Drug Information Journal, Vol. 35, 1087–1093, 2001
Printed in the USA.

0092-8615/2001

ADAPTIVE DESIGNS FOR CLINICAL TRIALS WITH HIGHLY SUCCESSFUL TREATMENTS*

ANASTASIA IVANOVA, PHD
Assistant Professor of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina
WILLIAM F. ROSENBERGER, PHD
Associate Professor of Mathematics and Statistics, Adjunct Associate Professor of Epidemiology and Preventive Medicine, University of Maryland, Baltimore County and Baltimore, Baltimore, Maryland

We compare the performance of two adaptive designs and equal allocation in a clinical trial with two highly successful treatments and binary outcomes. The measure of interest in the trial is the odds ratio. The goal of the adaptive design is to decrease the total number of failures compared to equal allocation while keeping the power at the same level. One design is based on sequential maximum likelihood estimation, the other on an urn model. We find that the urn model produces a better procedure than the sequential maximum likelihood approach and equal allocation, in that it yields fewer expected treatment failures, maintains the power of the asymptotic test, and is more powerful when the Fisher's exact test is used. We conclude that adaptive designs have attractive properties when both treatments are highly successful.

Key Words: Ethics; Measures of association; Optimal allocation; Power; Urn models

INTRODUCTION

RANDOMIZING PATIENTS in equal proportions to experimental and control therapies in a clinical trial has become so embedded in the culture of clinical research that it is a useful exercise to examine afresh the basis for equal allocation. For, if there were no substantial reasons for equal allocation to treatments, one could reasonably presume that patients would prefer to have a greater than 50% chance of being assigned the better treatment; and one could reasonably presume that physicians would prefer that their patients have a greater than 50% chance of being assigned the better treatment. In some diseases where recruitment of patients is especially difficult, advertising a 2 : 1 or 3 : 1 allocation favoring the experimental therapy can be a recruitment tool.

In the seminal text by Friedman et al. (1), two reasons for advocating equal allocation are given:

1. Equal allocation maximizes the power of tests of the treatment effect, and
2. Equal allocation is consistent with the state of equipoise that precedes the trial.

Let us examine both of these statements more carefully.

Equal allocation may, in fact, not maxi-

Reprint address: Professor Anastasia V. Ivanova, Department of Biostatistics, University of North Carolina at Chapel Hill, CB7420, Chapel Hill, NC 27599–7420.

*Grant support: For Professor Rosenberger, R29–DK51015–05 from the National Institute of Diabetes and Digestive and Kidney Diseases. This was an invited paper presented by the second author at the Pharmaceutical Research and Manufacturers Association/Food and Drug Administration Workshop, November 1, 2000, Washington, District of Columbia.

mize power in many cases. One common example can be seen from the standard power computation for the comparison of two normal means. Power is maximized when allocation is proportional to the standard deviations of each therapy. Consequently, when the variability in the experimental treatment group is larger than that of the control group, the allocation that assigns more patients to the experimental therapy will be more efficient.

In trials with binary response, where p_A is the probability of success on treatment A and p_B is the probability of response on treatment B, equal allocation to A and B will not necessarily minimize the variance of the standard test of the simple difference, $p_A - p_B$. Solving a simple optimization problem, we obtain the optimal allocation to maximize power. If N_A is the number of patients assigned to A out of n total, then the allocation proportion should be

$$\frac{N_A}{n} = \frac{\sqrt{p_A q_A}}{\sqrt{p_A q_A} + \sqrt{p_B q_B}},$$

where $q_A = 1 - p_A$ and $q_B = 1 - p_B$. This expression is known as *Neyman allocation*, and, as in the normal means example, allocation is proportional to the standard deviation for each treatment. In practice, Neyman allocation can only be implemented when p_A and p_B are known in advance, which is never the case. However, there have been sequential allocation rules proposed that tend to Neyman allocation as the total sample size n grows large (eg, 2). Sequential allocation rules that update according to accrued information are called *adaptive designs*; their function is to adaptively change allocation probabilities based on accruing patient data to satisfy some objective, such as attaining some optimal allocation in the limit. Many adaptive designs have been shown to yield more powerful tests than equal allocation (eg, 3,4).

The second basis for equal allocation is that it is consistent with the equipoise at the beginning of the clinical trial. Since investigators do not know which treatment is better, it is consistent with that uncertainty to give patients an equal chance of receiving either. Such an argument is fine at the beginning of the trial, but as we are monitoring the trial and we are observing that one of the treatments tends to be performing better than the other, should we continue equal allocation? Adaptive designs can mitigate this dilemma, in some sense. As data accruing begin to favor a treatment (albeit not conclusively enough to stop the trial and declare a "winner"), adaptive designs increase the probability that patients will be assigned to that treatment. Thus, adaptive designs can dynamically update the allocation according to our present knowledge.

