

RESEARCH

Open Access



Design considerations for Factorial Adaptive Multi-Arm Multi-Stage (FAST) clinical trials

Jonathan Beall^{1*} , Jordan Elm¹, Matthew W. Semler², Li Wang³, Todd Rice², Hooman Kamel⁴, William Mack⁵ and Akshitkumar M. Mistry⁶

Abstract

Background Multi-Arm, Multi-Stage (MAMS) clinical trial designs allow for multiple therapies to be compared across a spectrum of clinical trial phases. MAMS designs fall under several overarching design groups, including adaptive designs (AD) and multi-arm (MA) designs. Factorial clinical trials designs represent a combination of factorial and MAMS trial designs and can provide increased efficiency relative to fixed, traditional designs. We explore design choices associated with Factorial Adaptive Multi-Arm Multi-Stage (FAST) designs, which represent the combination of factorial and MAMS designs.

Methods Simulation studies were conducted to assess the impact of the type of analyses, the timing of analyses, and the effect size observed across multiple outcomes on trial operating characteristics for a FAST design. Given multiple outcomes types assessed within the hypothetical trial, the primary analysis approach for each assessment varied depending on data type.

Results The simulation studies demonstrate that the proposed class of FAST trial designs can offer a framework to potentially provide improvements relative to other trial designs, such as a MAMS or factorial trial. Further, we note that the design implementation decisions, such as the timing and type of analyses conducted throughout trial, can have a great impact on trial operating characteristics.

Conclusions Motivated by a trial currently under design, our work shows that the FAST category of trial can potentially offer benefits similar to both MAMS and factorial designs; however, the chosen design aspects which can be included in a FAST trial need to be thoroughly explored during the planning phase.

Keywords Clinical trial, Factorial design, Adaptive design, Multi-arm, Multi-stage

Introduction

Multi-Arm, Multi-Stage (MAMS) are an increasingly popular class of clinical trials designs which provide a framework where several therapies can be assessed across multiple clinical trial phases [1, 2]. Generally falling under the categories of adaptive designs (AD) and multi-arm (MA) designs, MAMS often provide increases in efficiency relative to traditional trial designs by allowing for multiple active treatment arms to be compared against a single shared control [2, 3].

Like MAMS designs, factorial clinical trials are another class of trial design which offer increases in efficiencies relative to traditional design by allowing for multiple

*Correspondence:

Jonathan Beall
bealljo@musc.edu

¹ Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

² Department of Medicine, Division of Allergy, Pulmonary, and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

³ Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, USA

⁴ Department of Neurology, Weill Cornell Medicine, New York, NY, USA

⁵ Department of Neurosurgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

⁶ Department of Neurosurgery, University of Louisville, Louisville, KY, USA



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

treatment arms to be compared against a single control arm [4, 5]. In order to achieve these efficiencies, factorial design randomize subjects within two or more domains of interest and often assume that there is no interaction effect present among the active treatment arm, with this claim of no interaction often needing some form of prior support. At the time of the final assessment of treatment effect, the analyses are completed with respect to the multiple randomizations across domains of interest [4–6].

Factorial Adaptive Multi-Arm Multi-Stage (FAST) designs represent the combination of factorial and MAMS design and offer benefits similar both of its' predecessors, namely the ability to compare multiple active treatment arms to a single shared control under a framework which spans multiple trial phases and allows for pre-specified adaptations [4]. While MAMS and factorial designs have each been thoroughly researched and utilized, a review of the literature suggest that FAST designs are both understudied and underutilized [4]. Previous work by White et al. (2022) explored a general framework for FAST (or factorial-MAMS) designs; however, their work focused on the practical advantages/disadvantages of the FAST design.

The purpose of this work is to build on the current literature by exploring the impact of the timing and type of assessments during the Phase II portion of a seamless Phase II/III FAST design on the final operating characteristics. While this work is motivated by a clinical trial which is currently under development, detailed below, the results will provide insight into the many aspects which need to be carefully considered during the design of a FAST clinical trial.

Motivating trial

This work is motivated by a multicenter, factorial, randomized clinical trial currently being planned to compare the effectiveness of different fluid therapies and a mineralocorticoid in the treatment of patients with subarachnoid hemorrhage from spontaneous rupture of an intracranial aneurysm. Specifically, two newer fluid therapies will be compared to the standard use of saline (control), and the use of mineralocorticoid will be compared to no use (control). The primary objective is to determine whether any of the newer fluid therapies and the use of mineralocorticoid improves the outcomes of patients with subarachnoid hemorrhage. The primary outcome of the study is the proportion of patients who are assessed to have a functional outcome score of 4 to 6, which corresponds to poor outcomes on an ordinal scale (called modified Rankin Scale, or mRS) that ranges from 0 to 6.

