University of Texas, MD Anderson Cancer Center

UT MD Anderson Cancer Center Department of Biostatistics
Working Paper Series

Year 2007 *Paper* 32

Comparing Methods of Tuning Adaptively Randomized Trials

John D. Cook*

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

http://biostats.bepress.com/mdandersonbiostat/paper32

Comparing Methods of Tuning Adaptively Randomized Trials

John D. Cook

Abstract

The simplest Bayesian adaptive randomization scheme is to randomize patients to a treatment with probability equal to the probability p that the treatment is better. We examine three variations on adaptive randomization which are used to compromise between this scheme and equal randomization. The first variation is to apply a power transformation to p to obtain randomization probabilities. The second is to clip p to live within specified lower and upper bounds. The third is to begin the trial with a burn-in period of equal randomization. We illustrate how each approach effects statistical power and the number of patients assigned to each treatment. We conclude with recommendations for designing adaptively randomized clinical trials.

Comparing methods of tuning adaptively randomized trials

John D. Cook*

January 9, 2007

Abstract

The simplest Bayesian adaptive randomization scheme is to randomize patients to a treatment with probability equal to the probability p that the treatment is better. We examine three variations on adaptive randomization which are used to compromise between this scheme and equal randomization. The first variation is to apply a power transformation to p to obtain randomization probabilities. The second is to clip p to live within specified lower and upper bounds. The third is to begin the trial with a burn-in period of equal randomization. We illustrate how each approach effects statistical power and the number of patients assigned to each treatment. We conclude with recommendations for designing adaptively randomized clinical trials.

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

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 Bayesian adaptive randomization scheme is to assign treatment i with probability

$$p = P(\theta_i > \theta_j \,|\, \text{data}). \tag{1}$$

We refer to this approach as simple adaptive randomization (SAR).

In [1] we explored the effect of the power transformation

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

for a given $\lambda \geq 0$. That is, rather than assigning treatment i with probability p, we first calculate p and give treatment i with probability $f(p, \lambda)$. We refer to this approach as PT(λ). Note that $f(p, 1) = p$ and so PT(1) = SAR. Also, $f(p, 0) = 1/2$ for all $0 < p < 1$ and so when $\lambda = 0$ the randomization scheme reduces to equal randomization (ER), that is, PT(0) = ER. We show in [1] that as we vary λ between 0 and 1 the operating characteristics of the adaptive randomization scheme change continuously between those of ER and SAR.

Another approach to compromising between ER and SAR is to simply clip the randomization probabilities by setting a minimum randomization probability $r \leq 1/2$. We refer to this approach as Clip(r). If p falls below r, treatment i is assigned with probability r. By symmetry, this implies that if the probability p rises above $1-r$ then the treatment is given with probability $1-r$. Moderate values of p, those between r and $1-r$, are left unchanged. Explicitly, the Clip(r) assigns treatment i with probability

$$g(p, r) = \max\left(r, \min(p, 1-r)\right).$$

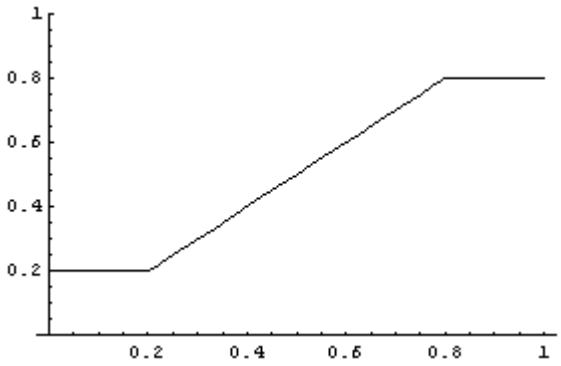

Figure 1: Clipping transformation $g(p, 0.2)$

Figure 1 graphs the transformation g for $r = 0.2$. Note that Clip(1/2) = ER and Clip(0) = SAR.

