Statistical Properties of Covariate-Adaptive Randomization Procedures for Trials with Unequal Treatment Allocation Ratios

Oleksandr Sverdlov[1], Yiwei Zhang[1]

[1]Novartis Pharmaceutical Corporation, One Health Plaza, East Hanover, NJ 07936

Abstract

Many modern randomized clinical trials involve multiple treatment arms and use designs to achieve treatment groups of unequal size. Some rationales for using unequal allocation include better statistical efficiency of treatment comparisons, cost-efficiency in situations when some treatment protocols are more expensive than others, and potential ethical constraints. Recently, several designs expanding covariate-adaptive randomization procedures to trials with unequal allocation have been proposed in the literature. In this paper we consider several restricted randomization designs which are fully randomized and maintain throughout the trial the target allocation ratio (if necessary, within each stratum). We compare operating characteristics of these designs through a simulation study. We study such issues as balance, randomness of treatment assignments, variations in the allocation ratio, and type I error and power of statistical tests.

Key Words: Allocation ratio preserving, Randomization design, Unequal allocation

1. Motivating Example

A randomized open-label proof-of-concept clinical trial is to be conducted to compare the effects of a high dose of an experimental drug (treatment 1) versus the reference (treatment 2) in chronic hepatitis C patients. A 2:1 active-placebo randomization rate is to be used, and the total planned study sample size is $n = 120$. It is thought that viral load measured at baseline is an important prognostic covariate. As such randomization will be stratified by viral load measured at screening ($\geq$ 800,000 IU/mL (5.903 $\log_{10}$) or $<$ 800,000 IU/mL (5.903 $\log_{10}$)). A priori the patient population proportion for each stratum is unknown and it is difficult to project how many study patients will be enrolled in each stratum.

The investigator emphasized the importance of maintaining the 2:1 treatment allocation ratio both within each stratum and overall in the trial. One strategy to achieve this goal is to use stratified permuted block design which achieves the desired allocation ratio after every cohort of 3 patients in each stratum; however since the study is open label, there may be a risk of selection bias with such a strategy. One may also consider increasing the block size to 6 or 9; however since the number of patients in each stratum is unknown a design with large block size may result in deviations from the desirable 2:1 allocation.

Which randomization design should be used in our trial example? A "good" design should be fully randomized and maintain the target allocation ratio within each stratum throughout the entire trial. In this paper we study several restricted randomization designs which can be used to achieve this goal. Although our motivating example is a two-arm trial, the discussed methodology is applicable in a general ($K \geq 2$)-arm trial setup. Throughout the paper, we focus on stratum-specific randomization procedures; in other words, each stratum utilizes a separate restricted randomization sequence. The methodology can be also extended to more complex covariate-adaptive randomization procedures; however this is not done in this paper and is left for the future work.

Therefore, the main focus of the current work is on statistical properties of restricted randomization procedures for ($K \geq 2$)-arm trials with pre-specified fixed unequal allocation which, if necessary, can be implemented within each stratum. We will attempt to identify a design that provides "best" tradeoff between balance (achieving the target allocation ratio) and randomness.

2. Restricted Randomization for Multi-Arm Trials with Fixed Unequal Allocation

Consider a clinical trial where n subjects must be sequentially randomized among $K \geq 2$ treatment groups to achieve some fixed allocation ratio $w_1 : w_2 : \ldots : w_K$, where w_i's are positive, not necessarily equal, integers with the greatest common divisor of 1. Let $\rho_i = \frac{w_i}{\sum_{k=1}^{K} w_k}$ denote the target treatment allocation proportions, where $0 < \rho_i < 1$ and $\sum_{k=1}^{K} \rho_k = 1$.

In *restricted randomization*, treatment allocation probabilities for any subject are conditional on history of past treatment assignments (Rosenberger and Lachin, 2002). Let P_{ij} denote the probability that subject j is assigned to treatment i ($0 < P_{ij} < 1$, $\sum_{i=1}^{K} P_{ij} = 1$). Let T_j denote the treatment assignment for the jth subject (T_j can take values $1, \ldots, K$). A general restricted randomization design is defined as

$$\begin{aligned} P_{ij} &= \Pr(\text{Subject } j \text{ is assigned to treatment } i) \\ &= \Pr\left(T_j = i \middle| T_1, \ldots, T_{j-1}\right), \; j = 2, 3, \ldots, n. \end{aligned} \qquad (1)$$

After n subjects have been randomized among the K treatments, the number of subjects assigned to treatment i is $N_i(n) = \sum_{j=1}^{n} \mathbf{1}\{T_j = i\}$, where $\mathbf{1}\{\cdot\}$ denotes an indicator function. In general, $N_i(n)$ are random with $\sum_{i=1}^{K} N_i(n) = n$.