Given the goal of testing two therapeutic domains, two newer fluid therapies and a mineralocorticoid, a 3×2 factorial trial was initially selected as an ideal design. However, the challenging aspect of this trial stems from the fluid intervention rather than the mineralocorticoid. Because of the acute and often unexpected changes experienced by subarachnoid hemorrhage patients that often lead to treatment with more than one fluid, "contamination" with other fluids in practice raises concerns about trial feasibility for the fluid interventions. Therefore, a seamless phase 2/3 testing is implemented for the fluid factors, while the mineralocorticoid simultaneously undergoes only phase 3 testing. This combination leads to a FAST design being optimal for the proposed questions of interest.

For the three fluids being tested, each have distinct electrolyte composition which generate distinct physiological effects in patients. For example, saline has the lowest bicarbonate and highest sodium content among the 3 fluids; therefore, fluid therapy with saline leads to lower plasma bicarbonate levels when compared to therapy with the newer fluids. We hypothesize that the therapeutic effect of fluids is mediated by changes in bicarbonate or sodium levels in the plasma. As such, bicarbonate and sodium can be considered outcomes of interest in the Phase II portion of the trial. Utilizing bicarbonate and sodium as biologic markers of therapeutic effect offers an opportunity for two assessments in the Phase II portion of the trial to increase overall trial efficiency.

An adaptation in Phase II would be to drop randomization into one of the newer fluids. Dropping one of the two newer fluid therapies is very desirable to conserve resources in Phase III testing. Whichever of the two newer fluids that generates the greatest effect in the plasma bicarbonate and sodium in the opposite direction compared to the expected levels produced by treatment with saline will be advanced. For example, the prior expectation is that, among the 3 fluid therapies considered, treatment with saline will result in the lowest bicarbonate levels. As such, the goal of the arm dropping assessment would be to compare the two novel fluid therapies to assess for which therapy produces the greatest bicarbonate levels, ultimately creating the greatest effect of treatment when compared to saline on bicarbonate levels. In the absence of a clear differences between the novel fluid therapies in both effects, one pre-selected newer therapy will be advanced by default. In the event that at least one significant difference is observed when comparing the novel fluid therapies, decision rules are explicitly defined for retaining either one or both of the therapies.

A second potential assessment in Phase II would be to determine the feasibility of the trial. Since the prior

expectation is that the electrolyte composition of the included fluids mediates changes patient to outcomes, if the expected differences when comparing the novel fluid therapies to saline on the electrolyte compositions is not observed, then the overall fluid assessment can be terminated.

Given the distinct goals for the fluid and mineralocorticoid domains, a FAST design is best suited to address all hypotheses to be tested and allow for increased efficiency relative to more traditional designs.

Study design

Following our motivating example, we explore a seamless Phase II/III FAST design which includes 2 interim assessments and a final assessment. The primary aim of this work is to examine the impact of the timing for the interim assessments on final operating characteristics. The interim analyses will constitute the Phase II portion of the trial with the final assessment representing the Phase III portion of the trial. Sample sizes for the interim analysis and final operating characteristics will be determined via simulation.

To accommodate the clinical aims of the hypothesized clinical trial, we propose a trial design which allows for multiple questions of interest to be jointly assessed through a FAST design. Henceforth, we will reference each question of interest as a domain. In general, we let $d = A, B, \dots \Delta$ reference the domains of interest with Δ representing the maximum number of domains. From our motivating trial, there are two domains of interest which we will reference as: $A = \text{Fluid}$ and $B = \text{Mineralocorticoid}$. Within in each domain, we will note treatment assignment within domain as d_t where $t = 0, 1, \dots \tau_d$ and τ_d represents the maximum number of treatment arms in domain d . Again utilizing our motivating trial as an example, there would then be three treatment arms in domain A , $A_0 = \text{Saline}$, $A_1 = \text{Fluid}_1$, and $A_2 = \text{Fluid}_2$, and two treatments in domain B , $B_0 = \text{Control}$ and $B_1 = \text{Mineralocorticoid}$. We further note r as the combination of treatment assignments across domains as a treatment regime with $r = 1, \dots, R = (\tau_1 \times \dots \times \tau_\Delta)$. If allowing all combination of treatment assignments across domains, then there are 6 possible treatment regimes for our motivating example: $r = 1, \dots, 6$ with regimes being $((A_0, B_0), (A_1, B_0), (A_2, B_0), (A_0, B_1), (A_1, B_1), (A_2, B_1))$.

