

# An Optimal Two-Period Multiarm Platform Design with New Experimental Arms Added During the Trial

HAITAO PAN\*, XIAOMENG YUAN<sup>1</sup>, AND JINGJING YE

---

## Abstract

Platform trials are multiarm clinical studies that allow the addition of new experimental arms after the activation of the trial. Statistical issues concerning “adding new arms”, however, have not been thoroughly discussed. This work was motivated by a “two-period” pediatric osteosarcoma study, starting with two experimental arms and one control arm and later adding two more pre-planned experimental arms. The common control arm will be shared among experimental arms across the trial. In this paper, we provide a principled approach, including how to modify the critical boundaries to control the family-wise error rate as new arms are added, how to re-estimate the sample sizes and provide the optimal control-to-experimental arms allocation ratio, in terms of minimizing the total sample size to achieve a desirable marginal power level. We examined the influence of the timing of adding new arms on the design’s operating characteristics, which provides a practical guide for deciding the timing. Other various numerical evaluations have also been conducted. A method for controlling the pair-wise error rate (PWER) has also been developed. We have published an R package, `PlatformDesign`, on CRAN for practitioners to easily implement this platform trial approach. A detailed step-by-step tutorial is provided in Appendix A.2.

KEYWORDS AND PHRASES: Platform trial, Adding new arms, Family-wise error rate, Multiarm trial.

---

## 1. INTRODUCTION

Two-period multiarm platform trials are defined as trials requiring two or more arms during the first period that have the ability to add a new experimental arm(s) during the second period. This paper was motivated by a recent pediatric osteosarcoma study at St. Jude Children’s Research Hospital (St. Jude). This trial includes two planned periods, before and after the addition of new experimental arms. During the first period, the study has two experimental arms and a common control arm. During the second period, two more experimental arms will be added. One reason for adding two additional arms is that not all potential treatments are available at the same time. Details about drug information are concealed because the study is still in the design stage.

The primary endpoint of this osteosarcoma study is progression-free survival; however, for the sake of simplicity, here we use the continuous endpoint to show our proposed methods. To be specific, we use Dunnett’s multiple correction method [7] to control the family-wise error rate (FWER) in the multiarm setting. We also adopt an optimal control-to-experimental arms allocation ratio rule, the *root-K rule* [22], to achieve a targeted marginal power while minimizing the overall sample size of the first period. Despite adding two new experimental arms during the second period, the goal of the design is to have the same targeted

marginal power and FWER as the trial with two experimental arms and one control without adding new arms. How to achieve this will be introduced in this paper.

This type of two-period multiarm trials has been discussed in Ren et al. [16] and Roig et al. [17]. In the former, the authors discussed a simplified version, with one experimental arm in the first period and a second experimental arm added later. Under this framework, Ren et al. presented statistical considerations, including type-I error control and power, as well as an optimal allocation ratio. In Ren et al., the total sample size is fixed and determined by a conventional three-arm design with equal randomization: each experimental arm has a marginal power of  $1 - \beta$  to detect a standardized treatment effect  $\Delta$  and a marginal type-I error controlled at  $\alpha$  (one-sided). Ren et al. discussed the optimal allocation and optimal timing of adding the new arm to maximize the disjunctive power of the study. However, marginal (pair-wise) power is often an interesting metric for possible registrational purposes. In Ren et al.’s method, the marginal power for each experimental arm cannot be maintained at its original level  $1 - \beta$ , mainly due to the fixed total sample size.

In Roig et al. [17], the authors assessed the robustness of model-based approaches to adjust for time trends when utilizing non-concurrent controls. The focus of that research is the consequences of incorporating a nonconcurrent control with various time-trend models and assumptions for

---

\*Corresponding author.

<sup>1</sup>The first two authors contributed equally.

the different arms. In our method, we use only the *concurrent* control data (i.e., patients are recruited and allocated to the control group after a new arm is added) and do not discuss how to use *nonconcurrent* control data (i.e., patients are allocated in the control arm before the new arm is open). Specifically, data from patients allocated to a new experimental arm are *only compared* to that from patients randomized to the control arm *contemporaneously*. For the non-concurrent control discussion, a good resource is the EU-PEARL webinar, “Non-concurrent controls in platform trials” where different multi-stakeholder perspectives on challenges and opportunities for the use of nonconcurrent control data are discussed. In a more general setting of how to leverage information from external or non-concurrent sources to potentially gain power and precision or reduce the sample size, especially based on Bayesian models, people can refer to Normington et al. [14]. It is worth noting that a recent paper [4] describes how to apply the estimands concept when adding new arms.

The two-period multiarm trial is a special case of the platform trial in which a new arm(s) is added only once after the start of a trial. The general platform design is defined as a multiarm multistage (MAMS) trial that adds and removes experimental arms during the trial course. For the general platform design, unlimited times of “adding” are allowed; therefore, a platform trial is also called a perpetual or non-ending trial. For the general platform design, there is a rich body of literature [20, 9, 5, 15, 3].

The rest of the paper is organized as follows. In Section 2, after defining the notations, we introduce the two-period K+M-experimental arm platform design methods. Specifically, we use a 2+2-experimental arm trial to illustrate the design’s components and the developed method in detail. A design to control the pairwise type I error rate (PWER) is also introduced. In Section 3, we briefly showcase how to use the R package `PlatformDesign` to design our motivating pediatric osteosarcoma study and other examples. Comprehensive numerical evaluations are presented in Section 4. Section 5 concludes the paper with a discussion.

## 2. METHODS

We discuss a general format of the two-period K+M-experimental arm platform design. The first period includes  $K$  experimental arms, and during the second period,  $M$  experimental arms will be added.  $K$  and  $M$  can be equal to 1. The second period includes two parts: an overlapping part and a non-overlapping part (see below for details). One common control arm is shared throughout the two periods.

We first introduce a K-experimental arm trial design upon which the K+M-experimental arm trial is based. The K-experimental arm trial we refer to in this paper is equivalent to a traditional K+1-arm trial, which has  $K$  experimental arms and one control arm. The K+M-experimental arm trial is based on the K-experimental arm trial, in the

sense that we will retain the same FWER (or PWER) and marginal power for the K+M-experimental arm trial as in the K-experimental arm trial, despite adding  $M$  new experimental arms during the second period of the K+M-experimental arm trial. Then we describe the proposed methods for the K+M-experimental arm trial, detailing how to add a new experimental arm(s) during the second period of the trial and determine the critical value and allocation ratios.

### 2.1 Design Components for the K-Experimental Arm Trial

In the K-experimental arm trial, we test  $K$  experimental arms against a control arm. We define  $X_{ki}$  as the treatment response of the  $i$ -th patient on arm  $k$  ( $k = 0, 1, \dots, K$ , where  $k = 0$  represents the control arm). We then assume that  $X_{ki} \sim N(\mu_k, \sigma_k^2)$  and the family of  $K$  null hypotheses to be tested is

$$H_{01} : \delta_1 = \mu_1 - \mu_0 \leq 0, \dots, H_{0K} : \delta_K = \mu_K - \mu_0 \leq 0.$$

We use  $\delta_k$  to denote the effect size for each experimental arm  $k$ ,  $k \in \{1, \dots, K\}$ . For simplification, we assume  $\sigma_0 = \sigma_1 = \dots = \sigma_K = \sigma$ , where  $\sigma$  is the common standard deviation. We also denote the standardized effect size using  $\Delta_k = \delta_k / \sigma$ .

To test the hypothesis  $H_{0k}$ ,  $k \in \{1, 2, \dots, K\}$  for experimental arm  $k$  versus control, we assume that a standardized test statistic,  $Z_k$ , is computed as

$$Z_k = \frac{\delta_k}{\sqrt{\text{Var}(\delta_k)}}, k \in \{1, \dots, K\}.$$

Under  $H_{0k}$ , the distribution of the Z-test statistics is standard normal,  $N(0, 1)$ . A marginal (or pair-wise) type-I error rate of level  $\alpha$  can be computed as  $1 - \Phi(z_{1-\alpha})$ , where  $\Phi(\cdot)$  is the standard normal probability distribution function.

If we assume that  $\sigma$  is unknown, then we would use the T-test statistic,  $T_k = \frac{\bar{X}_k - \bar{X}_0}{s\sqrt{1/n+1/n_0}}$ ,  $k \in \{1, \dots, K\}$ . Here,  $\bar{X}$  is the sample mean, and  $s$  is the sample standard deviation. The design parameters  $n$  and  $n_0$  are the numbers of patients enrolled in each experimental arm and the control arm (assuming an equal number of patients is recruited for each experimental arm). Under the null,  $T_k \sim t_{1,v}$  with  $v = n + n_0 - 2$ . We can then compute the marginal type-I error rate using the T distribution. In this paper, we will use the Z-test statistic to introduce the methods.

#### 2.1.1 Error Rate

For a set (or family) of hypotheses, a type-I error is defined as rejecting any true null hypothesis. In this paper, we use the Dunnett’s correction [7] to control the FWER in the strong sense, which means that the probability of rejecting any true null hypothesis is controlled at a pre-specified level for any possible values of  $(\delta_1, \dots, \delta_K)$ . The situation in which PWER (instead of FWER) is controlled is discussed

in Section 2.4. The guidance on multiplicity issues in clinical trials from the regulatory bodies (FDA 2017 and EMA 2012) states that controlling the family-wise type-I error in the strong sense is required for confirmatory trials.

To be explicit, by setting a *global null hypothesis*  $H_0^G$ ,

$$H_0^G : \delta_1 = \dots = \delta_k = \dots = \delta_K = 0,$$

Magirr et al. [13] showed that the FWER is maximized under  $H_0^G$ . Then, the FWER is defined as follows:

$$\text{FWER} = \Pr(\text{reject at least one } H_{0k}, k \in \{1, \dots, K\} | H_0^G) \quad (2.1)$$

Dunnett [7] provided an analytical formula to estimate the FWER when all the comparisons start and conclude at the same time. The FWER can be calculated using

$$\text{FWER}_D = 1 - \Phi_K(z_{1-\alpha_1}, \dots, z_{1-\alpha_1}; \Sigma_1), \quad (2.2)$$

where  $\Phi_K(\cdot; \Sigma_1)$  with  $\Sigma_1 = [\rho_{kk'}]$  is the standard  $K$ -variate normal probability distribution function, and  $\Sigma_1$  is a  $K$ -by- $K$  correlation matrix, with  $\rho_{kk'}$  denoting the correlation between  $Z_k$  and  $Z_{k'}$  at the final analysis. The  $z_{1-\alpha_1}$  is the critical value to control the  $\text{FWER}_D$  during the  $K$ -experimental arm trial; the subscript  $D$  refers to Dunnett's method.

### 2.1.2 Power

Sample sizes can be computed to control several types of power at specified levels. There are multiple definitions of power, depending on the objective of the trial in multiarm settings.

We use  $\omega = 1 - \beta$  to denote the *marginal power* (pair-wise) for a given experimental arm against the control.

The alternative hypothesis for each of the comparisons to be tested is

$$H_{11} : \delta_1 = \delta_1^1 > 0, \dots, H_{1K} : \delta_K = \delta_K^1 > 0.$$

In this paper, we focus on the *global alternative hypothesis*,  $H_1^G$ , which is given by

$$H_1^G : \delta_1 = \dots = \delta_k = \dots = \delta_K = \delta^* (> 0), \quad (2.3)$$

where  $\delta^*$  is the common effect size. Because we assume equal standard deviation (denoted as  $\sigma$ ) for each experimental arm, this is equivalent to  $H_1^G : \Delta_1 = \dots = \Delta_k = \dots = \Delta_K = \Delta (> 0)$ , where  $\Delta$  is the common standardized effect size. Then, we can define the following power based on  $H_1^G$ .

*Disjunctive* (any-pair) power ( $\Omega_{dis}$ ) is the probability of showing a statistically significant effect under the targeted effects for *at least* one comparison

$$\Omega_{dis} = \Pr(\text{reject at least one } H_{0k}, k \in \{1, \dots, K\} | H_1^G) \quad (2.4)$$

Of note, another popular alternative hypothesis is the *least favorable configuration* for experimental  $k \in \{1, \dots, K\}$ , which is given by  $H_1^{\text{LFC}_k} : \delta_k = \delta^*$ ,  $\delta_1 = \dots = \delta_K = 0$ . We will not explore this hypothesis in this paper.

*Conjunctive* (all-pairs) ( $\Omega_c$ ) power is the probability of showing a statistically significant effect under the targeted effects for *all* comparisons. The conjunctive power is computed as

$$\Omega_c = \Pr(\text{reject all } H_{0k} | H_1^G) \quad (2.5)$$

This power is optimistic, and we will not use it in the paper.

