Implementing Response-Adaptive Randomization in Multi-Armed Survival Trials

BASS Conference 2009

Alex Sverdlov, Bristol-Myers Squibb

Joint work with Yevgen Tymofyeyev (Merck)

Outline

Response-Adaptive Randomization

Response-adaptive randomization (RAR) procedures use the accruing information in the course of a clinical trial to change the allocation probabilities sequentially with the goal of assigning more patients to the better treatment

- Two treatments: A and B
- n patients enter the trial sequentially and must be randomized to either A or B
- Randomization sequence:

$$\mathbf{T}_n = (T_1, ..., T_n)', \quad T_j = 1, \ if \ A; \ = 0, \ if \ B$$

- Patients' responses: $\mathbf{Y}_n = (Y_1, ..., Y_n)'$
- Statistical model:

$$E(\mathbf{Y}_n) = f(\boldsymbol{\theta}|\mathbf{T}_n)$$

- RAR procedure:

$$\phi_{j+1} = \Pr(T_{j+1} = 1|\mathbf{T}_j, \mathbf{Y}_j), \ j = 1, 2, ..., n-1$$

Response-Adaptive Randomization for Survival Trials

Why implement RAR in survival trials?

- Outcomes are grave, hence it is ethical to assign more patients to a better treatment, if one exists
- It turns out in the survival response case allocations maximizing statistical efficiency are also skewed towards the better treatment
- Hence, more subjects may wish to participate
- Large sample sizes, hence asymptotic results apply

What are potential difficulties?

- Primary outcomes are time-to-event (PFS, OS), thus inherent delays in responses
- Recruitment may terminate before sufficient number of responses accrue to start benefiting from adaptation
- Censored data
- Logistical complexity, regulatory concerns

Survival Trials: Optimal Allocations

Consider a survival trial comparing two treatment arms, A and B

- T_k =survival time, exponential with mean θ_k
- C =censoring time
- $t_k = \min(T_k, C)$ and $\delta_k = 1_{\{t_k = T_k\}}$
- Based on samples of n_A and n_B patients, the m.l.e.'s are

$$\hat{\theta}_k = \frac{\sum_{i=1}^{n_k} t_{ik}}{\sum_{i=1}^{n_k} \delta_{ik}} = \frac{t}{r}, \quad k = A, B$$

- It can be shown that

$$E(\hat{\theta}_k) = \theta_k, \quad \text{var}(\hat{\theta}_k) = \frac{\theta_k^2}{n_k \epsilon_k},$$

where $\epsilon_k = \Pr(T_k \leq C)$ =prob. of death before censoring

- Zhang and Rosenberger (2007, JRSS C) derived an optimal allocation from minimizing a weighted sum of sample sizes subject to constraints on the variance:

$$\begin{aligned} \min \quad & n_A u(\boldsymbol{\theta}) + n_B v(\boldsymbol{\theta}), \\ \text{s.t.} \quad & \frac{\theta_A^2}{n_A \epsilon_A} + \frac{\theta_B^2}{n_B \epsilon_B} \leq V. \end{aligned}$$

- If $u(\boldsymbol{\theta}) = v(\boldsymbol{\theta}) = 1$, then one has $n_A/n_B = \theta_A\sqrt{\epsilon_B}/(\theta_B\sqrt{\epsilon_A})$ - *Neyman allocation* minimizing total sample size of the study

- If $u(\boldsymbol{\theta}) = \theta_A^{-1}$, $v(\boldsymbol{\theta}) = \theta_B^{-1}$, then one has $n_A/n_B = \sqrt{\theta_A^3\epsilon_B}/\sqrt{\theta_B^3\epsilon_A}$ - *"ethical" allocation* minimizing total expected hazard in the study

- Note that optimal allocations depend on $(\theta_A, \theta_B, \epsilon_A, \epsilon_B)$, which must be sequentially estimated, and a RAR can be used to "target" the desired allocations.