For the outcomes of interest, we will note outcomes by trial phase. Thus, let Y_{po} represent the o^{th} outcome of interest for the p^{th} trial phase, where $o = 1, 2, \dots, \theta$ and $p = 1, 2, \dots, \rho$. From our motivating example, the outcomes utilized in the Phase II portion of the trial represent biomarkers assumed to be affected by treatment and the Phase III outcome of interest represents a functional outcome. We will assume two analyses within the Phase

II portion of the trial and one in the Phase III portion of the trial. This is done to mirror the proposed design for the motivating trial. The Phase II portion of the trial will consist of 2 analyses: a feasibility analysis and an arm-dropping analysis.

Phase II

Arm dropping analysis

The goal of the arm-dropping analysis is to determine which, if any, of the treatment arms (novel fluid therapies) can be discontinued from enrollment. We let $\mu_{A_1}^{11}$ and $\mu_{A_2}^{11}$ represent the average value for outcome Y_{11} (bicarbonate) for subjects in arms 1 and 2 in domain A, respectively. Further, we let $\mu_{A_1}^{12}$ and $\mu_{A_2}^{12}$ represent the average value for outcome Y_{12} (sodium) for subjects in arms 1 and 2 in domain A, respectively. Then, utilizing independent t-tests by outcome, the non-control arms will be compared by assessing two hypotheses:

$$H_0 : \mu_{A_1}^{11} = \mu_{A_2}^{11} \text{ vs. } H_a : \mu_{A_1}^{11} \neq \mu_{A_2}^{11}$$

$$H_0 : \mu_{A_1}^{12} = \mu_{A_2}^{12} \text{ vs. } H_a : \mu_{A_1}^{12} \neq \mu_{A_2}^{12}$$

The results of these hypothesis tests will be used to determine which arms will continue to enroll. If there is a statistically significant difference observed for outcome Y_{11} , the treatment arm in domain A with the highest average outcome Y_{11} value will be retained. If there is a difference in outcome Y_{12} between the treatment arms in domain A, then the arm with the lowest outcome Y_{12} average will be retained. There is a possibility that neither arm will be dropped.

Feasibility analysis

The goal of the feasibility analysis is to assess, utilizing one or more biomarkers, if the expected effect of treatment is being observed on the biomarker(s). While the comparison for this example focuses on responses in key biomarkers, any appropriate Phase II assessment could be included in this analysis. Treatment arms will be compared using an independent one-sided t -test. Letting $\mu_{A_0}^{11}$ represent the average value for outcome Y_{11} for subjects in arm A_0 (the control arm for domain A) and $\mu_{A_{(1+2)}}^{11}$ represent the average outcome 11 for the pooled treatment arms from domain A, we then propose a one-sided test, with hypothesis below:

$$H_0 : \mu_{A_0}^{11} \geq \mu_{A_{(1+2)}}^{11} \text{ vs. } H_a : \mu_{A_0}^{11} \leq \mu_{A_{(1+2)}}^{11}$$

If statistical significance is not obtained, then the assessment of treatment effect within domain A will terminate and subjects will no longer randomized to any of the treatments within domain A.

For this analysis, all subject data are utilized in the comparison. Pooling data within A_1 and A_2 could result in an assessment which is under-powered if there is a differential effect of treatment on Y_{11} . Given this pooling, in addition to the treatment effect, the impact of this pooling is likely affected by the timing and ordering of the arm-dropping and feasibility assessments. The impact of these trial characteristics and their impact on overall operating characteristics will be assessed via simulation.

Phase III

For any trial where domain A proceeds beyond the Phase II portion of the trial, the final analysis will consist of an assessment comparing each treatment arm to control within domain. The final analysis will be conducted by constructing a generalized linear model where the outcome of interest is the Phase III outcome, assumed to be a binary outcome, and covariates in the model represent the treatment assignments for each subject. An appropriate correction for the multiple comparisons will be applied to α so that type 1 error is controlled at a 5% level [7]. Due to the arm dropping analysis, there are three possible scenarios for the final analysis: where a treatment in domain A is dropped, where no treatment in domain A is dropped, where domain A is terminated.

Domain A continued with one arm retained

In this scenario, while the majority of subjects will be randomized to one of the newer fluid arms or saline, a small proportion of subjects will be randomized to the dropped fluid arm prior to the arm dropping analysis. For the final analysis, subjects will be grouped as either having been randomized to saline or one of the newer fluid arms; that is, subjects randomized to either of the non-saline arms throughout the duration of the trial are pooled into a single fluid therapy arm for the final analysis.

