Adaptive designs to maximize power in clinical trials with multiple treatments

David Azriel, Paul D. Feigin

Technion - Israel Institute of Technology

January 17, 2013

The problem

- A, B, C are three treatments with unknown probabilities of success, p_A, p_B, p_C.
- A - placebo, B - standard treatment, C - new treatment.
- An n subjects experiment (binary response) is preformed in order to test the new therapy.
- **Target: to find a sequential design that maximizes the power of the relevant tests** keeping in mind the ethical goal of assigning a small fraction of subjects to the inferior treatment.

The problem (cont.)

- Let $\nu_A(n)$, $\nu_B(n)$, $\nu_C(n)$ be the fraction of subjects (*allocations*) assigned to treatments A, B, or C.
- Two steps:
  1. **Oracle**: to find the optimal allocation (asymptotically) **for a fixed design**.
  2. As the optimal allocation depends on unknown parameters, implementation requires a sequential design.

Preliminaries

At the end of the experiment, three one-sided hypotheses may be tested:

$$H_0 : \ p_A = p_B \text{ versus } H_1 : \ p_A < p_B \tag{1}$$
$$H_0 : \ p_A = p_C \text{ versus } H_1 : \ p_A < p_C \tag{2}$$
$$H_0 : \ p_B = p_C \text{ versus } H_1 : \ p_B < p_C. \tag{3}$$

We assume that $p_A < p_B < p_C$; two criteria of optimality are studied:

- **Correct selection:** to maximize the power (=probability that both are rejected) of (2) and (3) .
- **Correct order:** to maximize the power of (1) and (3).

Preliminaries (cont.)

At stage n, the estimators $\hat{p}_i(n)$, for $i = A, B, C$, are available.

$$\hat{p}_i(n) \stackrel{(d)}{=} \frac{Binomial(n\nu_i(n), p_i)}{n\nu_i(n)}.$$

For example, we reject the null in (1) if

$$\frac{\hat{p}_B(n) - \hat{p}_A(n)}{\hat{V}^{1/2}} \geq C_r;$$

$\hat{V}$ = an estimator of the variance; $C_r \geq 0$ = a critical value.
The power of correct selection is

$$P_{CS}(n) := P\left[\left\{\frac{\hat{p}_C(n) - \hat{p}_A(n)}{\hat{V}^{1/2}} \geq C_r\right\} \bigcap \left\{\frac{\hat{p}_C(n) - \hat{p}_B(n)}{\hat{V}^{1/2}} \geq C_r\right\}\right]$$

The power of correct order is

$$P_{CO}(n) := P\left[\left\{\frac{\hat{p}_B(n) - \hat{p}_A(n)}{\hat{V}^{1/2}} \geq C_r\right\} \bigcap \left\{\frac{\hat{p}_C(n) - \hat{p}_B(n)}{\hat{V}^{1/2}} \geq C_r\right\}\right].$$

Step 1 - optimal allocation for a fixed design

- Two different notions of optimality, related to Pitman and Bahadur efficiency, can be considered.
- In this case, Pitman's approach does not work (details are omitted).
- Bahadur's approach:
  - $P_{CS}(n) \approx 1 - e^{-C_1(\nu)n}$, $P_{CO}(n) \approx 1 - e^{-C_2(\nu)n}$.
  - We aim at finding the allocation that maximizes $C_1(\nu)$, $C_2(\nu)$.

Two treatments; Azriel, Mandel and Rinott (2012)

We would like to find allocations that maximize

$$P\left\{\frac{\hat{p}_B(n)-\hat{p}_A(n)}{\hat{V}^{1/2}} \geq C_r\right\}.$$

The standard argument is based on the following normal approximation:

$$P_{p_A,p_B}\left\{\frac{\hat{p}_B(n)-\hat{p}_A(n)}{\hat{V}^{1/2}} \geq C_r\right\} =$$

$$P_{p_A,p_B}\left\{\frac{\hat{p}_B(n)-\hat{p}_A(n)-(p_B-p_A)}{V^{1/2}} > \frac{C_r\hat{V}^{1/2}-(p_B-p_A)}{V^{1/2}}\right\}$$

$$\approx 1-\Phi\left(\frac{C_r\cdot\hat{V}^{1/2}-(p_B-p_A)}{V^{1/2}}\right)$$

$$\approx 1-\Phi\left(C_r-\frac{p_B-p_A}{V^{1/2}}\right)=1-\Phi\left(C_r-\frac{n^{1/2}(p_B-p_A)}{(nV)^{1/2}}\right).$$

The normal approximation is not valid because $V \approx C/n$. The error of the Normal approximation is in the order of $1/\sqrt{n}$ while the approximation itself is in the order of e^{-n}.

