Impact of Covariate Adaptive Allocation Procedures on Power and Validity in Small-Scale Clinical Studies

by

Michael Flanagin

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Master of Arts

(Biostatistics)

at the

University of Wisconsin – Madison

2018

Defended on July 4, 2018

Dissertation approved by the following members of the Final Oral Committee:

Amalia Magaret · Associate Professor of Biostatistics

Mike Leblanc · Associate Professor of Biostatistics

Abstract

This is where my abstract goes.

Acknowledgments

This is where the acknowledgments go.

Contents

Abstract i

Acknowledgments ii

Contents **iii**

List of Figures **v**

List of Tables **vi**

1 Background **1**

1.1 Randomization in RCTs 2

1.2 Restricted randomization 2

1.3 Dynamic (adaptive) randomization 3

1.4 Analysis methods 5

1.5 Aims 6

1.6 Measures to evaluate aims 7

2 Methods **8**

2.1 Data generation 9

2.2 Allocation procedures 10

2.3 Varied conditions 10

2.4 Analysis approaches 11

3 Simulation **13**

3.1 Binary Y, binary X 14

3.2 Binary Y, continuous X 14

3.3 Continuous Y, binary X 14

3.4 Continuous Y, continuous X 14

4 Results **15**

5 Discussion **16**

5.1 Limitations 17

Simulation management 17

List of Figures

List of Tables

1 Background

1.1 Randomization in clinical trials

Research to identify effective and/or efficacious interventions requires balance between scientific and logistic constraints. Randomization in clinical trials allows one to infer causation from associations in the presence of an appropriate experimental design, appreciating the limitations in identifying true causal relationships.

Oftentimes it is of interest to measure and adjust for known variables predictive of the outcome measurement (prognostic factors), to reduce confounding and to increase efficiency. On average, simple (or complete) randomization ensures the distribution of prognostic factors, measured and unmeasured, are balanced across groups. Chance imbalance in important prognostic factors may be seen as impacting the credibility of observed treatment effect estimates. The likelihood of imbalance increases as the sample size decreases or the number of prognostic variables increases. Especially in small studies, at any given point in the randomization process there could be a substantial imbalance in the number of patients assigned to each group. For this reason, restricted randomization techniques are often used to guarantee the sequential and overall imbalance in the number of participants in each group is controlled.

1.2 Restricted randomization

Blocked randomization is a restricted randomization approach that ensures current and overall imbalance in treatment assignments are controlled. Sequential balance is achieved by specifying a block size for which the sequence of (usually equal numbers of) treatment assignments is permuted and assigned to patients as they are enrolled. However, blocked randomization does not by itself guarantee overall balance in known prognostic variables of interest. For this reason, blocking is often combined with stratification to make study groups comparable with regard to specified stratifying

factors.

Stratified randomization is a procedure that separates the recruitment population into smaller subgroups (strata) where randomization, either simple or blocked, is performed. This property can be useful in multi-center trials, for instance, where it is of interest to account for between-center variability in patient outcomes due to unmeasured or unimportant factors *a priori*. While the randomization ratio is guaranteed to hold within pre-specified blocks of enrolled subjects, in small trials with many stratification factors one cannot assure accrued patients will fill the block for each subgroup, and randomization within strata alone will not ensure balance.

1.3 Dynamic (adaptive) randomization

Stratified block randomization is considered a static randomization method, as the probability of treatment assignment is not conditional on information on patients already enrolled. In contrast, adaptive (or dynamic) randomization approaches control imbalance by dynamically altering the randomization probability based on accrued patient information. In this thesis we consider covariate adaptive randomization procedures, which are a natural comparison to static randomization strategies intended to control imbalance of baseline prognostic factors across treatment groups. These procedures have been increasingly used as an alternative to stratified block randomization, particularly in small scale clinical trials with many prognostic factors.