### 2.1.3 Optimal Allocation Ratio

In a traditional two-arm randomized clinical trial in which the endpoint measured for both the control and experimental treatments has the same variance, the optimal allocation ratio between the two arms is 1:1, which maximizes the power. However, when there are multiple experimental arms compared to a control arm, the optimal allocation is no longer 1:1. If early stopping was implemented for an experimental arm, then the optimal allocation would be approximately  $\sqrt{K}$  patients (*root-K rule*) allocated to the control group for every patient allocated to a given experimental treatment [21]. Thus, as the number of experimental arms increases, the optimal allocation ratio also increases. The above result applies to the one-stage  $K$ -experimental arm design.

Based on the root- $K$  rule, we have the same allocation ratio ( $A_1 = \sqrt{K}$ ) across all experimental arms in  $K$ -experimental arm trial, thus,  $n_{0_1} = A_1 \times n_1$ ,  $A_1 \in (0, \infty)$ . Here,  $A_1$  is the allocation ratio for the control arm relative to the experimental arm. The design parameters  $n_1$  and  $n_{0_1}$  are the sample sizes of each of the  $K$  experimental arms and the control arm, respectively, during the  $K$ -experimental arm trial. In the first period of the  $K+M$ -experimental arm trial, the same allocation ratio,  $A_1$ , is kept.  $A_1 = n_{0_t}/n_t$ , as  $n_t$  and  $n_{0_t}$  are the sample size of each of the  $K$  experimental arms and the control arm, respectively, during the first period of the  $K+M$ -experimental arm trial, before the  $M$  new arms are added. (Figure 1.) Additionally, The correlation  $\rho_{kk'}$  between  $Z_k$  and  $Z_{k'}$  is  $A_1/(A_1 + 1)$  (For details, see Step 2 of Appendix A.2). If there is an equal allocation to the control and experimental arms, then,  $n_{0_1} = n_1$  and  $\rho_{kk'} = 0.5$ .

Other optimality criteria have also been proposed, including A-optimality, D-optimality, and E-optimality. Details can be found in Atkinson et al. [1]. Of note, for the MAMS design, Wason & Jaki [22] proposed a method to investigate the optimal allocation.

### 2.1.4 Design Summary for the $K$ -Experimental Arm Trial

In the  $K$ -experimental arm trial, there are  $K$  experimental arms and one common control arm. To control the FWER (e.g., at 0.025), equation (2.2) is used to derive the

critical value  $z_{1-\alpha_1}$ . Given the global alternative hypothesis  $H_1^G$  defined in formula (2.3), the required sample sizes for the control and each experimental arm are derived with the allocation ratio based on the root-K rule to obtain a desirable marginal power  $\omega_1$  (See Step 4 in Appendix A.2). The corresponding disjunctive power,  $\Omega_1$ , defined in equation (2.4), is calculated based on  $\omega_1$ , as described in Step 5 in Appendix A.2.

## 2.2 Design Components for the K+M-Experimental Arm Trial

At the end of the first period with  $K$  experimental arms,  $M$  experimental arms are allowed to be added, and the study enters the second period. The second period of the K+M-experimental arm trial has two parts. The first part is an overlapping duration in which  $K$  initial experimental arms and  $M$  new experimental arms overlap, and the second part is a non-overlapping duration in which only the  $M$  new experimental arms are open. Both parts share a common control arm. We use a 2+2-experimental arm trial (depicted in Figure 1) to introduce the notations used in the K+M-experimental arm trial. As its name suggests, the 2+2-experimental arm trial includes a first period in which there are two experimental arms, and a second period in which two new experimental arms are added. This is the setting of the St. Jude pediatric osteosarcoma trial.

In Figure 1, ‘Control’ denotes the common control arm, ‘Trt1’ and ‘Trt2’ denote the two initial experimental arms opened during the first period, and ‘Trt3’ and ‘Trt4’ refer to the two experimental arms opened during the second period of the 2+2-experimental arm trial (the right side of Figure 1).  $A_1$  is the randomization ratio of the control to Trt1 or Trt2 (determined by the root-K rule), and  $n_t$  is the “information time” when Trt3 and Trt4 are added. Specifically, the two arms are added when  $n_t$  patients have been enrolled in each of Trt1 and Trt2. Equivalently, the “information time” can be defined as  $n_{0_t} = \lceil A_1 n_t \rceil$ , the number of patients have been enrolled in the control arm when adding new arms, where  $\lceil \cdot \rceil$  means rounding up to the nearest integer. The information time ( $n_t$  and  $n_{0_t}$ ) should follow the two constraints:  $n_t \leq n_1$  and  $n_{0_t} \leq n_{0_1}$ .

The allocation ratio changes to  $A_2$  once the Trt 3 and 4 are added. During the overlapping stage, there are  $n_2 - n_t$  patients enrolled for each of the experimental arms and  $n_{0_2} - n_{0_t}$  patients enrolled for the control. Therefore,  $A_2 = (n_{0_2} - n_{0_t}) / (n_2 - n_t)$ . Determination of  $n_2$  and  $n_{0_2}$  will be introduced in the next section. After the overlapping stage (i.e., after Trt1 and Trt2 are stopped), Trt3 and Trt4 will continue to enroll until each has reached the required sample size of  $n_2$ . Therefore, both of these arms, need to enroll an additional  $n_t$  patients to “catch up” with Trt1 and Trt2 during the second part of the second period. In the same vein, the control arm will enroll an additional  $n_{0_t}$  patients to ensure the same number of concurrent controls across experimental arms. Therefore, the allocation ratio  $A_3$  is equal

to  $A_1$  after the completion of Trt1 and Trt2. Additionally, we denote the overall allocation ratio as  $A = n_{0_2} / n_2$ .

Because the 2+2-experimental arm trial has four overlapping experimental arms and therefore four test statistics, we can not use the critical value  $z_{1-\alpha_1}$  from the 2-experimental arm trial for the 2+2-experimental arm trial. For example, if we want to control the FWER, the critical value of the K+M-experimental arm trial,  $z_{1-\alpha_2}$ , should be computed based on the correlation matrix of  $K + M$  test statistics using formula (2.2) with  $K$  replaced by  $K + M$ .

## 2.3 Determination of the Optimal Allocation Ratio $A_2$ for the Overlapping Duration in a Two-Period K+M-Arm Trial

We need to first determine the critical value  $z_{1-\alpha_2}$  before we determine the optimal allocation ratio  $A_2$ . To calculate the  $z_{1-\alpha_2}$ , we need to determine the correlation matrix  $\Sigma_2 = [\rho_{kk'}]$  of the K+M-experimental arm trial. This can be derived as follows (see Appendix A.1 for the derivation):

$$\rho_{k,k'} = \frac{n_{0_{kk'}}}{\frac{n_{0_2}^2}{n_2} + n_{0_2}}.$$

Here,  $n_{0_{kk'}}$  is the number of shared controls between experimental arms  $k$  and  $k'$ . By Figure 1, if arms  $k$  and  $k'$  started at the same time, then  $n_{0_{kk'}} = n_{0_2}$ . Otherwise,  $n_{0_{kk'}} = n_{0_2} - n_{0_t}$ .

Once we have the  $\Sigma_2$ , we can use the following equation to find the updated critical value  $z_{1-\alpha_2}$ .

$$FWER = 1 - \int_{-\infty}^{z_{1-\alpha_2}} \int_{-\infty}^{z_{1-\alpha_2}} \dots \int_{-\infty}^{z_{1-\alpha_2}} \pi_Z(Z(z_1, \dots, z_K, z_{K+1}, \dots, z_{K+M}), 0, \Sigma_2) dz_1 dz_2 \dots dz_{K+M} \quad (2.6)$$

Then we can use  $z_{1-\alpha_2}$  to calculate the marginal power,  $\omega_2$ , and the disjunctive power,  $\Omega_2$ , of the K+M-experimental arm trial. (See more details at the end of Section 2.3.1.)

The goal of a two-period K+M-experimental arm platform design is to determine the minimum total sample size (denoted as  $N_2$ ) that can have the marginal power  $\omega_2$  and disjunctive power  $\Omega_2$  that are no less than their counterparts,  $\omega_1$  and  $\Omega_1$ , in the K-experimental arm, while controlling for FWER.

### 2.3.1 Admissible Set for Finding the Optimal Design(s)

It is easy to see  $z_{1-\alpha_2}$  can not be derived without  $n_2$  and  $n_{0_2}$ , as they are needed for computing the correlation matrix  $\Sigma_2$ . We define an admissible set for pairs of  $(n_2, n_{0_2})$  based on the following three constraints. The first two constraints are related to  $A_2$ , the allocation ratio after adding the new arms.

Here we have

$$A_2 = (n_{0_2} - n_{0_t}) / (n_2 - n_t) > 0,$$

where  $n_t$  and  $n_{0_t}$  are the numbers of enrolled patients in each of the experimental arms and the control arm, at the time

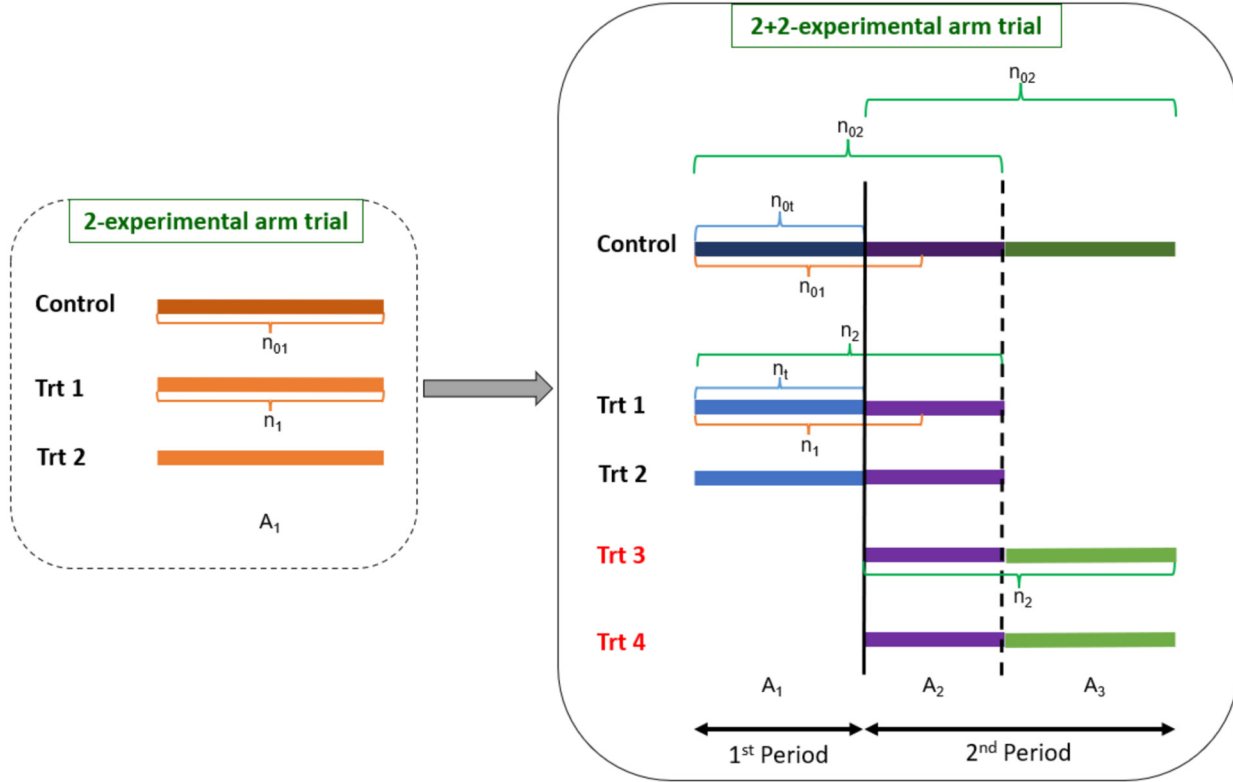


Figure 1: Schema of a two-period 2+2-experimental arm platform trial. The left part of the figure shows a traditional three-arm trial. In the context of this paper, we refer it as a two-experimental arm trial, as it has two experimental arms and one common control. The right part of this figure depicts the 2+2-experimental arm trial. During the first period, this trial has two experimental arms, Trt 1 and Trt 2 (light blue segments), and a control arm (dark blue segment). The vertical solid line separates the first and second periods of the trial and indicates the opening of two new experimental arms, Trt 3 and Trt 4. The dashed vertical line separates two parts of the second period and indicates the closing of Trt 1 and Trt 2. During the first part of the second period, the control arm (dark purple segment) is shared among the four experimental arms (light purple segments). During the second part of the second period, the control (dark green), Trt 3, and Trt 4 (light green) continue to accrue patients until reaching the planned sample sizes. The  $n_t$  and  $n_{0t}$  (blue brackets) indicate the numbers of patients enrolled in Trt 1, Trt 2, and the control, respectively, when Trt 3 and Trt 4 are added. The  $n_1$  and  $n_{01}$  (orange brackets) indicate the sample sizes for each of the two experimental arms and the control, respectively, in the 2-experimental arm trial. The  $n_2$  and  $n_{02}$  (green brackets) indicate sample sizes for each of four experimental arms and the concurrent control.  $A_1$  denotes the allocation ratio (control to experimental arm) during the first period.  $A_2$  denotes the allocation ratio during the first part of the second period, when all four experiments arms are open.  $A_3$  denotes the allocation ratio during the second part of the second period.

of adding the two new arms. The value of  $A_2$  needs to be a non-infinite positive number. In our 2+2 example, if  $n_t = 30$ , then  $n_{0t} = [A_1 * n_t] = 43$  (See Step 8 in Appendix A.2 for details).

