Comparison of dynamic block randomization and minimization in randomized trials: a simulation study

Lan Xiao[a], **Phillip W Lavori**[b], **Sandra R Wilson**[a,c], and **Jun Ma**[a]

[a]Department of Health Services Research, Palo Alto Medical Foundation Research Institute, Palo Alto, CA, USA

[b]Department of Health Research & Policy, Stanford University School of Medicine, Stanford, CA, USA

[c]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

Abstract

Background—Minimizing the imbalance of key baseline covariates between treatments is known to be very important to the precision of the estimate of treatment effect in clinical research. Dynamic randomization allocation techniques have been used to achieve balance across multiple baseline characteristics. However, empirical data are limited on how these techniques compare in terms of balance and efficiency. We are motivated by a newly funded randomized controlled trial, in which we have the option of choosing between two methods of randomization at the subject level: (1) randomizing individual subjects consecutively as they are enrolled, using Pocock and Simon's minimization method, and (2) simultaneously randomizing blocks of subjects once all subjects in a block have been enrolled, using a balance algorithm originally developed for cluster randomized trials.

Purpose—To compare dynamic block randomization and minimization in terms of balance on baseline covariates and statistical efficiency. Simple randomization was included as a reference.

Methods—A simulation study using data from a previous randomized controlled trial was conducted to compare balance statistics and the accuracy and power of hypothesis testing among the randomization methods.

Results—Dynamic block randomization consistently produced the best balance and highest power for various sample and treatment effect sizes, even after post-adjustment of the pre-specified baseline covariates in all three methods. Consistent with previous reports, minimization performed better in balance and power than simple randomization; however, the differences were noticeably smaller compared to those between dynamic block randomization and simple randomization.

Limitations—In this simulation study, we considered three sample sizes and two block sizes for a two-arm randomized trial. We assumed no interactions among the multiple baseline covariates.

Author for correspondence: Jun Ma, Department of Health Services Research, Palo Alto Medical Foundation Research Institute, 795 El Camino Real, Ames Building, Palo Alto, CA, USA. maj@pamfri.org.

It is necessary to evaluate how the results may vary when the simulation conditions are changed before drawing broader conclusions regarding comparisons between the randomization methods.

Conclusions—This study demonstrates that dynamic block randomization outperforms minimization with regard to achieving balance and maximizing efficiency. Nevertheless, the differences across the three randomization strategies are modest. The statistical advantages associated with dynamic block randomization need to be considered in relation to the planned sample size and the practical issues for its implementation in deciding the preferred method of randomization for a given trial (e.g., the time required to accrue blocks of subjects of adequate size as balanced against the need to commence intervention/treatment immediately in those randomized to that experimental condition).

Background

Randomized controlled trials (RCTs) are the most rigorous method of generating comparative efficacy and effectiveness evidence on different clinical interventions and necessarily occupy a central role in clinical research. One experimental condition (treatment) in such trials is generally considered to be the control or comparison condition and this condition often represents the current standard of care for the clinical problem being studied. The key requirements of RCTs are that they minimize bias, balance potential or known confounders, and hence ensure an efficient and unbiased comparison [1,2]. In an RCT, it is essential to assign subjects unpredictably and to balance the treatment groups on important baseline covariates.

Randomization ensures that each subject is assigned to each treatment independently of his/her baseline characteristics, measured or unmeasured, including characteristics that are the current values of potential outcomes of treatment. The use of randomization was clearly specified in The Design of Experiments (1935) by Fisher [3]. Simple randomization generally allocates subjects based on random numbers that ensure that each subject is assigned to the various treatments with equal probability. The element of unpredictability in an RCT is easily accomplished in simple randomization, but balance between treatments in terms of number of subjects and baseline characteristics cannot be guaranteed by this strategy and may not result even in moderatesized trials.

Various allocation techniques have been developed to minimize the potential imbalance on important baseline characteristics in different design settings. Minimization is a dynamic randomization technique that sequentially assigns subjects to treatment by attempting to minimize the total imbalance between treatments over multiple baseline covariates. The minimization method achieves marginal balance by looking at all of the selected baseline covariates for the previously assigned subjects and assigning the next subject to a treatment with a probability in favor of minimizing the overall imbalance across the covariates [4]. Use of nonextreme allocation probabilities (e.g., 2/3: 1/3 in a two-arm trial) maintains unpredictability.

Cluster randomized designs are useful for studies in which it is desirable to implement treatments at the level of a naturally occurring group of patients or other units (a block). Cluster randomized trials consider the variances at both unit and block levels [5]. Assuming

that all units within a block are enrolled prior to their randomization, the availability of complete knowledge of baseline covariates of all units can be used in order to achieve an optimal balance between treatments [6]. Raab and Butcher explored the importance of balancing covariates between treatments and provided several possible approaches to selecting a well-balanced allocation in cluster randomized trials. A natural extension of these approaches is their application to blocks of subjects as they are enrolled [7]. An important goal for this application is to balance between blocks as well as within. Carter and Hood programmed a balance algorithm in the R Software, which is freely available [8], that calculates the within and between block imbalance measure using baseline covariate information [7].

