Title

A pragmatic, adaptive clinical trial design for a rare disease: The FOcal Cerebral Arteriopathy Steroid (FOCAS) trial

Authors

Park, Yeonhee
Fullerton, Heather J
Elm, Jordan J

Publication Date

2019-11-01

DOI

10.1016/j.cct.2019.105852

Peer reviewed

Title:

A Pragmatic, Adaptive Clinical Trial Design for a Rare Disease:

The FOcal Cerebral Arteriopathy Steroid (FOCAS) Trial

Yeonhee Park, PhD[a]; Heather J. Fullerton, MD, MAS[b]; Jordan J. Elm, PhD[a]

[a]Department of Public Health Sciences, Medical University of South Carolina; Charleston, SC

[b]Departments of Neurology & Pediatrics, University of California, San Francisco; San Francisco, CA

Corresponding Author

Jordan Elm, PhD

Medical University of South Carolina

Department of Public Health Sciences

135 Cannon St. Suite 305 PO Box 250835

Charleston, SC 29425

Phone (843) 876-1605

elmj@musc.edu

Figures: 1

Tables: 2

Online Supplemental Material: Appendix

Abstract:

Background: Pediatric stroke investigators identified as their top research priority a clinical trial of corticosteroids for focal cerebral arteriopathy (FCA). However, FCA is both rare and an acute condition making it infeasible to enroll the large sample sizes needed for a standard, confirmatory clinical trial. We present a pragmatic approach to clinical trial design that may inform the approach to other rare disorders.

Methods: We surveyed pediatric stroke experts to determine the level of evidence that would impact their clinical management of FCA. Incorporating survey results, a randomized group sequential Bayesian adaptive design was proposed based on a quantitative radiologic outcome measure. The design sequentially determines based on accumulating information from the first patient whether intervention is better than control with high probability.

Results: Among 21 (100%) respondents, the probability of corticosteroid efficacy that would lead them to treat was 30% (median). The probability of efficacy that would make them unwilling to randomize (because they would feel all children should receive corticosteroids) was 70%. Simulation studies show that a total of 42 enrolled subjects controls the type I error rate at the desired level 0.2 and yields 80% power to detect the difference in mean change of 1.6.

Conclusions: Designs in rare diseases require special considerations; this is especially true for this childhood disease, which is both uncommon and acute. This design has incorporated expert consensus to establish the criteria for success, incorporates formal monitoring rules for safety, and futility or efficacy rules to stop early.

Keywords: Bayesian adaptive trial design, rare disease, pediatric stroke, Sequential monitoring rule.

Funding: This work was supported by a gift to UCSF by Marc and Lynne Benioff.

Introduction:

In a Delphi consensus process, pediatric stroke investigators identified a clinical trial of corticosteroids for the treatment of focal cerebral arteriopathy of childhood (FCA) as the highest research priority for their field.[1] FCA is a life-threatening inflammatory disease of cerebral arteries that progresses over days to weeks and causes arterial ischemic stroke in otherwise healthy children.[2] In the absence of clinical trial data, clinicians have begun to employ corticosteroid therapy.[3, 4] However, there are potential downsides: experimental evidence suggests that the post-stroke inflammatory response has beneficial (in addition to detrimental) effects, with roles in infarct resolution, and brain remodeling and repair.[5] Hence, pediatric stroke experts agree on the need for a clinical trial to guide FCA management.[1]

However, design of an FCA trial faces two major challenges: the rarity of FCA and its acuity. Although FCA is a common cause of childhood arterial ischemic stroke, ischemic stroke in children is rare, occurring in approximately 1.3 per 100,000 US children (non-neonates) per year.[6] Although the incidence of FCA has not been established, it causes ≈10% of childhood arterial ischemic strokes,[7] so has an annual incidence of ≈1.3 per 1,000,000 US children. Hence, in the US population of 74 million children, there are likely <100 annual FCA cases nationwide. The acuity of FCA presents additional challenges; while most rare disease clinical trials study chronic diseases and take advantage of established patient cohorts, an FCA trial needs to rapidly identify and treat acute incident cases, before the arteriopathy progresses.

