

# UCLA

## UCLA Previously Published Works

### Title

Mathematical programming tools for randomization purposes in small two-arm clinical trials: A case study with real data.

### Permalink

<https://escholarship.org/uc/item/31r755r9>

### Journal

Pharmaceutical Statistics, 23(6)

### Authors

Vazquez, Alan

Wong, Weng Kee

### Publication Date

2024

### DOI

10.1002/pst.2388

Peer reviewed

**MAIN PAPER**

# Mathematical programming tools for randomization purposes in small two-arm clinical trials: A case study with real data

Alan R. Vazquez<sup>1</sup>  | Weng-Kee Wong<sup>2</sup> 

<sup>1</sup>School of Engineering and Sciences, Tecnológico de Monterrey, Monterrey, Nuevo Leon, Mexico

<sup>2</sup>Department of Biostatistics, University of California, Los Angeles, California, USA

**Correspondence**

Weng-Kee Wong, Department of Biostatistics, University of California, Los Angeles, CA, USA.

Email: [wk Wong@ucla.edu](mailto:wk Wong@ucla.edu)

**Abstract**

Modern randomization methods in clinical trials are invariably adaptive, meaning that the assignment of the next subject to a treatment group uses the accumulated information in the trial. Some of the recent adaptive randomization methods use mathematical programming to construct attractive clinical trials that balance the group features, such as their sizes and covariate distributions of their subjects. We review some of these methods and compare their performance with common covariate-adaptive randomization methods for small clinical trials. We introduce an energy distance measure that compares the discrepancy between the two groups using the joint distribution of the subjects' covariates. This metric is more appealing than evaluating the discrepancy between the groups using their marginal covariate distributions. Using numerical experiments, we demonstrate the advantages of the mathematical programming methods under the new measure. In the supplementary material, we provide R codes to reproduce our study results and facilitate comparisons of different randomization procedures.

**KEYWORDS**

covariate-adaptive trial, energy distance, minimization method, prior information

## 1 | INTRODUCTION

The statistics literature is replete with randomization methods for clinical trials. Their main goal is to ensure that subjects are as comparable as possible between the treatment groups under study, so that inferences about the treatment effects are not biased by differences between the groups. In recent decades, adaptive randomization methods<sup>1</sup> have been increasingly used to achieve greater statistical efficiency and reduce costs for conducting a clinical trial. These methods assign subjects to treatment groups as they enroll in the trial using the accumulated data at hand.

Adaptive randomization methods are classified into three groups: covariate-adaptive, response-adaptive, and covariate-adjusted response-adaptive (CARA) methods.<sup>1</sup> Covariate-adaptive methods balance the distributions of the covariates (prognostic factors) of the subjects across the treatment groups. Response-adaptive methods use the responses from subjects who have received treatment to assign future subjects to treatment groups that appear to be

---

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *Pharmaceutical Statistics* published by John Wiley & Sons Ltd.

increasingly effective. CARA methods combine the ideas of covariate- and response-adaptive methods to generate the treatment groups of a clinical trial. We refer to Rosenberger and Sverdlov<sup>2</sup> and Rosenberger et al.<sup>3</sup> for complete reviews of these methods.

In this article, we study covariate-adaptive randomization methods to balance continuous covariates in small clinical trials with two treatment groups (i.e., two-arm clinical trials). These trials arise in practice due to budget limitations or because the disease under study is rare. They also arise due to ethical and physiological considerations that may be involved in specific cohorts of subjects, such as infants<sup>4</sup> or minority groups.<sup>5</sup> Evans and Ildstad,<sup>6</sup> Suresh,<sup>7</sup> and Hoare et al.<sup>8</sup> discussed challenging issues of the design and analysis of small clinical trials. Our takeaway is that these trials are practically relevant but their challenges are much less addressed in the literature.

A novelty of our study is that we use mathematical programming (MP; Bradley et al.<sup>9</sup>) to design clinical trials. MP is a systematic approach to optimizing an objective function in terms of decision variables that must satisfy a set of constraints. For this reason, it has been used frequently in engineering, operations research, and recently in the construction of various optimal designs for linear and nonlinear models<sup>10–15</sup> and fractional factorial designs,<sup>16,17</sup> as well as for analyzing data from screening experiments.<sup>18,19</sup>

In the MP literature, there are three adaptive randomization methods introduced by Williamson et al.,<sup>20</sup> Bertsimas et al.,<sup>21</sup> and Bhat et al.<sup>22</sup> The method of Williamson et al.<sup>20</sup> is response-adaptive and assigns subjects to treatment groups using a pre-computed assignment sequence, which maximizes the expected number of correct assignments throughout the clinical trial. The method of Bertsimas et al.<sup>21</sup> is covariate-adaptive and balances continuous covariates by matching the means and variances of their marginal distributions between groups. The method of Bhat et al.<sup>22</sup> is also covariate-adaptive and balances continuous covariates by minimizing the prediction variance of a linear regression model containing the treatment and covariate effects. An attractive feature of these methods is that they use the cumulative information on the covariate values of the subjects already in the trial to randomize the next patient.

In this article, we compare the MP method of Bertsimas et al.<sup>21</sup> with covariate-adaptive randomization methods that fall within the minimization framework (Section 2). We chose the method of Bertsimas et al.<sup>21</sup> over the method of Bhat et al.<sup>22</sup> because the former is model-free. In contrast, the method of Bhat et al.<sup>22</sup> assumes a linear model a priori, which may be incorrect, and consequently result in a suboptimal allocation scheme. To our knowledge, our comparison between minimization methods and Bertsimas et al.<sup>21</sup> method is new.

We compare the methods using two real two-arm clinical trials with up to 20 subjects and two or three continuous covariates. The performance of covariate-adaptive methods for such small trials is not addressed in the literature and so, our study fills that gap. Additionally, we propose a new measure of imbalance between two groups called the energy distance,<sup>23</sup> which compares the joint distribution of the covariates of the subjects. Our measure is more general than the current ones that aim to balance the marginal distributions of each covariate across groups. We conclude that MP methods are more attractive than standard covariate-adaptive methods for balancing the covariate distributions and group sizes in the small two-arm clinical trials we considered. We also show that all methods fail to balance at least one covariate in each trial, calling for method improvements for these cases. To facilitate the evaluation of the methods, we provide R codes to produce the results in the paper and may also be used to evaluate other methods.

The remainder of the article is organized as follows. In Section 2, we review representatives of covariate-adaptive methods to balance continuous covariate distributions in two-arm clinical trials. In Section 3, we present the method of Bertsimas et al.<sup>21</sup> and, in Section 4, we compare it with those representatives using two real clinical trials. In Section 5, we conclude with remarks and future work directions to further investigate the utility of MP as a competitive way of randomization in clinical trials.

## 2 | MINIMIZATION METHODS

Covariate-adaptive randomization methods are categorized into three groups: Stratified-block randomization, dynamic hierarchical randomization, and minimization.<sup>24</sup> Stratified-block randomization methods<sup>25</sup> are two-stage procedures in which subjects who enter the clinical trial are first assigned to a stratum using their covariates' values. Within each stratum, subjects are assigned to a treatment according to a randomization procedure, such as permuted-block randomization.<sup>1</sup> Dynamic hierarchical randomization<sup>26</sup> balances the subjects' covariates following a user-specified order of importance. That is, a subject is assigned to a treatment group to balance the most important covariates first and then the least important ones. Minimization methods<sup>27</sup> sequentially assign subjects to treatments to balance all covariates

simultaneously. Specifically, they assign a subject to the treatment group that would achieve the smallest degree of imbalance across all covariates of the subjects in the trial.

We focus on minimization methods for clinical trials with two treatments and continuous covariates. These methods can be further classified into two types depending on whether a linear regression model is assumed to link the mean response to the treatments and the subjects' covariates. A method that makes such an assumption is introduced by Atkinson.<sup>28–31</sup> This method seeks the allocation of subjects to groups that minimizes the variance of the estimated treatment effect in the linear regression model. Another method of this type is Bhat et al.<sup>22</sup> discussed previously. To limit the scope of this article, we do not include these methods in our study. Instead, we concentrate on minimization methods that do not assume a linear regression model. In particular, we study the methods of Pocock and Simon,<sup>32</sup> Nishi and Takaichi,<sup>33</sup> and Ma and Hu,<sup>34</sup> which are important model-free minimization methods.

We first discuss the general minimization framework. Next, we present the discrepancy measures used by Pocock and Simon,<sup>32</sup> Nishi and Takaichi,<sup>33</sup> and Ma and Hu,<sup>34</sup> and then the allocation rule for new subjects.

## 2.1 | The general framework

Minimization methods balance the marginal distributions of the covariates in the groups by minimizing a discrepancy measure. To sequentially allocate a total of  $N$  subjects with  $p$  covariates, minimization methods undertake the steps below.

Step 1. Recruit  $n_0 < N$  initial subjects and allocate them into one or two groups.

Step 2. For each new subject to be included in the trial, do the following:

- a. Construct two potential sets of groups by hypothetically assigning the new subject to groups one and two. In one set, group one contains the subject, while in the other set, group two contains that subject. Using a discrepancy measure, compare the two sets of groups in terms of the resulting marginal distributions of the covariates.
- b. Using the value of the discrepancy measure, determine the allocation of the subject.
- c. Update the treatment groups and re-start Step 2.

Step 3. Terminate if  $N$  subjects have been in the trial.

In Step 1, the value of  $n_0$  can be as small as one for some minimization methods, such as Pocock and Simon.<sup>32</sup> In this case, we randomly assign the initial subject to a group. More elaborate minimization methods, such as Nishi and Takaichi<sup>33</sup> and Ma and Hu,<sup>34</sup> need larger values of  $n_0$  and an initial allocation to both groups. This may be achieved using random allocation or permuted-block randomization,<sup>1</sup> which ensures that each group has the same number of subjects. In this way, these methods equipoise the two groups at the beginning of the sequential allocation process in Step 2.

In small clinical trials, the value of  $n_0$  should not be too large because the initial allocation method in Step 1 will overtake the covariate-adaptive allocation in Step 2, limiting the gain in balancing the covariates. Finding the best value of  $n_0$  is a practically relevant but challenging problem that we do not address here. Instead, we follow Ma and Hu<sup>34</sup> and use permuted-block randomization with blocks of size  $n_0/2$  and  $n_0 = 8$  for all minimization methods discussed here.

## 2.2 | Discrepancy measures

A discrepancy measure allows us to determine the assignment of the  $t$ -th subject ( $n_0 < t \leq N$ ) given that the previous  $t - 1$  subjects have already been assigned to the treatment groups. In what follows, we denote the treatment group as  $k \in \{1, 2\}$  and the  $p$  subjects' covariates as  $W_j, j = 1, \dots, p$ . The sizes of groups one and two before allocating the  $t$ -th subject are  $n_1$  and  $n_2$ , respectively.