In a two-arm RCT funded in 2009, we will evaluate the efficacy of a comprehensive behavioral weight loss intervention (focusing on diet, physical activity, and behavioral self-management) relative to usual care for the treatment of asthma in obese adults (the BE WELL trial; trial registration: NCT00901095). We plan to recruit cohorts of approximately 65 subjects who are obese and have inadequately controlled asthma and to recruit a total of five sequential cohorts (total $n = 324$). The intent is to randomize equal numbers of participants to the intervention and usual care arms. The intervention involves a 4-month intensive (weekly) group counseling program, followed by face-to-face and phone counseling of each participant for 8 months. A group intervention class will begin whenever the number of new patients randomized to the intervention arm who can attend the class reaches 10. One possible randomization strategy for this study would be to assign each subject to intervention or control consecutively using Pocock and Simon's minimization method. A second strategy would be to randomize subjects in, for example, blocks of 20 by applying the balance algorithm developed by Raab and Butcher (2001) for cluster randomized trials to take advantage of complete data on baseline covariates across the recruited sample prior to randomization.

There is limited empirical data on how these strategies compare in terms of balance and efficiency that could inform our decision making. We therefore undertook a simulation study to compare these strategies in terms of balance statistics and the accuracy and power for hypothesis testing. We also included simple randomization in the comparison because it has been used broadly as a reference method for allocation strategy comparisons.

Methods

Allocation strategies

Minimization—Minimization is designed to minimize marginal imbalance over multiple important baseline covariates as each consecutive treatment assignment is made. Pocock and Simon's minimization method requires continuous covariates to be categorized in order to calculate treatment imbalance [4,9]. For an arbitrary point in the succession of randomizations, denote n_{ijk} as the number of patients with level j of covariate i who have been previously assigned to treatment arm k ($i = 1,2, …,C$; $j = 1,2, …, J$; and $k = 1,2, …, K$, where C, J, and K are the numbers of covariates, levels of covariate i, and treatment arms, respectively). Let the next patient entering the trial have levels $r_1, r_2, …, r_C$ on the covariates $1, …,C$. Pocock and Simon proposed several ways of measuring the cumulative imbalance

on the previously assigned patients and after assignment of a new patient [4,9]. We chose the simplest form – marginal totals – as the imbalance measure in this study, which is defined as:

$$G_k = \sum_{i=1}^{C}\sum_{j=r_i} n_{ijk} \quad (1)$$

The G is a function that calculates the resulting overall imbalance of the treatment assignments if the subject is assigned to treatment k. The G scores corresponding to each treatment k can then be ranked from the smallest to the largest, and the treatment assignment that results in the least overall imbalance will be chosen with a high probability (p_k) so as to increase the chance of maximizing balance among the covariates. The choice of p_k determines the degree of balance and the predictability of treatment assignments.

Dynamic block randomization—Dynamic block randomization minimizes the imbalance over multiple important baseline covariates between treatment arms for all allocations within and between blocks. The imbalance criterion is defined as:

$$B = \sum_{i=1}^{C} w_i \left(\bar{X}_{1i} - \bar{X}_{2i} \right)^2 \quad (2)$$

where i is the i-th baseline covariate (i 1, …,C); $\bar{x_{1i}}$ and $\bar{x_{2i}}$ are the average of i-th covariate for treatment 1 and 2, respectively; and the weights (w) determine the relative contribution of each covariate to the imbalance score (B) [6]. Weights for covariates may be estimated from previous trials that have used the same set of baseline covariates. Alternatively, it is typical to standardize the covariates and assume equal weights for all, and we used this approach in this study. Nominal categorical covariates must be coded as orthogonal dummy variables in order to calculate the imbalance score.

The imbalance scores of all possible allocations within a block are calculated using expression (2). Calculation of the imbalance scores for second and subsequent blocks is conditional on the selected allocation of earlier blocks. Optimal allocations are defined as the set of allocations with the smallest 1000 B values for block sizes 17, with the smallest 100 B values for block sizes between 12 and 16, and with the lowest quarter of B values for block sizes between 8 and 11 [7]. An allocation is then randomly selected from the optimal allocations as the assignment for the block. Within each additional block, the number of units will be equally split between treatment arms if the block size is even (regardless of previous blocks), whereas if the block size is odd, the number of units allocated to each treatment arm in previous blocks will be considered to make the total number of units between treatments as equal as possible.

Simple randomization

In simple randomization, treatment assignments are made on a 1/2: 1/2 basis without regard to subjects' baseline characteristics. It is included for comparative purposes.