The final variation on adaptive randomization we consider is the use of a burn-in period of equal randomization. Let N be the maximum accrual of a trial in which the first n patients are randomized with probability 1/2 and the remaining $N - n$ are randomized to treatment i with probability p defined in equation 1. We denote this design by Burn(n/N). Note that Burn(1) = ER and Burn(0) = SAR.

The designs PT(t), Clip($(1-t)/2$), and Burn($1-t$) each interpolate between ER and SAR as t varies between 0 and 1. In the next section we will examine the operating characteristics of each family of designs as a function of t.

The simulation results presented in this paper were computed using the Adaptive Randomization software[1] available at [2].

[1]Most of the simulations in this report may be carried out with version 3.2.2 of the Adaptive Randomization software available at the time of writing. However, some simulations require extensions that will be included in the next version of the software.

2 Operating characteristics

To explore the operating characteristics of the three families of designs, we simulated a two-arm trial with a maximum of 80 patients. We assume the response probabilities θ_i are distributed *a priori* as beta(0.6, 1.4). We examine three scenarios. In each we assume the true probability of response on arm 1 is 0.2. The probabilities of response on arm 2 are 0.3, 0.4, and 0.5. We simulated each design 10,000 times for each value of t from 0 to 1 in increments of 0.1 and present the average behavior.

We simulated with and without an early stopping rule. When we apply a stopping rule, we stop early if at any point in the trial

$$P(\theta_i > \theta_j \,|\, \text{data}) > 0.95 \tag{2}$$

in which case we select 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 first present the results for the design without a stopping rule, accruing the maximum enrollment 80 patients in each simulation, and then present the results adding the stopping rule.

In each graph, t varies from 0 to 1 along the horizontal axis. The solid blue lines are for PT(t), the finely dashed red lines correspond are for Burn($1-t$), and the coarsely dashed black lines are for Clip($(1-t)/2$).

2.1 Simulation results without early stopping rule

Figure 2 shows the probability of concluding arm 2 is superior under Scenario 1 ($\theta_2 = 0.3$) as $\lambda = t$ varies. Figures 3 and 4 show the corresponding probabilities under Scenario 2 ($\theta_2 = 0.4$) and Scenario 3 ($\theta_2 = 0.5$) respectively.