Sverdlov and Tymofyeyev (2009) generalized to $K \geq 2$ treatments:

- Consider 2 distinct approaches to optimal allocations:
  - D_A-optimal design
  - Nonlinear programming optimal allocation rules
- Construct sequential RAR procedures to approach optimal allocations in the limit
- Compare the designs in terms of:
  - Variability
  - Imbalance
  - Power
  - Ethical criteria (number of deaths and total hazard)

Optimal Allocations for $K \geq 2$-Treatment Survival Trial

- We are interested in comparing $(K-1)$ experimental treatments to a "control"
- n_k - number of patients on treatment $k = 1, ..., K$, and $\sum_{k=1}^{K} n_k = n$
- A trial has recruitment period of fixed length $R > 0$ months, and the total duration $D > R$ months
- In the kth group one observes $t_k = \min(T_k, C, D - U)$ and $\delta_k = 1_{\{t_k = T_k\}}$, where

$$T_k = \text{survival time, exponential with mean } \theta_k$$
$$C = \text{censoring time, uniform over } (0, D)$$
$$U = \text{patient arrival time, uniform over } (0, R)$$

- Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_K)$.Given the responses (t_{ik}, δ_{ik}), $i = 1, ..., n_k$, $k = 1, ..., K$, we want to test

$$H_0 : \mathbf{A}^T\boldsymbol{\theta} = \mathbf{0} \quad \text{vs.} \quad H_A : \mathbf{A}^T\boldsymbol{\theta} \neq \mathbf{0},$$

where $\mathbf{A}^T$ is a $(K-1) \times K$ matrix of contrasts s.t. $\mathbf{A}^T\boldsymbol{\theta} = (\theta_2 - \theta_1, ..., \theta_K - \theta_1)$.

D_A-Optimal Design

- A D_A-optimal allocation vector $\boldsymbol{\rho}^* = (\rho_1^*, ..., \rho_K^*)$ minimizes $\det\{\mathbf{A}^T\mathbf{M}(\boldsymbol{\rho}, \boldsymbol{\theta})^{-1}\mathbf{A}\}$ over the set of probability distributions on the K-treatment design space.
- $\boldsymbol{\rho}^*$ is found using directional derivatives from the system of equations:

$$d_{\mathbf{A}}(k) = \frac{1}{\rho_k} - \frac{\epsilon_k/\theta_k^2}{\sum_{k=1}^K \rho_k(\epsilon_k/\theta_k^2)} = K - 1, \quad k = 1, ..., K. \tag{1}$$

- Result 1: *Assume that $\theta_1 \geq \theta_2 \geq ... \geq \theta_K$ and ϵ_k is a decreasing function of θ_k for each $k = 1, ..., K$. Then, for the D_A-optimal allocation solving (1), one has $\rho_1^* \geq \rho_2^* \geq ... \geq \rho_K^*$.* In addition, $0 \leq \rho_k^* \leq 1/(K-1)$ for $k = 1, ..., K$.
- Hence, D_A-optimal design is always "ethical" in the sense that it allocates greater proportions of subjects to more efficacious treatments

Optimal Allocations Based on Nonlinear Programming

- Define $\boldsymbol{\theta}_C = (\theta_2 - \theta_1, \theta_3 - \theta_1, ..., \theta_K - \theta_1)$. We are interested in testing

$$H_0 : \boldsymbol{\theta}_C = \mathbf{0} \quad \text{vs.} \quad H_A : \boldsymbol{\theta}_C \neq \mathbf{0}$$

  using Wald's test statistic

$$W_n = \hat{\boldsymbol{\theta}}_C^T \hat{\boldsymbol{\Sigma}}_n^{-1} \hat{\boldsymbol{\theta}}_C$$

- W_n is asymptotically (as $n_k \to \infty$) chi-square with $K-1$ degrees of freedom and noncentrality parameter

$$\phi(n_1, ..., n_K) = \boldsymbol{\theta}_C^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\theta}_C,$$

$$\boldsymbol{\Sigma}_n = \begin{pmatrix} \frac{\theta_2^2}{n_2 \epsilon_2} & 0 & \cdots & 0 \\ 0 & \frac{\theta_3^2}{n_3 \epsilon_3} & \cdots & 0 \\ \cdot & \cdot & \cdots & 0 \\ 0 & 0 & \cdots & \frac{\theta_K^2}{n_K \epsilon_K} \end{pmatrix} + \frac{\theta_1^2}{n_1 \epsilon_1} \mathbf{1}\mathbf{1}^T,$$

Following Tymofyeyev et al. (2007, JASA), we can formulate two optimization problems:

- Problem 1:

$$\begin{cases} \min_{n_1,\ldots,n_k} \sum_{j=1}^{K} w_j n_j, \\ \text{s.t.} \quad n_k \Big/ \sum_{j=1}^{K} n_j \geq B, \ k = 1, ..., K, \\ \phi(n_1, ..., n_K) \geq C, \end{cases}$$