Two treatments; Neyman allocation

Neyman allocation, i.e., $\nu_{\text{Neyman}} = \frac{\sqrt{p_A(1-p_A)}}{\sqrt{p_A(1-p_A)}+\sqrt{p_B(1-p_B)}}$ is recommended by many authors, since it minimizes $nV = \frac{p_A(1-p_A)}{\nu} + \frac{p_B(1-p_B)}{1-\nu}$.

(e.g., Brittain and Schlesselman (1982); Rosenberger et. al (2001); Hu and Rosenberger (2003); Bandyopadhyay and Bhattacharya (2006); Hu et. al (2006); Hu and Rosenberger (2006); Tymofyeyev et. al (2007); Biswas et. al (2010); Zhu and Hu (2010); Chambaz and van der Laan (2010)).

For example, Hu et. al (2006) state that “If one wishes to maximize the power of the usual test comparing two binomial probabilities, it is well-known that Neyman allocation, the ratio of the standard deviations, should be used".

Two treatments (cont.)

Theorem

Assume that $p_B > p_A$ *and* $\lim_n \nu_A(n) = \nu, \lim_n \nu_B(n) = 1 - \nu$ *for certain* $0 < \nu < 1$*; then for any* $C_r \geq 0$

$$\lim_n \frac{1}{n} \log \left\{ 1 - P\left(\frac{\hat{p}_B(n) - \hat{p}_A(n)}{\hat{V}^{1/2}} \geq C_r \right) \right\} = g(p_A, p_B, \nu),$$

where $g(p_A, p_B, \nu) := \log\{(1 - p_B)^{\nu}(1 - p_A)^{1-\nu} + p_B^{\nu} p_A^{1-\nu}\}$.

Let $\nu^* = \nu^*(p_A, p_B) := \arg\min_{\nu} g(\nu, p_A, p_B)$, ($g(\cdot, p_A, p_B)$ is convex).

Two treatments; numerical illustration.

Table: The optimal allocation ν^* for different parameters compared to Neyman allocation.

p_A	p_B	ν^*	ν_{Neyman}
0.5	0.8	0.518	0.556
0.5	0.65	0.504	0.512
0.6	0.75	0.510	0.531
0.7	0.75	0.505	0.514
0.7	0.85	0.521	0.562
0.7	0.9	0.535	0.604
0.85	0.95	0.541	0.621
0.5	0.9	0.542	0.625

Conclusion: For two treatments the optimal allocation is quite close to 0.5 and therefore adaptive design cannot significantly improve power in that case.
We shall see that for three treatments the situation is different.

Large deviations result - three treatments

Theorem

Assume that $\lim_n \nu_A(n) = \nu_A, \lim_n \nu_B(n) = \nu_B$ *and* $\lim_n \nu_C(n) = \nu_C$ *for certain* $\nu_A, \nu_B, \nu_C > 0$*; then for any* $C_r \geq 0$

$$\lim_n \frac{1}{n} \log\{1 - P_{CS}(n)\} = H_{CS}(\nu_A, \nu_B, \nu_C),$$

and,

$$\lim_n \frac{1}{n} \log\{1 - P_{CO}(n)\} = H_{CO}(\nu_A, \nu_B, \nu_C).$$

Definitions

Where

$$g(p_1, p_2, \nu) := \log\{(1-p_2)^{\nu}(1-p_1)^{1-\nu} + p_2^{\nu} p_1^{1-\nu}\}.$$

$$H_{CS}(\nu_A, \nu_B, \nu_C) = \max\{(\nu_B + \nu_C) g(p_B, p_C, \frac{\nu_B}{\nu_B + \nu_C}),$$
$$(\nu_A + \nu_C) g(p_A, p_C, \frac{\nu_A}{\nu_A + \nu_C})\}.$$

$$H_{CO}(\nu_A, \nu_B, \nu_C) = \max\{(\nu_B + \nu_C) g(p_B, p_C, \frac{\nu_B}{\nu_B + \nu_C}),$$
$$(\nu_A + \nu_B) g(p_A, p_B, \frac{\nu_A}{\nu_A + \nu_B})\}$$