Therefore, letting $I(X_{A_0i} = 1)$ represent an indicator variable for if subject i was randomized to saline, $I(X_{B_1i} = 1)$ represent an indicator variable for if subject i was randomized to mineralocorticoid, and $p_i = P(Y_{21} = 1)$, we can then construct the primary analysis model as below.

$$\text{logit}(p_i) = \beta_0 + \beta_1 I(X_{A_0i} = 1) + \beta_2 I(X_{B_1i} = 1)$$

To control the family-wise type I error in the strong sense at $\alpha = 0.05$, we will then assess for any significant treatment effect among the fluid and mineralocorticoid arms using a gatekeeping procedure:

$$H_0 : \beta_1 = \beta_2 = 0 \text{ vs. } H_a : \beta_1 \neq 0 \text{ or } \beta_2 \neq 0$$

If the null hypothesis for the test above is rejected, implying that either the pooled fluid arm or the

mineralocorticoid arm has a non-zero treatment effect relative to the saline/no mineralocorticoid arm, then individual tests comparing each arm to control will be completed.

$$H_0 : \beta_1 = 0 \text{ vs. } H_a : \beta_1 \neq 0$$

$$H_0 : \beta_2 = 0 \text{ vs. } H_a : \beta_2 \neq 0$$

As with the feasibility assessment in Phase II, it is worth noting that the comparisons above utilize the pooling approach so that all patient data inform the final analysis. This approach could reduce power if there is a different treatment effect within the pooled treatment arms; however, the impact of this effect is expected to vary with the timing of the assessments within Phase II. The relationship between the design parameters and the operating characteristics will be assessed in Phase III.

Domain A continued with both arms retained

If no arm is dropped, then the final analysis will also proceed using a gatekeeping approach. The primary analysis model including fluid arm and mineralocorticoid arm will be constructed as below. Unlike the previous scenario, if neither of the newer fluid arms are dropping, then we will construct the following model where $I(X_{A_1i} = 1)$ and $I(X_{A_2i} = 1)$ represent indicator variables for if subject i was randomized to either of the newer fluids and $I(X_{B_1i} = 1)$ represents an indicator variable for if subject i was randomized to mineralocorticoid.

$$\text{logit}(p_i) = \beta_0 + \beta_1 I(X_{A_1i} = 1) + \beta_2 I(X_{A_2i} = 1) + \beta_3 I(X_{B_1i} = 1)$$

To preserve α , a gatekeeping procedure will be used to assess for any significant treatment effect among the fluid and mineralocorticoid arms:

$$H_{01} : \beta_1 = \beta_2 = \beta_3 = 0 \text{ vs. } H_{a1} : \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

If the null hypothesis for the test above is rejected, then all pairwise combinations will be tested as below.

$$H_{02} : \beta_1 = \beta_2 = 0 \text{ vs. } H_{a2} : \beta_1 \neq 0 \text{ or } \beta_2 \neq 0$$

$$H_{03} : \beta_1 = \beta_3 = 0 \text{ vs. } H_{a3} : \beta_1 \neq 0 \text{ or } \beta_3 \neq 0$$

$$H_{04} : \beta_2 = \beta_3 = 0 \text{ vs. } H_{a4} : \beta_2 \neq 0 \text{ or } \beta_3 \neq 0$$

Depending on which of the null hypotheses for the tests above are rejected, then individual tests comparing each arm to control will be completed.

$$H_{05} : \beta_1 = 0 \text{ vs. } H_{a5} : \beta_1 \neq 0$$

$$H_{06} : \beta_2 = 0 \text{ vs. } H_{a6} : \beta_2 \neq 0$$

$$H_{07} : \beta_3 = 0 \text{ vs. } H_{a7} : \beta_3 \neq 0$$

For example, the null hypothesis H_{05} will only be assessed if H_{01} , H_{02} , and H_{03} are all rejected in their respective tests.

Domain A terminated at feasibility assessment

In this scenario, after the feasibility assessment, randomization will cease within domain A. As such, for the final analysis will consider only the treatment assignments for domain B. For the final analysis in this scenario, subjects will be grouped as either having been randomized to mineralocorticoid or no mineralocorticoid. Therefore, let $I(X_{B_1i} = 1)$ represent an indicator variable for if subject i was randomized to mineralocorticoid, we can then construct the primary analysis model as below.

$$\text{logit}(p_i) = \beta_0 + \beta_1 I(X_{B_1i} = 1)$$

Under this scenario, only a single parameter β_1 is being assessed; therefore, no methods for the preservation of α are required.