### 2.2.1 | Difference in covariate groups

The discrepancy measure of Pocock and Simon<sup>32</sup> uses the difference between the number of subjects with a specific covariate value in the two groups. To use this measure, we first transform each continuous covariate  $W_j$  into a discrete

variable, denoted by  $Z_j$ , with a user-specified number of categories. For simplicity, we assume that all transformed covariates  $Z_j$  have  $c$  categories, which can be created using the quantiles of the distribution of  $W_j$  at probabilities  $1/c, 2/c, \dots, (c - 1)/c$ .

Let  $n_{jlk}$  be the number of subjects in group  $k$  whose covariate  $Z_j$  is at level  $l \in \{1, \dots, c\}$ , before allocating the  $t$ -th subject. We denote the categories of the covariates of the  $t$ -th subject as  $r_1, r_2, \dots, r_p$ , and let  $n_{jr,k}$  be the number of subjects in group  $k$  with  $Z_j$  in category  $r_j$ , before allocating the subject. Pocock and Simon<sup>32</sup> calculate the individual discrepancy between the two groups for the  $j$ -th covariate as

$$\Delta d_j^{PS} = |(n_{jr_1} + 1) - n_{j,r_2}| - |n_{jr_1} - (n_{j,r_2} + 1)|.$$

A value of  $\Delta d_j^{PS}$  smaller than zero means that the allocation of the  $t$ -th subject to group one is better than that to group two. In this case, the difference between the number of subjects in category  $r_j$  in the groups is smaller than the difference that would be obtained if the subject were assigned to group two. If  $\Delta d_j^{PS}$  is larger than zero, assigning the  $t$ -th subject to group two is preferred. This is because this assignment reduces the gap between the number of subjects with covariate  $Z_j$  in category  $r_j$  in the groups, compared to assigning the subject to group one. Specifically, the discrepancy measure of Pocock and Simon<sup>32</sup> computes the overall imbalance among all covariates using

$$D_{PS} = \sum_{j=1}^p \Delta d_j^{PS}. \tag{1}$$

### 2.2.2 | Difference in means and variances

Nishi and Takaichi<sup>33</sup> compare the mean and standard deviation of the marginal distributions of the covariates in the two groups. Before allocating the  $t$ -th subject, we consider  $\bar{W}_{jk}$  and  $S_{jk}$  as the mean and standard deviation, respectively, of the  $j$ -th covariate in group  $k$ . For this covariate, we define the grand mean,  $\bar{W}_{j\cdot}$ , and grand standard deviation,  $S_{j\cdot}$ , across the current groups as

$$\bar{W}_{j\cdot} = \frac{n_1 \bar{W}_{j1} + n_2 \bar{W}_{j2}}{n_1 + n_2} \text{ and } S_{j\cdot} = \left( \frac{(n_1 - 1)S_{j1}^2 + (n_2 - 1)S_{j2}^2}{n_1 + n_2 - 2} \right)^{1/2}.$$

To assess the benefit of adding the  $t$ -th subject to a group, we calculate the hypothetical mean and standard deviation for the two groups including the subject. More specifically, we define  $\bar{W}_{j1}^{(+)}$  and  $\bar{W}_{j2}^{(+)}$  as the group means of the  $j$ -th covariate given that the  $t$ -th subject is included in groups one and two, respectively. The group standard deviations,  $S_{j1}^{(+)}$  and  $S_{j2}^{(+)}$ , of the  $j$ -th covariate are defined similarly. If group one includes the subject, the grand mean and grand standard deviation across the groups would be updated with

$$\bar{W}_{j\cdot}^{(1)} = \frac{(n_1 + 1)\bar{W}_{j1}^{(+)} + n_2 \bar{W}_{j2}}{n_1 + n_2 + 1} \text{ and } S_{j\cdot}^{(1)} = \left( \frac{n_1 S_{j1}^{(+)^2} + (n_2 - 1)S_{j2}^2}{n_1 + n_2 - 1} \right)^{1/2}.$$

If group two includes the subject, the updated mean and standard deviations are defined similarly and denoted as  $\bar{W}_{j\cdot}^{(2)}$  and  $S_{j\cdot}^{(2)}$ , respectively. That is, the coefficients of  $\bar{W}_{j2}$  and  $S_{j2}^2$  in  $\bar{W}_{j\cdot}^{(2)}$  and  $S_{j\cdot}^{(2)}$  carry the extra one.

Nishi and Takaichi<sup>33</sup> measure the goodness of allocation of the  $t$ -th subject to group  $k$  in terms of the  $j$ -th covariate using the function:

$$d_j(k) = |\bar{W}_{jk}^{(+)} - \bar{W}_{j\cdot}^{(+)}| - |\bar{W}_{jk} - \bar{W}_{j\cdot}| + |S_{jk}^{(+)} - S_{j\cdot}^{(+)}| - |S_{jk} - S_{j\cdot}|. \tag{2}$$

The first two elements on the right-hand side of (2) are the absolute distance between the new mean of group  $k$  and the new grand mean if subject  $t$  is in group  $k$ , and the absolute distance between the current mean of group  $k$  and the

current grand mean. A small difference between these elements is preferred because this indicates that the mean of group  $k$  with the new subject is closer to the grand mean. The last two elements of (2) are the absolute distance between the new group-specific and grand standard deviations if subject  $t$  is in group  $k$ , and the absolute distance between the current group-specific and grand standard deviations. Ideally, the difference between these elements is small too, as it implies that the standard deviation of group  $k$  with the new subject is close to the grand standard deviation.

We define the individual discrepancy in the  $j$ -th covariate as  $\Delta d_j^{NT} = d_j(1) - d_j(2)$ . A positive value of  $\Delta d_j^{NT}$  means that assigning the  $t$ -th subject to group two reduces the gap between the mean and standard deviations of the distributions of covariate  $j$  in the two groups, compared to assigning the subject to the other group. A negative value of  $\Delta d_j$  means that the assignment of the  $t$ -th subject to group one results in the smallest gap.

The overall discrepancy measure of Nishi and Takaichi<sup>33</sup> is

$$D_{NT} = \sum_{j=1}^p \Delta d_j^{NT} + \frac{n_1 - n_2}{n_1 + n_2}, \quad (3)$$

where the last term promotes subject assignments that create equal group sizes. So, the smaller the value of this term, the closer  $n_1$  is to  $n_2$ .

### 2.2.3 | Difference in distributions

The discrepancy measure of Ma and Hu<sup>34</sup> compares the marginal distributions of the covariates in the two groups using a Kernel density estimate.<sup>35</sup> It is assumed that the covariates of the subjects in the trial are standardized to have a mean of zero and a variance of one. The estimated density function for covariate  $W_j$  in group  $k$  is given by

$$\hat{f}_{jk}(w) = \frac{1}{n_k h(n_k)} \sum_{i=1}^{n_k} K\left(\frac{w - w'_{ijk}}{h(n_k)}\right)$$

where  $K(u)$  is the Kernel function,  $h(v)$  the bandwidth function, and  $w'_{ijk}$  the  $j$ -th covariate value of the  $i$ -th subject in group  $k$  in the standardized scale. Note that the density function is estimated using the covariate values of the  $t - 1$  subjects already in the trial. Ma and Hu<sup>34</sup> use the bandwidth function  $h(v) = v^{-0.2}$  and the normal Kernel  $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$ .

For the  $t$ -th subject, the following function defines an individual discrepancy in  $W_j$ :

$$\Delta d_j^{MH}(w'_{tj}) = \frac{n_1}{n} \hat{f}_{j1}(w'_{tj}) - \frac{n_2}{n} \hat{f}_{j2}(w'_{tj}), \quad (4)$$

where  $w'_{tj}$  is subject  $t$ 's  $j$ -th covariate value on the standardized scale. Equation (4) can then be interpreted as a weighted difference between the probability that the  $j$ -th covariate value of the  $t$ -th subject comes from treatment one or two, where the weights are the proportions of the groups. For example, if the groups are balanced and  $\Delta d_j^{MH}(w'_{tj}) > 0$ , the  $j$ -th covariate value of the new subject is more likely to follow the covariate distribution of group one, and therefore its allocation to group two is preferred to balance the distributions. If  $\Delta d_j^{MH}(w'_{tj}) < 0$ , the  $j$ -th covariate value of the new subject is more likely to follow the covariate distribution of group two and so, its allocation to group one is preferred.

The overall discrepancy measure of Ma and Hu<sup>34</sup> for all covariates is

$$D_{MH} = \sum_{j=1}^p \Delta d_j^{MH}(w'_{tj}). \quad (5)$$

## 2.3 | Decision rule

In the minimization framework, the final allocation of the new subject is typically determined using the biased-coin strategy of Efron.<sup>36</sup> To this end, we first set a user-specified probability,  $P_0$ , ranging from 1/2 to one. Next, we assign the

new subject to group one or two with probability  $P$  and  $1 - P$ , respectively, where  $P$  depends on  $P_0$  and  $D$ , the discrepancy measure value of the subject. Specifically, we have that

$$P = \begin{cases} P_0 & \text{if } D < 0, \\ 1 - P_0 & \text{if } D > 0, \\ 1/2 & \text{otherwise,} \end{cases} \quad (6)$$

where we compute  $D$  using either (1), (3), (5), or (10) for the method of Bertsimas et al.<sup>21</sup> discussed later.

The biased-coin strategy allows a minimization method to have randomness in allocating subjects to groups and maintain approximately balanced groups in terms of the covariates.<sup>37</sup> Allocation randomness reduces the possibility that the user, say, a physician, knows the allocation rule in advance and interferes with it. For example, the user may ignore the rule and assign the subject to the treatment that, in the user's opinion, would benefit the subject. In this case, the user commits selection bias,<sup>1</sup> which may invalidate the statistical conclusions of the trial.

If  $P_0 = 1/2$  in (6), the assignment of the subject to a group is completely random, thus avoiding selection bias. However, this value can cause a severe imbalance in the covariates between the groups at the end of the recruitment. A value of  $P_0$  equal to one results in groups with balance in their covariates, but a high risk of selection bias. This is because the assignment is deterministic when the discrepancy measure value of the subject differs from zero. A value of  $P_0$  between  $1/2$  and one is thus ideal as it gives the trade-off between covariate balance and allocation randomness. The value should be such that it biases the assignment towards the best group according to the discrepancy measure (to promote covariate balance) and provides a good degree of allocation randomness (to mitigate selection bias). Based on the simulation study of Toorawa et al.<sup>38</sup> on minimization methods, we use  $P_0 = 0.8$  unless otherwise stated.