- Problem 2:

$$\begin{cases} \max_{m_1,\ldots,m_k} \phi(m_1, ..., m_K), \\ \text{s.t.} \quad m_k \Big/ \sum_{j=1}^{K} m_j \geq B, \ k = 1, ..., K, \\ \sum_{j=1}^{K} w_j m_j \leq M, \end{cases}$$

where $B \in [0, 1/K]$ is a minimum desired proportion of patients for each treatment group

- It is easily checked that $\phi(\mathbf{n})$ is strictly concave and $\nabla\phi > 0$
- By Theorem 1 of Tymofyeyev et al. (2007), there exist the unique solutions to both Problems 1 and 2.
- Moreover, let $\mathbf{n}^* = (n_1^*, ..., n_K^*)$ be the optimum for Problem 1, and $\mathbf{m}^* = (m_1^*, ..., m_K^*)$ be the optimum for Problem 2. Then one has

$$\frac{n_k^*}{\sum_{j=1}^K n_j^*} = \frac{m_k^*}{\sum_{j=1}^K m_j^*} = \rho_k^*, \quad k = 1, ..., K.$$

Note : Problem 2 is a nonlinear optimization problem with linear constraints, which can be easily solved using optimization software

Choices of the vector of weights $\mathbf{w}$:

- If $\mathbf{w} = (1, ..., 1)$, then we are *maximizing power of Wald's test* for a total sample size n s.t. the proportion of patients in each treatment group is at least B.
- If $\mathbf{w} = (\theta_1^{-1}, ..., \theta_K^{-1})$, then we are *minimizing the total expected hazard* in the trial s.t. the constraints on B and $\phi(\mathbf{n})$
- If $\mathbf{w} = (\epsilon_1, ..., \epsilon_K)$, then we are *minimizing the expected number of deaths* in the trial s.t. the constraints on B and $\phi(\mathbf{n})$

Note : For $\mathbf{w} = (1, ..., 1)$, we have the analytical form of the optimal solution. It will be referred to as *NP-optimal allocation*

Allocation Surfaces

Consider $K = 3$ treatments, $R = 55$ and $D = 96$, $5 \leq \theta_A \leq 35$, $5 \leq \theta_B \leq 35$, $\theta_C = 17$.