Neyman allocation was obtained when the sample size was minimized for fixed power. In Rosenberger et al. (4), an optimal allocation for comparing two binomial success probabilities was determined according to the following criteria: for fixed power and size of the test, determine the design that minimizes the expected number of treatment failures (one can also deduce the optimal allocation from more general formulas given in [5]).

Optimum allocation was derived for three common measures of association: the simple difference, $p_A - p_B$, the relative risk, q_B/q_A, and the odds ratio, $p_A q_B / p_B q_A$. The optimal allocation N_A/n is given in Table 1. Each gives a different allocation from Neyman allocation, because the optimization criterion is different. These allocation proportions cannot be implemented in practice, since they depend on the unknown parameters.

We can use adaptive designs to approximate these allocation proportions for large n. A formal method to do this, by substituting sequentially computed estimates of p_A and p_B, is explored formally by Rosenberger et al. (4). An alternative approach, in which potential allocation schemes are less variable, is an adaptive design based on an urn model, called the *drop-the-loser rule* (Ivanova and Durham, unpublished data, 2001). While the decreased variability is attractive,

TABLE 1
Optimal Allocation (N_A/n) for Three Measures of Association (4)

Measure		Optimal Allocation
Simple difference	$p_A - p_B$	$\frac{\sqrt{p_A}}{\sqrt{p_A}+\sqrt{p_B}}$
Relative risk	$\frac{q_B}{q_A}$	$\frac{\sqrt{p_A}}{\frac{\sqrt{p_A}}{q_A}+\frac{\sqrt{p_B}}{q_B}}$
Odds ratio	$\frac{p_A q_B}{q_A p_B}$	$\frac{\frac{1}{\sqrt{p_A q_A}}}{\frac{1}{\sqrt{p_A q_A}}+\frac{1}{\sqrt{p_B q_B}}}$

the design does not approach optimal allocation except when p_A and p_B are large, that is, both treatments are highly successful.

There are many diseases for which a high success rate for therapy is expected. For example, in the cardiovascular disease COMPASS trial, the primary outcome of 30-day mortality was observed in only 5.7% of saruplase-treated patients and 6.7% of streptokinase-treated patients (6). In the important clinical trial assessing the effects of AZT in maternal-fetal HIV transmission, there was a 91.6% success rate for mothers on AZT versus 74.8% on placebo (7).

The focus of this paper is on binomial trials with immediate response. In practice, most clinical trials do not have immediate response, but adaptive designs are relevant when most outcomes are likely to be obtained during the recruitment period. In this case, allocation probabilities can be updated using the data available. An adaptive design is likely not to converge as quickly to the limiting allocation, but there will be some beneficial effects of the adaptive design. Other papers have explored properties of delayed response by simulating a priority queue under a stochastic delay mechanism (3,8,9), and have demonstrated that adaptive designs with delayed response have similar properties to those with immediate response. Adaptive designs are not relevant for long-term clinical trials when responses are observed months or years after a limited recruitment period.

We wish to compare procedures that are designed to minimize the expected number of failures for clinical trials comparing two highly successful treatments. In this context, we consider highly successful treatments to have success probabilities greater than 0.50. Our goal is to find an adaptive design with smaller variability and a test that yields reasonable size and power. A simulation study is an appropriate tool to investigate comparison of adaptive designs. Most theoretical comparisons would rely on asymptotic results that may not hold in the context of finite sample clinical trials. Another method, exact computations using networking algorithms, is computationally unfeasible for clinical trials of more than 75 patients. Therefore, in this paper, we present the results of a simulation study.

METHODS

Measures of Association

The investigator will typically choose an appropriate measure of association for the study. For small p_A and p_B, Figure 1 demonstrates that allocation according to the relative risk measure is similar to that for the simple difference measure. Although the allocation is similar, the power of the tests is not. Previous work (4) has found that a test based on the simple difference will yield better power than tests based on other measures.

In this paper, the measure of interest is the odds ratio, and we will focus on large values of p_A and p_B. We will compare the performance of two different adaptive designs and equal allocation using the same test based on the odds ratio.

