Journal of Biopharmaceutical Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597263

Adaptive survival trials

William F. Rosenberger[ab]; Padmanabhan Seshaiyer[a]
[a] Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, Maryland [b] Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, Maryland

To cite this Article Rosenberger, William F. and Seshaiyer, Padmanabhan(1997) 'Adaptive survival trials', Journal of Biopharmaceutical Statistics, 7: 4, 617 — 624
To link to this Article: DOI: 10.1080/10543409708835211
URL: http://dx.doi.org/10.1080/10543409708835211

Journal of Biopharmaceutical Statistics, 7(4), 617–624 (1997)

ORIGINAL ARTICLES

ADAPTIVE SURVIVAL TRIALS

William F. Rosenberger[1,2] *and Padmanabhan Seshaiyer*[1]

[1]Department of Mathematics and Statistics
University of Maryland, Baltimore County
1000 Hilltop Circle
Baltimore, Maryland 21250

[2]Department of Epidemiology and Preventive Medicine
University of Maryland School of Medicine
Baltimore, Maryland 21201-1549

Key words. Adaptive designs; Exponential distribution; Lognormal distribution; Logrank test; Simulation; Weibull distribution

Abstract

We present a design for adaptive survival trials, where the probability of randomization to one of two treatments is skewed away from 0.5 according to the current value of the logrank statistic. A formula mapping the logrank statistic onto [0,1] is given, which is then used to bias a coin used for randomization. Simulation evidence shows that the allocation scheme works well and offers a more ethical alternative when lifetime data are available from other patients during the recruitment period. Power is not adversely affected by the resulting unequal allocation. The usual test statistic appears to be standard normal under the proposed allocation scheme.

1. Introduction

Most phase III clinical trials of two treatments employ an equal allocation scheme. Such schemes are often unattractive to clinicians and volunteers, as they mandate a 50% chance of being assigned to the less effective treatment, provided a treatment effect exists. For many years, adaptive designs have been proposed as a compromise. As data accrue during the course of a clinical trial, this information is used to skew the allocation probabilities to favor the treatment performing better thus far. Recent trials employing adaptive designs (1,2) and several more in the works suggest renewed interest in these designs.

Most of the adaptive designs proposed in the literature, however, are limited to binary responses that are ascertainable fairly quickly (3). Such designs have focused on urn models (4,5), where balls of different colors in the urn represent different treatments. A patient is randomized by drawing a ball at random from the urn. Based on the patient's response (success/failure), balls are added to the urn according to some rule. Successes usually generate balls representing the same treatment, and failures alternate treatments. In this way, the urn composition is skewed to favor balls representing treatments that have been more successful thus far (or less failure-prone).

Other techniques for adaptive designs have been proposed. These include the family of adaptive biased coin designs (6), which change the probability of assignment to treatment after patient or group of patients according to some rule that incorporates previous responses. Recently, a specific adaptive biased coin design has been proposed for continuous outcomes (7). Treating a clinical trial as a two-armed bandit problem has been explored (8), as have Bayesian methods (9).

Many clinical trials are long-term survival trials, with a limited recruitment period, delayed response, censoring, and competing risks. In this paper, we propose an adaptive design for survival trials under these settings. When a patient is ready for randomization, a function of the current value of the logrank statistic is used to bias a coin, which is then used for randomization. This was first proposed, but never explored, by Rosenberger and Lachin (3). In this paper we perform a major simulation study to examine the operating characteristics of the design under different survival distributions.

Previous work on adaptive survival trials has appeared in the literature (10,11). The designs in these papers allocate deterministically based on numbers of deaths in each arm, and hence the designs are not fully randomized. They also employ a sequential stopping rule. Most clinical trials in the United States use a fixed sample design and interim monitoring. Our design assumes a fixed sample size and also skews the allocation probabilities according to the *relative* efficacy of the treatments, making it fully randomized. We recently became aware of a paper (12) that proposes using the Gehan statistic to develop an adaptive strategy for survival trials.