Therefore, the first two constraints are

$$n_{02} > n_{0t} = 43,$$

and

$$n_2 > n_t = 30.$$

We also need to set an upper limit for the total sample size of the K+M-experimental arm trial,  $N_2$ . A reasonable

upper limit implies that  $N_2$  should not exceed the required sample sizes (denoted as  $S$ ) for conducting two separate multiarm trials, i.e., a K-experimental arm trial and an M-experimental arm trial.

Based on formulae (A.2) and (A.3),

$$S = \frac{(z_{1-\alpha_1} + z_{1-\beta_1})^2}{\Delta^2} (1 + 2\sqrt{K} + K) + \frac{(z_{1-\alpha_1^*} + z_{1-\beta_1})^2}{\Delta^2} (1 + 2\sqrt{M} + M),$$

where  $z_{1-\alpha_1}$  and  $z_{1-\alpha_1^*}$  are the critical values for the K- and M-experimental arm trials, respectively. Therefore, the

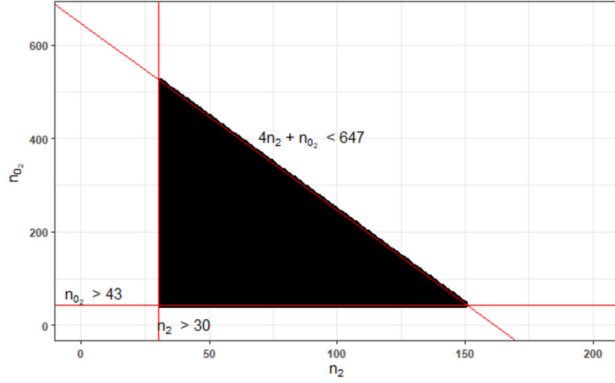


Figure 2: Admissible set of  $(n_2, n_{0_2})$  (shaded triangular region), when  $n_t = 30$  and  $n_{0_t} = 43$  in a two-period 2+2-experimental arm platform trial.

third constraint for  $(n_2, n_{0_2})$  is

$$N_2 = (K + M)n_2 + n_{0_2} + n_{0_t} < S.$$

In our “2+2” example,  $z_{1-\alpha_1} = z_{1-\alpha_1^*}$  as  $K = M = 2$ , and  $S = 2N_1 = 690$ . (See Step 4 in Appendix A.2 for derivation of  $N_1$ .)

Therefore, the third constraint for the “2+2” example is

$$N_2 = 4n_2 + n_{0_2} + n_{0_t} < 690.$$

Under the above three constraints, the admissible set of  $(n_2, n_{0_2})$  can be identified. Given  $n_t = 30$  and  $n_{0_t} = 43$ , we can obtain the feasible region (shaded area in Figure 2). Specifically, all integer pairs  $(n_2, n_{0_2})$  in this region are potential design candidates.

Once we have the feasible region for the pairs of  $(n_2, n_{0_2})$ , then we can compute the correlation matrix  $\Sigma_2$  of the  $K+M$  test statistics  $(Z_1, \dots, Z_K, Z_{K+1}, \dots, Z_{K+M})$ . Derivation of  $\Sigma_2$  in the “2+2” setting is presented in Step 9 of Appendix A.2. With the correlation matrix  $\Sigma_2$ , we can use equation (2.6) to find the marginal type I error  $\alpha_2$  for a specific pair  $(n_2, n_{0_2})$ .

With  $n_1, n_{0_1}, \alpha_1, \beta_1$ , and  $\alpha_2$ , we can use the following equation (2.7) to calculate the marginal power  $\omega_2 = 1 - \beta_2$  for each pair  $(n_2, n_{0_2})$  from  $z_{1-\beta_2}$ .

$$z_{1-\beta_2} = \sqrt{\frac{\frac{1}{n_1} + \frac{1}{n_{0_1}}}{\frac{1}{n_2} + \frac{1}{n_{0_2}}}} (z_{1-\alpha_1} + z_{1-\beta_1}) - z_{1-\alpha_2} \quad (2.7)$$

Next, we can derive the disjunctive power  $\Omega_2$  for each pair of  $(n_2, n_{0_2})$  by plugging  $\beta_2$  and  $\Sigma_2$  into the following equation.

$$\Omega_2 = 1 - \int_{-\infty}^{z_{\beta_2}} \int_{-\infty}^{z_{\beta_2}} \dots \int_{-\infty}^{z_{\beta_2}} \pi_Z(Z(z_1, z_2, \dots, z_{K+M}), 0, \Sigma_2) dz_1 dz_2 \dots dz_{K+M}$$

Based on the above procedure, we can compute the associated  $\omega_2$  and  $\Omega_2$  for all admissible pairs in the feasible region. In the 2+2-experimental arm example, the total number of  $(n_2, n_{0_2})$  pairs in the admissible set is 29,040, and we can compute  $\omega_2$  and  $\Omega_2$  for each of the pairs. Then, we can perform a two-step selection procedure to determine the “optimal” design(s):

1. We keep only the designs in which  $\omega_2 \geq \omega_1$  and  $\Omega_2 \geq \Omega_1$ , respectively (lower limits are decided in Step 7 of Appendix A.2).
2. Among the designs selected, we recommend the one(s) with the smallest sample size ( $N_2$ ) as the “optimal” design(s).

We demonstrate how to design a two-period 2+2-experimental arm platform design by using the R package PlatformDesign in Appendix A.2.

## 2.4 An Optimal K+M-Experimental Arm Design that Controls the PWER

We have described the method for designing a K+M-experimental arm trial to control the FWER, i.e., to control the multiplicity when many interventions are evaluated simultaneously against a common control. However, depending on the reason different treatments are included in the same platform trials, they may not be considered “a family” simply because they are included in the same trial. For master protocols like platform trials, if different experimental arms are included solely for operational efficiency (e.g., reducing the sample size of the control arm by using a shared control to save resources expended during recruitment), we would not necessarily need to perform multiplicity adjustment to control the FWER. Therefore, in this section we introduce an alternative version of the optimal K+M-experimental arm design controlling for the *pair-wise type I error rate* (PWER). The PWER is the probability of incorrectly rejecting the null hypothesis for the primary outcome in a particular experimental arm, regardless of outcomes in the other experimental arms. In this case, the critical value  $z_\alpha$  can be derived directly from the equation below:

$$z_\alpha = \Phi^{-1}(1 - \alpha), \quad (2.8)$$

where  $\alpha$  is a prespecified pair-wise type-I error for each comparison in the trial, which will not be changed due to adding new arms. That is,  $\alpha_1 = \alpha_2 = \alpha$ . Therefore, the main difference between the K+M-experimental arm trial designs controlling for FWER and PWER is that the latter does not use the Dunnett method to derive critical values. Instead, the design derives it directly from Equation 2.8. Notably, the upper limit  $S$  for the total sample size  $N_2$  when controlling for PWER, is constructed using the total sample sizes from the multiarm trials which also control the PWER. Other procedures are similar between the two versions of the design. For more details, see Appendix A.3 for an example to design a “2+2” trial controlling the PWER.

### 3. SOFTWARE EXAMPLE

We developed an R package, `PlatformDesign` for implementing the proposed two-period multiarm platform methods. In this section, we demonstrate our package with three examples of 1) a “2+2” trial with  $n_t = 30$ , 2) a “2+2” trial with  $n_t = 50$  and 3) a “1+3” trial with  $n_t = 30$ .

*Example 1* If no arms are to be added during the course of a study, we can use the functions `one_stage_multiarm()` to compute the sample sizes for the experimental and control arms. For instance, in a study planned to have only two experimental arms and one common control, given a FWER of 0.025 and marginal power of 80% and assuming the expected standardized effective size of 0.4, by using the following code, we found that the sample size for the control is 143, and that for each of the experimental arms is 101. Thus, the planned total sample size in the first period is 345.

```
> library(PlatformDesign)
> res1 <- one_stage_multiarm(K = 2, fwer = 0.025,
marginal.power = 0.8, delta = 0.4)
> res1$n0_1
#[1] 143
> res1$n1
#[1] 101
> res1$N1
#[1] 345
```

However for our pediatric osteosarcoma study, the plan is to add two new experimental arms during the trial. Assuming that the new arms will be added when 30 patients have been enrolled in each experimental arm, and that the study controls the FWER at 0.025 and achieves marginal power at 80%, we can use the following function `platform_design(...)` to find the optimal design(s).

```
> p122 <- platform_design(nt = 30, k = 2, M = 2, fwer = 0.025, marginal.power = 0.8, delta = 0.4)
flag.dmp == 0, lower limits of marginal and disjunctive power are both met
> p122
$design_karm
  K  n1 n0_1  N1  z_alpha FWER1  z_beta1  Power1  cor0 delta
1  2  101  143 345 2.220604 0.025 0.8416212 0.9222971 0.4142136 0.4

$designs
  n2 n0_2 nt n0t  nc  N2  A1  A2  cor1  cor2  critical_value2
15669 107 198 30 43 241 669 1.414214 2.012987 0.3508197 0.2746316 2.4752333
15994 106 202 30 43 245 669 1.414214 2.092105 0.3441558 0.2708949 2.475790
16315 105 206 30 43 249 669 1.414214 2.173333 0.3376206 0.2671464 2.476330
16632 104 210 30 43 253 669 1.414214 2.256757 0.3312102 0.2633910 2.476854

  marginal.power2 disjunctive.power2 FWER2 delta save
15669 0.8001100 0.9853799 0.025 0.4 21
15994 0.8003363 0.9857541 0.025 0.4 21
16315 0.8003878 0.9860900 0.025 0.4 21
16632 0.8002699 0.9863903 0.025 0.4 21

$flag_dp
[1] 0

$flag_mp
[1] 0

$flag_dmp
[1] 0
```

The first part of the outputs (`$design_karm`) contains the parameters for the  $K$ -experimental arm trial. The second part (`$designs`) contains the parameters for the  $K+M$ -experimental arm trial designed based on the former. From above, four designs are recommended (`$designs`), all of which meet the requirements, in terms of controlling the FWER and obtaining power levels equal to or greater than that of the  $K$ -experimental arm trial. If we choose design #16632 (the last row), then the sample sizes for each experimental arm and its corresponding concurrent control in

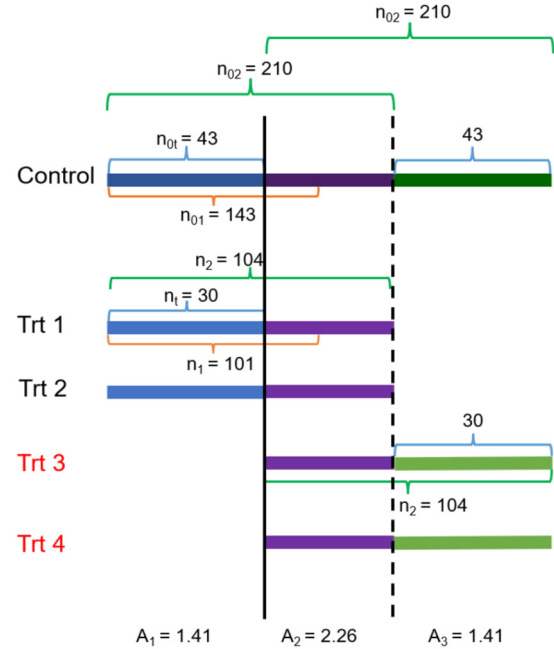


Figure 3: An example of adding two experimental arms to a two-experimental arm trial comparing Trt 1 and Trt2 to control. Key design parameters are shown. The vertical solid line represents when the new experimental arms (Trt3 and Trt 4) are added to the trial. The dashed vertical line represents when the Trt 1 and Trt 2 arms close to accrual. The blue brackets represent the “information time”, when the two new experimental arms are added. They indicate the number of patients enrolled in each of the two initial experimental arms and in the control at the time. The green brackets represent the sample sizes required per experimental arm and the corresponding control in the 2+2-experimental arm trial. The orange brackets indicate the sample sizes for each of the two experimental arms and the control for the 2-experimental arm trial without adding a new arm(s). The optimal allocation ratios ( $A_1$ ,  $A_2$ , and  $A_3$ ) for each period are shown at the bottom of the figure.

the 2+2-experimental arm trial are 104 and 210, respectively. The sample size for the entire control arm (including non-concurrent controls) is 253. Using this design,  $A_2$ , the allocation ratio in the first part of the second period is 2.26, control-to-experimental arm. Other parameters of this design are shown in Figure 3.