With a traditional trial design, the sample size is chosen such that the study has a certain amount of power for a minimum clinically important effect with the standard two-sided alpha level of 0.05. This allows us to be "95% confident" that if we observe a treatment effect, then it is not due to chance alone. Likewise, if we do not observe a treatment effect, it is not due to lack of power. For a trial of a rare disease like FCA, this approach may result in an infeasible sample size, accomplishable only with hundreds of enrolling sites and a long enrollment window.

Biomedical ethicists have argued that, for ethical reasons, the overall significance level should be adjusted in settings when clinicians would change clinical practice with a level of confidence lower than 95%.[8] A lower alpha level reduces the necessary sample size for a trial, and hence reduces the number of participants randomized to the inferior therapy. Others have suggested that approaches to sample size calculation ought to be different for rare diseases and need to take into account a variety of assumptions about the treatment effect, as well as input from the scientific community.[9] In this paper, we describe how we designed a FCA corticosteroid trial—the FOcal Cerebral Arteriopathy Steroid (FOCAS) trial—by seeking expert input to establish what would be considered a "good" probability that corticosteroids are better. The trial design operating characteristics are provided. This approach may be useful to investigators designing clinical trials for other rare diseases.

Material and methods:

We approached the FOCAS trial design with the goal of obtaining the level of evidence sufficient to change current clinical practice while providing a feasible approach for a rare disease in an acute setting. The proposed intervention is corticosteroid therapy (a standard 3-day course of IV methylprednisolone followed by oral prednisone) versus control group (standard care); both the treatment and control groups will receive standard therapy of oral aspirin and supportive care. The primary endpoint is the change in the FCA Severity Score (FCASS; a quantitative radiologic measure of the arteriopathy severity) from baseline to 1 month post enrollment.[2] Safety outcome is a binary indicator taking 1 for a safe intervention and 0 for toxic intervention which would lead to any serious adverse event. The maximum sample size will be 50 children to be enrolled at 25 sites including the loss-to-follow-up which will not exceed 20% approximately. The primary analysis is intention-to-treat and will include all randomized children. We adopt the idea from a Prospective Randomized Open, Blinded End-point (PROBE) design because

corticosteroids cause behavioral side effects that preclude double-blinding; the neuroradiologists will measure the FCASS blinded to both treatment arm and time.[10]

Expert Survey: To determine the appropriate criteria for success for the trial, we designed a survey of pediatric stroke experts (full survey available, Online Supplemental Material). The survey first established current practices concerning the use of IV corticosteroids for the treatment of FCA. We also queried, using a Likert scale, current perceptions regarding the safety and benefit of corticosteroid therapy. We then asked, "What probability of efficacy would lead you to treat FCA with corticosteroids?" We repeated this question with two definitions of efficacy: improved imaging biomarkers (arteriopathy progression and infarct volume) and improved 12-month neurological outcomes (using the Pediatric Stroke Outcome Measure, or PSOM). The survey provided examples: "If you could tell a family that the treatment has a 50% chance of improving their child's 12-month outcome, would that be enough for you to treat?" To establish stopping rules for the trial, we also asked what probability of efficacy would make them unwilling to randomize their next patient due to efficacy or due to futility, and asked about acceptable levels of serious adverse effects.

We applied the survey in-person, taking advantage of an investigator meeting for an NIH-funded multicenter observational study of childhood arterial ischemic stroke (the Vascular effects of Infection in Pediatric Stroke, or VIPS, study).[11] After distributing paper copies of the survey to 21 experts, the FOCAS PI (H.J.F.) verbally explained the intent of the survey and then guided the respondents through the individual questions. We then collected and compiled the survey responses, and used descriptive statistics to summarize the results.