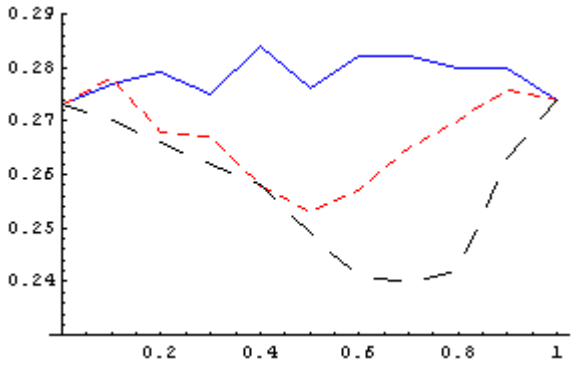

Figure 2: Scenario 1, no stopping rule, correct selection probability

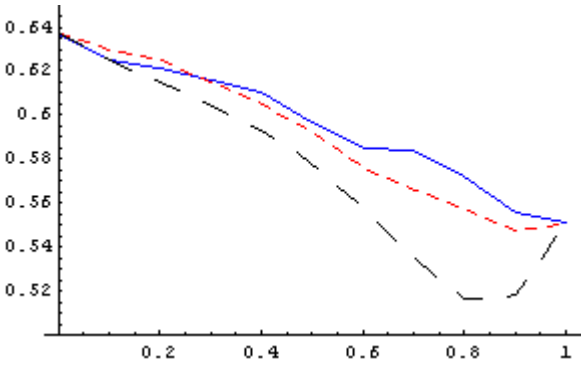

Figure 3: Scenario 2, no stopping rule, correct selection probability

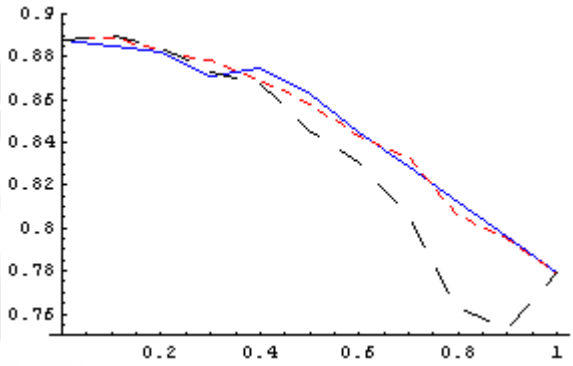

Figure 4: Scenario 3, no stopping rule, correct selection probability

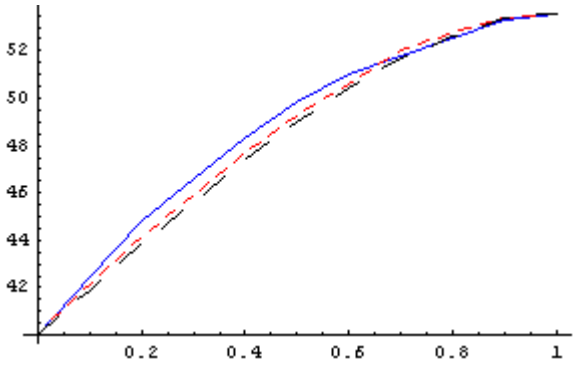

Figure 5: Scenario 1, no stopping rule, patients on superior arm

Note that the curve for PT(t) is generally on top. That is, the power transformation design generally reaches the correct conclusion more often than the corresponding burn-in or clipping design.

Note also that the curves for PT(t) and Burn($1-t$) are essentially monotone decreasing in Scenarios 2 and 3, apart from simulation noise, whereas the curves for Clip($(1-t)/2$) has a pronounced local minimum. This suggests that a small amount of clipping may result in less power than no clipping.

Figures 5, 6, and 7 show the average number of patients assigned to the superior arm under Scenarios 1, 2, and 3 respectively.

Note that each power transformation design puts more patients on the superior arm than corresponding burn-in or clipping design. The advantage of the power transformation increases as the probability of response on arm 2 increases. Note also that the curves for number of patients on the superior arm are smoother than the corresponding curves for probability of selecting the superior arm.

