

RESEARCH

Open Access



Comparison of Pocock and Simon's covariate-adaptive randomization procedures in clinical trials

Guogen Shan^{1*}, Yulin Li¹, Xinlin Lu¹, Yahui Zhang¹ and Samuel S. Wu¹

Abstract

When multiple influential covariates need to be balanced during a clinical trial, stratified blocked randomization and covariate-adaptive randomization procedures are frequently used in trials to prevent bias and enhance the validity of data analysis results. The latter approach is increasingly used in practice for a study with multiple covariates and limited sample sizes. Among a group of these approaches, the covariate-adaptive procedures proposed by Pocock and Simon are straightforward to be utilized in practice. We aim to investigate the optimal design parameters for the patient treatment assignment probability of their developed three methods. In addition, we seek to answer the question related to the randomization performance when additional covariates are added to the existing randomization procedure. We conducted extensive simulation studies to address these practically important questions.

Keywords Additional covariates, Allocation predictability, Covariate adaptive randomization, Imbalance score, Pocock and Simon, Statistical power

Introduction

Randomized controlled trials (RCTs) are critical in evaluating the effectiveness of new treatments and interventions in clinical research. Randomization is a fundamental element of clinical trials, primarily aimed at preventing selection bias and enhancing the validity and accuracy of the results [1–4]. When sample size is large enough, influential factors are more likely to be balanced approximately. However, in the case of a large number of influential factors, complete randomization (CR) carries a substantial risk of chance imbalance in these factors [5]. For trials with limited sample sizes, CR can lead to significant disparities in the number of participants across groups [6–8]. Such chance imbalances in baseline

covariates and group sample sizes can ultimately undermine the statistical power and may present challenges in comparing treatment groups and interpreting trial results. For instance, in clinical trials involving relatively small sample sizes, simple randomization methods have been shown to produce an unequal distribution of both participants and influential factors across treatment and control groups [9, 10].

Stratified block randomization (SBR) is frequently employed to mitigate imbalances in baseline covariates across groups. In a review article by Lin et al. [3], they found that SBR designs were used in the majority of trials (close to 70%). However, its capability is constrained to a limited array of factors. With an extensive number of strata, the efficacy of SBR in maintaining balance among treatment groups can be compromised as some strata may end up with very few participants [5, 10, 11]. When the global treatment balance is the target of a randomization procedure, some existing designs may be utilized, such as the permuted block design (PBD) [12],

*Correspondence:

Guogen Shan
gshan@ufl.edu

¹ Department of Biostatistics, University of Florida, Gainesville 32610, FL, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Efron's biased coin design [13], Wei's urn design [14]. Over recent decades, various innovative restricted randomization designs have been developed, building upon the biased coin design and the urn design. These include the big stick design, the biased coin design with imbalance tolerance, the Ehrenfest urn design and the block urn design [15–18]. These recent methods have demonstrated better performance than traditional methods, particularly in maintaining a delicate balance between imbalance score and allocation predictability.

Covariate adaptive randomization (CAR) is a popular minimization method to achieve balance over a broader spectrum of covariates [19–25]. The concept of CAR approach was first introduced with the focus on dynamically identifying the treatment that could minimize the overall imbalance across covariates [23]. Pocock and Simon later proposed a generalized and flexible CAR approach [26]. One of the primary features of their method is the incorporation of allocation probabilities [4, 27–29]. After temporarily assigning a newly recruited participant to all the available groups, each group ends up with its own total number of imbalances. Then, instead of strictly assigning the new participant to the group with lowest total number of imbalances, the participant will be allocated to that group with a probability [22]. This means that while the option that would minimize imbalance is weighted more heavily, there is still a chance that the participant could be allocated to other groups to improve randomness. The employment of the probabilistic element introduces another layer of randomization, which can be beneficial for reducing bias and confounding [30–33]. Another established method for achieving covariate balance is the dynamic hierarchical randomization which offers an alternative to minimization when there are too many stratification factors. Such approaches not only maintain randomness but also ensure balance across each covariate level, accommodating varying degrees of imbalance among different covariates [34].

In this article, we critically examine several research questions emerging from the application of Pocock and Simon's (PS) method in ongoing clinical trials [35]. First, Pocock and Simon proposed three methods to determine the treatment assignment probability after each new patient. A pivotal aspect of our research is to identify the optimal parameters within these methods that yield the best possible performance. Furthermore, our study delves into the exploration of the impact on the randomization method performance when adding new prognostic factors to the CAR procedure. Specifically, we aim to explore to which extent the overall performance can be enhanced through the incorporation of different number of factors. This will allow us to ascertain whether the inclusion of

more factors leads to increasingly accurate and reliable results, or if there are diminishing returns beyond a certain point. In addition, we proactively consider increasing the number of study sites to mitigate the risks associated with low patient recruitment rates. Lastly, our study also conduct a comparative analysis of statistical power between CR and PS designs.

The structure of this article is outlined as follows. In **Methods** section, we illustrate Pocock and Simon's minimization method. We describe the three formulas used for calculating treatment assignment probabilities, the calculation of treatment imbalance score and allocation predictability that evaluate randomization procedures. Then, in **Numerical study** section, we first run simulation studies to address the three research questions mentioned above. After that, we use data from a trial for patients with pleural infection to demonstrate the application of CAR designs and further examine our research questions. In **Discussion** section, we provide some comments about potential topics we can explore in the future.

Methods

For a study with K treatments, suppose there are I prognostic factors that need to be balanced to assess the treatment effect properly. We consider categorical prognostic factors in this article. Suppose J_i is the total possible level of the i -th prognostic factor, where $i = 1, 2, \dots, I$. For the i -th prognostic factor, let X_i be the level value: $X_i \in \{1, 2, \dots, J_i\}$. One example is the disease severity with three levels as mild, moderate, and severe.

When the i -th factor is considered, the number of already enrolled participants can be organized in a K by J_i contingency table. Suppose $N_{k,x_i|i}$ is the number of participants being assigned to treatment k whose i -th factor value is x_i . Then, $\sum_{k=1}^K \sum_{x_i=1}^{J_i} N_{k,x_i|i}$ is the sum of all numbers from the K by J_i table, and that is the total enrolled participants so far.

When a new participant is recruited for this study, that participant is ready for randomization after collecting the information of the I prognostic factors: $\{x_1, x_2, \dots, x_I\}$. If treatment k is assigned to this new participant, $N_{k,x_i|i}$ is increased by 1, and other values in the K by J_i table for the i -th factor remain the same. The values at the x_i -th column of the K by J_i table become

$$Y(x_i|i, k) = (N_{1,x_i|i}, \dots, N_{k-1,x_i|i}, N_{k,x_i|i} + 1, N_{k+1,x_i|i}, \dots, N_{K,x_i|i}).$$

These numbers are then used to calculate the imbalance value for the i -th factor after adding one more participant to treatment k . Following Pocock and Simon [26], that imbalance value is denoted as $d_{ik} = D(Y(x_i|i, k))$, where D is a function to calculate the imbalance value. In practice, three methods based on the range, variance, and standard deviation of $Y(x_i|i, k)$ can be used, and we refer

them to be as the range method, the var method, and the SD method to compute the imbalance value. A total of I imbalance values d_{ik} ($i = 1, 2, \dots, I$), can be calculated by using one of the three methods. These imbalance values are then used to calculate the total imbalance score for assigning this new participant to treatment k . These individual imbalance values can be combined in multiple approaches to obtain the total score. We consider an equal weight in calculating the total imbalance score during the randomization procedure

$$S_k = \sum_{i=1}^I d_{ik}.$$

These total imbalance scores: S_k ($k = 1, 2, \dots, K$), are sorted from the smallest to the largest, with the associated probabilities p_1 to p_K , where $p_1 \geq p_2 \geq \dots \geq p_K$, and $\mathbf{P} = \{p_1, p_2, \dots, p_K\}$. A treatment having a small value of imbalance score has a high probability in the treatment assignment. It should be noted that the computed imbalance score S_k is the covariate imbalance score.