Simulation

To evaluate the operating characteristics of the proposed design, specifically the impact of the timing of Phase II analysis on the overall operating characteristics of the trial, thorough simulation studies were conducted. For our simulations, we assume a total sample size of 1005 subjects with interim assessment sample sizes being assessed in simulation.

It is assumed that there are two phases of interest, $p = 1, 2$ with two outcomes of interest for $p = 1$ and one outcome of interest for $p = 2$. As such, for each treatment arm, there are 3 outcomes of interest: (Y_{11}, Y_{12}, Y_{21}) . Additionally, we assume that there two domains of interest $d = A, B$, with 3 treatment arms considered for domain A, (A_0, A_1, A_2) and 2 treatment arms considered for domain B, (B_0, B_1) .

We assume that $Y_{r11} \sim N(\mu_{r11}, \sigma_{r11})$ where μ_{r11} and σ_{r11} represent the mean and standard deviation for the r^{th} treatment regime, respectively. We similarly define Y_{r12} as Y_{r11} . For Y_{r21} , we assume that $Y_{r21} \sim \text{Binomial}(\pi_{r21})$, where $\pi_{r21} = P(Y_{r21} = 1)$. Thus, for the first phase of the study, $p = 1$, all outcomes are considered to be continuous and the outcome for the second phase of the study, $p = 2$, are assumed to be dichotomous. For simplicity, we do not include any covariates beyond treatment assignment.

Parameters varied in the simulations include timing of the feasibility assessment, timing of the arm-dropping

assessment, effect of treatments in domain A on each biomarker of interest, effect of treatments in domain A on the Phase III outcome, and effect of treatments in domain B on the Phase III outcome.

For timing of the feasibility and arm dropping analyses, the timing of each analysis is varied such that each can occur once outcomes from 90 subjects are available up to outcomes from 300 subjects, incrementing by 30. This allows for an assessment of both the effect of sample size for each analysis on the final operating characteristics and the effect that the order of the analyses has on the final operating characteristics. For each of the biomarkers associated with analyses in the Phase II portion of the trial, the impact of a treatment in domain A providing a mean change from baseline of either 0 or 10 units was assessed. For the Phase III portion of the trial, treatments in either domain can provide a 10% change in the primary outcome (Y_{21}) relative to the control group.

Simulations were conducted using R. For each simulation condition, 2500 simulated datasets were generated.

Results

Full simulation results can be seen in Additional files 1 and 2. Figure 1 presents the probability of retaining the correct arm during the arm dropping analysis, the probability of proceeding beyond the feasibility analysis, and the overall probability of detecting the treatment effect across the possible sample size requirements for the feasibility and arm-dropping analysis for a representative set of simulation conditions (Scenarios A, B and C). Of note, the ordering of the feasibility and arm-dropping analyses is determined by their respective timings/required sample sizes.

The graphs associated with row A represent the case where $Y_{11} : (CTL = 0, A_1 = -10, A_2 = -10)$, $Y_{12} : (CTL = 0, A_1 = 10, A_2 = -10)$, and $Y_{21} : (A_1 = 0.1, A_2 = 0.1, B_1 = 0)$. The opposing signs for the effect of treatment on Y_{11} and Y_{12} is clinically relevant for our motivating trial, as we expect one biomarker to increase the second to decrease with treatment; however, these opposing signs are inconsequential for the trial as both outcomes are continuous. In this case, the correct arm to retain during the arm-dropping analysis is A_2 . This is a reflection of the the arm-dropping algorithm where the arm(s) with the highest average outcome Y_{11} value and the lowest outcome Y_{12} value being retained if there is a significant difference. If no difference is found, the the trial will proceed with A_2 . Given that both A_1 and A_2 demonstrate the same mean shift from control for Y_{11} and Y_{12} , then there is no expected difference between A_1 and A_2 .

The graphs associated with row B represent the case where $Y_{11} : (CTL = 0, A_1 = 0, A_2 = -10)$, $Y_{12} :$