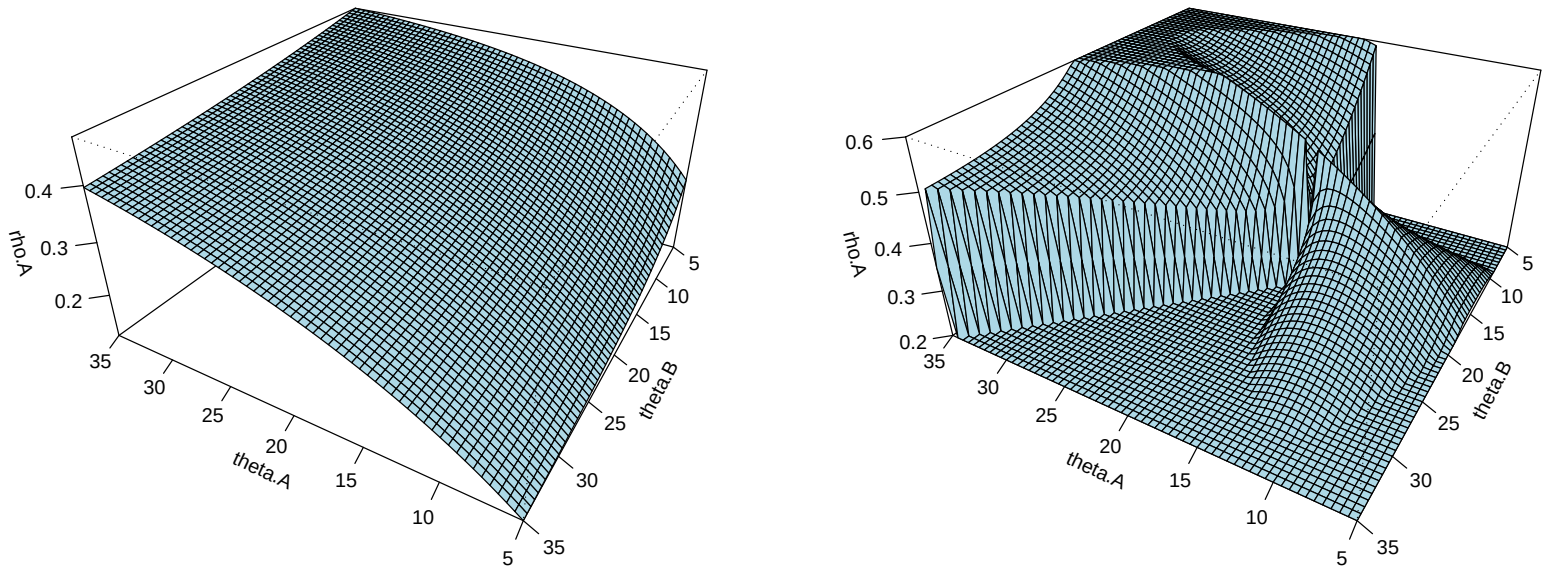


Figure: Allocation surfaces $\rho_A(\theta_A, \theta_B, 17)$ for the D_A-optimal allocation (left plot) and the NP-optimal allocation with $B = 0.2$ (right plot)

Operating characteristics of the designs:

- D_A-efficiency: Given a D_A-optimal design ξ^*, the D_A-eff. of any other design ξ is defined as

$$E(\xi) = \left\{ \frac{|\mathbf{A}^T\mathbf{M}^{-1}(\xi^*)\mathbf{A}|}{|\mathbf{A}^T\mathbf{M}^{-1}(\xi)\mathbf{A}|} \right\}^{1/(K-1)}$$

  A value of $E(\xi) = 0.95$ means that design ξ is 95% as efficient as ξ^*.

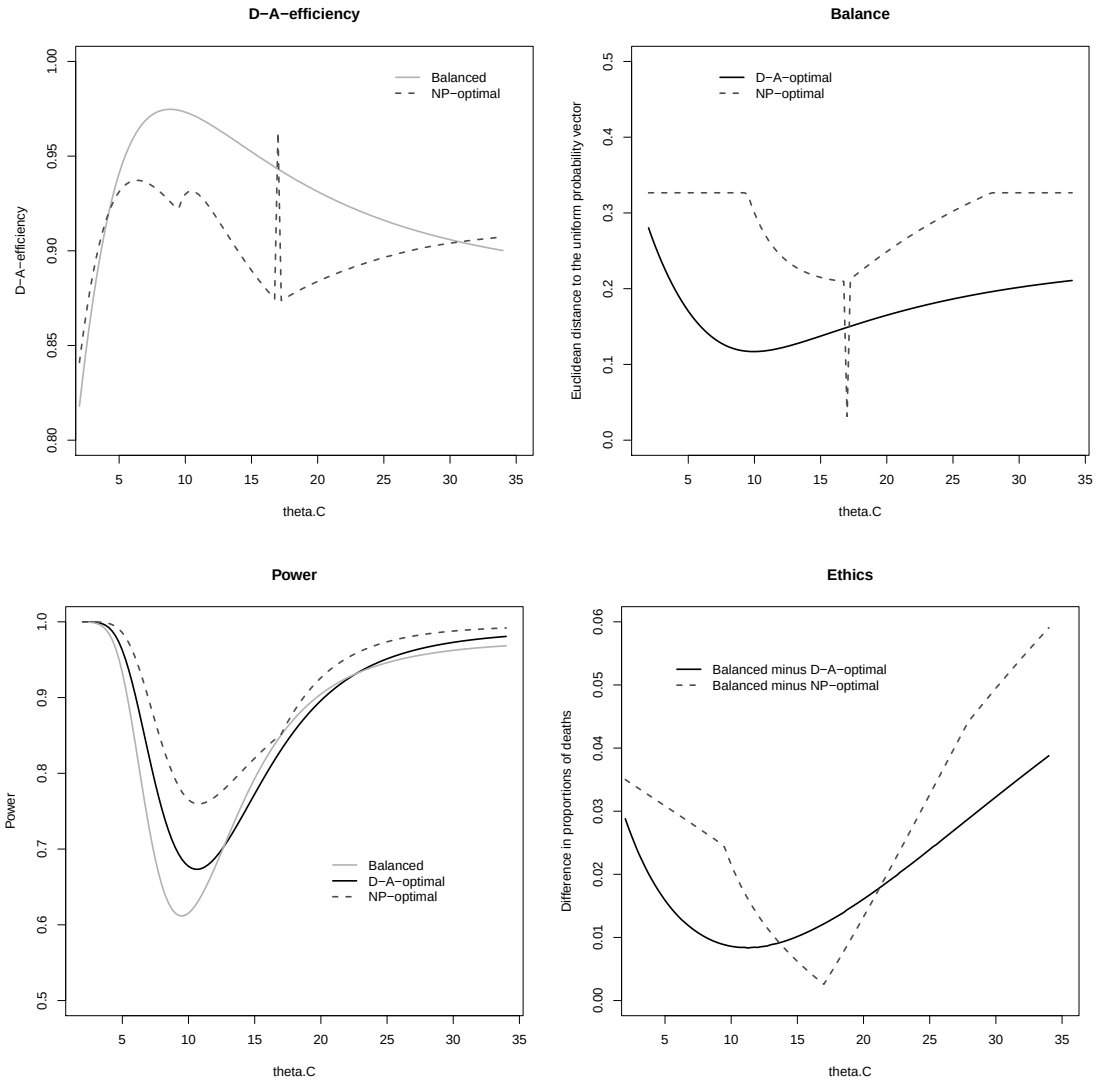
- Balance: Euclidean distance between a vector of allocation proportions and the vector of uniform probabilities $(1/K, ..., 1/K)$.