Properties of the optimal allocation

We now focus on $P_{CS}(n)$. Consider the case $C_r = 0$.

$$P_{CS}(n) = P\left[\{\hat{p}_C(n) \geq \hat{p}_B(n)\} \cap \{\hat{p}_C(n) \geq \hat{p}_A(n)\}\right];$$

$$1 - P_{CS}(n) = P[\{\hat{p}_B(n) > \hat{p}_C(n)\} \cup \{\hat{p}_A(n) > \hat{p}_C(n)\}]$$

$$= P\{\hat{p}_B(n) > \hat{p}_C(n)\} + P\{\hat{p}_A(n) > \hat{p}_C(n)\} - P[\{\hat{p}_B(n) > \hat{p}_C(n)\} \cap \{\hat{p}_A(n) > \hat{p}_C(n)\}].$$

The large deviations rate is determined by the maximum rate of the first two probabilities in the latter expression; this explains the form of H_{CS}, which is based on the result in two treatments.

Properties of the optimal allocation (cont.)

The optimal allocation is $\arg\min H_{CS}(\nu_A, \nu_B, \nu_C)$;

$$\nu^{CS} := \arg \min_{\nu_A,\nu_B,\nu_C} \max\{(\nu_B + \nu_C) g(p_B, p_C, \frac{\nu_B}{\nu_B + \nu_C}), \\ (\nu_A + \nu_C) g(p_A, p_C, \frac{\nu_A}{\nu_A + \nu_C})\}.$$

- By "minimax argument", under the optimal allocation the two terms in the curly brackets are equal.
- The same "effort" is made to distinguish between p_A and p_C and between p_B and p_C.
- The latter is harder to distinguish than the former and therefore ν_A^{CS} is smaller than ν_B^{CS}, ν_C^{CS}.

Numerical examples

Table: Comparison of optimal allocations.

p_A	p_B	p_C	ν_A^{CS}	ν_B^{CS}	ν_C^{CS}	ν_A^{CO}	ν_B^{CO}	ν_C^{CO}
0.1	0.2	0.5	0.119	0.415	0.466	0.449	0.497	0.054
0.1	0.6	0.9	0.039	0.519	0.442	0.100	0.492	0.408
0.2	0.7	0.8	0.008	0.507	0.485	0.012	0.505	0.483
0.3	0.5	0.7	0.066	0.471	0.462	0.290	0.421	0.290
0.3	0.5	0.8	0.105	0.459	0.436	0.433	0.460	0.107
0.4	0.5	0.7	0.128	0.433	0.439	0.466	0.472	0.062
0.5	0.6	0.7	0.071	0.467	0.462	0.315	0.416	0.269

Step 2 - adaptive design

The optimal allocations ν^{CS} and ν^{CO} depend on p_A, p_B, p_C, which are unknown, and in order to implement them an adaptive rule is needed.
Adaptive design:

- $x_n \in \{A, B, C\}$ is the treatment assigned to the n^{th} subjects and y_n denotes its binary response.
- $x_n \in \mathcal{F}_{n-1} := \sigma\{(x_1, y_1), (x_2, y_2), \ldots, (x_{n-1}, y_{n-1})\}$.
- We assume that $y_n | \mathcal{F}_{n-1} \sim \text{Bernoulli}(p_{x_n})$.

Adaptive design - RCS

Consider the function

$$\nu^{CS}(p_A, p_B, p_C) := \left(\nu_A^{CS}(p_A, p_B, p_C), \nu_B^{CS}(p_A, p_B, p_C), \nu_C^{CS}(p_A, p_B, p_C)\right).$$

At each stage n of the experiment the next subject is allocated according to

$$x_{n+1} = \begin{cases} A & \text{w. p.} \quad \hat{\nu}_A(n+1) := \nu_A^{CS}(\hat{p}_A(n), \hat{p}_B(n), \hat{p}_C(n)) \\ B & \text{w. p.} \quad \hat{\nu}_B(n+1) := \nu_B^{CS}(\hat{p}_A(n), \hat{p}_B(n), \hat{p}_C(n)) \\ C & \text{w. p.} \quad \hat{\nu}_C(n+1) := \nu_C^{CS}(\hat{p}_A(n), \hat{p}_B(n), \hat{p}_C(n)), \end{cases}$$

under some truncation. We call this design RCS - random correct selection.