Treatment assignment probability

Three formulas for \mathbf{P} were proposed by Pocock and Simon [26]. The first two formulas were developed by using the ordering of S_k values. The first one is relatively simple

$$p_1 = p \text{ and } p_k = \frac{1-p}{K-1}, k = 2, 3, \dots, K.$$

To reduce the overall imbalance score, p should be larger than $1/K$. In the illustrated example in Pocock and Simon [26], for a study with 3 treatments, p was chosen to be $2/3$. Once the value of p is chosen, the treatment assignment probabilities are determined. We refer to this approach as the PS_p approach. The second formula is

$$p_k = q - \frac{2(Kq-1)}{K(K+1)}k,$$

where $k = 1, \dots, K$, and q is a constant between $1/K$ and $2/(K-1)$. We refer to this as the PS_q approach.

In addition to the ordering of S_k values, the third formula used the S_k values in computing the treatment assignment probabilities as

$$p_k = \frac{1}{K-t} \left[1 - \frac{tS_k}{\sum S_k} \right], k = 1, \dots, K,$$

where t is a constant between 0 and 1. We refer to this approach as the PS_t approach. The last formula is relatively more complicated than the first two formulas.

Once the value of p in PS_p or q in PS_q is chosen, the probabilities are determined. However, in the last formula, the probabilities could be changed as the values of S_k are updated after each new patient.

In general, the PS_p approach assigns the treatment arm with the minimal total imbalance score a probability higher than the mean, while the remaining arms equally share the leftover probability; the PS_q approach distributes probabilities in a monotonically decreasing pattern based on the rank of the total imbalance score with a constant decrement value; the PS_t approach assigns treatment probabilities by inversely weighting them against the total imbalance scores for each arm and dynamically updates the probabilities with each new patient.

In the simulation studies, we included the following randomization methods: CR, stratified block randomization, the stratified big stick design (SBSD) by Zhao [25], and the hierarchical dynamic balancing randomization (HDBR) by Heritier et al. [36]. The SBSD design is a two-stage CAR randomization method that can improve balance and randomness of a trial as compared to the traditional stratified permuted block randomization. The HDBR design is a dynamic balancing randomization with the constraint on the importance ordering of factors [36]. The R codes can be downloaded from the GitHub page: https://github.com/AdaptiveDesign/CAR_randomization.

Treatment balance and allocation predictability

In randomized clinical trials, treatment balance and allocation predictability are two essential metrics used to evaluate the performance of randomization procedures. These two values are traditionally calculated at the end of the trial, after all participants have been randomized. Suppose $N_k(m)$ is the number of participants being assigned to treatment k after m patients enrolled, where $k = 1, \dots, K$ and $0 \leq N_k(m) \leq m$. Then, we have $\sum_{k=1}^K N_k(m) = m$. In order to calculate the final treatment imbalance score, we begin by applying the range method [37] to compute the imbalance value following the enrollment of each individual patient

$$D(m) = \max(\{N_k(m)\}_{k=1}^K) - \min(\{N_k(m)\}_{k=1}^K),$$

where $m = 1, 2, \dots, N$. Following the literature [38–40], we utilized the treatment imbalance score (IS) as

$$IS = \frac{1}{N} \sum_{m=1}^N \frac{D(m)^2}{m}.$$

The value of imbalance score represents the lost information [40]. A randomization procedure with a low value of imbalance score is preferable. The treatment

imbalance score, IS, is used to compare different randomization methods including the CR in which there is no covariate involved.

Allocation predictability (AP) is the probability of accurately predicting treatment assignments from the calculated imbalance score, under the assumption that the investigator consistently predicts the treatment with the higher likelihood of allocation [11, 36]. AP is a commonly used metric for assessing the lack of randomness of an allocation process, with a lower AP indicating a more unpredictable randomization procedure [8]. When the m -th participant is assigned to the treatment group with the lowest imbalance score, the guess is correct, denoted as $G_m = 1$. Otherwise $G_m = 0$ for wrong guess. For a two-arm study designed with complete randomization, the probability of $P(G_m = 1)$ is 50%. For a study with K treatments, allocation predictability can be defined as

$$AP = \left(\frac{1}{N} \sum_{m=1}^N P(G_m = 1) - \frac{1}{K} \right) \frac{K}{K-1}.$$

The range of $\frac{1}{N} \sum_{m=1}^N P(G_m = 1)$ is from 0 to 1. The quantity $\frac{K}{K-1}$ is added to the AP calculation to make sure

that upper limit of AP is 1. For a CR design, the AP value will be very close 0. Thus, it is preferable to have a randomization procedure with a low value of AP^2 .

In the simulation studies, we used rescaled IS and AP value (IS_r and AP_r) to calculate the weighted score for comparing different methods. The weighted score is defined as

$$\Omega = \sqrt{\frac{IS_r^2 + AP_r^2}{2}}.$$

When multiple values of the parameter (e.g., p in the PS_p method) are studied, the computed IS values are rescaled to 0 and 1 as the IS_r values. The AP_r values are calculated by using AP values in a similar approach. A randomization procedure having a small weighted score Ω is preferable.

Numerical study

We first run simulation studies to identify the optimal parameter of the three treatment assignment methods (PS_p , PS_q , and PS_t) having a good performance with regard to IS, AP and the weighted score Ω [41].