Adaptive Designs

We explore two types of adaptive designs, one based on *sequential maximum likelihood*

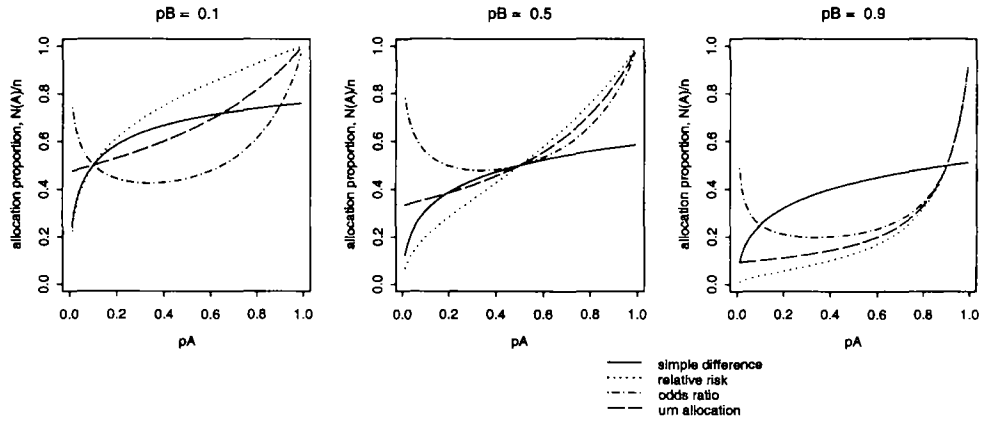

FIGURE 1. Optimal allocation rules, given in Table 1, plotted across values of p_A for $p_B = 0.1, 0.5, 0.9$. Also plotted is the limiting DL rule (urn) allocation.

estimation (SMLE) (sometimes referred to as *adaptive biased coin designs*) and the other based on an urn model, the drop-the-loser (DL) rule. These are described as follows:

- **SMLE:** When patient j is ready to be randomized, compute $\hat{p}_A(j-1)$ and $\hat{p}_B(j-1)$, the proportions of observed successes among the first $j-1$ patients. These estimates are then substituted into the optimal allocation desired. For example, from Table 1, for the odds ratio, we would compute

$$\frac{\frac{1}{\sqrt{\hat{p}_A(j-1)\hat{q}_A(j-1)}}}{\frac{1}{\sqrt{\hat{p}_A(j-1)\hat{q}_A(j-1)}}+\frac{1}{\sqrt{\hat{p}_B(j-1)\hat{q}_B(j-1)}}}.$$

  This quantity is then used as the probability that patient j will be assigned to treatment A. Hence, the treatment assignments are random and depend on previous patient outcomes. For this procedure, the proportion of patients assigned to treatment A converges to the optimal allocation as n becomes large (4), and
- **DL Rule:** Consider an urn containing balls of three types, labeled A, B, and C. Balls of type A and B represent treatments, and balls of type C are called *immigration balls*. When a patient is ready to be randomized, a ball is drawn at random. If it is type A or type B, the patient is assigned to treatment A or B, respectively. If the patient then has a failure on that treatment, the ball is not replaced. If the patient has a success on that treatment, the ball is replaced, and the urn remains unchanged. If the ball drawn is type C, the patient is not treated, and the ball is replaced in the urn, together with an additional type A ball and an additional type B ball. The patient will be treated with A or B depending on which treatment ball is drawn subsequently. The procedure iterates through all patients. Immigration balls ensure that the urn does not become empty during the course of the trial. Typically, an urn will begin with one immigration ball, and one or more balls representing each of the treatments. As was proved in Ivanova and Durham (unpublished data, 2001) in the limit, the proportion of patients assigned to A converges to

$$\frac{\frac{1}{q_A}}{\frac{1}{q_A}+\frac{1}{q_B}}.$$

Note that this allocation is not optimal for any measure, according to Table 1, but in Figure 1, we see that, for large values of p_A and p_B, this allocation is very close to optimal. Other urns, such as the randomized-play-the-winner rule (10), approach the same limit with similar rate, but have greater variability (Ivanova and Durham, unpublished data, 2001).

Hypothesis Tests

We examined two tests, the Fisher's exact test and an asymptotic test for the odds ratio:

- **Fisher's exact test:** This test is the most appropriate for testing the hypothesis that the log odds ratio is 0. However, it assumes that all margins are fixed, which is not appopriate for adaptive designs, because N_A and N_B are sufficient statistics for the odds ratio, as they absorb information on the treatment effect which can be detected by their distance from $n/2$. Nonetheless, we were curious to see if adaptive allocation made a difference with respect to size and power of the test, and
- **Asymptotic test:** In order to symmetrize the odds ratio measure, we will take the logarithmic transform. The usual test statistic of the log odds ratio, given by

$$Z = \frac{log\left(\frac{\hat{p}_A\hat{q}_B}{\hat{p}_B\hat{q}_A}\right)}{\sqrt{\frac{1}{N_A\hat{p}_A\hat{q}_A} + \frac{1}{N_B\hat{p}_B\hat{q}_B}}}$$

has an asymptotic standard normal distribution under both the SMLE (4) and the DL rule (11). We use a continuity corrected version of this test, where $\hat{p}_A$ and $\hat{p}_B$ are computed by adding 0.5 to the numerator and denominator.

