Maximizing Power and Minimizing Treatment Failures Using Randomization

William F. Rosenberger

Department of Mathematics and Statistics
University of Maryland, Baltimore County
Department of Epidemiology and Preventive Medicine
University of Maryland, Baltimore

Feifang Hu

Department of Statistics
University of Virginia

Randomization in Clinical Trials:

Theory and Practice

By William F. Rosenberger and John M. Lachin

John Wiley and Sons, 2002

Winner, best new book on mathematics or statistics in 2002 by Association of American Publishers

Clinical Trials:

•Complex, multiple objectives

- •maximizing power to detect clinically relevant differences
- •maximizing the individual patient's experience in the trial
- •minimizing total monetary cost of trial
- •etc.

Binary response

$p_A : Pr(success|A)$, $q_A = 1 - p_A$

$p_B : Pr(success|B)$, $q_B = 1 - p_B$

n_A : number of patients on A

n_B : number of patients on B, $n = n_A + n_B$

Consider the variance of the difference of two proportions $\widehat{p}_A - \widehat{p}_B$:

$$\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}.$$

If we want to find the ratio $R = n_A/n_B$ to minimize this variance (i.e., maximize power),

we simply rewrite the expression as

$$\frac{(R+1)p_A q_A}{Rn} + \frac{(R+1)p_B q_B}{n},$$

take the derivative with respect to R and equate to zero.

Neyman allocation: Maximizes the power of the test of simple difference $p_A - p_B$.

$R(p_A, p_B) = n_A/n_B = \sqrt{p_A q_A}/\sqrt{p_B q_B}$.

Can also be interpreted in this way: *For fixed variance of the test, minimize the total sample size.*

Problems:

•Cannot be implemented; don't know what p_A, p_B are.

•When $p_A + p_B > 1$, Neyman allocation assigns more patients to the inferior treatment.

Alternative optimality criterion: *For fixed power of the t-test, find the optimality criterion to minimize the expected number of treatment failures.*

So our optimization criterion:

minimize the expected number of failures, given by

$$n_A q_A + n_B q_B = \frac{R}{R+1} n q_A + \frac{1}{R+1} n q_B$$

subject to

$\text{avar}\{\widehat{p}_A - \widehat{p}_B\} = K$ for constant K.

This leads to the allocation ratio

$$R(p_A, p_B) = n_A/n_B = \sqrt{p_A}/\sqrt{p_B}.$$

Rosenberger, Stallard, Ivanova, Harper, and Ricks, 2001 (RSIHR allocation)

This allocation deals with power consideration, but also the patient's experience in the trial.

But we still don't know the unknown parameters.

Gene therapy is the new frontier for clinical research.

Very small clinical trials have been conducted in a number of areas of gene therapy:

-- X-linked severe combined immunodeficiency
-- ADA deficiency
-- Mucopolysaccharidosis
-- Familial hypercholesterolemia
-- Cystic fibrosis
-- Hemophilia
-- Chronic granulomatous disease

However, few placebo-controlled randomized clinical trials have been performed.

Example: Transfer of cystic fibrosis transmembrane conductance regulator (CFTR) gene to the nasal epithelium.

Placebo-controlled trial: 8 patients on therapy; 8 on placebo
Primary outcome: Correction of chloride abnormality after 1 week.

Results: Active therapy was 83.3% successful, placebo was 28.6% successful ($p < 0.05$).

In the cystic fibrosis trial, $p_A = 0.833$, $p_B = 0.286$, so Neyman allocation is

$R = 0.825$ or 45% to active, 55% to placebo.

Clearly this is unethical. Neyman allocation may maximize power, but is unethical if
$p_A + p_B > 1$.

RSIHR allocation would be 63:37 in favor of treatment. For the 16 patients, we would have 10:6 allocation, and 4 more patients would be assigned to the experimental therapy, with minimal loss in power.

Why do equal allocation anyway?

Friedman, Furberg, and DeMets (1981, p. 41): (1) Equipoise dictates an uncertainty about the treatments at the beginning of the trial that should be mirrored in a 50:50 allocation ratio; (2) maximizes power.