We are interested in randomization procedures that attempt to achieve, at least approximately, the target treatment numbers $n\rho_1, \ldots, n\rho_K$. In addition, if randomization is stratified and the number of subjects in each stratum is unknown at the trial outset, it is desirable that the treatment allocation proportions are close to the target proportions at every step; i.e. $j^{-1}N_i(j) \approx \rho_i$ for $i = 1, \ldots, K$ and $j \geq 2$.

3. Randomization Designs Studied in This Paper

In this paper we shall explore six competing randomization designs:

(i) Completely Randomized Design (CRD): Every subject is randomized to treatment groups with fixed probabilities:

$$P_{ij} = \rho_i, \quad i = 1, \dots, K. \qquad (2)$$

The CRD is the most random procedure, but the actual treatment numbers may deviate far from the target numbers.

(ii) Permuted Block Design (PBD): Treatment assignments are made at random within blocks of size $b = w_1 + w_2 + \cdots + w_K$. In each block, exactly w_i subjects are assigned to treatment i. The PBD, by construction, achieves exactly the required treatment numbers after every cohort of b patients, but it periodically results in treatment assignments that can be guessed with high probability which may lead to selection bias.

(iii) Block Urn Design (BUD): This design was proposed by Zhao and Weng (2011), to provide a more random design than the PBD. The design can be described as follows. Consider two urns, Active and Inactive. At the beginning of the trial, the Inactive urn is empty and the Active urn contains $b = \lambda W$ balls, where $W = \sum_{i=1}^{K} w_i$ is a "minimal balanced set" and w_i is the number of balls representing the ith treatment. The parameter λ is a small positive integer which determines the number of minimal balanced sets in the block of size b. Note that the BUD with $\lambda = 1$ is equivalent to the PBD(bsize=W).

The treatment assignments are made sequentially by drawing balls at random from the Active urn. If type i ball is drawn, treatment i is assigned and the ball is placed in the Inactive urn. The procedure is repeated until a minimal balanced set is collected in the Inactive urn, at which instant these W balls are returned into the Active urn. The described procedure is further repeated until the pre-specified number of subjects has been enrolled in the study.

Zhao and Weng (2011) provide a closed-form expression for the randomization probabilities for BUD. Let $N_{i,j-1}$denote the number of treatment i assignments among the first $(j-1)$ subjects, and let $k_{j-1} = \min_{1 \le i \le m}(\text{int}(N_{i,j-1}/w_i))$ denote the number of minimal balanced sets in previous assignments ($\text{int}(x)$ returns the greatest integer less than or equal to x). Then subject j is randomized to treatments with probabilities

$$P_{ij} = \frac{w_i\lambda + w_i k_{j-1} - N_{i,j-1}}{W\lambda + W k_{j-1} - (j-1)}, \quad i = 1, \dots, K. \qquad (3)$$

The BUD results in deterministic assignments less frequently than the PBD, and the occurrence of deterministic assignments for the BUD is random.

(iv) Drop-the-Loser Urn Design (DL): The DL design was developed by Ivanova (2003) in the context of multi-arm binary response trials with response-adaptive randomization. Here we consider its application for fixed unequal allocation.

Consider an urn containing balls of $K+1$ types: types $1, \dots, K$ represent treatments and type 0 is the immigration ball. Initially, the urn contains one immigration ball and w_i treatment balls. The treatment assignments are made sequentially by drawing balls at random from the urn. If a type i ball is drawn ($i = 1, \dots, K$), treatment i is assigned to the

subject and the ball is not replaced into the urn. If type 0 ball is drawn, it is replaced into the urn along with $aw_1, \ldots, aw_K$ balls of types $1, \ldots, K$. The parameter a is some positive integer (larger values of a imply greater amount of randomness in the experiment). The DL rule is a fully randomized procedure, and it is known to be asymptotically best (Hu et al. 2006).

(v) Doubly Adaptive Biased Coin Design (DBCD): The DBCD procedure was proposed by Hu and Zhang (2004) in the context of response-adaptive randomization. Here we consider its application for fixed unequal allocation.