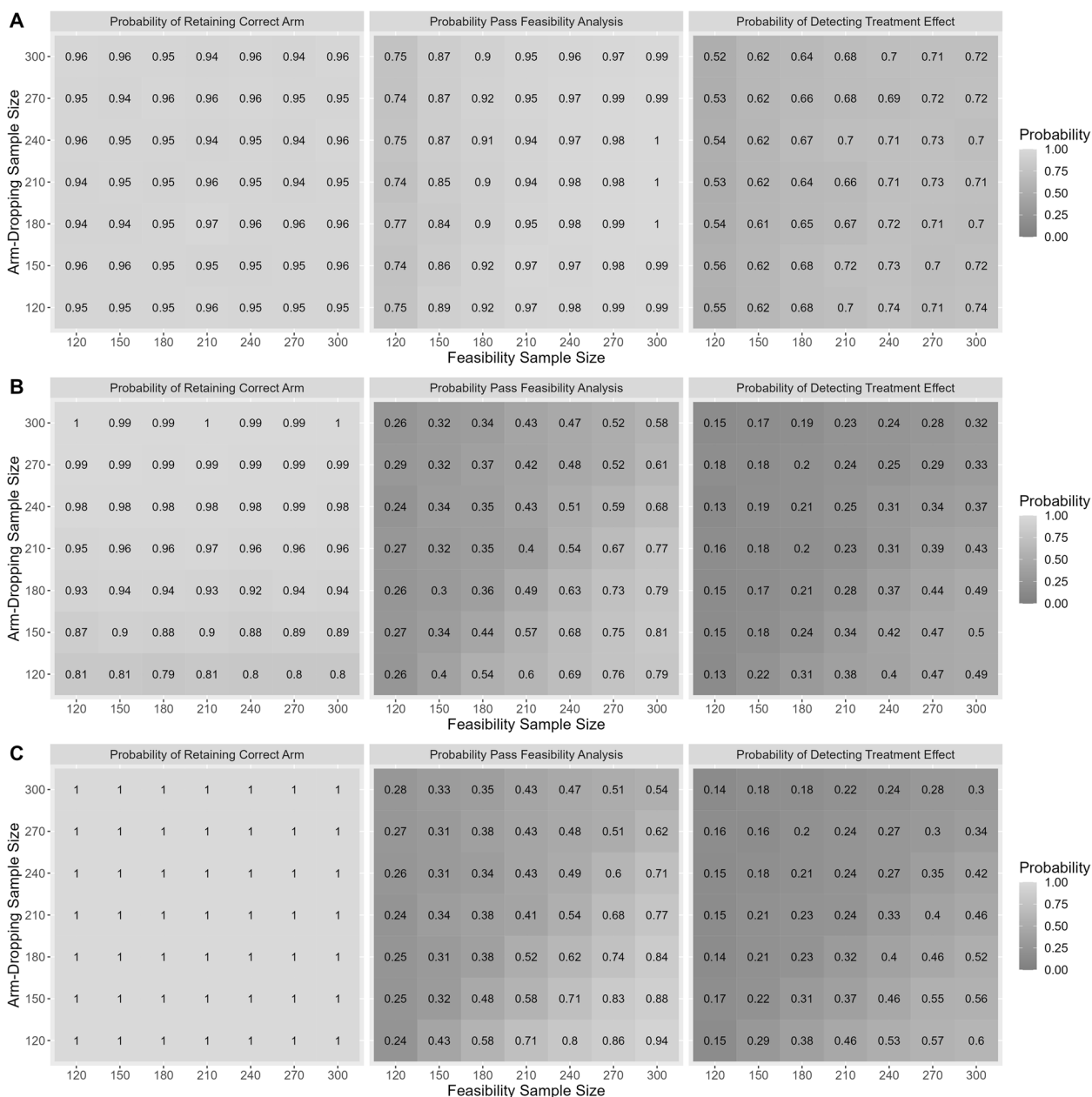


Fig. 1 Simulation results for three scenarios (A, B, C). The graphs associated with row A represent the case where $Y_{11} : (A_0 = 0, A_1 = -10, A_2 = -10)$, $Y_{12} : (A_0 = 0, A_1 = 10, A_2 = -10)$, and $Y_{21} : (A_1 = 0.1, A_2 = 0.1, B_1 = 0.0)$. The graphs associated with row B represent the case where $Y_{11} : (A_0 = 0, A_1 = 0, A_2 = -10)$, $Y_{12} : (A_0 = 0, A_1 = 0, A_2 = 10)$, and $Y_{21} : (A_1 = 0, A_2 = 0.1, B_1 = 0.0)$. The graphs associated with row C represent the case where $Y_{11} : (A_0 = 0, A_1 = -10, A_2 = 0)$, $Y_{12} : (A_0 = 0, A_1 = 10, A_2 = 0)$, and $Y_{21} : (A_1 = 0.1, A_2 = 0.0, B_1 = 0.0)$

($CTL = 0, A_1 = 0, A_2 = 10$), and $Y_{21} : (A_1 = 0, A_2 = 0.1, B_1 = 0)$. Unlike Scenario A, Scenario B represents the case where A_2 demonstrates a difference from control when A_1 is equivalent to control. In this case, the correct arm to retain during the arm-dropping analysis is A_2 .