- Power of Wald's test

- Difference in proportions of deaths $\sum_{k=1}^{K} \rho_k \epsilon_k$ between the balanced allocation and an optimal allocation

Theoretical Comparison of Design Characteristics

Consider $K = 3$ treatments, $R = 55$ and $D = 96$, $\theta_A = 8$, $\theta_B = 17$, $2 \leq \theta_C \leq 34$.



Implementing Response-Adaptive Randomization to "Target" Optimal Allocations

- Optimal allocations depend on unknown model parameters
- Given data from the first $(j-1)$ patients: $\mathbf{T}_{j-1}$ and $\mathbf{Y}_{j-1}$, compute

$$\hat{\boldsymbol{\rho}}(j-1) = (\hat{\rho}_1(j-1), ..., \hat{\rho}_K(j-1)) \quad \text{estimate of the target allocation}$$
$$N_i/(j-1), \quad i = 1, ..., K \quad \text{current treatment proportions}$$

- Randomize the jth patient to treatment k with probability (Doubly-Adaptive Biased Coin (DBCD), Hu and Zhang (2004), Ann. Statist.)

$$\psi_{jk} = \frac{\hat{\rho}_k(j-1)\left(\frac{\hat{\rho}_k(j-1)}{N_k/(j-1)}\right)^{\gamma}}{\sum_{i=1}^{K}\hat{\rho}_i(j-1)\left(\frac{\hat{\rho}_i(j-1)}{N_i/(j-1)}\right)^{\gamma}}, \quad k = 1, ..., K,$$

where $\gamma \geq 0$ is a parameter controlling the degree of randomness of an allocation procedure.

Role of γ:

- $\gamma = 0$ (the highest variability). One has sequential maximum likelihood procedure

$$\psi_{jk} = \hat{\rho}_k(j-1), \quad k = 1, ..., K.$$

- $\gamma = \infty$ (the smallest variability). The procedure is almost deterministic:

$$\psi_{jk} = 1, \quad \text{if treatment } k \text{ has maximum value of } \frac{\hat{\rho}_i(j-1)}{N_i/(j-1)}, \ i = 1, ..., K,$$

$$= 1/s, \quad \text{if } s \text{ treatments are ties in terms of } \frac{\hat{\rho}_i(j-1)}{N_i/(j-1)}, \ i = 1, ..., K,$$

$$= 0, \quad \text{otherwise}$$

- $\gamma = 2$ is recommended for use in practice (Rosenberger and Hu, 2004)

Asymptotic properties of the DBCD procedure:

- Hu and Zhang (2004) (assuming that responses are immediate) showed that if the allocation vector $\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{\theta})$ is continuous $\forall \boldsymbol{\theta}$, and $\boldsymbol{\rho}$ is twice cont.-diff. in a neighborhood of the true $\boldsymbol{\theta}^*$, then as $n \to \infty$

