University of Texas, MD Anderson Cancer Center

UT MD Anderson Cancer Center Department of Biostatistics
Working Paper Series

Year 2006 *Paper* 27

Understanding the Exponential Tuning Parameter in Adaptively Randomized Trials

John D. Cook*

*M. D. Anderson Cancer Center, cook@mdanderson.org

http://biostats.bepress.com/mdandersonbiostat/paper27

Understanding the exponential tuning parameter in adaptively randomized trials

John D. Cook*

December 13, 2006

Abstract

We examine the effect of a parameter λ used to calibrate how responsive randomization probabilities are to observed data in an adaptively randomized clinical trial. We define and motivate the parameter λ and demonstrate how varying this parameter effects the operating characteristics of example clinical trial designs.

1 Introduction

Let A and B be two treatments in an adaptively randomized trial, and let θ_i be the probability of response on arm i where i is A or B. The simplest adaptive randomization scheme is to assign arm i with probability

$$p = P(\theta_i > \theta_j \,|\, \text{data}).$$

*University of Texas M. D. Anderson Cancer Center, cook@mdanderson.org

A generalization of this approach is to raise this inequality probability to some power and renormalize, assigning arm i with probability

$$f(p, \lambda) = \frac{p^\lambda}{p^\lambda + (1-p)^\lambda}$$

for a given $0 \leq \lambda \leq \infty$. When $\lambda = 1$, $f(p, \lambda)$ is simply p.

The parameter λ is called a tuning parameter because it controls the degree to which the randomization probabilities are influenced by the data. When $\lambda = 0$, the method reduces to equal randomization. As $\lambda \to \infty$ the method approaches myopic optimization, assigning each patient the arm which looks best at the time the patient enrolls.

We first explore some of the elementary properties of the function $f(p, \lambda)$. Then we illustrate how λ effects the operating characteristics of a clinical trial. We will show that generally as λ increases, the number of patients treated on the superior arm increases while the probability of correctly selecting the superior arm decreases. However, for small values of λ, the number of patients on the superior arm increases rapidly while statistical power decreases only slowly.

The simulation results presented in this paper were computed using the Adaptive Randomization software available at [1].

2 Basic properties

To understand how the function f behaves, we start with a simple example. Suppose $P(\theta_A > \theta_B \,|\, \text{data}) = 0.7$ given the current data. If $\lambda = 1$, arm A would be assigned with probability 0.7. If $\lambda = 1/2$, arm A would be assigned with probability 0.60, and if $\lambda = 2$, arm A would be assigned with probability 0.84.

We now make some more general observations about the function f. First

of all, $f(1/2, \lambda) = 1/2$ for all values of λ. If two arms are equally effective, they will be assigned with equal probability, regardless of the tuning parameter value.

Next, note that for any p, $f(p, 0) = 1/2$. This means that setting $\lambda = 0$ corresponds to equal randomization: arms A and B are each assigned with probability 1/2 regardless of $P(\theta_i > \theta_j \,|\, \text{data})$. In this case the "adaptive" randomization algorithm is not adaptive at all.

Finally, note that $\lim_{\lambda \to \infty} f(p, \lambda)$ is 0 if $p < 1/2$ and 1 if $p > 1/2$. This says that setting $\lambda = \infty$ corresponds to myopic optimization: the arm with higher probability of response is assigned with probability 1. In this case the adaptive "randomization" is actually deterministic.

The above observations show that our tuning parameter could be called an *interpolation* parameter, interpolating between equal randomization and myopic optimization. For small values of λ, adaptive randomization behaves more like equal randomization and for large values of λ, adaptive randomization behaves more like myopic optimization. In practice, moderate values such as $\lambda = 1$ can give the best of both worlds: power and bias comparable to equal randomization, while also assigning significantly more patients to the superior arm on average. See [2] for analysis of adaptive randomization trials using $\lambda = 1$.

We give a few graphs illustrating how $f(p, \lambda)$ behaves between the extremes discussed above. First, we set $\lambda = 1/2$ and show how $f(\cdot, 1/2)$ increases small probabilities and decreases large probabilities in Figure 1, adding the line $y = x$ for comparison. Next, we set $\lambda = 2$ and show in Figure 2 how $f(\cdot, 2)$ makes small probabilities smaller and makes large probabilities larger. Finally, in Figures 3 and 4 we show the effect of varying values of λ for $p = 0.3$ and $p = 0.7$.