Properties of the RCS design

The properties of the RCS design are stated in the following theorem.

Theorem

The RCS design satisfies

I. $\hat{p}_A(n), \hat{p}_B(n), \hat{p}_C(n)$ *are strongly consistent, asymptotically normal and asymptotically independent;*

II. $\left(\hat{\nu}_A(n), \hat{\nu}_B(n), \hat{\nu}_C(n)\right) \overset{n\to\infty}{\longrightarrow} \nu^{CS}(p_A, p_B, p_C)$ *almost surely;*

III. $\left(\nu_A(n), \nu_B(n), \nu_C(n)\right) \overset{n\to\infty}{\longrightarrow} \nu^{CS}(p_A, p_B, p_C)$ *almost surely.*

The large deviation rate of the RCS

Consider the following randomized design

- Let $\pi_A, \pi_B, \pi_C > 0$ be such that $\pi_A + \pi_B + \pi_C = 1$.
- At stage n of the experiment we choose $x_n = i$ with probability π_i for $i = A, B, C$.

Theorem

For this design we have for any $C_r \geq 0$

$$\lim_n \frac{1}{n} \log\{1 - P_{CS}(n)\} = \max\{R(p_A, p_C, \pi_A, \pi_C), R(p_B, p_C, \pi_B, \pi_C)\},$$

where

$$R(p_1,p_2,\pi_1,\pi_2) := \max_{K \in [p_1,p_2]} \log\{f(K,p_1,\pi_1) + f(K,p_2,\pi_2) + 1 - (\pi_1 + \pi_2)\},$$

$$f(K,p,\pi) := \pi \left(\frac{p}{K}\right)^K \left(\frac{1-p}{1-K}\right)^{1-K}.$$

Comment

It can be shown that

$$R(p_1, p_2, \pi_1, \pi_1) \geq (\pi_1 + \pi_2) g(p_1, p_2, \frac{\pi_1}{\pi_1 + \pi_2}).$$

- The R term is minus the rate in a randomized design and the g term is that of a fixed design.
- Thus, the probability of an error decreases faster to zero in a fixed design than in a randomized design.
- Designs such as the biased coin of Efron that decrease the randomness may also cause increase of power.

Comment (cont.)

Table: Comparison of rates between the optimal fixed design (optimal) and the randomized design with optimal proportions (rand).

			correct selection		correct order	
p_A	p_B	p_C	optimal	rand	optimal	rand
0.1	0.2	0.5	0.0464	0.0419	0.0096	0.0089
0.1	0.6	0.9	0.0652	0.0325	0.0610	0.0501
0.2	0.7	0.8	0.0067	0.0046	0.0067	0.0051
0.3	0.5	0.7	0.0199	0.0178	0.0147	0.0146
0.3	0.5	0.8	0.0472	0.0407	0.0190	0.0181
0.4	0.5	0.7	0.0186	0.0179	0.0048	0.0046
0.5	0.6	0.7	0.0051	0.0050	0.0036	0.0036

The large deviation rate of the RCS (cont.)

The following theorem implies that the RCS design is as good as the randomized design with the optimal proportions.