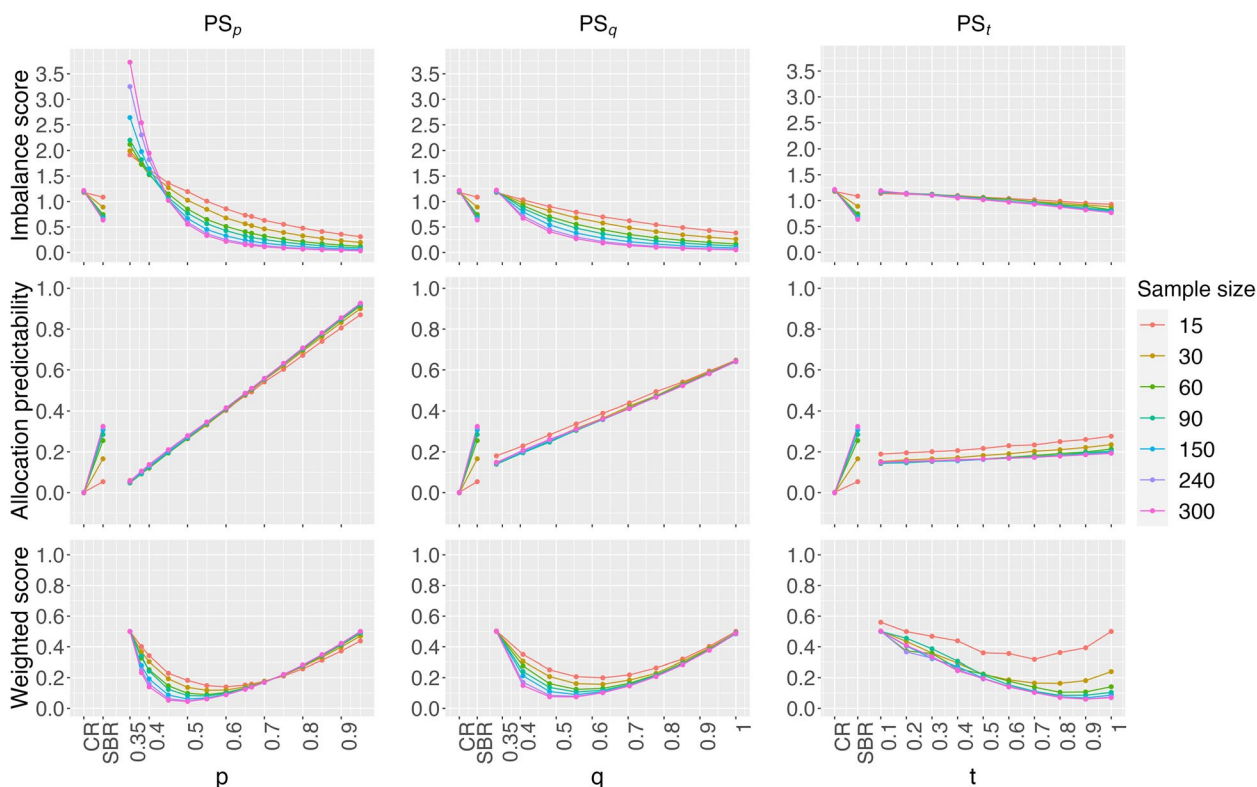


Fig. 1 Imbalance score, allocation predictability, and weighted score of the PS procedure based on three treatment assignment probability methods: PS_p , PS_q , and PS_t . These methods are compared with CR, and SBR with the block size of 6 for a study with 3 treatments and two influential factors

Optimal parameter in the PS randomization procedures

For a study with 3 treatments, let sample sizes per arm be: $n = 5, 10, 20, 30, 50, 80,$ and 100. Two influential factors are considered: factor 1 that follows a Bernoulli distribution with $p = 60\%$, and factor 2 following a multinomial distribution with three possible outcomes with probabilities of (30%, 20%, 50%).

Figure 1 shows the comparison between the three treatment assignment methods based on the range approach to calculate the covariate imbalance score. We used a block size of 6 in the SBR in this figure. For each given sample size, SBR could have a lower imbalance score than CR while SBR has larger AP values than CR. In the PS_p method, the parameter p is from 0.35 to 0.95 (left side). Imbalance score decreases as p goes up for each given sample size. As p gets large, the probability of a correct guess on treatment assignment becomes high. For that reason, allocation predictability is an increasing function of p . For weighted score Ω considering both rescaled imbalance and allocation predictability, the optimal p is found to be between 0.4 and 0.6 to have a good balance between imbalance and allocation predictability. When sample size is small (e.g., 15

per arm), the optimal p is near 0.6. That optimal value is reduced to 0.5 when $n > 100$ per arm. In Fig. 1, we observed similar results when PS_q was used to define the probability of treatment assignment. The optimal q value is close to 0.60 when $n < 100$, and that value is reduced to 0.5 when $n > 100$ as observed. For the PS_t method, the weighted score is larger than other methods when sample size is very small (e.g., 15 in a study). In other cases, the PS_t method has good performance as the computed IS and AP values are almost independent of the choosing t values. In Fig. 1, we found that t near 0.8 provide a good balance between IS and AP for the PS_t method.

The findings using other approaches (e.g., the var approach, and the SD approach) for the covariate imbalance score calculation, are similar to those observed in Fig. 1. In general, the SD approach has lower Ω values than the var approach. The SD approach considers the sample sizes in the imbalance score calculation. Based on that finding, we focus our comparisons on the range approach and the SD approach in computing the covariate imbalance score (S_k) in the PS methods to determine the treatment assignment during the study. Once the

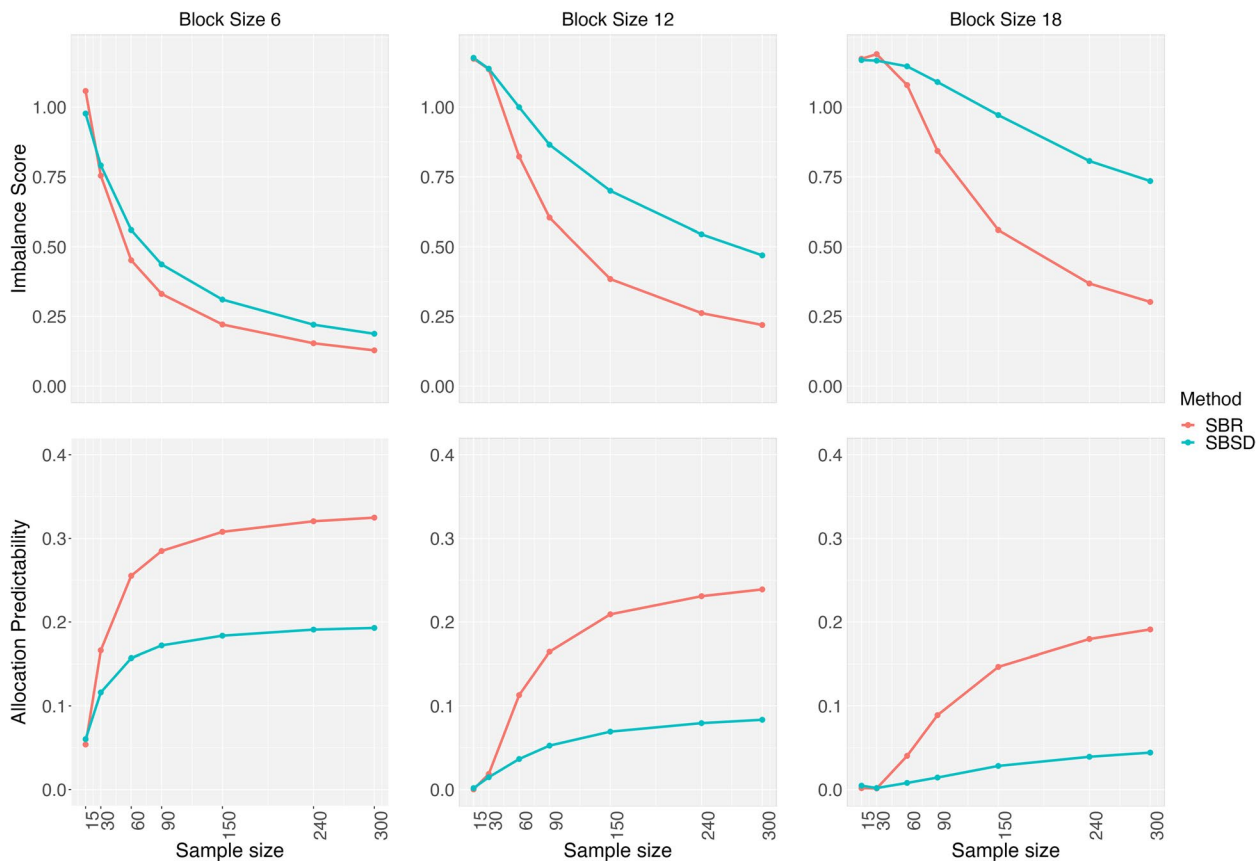


Fig. 2 Imbalance score and allocation predictability of the SBR and SBSD methods with the block sizes of 6, 12, and 18. SBR: stratified block randomization; SBSD: stratified big stick design

study is completed, the treatment imbalance score IS is computed to compare different methods regarding treatment imbalance.

In Fig. 1, we study the SBR method with the block size of 6. We further compare the SBR method with the block sizes of 6, 12, and 18, and include the SBSD method in Fig. 2 for comparison. The sample size is the same as that in Fig. 1. For the two methods, the IS value is a decreasing function of sample size, while that trend is reversed for the AP value. The SBR design has lower IS values than the SBSD method, and their difference gets larger as the block size goes up from 6 to 18. However, the SBSD method can improve the randomness of a trial by reducing a constant AP value for each given sample size. For a study with the sample size of 300, their difference in the AP value is close to 0.14 from these three plots.