### 3 | A MINIMIZATION METHOD ROOTED IN MATHEMATICAL PROGRAMMING

The covariate-adaptive method of Bertsimas et al.<sup>21</sup> consists of solving a subclass of MP problems called mixed-integer programming (MIP) problems.<sup>39</sup> In a MIP problem, the goal is to determine the values of a set of discrete and continuous decision variables to optimize a linear objective function while satisfying a set of constraints. A MIP problem formulation has the form:

$$\max_{\mathbf{x}} \quad \mathbf{c}^T \mathbf{x} \quad (7a)$$

subject to

$$\mathbf{G}\mathbf{x} = \mathbf{b}, \quad (7b)$$

$$\mathbf{H}\mathbf{x} \leq \mathbf{d}, \quad (7c)$$

$$x_i \in \mathbb{Z}, \quad \forall i \in \mathcal{J},$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  is an  $n \times 1$  vector of decision variables,  $\mathbf{c}$  is an  $n \times 1$  constant vector,  $\mathbf{G}$  is an  $m_1 \times n$  constant matrix,  $\mathbf{H}$  is an  $m_2 \times n$  constant matrix,  $\mathbf{b}$  is an  $m_1 \times 1$  constant vector,  $\mathbf{d}$  is an  $m_2 \times 1$  constant vector, and  $\mathcal{J}$  is a non-empty set of indices.

The above formulation is general and can be adapted to reflect the objective of a clinical trial. For example, in the context of non-adaptive designs for two-arm clinical trials, the entries of  $\mathbf{x}$  can take the value  $x_i = 1$  if the  $i$ -th subject is assigned to group one, and the value  $-1$  otherwise. If  $n$  is even, the equality constraint in (7b) can indicate that the group sizes are equal to  $n/2$ . This can be done by setting  $\mathbf{b} = \mathbf{0}$  and  $\mathbf{G} = \mathbf{1}_n^T$ , where  $\mathbf{1}_n$  is the  $n \times 1$  vector with all entries equal to 1. If  $n$  is odd, (7c) can indicate that the absolute difference in the group sizes does not exceed one. We can achieve this by setting  $\mathbf{d} = \mathbf{1}_2$  and  $\mathbf{H}$  to the  $2 \times n$  matrix  $(\mathbf{1}_n; -\mathbf{1}_n)^T$ . In this case, we should also set  $\mathbf{G} = \mathbf{0}_n^T$  to remove (7b), where  $\mathbf{0}_n$  is the  $n \times 1$  vector with all entries equal to 0. A goal of the MIP problem may be to balance the average



of the values of a covariate between the groups. To this end, the  $i$ -th entry of  $\mathbf{c}$  in (7a) is set to the covariate value of the  $i$ -th subject. The solution to this MIP problem is then the assignment of subjects to groups that minimizes the difference between the average covariate values between them and satisfies the constraints.

Commercial optimization software such as GUROBI, MOSEK, and BARON can solve several types of MIP problems using state-of-the-art optimization techniques.<sup>40</sup> Throughout the solution process, the software provide both the best feasible solution and the best bound for the objective function's optimal value obtained so far. If the objective value of the best solution is equal to the best bound, the solution is optimal.

We present the covariate-adaptive randomization method of Bertsimas et al.<sup>21</sup> and show that it falls within the minimization framework in Section 2.1. This close link between MP and minimization methods is new to the literature.

### 3.1 | Incorporating prior information on the covariates of future subjects

We denote the  $p \times 1$  covariate vector of the  $i$ -th subject as  $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{ip})^T$ , where  $w_{ij}$  is the  $j$ -th covariate value of this subject. We define the  $p \times 1$  sample mean vector  $\bar{\mathbf{w}}_t$  and the  $p \times p$  empirical covariance matrix  $\Sigma_t$  of the covariate vectors available before allocating the  $(t+1)$ -th subject as:

$$\bar{\mathbf{w}}_t = \frac{1}{t} \sum_{i=1}^t \mathbf{w}_i \text{ and } \Sigma_t = \frac{1}{t} \sum_{i=1}^t (\mathbf{w}_i - \bar{\mathbf{w}}_t)(\mathbf{w}_i - \bar{\mathbf{w}}_t)^T.$$

To incorporate prior information on future subjects, Bertsimas et al.<sup>21</sup> define the unknown covariate vectors for these subjects as  $\tilde{\mathbf{w}}_i$  and their entries as  $\tilde{w}_{ij}$ , with  $i = t+1, \dots, N$  and  $j = 1, \dots, p$ . These vectors are collected in the  $p \times (N-t)$  covariate matrix  $\tilde{\mathbf{W}} = [\tilde{\mathbf{w}}_{t+1}; \dots; \tilde{\mathbf{w}}_N]$  that belongs to the set

$$U_{\tilde{\mathbf{W}}} = \left\{ \tilde{\mathbf{W}} \in \mathbb{R}^{p \times (N-t)} \mid \tilde{\mathbf{w}}_i = \bar{\mathbf{w}}_t + (\Sigma_t)^{1/2} \boldsymbol{\epsilon}_i, i = t+1, \dots, N, \mathbf{E} \in U_E \right\},$$

where the  $p \times 1$  vector  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{ip})^T$  contains a random perturbation to the  $j$ -th covariate value of the  $i$ -th subject, denoted by  $\epsilon_{ij}$ . The  $p \times (N-t)$  matrix  $\mathbf{E} = [\boldsymbol{\epsilon}_{t+1}; \dots; \boldsymbol{\epsilon}_N]$  that collects all perturbation vectors belongs to the uncertainty set:

$$U_E = \left\{ \mathbf{E} \in \mathbb{R}^{p \times (N-t)} \mid \|\mathbf{E}\|_2 \leq \Gamma \sqrt{(N-t)p} \right\},$$

where  $\|\mathbf{E}\|_2$  is the Frobenious norm of  $\mathbf{E}$ .

The tuning parameter  $\Gamma$  in  $U_E$  controls how diffuse the information on the future subjects' covariate vectors is, with respect to the covariate vectors of the subjects in the trial. A small  $\Gamma$  value implies that the covariate vectors of the future subjects will be close to the covariate vector of the "average" subject in the trial. A large  $\Gamma$  value implies that future subjects' covariate vectors may heavily depart from this subject. Bertsimas et al.<sup>21</sup> suggest to choose a value of  $\Gamma$  between 1/2 and four. This recommendation is supported by their numerical experiments involving real and synthetic clinical trials.

### 3.2 | Problem formulation

The problem formulation of Bertsimas et al.<sup>21</sup> for covariate-adaptive allocation determines the assignments of current and future subjects that minimize the absolute difference in the means and variances of the covariates between the two treatment groups. It is assumed that  $N$  is even and that both groups will end up with  $N/2$  subjects. In their problem formulation, the assignment of the  $i$ -th subject to a group is expressed using the binary decision variable  $x_i$ , where  $i = t, \dots, N$ . The variable  $x_i$  takes the value of one if and only if the subject is assigned to group one. Otherwise, it takes a value of zero, and the subject is assigned to group two. Regarding the subjects already in the trial, we also denote their



group allocations using binary variables. Specifically, we use the binary variable  $\hat{x}_i$  which takes a value of one or zero if and only if the subject is in group one or two, respectively, where  $i = 1, \dots, t - 1$ .

Before allocating the  $t$ -th subject, we consider the mean of the  $j$ -th covariate from group  $k$ , denoted by  $\bar{W}_{jk}$ , which would be obtained at the end of the recruitment of subjects. Using the binary encoding for the group assignments, the difference in the mean of the  $j$ -th covariate between the groups is

$$\bar{W}_{j1} - \bar{W}_{j2} = \frac{2}{N} \left\{ \sum_{i=1}^{t-1} w_{ij}(2\hat{x}_i - 1) + w_{ij}(2x_t - 1) + \sum_{i=t+1}^N \tilde{w}_{ij}(2x_i - 1) \right\}.$$

Bertsimas et al.<sup>21</sup> compute an approximated variance of the  $j$ -th covariate under group  $k$ , denoted by  $V_{jk}$ , which would be obtained at the end of the assignment of subjects. In this approximation, the covariate vectors are centered around the  $j$ -th entry of the vector  $\bar{\mathbf{w}}_t$ , instead of the mean vector of all subjects in the trial. In this way, we limit the amount of uncertainty in the calculations of the group variances produced by the unknown covariate vectors of future subjects. The difference between approximated variances of the  $j$ -th covariate under groups one and two is

$$V_{j1} - V_{j2} = \frac{2}{N} \left\{ \sum_{i=1}^{t-1} (w_{ij} - \bar{w}_{ij})^2 (2\hat{x}_i - 1) + (w_{ij} - \bar{w}_{ij})^2 (2x_t - 1) + \sum_{i=t+1}^N (\tilde{w}_{ij} - \bar{w}_{ij})^2 (2x_i - 1) \right\},$$

where  $\bar{w}_{ij}$  is the  $j$ -th entry of  $\bar{\mathbf{w}}_t$ .

The problem formulation to obtain the optimal assignment of the  $t$ -th subject is

$$\min_{\mathbf{x}, \mathbf{w}, \mathbf{v}} \sum_{j=1}^p W'_j + \rho V_j. \quad (8a)$$

subject to

$$W'_j = |\bar{W}_{j1} - \bar{W}_{j2}|, \quad j = 1, \dots, p, \quad \forall \tilde{\mathbf{W}} \in U_{\tilde{\mathbf{w}}}. \quad (8b)$$

$$V_j = |V_{j1} - V_{j2}|, \quad j = 1, \dots, p, \quad \forall \tilde{\mathbf{W}} \in U_{\tilde{\mathbf{w}}} \quad (8c)$$

$$\sum_{i=1}^{t-1} \hat{x}_i + x_t + \sum_{i=t+1}^N x_i = \frac{N}{2}, \quad (8d)$$

$$x_i \in \{0, 1\}, \quad i = t, \dots, N. \quad (8e)$$

This problem formulation has  $N - t$  binary decision variables in  $\mathbf{x} = (x_t, x_{t+1}, \dots, x_N)^T$ ,  $2p$  continuous decision variables in  $\mathbf{W} = (W'_1, W'_2, \dots, W'_p)^T$  and  $\mathbf{V} = (V_1, V_2, \dots, V_p)^T$ ,  $2p$  constraints involving the unknown covariate vectors  $\tilde{\mathbf{w}}$  in (8b) and (8c), and a linear constraint in (8d). Technically, the problem formulation is a robust MIP problem,<sup>41</sup> because the values of some of the elements of (8b) and (8c) are uncertain.

The objective function in (8a) is a weighted sum of the absolute differences in the means and the approximated variances between the groups for all covariates. The weights are determined by the tuning parameter  $\rho$ , which controls the trade-off between the imbalance in the group means and the approximated group variances. A large value of  $\rho$  emphasizes a close match between the approximated variances of the group in the covariates, while a small value emphasizes a close match between the means of the groups.

The constraints in (8b) and (8c) give the absolute differences between the group means and the approximated group variances of the covariates, respectively. The constraint in (8d) states that the size of group one must be  $N/2$  at the end of the trial. It also implies that the size of group two must also be  $N/2$ . The constraints contained within (8e) ensure that the decision variables  $x_i$  are binary.