2. Methods

We use the usual Mantel formulation of the unnormalized logrank statistic to develop a mapping onto [0,1] that exceeds 0.5 if treatment A has been doing better thus far, and is less than 0.5 if treatment B has been doing better. Define $\{0 < \tau_1 < \cdots < \tau_K\}$ to be the ordered event times for all patients in the trial (where τ_i is the time from recruitment to event, i.e., allowing for staggered entry). At each event time τ_i, a 2×2 table is constructed with $\delta_i = 1$ if the event occurred on treatment A, 0 otherwise (in the continuous time model, ties are assumed not to occur). Let n_{ji} be the number of patients at risk on treatment j ($j = A, B$) immediately prior to the event time, and let $N_i = n_{Ai} + n_{Bi}$. Then under the hypergeometric model, the numerator of the logrank statistic computed at time τ is given by $\sum_{i=1}^{K(\tau)} \{\delta_i - n_{Ai}/N_i\}$, where, for convenience, we make K a function of τ (i.e., the total number of events occurring up to and including the time of computation, τ). When there is no censoring,

$$-n_A \sum_{i=1}^{K(\tau)} \frac{1}{N-i} \le \sum_{i=1}^{K(\tau)} \{\delta_i - n_{Ai}/N_i\} \le n_B \sum_{i=1}^{K(\tau)} \frac{1}{N-i},$$

where n_j is the total number randomized to treatment j up to time τ and $N = n_A + n_B$. [This bound is conservative if there is censoring.] Let $\mathcal{F}_{\tau-}$ be the history of the events and censorings to τ. Let $X(\tau) = 1$ if a patient is randomized to treatment A at time τ, and 0 if the patient is randomized to treatment B at time τ. We suggest the following mapping:

$$p_{\tau-} \equiv E\{X(\tau) \mid \mathcal{F}_{\tau-}\} = \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{K(\tau-)} \left\{\delta_i - \frac{n_{Ai}}{N_i}\right\}}{\max\{n_A, n_B\} \sum_{i=1}^{K(\tau-)} \frac{1}{N-i}}\right).$$

This idea is similar to that of Rosenberger (7) in dealing with immediate continuous outcomes using a nonparametric rank test. The denominator may be zero early in the trial. In this case, equal allocation should be used until sufficient data accrue (i.e., $p_{\tau-} = 1/2$).

3. Details of the Simulation

A major simulation was performed on a SUN workstation in C++. Random numbers were generated using an efficient algorithm (13) and exponential, lognormal, and Weibull survival times were generated using standard transformations (14). Each simulation was based on 10,000 replications. Both the adaptive allocation

scheme and equal allocation were simulated, so that operating characteristics could be compared.

The simulated clinical trial assumed a limited recruitment period, staggered entry, a fixed duration, and uniform censoring, characteristics of many survival trials. The duration of the trial was assumed to be 1.5936 (in arbitrary units), at which point patients who had not died or were not already censored were considered to be administratively censored. Recruitment was assumed to be uniform over [0,1]. Censoring was assumed to be uniform over [0,1.5936] and was added to a patient's recruitment time. The trial's duration of 1.5936 was chosen because if survival is exponentially distributed with parameter 1, the probability that a patient recruited at time 0 is censored equals 0.5 (15).

Under H_0, the survival distribution for both treatments was assumed to be either exponential with scale parameter 1, Weibull with shape parameter 4 and scale parameter 1, or lognormal with mean 0 and variance 1. Each distribution was examined for $n = 150$. Like censoring, the randomly generated survival time was added to a patient's recruitment time to obtain the patient's time of death. The first occurring time (censoring, death, or end of study) was taken and backed up to start at calendar time 0.

For alternative hypotheses, the control treatment was assumed to have the null distribution, while the parameters of the experimental therapy varied. For exponential survival, the scale parameter was varied; for the Weibull distribution, the shape parameter was fixed at 4 and the scale parameter was allowed to vary; for the lognormal distribution, the variance was assumed to be 1 with the mean varying. We used Latta's suggested values (15) for parameters under the alternative and n was set to yield reasonable power for comparison between adaptive and equal allocation. Simulated power was computed for $\alpha = 0.05$, two-sided.

4. Results

Table 1 examines the behavior of the logrank statistic under H_0. In particular, the proportion of logrank statistics that fell in the 0.01, 0.05, 0.10, and 0.20 tails (two-sided) is given. One can see that, for each distribution, the tail probabilities for adaptive and equal allocation are almost identical. This leads to the conclusion that the test statistic is standard normal under the adaptive allocation scheme under H_0. Table 2 supports this conclusion. The mean value and standard deviation of the 10,000 generated logrank statistics are given, as well as the mean proportion of patients assigned to A, N_A/n (averaged over 10,000 replications) and its standard deviation. The logrank statistic has mean 0 and variance 1 for both allocation rules. Also, the allocation proportions, N_A/n, are close to 1/2, although the variability is clearly larger for adaptive allocation.