Once we decide  $n_2 = 104$  and  $n_{0_2} = 210$ , the sample sizes for each of the experimental arms (Trt 1 to Trt 4) and the control arm in the first part of the second period are  $104 - 30 = 74$  and  $210 - 43 = 167$ , respectively. Accordingly, we can determine the sample sizes of Trt 3, Trt 4 and the control arm for the second part of the second period to be  $104 - 74 = 30$  and  $210 - 167 = 43$ , respectively. The optimal allocation for the second part of the second period  $A_3$  can be computed accordingly as  $A_3 = 43/30 = 1.41$ .

*Example 2* With the constraints on sample sizes described in Section 2.3.1, the optimal design(s) may not exist when  $n_t$ , i.e., the timing of adding new arm(s) is relatively “late”. For instance, in the above “2+2” example, if  $n_t = 50$ , no optimal design is identified if a marginal power of 80% needs to be maintained, as shown in the following code.

The `platform_design(.)` function returns *criteria indicators* (i.e., `flag.dp`, `flag.mp`, and `flag.dpmp`) to show if any optimal design exists, given FWER, marginal power, the timing of adding a new arm(s), and the number of experimental arms for each period in a K+M-experimental arm trial. If `flag.dpmp = 0`, the optimal design can maintain both marginal and disjunctive power levels no less than those in the K-experimental arm trial. Otherwise, the algorithm will check if we can find a design(s) that maintains either the marginal or disjunctive power. When  $n_t = 50$ , `flag.dpmp = 1`, and `flag.dp = 0`, it indicates that we can only find designs that keep the disjunctive power no less than its counterpart in the K-experimental arm trial. However, the marginal power in the designs found is less than 80%. The accompanying `warning message` conveys the same information.

```
> p122_nt50 <- platform_design(nt = 50, K = 2, M = 2, fwr = 0.025, marginal.power = 0.8, delta = 0.4)
Warning message:
In platform_design(nt = 50, K = 2, M = 2, fwr = 0.025, marginal.power = 0.8, :
  flag.mp == 1, lower limit of the marginal power can not be met, recommended design(s) selected from
  designs with disjunctive power >= Power1
> p122_nt50
$design_Karm
  K n1 n0_1 N1 z_alpha1 FWER1 z_beta1 Power1 cor0 delta
1 2 101 143 345 2.220604 0.025 0.8416212 0.9222971 0.4142136 0.4

$designs
  n2 n0_2 nt n0t nc N2 A1 A2 cor1 cor2 critical_value2 marginal.power2
5525 64 143 50 71 214 470 1.414214 5.142857 0.3091787 0.1556704 2.483325 0.5648199
5797 63 147 50 71 218 470 1.414214 5.846154 0.3000000 0.1551020 2.483719 0.5633370
6065 62 151 50 71 222 470 1.414214 6.666667 0.2910798 0.1542145 2.484102 0.5614494
  disjunctive.power2 FWER2 delta save
5525 0.9223835 0.025 0.4 220
5797 0.9224524 0.025 0.4 220
6065 0.9223364 0.025 0.4 220

$flag.dp
[1] 0

$flag.mp
[1] 1

$flag.dpmp
[1] 1
```

*Example 3* The PlatformDesign package can be used to design any K+M-experimental arm platform trial with K and M as positive integers. Here, we show a hypothetical example for designing a 1+3-experimental arm platform trial using this R package.

```
> p113 <- platform_design(nt = 30, K = 1, M = 3, fwr = 0.025, marginal.power = 0.8, delta = 0.4)
flag.dpmp == 0, lower limits of marginal and disjunctive power are both met
> p113
$design_Karm
  K n1 n0_1 N1 z_alpha1 FWER1 z_beta1 Power1 cor0 delta
1 1 99 99 198 1.959964 0.025 0.8416212 0.8 0.5 0.4

$designs
  n2 n0_2 nt n0t nc N2 A1 A2 cor1 cor2 critical_value2 marginal.power2
17916 105 204 30 30 234 654 1.2 32 0.3398058 0.2898344 2.473582 0.800046
  disjunctive.power2 FWER2 delta save
17916 0.9843551 0.025 0.4 27

$flag.dp
[1] 0

$flag.mp
[1] 0

$flag.dpmp
[1] 0
```

Based on the output above, one design with  $N_2 = 654$  is recommended. In addition, we can see that all the criteria indicators are equal to zero, implying that the criteria for both marginal and disjunctive power levels have been met.

More explanations of the above results and step-by-step instructions for using this package can be found in Appendix A.2. Details of how to use `Platform_Design(.)` and other functions can also be found in documents and vignettes of the R package PlatformDesign.

## 4. NUMERICAL EVALUATIONS

Unless otherwise specified, all numerical evaluations are conducted in the setting of a two-period 2+2-experimental arm trial (except for Figure 14 in Section 4.4), which controls the FWER at 0.025 and achieving a marginal power of 0.8, given the standardized effect size  $\Delta$  is 0.4. Therefore, the disjunctive power exceeds 0.922. In this section, we will examine the relations among various design parameters in the 2+2-experimental arm trial.

### 4.1 Correlations

We explored the relations between the correlations of Z-test statistics and the disjunctive power  $\Omega_2$  in the 2+2-experimental arm trial. Specifically, during the second period, two types of correlations occur. We denote  $\rho_1$  as the correlation between any pair of experimental arms that start at the same time, and  $\rho_2$  as the correlation between any pair of experimental arms that start at different times.

In the “2+2” example, the change in disjunctive power  $\Omega_2$  is driven by  $\rho_2$  (Figure 5) instead of  $\rho_1$  (Figure 4). The disjunctive power ( $\Omega_2$ ) decreases as  $\rho_2$  increases (Figure 5). Given a specified marginal power, an optimal design(s) may not exist if the timing of adding new arms is relatively late (i.e., the value of  $n_t$  is large). Therefore, in Figures 4 and 5, for  $n_t = 50, 60, 70$ , or 80, the lower limit of marginal power  $\Omega_2$  is 75%. We must choose a lower limit for  $\omega_2$  when  $n_t > 40$  to ensure that  $N_2 < S$ . More detailed reasoning for this is provided in Appendix A.2.

### 4.2 Influence of the Timing of Adding New Arms

To design a platform trial, we must know how the timing of adding a new arm(s) affects the design’s properties. Here we examine the relations between the timing of adding new arms (i.e., “information time”,  $n_t$ ) and various design parameters in the K+M-experimental arm trial (e.g., the total required sample size  $N_2$ , the disjunctive power  $\Omega_2$ , and the marginal type-I error rate  $\alpha_2$ ) using a “2+2” example.

As shown in Figure 6, the total sample size  $N_2$  increases with increased information time  $n_t$ . Thus, the earlier the timing of adding new arms, the more patients can be saved by conducting a 2+2-experimental arm trial compared to two separate 2-experimental arm trials (shown as a red line in Figure 6). For instance, if two experimental arms are added to the trial when  $n_t = 30$ , the total required sample size is 669. This means that 21 fewer patients are needed, keeping the FWER at 0.025 and marginal power of 80%, and assuming a standardized effect size of 0.4.



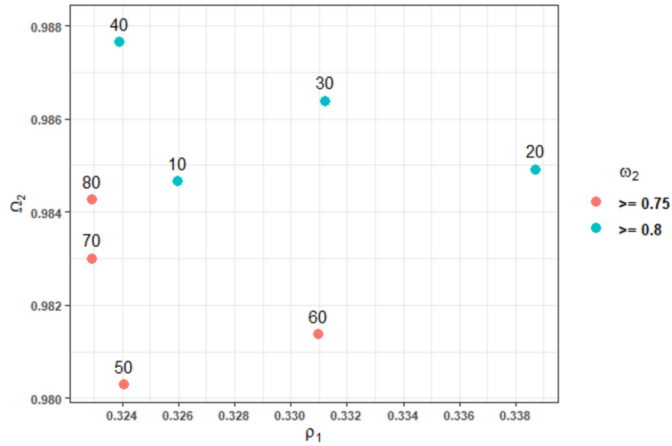


Figure 4: Relations between the correlation  $\rho_1$  and disjunctive power  $\Omega_2$  in a two-period 2+2 platform trial setting, given the FWER of 0.025, a disjunctive power level represented by  $\Omega_2 \geq 0.922$ , the marginal power represented  $\omega_2 \geq 0.8$  (for  $n_t = 10, 20, 30,$  or  $40$ ; blue dots) or  $\geq 0.75$  (for  $n_t = 50, 60, 70,$  or  $80$ ; red dots), and the standardized effect size  $\Delta = 0.4$ . The value associated with each dot is the corresponding  $n_t$  value.

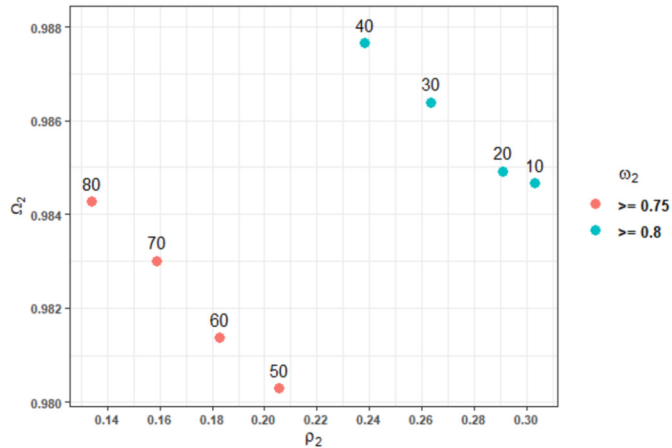


Figure 5: Relations between the correlation  $\rho_2$  and disjunctive power  $\Omega_2$  in a two-period 2+2 platform trial setting, given the FWER of 0.025, disjunctive power  $\Omega_2 \geq 0.922$ , the marginal power  $\omega_2 \geq 0.8$  (for  $n_t = 10, 20, 30,$  or  $40$ ; blue dots) or  $\geq 0.75$  (for  $n_t = 50, 60, 70,$  or  $80$ ; red dots), and the standardized effect size  $\Delta = 0.4$ . The value associated with each dot is the corresponding  $n_t$  value.

Figure 7 suggests that the disjunctive power  $\Omega_2$  also increases with the delay of adding new arms in the “2+2” scenario. This is expected, as the delay decreases the correlations between any pair of arms starting at different times ( $\rho_2$ ), caused by a briefer overlapping period. Therefore, experimental arms become more independent, which increases  $\Omega_2$ .

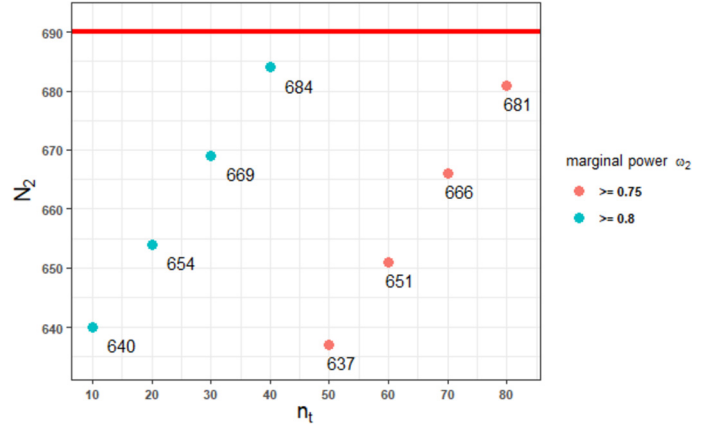


Figure 6: The timing of adding new arms ( $n_t$ ) affects the total sample size  $N_2$ , given the FWER of 0.025, disjunctive power  $\Omega_2 \geq 0.922$ , and the marginal power  $\omega_2 \geq 0.8$  (for  $n_t = 10, 20, 30,$  or  $40$ ; blue dots) or  $\geq 0.75$  (for  $n_t = 50, 60, 70,$  or  $80$ ; red dots), and the standardized effect size  $\Delta = 0.4$ . The value associated with each dot is the corresponding  $n_t$  value. The red line indicates the total sample size needed for conducting two separate 2-experimental arm trials. The values associated with each dot is the corresponding  $N_2$  value.