Initial developments in covariate adaptive methods aimed to reduce the probability of undesirable, albeit unlikely, allocation sequences which result in both overall treatment group imbalances and imbalances within subgroups defined by important prognostic factors. In this subsection we follow the historical development of covariate adaptive approaches with a brief discussion of the characteristics and performance of a few

selected methods.

Biased coin randomization introduced by Efron (1971) was the first randomization method to change the probability of assignment dynamically based on observed covariate values of accrued patients. Simple randomization is performed until the disparity reaches a prespecified limit, at which time the group with the least subjects is biased to have a greater probability of assignment.

Taves (1974) extended Efron's biased coin design to the context of small scale clinical trials, where it is of interest to constrain imbalance in multiple prognostic factors across treatment groups. Briefly, the method sequentially allocates incoming patients deterministically to the treatment category that minimizes the overall unweighted sum of covariate imbalance given the new assignment. The assignment is performed deterministically: assignment is randomized only when assignment to either treatment category results in the same imbalance. Pocock and Simon (1975) further generalized Taves' method to incorporate relative importance of prognostic factors by introducing weighting of covariate imbalances into the overall imbalance metric.

Signorini et. al 1993 extended earlier methods in order to induce balance both overall and within strata while avoiding investigator bias through unblinding. He proposed a tree-based method of dynamic balancing randomization (DBR) that evaluates imbalance for each prognostic factor in a nested fashion by their prespecified order of importance. The method flexibly allows for different levels of imbalance in different strata and ensures conditional balance, meaning that within each subgroup the ratio of treatment assignment is constrained within prespecified bounds. However, the method does not guarantee balanced group assignments will be achieved within each prognostic factor considered separately.

Heritier et. al (2005) modify Signorini et. al's DBR method to control imbalance marginally within each prognostic factor. For each accrued patient, the potential

imbalance for each treatment assignment is considered sequentially within each prognostic factor in decreasing order of importance. If the potential observed imbalance exceeds a prespecified threshold, assignment is performed deterministically (or forced) to the group which minimizes the imbalance. Heritier et. al suggested including non-deterministic allocation to reduce the number of forced allocations and prevent investigator unblinding.

Model-based approaches are another alternative approach to dynamic randomization, where the probability of treatment assignment is chosen to minimize the variance of the estimated treatment effect. Model-based methods can flexibly incorporate continuous prognostic factors without the need to dichotomize into groups, and can include interaction terms and balance prognostic factors even when the number of variables is large. Aickin (1998, 2001, 2009) proposed a model-based approach to covariate adaptive randomization, where a subjects' treatment assignment is based on maximizing the log-likelihood of the model.

1.4 Analysis methods

Analysis of trials using a covariate adaptive allocation (CAA) scheme must account for the randomization scheme to recover the precision gains conferred by inducing more balanced treatment groups with respect to chosen balancing factors. To obtain the correct variance term and significance level for the test statistic, one must consider all possible sequences of assignments which could have been made in repeated trials assuming no group differences in mean response. In most cases, ignoring the randomization procedure and using standard regression methods that implicitly assume complete randomization lead to larger variance estimates and conservative inference. Since CAA modifies the randomization scheme to induce similarity across treatment arms

relative to within arms, the efficiency gain can be realized using a nonparametric re-randomization approach for estimating standard errors (Simon and Simon 2011). Briefly, observed values and entry order are fixed, treatment assignments are reshuffled and the test statistic computed for each permutation.

1.5 Aims

The goal of the thesis is to address in both the binary and continuous outcome setting if covariate adaptive randomization (CAR) followed by standard asymptotic tests yield valid inference, and if so, to quantify the gains in precision relative to simple randomization (SR) or stratified block randomization (SBR). We will compare Heritier's modified DBR scheme to stratified block randomization and complete randomization, while comparing re-randomization based permutation tests to standard asymptotic tests in a simulation study. We will consider the setting of equal allocation to treatment assignment, no temporal trend (drift), binary predictors, and two outcome types (binary and continuous). Our objective is to identify any scenarios, if any, where minimization improves power relative to SBR or simple randomization. Contour plots of effect size by sample size will compare power across methods for various outcome types and conditions.