The following scatter plots, Figures 8-10, summarize the relationship be-

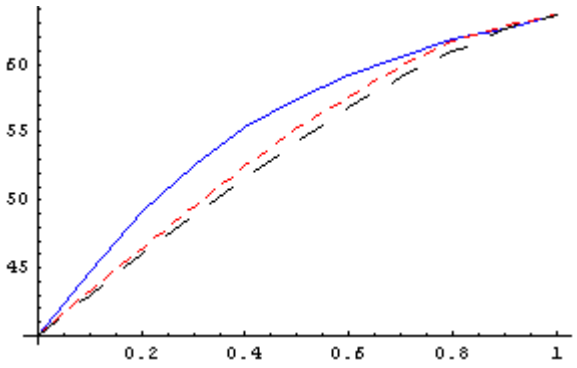

Figure 6: Scenario 2, no stopping rule, patients on superior arm

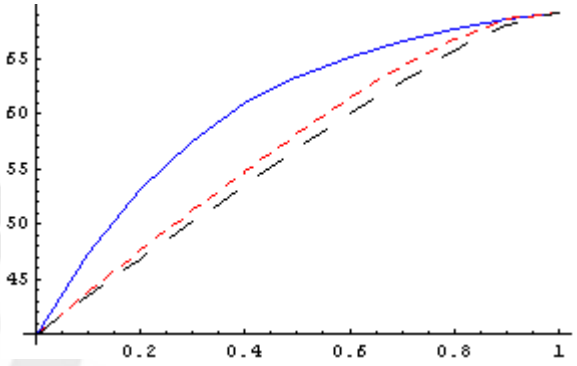

Figure 7: Scenario 3, no stopping rule, patients on superior arm

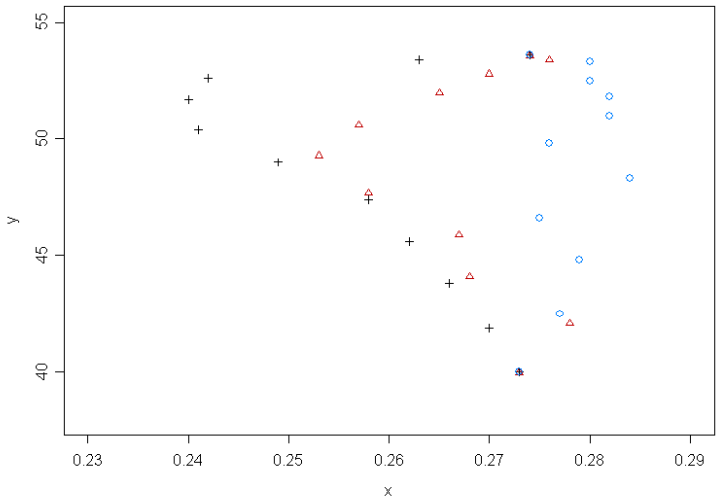

Figure 8: Scenario 1, no stopping rule, selection vs patients on superior arm

tween selection probability and number of patients on the superior arm for each scenario. The blue circles represent the power transformation, red triangles represent burn-in, and black crosses represent clipping.

2.2 Simulation results with early stopping rule

In this section we repeat the simulations of the previous section with the addition of the stopping rule given in equation 2.

Figures 11, 12, and 13 show the probabilities of correctly concluding that arm 2 is superior under Scenarios 1, 2, and 3 respectively.

The addition of the stopping rule leads to more jagged operating characteristic curves. However, the power transformation designs generally continue to select the superior arm more often than the burn-in or clipping designs.

