



Comments on the Controversy over Response-Adaptive Randomization

Rick Chappell, Ph.D.

Professor,

Department of Biostatistics and Medical Informatics

The Paul P. Carbone Comprehensive Cancer
Center

University of Wisconsin Madison

Joint work with Ted Karrison, University of Chicago

National Cancer Policy Forum Workshop

10/2007

Washington, DC



The debate over RAR has been long-running, dating (at least) back to papers by:

Zelen (“play the winner”, 1979)

Peto (1985)

Bartlett, *et al.* (ECMO, 1985)



So what is the problem?

Peto (1985):

“ . . . patients entering a trial are often not uniform in prognosis. For example, both in the European Coronary Bypass Trial and in the Medical Research Council’s 8th Acute Myeloid Leukaemia Trial **there was a statistically significant trend** towards better prognosis as time went by. If, therefore, there had also been a trend in allocation proportions towards the apparently better treatment, an appreciable bias might have been engendered.”

Rosenberger and Lachin (1993): valid inference requires

“The sequence of patients who arrive for entry into the trial represents samples drawn at random from two homogeneous probabilities, with **no drift** in the probabilities of success.”



How a Time Trend can Artificially Generate an Apparent Treatment Effect

1. Binary outcome variable (success or failure) observed within a relatively short time period from treatment initiation.
2. Initial allocation ratio is 1:1 to treatments *A* and *B*.
3. For ease of demonstration, randomization ratios are held constant within blocks of patients; initially, $n_A = n_B = 100$. Subsequent randomization ratios can vary such that $n_A \neq n_B$. If so, analysis is simplified if $n_A + n_B > 200$, but this is peripheral to the example. (Subscripts and other notation elided).



-
4. After each group of $n_A + n_B$ patients, the proportion of failures is compared between the two groups using a standard test statistic derived from the *accumulated* data. The subsequent group is assigned to treatment according to a typical RAR rule:

$ Z < 1$:	allocate in ratio 1:1
$1 \leq Z < 1.5$:	allocate in ratio 2:1
$1.5 \leq Z < 2$:	allocate in ratio 3:1

First group:

- 1:1 randomization - $100 + 100 = 200$ patients total
- Failure rate = 75% *in each arm*
- Slight but nonsignificant imbalance in failure rates

	S	F
Arm A	20	80
Arm B	30	70

80% vs. 70%, $Z = 1.03$

Second group:

- 2:1 randomization - $75 + 150 = 225$ patients total
- Failure rate = 50% *in each arm*
- Exact balance in failure rates

	S	F
Arm A	37	38
Arm B	75	75

From the accumulated evidence ($57/175$ vs. $105/250$), $Z = 1.87$;

“Evidence” for treatment difference increases despite exact parity in this group.

Third group:

- 3:1 randomization - $67 + 200 = 267$ patients total
- Failure rate = 50% *in each arm*
- Exact balance in failure rates

	S	F
Arm A	33	34
Arm B	100	100

From the accumulated evidence ($152/242 = 63\%$ vs. $245/450 = 54\%$),
 $Z = 2.04$;

“Significant Evidence” for overall treatment difference despite exact parity in this group also.

Time drift can lead to bias by confounding the treatment effect with time effects.

Although models that incorporate covariate effects can ameliorate this problem . . .

“. . . models are various and subjective, trend information can be sparse, and trials may be left with ambiguous and multiple adjusted answers, difficulties that are supposed to be prevented by randomization.”

Chappell and Karrison (LTE, 2007)

A stratified group-sequential approach (Karrison, Huo and Chappell, 2003) can address this bias at some cost in power. Few proposed methods use any correction for trend or indeed acknowledge the problem. Simulations typically assume zero trend.



Discussion

The advantage of response-adaptive designs is that they may reduce the number of patients assigned to the inferior treatment arm.

Why might one still be cautious about using them?

1. Many would argue that equal allocation mirrors the state of equipoise that exists at the beginning of the trial (Friedman, Furberg, and DeMets, 1981) and that, absent definitive evidence in favor of one treatment over another, it is neither efficient nor ethically appropriate to assign patients in a different ratio.
2. With the use of early stopping rules, the benefits from a response-adaptive design relative to equal allocation are greatly lessened; the ethical need for adaptation is obviated.



-
3. Peto (1985): “Most patients are not in any trial. . . . So, if trials are going to have any large impact on practice, the sooner they yield reliable results the better. This alone provides an important reason to keep the allocation probabilities approximately even, for statistical sensitivity will thereby be maximized.”
 4. RAR designs produce allocations for which standard Bayesian and frequentist estimates of treatment effects are biased in the presence of a time trend.



Conclusion

1. Results from response adaptive randomized phase II studies may be superior to those produced from nonrandomized designs.
2. Due to fears of bias from confounding with time trends, RAR cannot substitute for true randomization in confirmatory trials.



References

Bartlett RH, Roloff DW, Cornell RG, et al. (1985). Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study, *Pediatrics* 76:479-87.

Chappell R, Karrison T (2007). Letter to the Editor, *Statistics in Medicine*, in press.

Friedman LM, Furberg CD, DeMets DL (1981). *Fundamentals of Clinical Trials*. Boston: Wright PSG.

Karrison TG, Huo D, Chappell R (2003). A group sequential, response-adaptive design for randomized clinical trials, *Control Clin Trials* 24:506-22.

Peto R (1985). Discussion of papers by J.A. Bather and P. Armitage. *International Stat Review* 53:31-34.

Rosenberger WR, Lachin JM (1993). The use of response-adaptive designs in clinical trials, *Control Clin Trials* 14:471-84.

Zelen M (1979). A new design for randomized clinical trials, *N Engl J Med* 300:1242-6.