Overview of FOCAS trial design: We consider a multi-site, two-arm randomized clinical trial in which patients are enrolled in each site, and patients are randomized to either an intervention arm or control arm in a fixed ratio 1:1. For each enrolled patient, FCASS measurements are collected at baseline and 1 month to assess potential benefit of the

intervention based on the change of FCASS measurements, and are followed by 3 months to assess adequacy of 1 month of therapy by identifying relapses. Also, infarct volume at 1 month, arteriopathy relapse, serious adverse events, and prevalence of acute herpesvirus infection will be measured as secondary outcomes. The primary objective is to determine whether corticosteroid intervention improves the change of imaging outcomes (FCASS measurements) between baseline and 1 month of children with symptomatic FCA compared to the control. Our design has interim analyses to monitor safety, futility and efficacy. The interim analysis is performed at frequent looks, i.e., after every 10 subjects complete 1 month. The trial may terminate early if the intervention is superior or futile compared to the control. The proposed Bayesian monitoring rule based on the change from baseline to 1 month in FCASS measurements yields type I/II error by rejecting/accepting the null hypothesis, which states that there is no difference between intervention and control, assuming the null hypothesis is true. The errors are controlled by the test boundaries calibrated from simulations under the Bayesian framework. In addition to the early stopping for efficacy and futility, we monitor the safety of intervention and allow early stopping of the trial due to toxic intervention. Moreover, the FOCAS trial design identifies optimal duration for the intervention at the end of trial by assessing the probability to relapse for subjects showing an initial improvement, i.e., that the FCASS is worsening at 3 month compared to 1 month.

Statistical models: Index patients and sites by $i = 1, \dots, n_j$ and $j = 1, \dots, J$. Let G_{ij} be an experimental condition indicator denoting E or C for intervention or control, respectively, which the ith subjects in the jth site are treated. Let Y_{ij} denote an outcome of the ith subject in the jth site for the observed change of FCASS measurement from baseline to 1 month (i.e., 1 month FCASS measurement -baseline FCASS measurement) and Z_{ij} denote a binary safety outcome of the ith subject in the jth site with $Z_{ij} = 1$ indicating that the ith subject in the jth site did not experience serious adverse events. We model the efficacy outcome as

$$Y_{ij} = \alpha_j + \theta_{G_{ij}} + \varepsilon_{ij},$$

where α_j describes the effect at site j, $\theta_{G_{ij}}$ characterizes the effect of treatment which the ith patient in the jth site will be assigned (i.e., $G_{ij} = E$ or C), and the error term ε_{ij} follows a normal distribution with mean 0 and variance σ^2. The parameters $\alpha_j, j = 1, \dots, J$, θ_E, θ_C and σ^2 are estimated under a Bayesian framework. We assume that $\alpha_j, j = 1, \dots, J$ are independently distributed from a normal variable with mean a and variance ξ^2 and the treatment effect θ_g, $g = E$ or C follows a normal distribution with the mean μ_g and variance η_g^2. We also assign σ^2 an inverse-gamma distribution with shape parameter w_1 and rate parameter w_1/w_2. For rare disease trials, since conventional vague prior distributions are often problematic for small samples, it is required to more care for the selection of hyperparameters - $a, \xi, \mu_g, \eta_g, \gamma_1$ and γ_2. Appendix A provides the details of prior specification with reasonable hyperparameters for rare disease trial. We next model the safety outcome for intervention as

$$Z_{ij} \sim Ber\left(\pi_{G_{ij}}\right),$$

where $\pi_{G_{ij}}$ denotes the safety probability for treatment G_{ij}. In this study, we are mostly concerned about the safety of the intervention group, but our proposed safety monitoring approach can be applied to both arms. This approach will be reviewed in addition to ongoing review by independent medical safety monitor. We assign π_E a non-informative beta prior with shape parameters $\phi\nu$ and $\phi(1-\nu)$. The selection of hyperparameters ϕ and ν are suggested in Appendix A. Prior publications of FCA treatment with corticosteroids have not reported adverse effects. [3, 4] Furthermore, we do not expect that FCASS would be measured differently if there is a safety issue. So, it is reasonable to assume that Y_{ij} is independent of Z_{ij}. If there was a prior belief that safety and efficacy are correlated, then we could add a term for safety effect on the FCASS measurements in Y_{ij}. Lastly, under the Bayesian framework, we generate the sample