We are also interested if and when the answers to the above questions change when the effect size of prognostic factors is varied relative to the treatment effect, the baseline prevalence varies from 5% to 50% in the binary outcome setting, inference on treatment effect is performed using none (or a subset) of the prognostic variables, and when the sample size is varied. It is well known that ignoring the minimization design tends to yield conservative inference, and that adjusting for covariates used in the randomization scheme (balancing factors) recovers type I error rate to nominal

significance levels (Xu, Proschan, Lee 2016). Through comparison of estimated marginal and conditional treatment effects we seek to confirm this finding. We consider different sample sizes ranging from N=32 to 108 to compare CAR to SBR as small scale trials are the setting in which alternative randomization methods are considered. We seek to identify the specific conditions by which CAR confers a precision advantage, if any, relative to other methods to offset the operational complexities involved in implementing an adaptive allocation procedure. Our intent is to provide guidance to clinical researchers for determining under what settings covariate adaptive allocation provides precision gains relative to competing approaches as well as which analysis method yields valid tests with the most power.

Chapter 2 will introduce the notation used throughout the thesis, and Chapter 3 will discuss the design of the simulation study in further detail. The tables of simulation results will be presented in Chapter 4 and the key observations will be discussed in Chapter 5.

1.6 Measures to evaluate aims

For each combination of randomization scheme and analysis approach, we assess validity by estimating the nominal significance level of the test under the null hypothesis. We evaluate accuracy by estimating any potential bias and the coverage probability of confidence intervals, comparing those generated with standard regression methods (Wald-type) to permutation test quantile-based confidence intervals. Average standard error estimates will also be reported for analyses using standard regression methods. We evaluate efficiency by computing mean squared error (MSE) and power as a function of the true treatment effect size.

2 Methods

2.1 Data generation

Simulations were conducted of a two-arm randomized clinical trial with equal allocation (1:1 treatment:control). We conducted simulations for both continuous and binary outcome and covariate types, and varied the overall trial size from 32 to 96.

The outcome measure (Y) was simulated with a marginal prevalence of 10% or 50% in the binary setting to evaluate the potential impact of low numbers of observed outcomes on inference. Continuous outcomes were simulated as normally distributed with constant variance, with mean as a linear combination of the treatment assignment (Z), pre-specified risk factors (X), and observed entry time (T).

Binary risk factors were simulated such that their marginal prevalence was either 25% or 50%. Continuous risk factors (X) were generated under a standard normal distribution. The risk factors are modeled as independent. The effect sizes for treatment assignment (Z) and prognostic variables (X) were separately varied from none, low, medium, and high. Balancing factors refer to risk factors used in adaptive allocation procedures, for which it is desired to have comparability either within or between treatment groups. The exact type of balance desired informs the choice of imbalance metric minimized at each sequential allocation step. For instance, it may be of interest to ensure within-strata subgroups have approximately proportional treatment and control assignments (conditional balance), or that treatment groups are otherwise comparable with respect to pre-specified balancing factors (marginal balance).

Observed patient entry times occurred following a uniform distribution. In subsequent re-randomization analysis, patient entry order is considered fixed (see Section 2aiv on re-randomization tests). Temporal trends in outcome prevalence over the course of a study (drift) were modeled by varying the entry time effect size from none, mild,

to severe. For the binary response setting under severe drift, for instance, the drift effect size was chosen such that marginal outcome prevalence varies three-fold over the study period.

2.2 Allocation procedures

For each simulated set of observed entry times and prognostic factors, treatment group assignments were determined using three allocation procedures: complete randomization, stratified permuted block randomization with fixed block sizes equal to overall trial size divided by number of strata (defined by all combinations of balancing factor levels), and an adapted form of covariate adaptive randomization proposed by Heritier et. al (2005).