Table 1. Proportion of Logrank Statistics Falling in Left (L) or Right (R) Tail of Standard Normal Distribution Under H_0 ($n = 150$) (simulations based on 10,000 replications)

		Equal allocation		Adaptive	
Distribution	Tail probability (both sides)	L	R	L	R
Exponential	0.01	0.004	0.005	0.005	0.006
	0.05	0.023	0.026	0.026	0.026
	0.10	0.046	0.049	0.051	0.052
	0.20	0.098	0.101	0.101	0.103
Weibull	0.01	0.004	0.005	0.005	0.006
	0.05	0.022	0.025	0.027	0.025
	0.10	0.046	0.049	0.049	0.052
	0.20	0.096	0.098	0.099	0.100
Lognormal	0.01	0.005	0.004	0.005	0.006
	0.05	0.026	0.025	0.025	0.025
	0.10	0.049	0.051	0.051	0.052
	0.20	0.101	0.100	0.102	0.102

Table 2. Simulated Operating Characteristics Under H_0 (based on 10,000 replications)

Distribution	Allocation rule	Mean (N_A/n)	SD (N_A/n)	Mean (logrank)	SD (logrank)
Exponential	Adaptive	0.499	0.085	0.001	1.01
	Equal	0.500	0.041	0.010	0.99
Weibull	Adaptive	0.498	0.101	0.001	1.00
	Equal	0.500	0.041	0.003	0.99
Lognormal	Adaptive	0.500	0.083	0.001	1.01
	Equal	0.500	0.041	−0.002	1.00

Table 3 explores operating characteristics under various alternatives. Again the mean and standard deviation of N_A/n are given. Also given is the simulated power (the proportion of test statistics falling in the rejection region). One can see a clear advantage in terms of allocation proportions for adaptive allocation. In each case, 10–20% more patients are assigned to the better treatment under adaptive allocation. Power differs by only 1% for each alternative examined. One can conclude that the unbalanced allocation induced by the adaptive design does not sacrifice power.

Other simulations, besides the few reported here, were performed and similar results were observed.

5. Conclusions

Clinical trials necessitate continual compromises between individual and collective ethics (16). The adaptive design presented in this paper will allow the demands of

Table 3. Simulated Operating Characteristics Under H_A (based on 10,000 replications)

Distribution	Parameter	n	Allocation rule	Mean (N_A/n)	SD (N_A/n)	Simulated power
Exponential	1.6	150	Adaptive	0.569	0.087	0.49
			Equal	0.500	0.041	0.50
	0.625	150	Adaptive	0.430	0.075	0.61
			Equal	0.500	0.041	0.62
	1.2	800	Adaptive	0.532	0.046	0.46
			Equal	0.500	0.018	0.46
	0.833	800	Adaptive	0.467	0.043	0.62
			Equal	0.500	0.018	0.63
Weibull	1.6	30	Adaptive	0.558	0.114	0.73
			Equal	0.501	0.091	0.72
	0.625	30	Adaptive	0.390	0.114	0.98
			Equal	0.501	0.091	0.98
Lognormal	0.4	400	Adaptive	0.611	0.049	0.60
			Equal	0.500	0.025	0.60
	−0.4	400	Adaptive	0.406	0.047	0.63
			Equal	0.500	0.025	0.64

collective ethics to be satisfied. In particular, the design is fully randomized, provides minimal loss of power, allows for standard analyzes using the logrank statistic, and ensures a control group of adequate size, which can then be used for convincing follow-up studies. However, the gains toward individual ethics are substantial. For the cases we studied, between 10 and 20% of patients will be assigned to a more effective treatment, if one exists.

The nearly identical power under an adaptive design fits well with the conclusions of Sposto and Krailo (17), who found the logrank statistic to be equally or even slightly more powerful when there are modest imbalances. In fact, they found that as the treatment difference becomes larger, the larger the allowable imbalance becomes to attain similar power. This is precisely the scenario in which adaptive methods are most attractive.

Our design is relatively conservative, in that allocation proportions were not skewed more than 2:1 in any of the cases we explored. As stated above, this will allow for a control group of adequate size, if follow-up studies are necessary. However, the less conservative statistician may wish to see larger imbalances, particularly if the disease is especially grave. One could use a weighted logrank statistic to increase the imbalance, such as one that weights early deaths more heavily. One must be careful to determine if power will be impacted by a less conservative scheme.

Patient response is often confounded with covariates unrelated to treatment. It would be nice to adapt on the basis of an adjusted treatment effect to account for covariates deemed a priori to be important. The procedure described in this paper could be modified by using the score statistic for the treatment effect from a Cox

model rather than the logrank test in the mapping. In the case of no covariates, the score statistic is equivalent to the logrank test (with no censoring) (18). When a patient is recruited, the score statistic (unnormalized) from a Cox model is computed, and mapped to [0,1]. Such a model-based approach can also be used to adjust for any time trend in characteristics of recruited patients (19). The inability to address time trends has been identified as a pivotal drawback of adaptive designs (20).