from the posterior distribution, which is proportional to likelihood function multiplied by the prior distribution, using the Markov Chain Monte Carlo algorithm with Gibbs sampler.[12] We provide the detail of posterior distribution based on our statistical model in Appendix B.

Monitoring rules: The FOCAS trial design monitors efficacy, futility and safety. Suppose that the proposed design has $K-1$ interims and an interim analysis will be conducted after the enrollment of each cohort (e.g., every 10 subjects) to determine whether the trial should be terminated early for efficacy, futility and safety. The final analysis will be performed after a follow-up period of 3 months for the last subject. Let $Data_k$ denote the accumulating data at the kth analysis. To define monitoring rule, let δ denote the maximum of acceptable rate of any serious adverse event for intervention. The value of δ is pre-specified by the survey results for FOCAS trial design. Then, 1-δ denotes the minimum of acceptable rate to say the intervention is safe enough. Let c_1, c_2 and ε be pre-specified probability cutoffs obtained by preliminary simulation-based calibration. We searched the cutoffs in a range identified from the survey results to save several rounds of calibrations to obtain the target type I/II error rates. We provide a detailed explanation in Appendix C. For the interim analysis at $k = 1, \dots, K-1$, the decision rule are as follows:

1. Stop the trial for superiority of intervention over control if

$$P\{\theta_E < \theta_C | Data_k\} > c_1.$$

2. Stop the trial for futility of intervention over control if

$$P\{\theta_E < \theta_C | Data_k\} < c_2.$$

3. Stop the trial for safety if

$$P\{\pi_E \geq 1-\delta | Data_k\} < c_3.$$

If the trial is not stopped early, then at the last analysis (i.e., $k = K$), it is concluded that intervention is superior to control if

$$P\{\theta_E < \theta_C | Data_K\} > c_1.$$

Our monitoring rule follows an adaptive group sequential method in order to allow us to terminate the trial early for superiority, futility or safety based on the accumulating data.[13-16] The similar approach monitoring futility and toxicity was considered in Bayesian adaptive oncology clinical trials to identify a set of admissible doses – acceptably safe and efficacious doses- for the dose finding problem.[17-19] Also, the futility monitoring rule was considered a lot in phase II oncology trials based on the response probability.[20, 21] However, the adaptive sequential rule incorporating efficacy, futility and safety was never used for pediatric Strokes, and we propose a tailored monitoring rule for pediatric Stokes.

Consideration of duration of intervention: We define arteriopathy relapse in corticosteroid intervention arm as a worsening FCASS measurement from 1 month to 3 months after initial improvement (i.e., reduction in FCASS from baseline to 1 month). Let F_1 and F_3 denote the FCASS measurement at 1 month and 3 months, respectively. Let $d = F_1 - F_3$. If any 2 subjects with d<0 are observed, then the duration of corticosteroid intervention arm will be extended to 3 months. The primary outcome will not be changed. Because relapse is so rare, the intervention duration is likely to be extended if we observed a relapse in only 2 subjects.