### 3.3 | Reformulation as a minimization method

Bertsimas et al.<sup>21</sup> show that the problem formulation in (8a)–(8e) can be reformulated as a standard MIP problem, whose solution involves the evaluation of two feasible solutions only. More specifically, we need to evaluate the two sets of potential groups that would result if the  $t$ -th subject is allocated to group one or two. Here, we show that the covariate-adaptive randomization method of Bertsimas et al.<sup>21</sup> is actually a minimization method.

To see this, we first let  $\mathbf{v}_j$  be the  $j$ -th row of  $(\Sigma_t)^{1/2}$ . In the alternative problem formulation of Bertsimas et al.,<sup>21</sup> the absolute difference between the group means for the  $j$ -th covariate in (8b) reduces to

$$\frac{N}{2} W'_j = \left| \sum_{i=1}^{t-1} (w_{ij} - \bar{w}_{ij})(2\hat{x}_i - 1) + (w_{ij} - \bar{w}_{ij})(2x_t - 1) \right| + \Gamma \|\mathbf{v}_j\|_2 (N - t) \sqrt{p}, \quad (9)$$

where  $x_t \in \{0, 1\}$  is the only decision variable. The uncertainty in the covariate values of future subjects is in the last term of (9).

The expression for the absolute difference between the approximated group variances in the  $j$ -th covariate ( $V_{j\cdot}$ ) is slightly more complicated; see Bertsimas et al.<sup>21</sup> There is one expression for the case where  $p = 1$ , and another for the case where  $p \geq 2$ . We adopt the latter because our focus is on balancing several covariates simultaneously. In this case,  $V_{j\cdot}$  is the maximum of the quantities:

$$\begin{aligned} & \frac{2}{N} \sum_{i=1}^{t-1} (w_{ij} - \bar{w}_{ij})^2 (2\hat{x}_i - 1) + (w_{ij} - \bar{w}_{ij})^2 (2x_t - 1) + \tilde{\Gamma} \|\mathbf{v}_j\|_2^2 I(k - n_1 - x_t - 1) \\ & \text{and} \\ & -\frac{2}{N} \sum_{i=1}^{t-1} (w_{ij} - \bar{w}_{ij})^2 (2\hat{x}_i - 1) - (w_{ij} - \bar{w}_{ij})^2 (2x_t - 1) - \tilde{\Gamma} \|\mathbf{v}_j\|_2^2 I(k - n_2 - x_t - 1), \end{aligned}$$

where  $\tilde{\Gamma} = \Gamma^2 (N - t)p$ ,  $n_1 = \sum_{i=1}^{t-1} \hat{x}_i$ ,  $n_2 = t - 1 - \sum_{i=1}^{t-1} \hat{x}_i$ , and  $I(z)$  is an indicator function that equals one if  $z \geq 0$  and zero otherwise. As with the difference in group means, the only decision variable in these expressions is  $x_t$  and the uncertainty in the future covariate values is in their last terms.

Since the minimization of  $V_{j\cdot}$  is straightforward, Bertsimas et al.<sup>21</sup> use the difference in the group (approximated) standard deviations instead, so as to put the two competing objectives on the same scale and facilitate an intuitive choice of the tuning parameter  $\rho$ . Therefore, the discrepancy measure of these authors is

$$D_B = \sum_{j=1}^p W'_j + \rho \sum_{j=1}^p V_j^{1/2}. \quad (10)$$

Using numerical experiments, Bertsimas et al.<sup>21</sup> found that  $\rho = 6$  generally leads to a good trade-off between the balance in terms of the group means and variances.

The minimization method follows the steps in Section 2.1 with few modifications. In Step 2, for each new subject to be included in the trial, we set a value of the uncertainty parameter  $\Gamma$  at random from the interval  $[1/2, 4]$ . The numerical experiments of Bertsimas et al.<sup>21</sup> show that choosing the value of  $\Gamma$  independently at random at each time step protects against selection bias. Since the selection process of  $\Gamma$  incorporates randomness in the decision process, we determine the final allocation of the subject using  $P_0 = 1$  in (6).

## 4 | NUMERICAL COMPARISONS

In this section, we compare the minimization method of Bertsimas et al.<sup>21</sup> with those in Section 2 using two real clinical trials. First, we introduce the metrics to assess the balance in the covariate distributions across the groups and the randomness in the treatment allocation of the methods. Next, we show the performance of the methods for each trial separately and end the section with a discussion.

We refer to the minimization methods of Pocock and Simon,<sup>32</sup> Nishi and Takaichi,<sup>33</sup> Ma and Hu<sup>34</sup> and Bertsimas and Weismantel<sup>39</sup> as PS, NT, MH, and BKW, respectively. For the PS method, we use  $c = 3$  categories for the surrogate discrete covariates. The supplementary materials accompanying the article have an R implementation of all the minimization methods discussed here.

## 4.1 | Evaluation metrics

### 4.1.1 | Covariate balance

We use several evaluation metrics to measure the balance in the distributions of the covariates between the two treatment groups. They include the absolute difference in the size of the groups, the absolute difference between the means of the  $j$ -th covariate across the groups, and the absolute difference between the standard deviations of the  $j$ -th covariate across the groups. Ideally, the values of these metrics are equal or close to zero.

The differences between the individual means and standard deviations for each individual covariate are marginal metrics, because they measure the discrepancy between the marginal distributions of the covariates in the two groups. However, they do not measure the discrepancies between the joint distributions of the covariates that generate the data. Balance in the joint covariate distribution is more general than balance in the marginal distribution of each covariate among the groups. This is because a joint covariate distribution balance implies a balance in each marginal covariate distribution, regardless of whether the covariates are dependent or independent. In contrast, a balance in each marginal covariate distribution may only achieve a balance in the joint covariate distributions of the groups when the covariates are independent.

To measure the discrepancies in the joint distribution of the covariates among the groups, we use the energy distance.<sup>23</sup> Let  $\mathbf{W}_k$  be the  $N_k \times p$  covariate matrix for the  $k$ -th group and  $\mathbf{w}_{ik}$  be the  $i$ -th row of this matrix, where  $N_k$  is the size of the  $k$ -th group at the end of the trial and  $k = 1, 2$ . The energy distance for two independent random samples is

$$E = \frac{2}{N_1 N_2} \sum_{u=1}^{N_1} \sum_{v=1}^{N_2} \|\mathbf{w}_{u1} - \mathbf{w}_{v2}\| - \frac{1}{N_1^2} \sum_{u=1}^{N_1} \sum_{l=1}^{N_1} \|\mathbf{w}_{u1} - \mathbf{w}_{l1}\| - \frac{1}{N_2^2} \sum_{v=1}^{N_2} \sum_{m=1}^{N_2} \|\mathbf{w}_{v2} - \mathbf{w}_{m2}\|, \quad (11)$$

where  $\|\mathbf{u}\|$  is the  $L_2$ -norm of vector  $\mathbf{u}$ .

Szekely and Rizzo<sup>42</sup> use the statistic  $E$  to test the homogeneity of two multivariate joint distributions. Using a permutation test, they reject the hypothesis of equal joint distributions when the value of  $E$  is large. Therefore, in our context, low values of the energy distance indicate that the joint covariate distributions of the groups are similar. To the best of our knowledge, the energy distance is the first metric to evaluate minimization methods in terms of the balance in the joint covariate distributions between the groups they produce.

The energy distance can assess the balance in the distributions of many covariates simultaneously using a single number. This is in contrast to marginal metrics such as the absolute difference between the group means or group standard deviations of the covariates. To evaluate the quality of a clinical trial in terms of these metrics, we must compare several values simultaneously, which can be cumbersome in the presence of many covariates. Therefore, the energy distance provides the user with a simple yet effective summary of the covariate balance in a trial, compared to marginal metrics.

### 4.1.2 | Allocation randomness

We evaluate the randomness of allocation of a minimization method using the Correct Guess (CG) probability.<sup>43,44</sup> Recall that the sizes of groups one and two before allocating the  $t$ -th subject are  $n_1$  and  $n_2$ , respectively, and that, in the problem formulation of Bertsimas et al.,<sup>21</sup>  $x_t$  is one if the  $t$ -th subject is assigned to group one and it is zero otherwise. At time step  $t$ , the CG probability is

$$\text{CG}_t = \begin{cases} 1 & \text{if } (n_1 < n_2 \text{ and } x_t = 1) \text{ or } (n_1 > n_2 \text{ and } x_t = 0) \\ 1/2 & \text{if } (n_1 = n_2) \\ 0 & \text{if } (n_1 > n_2 \text{ and } x_t = 1) \text{ or } (n_1 < n_2 \text{ and } x_t = 0) \end{cases}.$$

The value of  $CG_t$  is one if the group sizes are unequal and we can determine the assignment of the  $t$ -th subject using the smallest group size. If the group sizes are unequal but we cannot determine the assignment using this rule, the value of  $CG_t$  is zero. If the group sizes are equal at time  $t$ ,  $CG_t$  is  $1/2$ .

We compute the CG probability for all treatment allocations in Step 2 of the minimization framework in Section 2.1. That is, we calculate  $CG_t$  for  $t = n_0 + 1, n_0 + 2, \dots, N$ . We use the mean of the  $CG_t$  values as an overall measure of allocation randomness.<sup>44</sup> We prefer a low mean CG probability because this implies a low mean probability of guessing the assignment of the next subject using the previous allocations. Simple randomization—in which we assign each subject to a group completely at random and independently of the previous assignments—has a mean CG probability of  $1/2$ .

## 4.2 | The pembrolizumab clinical trial

Sundahl et al.<sup>45</sup> studied the safety of pembrolizumab when combined with sequential or concomitant body radiotherapy in metastatic bladder cancer. The response under study was the dose-limiting toxicity measured on a specific scale. The clinical trial consisted of 18 subjects whose prognostic factors included age, sex, hemoglobin concentration, a modified proportion score of PD-L1, smoking status, among others. The subjects were enrolled into two equally sized groups. Groups one and two were administered pembrolizumab using sequential and concomitant body radiotherapy, respectively.

We compare the actual assignment of the subjects to the two groups of Sundahl et al.<sup>45</sup> with the minimization methods to balance the continuous covariates: age, the modified proportion score of PD-L1, and hemoglobin concentration. To evaluate the methods, we standardized the covariates to have a mean of zero and a standard deviation of one. We then applied each method 1000 times to balance the covariates in the standardized set.

Figures 1–6 show the results of the simulations. Specifically, Figures 1, 5, and 6 show the distribution of the absolute difference between group sizes, the energy distance, and the mean CG probability, respectively, obtained by the methods. Figures 2, 3, and 4 show the distribution of the difference in the mean and standard deviations of the groups for age, modified proportion score of PD-L1, and hemoglobin concentration, respectively. Figures 2–5 have a horizontal dashed line placed at the corresponding metric values observed in the clinical trial.