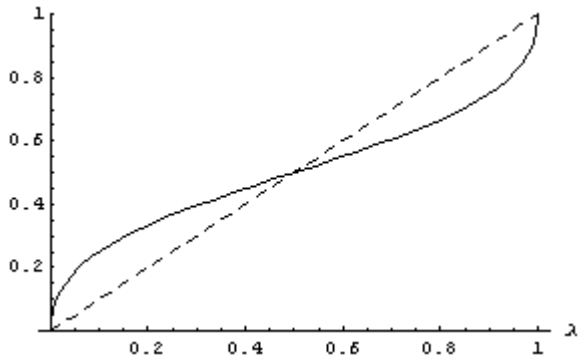

Figure 1: $\lambda = 1/2$, p varying

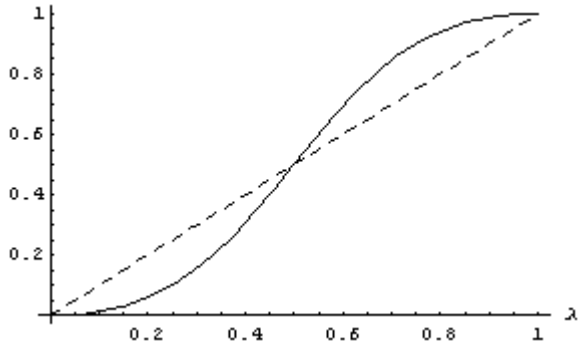

Figure 2: $\lambda = 2$, p varying

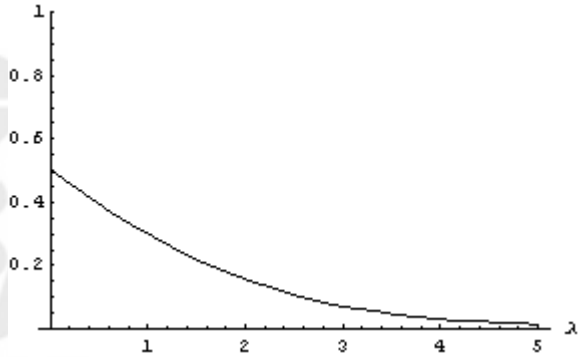

Figure 3: $p = 0.3$, λ varying

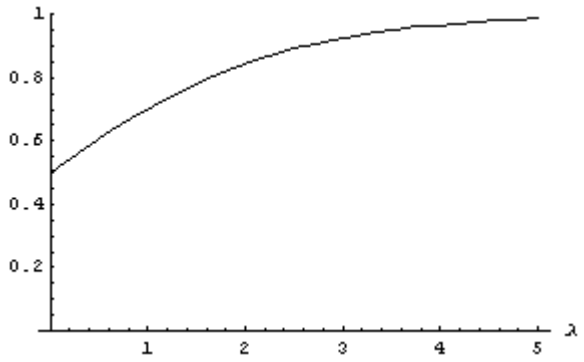

Figure 4: $p = 0.7$, λ varying

3 Operating characteristics

To illustrate the effect of λ on operating characteristics, we simulated a two-arm trial with a maximum of 80 patients. We assumed the response probabilities θ_i are distributed *a priori* as beta(0.6, 1.4). If at any point in the trial

$$P(\theta_i > \theta_j \,|\, \text{data}) > 0.95$$

we end the trial, selecting treatment i as the superior treatment. If we reach the maximum number of patients without either arm satisfying the above inequality, we declare the trial inconclusive. We simulated the trial 10,000 times for each value of λ from 0 to 4 in increments of 0.1 and present the average behavior.

In each scenario, we assume the true probability of response on arm 1 is 0.2. In Scenario 1, the true probability of response on arm 2 is 0.3. In Scenario 2, the true probability of response on arm 2 is 0.4.

We assume patient responses are recorded immediately. It would be more realistic to model the arrival of patients as a Poisson process and a delay in observing patient outcomes. (The Adaptive Randomization software from [1]

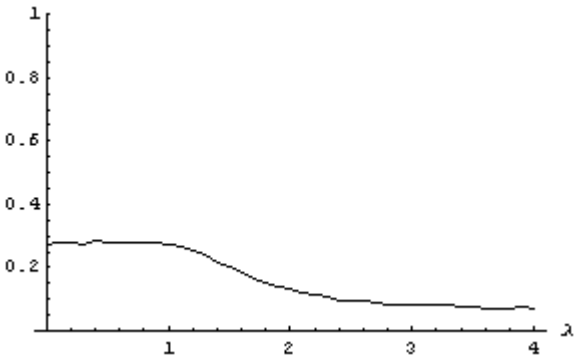

Figure 5: Scenario 1, no stopping rule, correct selection probability

supports such simulations.) However, arrival rates and observation windows are peculiar to a specific clinical trial. We avoid these issues in order to present generic examples.

We first present the results for the design without a stopping rule, accruing the maximum enrollment 80 patients in each simulation. This is intended, not to recommend trials without stopping rules, but to show the limiting behavior of as stopping rules become more difficult to satisfy. We follow this by including the stopping rule above. This rule stops trials rather often and so is complementary to the example without a stopping rule.