Let $(\rho_1, \ldots, \rho_K)$ denote the vector of target allocation proportions ($\rho_i = \frac{w_i}{\sum_{k=1}^{K} w_k}, i = 1, \ldots, K$). Initial treatment assignments are made completely at random (2) until each group has at least one subject. Subsequent treatment assignments are made as follows. Let $N_{i,j-1}$denote the number of treatment i assignments among the first $(j-1)$ subjects. Then the treatment randomization probabilities for the jth subject are

$$P_{ij} = \frac{\rho_i \left(\frac{\rho_i}{N_{i,j-1}/(j-1)}\right)^{\gamma}}{\sum_{k=1}^{K} \rho_k \left(\frac{\rho_k}{N_{k,j-1}/(j-1)}\right)^{\gamma}}, \; i = 1, \ldots, K, \qquad (4)$$

where $\gamma \geq 0$ is a user-defined parameter controlling the degree of randomness ($\gamma = 0$ is most random and $\gamma \to \infty$ is almost deterministic procedure). Importantly, treatment allocation proportions of the DBCD procedure follow an asymptotically normal distribution: $\sqrt{j}\left(N_{ij}/j - \rho_i\right) \to N\left(0, (1+2\gamma)^{-1}\rho_i(1-\rho_i)\right), \; i = 1, \ldots, K$.

(vi) Minimum Quadratic Distance Constrained Balance Randomization (MinQD): The design was proposed by Titterington (1983), in the context of multi-arm randomized trials with covariate-adaptive randomization and balanced allocation. Here we consider an extension of this procedure to clinical trials with unequal allocation.

Consider a point in the trial when $j-1$ subjects have been randomized among the K treatments, and let $(N_{1,j-1}, \ldots, N_{K,j-1})$ denote the corresponding treatment numbers $(\sum_{i=1}^{K} N_{i,j-1} = j-1)$. The randomization rule for the jth subject is as follows.

- For $k = 1, \ldots, K$, compute B_k, the hypothetical "lack of balance" which results from assigning the jth subject to treatment k: $B_k = \max_{1 \leq i \leq K} |N_{ij}^k/j - \rho_k|$, where $N_{ij}^k = \begin{cases} N_{i,j-1} + 1, & if \; i = k; \\ N_{i,j-1}, & if \; i \neq k. \end{cases}$
- The treatment randomization probabilities for the jth subject $(P_{1j}, \ldots, P_{Kj})$ are determined as a solution to the constrained optimization problem:

  $\text{minimize}_{P_{1j}, \ldots, P_{Kj}} \sum_{i=1}^{K} \left(P_{ij} - \rho_i\right)^2$
  subject to $\sum_{i=1}^{K} B_i P_{ij} \leq \eta B_{(1)} + (1-\eta) \sum_{i=1}^{K} B_i \rho_i$
  and $\sum_{i=1}^{K} P_{ij} = 1; \quad 0 \leq P_{ij} \leq 1, \; i = 1, \ldots, K, \qquad (5)$

Essentially, the idea is to quadratic distance between the vector of randomization probabilities and the target allocation vector subject to a constraint on expected "lack of

balance". The user-defined parameter η ($0 \leq \eta \leq 1$) that controls the degree of randomness in the experiment ($\eta = 0$ is most random and $\eta = 1$ is almost deterministic procedure).

4. Design Performance Metrics

4.1 Balance and Randomness

Two important characteristics of a restricted randomization procedure are the *Imbalance* and the *Forcing Index* (measure of lack of randomness).

For a randomization design with n subjects, treatment imbalance can be defined as a distance, in some metric, between the achieved allocation $(N_1(n), \dots, N_K(n))$ and the target allocation $(n\rho_1, \dots, n\rho_K)$. In Euclidean metric, imbalance is defined as

$$Imb(n) = \sqrt{\sum_{i=1}^{K}(N_i(n) - n\rho_i)^2} \qquad (6)$$

Small values of $Imb(n)$ are desirable; $Imb = 0$ corresponds to a perfectly balanced trial.

A Forcing Index for the j-th subject (FI_j) is defined as the distance from vector of treatment randomization probabilities for the subject $\boldsymbol{P}_j = (P_{1j}, \dots, P_{Kj})$ to the "completely random" vector of randomization probabilities $\boldsymbol{\rho} = (\rho_1, \dots, \rho_K)$, i.e. $FI_j = \sqrt{\sum_{i=1}^{K}\left(P_{ij} - \rho_i\right)^2}$. If $FI_j = 0$, then the treatment assignment for patient j is made completely at random. A Forcing Index for the design is defined as (Heritier et al. 2005)

$$FI(n) = n^{-1}\sum_{j=1}^{n} FI_j \qquad (7)$$

The smaller $FI(n)$ is, the less predictable is the randomization procedure (the value of zero corresponds to complete randomization).