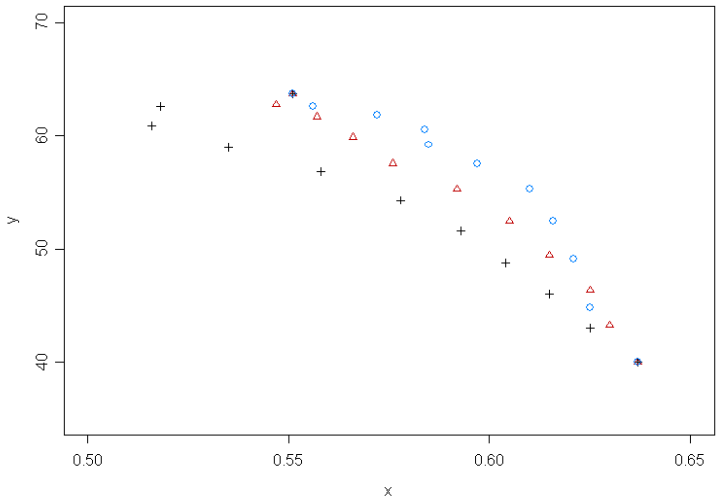

Figure 9: Scenario 2, no stopping rule, selection vs patients on superior arm

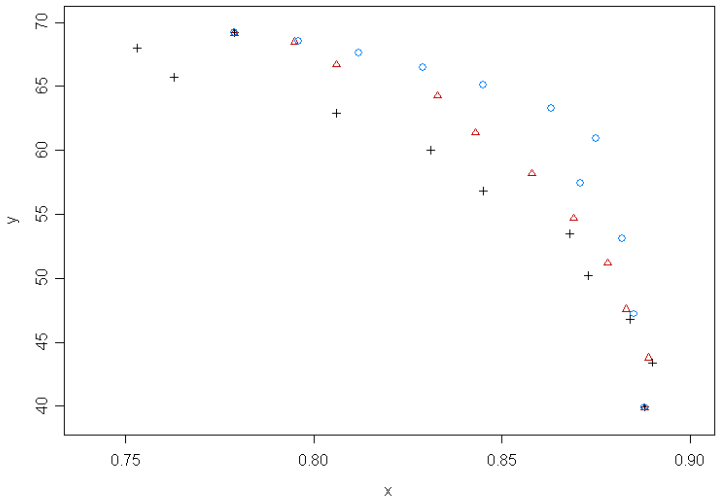

Figure 10: Scenario 3, no stopping rule, selection vs patients on superior arm

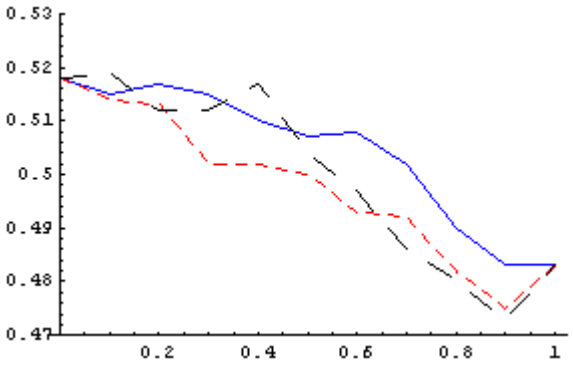

Figure 11: Scenario 1, early stopping rule, correct selection probability

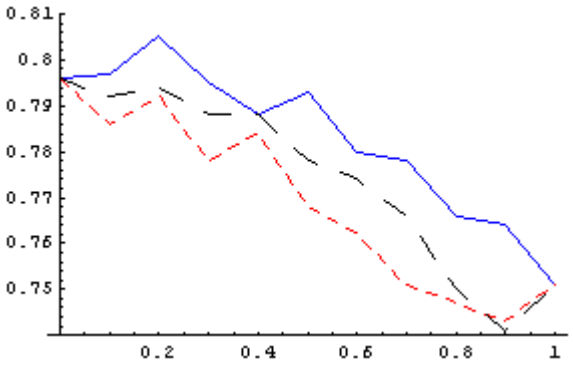

Figure 12: Scenario 2, early stopping rule, correct selection probability

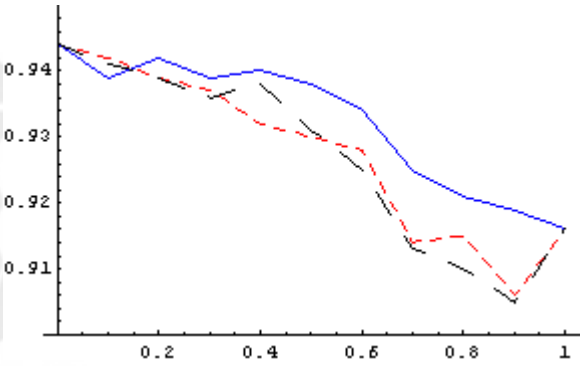

Figure 13: Scenario 3, early stopping rule, correct selection probability

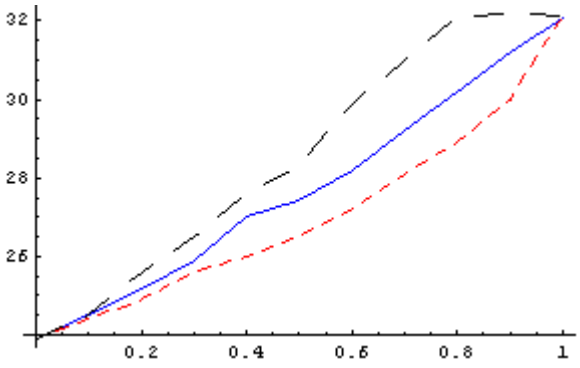

Figure 14: Scenario 1, no stopping rule, patients on superior arm

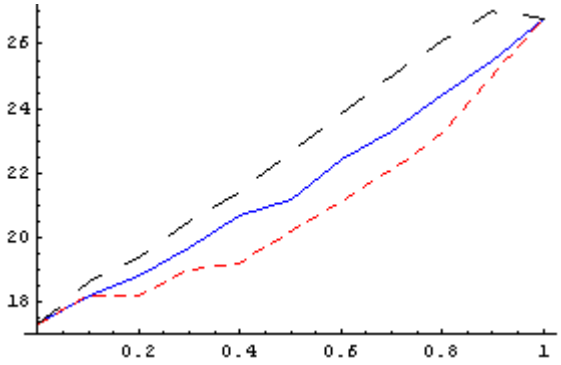

Figure 15: Scenario 2, no stopping rule, patients on superior arm

Figures 14, 15, and 16 show the average number of patients assigned to the superior arm under Scenarios 1, 2, and 3 respectively.

The curves representing the number of patients assigned to the superior arm are quite different when a stopping rule is added. Without a stopping rule, the power transformation approach consistently put the most patients on the superior arm. Burn-in and clipping were close to each other, with burn-in doing slightly but consistently better. But now using a stopping rule, the clipping method places the most patients on the superior arm, followed by power