Simulation study—We evaluated the three allocation strategies in a simulated two-arm trial using data on the five baseline covariates for randomization in the BOAT trial, which investigated the efficacy of a shared treatment decision-making approach to asthma care management in comparison to usual care and an active control condition (a conventional care management approach)[10]. BOAT preceded BE WELL and similarly targeted adults in the same health care system whose asthma was not well-controlled; BE WELL specifically focuses on obese individuals. In our simulation (Flowchart in Figure 1), we used the original five baseline covariates in the BOAT trial: sex (2 levels: male and female), prior asthma hospitalization (2 levels: none and 1), ethnicity (2 levels: White and non-White), age (3 levels: 18–34, 35–50 and 51), and asthma controller medication use (3 levels: none, 1–3 days per week, and 4 days per week). Table 1 shows the distributions of the covariates. The sample sizes considered were 40, 60, and 80, with effect sizes of 0, 0.2, 0.5, and 0.8, respectively. The simulation data set consisted of 1,000 samples of 80 subjects randomly selected from the 612 BOAT participants, and then the first 40 or 60 of each sample were used to create 1000 samples of 40 or 60 subjects.

Because there is no obvious decision rule for optimizing the choice of p_k for any individual trial, we evaluated the performance of minimization at various p_k values found in the literature for assigning the next subject to the treatment arm with smaller value G_k (i.e., $p_k = 2/3$ [11], 3/4 [9], 0.8 [12], 0.9 [12], 1 [4]). We started the minimization method with simple randomization of the first 10 patients, a strategy that is often used to prevent guessing of assignments [13]. We carried out dynamic block randomization as follows. Given that the samples of 40, 60, or 80 subjects were to be randomized in sequential blocks of 20, with all baseline covariates collected prior to randomization, simple balanced randomization was first performed to assign 10 subjects to the treatment arm and the remaining 10 subjects to the control arm. The number of possible allocations was the number of combinations of choosing 10 out of 20 ($n = 184,756$). The imbalance score of each of these allocations was calculated and they were all ordered from the smallest to the largest score. From the 1000 allocations with the lowest imbalance scores, one was randomly chosen and assigned to the first block. The same steps were repeated for each subsequent block of 20 subjects, except that calculation of the imbalance scores for the new block was conditioned on the selected allocation of earlier blocks. To assess the effect of block size on performance, we also ran dynamic block randomization in blocks of 10 subjects.

For simple randomization, the subjects were assigned in a 1: 1 ratio to the treatment and control arms based on random numbers generated using PROC PLAN in SAS Enterprise Guide 4.2 (SAS Institute, Cary, NC).

Six measures were used comparing the three allocation strategies:

1) Imbalance score (B): Imbalance scores were computed using expression (*2*).

2) Marginal imbalance ($\bar{b}_M$, Maxb_M): For a given baseline covariate, treatment comparisons need to be examined marginally within each categorical level of the covariate [14]. Let n_{ijk} represent the number of subjects who are assigned to treatment k ($k = 1,2$) and who have level j of covariate i as previously denoted. Define

$$b_{Mij} = \frac{|n_{ij1} - n_{ij2}|}{\sum\limits_{k=1}^{2} n_{ijk}} \quad (3)$$

as the imbalance within the marginal subgroup of subjects with level j of covariate i [15]. Define $\bar{b}_M = \overline{b_{Mij}}$ and $\mathrm{Max}b_M = \max_{ij}\{b_{Mij}\}$. $\mathrm{Max}b_M$ is a better indicator of the undesirable situation in which most covariate levels are well balanced while a few are highly imbalanced [15].

3) Significance of imbalance (I_{Pi}): Pearson chi-square tests were performed to eval uate the association between each baseline covariate and treatment. I_{Pi} indicates the significance of the test of association for covariate i ($I_{Pi} = 1$, where p-value <0.05; $I_{Pi} = 0$, otherwise).

4) Power: To evaluate the impact of balancing properties in the design stage on the power of testing for the treatment effect, we considered effect sizes of 0, 0.2, 0.5, and 0.8 assuming standard normal random error and pre-specified fixed effects for the five covariates. An effect size of 0 was selected to evaluate the type I error rate. Effect sizes of 0.2, 0.5, and 0.8 were selected as they conventionally represent small, medium, and large effect sizes, respectively [16]. The following simulation model was used:

$$Y = \beta_0 + \beta_1 \boldsymbol{TRT} + \sum_{i=1}^{C} \beta_{i+1} \boldsymbol{X}_i + e \quad (4)$$

where Y is the outcome vector; $\boldsymbol{TRT}$ is the vector for treatment assignment; $\boldsymbol{X}_i$ is the corresponding vector for baseline covariate i; and βs are the coefficients for all fixed effects; and $e \sim N(\boldsymbol{0, I})$. The β coefficients for each level of the five covariates were selected from the range 0 to 1 for simulating $\mathbf{Y}$ and then were re-estimated to evaluate the significance of treatment difference. All tests were two-sided at the 5% significance level. The estimated type I error rate, $\hat{\alpha}$ was the percentage of significant results out of 1000 total samples using Y simulated given an effect size of 0, and the estimated power, $1 - \hat{\beta}$, was such a percentage using Y simulated given effect sizes of 0.2, 0.5, and 0.8. In addition, the ideal situation in which there was no confounding but only a pure treatment effect was also simulated by setting the baseline covariate βs to 0 for all possible combinations of the effect sizes and sample sizes, and power estimates based on independent t-tests were used as the reference. To quantify the increase in efficiency of dynamic block randomization and minimization compared with simple randomization, we also calculated the gain in required sample size based on the power estimates for a sample size of 60 and an effect size of 0.8 but assuming a pure treatment effect.