In general, procedures with low imbalance have high forcing index and vice versa. Atkinson (2002) suggested plotting imbalance versus lack of randomness for a range of sample sizes to find *admissible* randomization procedures.

4.2 The Allocation Ratio Preserving (ARP) Property

Kuznetsova and Tymofyeyev (2014) discuss a very important feature of a restricted randomization procedure targeting unequal allocation—the ARP property. Let $\boldsymbol{\rho} = (\rho_1, \dots, \rho_K)$ denote the target allocation. Let Ω denote the set of all allocation sequences $\boldsymbol{\omega} = \{\omega_1, \dots, \omega_n\}$ the allocation procedure can produce. Here $\omega_j = i$ if treatment i is assigned at the jth allocation ($j = 1, \dots, n$). Let $p(\boldsymbol{\omega})$ denote the probability with which sequence $\boldsymbol{\omega}$ occurs in Ω.

A randomization procedure is said to possess the ARP property if for all $j = 1, \dots, n$ and $i = 1, \dots, K$,

$$\sum_{\boldsymbol{\omega}\in\Omega} p(\boldsymbol{\omega})\mathbf{1}\{\omega_j = i\} = \rho_i. \qquad (8)$$

The left-hand side of (8) represents the unconditional probability of allocating treatment i at the jth allocation.

In essence, an ARP procedure maintains the unconditional allocation ratio at each step, whereas a non-ARP procedure has variations in the allocation ratio from allocation to allocation which may provide a potential for selection, evaluation and accidental bias even in double-blind studies (Kuznetsova and Tymofyeyev, 2014). As such, an ARP procedure is desired.

4.3 Power and Type I Error

In practice, it is important that the randomization design results in valid and powerful statistical inference. Validity means that the type I error rate is maintained at the nominal level (e.g. 5%), and high power of statistical test is desired to detect treatment difference when such exists.

5. Simulation Study

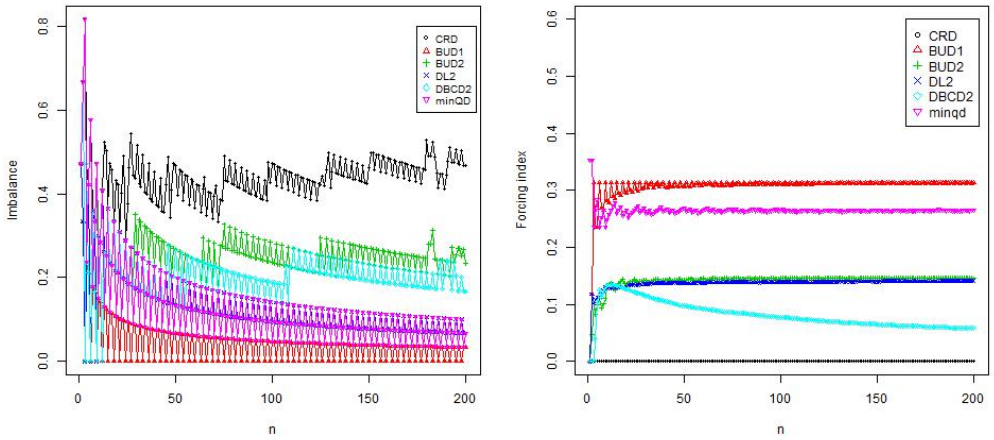
A simulation study was performed to compare the performance of six randomization designs described in Section 2, for a two-arm trial with 2:1 allocation, for a range of sample sizes up to 200. Specifically, the designs are: (i) CRD; (ii) PBD(bsize=3); (iii) BUD($\lambda = 2$); (iv) DL($a = 2$); (v) DBCD($\gamma = 2$); and (vi) MinQD($\eta = 0.50$). For each design and each sample size, the operating characteristics were computed based on 10,000 simulation runs.