Simulation Details

Simulations were programmed in S-Plus and were run for 5000 replications for a two-sided test with nominal $\alpha = 0.05$. We simulated power and size of the test for $p_A(p_B) = 0.6$ to 0.9 by 0.1 and also computed the simulated expected number of failures and the standard deviation. We simulated the DL rule and the SMLE procedures, along with equal allocation. For the adaptive procedures, the first six patients were randomly allocated in equal numbers to A and B, to force balance. For the SMLE, these data were then used to compute the probability of assignment for the seventh patient. If $\hat{p}_A(\hat{p}_B)$ was 0 or 1, we allocated with probability 1/2. For the DL rule, the initial urn composition for the seventh patient consisted of one immigration ball, one ball each representing treatments A and B, one additional ball of type A for each success observed in the first three patients treated on A, and one additional ball of type B for each success observed in the first three patients treated on B. Since, for equal allocation, N_A follows a binomial distribution, the expected value and standard deviation of the number of failures can be computed directly from the binomial probability function. Therefore, values in the table for equal allocation are not simulated, but are computed directly. We determined the appropriate sample sizes to yield approximate power 0.80 under equal allocation, to facilitate comparisons.

RESULTS

Table 2 gives the expected number of treatment failures and the standard deviation for the SMLE approach, DL rule, and equal allocation. One can see that the SMLE and DL rule are very similar, with the SMLE having slightly fewer failures and always being better, with respect to expected treatment failures, than equal allocation when $p_A \neq p_B$. The DL rule is generally less variable than the SMLE approach. The adaptive procedures tend to yield savings from four to six patients.

Tables 3 and 4 give the size and power of the Fisher's exact test and the asymptotic test, Z, defined earlier, respectively. While the Fisher's exact test preserves the size of

TABLE 2
Simulated Expected Number of Failures (S.D.) for the DL Rule (*D*), SMLE (*S*), and Equal Allocation (*E*). (Results for equal allocation are computed directly from the binomial distribution.)

p_A	p_B	n	*D*	*S*	*E*
0.6	0.6	200	80.1 (6.9)	80.2 (7.0)	80.0 (6.9)
0.7	0.7	200	60.1 (6.5)	59.9 (6.6)	60.0 (6.5)
0.8	0.8	200	40.0 (5.6)	40.0 (5.6)	40.0 (6.7)
0.9	0.9	200	20.0 (4.2)	20.0 (4.2)	20.0 (4.2)
0.6	0.7	712	244.1 (12.8)	245.6 (13.3)	249.2 (12.7)
0.7	0.8	584	140.6 (10.5)	141.3 (11.1)	146.0 (10.4)
0.8	0.9	394	54.1 (6.7)	53.4 (7.5)	59.1 (7.0)
0.6	0.8	162	44.1 (5.8)	44.6 (6.4)	48.6 (5.7)
0.7	0.9	122	20.6 (3.9)	20.0 (4.8)	24.4 (4.3)
0.6	0.9	64	12.9 (3.0)	11.9 (3.8)	16.0 (3.3)

$\alpha = 0.05$, 5000 replications.

the test, it is slightly conservative and less powerful than the asymptotic test. For large values of p_A and p_B, the test is very conservative with equal allocation. For the asymptotic test, we see that the DL rule is slightly more powerful than the SMLE approach, especially when the differences are larger. In every case, the DL has power within 1% of equal allocation.

The reason that the DL rule is more powerful than the SMLE, in general, is that the DL rule is less variable. For instance, when $p_A = p_B = 0.9$, the expected allocation proportion for the DL rule is 0.50 with standard deviation 0.06. For the SMLE, the expected allocation proportion is the same, with standard deviation 0.12

CONCLUSIONS

The DL rule is invariant to the choice of measure of association, and the limiting allocation is not optimal. Nonetheless, we conclude that for clinical trials with highly suc-

TABLE 3
Simulated Power (Size) of the Fisher's Exact Test for the DL Rule (*D*), SMLE (*S*), and Equal Allocation (*E*)

p_A	p_B	n	*D*	*S*	*E*
0.6	0.6	200	0.04	0.04	0.04
0.7	0.7	200	0.04	0.05	0.04
0.8	0.8	200	0.04	0.04	0.03
0.9	0.9	200	0.04	0.04	0.03
0.6	0.7	712	0.78	0.77	0.78
0.7	0.8	584	0.77	0.78	0.77
0.8	0.9	394	0.78	0.76	0.76
0.6	0.8	162	0.76	0.75	0.75
0.7	0.9	122	0.77	0.73	0.74
0.6	0.9	64	0.78	0.71	0.74

$\alpha = 0.05$, 5000 replications.