For the covariate-adaptive randomization procedure, the maximum imbalance of treatment to control assignments (overall and within strata defined by each balancing factor level, considered separately) was set to 2. The allocation biasing probability, or the probability of assigning patient to treatment minimizing the imbalance measure when prospective imbalance meets or exceeds a prespecified threshold, was chosen as 0.7 to minimize the effect of non-deterministic and forced allocations on inference.

2.3 Varied conditions

The following table describes the conditions varied in the binary outcome setting. Note: in the continuous outcome setting, the effect sizes are modified to represent comparable differences in means to the given odds ratios.

2.4 Analysis approaches

For each allocation procedure we estimated the treatment effect, adjusted and unadjusted for balancing factors. To evaluate power, coverage, and type I error control, we report the associated linear and logistic regression model-based p-values, standard errors and Wald-type confidence intervals, based on a two-sided type I error threshold (alpha) of 0.05. For balancing purposes, continuous balancing factors were first dichotomized by their population median, and later parameterized as continuous in the adjusted analysis.

The simulation model is of the form

The adjusted regression model is

The unadjusted model

To compare the bias and power based on re-randomization analysis to that based on standard regression techniques, we compute power, coverage, and level and conducted re-randomization based inference for each simulated trial following covariate adaptive randomization. Re-randomization is a permutation-based method for estimating uncertainty and follows from the generally accepted sentiment to 'analyze as you randomize'. The approach considers the outcomes (Y), balancing factors (X), and observed entry time (T) as fixed and repeats the allocation procedure multiple times, each generating a new sequence of treatment assignments (. Regression estimates are computed under each re-randomized treatment allocation sequence, and re-randomization based 95% confidence intervals are generated using the 2.5th and 97.5th quantile of the re-randomization-based treatment effect estimates. Re-randomization based p-values are estimated using the observed proportion of re-randomized allocation sequences yielding treatment effect estimates as or more extreme than the observed treatment effect.

Each simulation model configuration was simulated 10,000 trials, for which the re-randomization procedure was repeated 500 times.

3 Simulation

3.1 Binary Outcome, Binary Predictors

3.2 Binary Outcome, Continuous Predictors

3.3 Continuous Outcome, Binary Predictors

3.4 Continuous Outcome, Continuous Predictors

4 Results

5 Discussion

6 Limitations

Due to many unforeseen factors, we ran into these complications:

6.1 Simulation management

For each simulation, we had to keep track of the following:

- Prognostic factors matrix $\mathbf{X}$
- Entry time vector $\mathbf{T}$
- Allocation sequence vectors $\mathbf{Z}$, each with their associated
- Outcome measure vector $\mathbf{Y}$
- Model-based regression estimates:
  - Estimate $\hat{\beta}$
  - Standard error $\hat{SE}(\hat{\beta})$
  - P-value p_T
  - T-statistic t_obs
  - Confidence interval $\hat{CI}(\hat{\beta})$
- Re-randomization based regression estimates, including:
  - Re-randomized allocation sequence vectors $\mathbf{Z}$, each with their associated estimate $\hat{\beta}$

We structured the simulation around using the **"simulator"** R package, which handles file management (saving/loading data), RNG seeds for parallel processing, and sequence of simulation steps.

Issues

Memory management was poor: the simulator stored data as so:

- each **Model** object contains the simulation parameters

- each **Draw** object contains the prognostic factor and entry times **(X, T)** for all subjects
- the first **Output** object contains the allocation sequence vector and outcome vector **(Z, Y)** for all subjects
- the second **Output** object contains the model-based regression estimates ($\hat{\beta}$, $\hat{SE}(\hat{\beta})$, p_T, t_obs, $\hat{CI}(\hat{\beta})$)
- the third **Output** object contains the re-randomization based regression estimates ($\hat{\beta}$, $\hat{SE}(\hat{\beta})$, p_T, t_obs, $\hat{CI}(\hat{\beta})$)