We also studied the effect of the number of study sites on weighted score in Fig. 3, with the first factor as the number of study sites, and the second factor follows a multinomial distribution with three possible outcomes with probabilities of (30%, 20%, 50%). In general, imbalance score is a decreasing function of p or q in the PS_p or PS_q method, while that trend is reversed for allocation predictability. A study with a large number of study sites (e.g., 20 study sites) often has worse weighted score Ω than studies with less number of sites.

The weighted score decreases more when the number of study sites is increased from 2 to 5, than the cases when the number of study sites is increased from 5 to 10 or from 10 to 20.

Additional factors in CAR procedures

Another research question we tried to address in this article is to investigate the IS and AP performance when additional prognostic factors are added to CAR procedures. In Fig. 4, we presented the IS and AP values as a function of the total sample based on the range method (left) and the SD method (right) in calculating the covariate imbalance score S_k for a study with 3 treatments: the $PS_{p=0.5}$ method (rows 1 and 2), the $PS_{q=0.5}$ method (rows 3 and 4), and the $PS_{t=0.8}$ method (rows 5 and 6). Here, $PS_{p=0.5}$ has the probabilities as $(p_1, p_2, p_3) = (50\%, 25\%, 25\%)$ while $(p_1, p_2, p_3) = (42\%, 33\%, 25\%)$ in $PS_{q=0.5}$. In the $PS_{t=0.8}$ method, the probability (p_1, p_2, p_3) is not equal for each new patient. It can be seen that as compared to $PS_{q=0.63}$, $PS_{p=0.6}$ has a higher probability to assign a new patient to the treatment having the lowest imbalance score. In addition to the two aforementioned influential factors (one binary, one multinomial with three levels), we added another four factors having 2 or 3 levels, with a total of 6 factors.

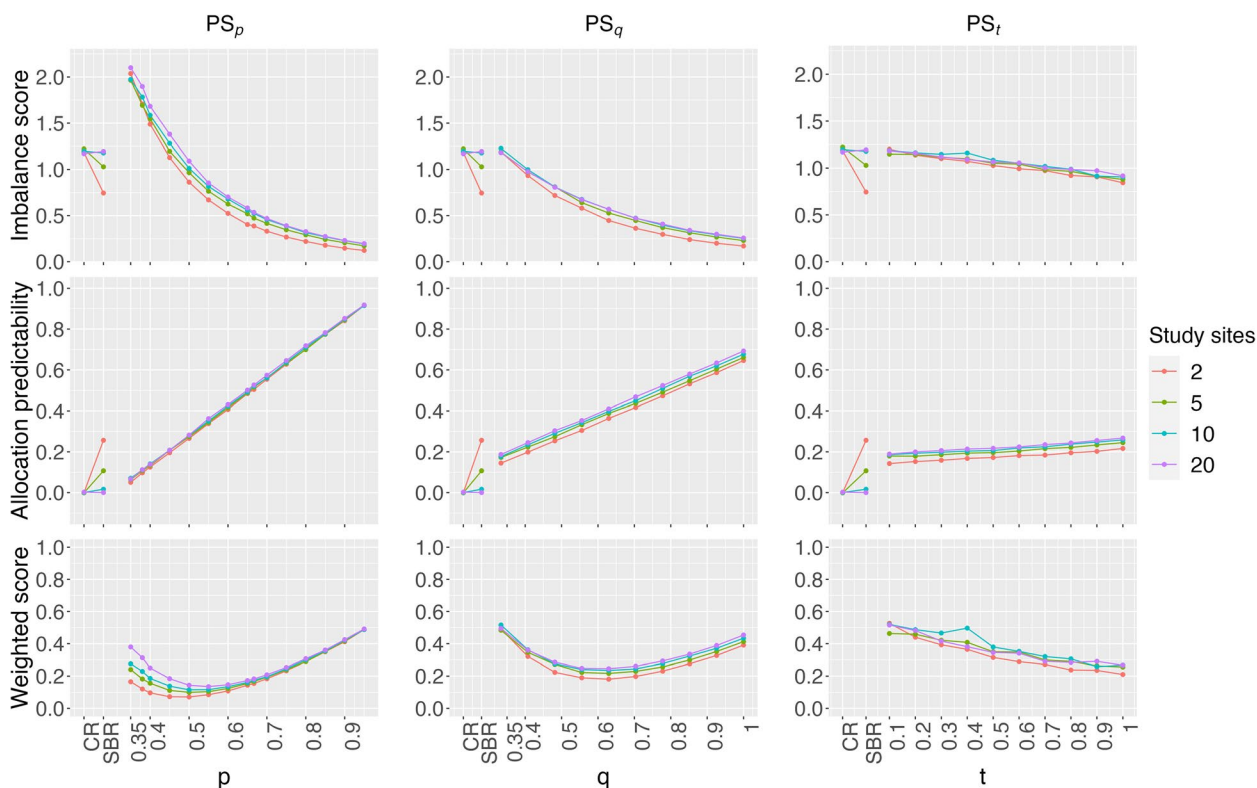


Fig. 3 For a 3-arm study with the total sample size of 60, imbalance score, allocation predictability, and weighted score are plotted as a function of the parameters in the three PS methods, when the number of study sites is increased from 2 to 20