Descriptive statistics such as means and quartiles for the imbalance score (B), and marginal imbalance ($\bar{b}_M$, $\mathrm{Max}b_M$), and the number of baseline covariates that were significantly different ($p < 0.05$) by treatment arm were summarized and compared across the three randomization allocation strategies for all of the simulated sample sizes.

Results

Balance

Tables 2 and 3 show the distribution statistics of imbalance scores (B) and marginal imbalance measures $\bar{b}_M$ and $\text{Max}b_M$ for 1000 samples of sizes 40, 60, and 80, respectively. Among the three allocation strategies, dynamic block randomization in blocks of 20 consistently had the lowest means and quartiles of all of the measures. The means and quartiles of the imbalance measures for minimization improved with increasing probabilities (p_k) and approached, but never outperformed, those of dynamic block randomization with a block size of 20. When the block size was reduced to 10, the imbalance measures of dynamic block randomization were comparable to those of minimization at $p_k = 0.9$. As expected, simple randomization had the worst imbalance according to all the measures. Also, as expected, all imbalance measures decreased with increasing sample sizes, regardless of allocation strategy.

Table 4 shows the number of significant p-values comparing five baseline covariates between treatments ($E_i I_{pi}$) for all 1000 samples of sizes 40, 60, and 80, respectively. Dynamic block randomization produced no significant results. Minimization generated some significant comparisons for $p_k = 2/3$, 3/4, 0.8, and 0.9 but none for $p_k = 1$. The numbers of significant p-values for simple randomization were around the expected 250 for 5000 tests (5 covariates * 1000 samples) of a true null at 5% alpha.

Power

As shown in Table 5 (Panel a), all allocation strategies resulted in type I error rates around the conventional 5% level across the three sample sizes, assuming no treatment effect. The exact 95% confidence interval for an observed error rate of 5% from 1000 simulations is 3.7–6.5%, and none of the type I error rates in Table 5 (Panel a) was outside this range. Therefore, The fact that the observed percentages are within the expected confidence interval is consistent with the expectation that adjusting for the baseline covariates achieved the 5% type I level for all three sample sizes using any of the allocation strategies.

Table 5 (Panel b) shows that dynamic block randomization with a block size of 20 resulted in the highest power for all three sample sizes and all effect sizes. For sample sizes of 80 and 60, the levels of power achieved by dynamic block randomization with a block size of 20 were comparable to the 'no-confounding' reference levels, which were based on independent t-tests assuming zero β coefficients for the covariates. Changes in the power estimates were small when the block size was reduced to 10. Compared to simple randomization, the increased power associated with dynamic block randomization at either block size was equivalent to a gain in sample size of 20% for a sample size of 60 and an effect size of 0.8. Minimization resulted in similar power estimates at various p_k values, and, for a sample size of 60 and an effect size of 0.8, the gains in sample size relative to simple randomization ranged from 8% to 12% at various p_k values.

Discussion

To the best of our knowledge, there has been no prior published methodological report that provides a direct comparison between randomization using the minimization method and dynamic block randomization using an algorithm that originated in cluster randomized trials, probably due to the presumed difference in the unit of randomization. In this study, one balance algorithm used in cluster randomized trials was extended to a practical example in which the subject is the unit of randomization.

In our case study, we were particularly interested in comparing dynamic block randomization for a block size of 20 with Pocock and Simon's minimization method at varying allocation probabilities, p_k. We found that dynamic block randomization at that block size outperformed minimization in terms of both balance and power. Compared with minimization, dynamic block randomization achieved a higher level of balance, as indicated by the total imbalance score, average marginal imbalance, and maximum marginal imbalance, while preserving the unpredictability of the allocation. The balance achieved by dynamic block randomization at that sample and block size was notably better than that achieved by minimization, even for the extreme $p_k = 1$ where there is no unpredictability.

At the sample sizes evaluated, our model-based analyses showed that dynamic block randomization resulted in greater power for detecting treatment effects than minimization, that is larger increases in the effective sample size, when both methods were compared with simple randomization. Covariance adjustment is more efficient, statistically, when the covariates are more balanced [6,17]. The increase in balance decreases the standard error of the covariance-adjusted, estimated treatment effect, which has the effect of increasing the sample size and hence the power [17]. The pursuit of balance on covariates is not only important for statistical efficiency in smaller trials, as shown in our results, but also is desirable in larger trials in that it tends to increase the credibility of the trial, and to maximize the knowledge gained from all analyses in the trial, not just those relating to the primary measure of efficacy or to the sample as a whole [18]. The relative advantage of dynamic block randomization over minimization, for example, decreases as sample size increases, and at sizes where its advantages are not large, the delay in randomization necessary to accumulate blocks of, for example, 20 patients, may offset the gains associated with blocks.

