posterior probabilities.
  - Allocates slightly fewer patients into superior treatments.

Acknowledgements

- Contributors: Department of Biostatistics, VCU.
  - Ghalib Bello: Graduate Student
  - Lauren Grant: Graduate Student
- Contributors: Bone Marrow Transplant Division, Massey Cancer Center, VCU Medical Center.
  - John M. McCarty, M.D.: Principal Investiagor
  - Amir A. Toor, M.D.: Co-Investigator
  - Cathy Roberts, Ph.D.: Project Coordinator
- This work was supported in part by:
  - Grant MCC-12889 through the Genzyme Corporation
  - Biostatistics Shared Resource of the VCU Massey Cancer Center, funded from NIH-NCI Cancer Center Support Grant P30 CA016059.

References

- Berry, D.A. (2001). Adaptive trials and Bayesian statistics in drug development (with discussion). *Biopharmaceutical Report* 9(2): 1–11.
- Berry, D.A. (2004). Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science* 19(1): 175–187.
- Thompson, W.R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25(3-4): 285–294.
- Bather, J.A. (1981). Randomized allocation of treatments in sequential medical trials. *Journal of the Royal Statistical Society, Series B* 43(3): 265–292.
- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society* 58(5): 527–535.
- Chang, M. (2008). *Adaptive design theory and implementation using SAS and R*. Chapman & Hall/CRC, New York.
- Zelen, M. (1969). Play the winner rule and the controlled clinical trial. *Journal of the American Statistician* 64: 131–146.

References

- Chaganty, N.R., Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika* 93(1): 197–206.
- Berry, D.A., Eick, S.G. (1995). Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine* 14: 231–246.
- Berry, D.A. (2006). Bayesian clinical trials. *Nature* 5: 27–36.
- Simon, R. (1977). Adaptive treatment assignment methods and clinical trials. *Biometrics* 33: 743–749.
- Thall, P.F., Wathen, J.K. (2007). Practical Bayesian adaptive randomization in clinical trials. *European Journal of Cancer* 43(5): 859–866.
- Rosenberger, W.F., Stallard, N., Ivanova, A., Harper, C.N., Ricks, M.L. (2001). Optimal adaptive designs for binary response trials. *Biometrics* 57: 909–913.
- Berry, D.A. (2010). Adaptive clinical trials: the promise and the caution. *Journal of Clinical Oncology* 29(6): 606-609.
- Korn, E.L., Freidlin, B. (2011). Outcome-adaptive randomization: is it useful? *Journal of Clinical Oncology* 29: 771-776.

Thank You

Questions?