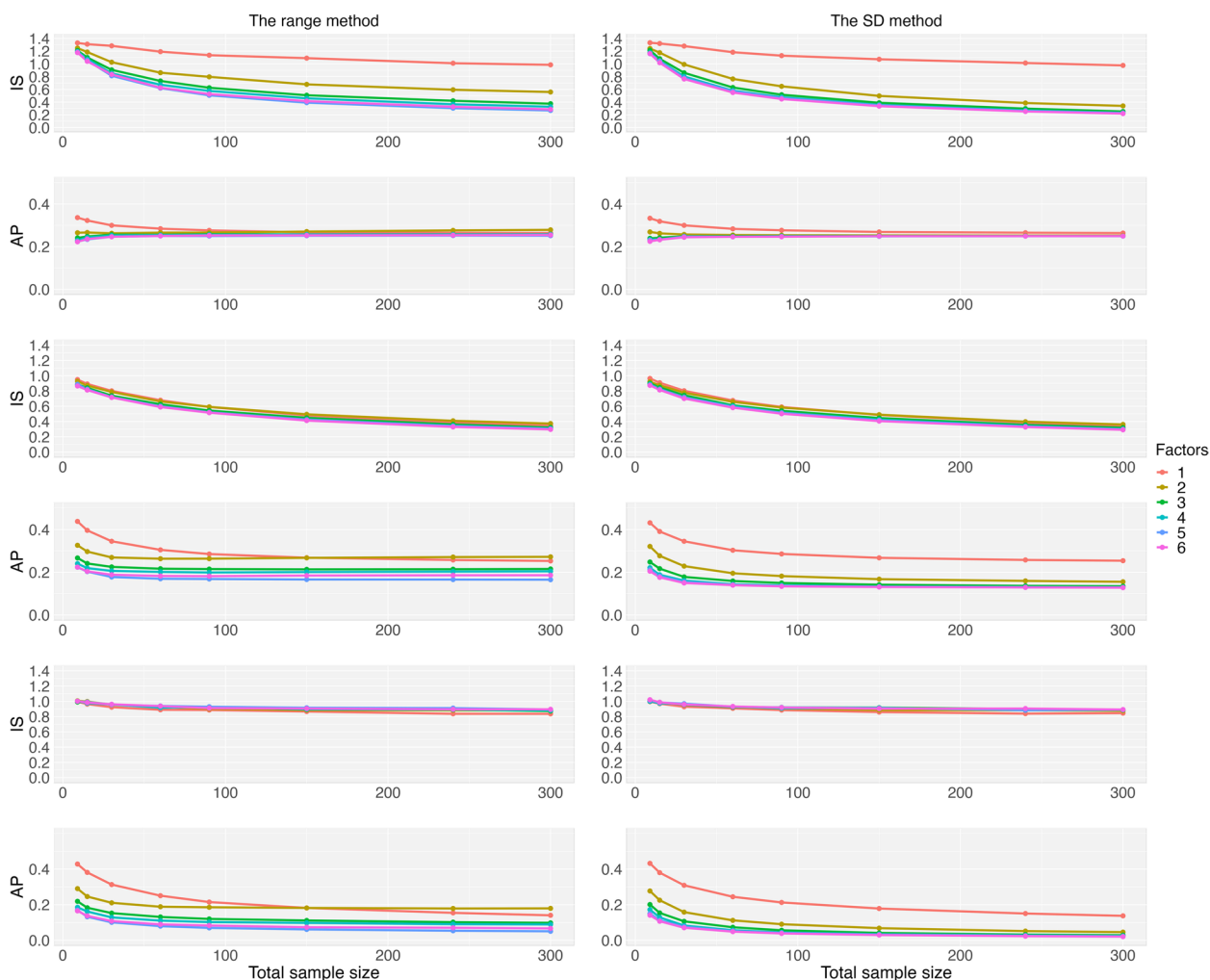


Fig. 4 Imbalance score and allocation predictability of the three PS methods: the $PS_p=0.5$ method (rows 1 and 2), the $PS_q=0.5$ method (rows 3 and 4), and the $PS_t=0.8$ method (rows 5 and 6), as a function of the total sample size and the number factors in the CAR for a study with 3 treatments. The range approach (left) and the SD approach (right) are used in the covariate imbalance score calculation

In Fig. 4, we found that the IS values (rows 1, 3, and 5) decrease as sample size goes up. The curves are relatively flat for the PS_t method as compared to the other two equal allocation methods. When sample size is not too small with more than 1 factor, the two equal allocation methods have smaller IS values than the PS_t method. However, the PS_t method has lower AP values than the other two methods. When sample size is 300, the average of the IS values are 0.56 and 0.52 for the range method and the SD method, respectively. We also found that as compared to the range method, the SD method can reduce the average of AP values by 20% from these data. The SD method generally has better performance than the range method.

We further compared the PS_t method and the HDBR design in Fig. 5 with regards to IS and AP as the number of factors increases when the range method was used

in the covariate imbalance score calculation. The HDBR design by Heritier et al. [36] aimed to maintain the marginal balance over important factors. It is expected that the HDBR design can reduce the IS values as compared to the PS method. The randomness index AP increases as sample sizes go up for the HDBR design. When sample size is 100 or above, the AP values of the HDBR design are larger than those for the PS_t method. If a low IS value is more important than the AP value, the HDBR design is a great randomization method to be utilized. Otherwise, the PS_t method could be utilized to have low AP values.

Statistical power comparison

We compared the statistical power between CR and the CAR designs with $PS_p=0.6$ with the sample size of 200 per arm in a two-arm randomized trial with the expected effect size of $0.4/\sqrt{2} = 0.28$. The sample size

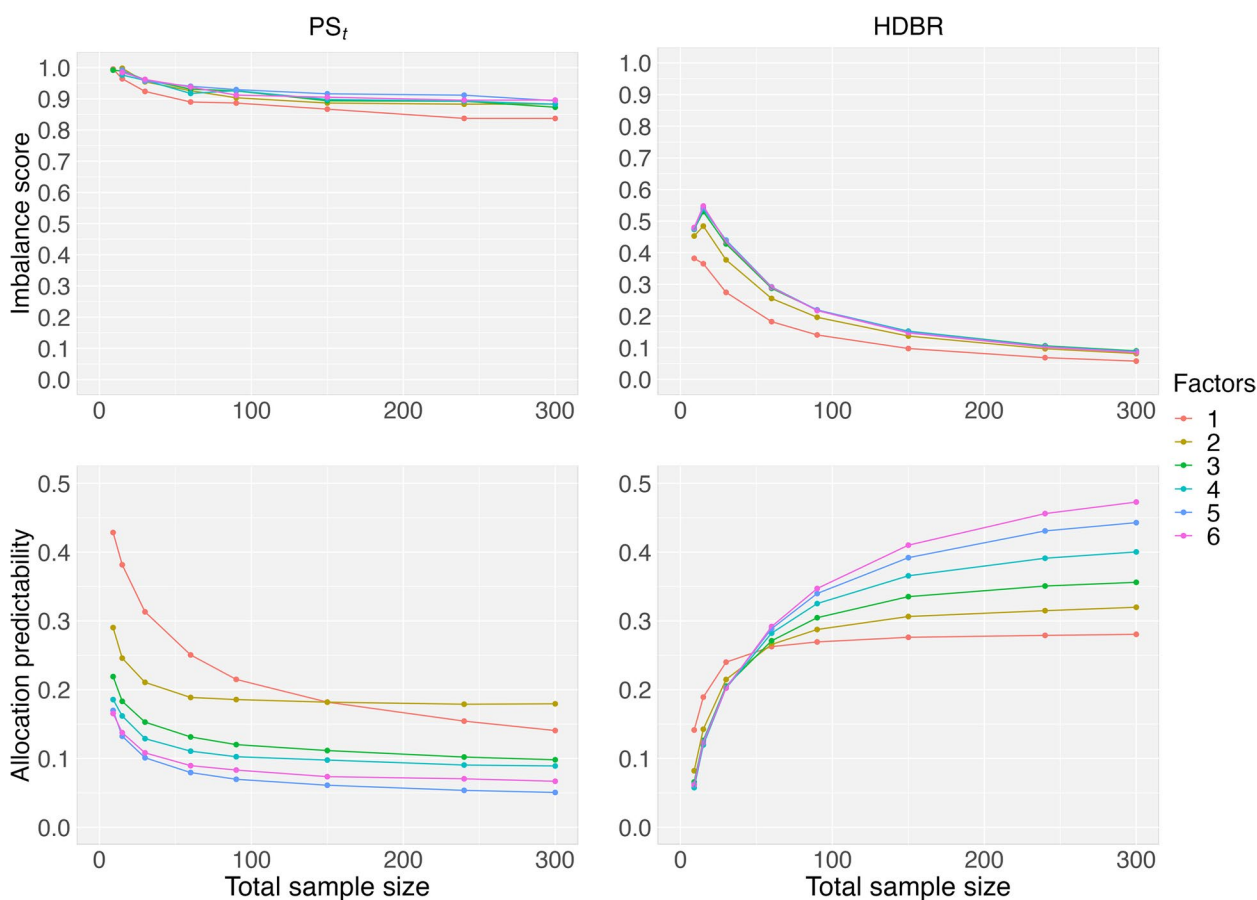


Fig. 5 Imbalance score and allocation predictability of the PS_t method and the HDBR method as a function of the total sample size and the number factors in the CAR for a study with 3 treatments. HDBR: hierarchical dynamic balancing randomization

was determined to detect the effect size of 0.28 to achieve the statistical power of 80% when $\alpha = 0.05$. Suppose the first two factors are: binomial with $p=50\%$, and multinomial with 3 outcomes having the equal probability. The correlation between the binomial factor and the outcome is $\rho_1 = 0.3$.

