University of Texas, MD Anderson Cancer Center

UT MD Anderson Cancer Center Department of Biostatistics
Working Paper Series

Year 2007 *Paper* 39

The Effect of Population Drift on Adaptively Randomized Trials

John D. Cook*

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

The Effect of Population Drift on Adaptively Randomized Trials

John D. Cook

Abstract

Adaptively randomized trials aim to treat patients in clinical trials more effectively by increasing the probability of assigning treatments that appear to have a higher probability of response. Studies of adaptive randomization to date have assumed constant probabilities of response on each treatment. This paper examines the effect of response probabilities that change over time due to population drift.

The effect of population drift on adaptively randomized trials

John D. Cook*

December 14, 2007

Abstract

Adaptively randomized trials aim to treat patients in clinical trials more effectively by increasing the probability of assigning treatments that appear to have a higher probability of response. Studies of adaptive randomization to date have assumed constant probabilities of response on each treatment. This paper examines the effect of response probabilities that change over time due to population drift.

1 Introduction

Clinical trials implicitly assume static population characteristics, and yet populations may change over time. Change could be gradual, such as an increasing proportion of patients being treated in early stages of their disease due to advances in diagnosis. Change could also be sudden, as when a new center joins a multi-center trial. The nature of change in the patient population could be complex. Because patients often enroll in clinical trials when standard treatments have been ineffective, changes in the standard of care cause changes in the population enrolling in clinical trials: patients are implicitly selected by resistance to different treatments. Often population characteristics change so slowly that they are not of immediate concern. However, it is possible for a population to drift significantly during the course of a multi-year clinical trial. This paper addresses the effect of significant population drift on adaptively randomized clinical trials.

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

We limit the term "population drift" to unanticipated changes in the characteristics of patients enrolling in a given clinical trial that impact the efficacy of the treatments under consideration. The causes of these changes are necessarily unknown; otherwise they would be accounted for in the statistical design. For example, if one suspected that the proportion of patients in each disease stage would vary during the trial, and that staging would impact the relative efficacy of the treatments being tested, one could restrict eligibility by disease stage or create a statistical model which incorporates disease stage as a covariate.

Since we limit our attention to unanticipated population drift, we do not consider attempts to directly model the nature of the drift. One may anticipate that some sort of population drift is possible during a trial, but it is unreasonable to assume that one could determine the form of this drift in advance, anticipating the form of an unanticipated effect.

We may say that a treatment becomes more or less effective over time, though strictly speaking the treatment does not change; only the population to which the treatment is administered changes. However, since we are assuming the cause of population drift is unknown, we are justified in using phenomenological language and speaking of the treatment changing.

2 Population drift in randomized trials

Population drift is problematic for randomized trials, whether using equal randomization (ER) or outcome-adaptive randomization (AR). The primary concern is cross-over, changes in the effectiveness of treatments relative to each other. For example, an early stopping rule may select the treatment that performed better at the beginning of the trial, while another treatment would have been selected had the trial continued longer.

If population drift is present but not recognized, a randomized trial will select with high probability the treatment which *was* more effective during the trial, not the treatment which *will be* more effective after the trial ends.

If population drift is recognized, difficult questions arise. Which treatment should be selected for future patients? The treatment that appeared to be more effective at the end of the trial? Would it be appropriate to extrapolate population trends and speculate which treatment will be most effective by the time this treatment is in common use? Would a regulatory

agency approve a treatment based on anticipated changes to a population due to an unknown cause? (Recall that we assume the cause of patient drift is unknown, otherwise the cause would have been built into either the entry criteria or the statistical model.) Should a post hoc analysis be done to try to determine the cause of the drift? If the selection of the best treatment depends on how one handles drift, the trial will likely be inconclusive and a subsequent trial will be necessary to resolve the questions raised.

How should AR and ER designs be compared when there is substantial population drift? When there is no population drift, one could examine correct selection probability, as in [2]. However, comparing correct selection probability is problematic when it is not clear what the correct selection should be. Indeed, if substantial population drift is likely to result in an inconclusive trial, as we have argued above, then selection probabilities are moot. Instead we compare AR and ER trials on the basis of patient benefit. The motivation for adaptive randomization is to treat patients more effectively in a clinical trial [1]. But *could an adaptive randomization design do more harm than good* if the population is changing? One could imagine an AR design treating patients less effectively than an analogous ER design if the former persists in assigning more patients to what was once the more effective treatment after conditions have changed.

When there is no population drift, AR designs typically lead to more patients being assigned to the more effective treatment. When there is population drift, particularly cross-over, one can no longer speak of "the most effective treatment" without reference to time, and so the comparison must be based on the number of patients treated effectively. One design is better than another if it assigns on average more patients to what was the better treatment at the time the patient was treated.

We will present simulation studies to examine whether AR continues to treat patients more effectively than ER under a variety of scenarios.

3 Simulation studies

All simulations in this section are based on a two-arm trial of 500 patients. To simplify matters, we do not include an early stopping rule and we assume patient outcomes are known immediately upon treatment. We assume patients arrive according to a Poisson process. Drift is specified as a function of arrival time, not accrual number. For patients arriving at a time past the

expected duration of the trial, response probabilities are held constant at the value specified for the expected end of the trial.

In each scenario, we compare the number of patient responses on the AR design and an ER design. We also compare the number of responses to an AR design with constant probabilities of response on each arm given by the average response probability over time.

Let θ_i be the probability of response on arm i where i is 1 or 2. We assume each θ_i is distributed *a priori* as beta(2, 3) and assign each arm with probability equal to the posterior probability that it is the better arm. Specifically, the adaptive randomization process used here assigns Arm 1 with probability

$$p = P(\theta_1 > \theta_2 \,|\, \text{data}).$$

Each AR trial is simulated 10,000 times. Expected patient responses for ER trials are calculated multiplying the average response probability over time by the number of patients.