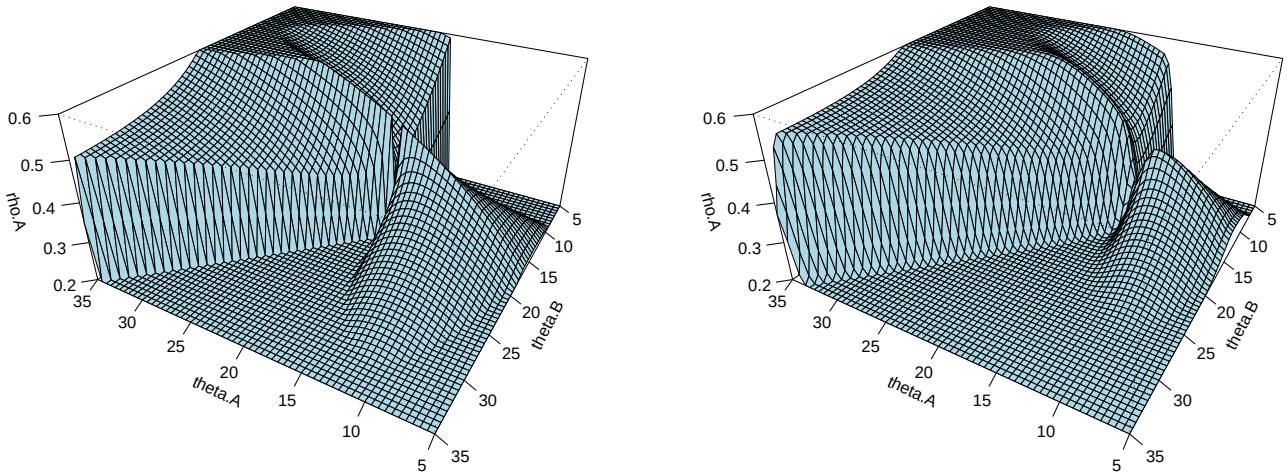
$$\mathbf{N}(n)/n \to \boldsymbol{\rho}^* \quad \text{a.s.}$$
$$\sqrt{n}(\mathbf{N}(n)/n - \boldsymbol{\rho}^*) \to \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{in distribution,}$$

  where $\boldsymbol{\rho}^* = \boldsymbol{\rho}(\boldsymbol{\theta}^*)$ and $\boldsymbol{\Sigma}$ is a known expression.

- Hu et al. (2008) justified the above asymptotic properties of the DBCD procedure for cases when responses are moderately delayed (e.g. exponential delays)

In our case:

- $\boldsymbol{\rho}$ for D_A-optimal allocation is nice and smooth, and hence asymptotic results of Hu and Zhang (2004) apply
- $\boldsymbol{\rho}$ for NP-optimal allocation is discontinuous for certain values of $\boldsymbol{\theta}$. We "smooth" $\boldsymbol{\rho}$ using a standard K-variate Gaussian kernel as follows:



and target the "smoothed" NP-optimal allocation using the DBCD procedure

Simulation Study

- 3 procedures:
  - Complete randomization (CRD)
  - DBCD with $\gamma = 2$ targeting D_A-optimal allocation (D_A)
  - DBCD with $\gamma = 2$ targeting "smoothed" NP-optimal allocation (NP)

- $K = 3$ treatments, $R = 55$ months, $D = 96$ months

- Two choices of θ_A ("control" treatment):
  - $\theta_A = 8.5$ ("poor" survival); $\epsilon_A = 0.91$
  - $\theta_A = 24$ (2-year survival); $\epsilon_A = 0.74$

- Sample size n is chosen s.t. CRD has 90% power (when $\theta_A = 8.5$), or 80% power (when $\theta_A = 24$) under a given alternative

- 10000 simulations of a trial with n patients for each experimental scenario using R

Table: Theoretical optimal designs for D_A-optimal allocation (D_A), and a Gaussian-smoothed nonlinear programming optimal allocation (NP) with $B = 0.2$

Scenario	n	$(\theta_A, \theta_B, \theta_C)$	D_A	NP
Ia	136	(8.5, 17, 17)	(.22, .39, .39)	(.32, .34, .34)
IIa	162	(8.5, 8.5, 17)	(.29, .29, .43)	(.20, .20, .60)
IIIa	84	(8.5, 25, 17)	(.19, .44, .37)	(.23, .57, .20)
IVa	136	(8.5, 8.5, 8.5)	(1/3, 1/3, 1/3)	(1/3, 1/3, 1/3)
Ib	516	(24, 34, 34)	(.26, .37, .37)	(.40, .30, .30)
IIb	567	(24, 24, 34)	(.30, .30, .40)	(.20, .20, .60)
IIIb	213	(24, 48, 34)	(.24, .42, .34)	(.26, .54, .20)
IVb	516	(24, 24, 24)	(1/3, 1/3, 1/3)	(1/3, 1/3, 1/3)

Table: Simulated allocation proportions $\mathbf{N}/n = (N_A/n, N_B/n, N_C/n)$ and their standard deviations (S.D.)