Regarding the absolute difference between the group sizes, Figure 1 shows that the median absolute difference is two for the PS, NT, and MH methods. In contrast, the group absolute differences obtained by the BMW method are all equal to zero. This is because, in contrast to the other methods, the BMW method enforces group size balance.

Figure 2A shows that all minimization methods improve the actual group assignment for the absolute difference between group means of age. This is because at least 75% of the group mean differences for all methods are below the dashed line. The BKW method has a lower median and third quartile of group mean differences than the rest.

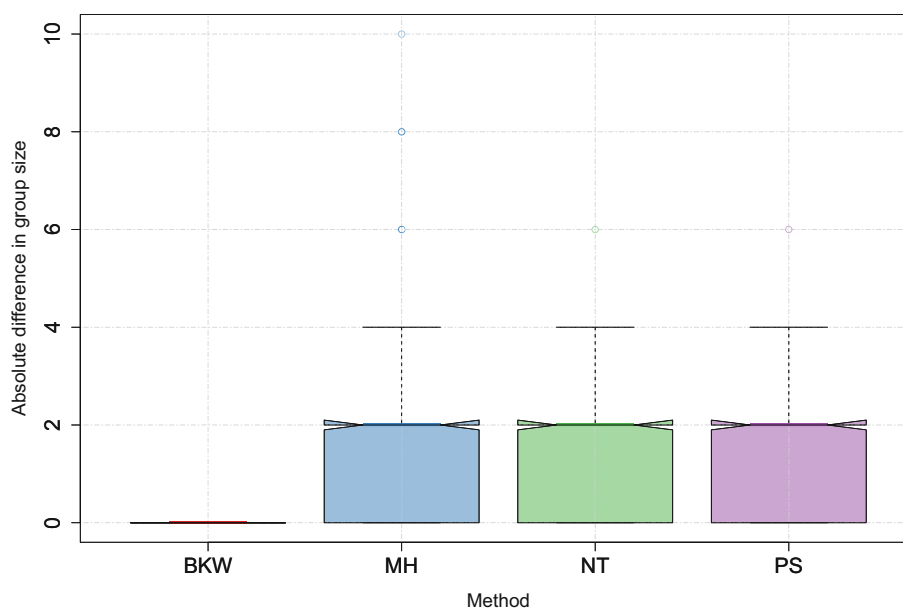
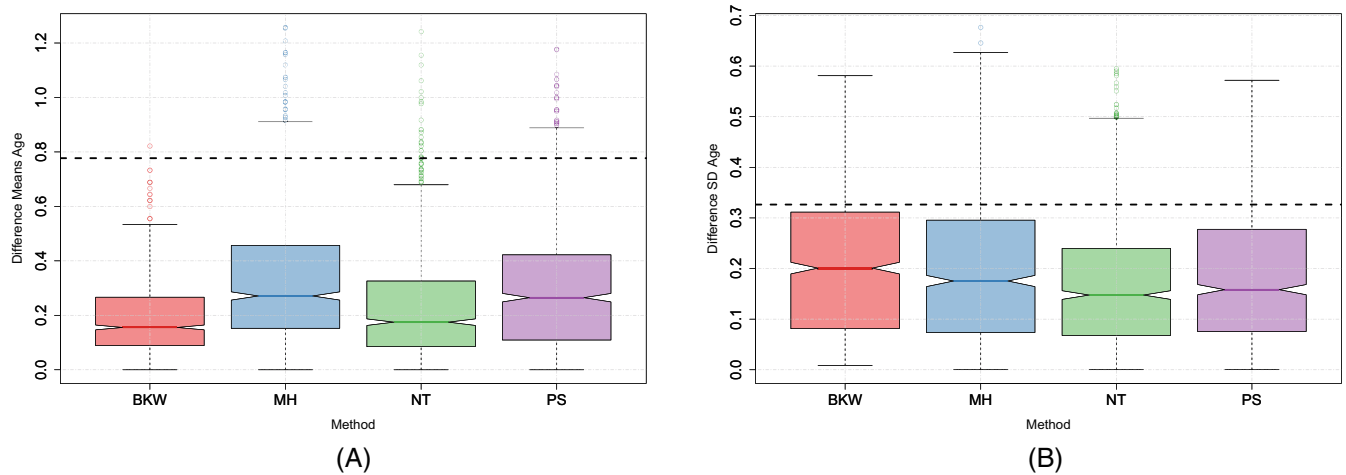
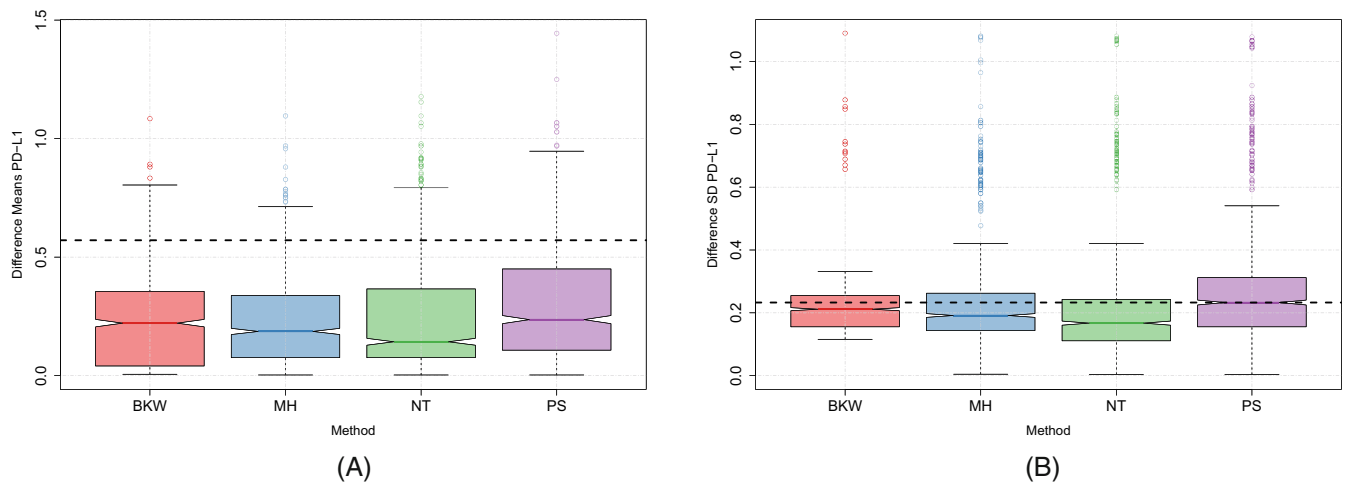


FIGURE 1 Boxplots of the absolute difference in group sizes obtained for the pembrolizumab trial.



**FIGURE 2** Boxplots of the absolute differences in group means and standard deviations for age in the pembrolizumab trial. The dashed line shows the actual absolute differences. (A) Absolute difference in means, (B) Absolute difference in standard deviations.

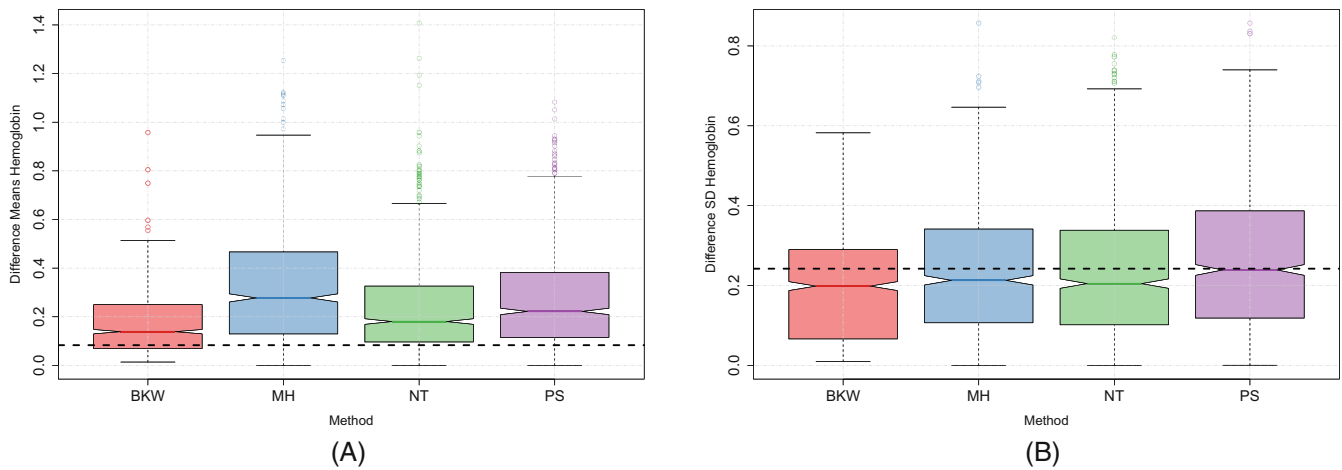


**FIGURE 3** Boxplots of the absolute differences in group means and standard deviations for PD-L1 in the pembrolizumab trial. The dashed line shows the actual absolute differences. (A) Absolute difference in means, (B) Absolute difference in standard deviations.

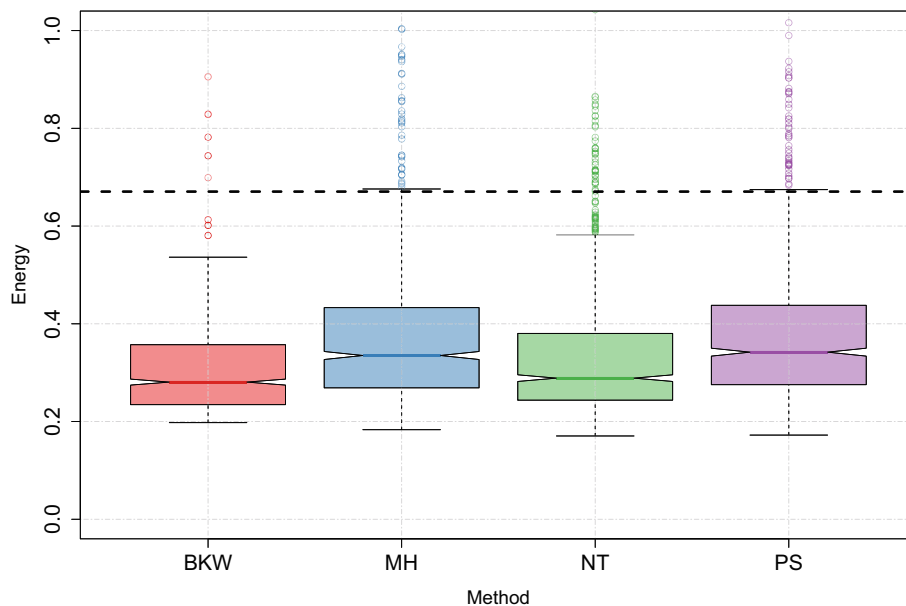
Figure 2B shows that all methods are better than the actual group assignment to balance the standard deviations of age, because at least 75% of their absolute differences in the group standard deviations are below that obtained from the actual assignment. The NT method is better than the others because both its median and third quartile of group mean differences are smaller than those of the other methods.