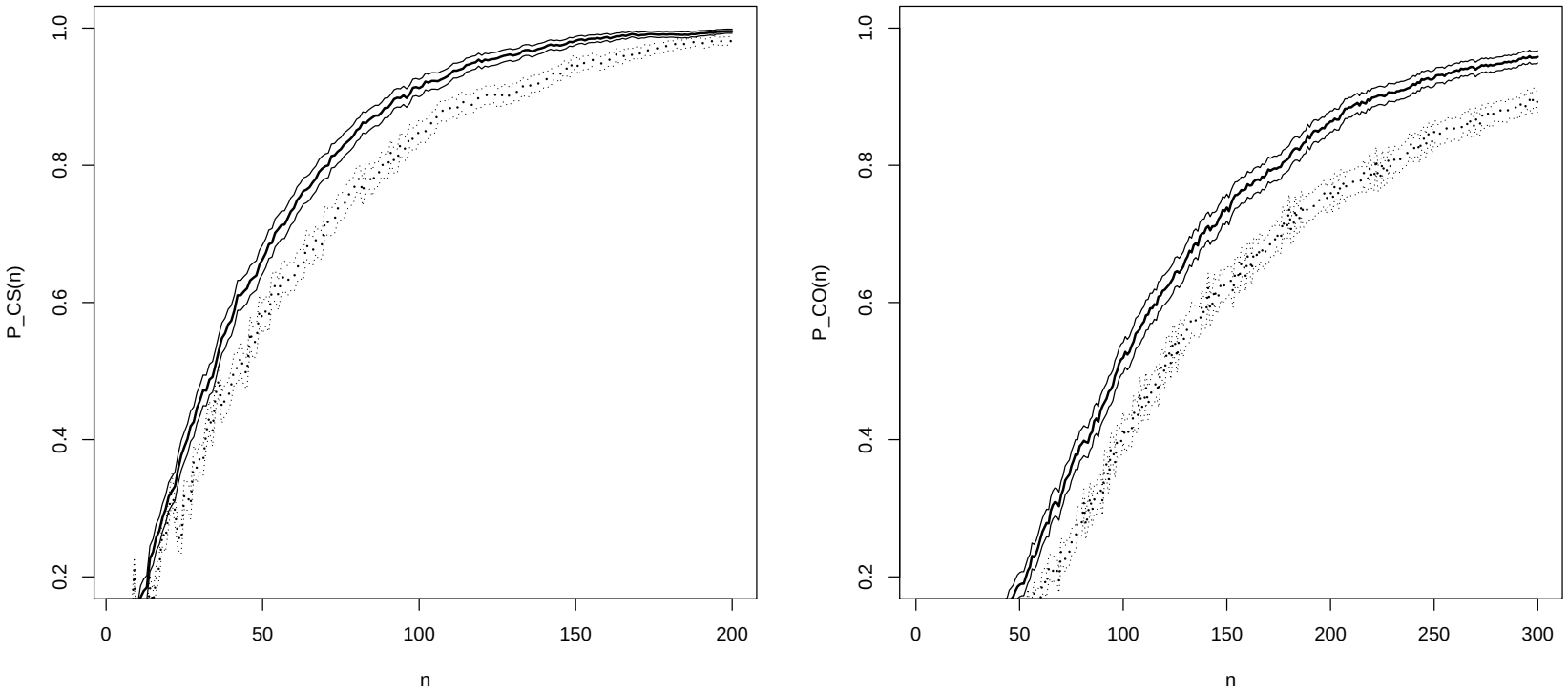
Theorem

The RCS design satisfies for any $C_r \geq 0$

$$\limsup_n \frac{1}{n} \log\{1 - P_{CS}(n)\} \leq \max\{R(p_A, p_C, \nu_A^{CS}, \nu_C^{CS}), R(p_B, p_C, \nu_B^{CS}, \nu_C^{CS})\},$$

The theorem states that the lim sup is smaller than the limit in the randomized design. The other direction seems also true but we could not find a formal proof for this claim.