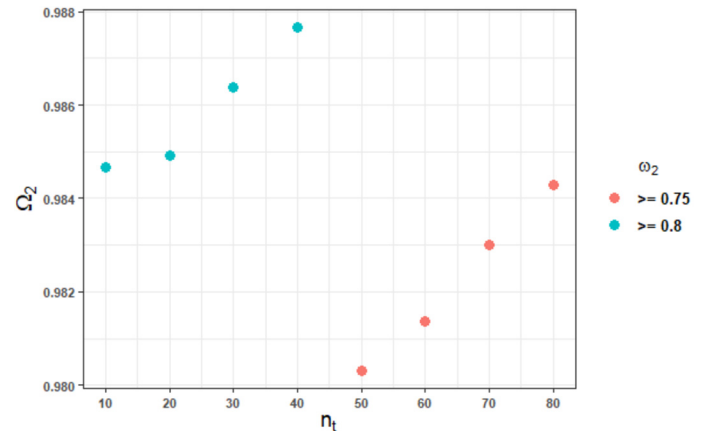


Figure 7: The timing of adding new arms ( $n_t$ ) affects the disjunctive power ( $\Omega_2$ ). Given FWER = 0.025, the marginal power  $\omega_2 \geq 0.8$  (for  $n_t = 10, 20, 30,$  or  $40$ ; blue dots) or  $\geq 0.75$  (for  $n_t = 50, 60, 70,$  or  $80$ ; red dots), and the standardized effect size  $\Delta = 0.4$ .

We also examined the relations between the marginal type-I error rate ( $\alpha_2$ ) and the timing of adding new arms ( $n_t$ ) in the “2+2” example. The marginal type-I error rate  $\alpha_2$  decreases when  $n_t$  increases (Figure 8), though this change is negligible (range of  $\alpha_2$ , 0.00650 to 0.00665). This finding indicates that the timing of adding new experimental arms to an existing platform protocol has a minimal impact on the marginal type-I error rate.

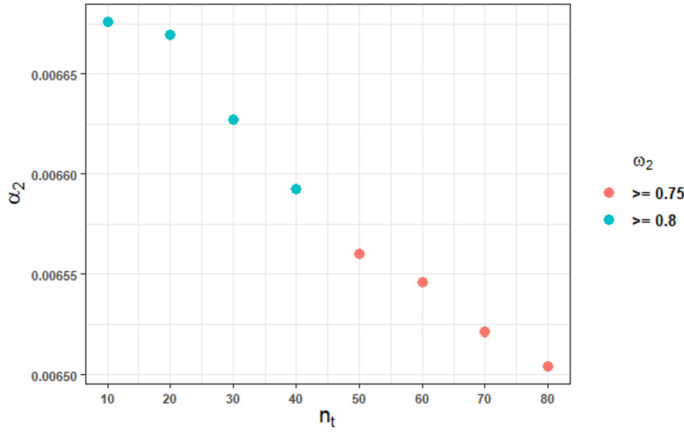


Figure 8: The timing of adding new arms ( $n_t$ ) has a minimal impact on the marginal type-I error rate ( $\alpha_2$ ) in the “2+2” example.

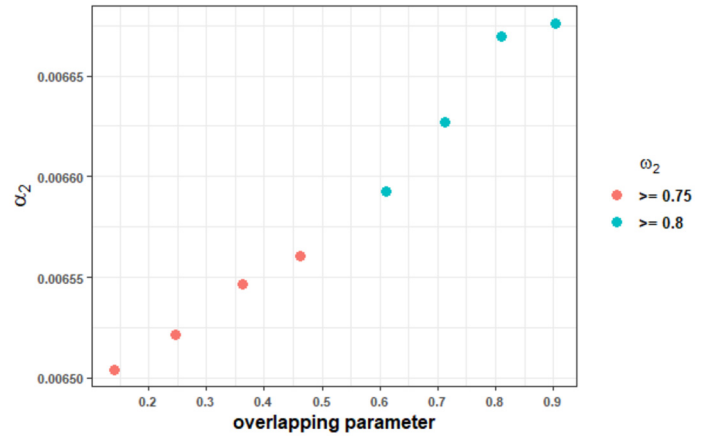


Figure 10: The relation between the overlapping parameter and the marginal type-I error  $\alpha_2$ .

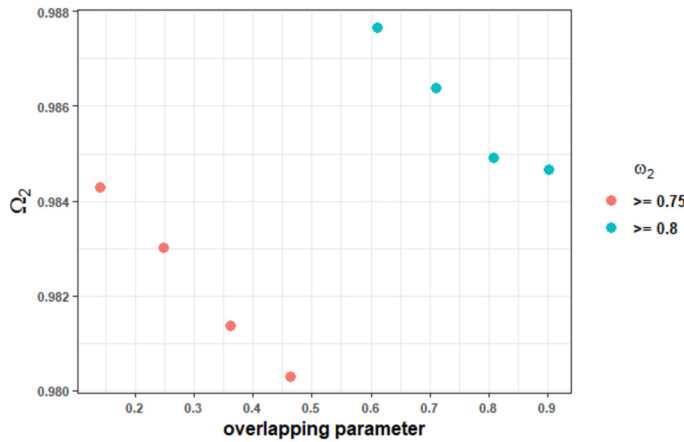


Figure 9: Relations between the overlapping parameter and the disjunctive power  $\Omega_2$ .

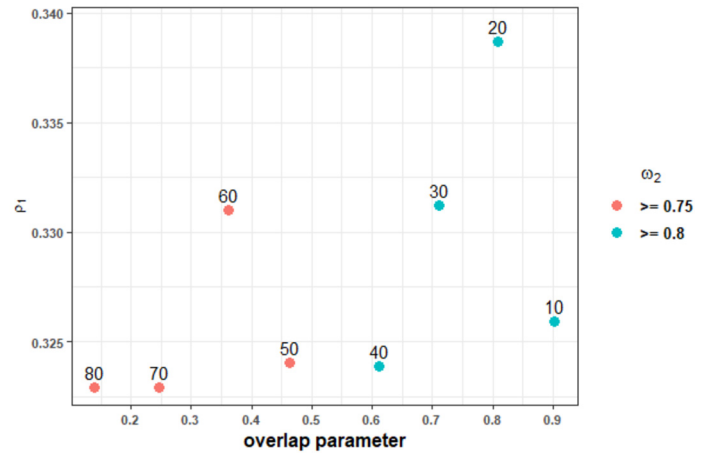


Figure 11: The relation between the correlation  $\rho_1$  and overlapping parameter. The value associated with each dot is the value of  $n_t$ .

### 4.3 Overlapping Parameter

We define an overlapping parameter as  $\frac{n_2 - n_t}{n_2}$ , which represents the percentage of patients in an experimental arm who are enrolled during the overlapping stage. We explored the relations between the overlapping parameter and various design parameters, including the disjunctive power  $\Omega_2$  and the marginal type-I error rate  $\alpha_2$  in the “2+2” scenario.

As illustrated in Figure 9, the disjunctive power  $\Omega_2$  decreases as the overlapping parameter increases. This is the opposite of what we observed for the relation between  $n_t$  and  $\Omega_2$ .

Unlike the relation with  $\Omega_2$ , the marginal type I error  $\alpha_2$  increases with the increase in the overlapping parameter (Figure 10).

We also explore the relation between the overlapping parameter and test statistics correlations for two types of correlations. In Figure 11, there is no obvious trend between

the overlapping parameter and  $\rho_1$ , but in Figure 12, there is a positive trend between the overlapping parameter and  $\rho_2$ .

### 4.4 Optimal Overall Allocation Ratio $A$

The overall allocation ratio (defined as  $A = n_{0_2}/n_2$ ) stays very close to the value of  $\sqrt{4} = 2$  with various  $n_t$  (Figure 13). The value of  $A$  ranges from 1.95 to slightly less than 2.10.

Given the timing of adding new arms at  $n_t = 30$  and choosing only the optimal design with the largest disjunctive power, we explored how the overall allocation ratio  $A$  changes with varied  $K = 1, 2, 3, 4$ , or 5 and  $M = 1, 2, 3, 4$ , or 5. From Figure 14, given the same  $M$ , the overall allocation ratio  $A$  increases if  $K$  increases. Given the same  $K$ ,  $A$  increases as  $M$  increases.

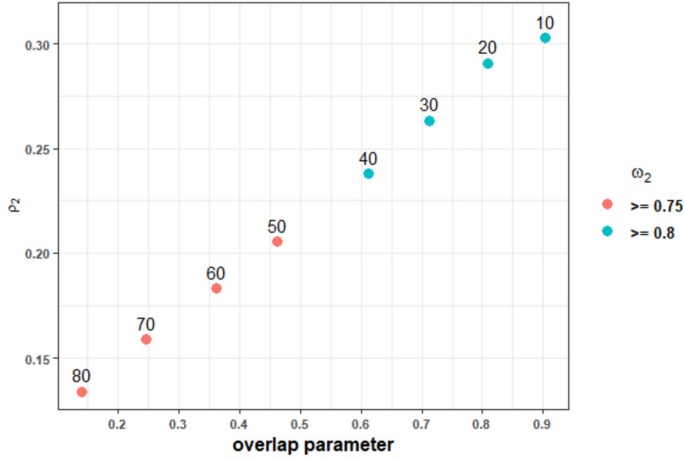


Figure 12: The relation between the correlation  $\rho_2$  and overlapping parameter. The value associated with each dot is the value of  $n_t$ .

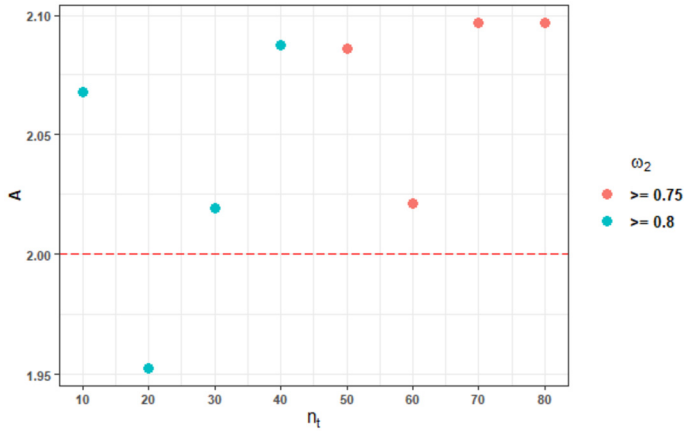


Figure 13: The relation between the overall allocation ratio  $A = n_{0_2}/n_2$  and the timing of adding new arms ( $n_t$ ) in a two-period 2+2-experimental arm platform trial. The red dashed line represents the optimal allocation ratio used in the first period, based on the root-K method.

### 5. CONCLUSION

The popularity of platform trials has increased in recent years. However, due to the complexity of such trial designs, many design-related questions remain, and the use of platform trials is still limited, especially in the confirmatory late-phase setting. To facilitate the use of platform trials, we propose an optimal design for two-period multiarm platform trials, in terms of minimizing the total sample size to control the FWER or PWER. Instead of adding new arms without end, this type of trial considers two periods, before and after new experimental arms are added. Each period can have one or more experimental arms, and a common control arm is shared by both periods. A two-period multiarm platform trial is usually very useful in the setting of a

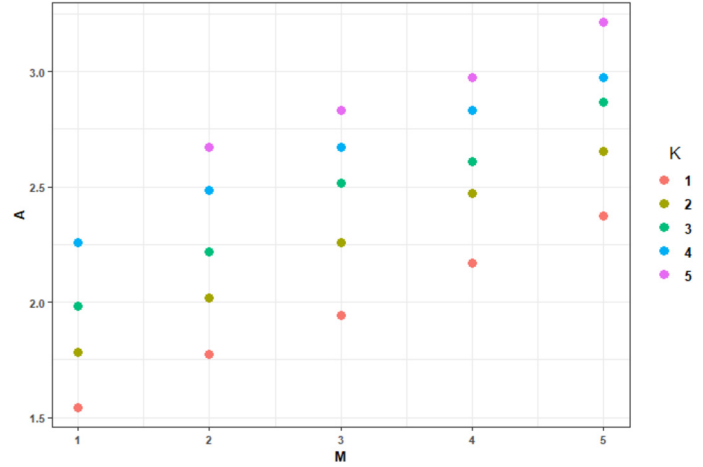


Figure 14: Relations between the overall allocation ratio  $A = n_{0_2}/n_2$  and the numbers of experimental arms initially opened ( $K = 1, 2, 3, 4,$  or  $5$ ) and added later ( $M = 1, 2, 3, 4,$  or  $5$ ), given  $n_t = 30$ ,  $FWER = 0.025$ , disjunctive power  $\Omega_2 \geq 0.922$ , the marginal power  $\omega_2 \geq 0.8$ , and the standardized effect size  $\Delta = 0.4$ . The results are shown for each combination of  $K$  and  $M$ , where the optimal design with the greatest  $\Omega_2$  is presented. In the “5+1” scenario, the optimal design does not exist due to sample size constraints and the prespecified goal for the marginal power to be at least 80%.

single institution and is a special type of MAMS platform.