Regarding the group means for PD-L1, Figure 3A shows that all minimization methods improve the actual assignment of groups, since 75% of their absolute group mean differences are below the observed in the clinical trial. In this case, the NT method is the most successful because it has a lower median absolute group difference than the others. Regarding the group standard deviations for PD-L1, Figure 3B shows that at least 50% of the group absolute differences obtained from the methods are smaller than the actual group difference in the clinical trial. The NT method is again better than the rest because its median absolute difference in group standard deviation is smaller than the others.

Regarding hemoglobin concentration, Figure 4A shows that, to a large extent, all methods have greater absolute group mean differences than the one observed in the clinical trial. This is because either the first or second quartile of each method is above the dashed line in the figure. For the absolute group difference of the standard deviations, Figure 4B shows that about 50% of the group differences for all methods are below the actual group difference observed. In any case, the BKW method performs slightly better than the other methods for this covariate, as its median group differences are smaller than the others.



**FIGURE 4** Boxplots of the absolute differences in group means and standard deviations for hemoglobin in the pembrolizumab trial. The dashed line shows the actual absolute differences. (A) Absolute difference in means, (B) Absolute difference in standard deviations.



**FIGURE 5** Boxplots of the energy distance values for the pembrolizumab trial. The dashed line shows the energy distance value observed in the trial.

Figure 5 shows that all methods are better at balancing the joint distribution of the covariates between the two groups, than the actual group assignment. This is because at least 75% of the energy distance values for all methods are smaller than 0.671, which was observed from the actual group assignment. The BKW and NT methods are the best in terms of the energy distance, since they have a lower median energy distance value than the rest. However, the BKW method has slightly lower median and third quartile values than the NT method.

Regarding randomness in assigning subjects to groups, Figure 6 shows that the median values of the mean CG probability are the same for all methods. However, the NT method tends to have a smaller mean CG probability than the rest because its first and third quartiles are smaller than those of the others. In any case, the first quartile of the mean CG probability is at least 1/2 for all methods.

### 4.3 | The infant spasms clinical trial

Chiron et al.<sup>46</sup> studied the effect of vigabatrin on spasms due to tuberous sclerosis in infants. To this end, they conducted a clinical trial with 22 infants whose prognostic factors included age, sex, and duration and frequency of spasms.

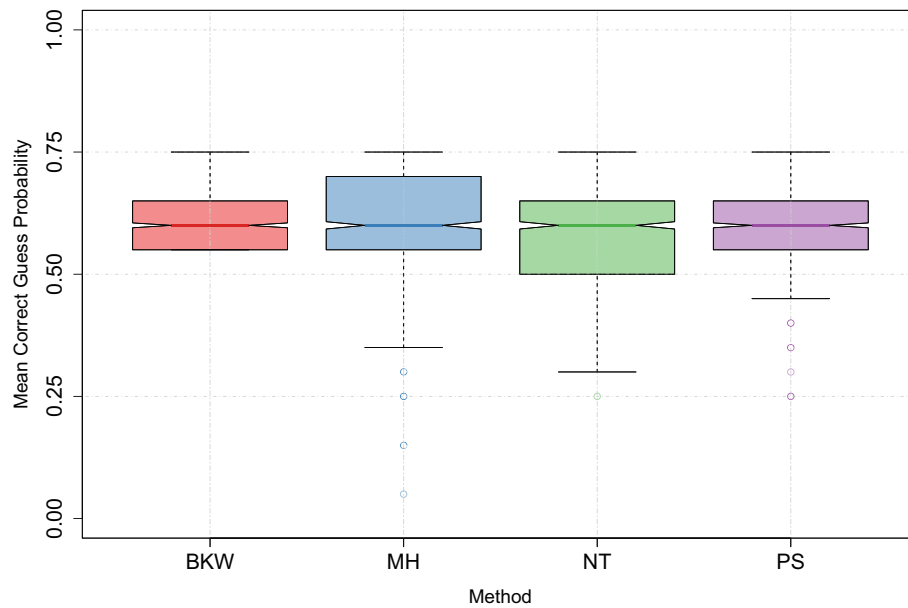


FIGURE 6 Boxplots of the mean correct guess probability values for the pembrolizumab trial.

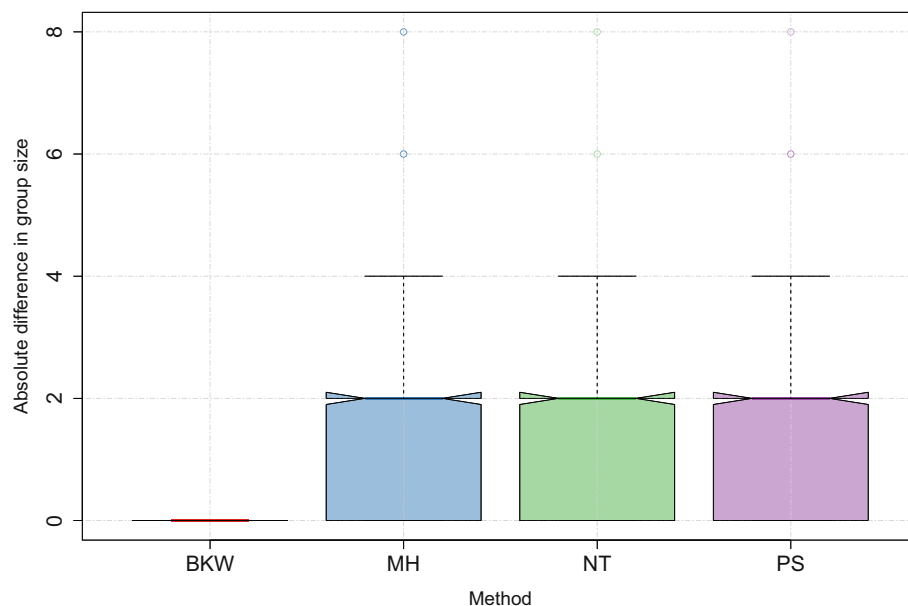
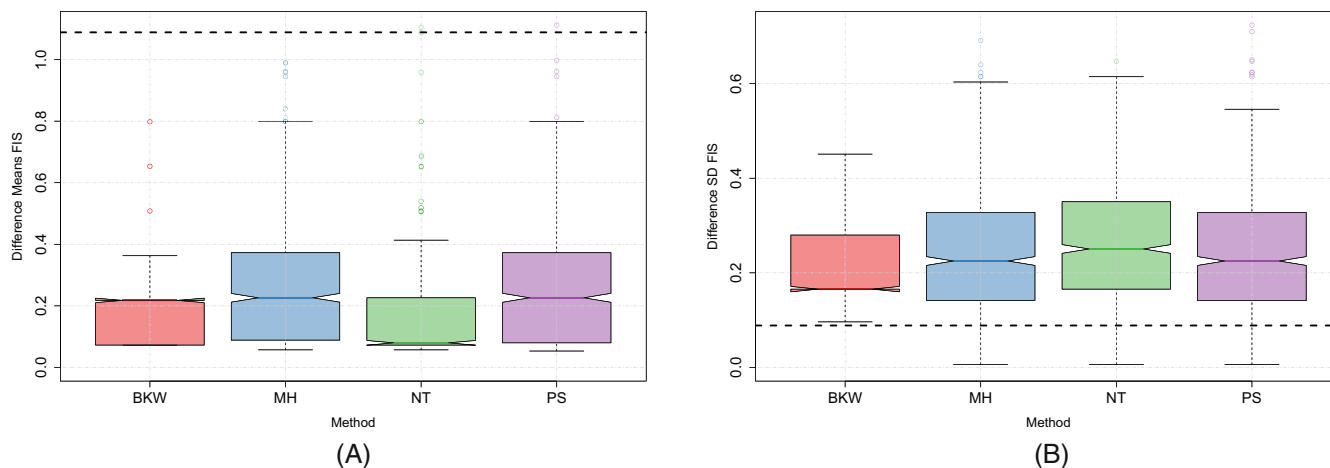


FIGURE 7 Boxplots of the absolute difference in group sizes obtained for the infant spasms trial.

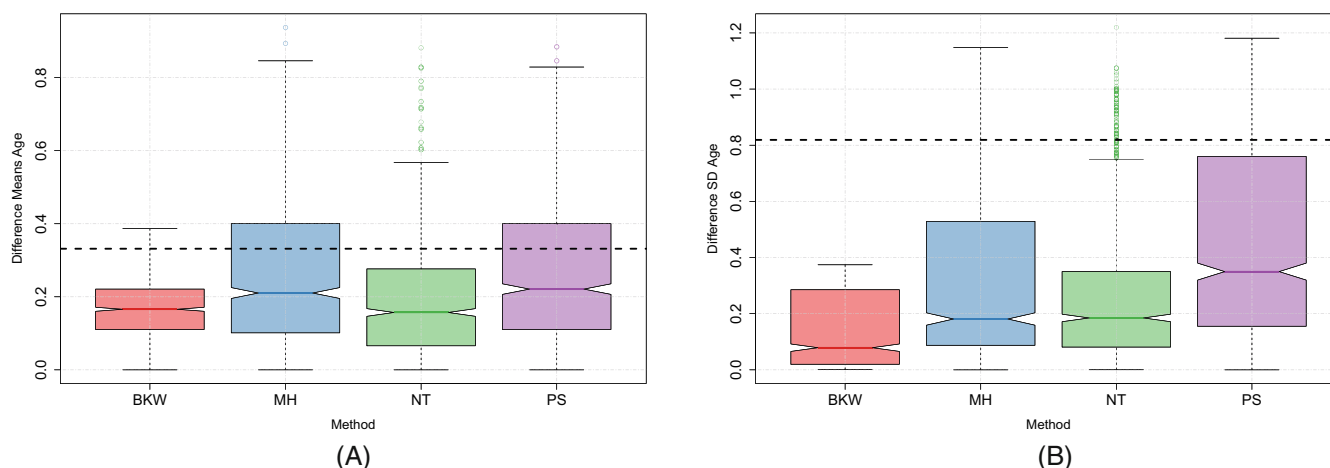
As an alternative treatment, Chiron et al.<sup>46</sup> used hydrocortisone which is a standard steroid. In the trial, the infants were enrolled into two equally sized groups. One group was treated with vigabatrin while the other was treated with hydrocortisone. To assess the efficacy of the treatments, the authors used the time to disappearance of the spasms, the tolerability to the treatment, the evolution of the development quotient, among other responses.