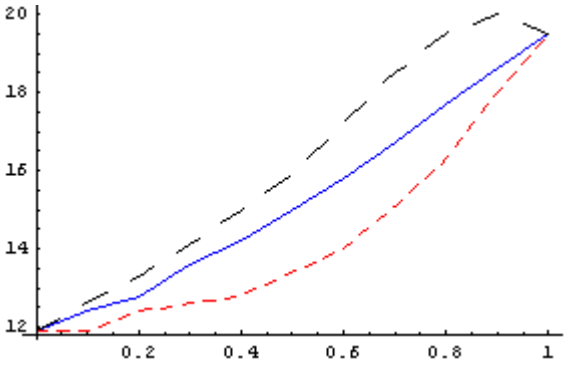

Figure 16: Scenario 3, no stopping rule, patients on superior arm

transformation, and then by burn-in.

The following scatter plots, Figures 17 - 19, summarize the relationship between selection probability and number of patients on the superior arm. As before, the blue circles represent the power transformation, red triangles represent burn-in, and black crosses represent clipping.

3 Discussion

Power transformation, burn-in, and clipping are three generalizations of the simplest Bayesian adaptive randomization design. Each contains a parameter that can be used to create designs whose operating characteristics are intermediate between those of ER and SAR. In each method, the probability of correctly selecting the better arm at the end of the trial increases as one gets closer to ER. The expected number of patients treated on the superior arm increases as one gets closer to SAR. One may use any one of these methods to design an adaptively randomized trial with properties somewhere between those of ER and SAR, according to one's trade-off between statistical power and treating