In this paper, to meet registrational purposes, we systematically described how to control the FWER or PWER when adding new arms, re-estimate the sample size to achieve the desirable power, and determine the optimal allocation ratio. Numerical evaluations were conducted to comprehensively examine the properties of the proposed design. The advantage of this design over conducting separate multiarm trials is that we can reduce the sample size and use a shared infrastructure. We also provide a step-by-step tutorial in Appendix A.2, that demonstrates how to use the R package PlatformDesign.

In this paper, we considered conducting the main analyses using only the concurrent controls. Because osteosarcoma is a relatively rare disease, patient accrual can take a long time. Therefore, we need to be careful about the potential changes in treatment effect over time. For clinical studies with relatively faster accrual rates, the difference between the two periods may not be substantial. In those cases, a nonconcurrent control may still be used. Nevertheless, in the design of our pediatric osteosarcoma study, we plan to include all control arm data for sensitivity analyses (i.e., to increase the estimation precision and power). There are three rationales for using a pooled control arm: (1) Pediatric osteosarcoma is a rare disease, so patients are scarce. (2) If the timing of adding new arms is early and the medical landscape is stable, there is little concern about any potential shift in the treatment effect over time. (3) The

nonconcurrent control is essentially part of the control arm. Those patients are enrolled in the same study, screened with the same inclusion/exclusion criteria, and participate at the same institution, just like the concurrent controls. How to use nonconcurrent controls has been described in many papers [15, 6, 19, 18].

We also examined the timing of adding new arms in platform trials because practical guidance about deciding the timing of adding and closing arms will help increase the uptake of this approach. However, we have focused primarily on the statistical aspects of adding arms. The optimal timing of adding (or closing) arms in platform trials depends on the clinical context, the nature of the interventions, and the capability of stakeholders to deliver amendments [12]. It should also be noted that whether to add a new treatment arm to a multiarm study in the two-period setting (called two-stage in [11]) has been discussed based on a decision-theoretic framework.

Future works will involve extending the current method to a multiperiod, multiarm, multistage setting. The design will include more than two periods and will be not only multiarm but also multistage, allowing early closing of arms or graduating during the interim analyses. In this paper, we do not consider the presence of time trends, which is an effect of a treatment (either an experimental or control treatment) that may vary with time as the study period of platform trials is often longer than that of fixed trials. This happens, for example, when there is a learning curve amongst the study personnel or when standard care changes over time. In the future, we may use models that incorporate time trends in the proposed framework. We also may study how to use nonconcurrent control arm data when time trends are considered.

## APPENDIX A

### A.1 Derivation of the Correlation Between Test Statistics

For a comparison between the experimental arm  $k$  ( $k = 1, 2, 3, \dots$ ) and the control arm, we have the test statistics

$$Z_k = \frac{\bar{X}_k - \bar{X}_{0_k}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{n_0}}}$$

where  $n$  refers to the number of patients in each experimental arm and  $n_0$  refers to that in its control.  $\bar{X}_k$  and  $\bar{X}_{0_k}$  are sample means for experimental arm  $k$  and its corresponding control, respectively. We assume standard deviation  $\sigma$  is equal for all arms.

The correlation between the test statistics of experimental arms  $k$  and  $k'$  is

$$\rho_{k,k'} = Cov\left(\frac{\bar{X}_k - \bar{X}_{0_k}}{\sqrt{\frac{1}{n} + \frac{1}{n_0}}}, \frac{\bar{X}_{k'} - \bar{X}_{0_{k'}}}{\sqrt{\frac{1}{n} + \frac{1}{n_0}}}\right) / \sigma^2$$

where  $\bar{X}_k$  ( $\bar{X}_{k'}$ ) and  $\bar{X}_{0_k}$  ( $\bar{X}_{0_{k'}}$ ) are sample means of experimental arm  $k$  ( $k'$ ) and its associated control, respectively. Although for both arms  $k$  and  $k'$ , the control sample size is  $n_0$  patients, the patients enrolled are not necessary the same individuals. Therefore, for arms  $k$  and  $k'$ , we use  $\bar{X}_{0_k}$  and  $\bar{X}_{0_{k'}}$ , respectively, to denote their sample means.

Let

$$\kappa = \frac{1}{\frac{1}{n} + \frac{1}{n_0}}$$

Then,

$$\begin{aligned} Cov\left(\frac{\bar{X}_k - \bar{X}_{0_k}}{\sqrt{\frac{1}{n} + \frac{1}{n_0}}}, \frac{\bar{X}_{k'} - \bar{X}_{0_{k'}}}{\sqrt{\frac{1}{n} + \frac{1}{n_0}}}\right) &= \kappa Cov(\bar{X}_{0_k}, \bar{X}_{0_{k'}}) \\ &= \kappa Cov\left(\frac{\sum_{i=1}^{n_0} X_{0_{ki}}}{n_0}, \frac{\sum_{j=1}^{n_0} X_{0_{k'j}}}{n_0}\right) \\ &= \kappa \frac{1}{n_0^2} Cov\left(\sum_{i=1}^{n_0} X_{0_{ki}}, \sum_{j=1}^{n_0} X_{0_{k'j}}\right) \\ &= \kappa \frac{1}{n_0^2} n_{0_{kk'}} \sigma^2 \end{aligned}$$

where  $n_{0_{kk'}}$  is the number of the shared controls between the experimental arms  $k$  and  $k'$ . The first equality satisfies because  $\bar{X}_k$  and  $\bar{X}_{k'}$  are independent.  $\bar{X}_k$  and  $\bar{X}_{0_{k'}}$  ( $\bar{X}_{k'}$  and  $\bar{X}_{0_k}$ ) are also independent, while  $\bar{X}_{0_k}$  and  $\bar{X}_{0_{k'}}$  share the overlapping control data.

Therefore,

$$\rho_{k,k'} = \frac{1}{n_0^2} n_{0_{kk'}} = \frac{1}{\frac{1}{n} + \frac{1}{n_0}} \frac{n_{0_{kk'}}}{n_0^2} = \frac{n_{0_{kk'}}}{\frac{n_0^2}{n} + n_0} \quad (A.1)$$

### A.2 Step-by-Step Explanation of the Proposed Design Using the R Package PlatformDesign

The following steps contain two parts: 1) Steps 1 to 5 derive the design parameters in the K-experimental arm trial upon which the K+M-experimental arm trial is based. 2) Steps 6 to 14 illustrate how the design parameters are calculated for the K+M-experimental arm trial, to control the FWER and marginal power at their pre-specified levels.

*Step 1: Initial Setup* Four design parameters for the K-experimental arm trial should be pre-specified: the number of experimental arms ( $K$ ), the family-wise error rate ( $FWER_1$ ), the marginal type-II error ( $\beta_1$ ), and the allocation ratio (control-to-each experimental arm, denoted as  $A_1$ ). In our method, we use  $A_1 = \sqrt{K}$ , according to the K-root optimal allocation rule. In the following code, we assume  $K = 2$ ,  $FWER_1 = 0.025$ ,  $\beta_1 = 0.2$ , and  $A_1 = \sqrt{2}$ . In addition,  $z_{\beta_1}$  ( $z\_beta1$  in the following code) is the corresponding critical value for the power of  $1 - \beta_1$ .

```

K <- 2
FWER_1 <- 0.025
beta1 <- 0.2
z_beta1 <- qnorm(1-beta1) #z_(1-beta1)
A1 <- sqrt(K)

```

*Step 2: Correlation Matrix 1* We use  $Z_1$  and  $Z_2$  to denote the two test statistics for comparing each experimental arm to the control in the K-experimental arm trial. Given  $A_1$ , the correlation between  $Z_1$  and  $Z_2$  (denoted as  $\rho_0$ ) and the correlation matrix (denoted as  $\Sigma_1$ ) can be calculated as below.

First, by plugging  $n_1$  and  $n_{0_1}$  into formula (A.1), we have

$$\rho_0 = \frac{n_{0_{kk'}}}{\frac{(n_{0_1})^2}{n_1} + n_{0_1}}$$

where  $n_1$  is the number of patients in each experimental arm, and  $n_{0_1}$  is the number of patients in the control arm in the K-experimental arm trial.

Because the two experimental arms share a common control arm and  $n_{0_{kk'}} = n_{0_1}$ , the correlation of  $Z_1$  and  $Z_2$  can be computed as

$$\rho_0 = \frac{1}{(n_{0_1}/n_1 + 1)} = \frac{1}{(A_1 + 1)}$$

In our “2+2”-experimental arm example, where  $K = 2$ , we have

$$\Sigma_1 = \begin{bmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.4142 \\ 0.4142 & 1 \end{bmatrix}$$

Based on the above derivations, the function `one_stage_multiarm(.)` can be used to find  $\rho_0$  and the correlation matrix  $\Sigma_1$  as shown below.

```

multi <- one_stage_multiarm(K = 2, fwer = 0.025,
marginal.power = 0.8, delta = 0.4)
corMat1 <- multi$corMat1
corMat1
##           [,1]           [,2]
## [1,] 1.0000000 0.4142136
## [2,] 0.4142136 1.0000000

```

*Step 3: Critical Value 1* Given  $K$ ,  $\Sigma_1$ , and  $FWER_1$ , we can determine the associated critical value (denoted as  $z_{1-\alpha_1}$ ) for the marginal type-I error rate in the two-experimental arm trial (denoted as  $\alpha_1$ ) based on the following equation:

$$FWER_1 = 1 - \int_{-\infty}^{z_{1-\alpha_1}} \int_{-\infty}^{z_{1-\alpha_1}} \dots \int_{-\infty}^{z_{1-\alpha_1}} \pi_Z(Z(z_1, z_2, \dots, z_K), 0, \Sigma_1) dz_1 dz_2 \dots dz_K.$$

This calculation can also be achieved using the function `one_stage_multiarm`.

```

multi$z_alpha1
# [1] 2.220604

```

*Step 4: Sample Sizes 1* Given  $\beta_1$ ,  $A_1 = \sqrt{K}$ , an effective standardized effect size  $\Delta$  (assumed to be 0.4), and  $z_{1-\alpha_1}$  derived from the above **Step 3** (`$z_alpha1`), we can derive the required sample sizes for the experimental ( $n_1$ ) and control arms  $n_{0_1} (= A_1 n_1 = \sqrt{K} n_1)$ , respectively, as shown below.

We have

$$z_{1-\alpha_1} + z_{1-\beta_1} = \frac{\mu_i - \mu_0}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{\sqrt{K} n_1}}} = \frac{\Delta}{\sqrt{\frac{1}{n_1} + \frac{1}{\sqrt{K} n_1}}}$$

therefore,

$$n_1 = \frac{(z_{\alpha_1} + z_{\beta_1})^2}{\Delta^2} \left(1 + \frac{\sqrt{K}}{K}\right) \quad (\text{A.2})$$

and

$$n_{0_1} = \frac{(z_{\alpha_1} + z_{\beta_1})^2}{\Delta^2} (\sqrt{K} + 1) \quad (\text{A.3})$$

Thus, the total sample size of the K-experimental arm trial is

$$N_1 = K n_1 + n_{0_1}$$

We can use the function `one_stage_multiarm(.)` to determine the required sample sizes for the two-experimental arm trial. Based on the outputs below, 101 patients are needed for each experimental arm, and 143 patients are needed for the control arm. The total sample size is 345 patients.

```

multi <- one_stage_multiarm(K = 2, fwer = 0.025,
marginal.power = 0.8, delta = 0.4)
multi$n1
multi$n0_1
multi$N1
# $n1
# [1] 101
# $n0_1
# [1] 143
# $N1
# [1] 345

```

*Step 5: Disjunctive Power 1* Given  $\beta_1$  and  $\Sigma_1$  computed above, we can derive the disjunctive power,  $\Omega_1$ , based on the following equation:

$$\Omega_1 = 1 - \int_{-\infty}^{z_{\beta_1}} \int_{-\infty}^{z_{\beta_1}} \dots \int_{-\infty}^{z_{\beta_1}} \pi_Z(Z(z_1, z_2, \dots, z_K), 0, \Sigma_1) dz_1 dz_2 \dots dz_K.$$

This result is also included as part of the output from the function `one_stage_multiarm(.)`, i.e., `$Power1`. Here, the computed disjunctive power is 0.922.

```

multi$Power1
# [1] 0.9222971

```

From Steps 1 to 5, we demonstrated how to derive the marginal type-I error rate ( $\alpha_1$ ), the sample size for each experimental arm ( $n_1$ ), the sample size for the control arm ( $n_{0_1}$ ), and the disjunctive power ( $\Omega_1$ ) in the K-experimental arm trial, given  $K$ ,  $FWER_1$ , marginal power  $1 - \beta_1$ , and the standardized effect size  $\Delta$ .