Under the null hypothesis, both Neyman and RSIHR allocation would yield 50:50 allocation at the beginning of the trial.

But we could update an estimate of p_A and p_B at certain points in the trial, and adapt the allocation accordingly.

Melfi, Page, and Geraldes (2001): Sequential maximum likelihood estimation (SMLE):

After $j-1$ patients have responded, compute $\widehat{p}_A(j-1)$, $\widehat{p}_B(j-1)$, and $\widehat{R} = R(\widehat{p}_A(j-1), \widehat{p}_B(j-1))$. Then randomly assign treatment A to patient j with probability $\widehat{\rho} = \widehat{R}/(1+\widehat{R})$.

Then we would expect that the random proportion of patients assigned to A, $N_A(n)/n$ would converge to ρ (true under certain conditions).

Example: after 10 patients, we have $\widehat{p}_A$ = 0.4 and and $\widehat{p}_B$ = 0.6. Then the 11th patient would be assigned to A with probability $\sqrt{0.4}/(\sqrt{0.4}+\sqrt{0.6}) = 0.45$

Urn Models have been used because in adaptive designs, but these are not based on any optimality criteria.

Wei/Durham procedure:

Two treatments, A and B

Begin with α_A and α_B balls of types A and B in an urn

Draw A: assign patient to A
replace ball
add 1 type A ball if treatment A is successful
add 1 type B ball if treatment A is failure

Draw B: assign patient to B
replace ball
add 1 type B ball if treatment B is successful
add 1 type A ball if treatment B is failure

Some nice properties:

•Asymptotically, allocation ratios tend to

$N_A/N_B \rightarrow q_B/q_A$ (relative risk)

Ivanova's procedure (Ivanova, 2002):

Urn contains balls of type A, type B, and type 0.

Type A or B drawn: assign that treatment.
Success: ball is replaced, urn unchanged.
Failure: ball is not replaced.

Type 0 ball drawn: no subject is treated.
ball is replaced along with one
of type A and one of type B.

Same limiting allocation as RPW, but reduced variability.

Example: Suppose we start the trial with 1 ball of each type in the urn. The first patient is ready to be randomized and a type A ball is drawn (with probability 1/3). The patient is a success, and the ball is replaced. The probability of assignment to A is still 1/3. Suppose for patient 2, a type 0 ball is drawn. That ball is returned to the urn along with an additional type A and type B ball. Another ball is drawn. This time the probability that patient 2 will be assigned to A is 2/5 and to B is 2/5. Suppose a type A ball is drawn and we have a treatment failure. The ball is not returned to the urn so the probability that patient 3 will be assigned to A is now 1/4 and to B is 1/2.

Eisele's procedure, Eisele (1994).

Let $g(x,\rho) = \dfrac{\rho(\rho/x)^\gamma}{\rho(\rho/x)^\gamma + (1-\rho)((1-\rho)/(1-x))^\gamma}$

(Hu and Zhang, 2002).

Allocate to treatment A with probability $g(N_A(j-1)/(j-1), \widehat{\rho})$.

Note that when $\gamma = 0$, we have the SMLE procedure. So this can be considered a generalization of that procedure. We found $\gamma = 2$ to have particularly good properties.

Again, under certain conditions, $N_A(n)/n \to \rho$.

Example: Suppose we have assigned 9 patients, 5 to A and 4 to B. We have observed a success rate of $\widehat{p}_A$ = 3/5 and and $\widehat{p}_B$ = 1/4. If we are interested in RSIHR allocation, we can compute
$\widehat{\rho} = \sqrt{3/5}/(\sqrt{3/5} + \sqrt{1/4}) = 0.6077.$
Then the probability of assigning the 10th patient to A ($\gamma = 2$) is computed as

$$\text{P(A)} = \frac{0.6077(9\times 0.6077/5)^2}{0.6077(9\times 0.6077/5)^2 + 0.3923(9\times 0.3923/4)^2} = 0.704.$$

Response-adaptive randomization procedures have some advantages:

1. They are fully randomized, and protect against biases (Rosenberger and Lachin, 2002)

2. They reflect the magnitude of the treatment effect thus far.

3. They can be used as a recruitment tool.

Some have argued that one must make assignments with probability 1 to either A or B if one treatment is performing better. This was the principal component of much adaptive designs research through the 60's and 70's.