TABLE 4
Simulated Power (Size) of the Asymptotic Test (*Z*) for the DL Rule (*D*) SMLE (*S*), and Equal Allocation (*E*) for $n = 100$

p_A	p_B	n	*D*	*S*	*E*
0.6	0.6	200	0.05	0.05	0.05
0.7	0.7	200	0.04	0.05	0.05
0.8	0.8	200	0.05	0.05	0.05
0.9	0.9	200	0.05	0.05	0.04
0.6	0.7	712	0.79	0.78	0.80
0.7	0.8	584	0.79	0.79	0.80
0.8	0.9	394	0.80	0.79	0.80
0.6	0.8	162	0.78	0.77	0.80
0.7	0.9	122	0.79	0.77	0.80
0.6	0.9	64	0.80	0.75	0.80

$\alpha = 0.05$, 5000 replications.

cessful treatments, the DL rule behaves quite well and is less variable than the SMLE approach. We also recommend the Z test, as it preserves type I error fairly well and has higher power than the Fisher's exact test, although the latter did surprisingly well, given that the test is conditional on sufficient statistics.

We found that adaptive designs are most useful when they are most needed: when the differences between p_A and p_B are larger. This is especially true when we compare $p_A = 0.9$ and $p_B = 0.6$. In this case, there is little to recommend equal allocation, as there is a large increase in the expected number of failures. It appears that equal allocation is only particularly effective for midrange values of p_A and p_B. A good rule of thumb would be between 0.4 and 0.6. We can see this more clearly in Figure 1, where the optimal allocation flattens out and is very well approximated by equal allocation.

There is a clear advantage to adaptive designs that cannot be measured in the tables. If a study is powered for a small difference, but the true treatment difference is unexpectedly larger, much can be gained, in terms of expected treatment failures, by having employed the adaptive procedure rather than equal allocation.

We did not explore the DL rule for small success rates because it does not approximate well the optimal allocation (see Figure 1). In this case, for the simple difference measure, the SMLE performs better than equal allocation in most cases (4).

We conclude that adaptive designs should be considered in planning clinical trials when both treatments are presumed highly successful. In particular, the DL rule will increase power over equal allocation and allow fewer treatment failures, on average.

REFERENCES

1. Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials*. Boston, MA: Wright PSG; 1981, 41.
2. Melfi V, Page C. Variability in adaptive designs for estimation of success probabilities. In *New Developments and Applications in Experimental Design*. Flournoy N, Rosenberger WF, Wong WK, eds. Hayward, CA: Institute of Mathematical Statistics: 1998, 106–114.
3. Ivanova AV, Rosenberger WF. A comparison of urn designs for randomized clinical trials of $K > 2$ treatments. *J Biopharm Stat*. 2000;10:93–107.
4. Rosenberger WF, Stallard N, Inanova A, Harper C, Ricks M. Optimal adaptive designs for binary response trials. *Biometrics*. 2001;57:833–837.
5. Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, FL: Chapman and Hall/CRC; 2000, 329.
6. The COMPASS Investigators. Randomized double-blind study comparing saruplase with streptokinase in acute myocardial infarction: the COMPASS equivalence trial. *J Am Coll Cardiol*. 1998;31:487–493.
7. Connor EM, Sperling RS, Gelber R, Kiselev P, Scott G, O'Sullivan MJ, VanDyke R, Bey M, Shearer W, Jacobson RL, Jiminez E, O'Neil E, Bazin B, Delfraissy J, Culnane M, Coombs R, Elkins M, Moye J, Stratton P, Balsley J for the Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. *New Engl J Med*. 1994;331:1173–1180.
8. Rosenberger WF, Seshaiyer P. Adaptive survival trials. *J Biopharm Stat*. 1997;7:617–624.
9. Rosenberger WF. Randomized play-the-winner clinical trials: review and recommendations. *Control Clin Trials*. 1999;20:328–335.
10. Wei LJ, Durham S. The randomized play-the-winner rule in medical trials. *J Am Stat Assoc*. 1978;73: 840–843.
11. Ivanova AV, Rosenberger WF, Durham SD, Flournoy N. A birth and death urn for randomized clinical trials: asymptotic results. *Sankhya B*. 2000;62:104–118.