We begin by considering several scenarios in which response probabilities change linearly over time. We then consider scenarios in which the response probability on one arm is given by a step function.

3.1 Linear drift scenarios

We consider five scenarios with linearly varying probabilities of response. In the "rising tide" scenario, the probability of response on Arm 1 begins at 0.3 and rises to 0.4 by the end of the trial while the probability of response on Arm2 begins at 0.4 and rises to 0.5. In the "falling tide" scenario, the drift on each arm is reversed.

In the "catch up" scenario, the probability of response on Arm 1 increases from 0.3 to 0.5 while the probability of response on Arm 2 is constantly 0.5. In the "fall behind" scenario, the drift on Arm 1 is reversed. For the "cross over" scenario, the probability of response increases from 0.3 to 0.5 on Arm 1, but decreases from 0.5 to 0.3 on Arm 2.

The average number of patient responses for the linear drift scenarios are given in the following table.

Scenario	Average AR	Drift AR	ER
Rising tide	216.4	216.9	200
Falling tide	216.4	215.6	200
Catch up	241.2	243.4	225
Fall behind	241.2	237.7	225
Cross over	199.9	202.5	200

In each scenario, the AR design treats more patients effectively than does the ER design. Also, population drift has little effect on the AR results compared to fixing the response probabilities at their average values.

3.2 Jump discontinuity scenarios

In this section the probability of response on Arm 1 begins at 0.3 and remains constant until jumping to 0.5 and remaining constant for the remainder of the trial. The probability of response on Arm 2 remains fixed at 0.4. We vary the location of the discontinuity in increments of 5% of the expected duration of the trial. As the location of discontinuity increases, Arm 1 has a lower response rate for a longer amount of time and so the expected number of patients treated effectively decreases.

As Figure 1 illustrates, AR treats more patients effectively than ER in every scenario. The advantage of AR is greatest when the discontinuity occurs either early or late in the trial and is at a minimum when the break occurs around 40% through the trial. Also, the advantage of AR is not symmetric about this minimum. Instead, the method does best, relative to ER, when the break occurs late in the trial.

The table below gives the values plotted in Figure 1 but also includes simulation results for AR with constant probability of response. Notice that the AR results with and without drift are more similar to each other than either is to ER. In other words, the choice of AR vs ER has a greater effect than whether there is population drift.

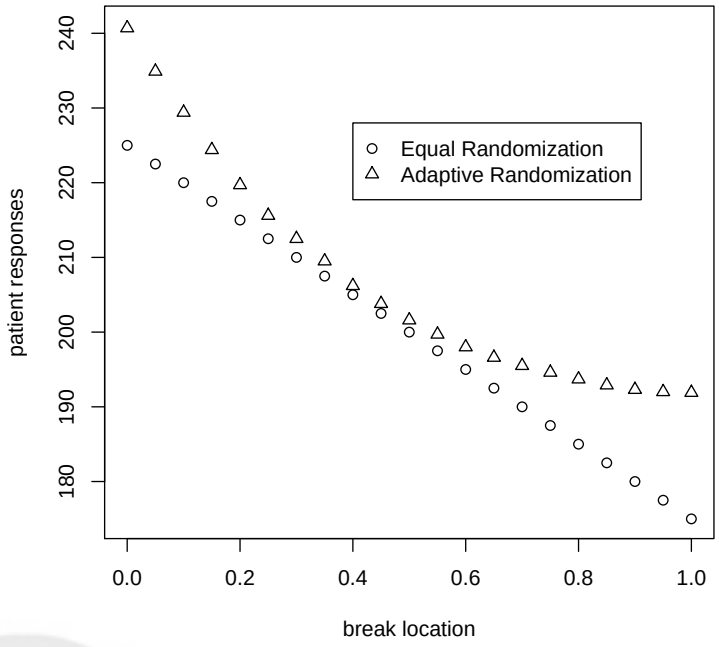

Figure 1: Comparing AR and ER with discontinuous drift

Break location	Average AR	Drift AR	ER
0.00	240.5	240.7	225.0
0.05	235.3	234.9	222.5
0.10	230.8	229.4	220.0
0.15	225.9	224.4	217.5
0.20	221.2	219.7	215.0
0.25	216.7	215.6	212.5
0.30	212.7	212.5	210.0
0.35	208.9	209.5	207.5
0.40	205.7	206.2	205.0
0.45	202.8	203.8	202.5
0.50	199.7	201.6	200.0
0.55	197.8	199.7	197.5
0.60	196.0	198.0	195.0
0.65	194.6	196.6	192.5
0.70	193.5	195.5	190.0
0.75	192.6	194.6	187.5
0.80	192.1	193.7	185.0
0.85	191.8	192.9	182.5
0.90	191.5	192.3	180.0
0.95	191.4	192.0	177.5
1.00	191.4	191.9	175.0

4 Discussion

Our simulation results also suggest that population drift has relatively little effect on adaptive randomization, whether the drift is continuous or discontinuous; replacing varying response probabilities with their average values had little impact on the number of patients treated effectively.

In all scenarios examined here, the adaptive randomization design treated more patients effectively than the corresponding equal randomization design. These results suggest that the advantage of adaptively randomized trials treating patients more effectively is robust to population drift.

References

[1] Donald A. Berry and Stephen G. Eick (1995). Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine*, **14**, 231-246.

[2] J. Kyle Wathen and John D. Cook. Power and bias in adaptively randomized clinical trials (2006). M. D. Anderson Technical Report UTMDABTR-002-06.