Results:

Of 21 pediatric stroke investigators present at the VIPS investigator meeting, all (100%) participated in the survey. All respondents were pediatric stroke clinical experts; 20 were child neurologists and one was a pediatric hematologist. Seventeen (81%) were from U.S. institutions; three were Canadian and one was Swiss. The majority (13/21, 62%) currently treat

some cases of FCA with corticosteroids. Among those 13 who sometimes treat, the most common criterion for treatment was progressive FCA severity (11/13 respondents), followed by recurrent ischemic events (9/13 respondents). Regarding the statement, "I think corticosteroids most likely benefit children with FCA": 4% strongly agreed, 44% agreed, and 52% were neutral (none disagreed or strongly disagreed). Regarding the statement, "I think corticosteroids most likely harm children with FCA": 9% strongly disagreed, 74% disagreed, and 17% were neutral (none agreed or strongly agreed). All (100%) agreed they would be willing to enroll patients in a trial that randomized children to corticosteroids versus standard therapy.

If efficacy is measured by imaging biomarkers, the probability of efficacy that would lead an investigator to treat was a mean of 38% (median 30%; range 10% to 90%). If efficacy is measured by 12-month PSOM (neurological outcome), the probability of efficacy that would lead an investigator to treat was a mean of 34% (median 25%; range 10% to 90%). With either metric, only a single respondent indicated that they would require a probability of efficacy >80% in order to treat. The probability of efficacy (by either metric) that would make them unwilling to randomize because of efficacy (they would feel all children should receive the effective intervention) was a median 70%. The most common response was 80% (7 of 21 respondents), and only a single investigator (4.8% of 21 respondents) provided a response greater than 80%. Hence, we chose a threshold of c_1=81% for the probability of that corticosteroids are better than control. The probability of efficacy (by either metric) that would make them unwilling to randomize because of futility (they would feel no children should receive the futile intervention) was a mean of 17% (median 10%). Hence, we chose a threshold probability of efficacy of c_2=15% for our futility rule. The maximum acceptable proportion of severe/life-threatening adverse effects (assuming corticosteroids are effective) was a mean of 4% (range 1 to 15%). Hence we chose 8% as δ for our safety monitoring rule, and set c_3=20%.

Simulation study: We evaluated the operating characteristics of the FOCAS trial design using simulations. We assume enrollment occurred at 21 sites (i.e., 2 patients/site) that all 42 subjects reached the final visit. The maximum sample size 42 for a trial in simulation takes into account the dilution of the treatment effect due missing data from the planned sample size for FOCAS trial (N=50). Patients arrived according to a Poisson process with the accrual rate of 0.5 patients per year and were equally randomized to receive either intervention or control. Three interim analyses were planned when the first 10, 20, and 30 enrolled patients completed 1 month and the FCASS was measured. A final analysis was performed after the last patient completed follow-up. We generated the data according to the scenarios described in the first column of Table 1 with the expected change of FCASS measured between baseline and 1 month for control (Δ_C) and intervention (Δ_E). Based on FCASS data from the VIPS cohort, excluding two patients treated with IV steroids, the mean change in FCASS from baseline to 1 month was 3.8 (standard deviation=3) for the control group.[2] The simulation setting follows the same control group mean throughout and considers a range of effects for the intervention group with an equal variance for both groups. In other words, for the *i*th patient in site *j* which was assigned to intervention, the change of FCASS between baseline and 1 month follows a normal distribution with mean Δ_E and standard deviation 3. Serious adverse events (SAE) were generated from a Bernoulli distribution with a rate of 4%.

We set the type I error rate $\alpha = 0.2$ and power 80% to detect the difference in mean change of 1.6. Using an empirical approach to calibrate the cutoffs for the Bayesian monitoring rule (see Appendix C), $c_1 = 0.81$ and $c_2 = 0.15$ were identified in order to achieve our desired type I and II error rates. In addition, we identified c_3=0.2 as an appropriate safety monitoring cutoff (figure 1). Table 1 shows the results based on 10,000 simulations. The FOCAS trial design preserved the overall type I error rate at the level of 0.195 (see the second scenario) and yielded power 81.3% when the difference in mean change of FCASS between corticosteroids and

control was 1.6 (i.e., the fourth scenario). The proposed design is compared with a conventional design using a one-sided two-sample t-test statistics with a level of significance 0.2 at the end of the trial. The conventional design has no interim monitoring, and the total planned number of patients are required to be enrolled. However, our proposed design identifies a treatment effect earlier and we save 30-55% of patients. This is a remarkable benefit of the proposed design especially for a rare disease.