Statistical efficiency is also affected by model assumptions that may not be robust to violations. Dynamic block randomization has the additional advantage of providing a basis for (nonparametric) randomization inference, whereas the minimization method does not have this benefit due to its conditional independence criterion. Significance that can be assessed without unverifiable assumptions, what Tukey termed 'platinum standard significance' [2], requires that the clinical trial be randomized in such a manner that one assignment is chosen from a list of pre-defined, acceptable assignments with known (often equal) probability. Dynamic block randomization meets this requirement because it defines a set of optimal allocations based on the imbalance score, *B*, and assigns equal probability to each allocation [6].

For the purpose of comparability, we used categorized baseline covariates in the simulation. However, unlike Pocock and Simon's minimization method [4,9], the balance algorithm for dynamic block randomization that we applied is capable of handling continuous covariates as well. Where a continuous covariate is more predictive of outcomes than a given categorization of that covariate, dynamic block randomization would have an added advantage over this minimization method.

To assess the impact of block size on performance, we also investigated performance with a block size of 10. Compared with a block size of 20, this smaller block size resulted in somewhat inferior balance but comparable efficiency. Nevertheless, even at a block size of 10, dynamic block randomization still achieved a level of balance similar to that of the next best performing minimization method at $p_k = 0.9$, while preserving a higher level of unpredictability. The differences in balance associated with block size are driven by the sampling space, which is determined by the optimal imbalance score (B) values for each block and by the number of blocks. As mentioned in the Methods, the recommended threshold for optimal imbalance score values is the smallest 1000 B values for block sizes 17, the smallest 100 values for block sizes between 12 and 16, and the lowest quarter for block sizes between 8 and 11. The B values for subsequent blocks are conditional on the first block [7]. For a given sample size, a smaller sampling space gives rise to better balance but relatively less concealment.

As indicated above, the performance of randomization methods also varies with sample size. Over the range of sample sizes from 40 to 80, we demonstrated that, not surprisingly, balance and power improved with sample size for all three randomization methods, and that, for dynamic block randomization, a block size of 20 consistently performed better than a block size of 10. Also, as expected, the magnitude of the improvement, and the advantages of dynamic block randomization over minimization, diminished with sample size. For instance, the decreases in the imbalance measures were notably smaller from a sample size of 60–80 compared to from a sample size of 40–60. Further increases in sample size would likely lead to imbalance measures eventually becoming asymptotic to 0 for all three randomization approaches, although at different rates. At sample sizes of 60 and 80, the efficiency achieved by dynamic block randomization with a block size of 20 approached that of independent *t*-tests assuming a pure treatment effect with no confounding. Studies of additional block and sample sizes will provide further insights into the impact of these two critical parameters on the performance of dynamic block randomization.

In this simulation, we did not consider interactions among covariates, which may affect response to treatment. It would be impractical to balance for all covariate interactions of any order in most clinical trials [4]. Nevertheless, both dynamic block randomization and minimization can incorporate a first-order interaction between two categorical covariates by creating a new variable whose levels correspond to all combinations of the two covariates. Additionally, dynamic block randomization can also account for first-order interactions between categorical and continuous covariates and between two continuous covariates. Future studies are needed to investigate the impact of interactions among covariates on the performance of these randomization strategies. Future work is also needed to examine the

use of dynamic block randomization in randomized trials with three or more arms, whereas we assumed only two treatments in the current study.

Dynamic block randomization was originally developed for cluster randomized trials in which the unit of randomization is a naturally occurring group of units (e.g., physician practices, clinics, hospitals, households, schools, communities). In this article, we demonstrate that this method can be readily adapted to clinical trials where, even though the unit of randomization is the individual, it is practical to accrue groups of individuals before randomizing them. The motivation to do this may be somewhat greater when the intervention is one that is implemented in groups, as is often the case in behavioral intervention trials. The method is also applicable in health services research studies for chronic disease prevention and management, in which potential participants are pre-identified using information in medical records and hence can be recruited, enrolled, and randomized in cohorts in relatively brief time intervals of a few weeks duration. In our ongoing clinical trials [19,20], we find that recruitment in cohorts provides practical advantages beyond the design advantages of allowing dynamic block randomization, in that it focuses the efforts of recruitment personnel and increases the efficiency and effectiveness of the recruitment process. Another example of a reasonable application is in animal studies where there are important baseline covariates (e.g., weight), as the animals are frequently examined in batches. However, the statistical advantages of dynamic block randomization, even over minimization, for small to moderate sample sizes, are sufficiently compelling that consideration of this approach is warranted even for interventions/treatments that are delivered on an individual basis. Benefiting from complete knowledge of baseline covariates across the recruited sample, dynamic block randomization consistently produces better balance and power while preserving a high level of unpredictability, compared with minimization. However, its statistical gains need to be considered in relation to the practical issues for its implementation in deciding the preferred method of randomization for a given trial.