We compare the actual group assignment of Chiron et al.<sup>46</sup> with minimization methods to balance the continuous covariates: frequency of infantile spasms (FIS) and age. Similarly to the pembrolizumab trial, we standardized these covariates to have a mean of zero and a standard deviation of one, and executed each method 1000 times. Figures 7–11 visualize the simulation results for the infant spasms trial. Figures 7, 10, and 11 are similar to Figures 1, 5, and 6 for the pembrolizumab trial. Figures 8 and 9 show the distribution of the difference in the mean and standard deviations of the groups for the FIS and age, respectively. Figures concerning the balance in covariate distributions have a horizontal dashed line placed at the observed metric values in the trial.





**FIGURE 8** Boxplots of the absolute differences in group means and standard deviations for FIS in the infant spasms trial. The dashed line shows the actual absolute differences. (A) Absolute difference in means, (B) Absolute difference in standard deviations.



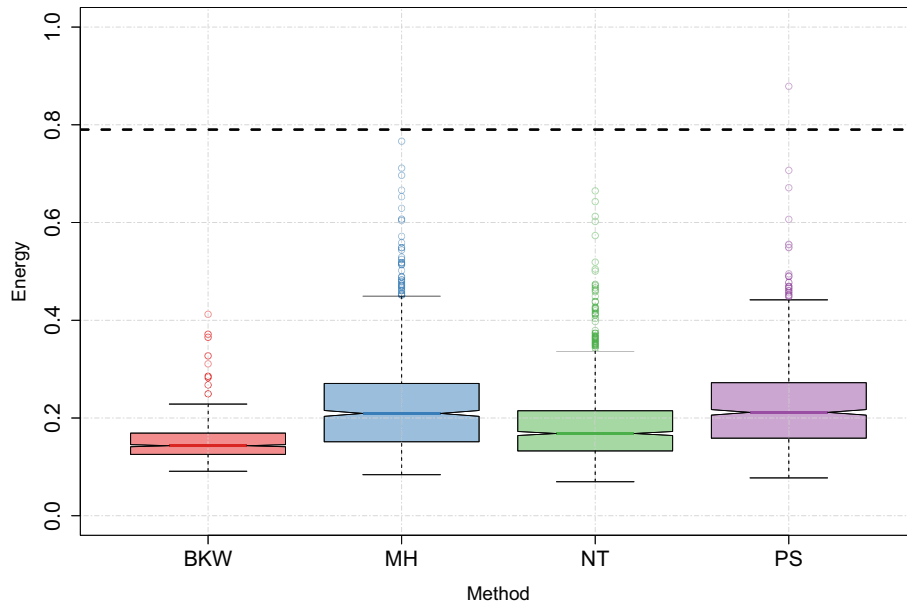
**FIGURE 9** Boxplots of the absolute differences in group means and standard deviations for age in the infant spasms trial. The dashed line shows the actual absolute differences. (A) Absolute difference in means, (B) Absolute difference in standard deviations.

Regarding the balance in the group sizes, Figure 7 shows that the BKW method is better than the others because all its group size differences are zero. In contrast, the other methods have a median absolute difference between group sizes equal to two.

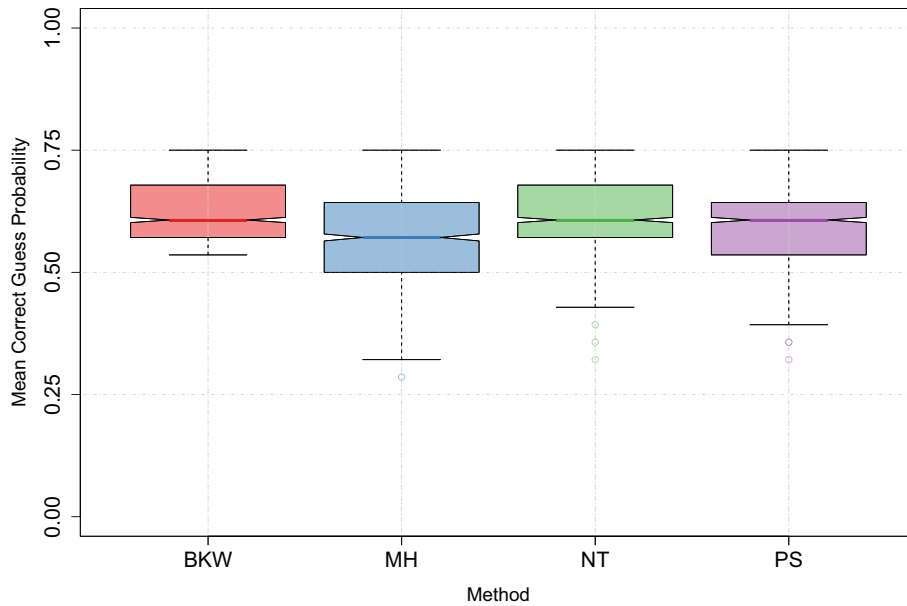
Figure 8A shows that all minimization methods improve the actual group assignment in terms of the mean of the FIS. This is because virtually all their absolute differences are smaller than the actual one indicated by the dashed line in the figure. The NT method has the smallest median absolute group difference. Regarding the absolute difference in the standard deviations for the FIS, Figure 8B shows that the BKW method has smaller first and second quartiles than those of the other methods. However, none of the methods outperforms the actual group assignment in this case. This is because the first quartiles of the absolute group differences of all methods are greater than the difference observed in the clinical trial.

Figure 9 shows that the BKW and NT methods are generally better than the others in terms of balancing age, since at least 75% of their group differences are below the differences observed in the trial. For the absolute difference in group means, the NT method has the lowest median, but the BKW has the smallest third quartile and dispersion; see Figure 9A. For the differences between group standard deviations, the BKW method has a median and third quartile that are smaller than those of the other methods; see Figure 9B.

Figure 10 shows that all methods are better than the actual clinical trial in terms of the energy distance. This is because the energy distance values of all the methods are smaller than the observed energy distance value, which was 0.79. The BKW method is the best in this case, as its first and third quartiles are smaller than those of the other methods.



**FIGURE 10** Boxplots of the energy distance values for the infant spasms trial. The dashed line shows the energy distance value observed in the trial.

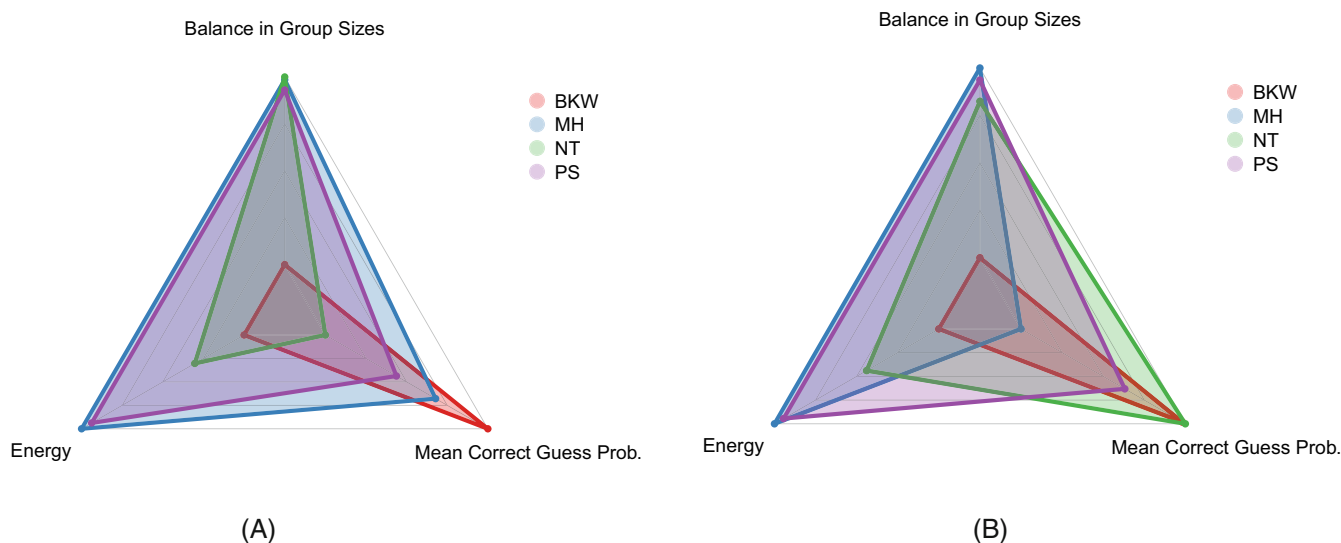


**FIGURE 11** Boxplots of the mean correct guess probability values for the infant spasms trial.

Figure 11 shows that the MH method has smaller mean CG probability values than the other methods. This is because all its quartiles are smaller than those of the others. However, none of the methods succeed in minimizing the mean CG probability, since more than 75% of their values are higher than or equal to  $1/2$ .

#### 4.4 | Discussion

In Sections 4.2 and 4.3, we provided a detailed comparison of the minimization methods in terms of the balance in the group sizes and the marginal and joint distributions of the covariates, as well as the randomness in the assignment of subjects to groups. We presented the analysis separately for each evaluation metric. To help users select a minimization



**FIGURE 12** Radar plot of the average values for the absolute difference in group sizes, energy distance, and the mean correct guess probability values for the pembrolizumab and infant spasms trial. (A) Pembrolizumab trial, (B) Infant spasms trial.

method, we constructed a radar plot<sup>47</sup> to visualize the overall performance of the methods in the trials. More specifically, the radar plot shows the average values for the absolute difference in group sizes, energy distance, and mean CG probability obtained by the methods. Each corner of the plot corresponds to the highest average value of a metric among all methods. Since lower average metric values are preferred, the points of a method should be close to the center of the plot.

Figure 12A,B shows the radar plot for the pembrolizumab and infant spasms trial, respectively. On average, the BKW method outperforms the other methods in terms of the absolute difference in group sizes and covariate balance for both trials. This is because their average values for these metrics are closer to the center of the plot than those of the other methods. Therefore, we recommend the BKW method if the goal is to balance the distributions of the covariates. However, the BKW method has higher average values of the mean CG probability than the other methods in both trials. If the goal is to minimize the mean CG probability, we recommend the NT or MH method because they have the smallest average mean CG probability values for the pembrolizumab and infant spasms trial, respectively.

The conclusion of Figure 12 on the BKW method agrees with the literature on minimization methods, since balance and allocation randomness are conflicting objectives.<sup>1,2</sup> In other words, a minimization method with a good balance in group sizes and covariate distributions tends to have a low degree of randomness in the assignment of subjects to groups. In contrast, a method with a high degree of randomness of allocation typically performs poorly in terms of balancing group sizes and covariate distributions. In any case, the summary of the covariate balance given by the energy distance allows us to easily see this phenomenon using the radar plots in Figure 12. We therefore recommend using the energy distance to evaluate the performance of minimization methods in terms of covariate balance.