3.1 Simulation results without early stopping rule

Figure 5 shows the probability of correctly selecting the superior arm under Scenario 1 as λ varies. The correct selection probability is essentially constant for λ between 0 and 1 then degrades for larger values. For $\lambda = \infty$ the correct selection probability is 0.052.

Figure 6 shows the expected number of patients treated on the superior arm

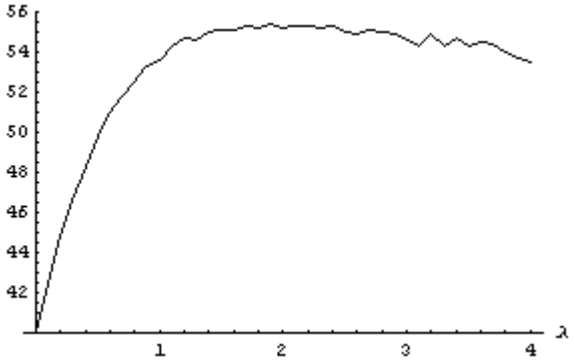

Figure 6: Scenario 1, no stopping rule, patients on superior arm

also for Scenario 1. Since there is no stopping rule, the balance of the 80 patients would be treated on the inferior arm. Aside from simulation noise, the curve rises rapidly to a peak and then slowly declines. The number of patients on the superior arm with $\lambda = \infty$ is 51.0.

Figures 7 and 8 are the corresponding results under Scenario 2. The correct selection probability continues to decline for larger values of λ, reaching 0.115 at $\lambda = \infty$. As in Scenario 1, the number of patients on the superior arm rises rapidly with λ, reaches a plateau and slowly declines. For $\lambda = \infty$, the average number of patients on the superior arm is 59.1.

3.2 Simulation results with early stopping rule

We now add the possibility of early stopping and consider the same two scenarios as above.

Figure 9 shows the probability of correctly selecting the superior arm under Scenario 1 as λ varies. At $\lambda = \infty$ the selection probability is 0.135.

Figure 10 shows the expected number of patients treated on the superior

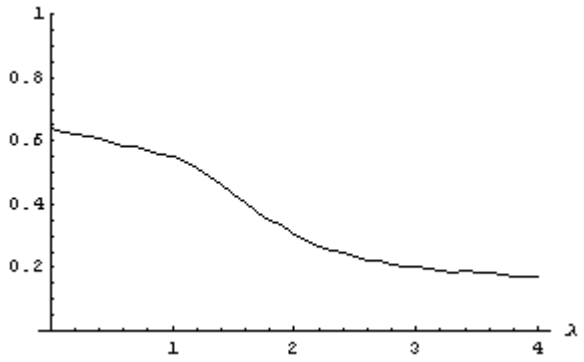

Figure 7: Scenario 2, no stopping rule, correct selection probability

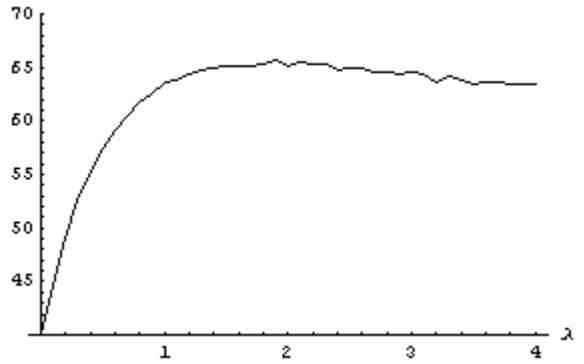

Figure 8: Scenario 2, no stopping rule, patients on superior arm

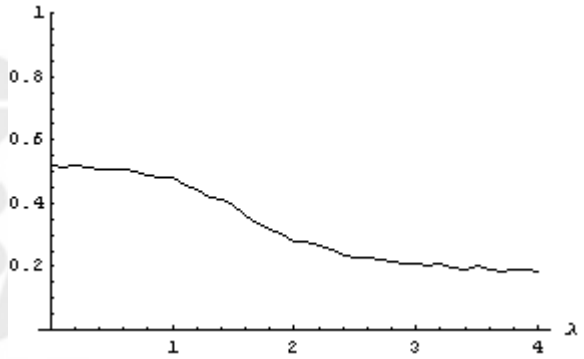

Figure 9: Scenario 1 with stopping rule, correct selection probability

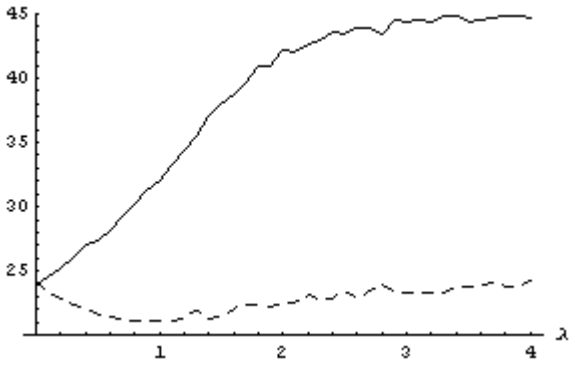

Figure 10: Scenario 1 with stopping rule, patients on each arm

arm also for Scenario 1. The dashed line corresponds to the inferior arm and the solid line corresponds to the superior arm. The number of patients assigned to the inferior arm reaches a minimum around $\lambda = 1$ and then increases. The number assigned to the superior arm increases to a plateau and slowly declines. At $\lambda = \infty$, an average of 27.9 patients are treated on the inferior arm and 43.1 on the superior arm.