The graphs associated with row C represent the case where $Y_{11} : (CTL = 0, A_1 = -10, A_2 = 0)$, $Y_{12} : (CTL = 0, A_1 = 10, A_2 = 0)$, and $Y_{21} : (A_1 = 0.1, A_2 = 0.0, B_1 = 0)$. Similar to Scenario B, Scenario C represents the case where A_1 demonstrates a difference

from control when A_2 is equivalent to control. As such, the correct arm to retain during the arm-dropping analysis is A_1 .

Figure 1 shows that, for the arm dropping analysis, the probability of retaining the correct appears to be invariant to the sample size when the analysis is conducted for Scenarios A and C; however, in Scenario B the probability of retaining the correct arm increases with larger sample sizes. Further, it appears that the probability of retaining the correct arm is approximately constant for all feasibility analysis sample sizes.

For the feasibility analysis, in general, the larger the sample size at the time of the analysis, the greater the probability of proceeding. However, unlike the arm-dropping analysis, the timing of the arm-dropping analysis does appear to impact the results of the feasibility analysis. Specifically, when the sample size requirement is larger for the feasibility analysis, increasing the sample size required for the arm-dropping analysis appears to decrease the probability of proceeding to the Phase III portion of the trial.

Lastly, we see that the power to detect the treatment effect in Phase III is heavily dependent upon the timing of the feasibility and arm-dropping analyses. In general, overall study power appears to be maximized when the sample size for the feasibility analysis is large and when the sample size for the arm-dropping analysis is small. These results, which are similar to other conditions tested in simulation presented in Additional files 1 and 2, indicate that the Phase II analyses should be ordered such that the arm-dropping analysis occur before the feasibility analysis.

Table 1 presents the characteristics across Phase III treatment effect conditions if the arm-dropping analysis were to occur once 150 outcomes are obtained and the feasibility analysis were to occur once 300 outcomes are obtained for scenarios where the arms which demonstrate an effect on the Phase III outcomes also demonstrate a treatment effect on both of the Phase II outcomes. Table 1 displays the likelihood of retaining the correct treatment arm (or both), the probability of continuing

beyond the feasibility assessment, the probability of declaring each possible combination of treatment arms across domain as successful in Phase III, the overall study power, and the overall study type I error. This table demonstrates that type I error (row $(A_1 = 0, A_2 = 0, B_1 = 0)$) is approximately 5%. While the point estimate demonstrates a slight inflation in the type I error, potentially due to the decisions made in the Phase II portion of the trial, the monte carlo error does not rule out a type I error rate of 5%. Further, we see that when at least one treatment arm within each domain has a non-zero effect, that power is expected to exceed 90%. However, as expected, when there is a differential effect between the Phase II and Phase III outcomes or when only a single treatment arm is efficacious for both phases, then the overall study power decreases (as observed in Additional files 1 and 2).

Conclusions

The class of FAST trial designs offers a framework which can provide increased efficiencies relative to standard and novel trial designs, such as a MAMS or factorial trial. Simulation results indicate that ordering the assessments for Phase II such that the arm dropping analysis is set to occur once 150 outcomes are obtained and the feasibility assessment is set to occur once 300 outcomes are obtained provides a trial which has excellent power and near 5% type I error. The utilization of a FAST trial design for our motivating sample results in a design which has excellent power and controlled type I error, while be able to address numerous clinical question of interest. It is likely a more sophisticated design than that of a simplistic 3×2 would produce a more optimal comparison; however, for simplicity a 3×2 traditional factorial trial with 85% power, 2.5% type I error per comparison in domain A, 5% type I error in domain B, and 10% loss to follow up would require approximately 1485 patients. This represents an increase of nearly 500 patients to obtained operating characteristics which are still inferior to those observed in our simulation results. While this nearly 50% increase in sample size requirement is observed in this comparison and more complicated designs could reduce

Table 1. Simulation results for the scenario listed in the column “Condition.” While “Condition” represents the effect of treatment on the primary Phase III outcome, for any treatment with a non-zero effect on the primary Phase III outcome, there is also a non-zero treatment effect on both Phase II outcomes. The columns grouped under the “Phase III Analysis” label represent the probability of declaring success for that collection of treatments. For example, in the last row under the column “A2:B1” the value is 0.73 (0.009), which is interpreted as there being a 73% chance of declaring success for A2 and B1 in this simulation scenario with a Monte Carlo error of 0.009. The column “Type I Error” represents the family-wise type I error rate