In summary, the function `one_stage_multiarm(.)` in R package `PlatformDesign` can complete steps 1 to 5 at once. Below are the outputs generated by this function.

```
one_stage_multiarm(K = 2, fwer = 0.025,
  marginal.power = 0.8, delta = 0.4)
# $n1
# [1] 101
# $n0_1
# [1] 143
# $N1
# [1] 345
# $z_alpha1
# [1] 2.220604
# $z_beta1
# [1] 0.8416212
# $Power1
# [1] 0.9222971
# $corMat1
#           [,1] [,2]
# [1,] 1.0000000 0.4142136
# [2,] 0.4142136 1.0000000
```

On the basis of the above steps, we now introduce our proposed methods for the K+M-experimental arm trial in Steps 6 to 14.

*Step 6: Timing of Adding New Arms* Timing is the first component to consider when planning to add new experimental arms to a platform trial. In this paper, we use the number of patients enrolled in each of the experimental arms ( $n_t$ ) at the time new arms are added to define the timing. By this definition, the number of patients enrolled in the control arm when new arms are added is  $n_{0_t} = [A_1 n_t]$ .

The code below follows a scenario in which 30 patients have enrolled in each experimental arm when new arms are added.

```
nt <- 30
nt
## [1] 30
n_0t <- ceiling(nt*A1)
n_0t
## [1] 43
```

*Step 7: Initial Setup 2* Then we need to decide the family-wise error rate in the K+M-experimental arm trial (denoted as  $FWER_2$ ). In this paper, we control the  $FWER_2$  at the same level as the  $FWER_1$ . With  $FWER_2$ , we can calculate the marginal type-I error rate (denoted as  $\alpha_2$ , which is always smaller than  $\alpha_1$ ). Then we can calculate the updated marginal power,  $\omega_2$ , based on  $\alpha_2$ . Lastly, we can calculate the disjunctive power,  $\Omega_2$ , using  $\omega_2$ . The details will be described in the following steps.

Beside controlling for FWER, the goal of this two-period K+M-experimental arm platform design is to minimize the sample size ( $N_2$ ), while keeping the marginal power ( $\omega_2$ )

and disjunctive power ( $\Omega_2$ ) no less than their counterparts in the K+M-experimental arm trial ( $\omega_1$  and  $\Omega_1$ ). That is, we set the lower limit of  $\omega_2$  (denoted as  $\omega_{2min}$ ) to be 0.8, and the lower limit of  $\Omega_2$  (denoted as  $\Omega_{2min}$ ) to be 0.922 in our “2+2” example. These two limits will be used to select the recommended optimal design(s) (details shown in Step 13).

```
FWER_2 <- FWER_1
FWER_2
## [1] 0.025
omega2_min <- 1-beta1
omega2_min
## [1] 0.8
Omega2_min <- multi$Power1
Omega2_min
## [1] 0.9222971
```

*Step 8: Admissible Set* Because we need to keep  $FWER_2$  equal to  $FWER_1$  when adding new arms, we must update  $n_1$  to  $n_2$  and  $n_{0_1}$  to  $n_{0_2}$  for each experimental arm and its concurrent control (see Figures 1 and 3). Here  $n_2$  and  $n_{0_2}$  are the sample sizes for each experimental arm and its concurrent control in the K+M-experimental arm trial.

We define an admissible set for the pairs of  $(n_2, n_{0_2})$  based on the following three constraints. The first two constraints for  $(n_2, n_{0_2})$  are related to the allocation ratio after adding the new arms. This ratio is denoted as  $A_2$ . To reiterate, in the first period with two experimental arms (before adding the new arms), the allocation ratio is  $A_1$ . Once the two new experimental arms are added, we need an updated allocation ratio  $A_2$  to achieve the desired design properties (i.e., control the FWER and achieve the marginal power). After the two initial experimental arms stopped recruiting, the trial will again have only two experimental arms. Therefore, the allocation ratio  $A_2$  will revert to  $A_1$ .

Here we have

$$A_2 = (n_{0_2} - n_{0_t}) / (n_2 - n_t) > 0,$$

where  $n_t$  and  $n_{0_t}$  are the numbers of patients for each of the experimental arms and the common control at the time of adding new arms. The value of  $A_2$  needs to be a non-infinite positive number. For example, the first two constraints in our “2+2” example are

$$n_{0_2} > n_{0_t} = 43$$

and

$$n_2 > n_t = 30.$$

We also need to set an upper limit for the total sample size  $N_2$  of the K+M-experimental arm trial. A reasonable upper limit is that  $N_2$  should not exceed the required sample sizes (denoted as  $S$ ) of separately conducting two multiarm trials, i.e., a K-experimental arm trial and an M-experimental arm trial.

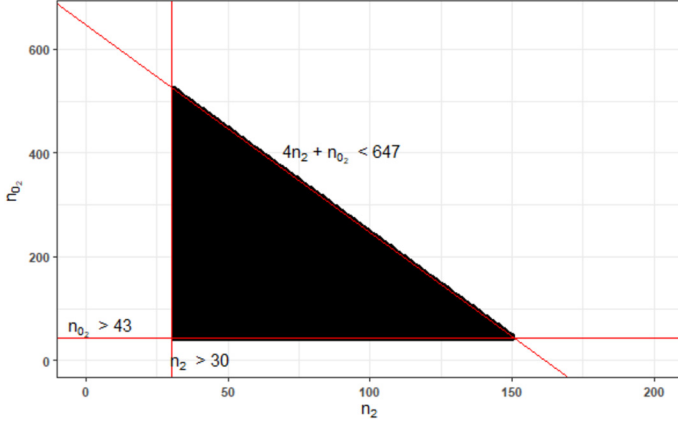


Figure 15: The admissible set of  $(n_2, n_{0_2})$  when  $n_t = 30$  in a two-period 2+2-experimental arm platform trial.

Based on formulae (A.2) and (A.3),

$$S = \frac{(z_{1-\alpha_1} + z_{1-\beta_1})^2}{\Delta^2} (1 + 2\sqrt{K} + K) + \frac{(z_{1-\alpha_1^*} + z_{1-\beta_1})^2}{\Delta^2} (1 + 2\sqrt{M} + M).$$

where  $z_{1-\alpha_1}$  and  $z_{1-\alpha_1^*}$  are the critical values for the K- and M-experimental arm trial, separately. Therefore, the third constraint for  $(n_2, n_{0_2})$  is

$$N_2 = (K + M)n_2 + n_{0_2} + n_{0_t} < S.$$

In our “2+2” example,  $z_{1-\alpha_1} = z_{1-\alpha_1^*}$  as  $K = M = 2$ , and  $S = 2N_1 = 690$ . Therefore, a third constraint for  $(n_2, n_{0_2})$  is

$$N_2 = 4n_2 + n_{0_2} + n_{0_t} < 690$$

Under these three constraints, the admissible set of  $(n_2, n_{0_2})$  can be identified using the function `admiss(.)` (integer points in the triangular region in Figure 15). The data set `pair3` contains all  $(n_2, n_{0_2})$  pairs satisfying the three constraints introduced above.

```
S <- 690 # Define S
pair3 <- admiss(n1=101, n0_1=143, nt=30, ntrt=4, S=690)
```

*Step 9: Correlation Matrix 2* For each pair of  $(n_2, n_{0_2})$  in the admissible set, the correlation matrix  $\Sigma_2$  of the four test statistics ( $Z_1, Z_2, Z_3$ , and  $Z_4$ ) can be derived based on equation (A.1) by plugging in  $n_2$  and  $n_{0_2}$ ,

$$\rho_{k,k'} = \frac{n_{0_{kk'}}}{\frac{(n_{0_2})^2}{n_2} + n_{0_2}}$$

Specifically, between the two initial experimental arms (and between the two added arms), the shared control is

$n_{0_{kk'}} = n_{0_2}$ . Therefore, the correlation of Z statistics between the two initially opened experimental arms (and between the two added arms) is

$$\rho_1 = \frac{1}{(n_{0_2}/n_2 + 1)}. \quad (\text{A.4})$$

The number of shared controls between one initially opened and one newly added experimental arm is  $n_{0_{kk'}} = n_{0_2} - n_{0_t}$ . Therefore, the correlation of the Z test statistics between these two experimental arms is

$$\rho_2 = \frac{n_{0_2} - n_{0_t}}{(n_{0_2}^2/n_2 + n_{0_2})}. \quad (\text{A.5})$$

In our “2+2” example, we have the  $\Sigma_2$  as

$$\Sigma_2 = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_2 \\ \rho_1 & 1 & \rho_2 & \rho_2 \\ \rho_2 & \rho_2 & 1 & \rho_1 \\ \rho_2 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

*Step 10: Critical Value 2* Now we can use  $FWER_2$  and  $\Sigma_2$  to calculate the marginal type-I error  $\alpha_2$  and the corresponding critical value  $z_{1-\alpha_2}$  for each pair of  $(n_2, n_{0_2})$  in the admissible set (found in **Step 8**) by using the following equation.

$$FWER_2 = 1 - \int_{-\infty}^{z_{1-\alpha_2}} \int_{-\infty}^{z_{1-\alpha_2}} \dots \int_{-\infty}^{z_{1-\alpha_2}} \pi_Z(Z(z_1, z_2, \dots, z_{K+M}), 0, \Sigma_2) dz_1 dz_2 \dots dz_{K+M} \quad (\text{A.6})$$

*Step 11: Marginal Power 2* With  $n_1, n_{0_1}, \alpha_1, \beta_1$ , and  $\alpha_2$ , we can use the following equation (A.7) to calculate the marginal power  $\omega_2 = 1 - \beta_2$  for each pair of  $(n_2, n_{0_2})$  from  $z_{1-\beta_2}$ .

$$z_{1-\beta_2} = \sqrt{\frac{\frac{1}{n_1} + \frac{1}{n_{0_1}}}{\frac{1}{n_2} + \frac{1}{n_{0_2}}}} (z_{1-\alpha_1} + z_{1-\beta_1}) - z_{1-\alpha_2} \quad (\text{A.7})$$

*Step 12: Disjunctive Power 2* With the marginal type-II error  $\beta_2$  and  $\Sigma_2$ , we can derive the disjunctive power  $\Omega_2$  for each pair of  $(n_2, n_{0_2})$  by using equation (A.8).

$$\Omega_2 = 1 - \int_{-\infty}^{z_{\beta_2}} \int_{-\infty}^{z_{\beta_2}} \dots \int_{-\infty}^{z_{\beta_2}} \pi_Z(Z(z_1, z_2, \dots, z_{K+M}), 0, \Sigma_2) dz_1 dz_2 \dots dz_{K+M} \quad (\text{A.8})$$

*Step 13: Design Selection* In our “2+2” example, we calculate  $\omega_2$  and  $\Omega_2$  from all 29,040 pairs of  $(n_2, n_{0_2})$  in the entire admissible set. We then perform a 2-step selection procedure to obtain the recommended design(s):

1. We keep only the designs with  $\omega_2 \geq \omega_1$  and  $\Omega_2 \geq \Omega_1$ .
2. Then, among the selected designs, we choose the ones with the smallest  $N_2$ .

Given  $n_t$ ,  $K$ ,  $M$ ,  $FWER_1$ ,  $\omega_1$ , and  $\Delta$ , the function `platform_Design(.)` can provide the optimal K+M-experimental arm trial designs with a minimum total sample size among designs having  $\omega_2$  (marginal power) and  $\Omega_2$  (disjunctive power) no less than their counterparts in the K-experimental arm trial.

```
design <- platform_design(nt = 30, K = 2, M = 2,
  fwer = 0.025, marginal.power = 0.8, delta = 0.4)
design$designs
```

The first part of the outputs (`$design_Karm`) contains the parameters for the K-experimental arm trial. The second part (`$designs`) contains the parameters for the K+M-experimental arm trial designed based on the former. From above (`$designs`), we can see it is possible to have multiple recommended designs that have the same total sample size  $N_2$ . We provide a full list of useful parameters for each of the recommended optimal designs.