Scenario (% resp.)		CRD	D_A	NP
Ia (77%)	$\mathbf{N}/n$	(.33, .33, .34)	(.28, .36, .36)	(.29, .36, .35)
	S.D.	(.04, .04, .04)	(.04, .04, .04)	(.04, .09, .09)
IIa (81%)	$\mathbf{N}/n$	(.33, .33, .34)	(.31, .31, .38)	(.27, .26, .47)
	S.D.	(.04, .04, .04)	(.03, .03, .03)	(.05, .05, .07)
IIIa (74%)	$\mathbf{N}/n$	(.33, .33, .34)	(.28, .37, .35)	(.27, .41, .32)
	S.D.	(.05, .05, .05)	(.05, .05, .05)	(.05, .10, .09)
IVa (85%)	$\mathbf{N}/n$	(.33, .33, .34)	(.33, .33, .34)	(.33, .33, .34)
	S.D.	(.04, .04, .04)	(.04, .04, .04)	(.07, .07, .07)
Ib (63%)	$\mathbf{N}/n$	(.33, .33, .34)	(.31, .35, .35)	(.32, .34, .33)
	S.D.	(.02, .02, .02)	(.02, .02, .02)	(.04, .06, .06)
IIb (65%)	$\mathbf{N}/n$	(.33, .33, .34)	(.32, .32, .36)	(.31, .30, .39)
	S.D.	(.02, .02, .02)	(.02, .02, .02)	(.05, .05, .06)
IIIb (60%)	$\mathbf{N}/n$	(.33, .33, .34)	(.30, .36, .34)	(.29, .40, .31)
	S.D.	(.03, .03, .03)	(.03, .03, .03)	(.04, .08, .07)
IVb (63%)	$\mathbf{N}/n$	(.33, .33, .34)	(.33, .33, .34)	(.33, .33, .34)
	S.D.	(.02, .02, .02)	(.02, .02, .02)	(.05, .05, .05)

Table: Power and error rates

Scenario	n	CRD	D_A	NP-1
Ia	136	0.902	0.891	0.891
IIa	162	0.902	0.919	0.942
IIIa	84	0.897	0.905	0.903
IVa	136	0.044	0.054	0.051
Ib	516	0.805	0.799	0.796
IIb	567	0.801	0.821	0.820
IIIb	213	0.801	0.816	0.823
IVb	516	0.049	0.048	0.050

Table: Total number of deaths in the study (S.D.)

Scenario	CRD	D_A	NP-1
Ia	115 (4)	115 (4)	115 (4)
IIa	142 (4)	142 (4)	140 (4)
IIIa	69 (4)	68 (4)	68 (4)
IVa	124 (3)	124 (3)	124 (3)
Ib	349 (11)	348 (11)	349 (11)
IIb	402 (11)	400 (11)	398 (11)
IIIb	137 (7)	135 (7)	134 (7)
IVb	382 (10)	382 (10)	382 (10)

Table: Total hazard (in patients per months) in the study (S.D.)

Scenario	CRD	D_A	NP-1
Ia	10.9 (1.1)	10.5 (1.0)	10.5 (1.1)
IIa	16.1 (1.5)	15.7 (1.4)	14.9 (1.4)
IIIa	6.3 (0.9)	5.9 (0.8)	5.8 (0.8)
IVa	16.4 (1.5)	16.4 (1.5)	16.4 (1.5)
Ib	17.4 (0.9)	17.2 (0.9)	17.3 (1.0)
IIb	21.4 (1.1)	21.2 (1.1)	21.0 (1.1)
IIIb	6.6 (0.6)	6.5 (0.6)	6.4 (0.6)
IVb	21.6 (1.1)	21.6 (1.1)	21.6 (1.1)

Redesigning a Phase III Survival Trial

- A randomized phase III clinical trial to compare the 3-yr survival rates of patients with locally advanced head and neck cancer (HNC) treated with standard fractionated RT alone (arm A) or RT+cisplatin (arm B) or RT+carboplatin (arm C) (Fountzilas et al. (2004) *Medical Oncology* 21(2), 95-107)

- From Jan-1995 to Jul-1999, $n = 124$ patients with proven locally advanced HNC were equally randomized to treatments using stratified blocks (randomization was centralized)

	A($n = 41$)	B($n = 45$)	C($n = 38$)
Died	36	23	23
Censored(%)	5 (0.12)	22 (0.49)	15 (0.39)
Survival rate at 3 yr (%)	17.5	52.0	42.0
Survival rate at 5 yr (%)	9.0	52.0	38.0
Median TTP (months)	6.3	45.2	17.7
OS (months)	12.2	48.6	24.5