We wish to compare expected treatment failures for these four procedures, and power of the test. Heretofore, we could only do this by simulation.

Hu and Rosenberger (2003, JASA)

Power depends on

1. the allocation target.

Procedures that target RSIHR or Neyman will be more powerful than urn procedures.

2. the variability of the procedure.

The most variable procedure is Wei/Durham's procedure. We have shown that Ivanova's procedure has the lowest variability that can possibly be achieved by any randomized response-adaptive procedure. (Can also be attained by deterministic procedures, such as Eisele's procedure as $\gamma \to \infty$ or Zelen's play-the-winner rule).

3. how fast the procedure converges to the target allocation.

Ivanova's is slowest to converge. Eisele's is fastest to converge.

So which procedure is better? Guiding principle: the procedure must be fully randomized.

CONCLUSIONS:

Eisele's procedure is always better than plain old complete randomization!

Table 1: *Simulated power and expected treatment failures (S.D.) for complete randomization and three response-adaptive randomization procedures. 10,000 replications.*

			Complete		Wei/Durham		Ivanova		Eisele	
p_A	p_B	n	Power	Failures	Power	Failures	Power	Failures	Power	Failures
0.9	0.1	17	91	9 (2.1)	86	6 (2.1)	89	6 (1.6)	91	8 (2.1)
0.9	0.3	38	90	15 (3.0)	83	11 (3.0)	85	11 (2.3)	89	13 (2.3)
0.9	0.5	96	90	29 (4.5)	83	21 (4.7)	83	20 (3.6)	90	26 (3.5)
0.9	0.7	400	91	80 (8)	85	67 (9)	83	63 (7)	90	78 (7)
0.9	0.8	1600	91	240 (14)	88	221 (17)	86	215 (14)	91	237 (14)
0.7	0.3	78	90	39 (4.4)	88	35 (4.8)	87	33 (4.7)	90	35 (3.9)
0.7	0.5	368	90	147 (9)	89	139 (10)	89	139 (10)	90	144 (9)
0.5	0.4	1200	89	660 (17)	89	655 (18)	89	655 (17)	89	657 (17)
0.3	0.1	150	90	120 (5)	90	118 (5)	90	118 (5)	91	115 (5)
0.2	0.1	480	91	408 (8)	91	407 (8)	91	407 (8)	91	404 (8)

REFERENCES

[1] Rosenberger WF, Stallard N, Ivanova A, Harper C, Ricks M. Optimal adaptive designs for binary response trials. Biometrics 2001;57:909-913.

[2] Friedman LM, Furberg CD, DeMets DL. Fundamentals of Clinical Trials. Boston: Wright PSG, 1981.

[3] Rosenberger WF, Lachin WF. Randomization in Clinical Trials. New York: Wiley, 2002.

[4] Hu F, Rosenberger WF. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. Journal of the American Statistical Association 2003 (in press).

[5] Rosenberger WF. Urn models and sequential design. Sequential Analysis 2002;21:1-42 (with discussion).

[6] Eisele JR. The doubly adaptive biased coin design for sequential clinical trials. Journal of Statistical Planning and Inference 1994;38:249-261.

[7] Hu F, Zhang L-X. Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. Annals of Statistics 2004, in press.

[8] Hu F, Rosenberger WF. Asymptotically best response-adaptive randomization procedures. Submitted.

[9] Ivanova AV. A play-the-winner type urn model with reduced variability. Metrika 2003, in press.

[10] Wei LJ, Durham S. The randomized play-the-winner rule in medical trials. Journal of the American Statistical Association 1978;73:840-843.