The method of Bertsimas et al.<sup>21</sup> has a tuning parameter  $\rho$  that controls the trade-off between the balance in the group means and the group standard deviations of the covariates. In our study, we followed their recommendation and used  $\rho = 6$ , thus favoring the balance in the group standard deviations over that of the group means. Using additional experiments (not shown here), we studied the effect of the tuning parameter on the balance of marginal covariate distributions, the energy distance, and the correct guess probability for the two trials. We found that decreasing the value of  $\rho$  generally results in a better balance between the group means of the covariates at the expense of an increase in the imbalance between the group standard deviations. The energy distance and the correct guess probability were not severely affected by the value of  $\rho$ . In any case, none of the values that we tested significantly improved all metrics simultaneously for both trials. We conclude that  $\rho = 6$  works well for the clinical trials discussed here.

## 5 | CONCLUDING REMARKS

We studied the mathematical programming method of Bertsimas et al.<sup>21</sup> to balance continuous covariates in small two-arm clinical trials. We showed that this method belongs to the class of minimization methods that are well known in

the literature on clinical trials. We compared the method with the standard minimization methods of Pocock and Simon,<sup>32</sup> Nishi and Takaichi,<sup>33</sup> and Ma and Hu,<sup>34</sup> in terms of how well they balance the marginal and joint distributions of the covariates between the groups, and the randomness in the allocation of future subjects. For our two case studies, we show that the method of Bertsimas et al.<sup>21</sup> outperforms the other methods in terms of the balance in the group sizes and covariate distributions. However, the method has a low degree of randomness in assigning subjects to groups, which may result in selection bias. Remarkably, our study also shows that none of the methods is entirely successful for our case studies, as they did not balance the marginal distribution of one of their variables. Therefore, better methods are needed to balance the covariate distributions in small trials.

The radar plots in Section 4.4 allow us to compare the minimization methods in terms of three important dimensions: balance in group sizes, covariate balance, and randomness in group assignment. However, another important dimension of a small clinical trial is its ethics.<sup>48,49</sup> For example, when studying a rare disease, a clinical trial in which the most promising treatment is assigned more frequently to subjects than the other treatment is more attractive in terms of ethics, compared to a completely randomized trial without selection bias. However, Rosenberger and Sverdlov<sup>2</sup> show that it is difficult to develop a clinical trial with a high degree of randomization that is in accordance with ethics, because these are conflicting objectives. To our knowledge, there are no mathematical programming approaches to obtain clinical trials with a clear trade-off between these dimensions. Therefore, more research is needed in this direction. To this end, we can explore multi-objective mathematical programming<sup>50,51</sup> to obtain a list of solutions (i.e., clinical trials) that are good in at least one of the objectives and competitive in the others.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest in this work.

## DATA AVAILABILITY STATEMENT

Data are simulated in our paper.

## ORCID

Alan R. Vazquez  <https://orcid.org/0000-0002-3658-0911>

Weng-Kee Wong  <https://orcid.org/0000-0001-5568-3054>

## REFERENCES

1. Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and Practice*. 2nd ed. Wiley Series in Probability and Statistics; 2016.
2. Rosenberger WF, Sverdlov O. Handling covariates in the design of clinical trials. *Stat Sci*. 2008;23:404-419.
3. Rosenberger WF, Sverdlov O, Hu F. Adaptive randomization for clinical trials. *J Biopharm Stat*. 2012;22:719-736.
4. Cornu C, Kassai B, Fisch R, et al. Experimental designs for small randomised clinical trials: an algorithm for choice. *Orphanet J Rare Dis*. 2013;8:1-12.
5. Balcazar HG, Byrd TL, Ortiz M, Tondapu SR, Chavez M. A randomized community intervention to improve hypertension control among mexican americans: using the promotoras de salud community outreach model. *J Health Care Poor Underserved*. 2009;20:1079-1094.
6. Evans CHJ, Ildstad ST. *Small Clinical Trials: Issues and Challenges*. National Academy Press; 2001.
7. Suresh KP. An overview of randomization techniques: an unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011;4:8-11.
8. Hoare ZSJ, Whitaker CJ, Whitaker R. Introduction to a generalized method for adaptive randomization in trials. *Trials*. 2013;14:1-8.
9. Bradley SP, Hax AC, Magnanti TL. *Applied Mathematical Programming*. Addison-Wesley; 1977.
10. Duarte BMP, Granjo JFO, Wong WK. Optimal exact designs of experiments via mixed integer nonlinear programming. *Stat Comput*. 2020;30:93-112.
11. Duarte BMP, Wong WK. A semi-infinite programming based algorithm for finding minimax  $D$ -optimal designs for nonlinear models. *Stat Comput*. 2014;24:1063-1080.
12. Duarte BMP, Wong WK. Finding Bayesian optimal designs for nonlinear models: a semidefinite programming-based approach. *Int Stat Rev*. 2015;83:239-262.
13. Duarte BMP, Wong WK, Atkinson AC. A semi-infinite programming based algorithm for determining  $T$ -optimum designs for model discrimination. *J Multivar Anal*. 2015;135:11-24.
14. Duarte BMP, Wong WK, Dette H. Adaptive grid semidefinite programming for finding optimal designs. *Stat Comput*. 2018;28:441-460.
15. Vazquez AR, Wong WK, Goos P. Constructing two-level QB-optimal screening designs using mixed-integer programming and heuristic algorithms. *Stat Comput*. 2023;33:1-18.
16. Bulutoglu DA, Margot F. Classification of orthogonal arrays by integer programming. *J Stat Plan Inference*. 2008;138:654-666.

17. Grömping U, Fontana R. An algorithm for generating good mixed level factorial designs. *Comput Stat Data Anal.* 2019;137:101-114.
18. Phoa FK, Pan YH, Xu H. Analysis of supersaturated designs via the Dantzig selector. *J Stat Plan Inference.* 2009;139:2362-2372.
19. Vazquez AR, Schoen ED, Goos P. A mixed integer optimization approach for model selection in screening experiments. *J Qual Technol.* 2021;53:243-266.
20. Williamson SF, Jacko P, Villar SS, Jaki T. A Bayesian adaptive design for clinical trials in rare diseases. *Comput Stat Data Anal.* 2017;113:136-153.
21. Bertsimas D, Korolko N, Weinstein AM. Covariate-adaptive optimization in online clinical trials. *Oper Res.* 2019;67:1150-1161.
22. Bhat N, Farias VF, Moallemi CC, Sinha D. Near-optimal A-B testing. *Manag Sci.* 2020;66:4477-4495.
23. Székely GJ, Rizzo ML. Energy statistics: a class of statistics based on distances. *J Stat Plan Inference.* 2013;143:1249-1272.
24. Lin Y, Zhu M, Su Z. The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials. *Contemp Clin Trials.* 2015;45:21-25.
25. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol.* 1999;52:19-26.
26. Leung O, Signorini D, Simes J, Beller E, Gebiski V, Callaghan T. Dynamic balance randomization for clinical trials. *Control Clin Trials.* 1991;12:3729-3741.
27. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. *Clin Pharmacol Ther.* 1974;15:443-453.
28. Atkinson AC. Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika.* 1982;69(1):61-67.
29. Atkinson AC. Optimum biased-coin designs for sequential treatment allocation with covariate information. *Stat Med.* 1999;18:1741-1752.
30. Atkinson AC. The comparison of designs for sequential clinical trials with covariate information. *J R Stat Assoc Ser A.* 2002;165:349-373.
31. Atkinson AC. The distribution of loss in two-treatment biased-coin designs. *Biostatistics.* 2003;4:179-193.
32. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics.* 1975;31:103-115.
33. Nishi T, Takaichi A. An extended minimization method to assure similar means of continuous prognostic variables between treatment groups. *Jpn J Biom.* 2003;24:43-55.
34. Ma Z, Hu F. Balancing continuous covariates based on Kernel densities. *Contemp Clin Trials.* 2013;34:262-269.
35. Scott DW. *Multivariate Density Estimation: Theory, Practice, and Visualization.* 2nd ed. Wiley; 2015.
36. Efron B. Forcing a sequential experiment to be balanced. *Biometrika.* 1971;58:403-417.
37. Berger VW, Bour LJ, Carter K, et al. A roadmap to using randomization in clinical trials. *BMC Med Res Methodol.* 2021;21:1-24.
38. Toorawa R, Adena M, Donovan M, Jones S, Conlon J. Use of simulation to compare the performance of minimization with stratified blocked randomization. *Pharm Stat.* 2009;8:264-278.
39. Bertsimas D, Weismantel R. *Optimization Over Integers.* Dynamic Ideas; 2005.
40. Jünger M, Naddef D, Pulleyblank WR, et al. *50 Years of Integer Programming 1958–2008: From the Early Years to the State-of-the-Art.* Springer; 2010.
41. Ben-Tal A, Nemirovski A, Roos C. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM J Optim.* 2003;13:535-560.
42. Székely GJ, Rizzo ML. Testing for equal distributions in high dimension. *InterStat.* 2004;10:1249-1272.
43. Blackwell D, Hodges JL. Design for the control of selection bias. *Ann Math Stat.* 1957;28:449-460.
44. Zhao W, Weng Y, Wu Q, Palesch Y. Quantitative comparison of randomization designs in sequential clinical trials based on treatment balance and allocation randomness. *Pharm Stat.* 2012;11:39-48.
45. Sundahl N, Vandekerckhove G, Decaestecker K, et al. Randomized phase 1 trial of pembrolizumab with sequential versus concomitant stereotactic body radiotherapy in metastatic urothelial carcinoma. *Eur Urol.* 2019;75:701-711.
46. Chiron C, Dumas C, Jambaqué I, Mumford J, Dulac O. Randomized trial comparing vigabatrin and hydrocortisone in infantile spasms due to tuberous sclerosis. *Epilepsy Res.* 1997;26:389-395.
47. Chambers JM, Cleveland WS, Kleiner B, Tukey PA. *Graphical Methods for Data Analysis.* CRC Press; 1983.
48. Garattini S, Bertele V. Ethics in clinical research. *J Hepatol.* 2009;51:792-797.
49. Nardini C. The ethics of clinical trials. *Ecancermedicalscience.* 2014;8:387.
50. Alves MJ, Clímaco J. A review of interactive methods for multiobjective integer and mixed-integer programming. *Eur J Oper Res.* 2007;180:99-115.
51. Özpeynirci Ö, Köksalan M. An exact algorithm for finding extreme supported nondominated points of multiobjective mixed integer programs. *Manag Sci.* 2010;56:2302-2315.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Vazquez AR, Wong W-K. Mathematical programming tools for randomization purposes in small two-arm clinical trials: A case study with real data. *Pharmaceutical Statistics.* 2024;23(6):794-812. doi:10.1002/pst.2388