In Fig. 6, the statistical power was presented as a function of ρ_3 which is the correlation between outcome and the third factor, with ρ_4 from 0.1 to 0.7. The correlation between the multinomial factor and the outcome was assumed to be $\rho_2 = 0.2$ (the first row) and 0.1 (the second row). We calculated the statistical power by using the $PS_{p=0.6}$ with 2, 3, and 4 factors as covariates. The simulated power is very close to the nominal level of 80% for the CR. When correlation is low (e.g., $\rho_3 = \rho_4 = 10\%$), the statistical power of CAR designs is similar to that of CR. As correlation goes up (ρ_4 from 0.1 to 0.7), the statistical power could be increased by more than 10% with the simulated statistical power being above 90%.

The statistical power gain using the CAR design with two factors can consistently improve more than 2% as

compared to the CR design. The statistical power of the CAR designs including 3 factors or more is an increasing function of ρ_3 . When ρ_3 goes up to 0.7, the statistical power could be above 90% at the nominal level of 80%. When ρ_4 is small (e.g., $\rho_4 = 0.1$), the statistical power gain is very limited by adding the fourth factor to the model with 3 factors. However, when ρ_4 is medium to large (e.g., 0.4, 0.7), it may further increase the statistical power by including that additional factor in the CAR design. We observe similar findings in comparing statistical power based on $PS_{p=0.6}$ with 2, 3, and 4 factors, as a function of ρ_3 .

We further compare the statistical power as p in the PS_p method and t in the PS_t method go up in Fig. 7. The statistical power was plotted as a function of ρ_3 given $\rho_1 = 0.3$, $\rho_2 = 0.2$, and $\rho_4 = 0.4$. Five different p values (0.55 to 0.8) and 5 t values (0.6 to 0.95) were studied. When two factors are used in CAR designs (first row), the statistical power is not sensitive to the value of ρ_3 . When we have three or four factors (the middle row, and the bottom row), the statistical power is an increasing

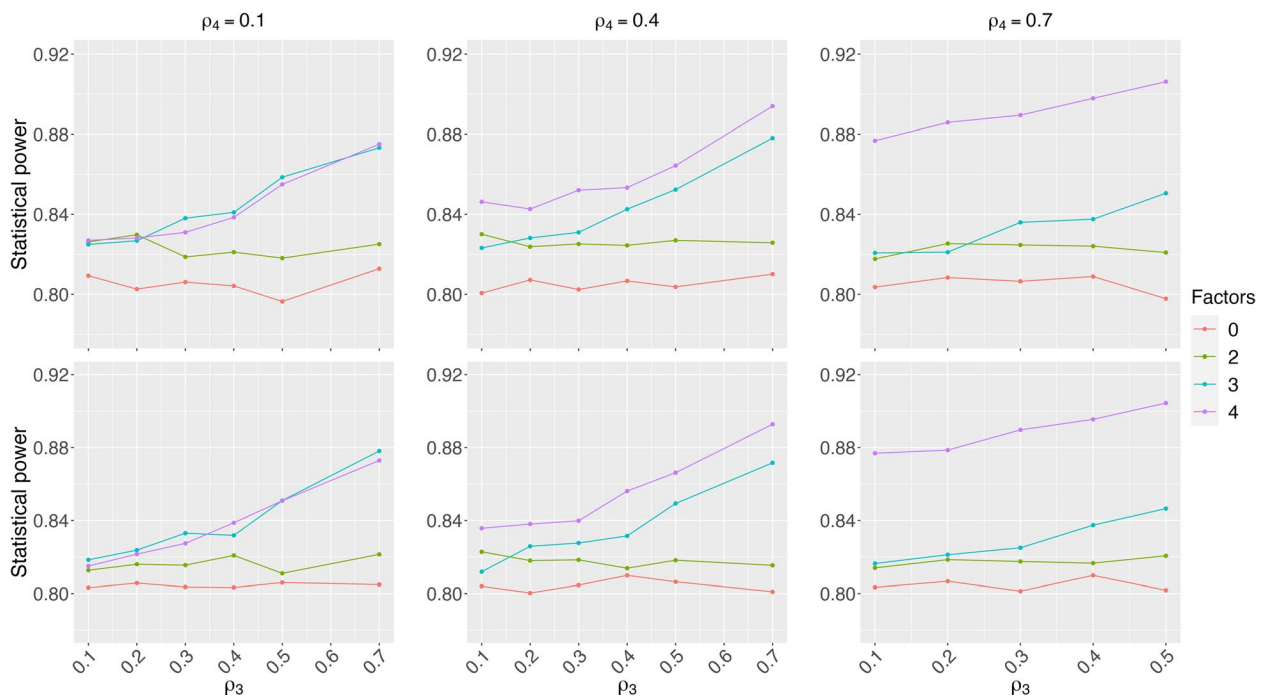


Fig. 6 Statistical power of a two-arm CAR design as a function of ρ_3 , given ρ_4 from 0.1 to 0.7. In the CAR design, the treatment assignment probability was set as 60%. The first and the second row has the correlation $\rho_2 = 0.2$ and 0.1. When the number of factor is zero, it is a CR design

function of ρ_3 . When $\rho_3 = 0.4$, the statistical power may be increased by 4% when two additional factors are added to a model having two existing factors. For the PS_p method, the p values between 0.6 to 0.7 have good performance with regards to the statistical power in many configurations. For the PS_t method, a high value of t (e.g., 0.8 or above) is associated with a large statistical power. Given the ρ_3 value and the number of factors, the statistical power difference between these considered methods is small which is around 2%.

An example

We used data from the Second Multi-centre Intra-pleural Sepsis Trial (MIST2) to illustrate the application of the CAR designs [42, 43]. The MIST2 trial was a four-arm randomized trial to investigate the efficacy of intra-pleural tissue plasminogen activator (t-PA) and DNase for patients with pleural infection. The primary outcome was the change in pleural opacity at day 7 from day 1. We used the estimated correlation between the primary outcome and three categorical factors: hospital-acquired infection ($\rho_1 = 0.12$), large tube size ($\rho_2 = 0.16$), and drain present ($\rho_3 = 0.27$). These correlation coefficients were presented in the article by Kahan et al. [43].

Suppose we are going to compare a new treatment with the gold standard: a randomized two-arm study. To

detect a medium effect size of 0.39 in an early phase trial, the sample size per group is estimated as 105 (a total of 210 patients) based on the two-sample t-test. Given that sample size, the simulated statistical power from 20,000 simulations is 79.95% in the statistical model without controlling for other covariates. For the CAR designs, the statistical power is between 81.2% and 82.2% when these three factors are considered in the randomization and adjusted in the primary analyses. The highest power is obtained when the $PS_{q=0.63}$ approach in conjunction with the SD method in calculating the imbalance score is used in the CAR design.