Figure 1 displays interplay between balance and randomness. Figure 1A is a plot of median imbalance versus sample size. The ranking of the procedures (most balanced to least balanced) is as follows: PBD(bsize=3); DL($a = 2$) and MinQD($\eta = 0.50$); DBCD($\gamma = 2$); BUD($\lambda = 2$); and CRD. Figure 1B shows median Forcing Index versus sample size. The most random design is CRD ($FI = 0$) and the least random design is PBD(bsize=3) ($FI \approx 0.3$). Interestingly, DL($a = 2$) and BUD($\lambda = 2$) designs are very similar in terms of randomness ($FI \approx 0.13$), and MinQD($\eta = 0.50$) design has a much higher degree of determinism ($FI \approx 0.26$). At the same time, DBCD($\gamma = 2$) is similar to DL($a = 2$) and BUD($\lambda = 2$) designs for small sample sizes, and it becomes more random as sample size increases. Figure 1C shows a plot of the distance from (Median Imbalance, Median Forcing Index) to (0,0). For an "admissible" design, this distance should be small. From Figure 1C, it is clear that the DL($a = 2$) rule has best overall performance: for sample sizes less than 50, it outperforms all designs except for the PBD(bsize=3); however for sample sizes greater than 50, the DL($a = 2$) rule is a clear winner.

Figure 2 shows simulated unconditional probability of assigning treatment 1 for five randomization designs (PBD(bsize=3), BUD($\lambda = 2$), DL($a = 2$), DBCD($\gamma = 2$), and MinQD($\eta = 0.50$)) for sample size up to $n = 120$ (the CRD is not considered here as it possesses the ARP property by definition). For an ARP design, the randomization probability for treatment 1 should be ~0.67 for any sample size. One can see from Figure 2 that MinQD($\eta = 0.50$)) design is a non-ARP procedure as it has fluctuations in the allocation probability. However, the other four designs maintain the allocation probability close to the target value of 2/3 (ARP procedures).

Finally, Table 1 shows simulated type I error and power of a 2-sample t-test with the fixed total sample size $n = 24$ and 2:1 allocation for six randomization designs. For each design, data (responses) were simulated according to the model: $Y_j = (1 - \delta_j)\mu + \varepsilon_j$, $j = 1, \ldots, 24$, where $\delta_j = 1$ (or 0) if the jth subject is assigned to treatment 1 (or 2) and ε_j are independent error terms, with standard normal distribution. From Table 1, all designs approximately maintain the type I error at 5% (the column $\mu = 0$). The DL rule is generally most powerful among the six designs studied (the columns $\mu = 0.5$, $\mu = 1$ and $\mu = 1.5$).

(A) Median Imbalance **(B) Median Forcing Index**



(C) Distance from (Median Imbalance, Median Forcing Index) to (0,0)

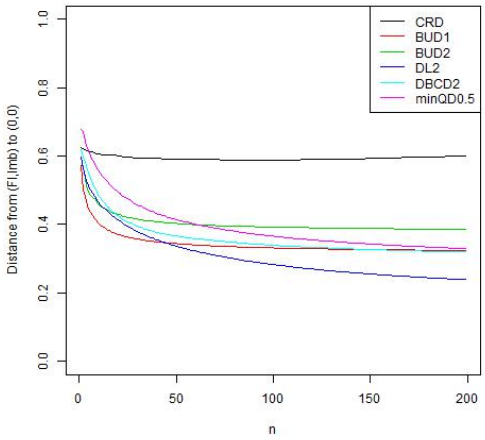
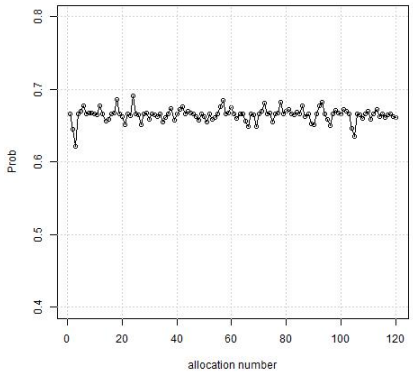
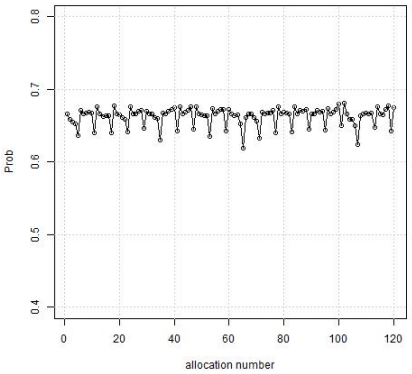


Figure 1: Imbalance and Forcing Index of six randomization designs.