Condition	Arm Dropping Analysis P(Correct Decision)	Feasibility Analysis P(Continue)	Phase III Analysis						Overall Operating Characteristics			
			A2:A1:B1	A2:A1	A2:B1	A1:B1	A2	A1	B1	None	Power	Type I Error
$A_1 = 0, A_2 = 0, B_1 = 0$	0.95(0.004)	0.05(0.005)	0(0)	0(0)	0(0)	0(0)	0(0.001)	0(0)	0.05(0.004)	0.95(0.004)	NA(NA)	0.05(0.004)
$A_1 = 0, A_2 = 0, B_1 = -0.1$	0.95(0.004)	0.05(0.004)	0(0)	0(0)	0(0.001)	0(0)	0(0)	0(0)	0.89(0.006)	0.11(0.006)	0.89(0.006)	0(0.001)
$A_1 = 0.1, A_2 = 0, B_1 = -0.1$	0.88(0.007)	0.81(0.008)	0(0)	0(0)	0(0.001)	0.54(0.01)	0(0)	0.07(0.005)	0.34(0.009)	0.05(0.004)	0.95(0.004)	0(0.001)
$A_1 = 0, A_2 = 0.1, B_1 = -0.1$	1(0)	0.89(0.006)	0(0)	0(0)	0.65(0.01)	0(0)	0.08(0.005)	0(0)	0.24(0.009)	0.03(0.004)	0.97(0.004)	0(0)
$A_1 = 0.1, A_2 = 0.1, B_1 = 0$	0.95(0.004)	1(0.001)	0(0)	0(0)	0.05(0.004)	0(0.001)	0.69(0.009)	0.04(0.004)	0.01(0.002)	0.21(0.008)	0.73(0.009)	0.06(0.005)
$A_1 = 0.1, A_2 = 0.1, B_1 = -0.1$	0.95(0.005)	0.99(0.002)	0(0.001)	0(0)	0.73(0.009)	0.04(0.004)	0.1(0.006)	0.01(0.002)	0.11(0.006)	0.02(0.003)	0.98(0.003)	NA(NA)

this deficit, this result shows the substantial increase in efficiency that can be obtained when using a FAST trial.

The presented work is directed towards the motivating example; however, as demonstrated by the simulation studies, operating characteristics can be negatively impacted if the order and timing of key analyses are improperly set, ultimately resulting in a loss of the increased efficiency relative to more standard designs. As such, a thorough study of the operating characteristics of the trial should be assessed through carefully constructed simulation studies if one is considering a FAST trial for their clinical question of interest.

Motivated by a trial currently under design, simulation studies demonstrated the numerous aspects of a FAST trial which need to be considered during the planning phase, such as the timing/sample size requirements for each assessment.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13063-024-08400-6>.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

Not applicable.

Authors' contributions

JB, AM, and JE drafted the manuscript. All authors read and approved the final version of the manuscript.

Funding

This work was supported by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under Award Number 5U01NS087748-09. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Availability of data and materials

The results presented are derived from simulated data, as such they are not available.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Ghosh P, Liu L, Mehta C. Adaptive multiarm multistage clinical trials. *Stat Med*. 2020;39:1084–102.
2. Lin J, Bunn V. Comparison of multi-arm multi-stage design and adaptive randomization in platform clinical trials. *Contemp Clin Trials*. 2017;54:48–59.
3. Wason J, Magirr D, Law M, Jaki T. Some recommendations for multi-arm multi-stage trials. 2012;25:716–27.
4. White I, Choodari-Oskoei B, Sydes M, Kahan B, McCabe L, Turkova A, Esmail H, Gibb D, Ford D. Combining factorial and multi-arm multi-stage platform designs to evaluate multiple interventions efficiently. *Clin Trials*. 2022;19:432–41.
5. Jaki T, Vasileiou T. Factorial versus multi-arm multi-stage designs for clinical trials with multiple treatments. *Stat Med*. 2016;36:563–80.
6. Sydes M, Parmar M, James N, Clarke N, Dearnaley D, Mason M, Morgan R, Sanders K, Royston P. Issues in applying multi-arm multi-stage methodology to a clinical trial in prostate cancer: the MRC STAMPEDE trial. *Trials*. 2009;10:1–16.
7. Wang D, Li Y, Liu X, Fo B, Lin Y, Larsen L, Offen W. Overview of multiple testing methodology and recent development in clinical trials. *Contemp Clin Trials*. 2009;45:13–20.

Publisher's Note

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

Received: 30 January 2024 Accepted: 14 August 2024

Published online: 12 September 2024