Discussion

From simulation studies, we found that the increase in statistical power depends on the correlation between covariates and the outcome. Specifically, adjusting for covariates that are strongly correlated with the outcome leads to a greater reduction in the standard errors for the treatment effect, and therefore a larger increase in the statistical power [44]. When adjusting for key confounders in the context of binary or survival outcomes though, it is expected that the standard errors would increase. However, this can be counterbalanced by an increase in the estimated treatment effect, which ultimately contributes to enhanced power [45–48]. It should be noted that in the current study, our focus was exclusively on

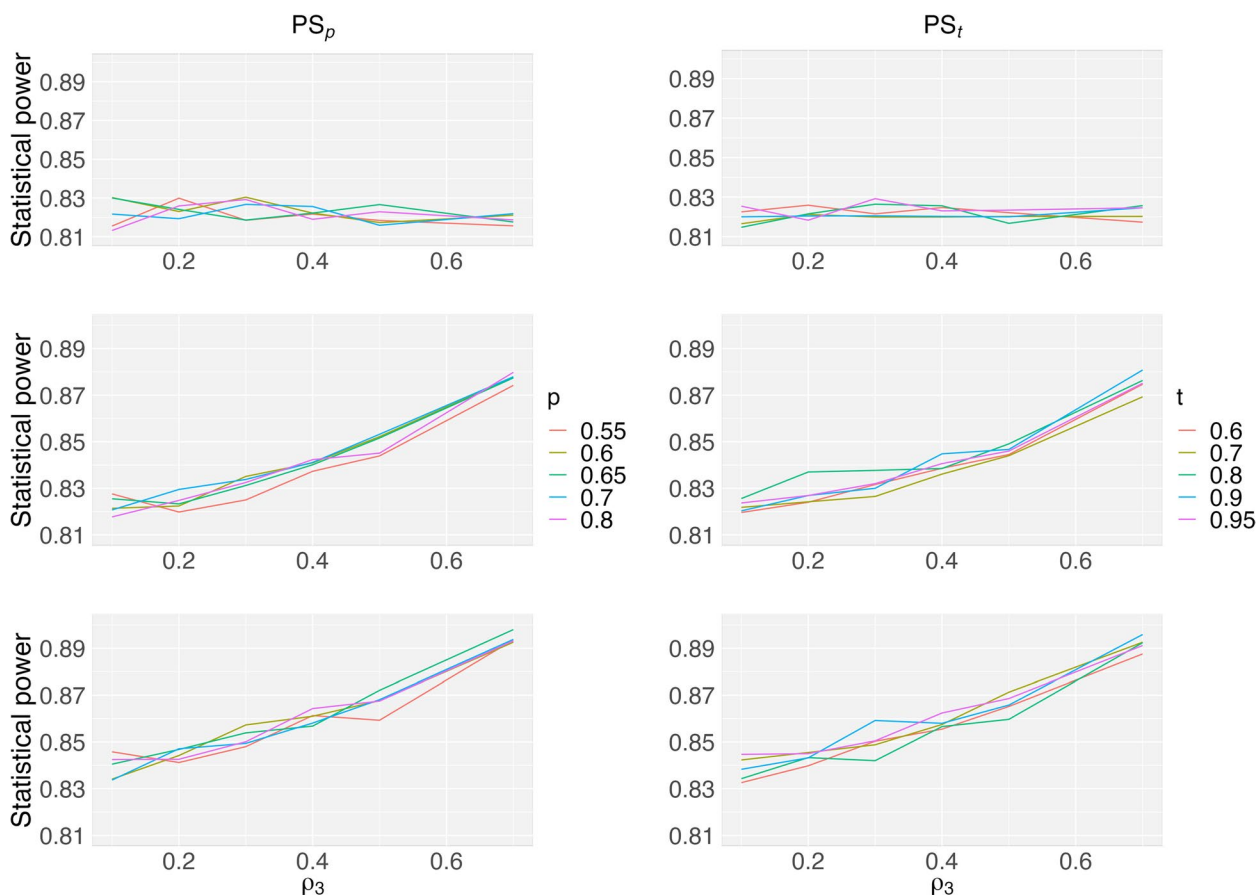


Fig. 7 Statistical power of a two-arm CAR design as a function of ρ_3 , given $\rho_1 = 0.3, \rho_2 = 0.2$, and $\rho_4 = 0.4$. Five different p values (0.55 to 0.8) in the PS_p method and five t values (0.6 to 0.95) in the PS_t method were studied, for two, three, and four factors (row 1, row 2, and row 3)

continuous outcomes. However, binary and survival outcomes represent areas of significant interest that have not been explored in this work. These outcome types could serve as a focus for future research endeavors.

Additionally, when we compared the statistical power between CAR designs and CR designs, there were some methodological disparities that merited attention as we implemented different strategies for covariate control. Specifically, we did not adjust for any covariates when analyzing the statistical power of CR designs, whereas for the CAR designs, we did control for a varying number of prognostic factors. This discrepancy introduces a layer of complication to the interpretation of the results, as it becomes challenging to determine the exact effect of the CAR methods on the observed increase in statistical power. It raises the question of whether the power increase is genuinely attributable to the effectiveness of CAR designs, or if it is merely the result of the inclusion of additional covariates in the statistical model. Therefore, discerning the specific contribution of CAR

methods to the improvement of statistical power remains an intricate issue. Future research could compare the statistical power by conducting identical data analyses while varying only the randomization methods employed. This would provide a clearer understanding of the influence of different randomization techniques on the resulting statistical power.

Acknowledgements

We would like to thank the comments from four reviewers and associate editor, that help us to improve the manuscript.

Authors' contributions

The idea of this article was originally developed by GS and SW. GS, YL, YZ, and XL conducted simulation studies. All authors drafted the manuscript and approved the final version.

Funding

Shan's research is partially supported by grants from the National Institutes of Health: R03CA248006, R03AG083207, and R01AG070849.

Availability of data and materials

Not applicable. This is a manuscript to develop novel statistical approaches, therefore, no real data is involved.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 September 2023 Accepted: 16 January 2024