An underlying assumption to all research in adaptive designs is that some information on previous patients' responses be available at recruitment. This is not the case in some long-term survival trials, where the recruitment period ends before any responses become available. The methods discussed in this paper are not applicable in that scenario. Extending the recruitment period for these trials may be possible and desirable in order to apply this methodology. In some studies where the primary outcome is not ascertainable in a reasonable time, it may be possible to use a surrogate measure on which to adapt. For example, in the fluoxetine trial (2), score on the Hamilton Depression Scale was used as a surrogate, and the adaptive design was based on those scores. It is not clear how to analyze the primary outcome variable when a surrogate is used for adaptation, but this is a viable way of adapting when it is not possible to use the primary outcome.

A number of interesting theoretical questions are evident, but will not be addressed in this paper. These include:

1. Finding the asymptotic properties of N_A/n.
2. Rigorously proving the asymptotic normality of the logrank statistic under a dependent allocation scheme, along with finding the correct permutational variance.
3. Extending item (2) to account for sequentially computed logrank statistics arising from an interim monitoring plan.
4. Exploring early stopping in the context of a dependent allocation scheme. Potentially, monitoring N_A/n could be used to determine a stopping rule.

These are very challenging questions, and likely will involve counting processes (21). However, for the practitioner, our simulations show that the design used should be of little consequence in the analysis.

Finally, we note that simulations are a very useful tool in designing appropriate studies. To facilitate this, our simulation programs can be found on the internet via ftp://ftp.math.umbc.edu/pub/padhu/Patient.tar.

Acknowledgments

This research was supported by Grant R29-DK51017-02 from the National Institute of Diabetes and Digestive and Kidney Diseases. The authors thank Ram V. Krishan for helpful suggestions on the data structures used in the simulation.

References

1. Rout CC, Rocke DA, Levin J, Gouws E, Reddy D: A reevaluation of the role of crystalloid preload in the prevention of hypotension associated with spinal anesthesia for elective cesarean section. *Anesthesiology* 79:262–269, 1993.
2. Tamura RN, Faries DE, Andersen JS, Heiligenstein JH: A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *J Am Statist Assoc* 89:768–776, 1994.
3. Rosenberger WF, Lachin JM: The use of response-adaptive designs in clinical trials. *Controlled Clin Trials* 14:471–484, 1993.
4. Zelen M: Play the winner rule and the controlled clinical trial. *J Am Statist Assoc* 64:131–146, 1969.
5. Wei LJ, Durham S: The randomized play-the-winner rule in medical trials. *J Am Statist Assoc* 73:840–843, 1978.
6. Wei LJ: The adaptive biased coin design for sequential experiments. *Ann Statist* 6:92–100, 1978.
7. Rosenberger WF: Asymptotic inference with response-adaptive treatment allocation designs. *Ann Statist* 21:2098–2107, 1993.
8. Berry DA, Fristedt B: *Bandit Problems: Sequential Allocation of Experiments*. Chapman and Hall, New York, 1985.
9. Cornfield J, Halperin M, Greenhouse SW: An adaptive procedure for sequential clinical trials. *J Am Statist Assoc* 64:759–770, 1969.
10. Flehinger BJ, Louis TA: Sequential treatment allocation in clinical trials. *Biometrika* 58:419–426, 1971.
11. Louis TA: Sequential allocation in clinical trials comparing two exponential survival curves. *Biometrics* 33:627–634, 1977.
12. Yao Q, Wei LJ: Play the winner for phase II/III clinical trials. *Statist Med* 15:2413–2423, 1996.
13. L'Ecuyer P: Efficient and portable combined random number generators. *Commun ACM* 31:742–751, 1988.
14. Law AM, Kelton WD: *Simulation Modeling and Analysis*. McGraw-Hill, New York, 1982.
15. Latta RB: A Monte Carlo simulation of some two-sample rank tests with censored data. *J Am Statist Assoc* 76:713–719, 1981.
16. Palmer CR: Ethics and statistical methodology in clinical trials. *J Med Ethics* 19:219–222, 1993.
17. Sposto R, Krailo MD: Use of unequal allocation in survival trials. *Statist Med* 6:119–125, 1987.
18. Kalbfleisch JD, Prentice RL: *The Statistical Analysis of Failure Time Data*. Wiley, New York, 1980.
19. Coad DS: Sequential tests for an unstable response variable. *Biometrika* 78:113–121, 1991.
20. Altman DG, Royston JP: The hidden effect of time. *Statist Med* 7:629–637, 1988.
21. Andersen PK, Borgan O, Gill RD, Keiding N: *Statistical Models Based on Counting Processes*. Springer, New York, 1992.