Simulations

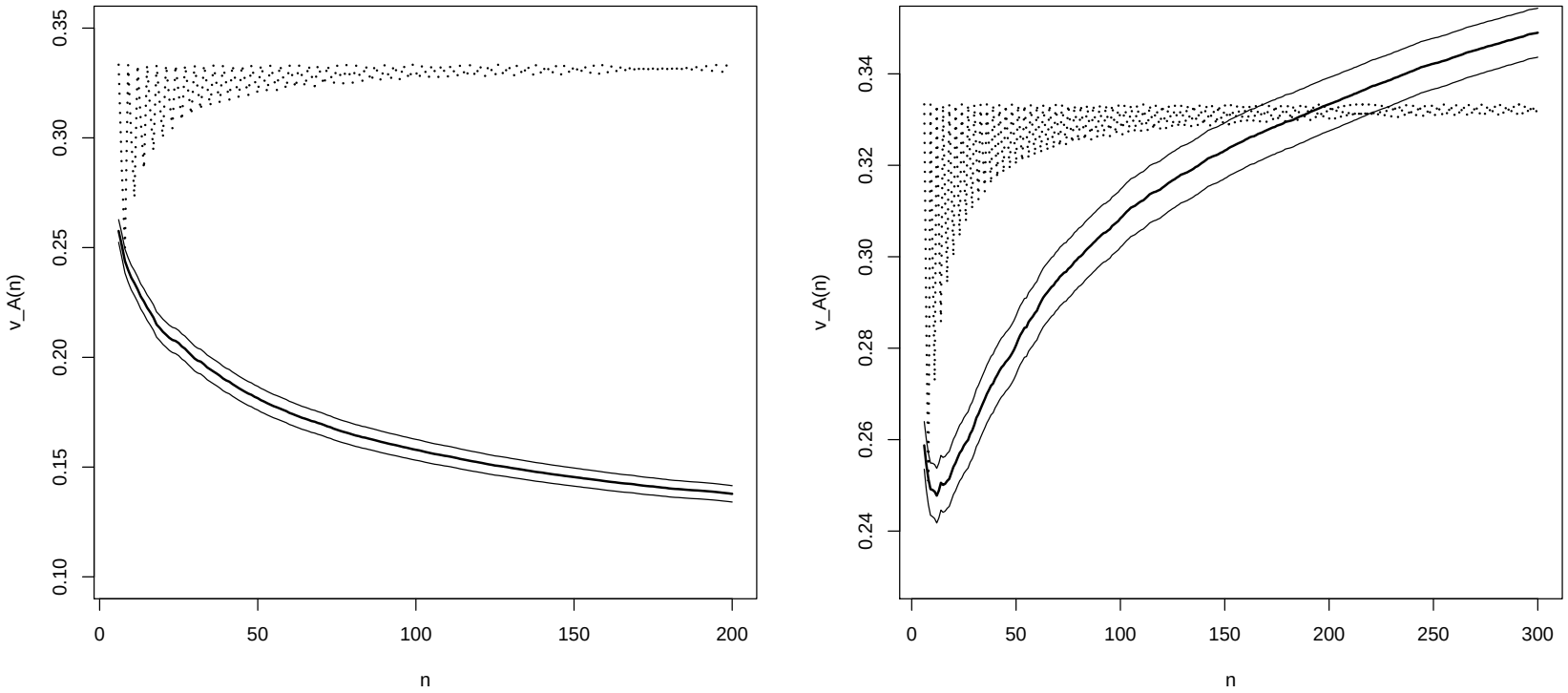


(a) $P_{CS}(n)$ in RCS and BD

(b) $P_{CO}(n)$ in RCO and BD

Figure: Estimates of $P_{CS}(n)$, $\nu_A(n)$, $P_{CO}(n)$ for $p_A = 0.3, p_B = 0.5, p_C = 0.8$, based on the mean of 2000 simulations. The solid line presents the RCS or RCO designs and dotted line is BD; the thin line indicates a 95% CI.

Simulations (cont.)



(a) $\nu_A(n)$ in RCS and BD

(b) $\nu_A(n)$ in RCO and BD

Figure: Estimates of $P_{CS}(n)$, $\nu_A(n)$, $P_{CO}(n)$ for $p_A = 0.3, p_B = 0.5, p_C = 0.8$, based on the mean of 2000 simulations. The solid line presents the RCS or RCO designs and dotted line is BD; the thin line indicates a 95% CI.

Extensions

- We considered the Wald test; the same results hold also for other tests such as log-risk and arcsin transformation.
- The generalization for more than three treatment is simple. For four treatments, for example, the maximum is computed over three terms rather than two.
- Similar results can be obtained for normal responses as well as other distributions (with m.g.f).

Normal responses

Table: Comparison of optimal allocations for normal responses.

μ_A	μ_B	μ_C	σ_A^2	σ_B^2	σ_C^2	ν_A^{CS}	ν_B^{CS}	ν_C^{CS}	ν_A^{CO}	ν_B^{CO}	ν_C^{CO}
1	2	3	1	1	1	0.07	0.46	0.47	0.29	0.41	0.29
1	2	3	2	1	1	0.12	0.43	0.44	0.42	0.37	0.21
1	2	3	4	1	1	0.22	0.38	0.40	0.56	0.31	0.14
1	2	3	3	2	1	0.12	0.51	0.37	0.44	0.41	0.15
1	2	3	1	2	3	0.03	0.44	0.54	0.15	0.41	0.44

Implication

- Strand et al. (2004) analyze two double-blind randomized trials comparing leflunomide (treatment) with placebo and sulfasalazine, or methotrexate, (standard) in active rheumatoid arthritis.
- In both trials the patients were randomized in a 2:3:3 ratio to placebo, standard and treatment.
- Primary outcome measures in both trials were ACR20 responses after several months. This criterion is binary and is defined as 20% or more improvement in at-least four out of six selected measures.

Implication (cont.)