As noted by Carter and Hood (2008) [7], for block sizes exceeding 20, the total number of enumerated allocations becomes computationally intense exponentially. Therefore, the feasibility of randomizing more than 20 subjects in a block in trials where this is desired will depend on the amount of available RAM, in addition to logistic considerations (e.g., the time required to accrue blocks of adequate size as balanced against the need to commence intervention/treatment in those randomized to that experimental condition).

Acknowledgments

References

1. Kalish LA, Begg CB. Treatment allocation methods in clinical trials: a review. Stat Med. 1985; 4:129–44. [PubMed: 3895341]

2. Tukey JW. Tightening the clinical trial. Control Clin Trials. 1993; 14:266–85. [PubMed: 8365193]

3. Fisher RA. The Design of Experiments. Oliver and Boyd, Edinburgh. 1935

4. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics. 1975; 31:103–15. [PubMed: 1100130]

5. Varnell SP, Murray DM, Janega JB, Blitstein JL. Design and analysis of group-randomized trials: a review of recent practices. Am J Public Health. 2004; 94:393–99. [PubMed: 14998802]

6. Raab GM, Butcher I. Balance in cluster randomized trials. Stat Med. 2001; 20:351–65. [PubMed: 11180306]

7. Carter BR, Hood K. Balance algorithm for cluster randomized trials. BMC Med Res Methodol. 2008; 8:65. [PubMed: 18844993]

8. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Avaibale at: http://www.R-project.org [accessed 22 November 2010]

9. Pocock SJ. Clinical Trials, A Practical Approach. John Wiley and Sons, Chichester. 1991

10. Wilson SR, Strub P, Buist AS, et al. Shared treatment decision making improves adherence and outcomes in poorly controlled asthma. Am J Resp Crit Care Med. 181:566–77. [PubMed: 20019345]

11. Efron B. Forcing sequential experiment to be balanced. Biometrika. 1971; 58:403–17.

12. Hagino A, Hamada C, Yoshimura I, et al. Statistical comparison of random allocation methods in cancer clinical trials. Control Clin Trials. 2004; 25:572–84. [PubMed: 15588744]

13. Zielhuis GA, Straatman H, van 't Hof-Grootenboer AE, et al. The choice of a balanced allocation method for a clinical trial in otitis media with effusion. Stat Med. 1990; 9:237–46. [PubMed: 2343219]

14. Mosteller, F.; Gilbert, JP.; Mcpeek, B. Controversies in design and analysis of clinical trials. In: Shapiro, SH.; Louis, TA., editors. Clinical Trials: Issues and Approaches. Marcel Dekker Inc; New York, NY: 1983. p. 13-64.

15. Dror S, Faraggi D, Reiser B. Dynamic treatment allocation adjusting for prognostic factors for more than two treatments. Biometrics. 1995; 51:1338–43. [PubMed: 8589225]

16. Cohen J. A power primer. Psychol Bull. 1992; 112:155–159. [PubMed: 19565683]

17. Greevy R, Lu B, Silber JH, Rosenbaum P. Optimal multivariate matching before randomization. Biostatistics. 2004; 5:263–75. [PubMed: 15054030]

18. McEntegart D. The pursuit of balance using stratified and dynamic randomization techniques: an overview. Drug Inf J. 2003; 37:293–308.

19. Ma J, King AC, Wilson SR, et al. Evaluation of lifestyle interventions to treat elevated cardiometabolic risk in primary care (E-LITE): a randomized controlled trial. BMC Fam Pract. 2009; 10:71. [PubMed: 19909549]

20. Ma J, Strub P, Camargo Jr CA, et al. The Breathe Easier through Weight Loss Lifestyle (BE WELL) intervention: a randomized controlled trial. BMC Pulm Med. 2010; 10:16. [PubMed: 20334686]

BOAT Trial (total n=612)
Baseline covariates: sex, prior asthma hospitalization, ethnicity, age, and asthma controller medication use

Step 1. 80 subjects were randomly drawn, from which three samples were formed: the first 40, first 60, and all 80

Step 2. Three randomization methods were applied to each of the three samples
a. Dynamic block randomization in blocks of 20 and 10 subjects, respectively
b. Minimization, with five allocation probabilities (p_k) of 2/3, 3/4, 0.8, 0.9, 1, respectively
c. Simple randomization

Step 3. Imbalance score (B), marginal imbalance ($\bar{b}_M$, Maxb_M), and the number of baseline covariates that were significantly different (p<0.05) by treatment arms were calculated for the 3*8=24 combinations of sample size and randomization allocation

Step 4. The outcomes (Y) for the 24 combinations were simulated for each of the 24 combinations, for four treatment effect sizes (0, 0.2, 0.5, and 0.8), a unit variance error, and one choice of β coefficient for each level of the five baseline covariates (all βs were selected from the range 0 to 1)

Step 5. The β coefficients were re-estimated using the same simulation model as in step 4, and the treatment effect significance for evaluating the power in each of the 24 combinations was recorded and compared to that by independent t-tests of pure treatment effect while assuming all baseline covariate βs to be zero.