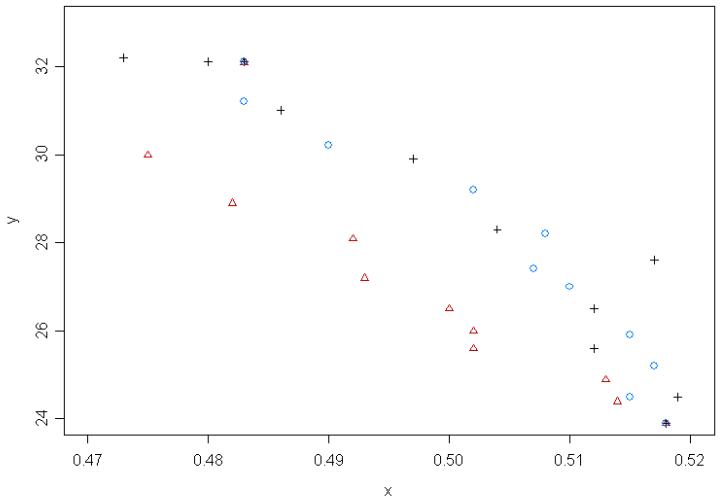

Figure 17: Scenario 1, no stopping rule, selection vs patients on superior arm

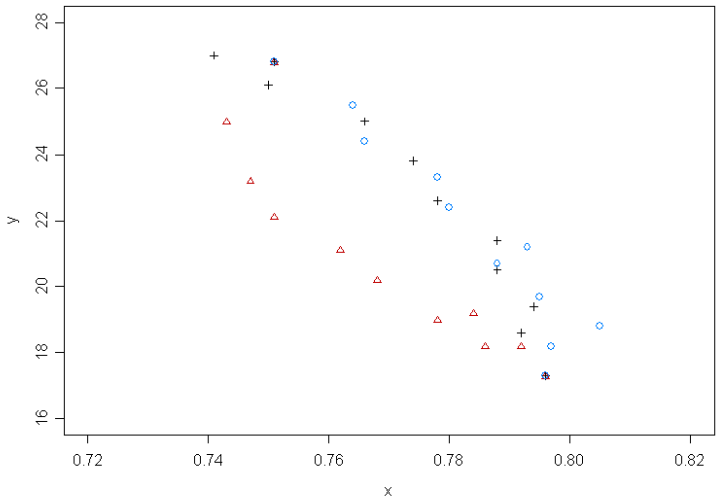

Figure 18: Scenario 2, no stopping rule, selection vs patients on superior arm

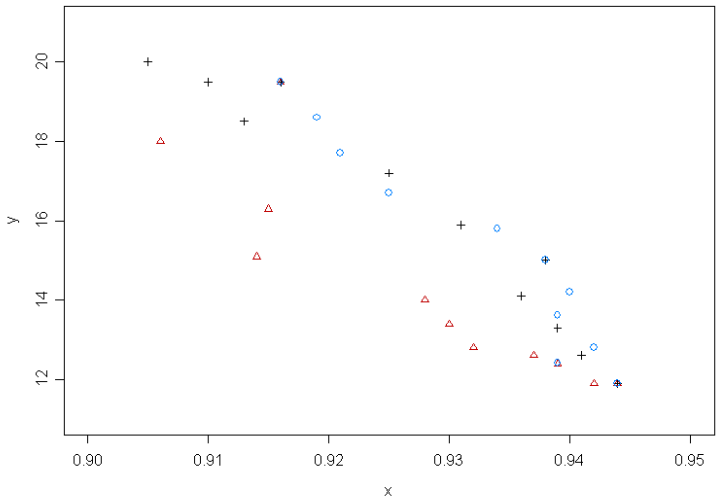

Figure 19: Scenario 3, no stopping rule, selection vs patients on superior arm

patients in the trial most effectively.

The simulation results presented in this report suggest that of these three approaches, the power transformation approach puts the most patients on the superior arm for a given statistical power. Absent a stopping rule, the power transformation clearly performs better. With the stopping rule examined here, the power transformation and clipping appear to do roughly equally well, with burn-in doing worse. The power transformation also has the advantage of being more general; this report has only examined designs of the form PT(λ) with $\lambda \leq 1$ though designs with $\lambda > 1$ are possible. See [1] for an exploration of the operating characteristics of PT(λ) designs over the range $0 \leq \lambda \leq \infty$.

We would suggest one begin exploring clinical trial designs by simulating a PT(0) and a PT(1) design. The PT(0) design will give the most statistical

power. The PT(1) design will give an idea of how many more patients can be assigned to the superior arm, and at what loss in power, under given scenarios. To assign even more patients to the superior arm than the PT(1) design, consider PT(2). If one wants a combination of power and treatment imbalance somewhere between PT(0) and PT(1), linear interpolation on λ will give an initial guess at an acceptable value of λ. One may also wish to simulate hybrid designs, for example, using a burn-in period with a PT(λ) design.

References

[1] John D. Cook, "Understanding the Exponential Tuning Parameter in Adaptively Randomized Trials" (December 2006). *UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series.* Working Paper 27. `http://www.bepress.com/mdandersonbiostat/paper27`.

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