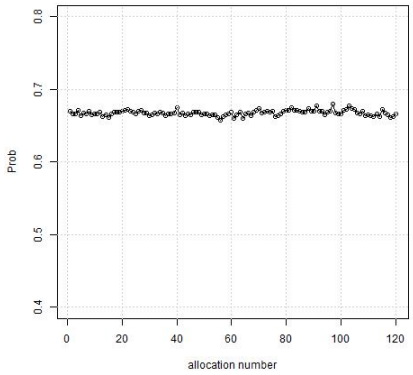
(A) PBD(bsize=3)



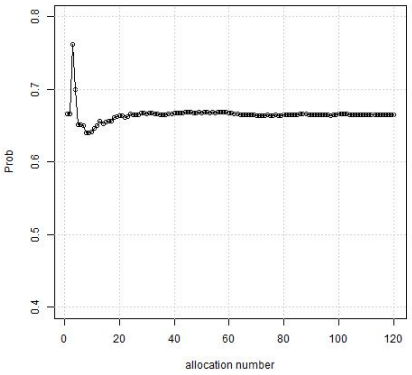
(B) BUD($\lambda = 2$)



(C) DL($a = 2$)



(D) DBCD($\gamma = 2$)



(E) MinQD($\eta = 0.50$)

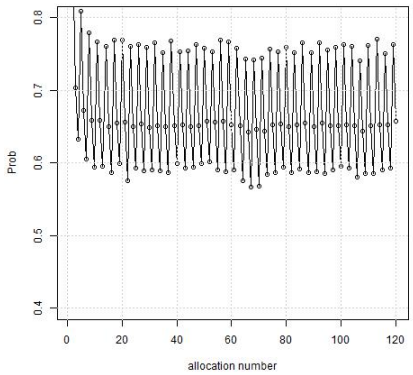


Figure 2: Simulated unconditional allocation probability for treatment 1.

Table 1: Type I error and power for six randomization designs

	$\mu = 0$	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$
CRD	0.047	0.192	0.584	0.900
PBD(bsize=3)	0.047	0.166	0.588	0.915
BUD($\lambda = 2$)	0.049	0.191	0.608	0.913
DL($a = 2$)	0.050	0.200	0.626	0.925
DBCD($\gamma = 2$)	0.045	0.188	0.578	0.911
MinQD($\eta = 0.50$)	0.057	0.207	0.581	0.892

6. Concluding Remarks

In this paper we studied statistical properties of (stratified) restricted randomization designs with unequal allocation targets. These designs attempt to maintain some pre-specified fixed unequal allocation ratio throughout the trial. We focused on a simple case of two treatment arms and 2:1 allocation; our findings provide important insight into the performance of various randomization designs. Further research and simulation studies are warranted for ($K > 2$)- arm trials with unequal allocation.

We found that the drop-the-loser rule of Ivanova (2003) applied with fixed unequal allocation provides the best tradeoff between balance and randomness for a range of sample sizes. Importantly, this procedure possesses the ARP property (the unconditional allocation probability is held constant at each step) and the design results in valid and powerful inference using two-sample t-test. In our ongoing and future work we plan to consider covariate-adaptive extensions of the DL rule to multi-arm trials with discrete and continuous covariates and unequal allocation targets.

References

Atkinson AC (2002) The comparison of designs for sequential clinical trials with covariate information. *Journal of the Royal Statistical Society. Series A* **165(2)**, 349-373.

Heritier S, Gebski V, Pillai A (2005) Dynamic balancing randomization in controlled clinical trials. *Statistics in Medicine* **24**, 3729-3741.

Hu F, Rosenberger WF, Zhang LX (2006) Asymptotically best response-adaptive randomization procedures. *Journal of Statistical Planning and Inference* **136**, 1911-1922.

Hu F, Zhang LX (2004) Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics* **32(1)**, 268-301.

Ivanova A (2003) A play-the-winner-type urn design with reduced variability. *Metrika* **58**, 1-13.

Kuznetsova O, Tymofyeyev Y (2014) Covariate-adaptive randomization procedures with unequal allocation. Accepted for the edited volume "Modern adaptive randomized clinical trials: statistical, operational, and regulatory aspects".

Rosenberger WF, Lachin JM (2002) *Randomization in Clinical Trials: Theory and Practice*, New York: Wiley.

Titterington DM (1983) On constrained balance randomization for clinical trials. *Biometrics* **39(4)**, 1083-1086.

Zhao W, Weng Y (2011) Block urn design—A new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemporary Clinical Trials* **32**, 953-961.