Step 6. Steps 1-5 were repeated 1000 times to produce the results in the tables

Figure 1.
Simulation flowchart

Table 1

Distributions of the covariates selected from the BOAT trial

Balancing variable	All ($n = 612$) n (%)
Sex	
Female	266 (43.5)
Male	346 (56.5)
Prior asthma hospitalization	
None	396 (64.7)
1	216 (35.3)
Ethnicity	
White	379 (61.9)
Non-White	233 (38.1)
Age category	
30–50	120 (19.6)
51–70	256 (41.8)
>70	236 (38.6)
Asthma controller medication use	
None	135 (22.1)
1–3 days per week	123 (20.1)
4 days per week	354 (57.8)

Table 2

Comparisons of total imbalance scores B by allocation strategies

		Mean	0%	25%	50%	75%	100%
Panel (a): $n = 80$							
Dynamic block randomization							
	Block size = 20	0.010	0	0.008	0.010	0.013	0.021
	Block size = 10	0.038	0.005	0.028	0.038	0.049	0.097
Minimization							
	$P_k = 2/3$	0.155	0.008	0.080	0.130	0.204	0.747
	$P_k = 3/4$	0.090	0.004	0.048	0.073	0.117	0.482
	$P_k = 0.8$	0.067	0.004	0.035	0.056	0.084	0.380
	$P_k = 0.9$	0.037	0.002	0.021	0.032	0.048	0.177
	$P_k = 1$	0.024	0.003	0.015	0.021	0.031	0.084
Simple randomization		0.349	0.018	0.196	0.305	0.458	1.703
Panel (b): $n = 60$							
Dynamic block randomization							
	Block size = 20	0.020	0	0.015	0.021	0.025	0.039
	Block size = 10	0.065	0.004	0.048	0.064	0.081	0.153
Minimization							
	$P_k = 2/3$	0.249	0.016	0.129	0.213	0.331	1.073
	$P_k = 3/4$	0.149	0.005	0.080	0.122	0.194	0.933
	$P_k = 0.8$	0.119	0.004	0.062	0.098	0.153	0.539
	$P_k = 0.9$	0.067	0.004	0.037	0.056	0.086	0.276
	$P_k = 1$	0.044	0.003	0.026	0.040	0.055	0.275
Simple randomization		0.459	0.041	0.261	0.399	0.611	1.894
Panel (c): $n = 40$							
Dynamic block randomization							
	Block size = 20	0.047	0	0.035	0.045	0.059	0.118
	Block size = 10	0.140	0.010	0.104	0.136	0.174	0.657
Minimization							
	$P_k = 2/3$	0.445	0.027	0.251	0.386	0.570	2.109
	$P_k = 3/4$	0.308	0.026	0.165	0.265	0.406	1.293
	$P_k = 0.8$	0.245	0.023	0.132	0.205	0.317	1.427
	$P_k = 0.9$	0.161	0	0.087	0.138	0.208	1.103
	$P_k = 1$	0.101	0.010	0.062	0.089	0.127	0.452
Simple randomization		0.703	0.057	0.398	0.630	0.913	2.461

Table 3

omparisons of average and maximum marginal imbalance by allocation strategies

	$\bar{b_M}$						$\text{Max} b_M$					
	Mean	0%	25%	50%	75%	100%	Mean	0%	25%	50%	75%	100%
Panel (a): $n = 80$												
Dynamic block randomization												
Block size = 20	0.02	0	0.01	0.02	0.02	0.04	0.04	0	0.03	0.04	0.05	0.12
Block size = 10	0.04	0.01	0.03	0.04	0.05	0.07	0.08	0.03	0.06	0.08	0.10	0.24
Minimization												
$P_k = 2/3$	0.07	0.02	0.06	0.07	0.09	0.19	0.14	0.03	0.11	0.13	0.17	0.40
$P_k = 3/4$	0.06	0.02	0.04	0.05	0.07	0.15	0.11	0.02	0.08	0.11	0.13	0.31
$P_k = 0.8$	0.05	0.01	0.04	0.05	0.06	0.12	0.10	0.02	0.07	0.09	0.12	0.25
$P_k = 0.9$	0.04	0	0.03	0.03	0.04	0.09	0.07	0.02	0.05	0.07	0.09	0.22
$P_k = 1$	0.03	0	0.02	0.03	0.03	0.06	0.06	0.02	0.04	0.06	0.07	0.15
Simple randomization	0.10	0.02	0.08	0.10	0.13	0.24	0.20	0.04	0.15	0.20	0.24	0.47
Panel (b): $n = 60$												
Dynamic block randomization												
Block size = 20	0.03	0	0.02	0.03	0.03	0.06	0.06	0	0.04	0.06	0.08	0.19
Block size = 10	0.05	0	0.04	0.05	0.06	0.10	0.11	0.02	0.08	0.10	0.13	0.33
Minimization												
$P_k = 2/3$	0.09	0.02	0.07	0.09	0.11	0.21	0.18	0.03	0.13	0.17	0.22	0.47
$P_k = 3/4$	0.07	0.01	0.06	0.07	0.09	0.16	0.15	0.03	0.11	0.14	0.18	0.39
$P_k = 0.8$	0.07	0.01	0.05	0.06	0.08	0.15	0.13	0.03	0.10	0.12	0.16	0.36
$P_k = 0.9$	0.05	0	0.04	0.05	0.06	0.12	0.10	0.02	0.07	0.10	0.12	0.25
$P_k = 1$	0.04	0	0.03	0.04	0.05	0.09	0.08	0.02	0.06	0.08	0.10	0.26
Simple randomization	0.12	0.03	0.10	0.12	0.15	0.24	0.23	0.06	0.18	0.23	0.28	0.54
Panel (c): $n = 40$												
Dynamic block randomization												
Block size = 20	0.04	0	0.03	0.04	0.05	0.11	0.10	0	0.06	0.08	0.12	0.37
Block size = 10	0.07	0.01	0.06	0.07	0.09	0.16	0.16	0.03	0.11	0.15	0.19	0.48