Figures 11 and 12 are the corresponding results under Scenario 2. The correct selection probability declines to 0.244 at $\lambda = \infty$. In Figure 12, as in Figure 10 above, the dashed line corresponds to the inferior arm and the solid line corresponds to the superior arm. As in Scenario 1, the number of patients on the inferior arm reaches a minimum and then increases. The number of patients on the superior arm rises sharply as before, but appears approximately constant once it reaches a maximum. At $\lambda = \infty$, 19.7 patients would be treated on the inferior arm and 45.0 on the superior arm.

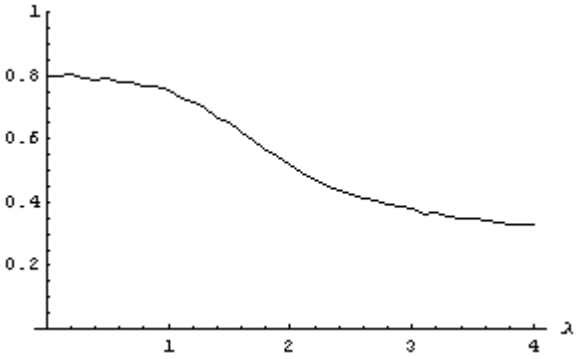

Figure 11: Scenario 2 with stopping rule, correct selection probability

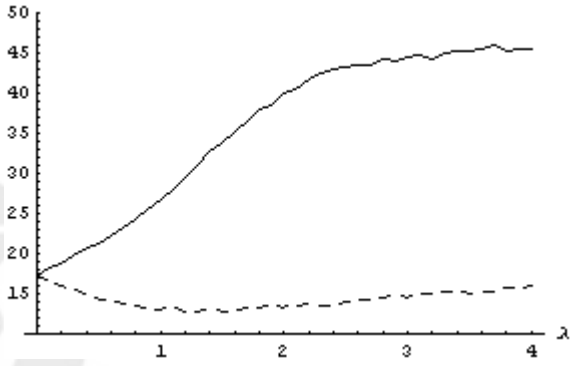

Figure 12: Scenario 2 with stopping rule, patients on each arm

4 Discussion

Adaptive randomization designs range between equal randomization when $\lambda = 0$ and myopic optimization when $\lambda = \infty$. The examples given here suggest that as λ increases, the probability of correctly selecting the superior arm decreases, though slowly at first. The number of patients treated on the superior arm increases rapidly, reaches a plateau, and slowly declines. Values of λ near 1 result in power comparable to that of equal randomization while assigning substantially more patients to the superior treatment. In some cases the number of patients assigned to the superior treatment could be increased still more by using a somewhat larger value of λ, though at a substantial loss of power. Very large values of λ have little to recommend them since one may achieve more power and assign more patients to the better arm by using a moderate value.

While the number of patients assigned to the superior arm generally increases as λ increases, it does not follow that the number of patients assigned to the inferior decreases since the expected sample size also increases with λ.

In our experience, the most commonly used value of λ is 1, though we have seen values of λ as large as 2. Some trials use a smaller value such as $\lambda = 1/2$. It is common to have a burn-in period of equal randomization before starting to imbalance the randomization probabilities. In this case, the tuning parameter is actually a function λ of the patient accrual number n. If N is the number of patients to randomize equally then $\lambda(n) = [n > N]\lambda_0$ where λ_0 is the nominal tuning parameter. Wathen [3] has proposed allowing λ to increase gradually as a function of n rather than as a step function.

The choice of λ must be made in the context of a specific clinical trial and evaluated by a simulation study. We suggest one begin by comparing the

results of simulations using $\lambda = 0$ and $\lambda = 1$. Then the results presented here will suggest whether one should consider values of λ greater or less than 1.

References

[1] M. D. Anderson Cancer Center Department of Biostatistics software download site. http://biostatistics.mdanderson.org/SoftwareDownload/

[2] J. Kyle Wathen and John Cook (2006), *Power and bias in adaptively randomized clinical trials* Available at

http://www.mdanderson.org/pdf/biostats_utmdabtr_002_06.pdf

[3] J. Kyle Wathen (2004), personal correspondence.