- Given this data, assume that ITT survival times are exponentially distributed with means $\theta_A = 8.5$, $\theta_B = 34$, and $\theta_C = 17$
- Trial duration $D = 96$ months, recruitment period $R = 55$ months
- Theoretical D_A-optimal design is $\boldsymbol{\rho}^*(\boldsymbol{\theta}) = (0.18, 0.46, 0.36)$
- Theoretical "smoothed" NP-optimal design is $\boldsymbol{\rho}^*(\boldsymbol{\theta}) = (0.20, 0.60, 0.20)$
- Simulate 10000 trials with $n = 124$ patients and 3 randomization procedures

	CRD	D_A	NP
		$n = 124$	
$\mathbf{N}(n)/n$	(.33, .33, .33)	(.27, .39, .34)	(.26, .46, .28)
S.D.($\mathbf{N}(n)/n$)	(.04, .04, .04)	(.04, .04, .04)	(.04, .09, .08)
$D(n)$ (S.D.)	98 (5)	96 (5)	94 (5)
$H(n)$ (S.D.)	8.7 (1.0)	8.0 (0.9)	7.7 (1.0)
Power	> 0.99	> 0.99	> 0.99
		$n = 66$	
$\mathbf{N}(n)/n$	(.33, .33, .33)	(.28, .38, .34)	(.26, .43, .31)
S.D.($\mathbf{N}(n)/n$)	(.06, .06, .06)	(.05, .05, .05)	(.05, .10, .09)
$D(n)$ (S.D.)	52 (3)	51 (3)	50.5 (4)
$H(n)$ (S.D.)	4.8 (0.8)	4.4 (0.7)	4.3 (0.7)
Power	.900	.920	.932
		$n = 63$	
$\mathbf{N}(n)/n$	(.33, .33, .33)	(.28, .38, .34)	(.26, .43, .31)
S.D.($\mathbf{N}(n)/n$)	(.06, .06, .06)	(.06, .06, .05)	(.05, .10, .09)
$D(n)$ (S.D.)	50 (3)	49 (3)	48 (3)
$H(n)$ (S.D.)	4.5 (0.8)	4.2 (0.7)	4.1 (0.7)
Power	.880	.900	.917

Conclusions

- The DBCD procedures targeting optimal allocations assign greater proportions of patients to more efficacious treatments
- On average, these procedures result in $1-4$ fewer deaths, and smaller total hazard than the balanced design
- Also, in most of the cases, for a given sample size these procedures are $1\% - 3\%$ more powerful than the balanced design
- Under H_0, all procedures reduce to the balanced allocation

Overall conclusion: The DBCD procedures can be good alternatives to the balanced designs in clinical trials with grave outcomes, such as in survival trials. The total sample size can be reduced without sacrificing power, which implies extra savings in the study cost, reduction of the risk of exposing subjects to less efficient therapies, and reduction of the total number of deaths in the trial.

Possible Extensions

- Optimal allocations for other response distributions, such as Weibull, or lognormal
- Incorporating covariates and developing covariate-adjusted response-adaptive (CARA) randomization procedures
- Bivariate response (efficacy + toxicity) in application to dose-finding (phase II) survival trials

Thank You for Your Attention!

Literature:

Founzilas, GF, et al (2004) Concomitant radiochemotherapy vs radiotherapy alone in patients with head and neck cancer. A Hellenic Cooperative Oncology Group phase III study , *Medical Oncology* **21(2)**, 95-107.

Hu, F and Zhang, L-X (2004) Asymptotic properties of doubly adaptive biased coin designs for multicenter clinical trials, *Annals of Statistics* **32**, 268-301.

Hu, F. and Zhang, L.-X., Cheung, S.H. and Chan, W.S. (2008) Doubly adaptive biased coin designs with delayed responses. *The Canadian Journal of Statistics* **36 (4)**, 541-559.

Hu, F and Rosenberger, WF (2006) *The Theory of Response-Adaptive Randomization in Clinical Trials,* New York, Wiley.

Rosenberger, WF and Lachin, J (2002) *Randomization in Clinical Trials. Theory and Practice,* New York, Wiley.

Rosenberger, WF and Seshaiyer, P (1997) Adaptive survival trials, *Journal of Biopharmaceutical Statistics* **7(4)**, 617-624.

Sverdlov, A and Tymofyeyev, Y (2009) Implementing response-adaptive randomization in multi-armed survival trials, Working paper.

Zhang, L and Rosenberger, WF (2007) Response-adaptive randomization for survival trials: the parametric approach, *Applied Statistics* **56(2)**, 153-165.