	$\overline{b_M}$							$\text{Max}b_M$						
	Mean	0%	25%	50%	75%	100%		Mean	0%	25%	50%	75%	100%
Minimization													
$P_k = 2/3$	0.13	0.03	0.10	0.13	0.15	0.31		0.25	0.05	0.19	0.24	0.29	0.58
$P_k = 3/4$	0.11	0	0.08	0.10	0.13	0.27		0.21	0.10	0.15	0.20	0.25	0.52
$P_k = 0.8$	0.10	0	0.07	0.09	0.11	0.21		0.19	0.10	0.14	0.18	0.23	0.52
$P_k = 0.9$	0.08	0	0.06	0.08	0.09	0.16		0.16	0	0.12	0.15	0.19	0.44
$P_k = 1$	0.06	0	0.05	0.06	0.07	0.16		0.13	0	0.10	0.12	0.16	0.42
Simple randomization	0.15	0.04	0.12	0.15	0.19	0.31		0.29	0.07	0.22	0.28	0.36	0.67

Table 4

Numbers of significant *p*-values comparing the base line covariates between treatment arms

Method	Number of significant *P*-value <0.05		
	n = 80	*n* = 60	*n* = 40
Dynamic block randomization			
Block size = 20	0	0	0
Block size = 10	0	0	0
Minimization			
$P_k = 2/3$	22	33	68
$P_k = 3/4$	2	6	16
$P_k = 0.8$	0	0	11
$P_k = 0.9$	0	0	2
$P_k = 1$	0	0	0
Simple randomization	248	225	262

Table 5

Comparisons of type I error rates, power, and gains in sample size by allocation strategies

Panel (a)

	Type I error rate		
	$n = 80$	$n = 60$	$n = 40$
Dynamic block randomization			
Block size = 20	4.2%	4.1%	5.0%
Block size = 10	5.1%	3.7%	4.5%
Minimization			
$P_k = 2/3$	4.5%	4.3%	4.0%
$P_k = 3/4$	4.2% 3.9%	3.9%	4.5%
$P_k = 0.8$	4.9%	5.0%	4.5%
$P_k = 0.9$	5.0%	5.0%	4.8%
$P_k = 1$	4.9%	5.3%	4.5%
Simple randomization	4.8%	4.9%	4.5%

Panel (b)

	Power ($n = 80$)			Power ($n = 60$)			Power ($n = 40$)			Gain in sample size
Effect size	0.8	0.5	0.2	0.8	0.5	0.2	0.8	0.5	0.2	$n = 60$ and effect size = 0.8
No-confounding reference	94.7%	60.9%	14.3%	87.3%	49.0%	11.8%	71.6%	35.2%	9.2%	124%
Dynamic block randomization										
Block size = 20	95.1%	60.6%	16.0%	87.0%	48.8%	11.8%	66.8%	32.7%	8.6%	120%
Block size = 10	95.3%	60.0%	13.0%	86.9%	45.5%	10.0%	68.1%	31.0%	9.3%	120%
Minimization										
$P_k = 2/3$	92.9%	58.3%	12.3%	84.1%	44.0%	9.9%	61.7%	27.8%	8.1%	112%
$P_k = 3/4$	93.4%	57.9%	14.5%	84.7%	45.4%	11.4%	63.3%	28.4%	8.2%	112%
$P_k = 0.8$	94.0%	58.9%	15.0%	83.9%	44.8%	10.9%	63.7%	30.5%	8.1%	112%
$P_k = 0.9$	94.4%	57.8%	12.8%	84.9%	46.7%	10.4%	64.8%	30.0%	7.7%	112%
$P_k = 1$	94.2%	58.5%	11.4%	83.6%	43.9%	10.3%	62.5%	27.7%	7.5%	108%
Simple randomization	91.6%	54.6%	11.7%	80.7%	43.6%	11.3%	59.2%	27.3%	7.2%	100%