For example, if we choose design # 15669 for this 2+2-experimental arm trial, the corresponding critical value for controlling the  $FWER$  at 0.025 is 2.475. The marginal power is 0.8, and the disjunctive power is 0.985, both of which are not less than their counterparts in the 2-experimental arm trial. The required total sample size is  $N_2 = 669$ . Among the 669 patients, in the first period the sample sizes for each experimental arm and the control are  $n_t = 30$  and  $n_{0_t} = 43$ , with an allocation ratio of  $A_1 = 1.414$ . Once the two additional experimental arms are added, the optimal allocation ratio changes to  $A_2 = 2.01$  for the overlapping stage of the second period. The allocation ratio will revert to  $A_1$  once the two initial experimental arms close to accrual. Through the entire 2+2-experimental arm trial, the sample size for each experimental arm is  $n_2 = 107$ . The sample size for the concurrent control of each experimental arm is  $n_{0_2} = 198$ . The sample size for the entire control arm, concurrent and non-concurrent combined, is  $n_c = 241$ . The reduction in the total sample size, compared to two separate 2-experimental arm trials, is 21.

*Step 14: Final Decision* As we can see from Step 13, the total sample size is the same for all four recommended designs. However, the other parameters can be different. Therefore, we can choose a final design based on our needs, according to the other parameters. For example, if we want a design with the largest disjunctive power  $\Omega_2$ , then our final choice is the design # 16632 in Figure 16.

*Note* If  $\omega_{2_{min}}$  and  $\Omega_{2_{min}}$  in Step 7 cannot be met at the same time, the algorithm in `platform_Design(.)` will return the designs with the smallest  $N_2$  but only satisfying one of the two limits. If we do not accept the result or if neither power level is reached, we can choose from the three options below:

- Go back to Step 7 and decrease the value of  $\omega_{2_{min}}$ . After that, repeat Steps 8 to 14 again. This can be done only if a marginal power less than  $\omega_1$  is acceptable, which partially compromises the goal of the design.

```
> p122 <- platform_design(nt = 30, K = 2, M = 2, fwer = 0.025, marginal.power = 0.8, delta = 0.4)
flag.dpmp == 0, lower limits of marginal and disjunctive power are both met
> p122
$design_Karm
  K  n1 n0_1  N1 z_alpha1 FWER1  z_beta1  Power1  cor0 delta
1 2 101 143 345 2.220604 0.025 0.8416212 0.9222971 0.4142136 0.4

$designs
  n2 n0_2 nt n0t  nc  N2  A1  A2  cor1  cor2 critical_value2
15669 107 198 30 43 241 669 1.414214 2.012987 0.3508197 0.2746316 2.475233
15994 106 202 30 43 245 669 1.414214 2.092105 0.3441558 0.2708949 2.475790
16315 105 206 30 43 249 669 1.414214 2.173333 0.3376206 0.2671464 2.476330
16632 104 210 30 43 253 669 1.414214 2.256757 0.3312102 0.2633910 2.476854
  marginal.power2 disjunctive.power2 FWER2 delta save
15669 0.8001100 0.9853799 0.025 0.4 21
15994 0.8003363 0.9857541 0.025 0.4 21
16315 0.8003878 0.9860900 0.025 0.4 21
16632 0.8002699 0.9863903 0.025 0.4 21

$flag.dp
[1] 0

$flag.mp
[1] 0

$flag.dpmp
[1] 0
```

Figure 16: Recommended optimal designs when  $n_t = 30$  in a 2+2-experimental arm trial.

- Go back to Step 6 to set up a smaller  $n_t$  (and  $n_{0_t}$ ). This will increase the overlap between the initial and added experimental arms. The rationale is that the later the new arms are added, the less likely we can find designs satisfying both limits defined in Step 7. After that, repeat Steps 8 to 14 again. This approach is feasible only if the situation allows us to change the timing of adding new arms.
- Consider controlling for PWER instead of FWER, as illustrated in Appendix A.3.

### A.3 Planning a 2+2-Experimental Arm Trial that Controls the PWER

If a study aims to control the PWER, we can simply use the function `platform_Design(.)` with argument `pwer` instead of `fwer` to determine the design parameters. For example, if we plan to add 2 new experimental arms when 30 patients have already been enrolled in each of the 2 initial experimental arms, given the pair-wise type-I error controlled at 0.025 and the marginal power equal to 0.8, we can use the following code to calculate the design parameters, as provided in the results. Here, five optimal designs are recommended and each row is an individual design. Notably, we can save 87 patients with this design compared to 2 separate multiarm trials.

```
> platform_design2(nt=30, K=2, M=2, pwer=0.025,
  marginal.power=0.8, delta=0.4, seed=123)
```

The main difference between using `pwer` instead of `fwer` in `platform_Design(.)` is that it does not use the Dunnett method to derive critical values. Instead, it calculates that directly from the user-defined pair-wise type-I error. Notably, the upper limit  $S$  for the total sample size  $N_2$  that is used to find the admissible set when controlling for PWER is constructed using the multiarm trials (one K- and one M-experimental arm trial), which are also controlling for PWER in the function `platform_Design(.)`. The sample



```

> p122_p <- platform_design(nt = 30, K = 2, M = 2, pwer = 0.025, marginal.power = 0.8, delta = 0.4)
flag.dmp == 0, lower limits of marginal and disjunctive power are both met
> p122_p
$design_karm
  K n1 n0_1 N1 z_alpha1 FWER1 z_beta1 Power1 cor0 delta
1 2 84 119 287 1.959964 0.04647892 0.8416212 0.9222971 0.4142136 0.4

$designs
  n2 n0_2 nt n0t nc N2 A1 A2 cor1 cor2 critical_value2 marginal.power2
7678 76 140 30 43 183 487 1.414214 2.108696 0.3518519 0.2437831 1.959964 0.8001424
7945 75 144 30 43 187 487 1.414214 2.244444 0.3424658 0.2402017 1.959964 0.8005861
8208 74 148 30 43 191 487 1.414214 2.386364 0.3333333 0.2364865 1.959964 0.8007312
8467 73 152 30 43 195 487 1.414214 2.534884 0.3244444 0.2326608 1.959964 0.8005900
8722 72 156 30 43 199 487 1.414214 2.690476 0.3157895 0.2287449 1.959964 0.8001734

disjunctive.power2 FWER2 delta save
7678 0.9867451 0.08807302 0.4 87
7945 0.9871901 0.08823059 0.4 87
8208 0.9875784 0.08848706 0.4 87
8467 0.9879145 0.08868938 0.4 87
8722 0.9882024 0.08888085 0.4 87

$flag.dp
[1] 0

$flag.mp
[1] 0

$flag.dmp
[1] 0

```

Figure 17: Recommended optimal designs when  $n_t = 30$  in a 2+2-trial that controls for PWER.

sizes for the multiarm trials (controlling for PWER) can also be calculated with the function `one_stage_multiarm(.)`. Other aspects of the algorithms are similar between the two applications of the `platform_Design(.)` function.

## ACKNOWLEDGEMENTS

We thank the associate editor and the reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. The authors thank Dr. Yixin Ren for her helpful conversations and Dr. Angela McArthur for the scientific editing of this manuscript.

## FUNDING

This research was supported by the NCI Comprehensive Cancer Center grant (P30 CA021765) and the American Lebanese Syrian Associated Charities (ALSAC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors have declared no conflict of interest. Data sharing is not applicable to this article, as no new data were created or analyzed in this study.

Accepted 10 November 2022

## REFERENCES

- ATKINSON, A., DONEV, A. and TOBIAS, R. *Optimum experimental designs, with SAS* **34**. OUP, Oxford (2007). [MR2323647](#)
- ANGUS, D. C., ALEXANDER, B. M., BERRY, S., BUXTON, M., LEWIS, R., PAOLONI, M. and WOODCOCK, J. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nature Reviews Drug Discovery* **18**(10) 797–808 (2019)
- BURNETT, T., MOZGUNOV, P., PALLMANN, P., VILLAR, S. S., WHEELER, G. M. and JAKI, T. Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs. *Bmc medicine* **18**(1) 1–21 (2020)
- COLLIGNON, O., SCHIEL, A., BURMAN, C. F., RUFIBACH, K., POSCH, M. and BRETZ, F. Estimands and Complex Innovative Designs. *Clinical Pharmacology & Therapeutics* (2022)
- The Adaptive Platform Trials Coalition. Author Correction: Adaptive platform trials: definition, design, conduct and reporting considerations. *Nature reviews. Drug discovery* **18**(10) 808 (2019)
- DODD, L. E., FREIDLIN, B. and KORN, E. L. Platform trials—beware the noncomparable control group. *New England Journal of Medicine* **384**(16) 1572–1573 (2021)
- DUNNETT, C. W. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**(272) 1096–1121 (1955)
- EU-PEARL webinar – “Non-concurrent controls in platform trials”. [https://eu-pearl.eu/workshops/non-concurrent-controls-in-platform-trials/?utm\\_content=196819571&utm\\_medium=social&utm\\_source=twitter&hss\\_channel=tw-906901452](https://eu-pearl.eu/workshops/non-concurrent-controls-in-platform-trials/?utm_content=196819571&utm_medium=social&utm_source=twitter&hss_channel=tw-906901452)
- AGENCY, E. M. Points to consider on multiplicity issues in clinical trials. Doc. Ref.: CPMP/EWP/908/99 (2002)
- GUIDANCE, D. Multiple endpoints in clinical trials guidance for industry. Center for Biologics Evaluation and Research (CBER) (2017)
- LEE, K. M., WASON, J. and STALLARD, N. To add or not to add a new treatment arm to a multiarm study: A decision-theoretic framework. *Statistics in Medicine* **38**(18) 3305–3321 (2019). <https://doi.org/10.1002/sim.8194>. [MR3979810](#)
- LEE, K. M., BROWN, L. C., JAKI, T., STALLARD, N. and WASON, J. Statistical consideration when adding new arms to ongoing clinical trials: the potentials and the caveats. *Trials* **22**(1) 1–10 (2021)
- MAGIRR, D., JAKI, T. and WHITEHEAD, J. A generalized Dunnett test for multi-arm multi-stage clinical studies with treatment selection. *Biometrika* **99**(2) 494–501 (2012). <https://doi.org/10.1093/biomet/ass002>. [MR2931269](#)
- NORMINGTON, J., ZHU, J., MATTIELLO, F., SARKAR, S. and CARLIN, B. An efficient Bayesian platform trial design for borrowing adaptively from historical control data in lymphoma. *Contemporary clinical trials* **89**, 105890 (2020)
- PARK, J. J., HARARI, O., DRON, L., LESTER, R. T., THORLUND, K. and MILLS, E. J. An overview of platform trials with a checklist for clinical readers. *Journal of clinical epidemiology* **125**. 1–8 (2020)
- REN, Y., LI, X. and CHEN, C. Statistical considerations of phase 3 umbrella trials allowing adding one treatment arm mid-trial. *Contemporary Clinical Trials* **109**, 106538 (2021)
- ROIG, M., KROTKA, P., BURMAN, C. F., GLIMM, E., HEES, K., JACKO, P. and POSCH, M. On model-based time trend adjustments in platform trials with non-concurrent controls. *BMC Medical Research Methodology* **22**, 228 (2022)
- SAVILLE, B. R., BERRY, D. A., BERRY, N. S., VIELE, K. and BERRY, S. M. The bayesian time machine: Accounting for temporal drift in multi-arm platform trials. *Clinical Trials* (2022). <https://doi.org/10.1177/17407745221112013>.
- SRIDHARA, R., MARCHENKO, O., JIANG, Q., PAZDUR, R., POSCH, M., BERRY S., . and LU, C. Use of nonconcurrent common control in master protocols in oncology trials: report of an American statistical association biopharmaceutical section open forum discussion. *Statistics in Biopharmaceutical Research* **14**(3) 353–357 (2022)
- VENTZ, S., CELLAMARE, M., PARMIGIANI, G. and TRIPPA, L. Adding experimental arms to platform clinical trials: randomization procedures and interim analyses. *Biostatistics* **19**(2) 199–215 (2018). <https://doi.org/10.1093/biostatistics/kxx030>. [MR3799612](#)
- WASON, J., MAGIRR, D., LAW, M. and JAKI, T. Some recommendations for multi-arm multi-stage trials. *Statistical methods in medical research* **25**(2) 716–727 (2016). <https://doi.org/10.1177/0962280212465498>. [MR3489662](#)
- WASON, J. M. and JAKI, T. Optimal design of multi-arm multi-stage trials. *Statistics in medicine* **31**(30) 4269–4279 (2012). <https://doi.org/10.1002/sim.5513>. [MR3040080](#)

Haitao Pan. Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, 38105, USA. E-mail address: [haitao.pan@stjude.org](mailto:haitao.pan@stjude.org)

Jingjing Ye. SDS Data Science and Operational Excellence (DSOE), BeiGene, Ltd., Cambridge, MA 02142, USA. E-mail address: [jingjing.ye@beigene.com](mailto:jingjing.ye@beigene.com)

Xiaomeng Yuan. Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, 38105, USA. E-mail address: [xiaomeng.yuan@stjude.org](mailto:xiaomeng.yuan@stjude.org)