This kind of three arms clinical trials are referred to as “active control equivalence trials” in Liu (2000).
To establish equivalence it is required that the hypotheses below will be rejected:

$$H_0: \ p_A = p_B \text{ versus } H_1: \ p_A < p_B$$
$$H_0: \ p_A = p_C \text{ versus } H_1: \ p_A < p_C$$
$$H_0: \ p_C \leq p_B - \delta \text{ versus } H_1: \ p_B - \delta < p_C,$$

for some small $\delta > 0$ (e.g. $\delta = 0.05$).

Implication (cont.)

Table: Optimal allocation for active control equivalence trials.

p_A	p_B	p_C	ν_A^*	ν_B^*	ν_C^*
0.29	**0.55**	**0.57**	0.017	0.492	0.490
0.3	0.5	0.6	0.150	0.440	0.410
0.35	0.55	0.6	0.065	0.471	0.464
0.25	0.6	0.58	0.002	0.499	0.499
Actually used			0.25	0.375	0.375

The optimal allocation is both more powerful and more ethical than the allocation actually used.

Discussion

- For **two treatment** adaptive designs can lead, at best, to a practically negligible improvements in terms of power.
- For three treatments the situation is different.
- An adaptive rule can be optimal for power maximization and also assigns a small fraction of subjects to the placebo treatment in certain situations.
- The optimal allocations are generally both more powerful and more ethical than standard allocations.

Discussion (cont.)

Consider three statisticians that:

1. Knows the true parameters but can use this knowledge only for the purpose of designing the experiment and not for inference.
2. Also knows the parameters but he can use only randomized designs, that is, at each stage he assigns $x_n = i$ with probability π_i for $i = A, B, C$.
3. Does not know the parameters and uses an adaptive design.

We showed that, in terms of the large deviation rate of an error, the first statistician can outperform the others and the second and third statisticians can perform equally well.

The end.

End.

Bibliography

- Azriel D., Mandel M., Rinott Y. (2012). Optimal allocation to maximize power of two-sample tests for binary response. *Biometrika*, **99**, 101 – 113.
- Bandyopadhyay U. and Bhattacharya, R. (2006). Adaptive Allocation and Failure Saving in Randomised Clinical Trials, *Journal of Biopharmaceutical Statistics*, **16** 817 - 829.
- Biswas A., Mandal S., Bhattacharya R. (2010). Multi-treatment optimal response-adaptive designs for phase III clinical trials, *Journal of the Korean Statistical Society*, in press.
- Brittain E., Schlesselman J.J. (1982). Optimal Allocation for the Comparison of Proportions, *Biometrics*, **38**, 1003 – 1009.
- Chambaz A., van der Laan M.J. (2010). Targeting The Optimal Design In Randomized Clinical Trials With Binary Outcomes And No Covariate *U.C. Berkeley Division of Biostatistics Working Paper Series*, **258**.

Bibliography(cont.)

- Hu F., Rosenberger W.F. and Zhang L. (2006). Asymptotically best response-adaptive randomization procedures, *Journal of Statistical Planning and Inference*, **136** 1911 - 1922.
- Liu J.P. (2000). Therapeutic equivalence. *Encyclopedia of Biopharmaceutical Statistics*, Marcel Dekker: New York.
- Hu F., Rosenberger W.F. (2006). *The theory of response-adaptive randomization in clinical trials*, Wiley, New York.
- Hu F., Rosenberger W.F. (2003), Optimality, Variability, Power: Evaluating Response-Adaptive Randomization Procedures for Treatment Comparisons, *Journal of the American Statistical Association*, **98**, 671–678.

Bibliography(cont.)

- Rosenberger W.F., Stallard N., Ivanova A., Harper C. N., and Ricks M. L. (2001). Optimal Adaptive Designs for Binary Response Trials, *Biometrics*, **57** 909 – 913.
- Strand V., Cohen S., Crawford B., Smolen J.S., Scott D.L. (2004). Patient-reported outcomes better discriminate active treatment from placebo in randomized controlled trials in rheumatoid arthritis, *Rheumatology*, **43**, 640 - 647.
- Tymofyeyev Y., Rosenberger W.F. and Hu F. (2007). Implementing Optimal Allocation in Sequential Binary Response Experiments, *Journal of the American Statistical Association*, **102** 224 – 234.
- Zhu H. and Hu F. (2010). Sequential Monitoring of response-adaptive randomized clinical trials, *The Annals of Statistics* **38**, 2218 - 2241.