Published online: 25 January 2024

References

1. Spieth PM, Kubasch AS, Penzlin AI, Illigens BMW, Barlinn K, Siepmann T. Randomized controlled trials - a matter of design. *Neuropsychiatr Dis Treat*. 2016;12:1341–9. <https://doi.org/10.2147/NDT.S101938>.
2. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research. *BJOG Int J Obstet Gynecol*. 2018;125(13):1716. <https://doi.org/10.1111/1471-0528.15199>.
3. Lin Y, Zhu M, Su Z. The pursuit of balance: An overview of covariate-adaptive randomization techniques in clinical trials. *Contemp Clin Trials*. 2015;45:21–5. <https://doi.org/10.1016/J.CCT.2015.07.011>.
4. Therneau TM. How many stratification factors are “too many” to use in a randomization plan? *Control Clin Trials*. 1993;14(2):98–108. [https://doi.org/10.1016/0197-2456\(93\)90013-4](https://doi.org/10.1016/0197-2456(93)90013-4).
5. Azher RA, Wason JMS, Grayling MJ. A Comparison of Randomization Methods for Multi-Arm Clinical Trials. *Stat Biopharm Res*. 2023;1–13. <https://doi.org/10.1080/19466315.2023.2238645>.
6. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet*. 2002;359(9306):614–8. [https://doi.org/10.1016/S0140-6736\(02\)07750-4](https://doi.org/10.1016/S0140-6736(02)07750-4).
7. Grayling MJ, Dimairo M, Mander AP, Jaki TF. A Review of Perspectives on the Use of Randomization in Phase II Oncology Trials. *J Natl Cancer Inst*. 2019;111(12):1255–62. <https://doi.org/10.1093/JNCI/DJZ126>.
8. Ryznyk Y, Sverdlov O. A comparative study of restricted randomization procedures for multiarm trials with equal or unequal treatment allocation ratios. *Stat Med*. 2018;37(21):3056–77. <https://doi.org/10.1002/sim.7817>.
9. Lim CY, In J. Randomization in clinical studies. *Korean J Anesthesiol*. 2019;72(3):221–32. <https://doi.org/10.4097/kja.19049>.
10. Kang M, Ragan BG, Park JH. Issues in outcomes research: An overview of randomization techniques for clinical trials. 2008. <https://doi.org/10.4085/1062-6050-43.2.215>.
11. Atkinson AC. The comparison of designs for sequential clinical trials with covariate information. *J R Stat Soc: Ser A (Stat Soc)*. 2002;165(2):349–73. <https://doi.org/10.1111/1467-985X.00564>.
12. Hill AB. The clinical trial. *Br Med Bull*. 1951;7(4):278–82. <https://doi.org/10.1093/oxfordjournals.bmb.a073919>.
13. Efron B. Forcing a sequential experiment to be balanced. *Biometrika*. 1971;58(3):403–17. <https://doi.org/10.1093/biomet/58.3.403>.
14. Wei LJ. A class of designs for sequential clinical trials. *J Am Stat Assoc*. 1977;72(358):382–6. <https://doi.org/10.1080/01621459.1977.10481005>.
15. Soares JF, Wu CFJ. Some Restricted randomization rules in sequential designs. *Commun Stat Theory Methods*. 1983;12(17):2017–34. <https://doi.org/10.1080/03610928308828586>.
16. Chen YP. Biased coin design with imbalance tolerance. *Commun Stat Stoch Models*. 1999;15(5):953–75. <https://doi.org/10.1080/15326349908807570>.
17. Chen YP. Which design is better? Ehrenfest urn versus biased coin. *Adv Appl Probab*. 2000;32(3):738–49. <https://doi.org/10.1239/aap/1013540242>.
18. Zhao W, Weng Y. Block urn design - A new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemp Clin Trials*. 2011;32(6):953–61. <https://doi.org/10.1016/j.cct.2011.08.004>.
19. Smith RL. Sequential Treatment Allocation Using Biased Coin Designs. *J R Stat Soc Ser B (Methodological)*. 1984;46(3):519–43. <https://doi.org/10.1111/j.2517-6161.1984.tb01323.x>.
20. Freedman LS, White SJ. On the Use of Pocock and Simon's Method for Balancing Treatment Numbers over Prognostic Factors in the Controlled Clinical Trial. *Biometrics*. 1976;32(3):691. <https://doi.org/10.2307/2529759>.
21. Mukherjee A, Grayling MJ, Wason JMS. Adaptive Designs: Benefits and Cautions for Neurosurgery Trials. *World Neurosurg*. 2022;161:316–22. <https://doi.org/10.1016/j.wneu.2021.07.061>.
22. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: A review. *Control Clin Trials*. 2002;23(6):662–74. [https://doi.org/10.1016/S0197-2456\(02\)00242-8](https://doi.org/10.1016/S0197-2456(02)00242-8).
23. Taves DR. Minimization: A new method of assigning patients to treatment and control groups. *Clin Pharmacol Ther*. 1974;15(5):443–53. <https://doi.org/10.1002/cpt.1974155443>.
24. Lin Y, Su Z. Balancing continuous and categorical baseline covariates in sequential clinical trials using the area between empirical cumulative distribution functions. *Stat Med*. 2012;31(18):1961–71. <https://doi.org/10.1002/sim.5363>.
25. Zhao W. A better alternative to stratified permuted block design for subject randomization in clinical trials. *Stat Med*. 2014;33(30):5239–48. <https://doi.org/10.1002/sim.6266>.
26. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975;31(1):103–15.
27. Proschan M, Brittain E, Kammerman L. Minimize the Use of Minimization with Unequal Allocation. *Biometrics*. 2011;67(3):1135–41. <https://doi.org/10.1111/j.1541-0420.2010.01545.x>.
28. Han B, Enas NH, McEntegart D. Randomization by minimization for unbalanced treatment allocation. *Stat Med*. 2009;28(27):3329–46. <https://doi.org/10.1002/sim.3710>.
29. Kuznetsova OM, Tymofyeyev Y. Preserving the allocation ratio at every allocation with biased coin randomization and minimization in studies with unequal allocation. *Stat Med*. 2012;31(8):701–23. <https://doi.org/10.1002/sim.4447>.
30. Taves DR. The use of minimization in clinical trials. *Contemp Clin Trials*. 2010;31(2):180–4. <https://doi.org/10.1016/j.cct.2009.12.005>.
31. Lauzon SD, Zhao W, Nietert PJ, Ciolino JD, Hill MD, Ramakrishnan V. Impact of minimal sufficient balance, minimization, and stratified permuted blocks on bias and power in the estimation of treatment effect in sequential clinical trials with a binary endpoint. *Stat Methods Med Res*. 2022;31(1):184–204.
32. Kuznetsova OM, Tymofyeyev Y. Hierarchical Dynamic Allocation Procedures Based on Modified Zelen's Approach in Multiregional Studies With Unequal Allocation. *J Biopharm Stat*. 2014;24(4):785–801.
33. Shan G, Banks S, Miller JB, Ritter A, Bernick C, Lombardo J, et al. Statistical advances in clinical trials and clinical research. *Alzheimers Dement Transl Res Clin Interv*. 2018;4:366–71.
34. Signorini DF, Leung O, Simes RJ, Beller E, Gebiski VJ, Callaghan T. Dynamic balanced randomization for clinical trials. *Stat Med*. 1993;12(24):2343–50. <https://doi.org/10.1002/sim.4780122410>.
35. Maciel CB, Barlow B, Lucke-Wold B, Gobinathan A, Abu-Mowis Z, Peethala MM, et al. Acute Headache Management for Patients with Subarachnoid Hemorrhage: An International Survey of Health Care Providers. *Neurocrit Care*. 2022. <https://doi.org/10.1007/s12028-022-01571-7>.
36. Heritier S, Gebiski V, Pillai A. Dynamic balancing randomization in controlled clinical trials. *Stat Med*. 2005;24(24):3729–41. <https://doi.org/10.1002/sim.2421>.
37. Jin M, Polis A, Hartzel J. Algorithms for minimization randomization and the implementation with an R package. *Commun Stat Simul Comput*. 2021;50(10):3077–87. <https://doi.org/10.1080/03610918.2019.1619765>.
38. Atkinson AC, Fedorov VV. *Optimal Design of Experiments*. Wiley StatsRef: Statistics Reference Online. 2014. <https://doi.org/10.1002/9781118445112.stat00911>.
39. Zagoraiou M. Choosing a covariate-adaptive randomization procedure in practice. *J Biopharm Stat*. 2017;27(5):845–57. <https://doi.org/10.1080/10543406.2017.1289944>.
40. Atkinson AC. Selecting a biased-coin design. *Stat Sci*. 2014;29(1):144–63. <https://doi.org/10.1214/13-STS449>.

41. Berger VW, Bour LJ, Carter K, Chipman JJ, Everett CC, Heussen N, et al. A roadmap to using randomization in clinical trials. *BMC Med Res Methodol.* 2021;21(1):1–24.
42. Rahman NM, Maskell NA, West A, Teoh R, Arnold A, Mackinlay C, et al. Intrapleural Use of Tissue Plasminogen Activator and DNase in Pleural Infection. *N Engl J Med.* 2011;365(6):518–26. <https://doi.org/10.1056/nejmoa1012740>.
43. Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: An assessment of 12 outcomes from 8 studies. *Trials.* 2014;15(1):1–7. <https://doi.org/10.1186/1745-6215-15-139/FIGURES/2>.
44. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med.* 2002;21(19):2917–30. <https://doi.org/10.1002/sim.1296>.
45. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Control Clin Trials.* 1998;19(3):249–56. [https://doi.org/10.1016/s0197-2456\(97\)00147-5](https://doi.org/10.1016/s0197-2456(97)00147-5).
46. Grayling MJ, Wheeler GM. A review of available software for adaptive clinical trial design. SAGE Publications Ltd; 2020. <https://doi.org/10.1177/1740774520906398>.
47. Robinson LD, Jewell NP. Some Surprising Results about Covariate Adjustment in Logistic Regression Models. *Int Stat Rev / Revue Internationale de Statistique.* 1991;59(2):227–40.
48. Hernández AV, Eijkemans MJC, Steyerberg EW. Randomized controlled trials with time-to-event outcomes: how much does prespecified covariate adjustment increase power? *Ann Epidemiol.* 2006;16(1):41–8. <https://doi.org/10.1016/j.annepidem.2005.09.007>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.