Sensitivity analysis: We evaluated the proposed design for the early stopping when the SAE rate varies with 20% and 25%, which are larger than the threshold δ= 0.08, to consider a case where the rate is beyond the maximum of survey results accepting the safety issues. We compared the results with the SAE rate 4%. The left panel of Figure 1 shows that the trial stops earlier for high SAE rate with 73-86%, because it clearly detects serious events from 8%. It saves more than 50% of patients from the toxic drug. Also, when the drug is safe (i.e., the SAE rate is 4%), we almost never stopped early and our trial avoids erroneous decision for safety monitoring. The right panel of Figure 1 shows the results of the FOCAS trial design when there is a site effect or when the site effect is more variable compared to the fourth scenario of Table 1. Note that the site effect was generated from a normal distribution with mean 0 and standard deviation τ, and the FCASS change is now generated from $Y_{ij} \sim N(\Delta_{G_{ij}}, \tau^2 + 3^2)$, where G_{ij} is either E or C. Since the sample size was calculated with τ^2=0, the rejection probability which indicates the power under the fourth scenario in Table 1 becomes smaller as the variability of the site effect increases. However, the mean sample size is mostly similar regardless of the variability of the site effect. Moreover, we investigated the operating characteristic of the proposed design when the errors are not normally distributed. We generated the data under the same setting as Table 1, but with two different non-normal error distributions, including t distribution with degrees of freedom 2.25 and uniform distribution on interval (-5.2, 5.2). Those parameters are selected to obtain the same variability as Table 1 but the distribution has fat-tail

or flat-tail, respectively. The results are presented in Table 2. We notice that a moderate departure from normality does not cause a big difference in the operating characteristics. Both type I and II errors are preserved at the nominal level.

Discussion:

We present a pragmatic, adaptive clinical trial design, with criteria for trial success based on the level of evidence needed to impact clinical practice. We have demonstrated the trial operating characteristics under a variety of assumptions.

In the VIPS observational study, the well-performing sites enrolled an average of one FCA case per site per year.[7] If enrollment in FOCAS, as an interventional trial, is half of enrollment in VIPS, then patients would enroll 0.5 cases/site/year, or 2 cases/site over a 4-year enrollment period. We need 25 sites to enroll 50 patients over four years. The proposed design allows for a feasible study, accomplishable within the typical 5 year NIH funding period.

We propose that several factors make this pragmatic approach reasonable for the FOCAS trial design: (1) FOCAS will study standard doses of old medications with well-established safety profiles commonly used for other pediatric disorders. (We will not seek a new FDA label.) (2) Alternative therapies for FCA do not exist, and untreated FCA often progresses. (3) Most pediatric stroke experts already employ corticosteroids for select cases of FCA, and perceive it to be safe. Hence, to be convinced to treat FCA with corticosteroids, clinicians require a lower than usual level of confidence that the treatment is effective.

We also propose that this approach is more ethical, minimizing the number of children who would ultimately receive the inferior therapy. We anticipate that this assurance—that we will minimize unnecessary randomization—will improve participation in the clinical trial. It would be less likely to enroll children into a trial designed to prove efficacy at a higher level than what they feel necessary for clinical practice.

Conclusions: Other rare diseases may be similar to FCA in that clinicians would be comfortable with a lower level of confidence regarding treatment efficacy because of either the severity of disease or the safety of the intervention. This approach to trial design— applying the probability of efficacy needed to change clinical practice based on expert consensus and incorporating frequent opportunities to stop early—may help make the study of interventions for rare disease more feasible.

References:

1. Steinlin M, O'Callaghan F, Mackay MT. Planning interventional trials in childhood arterial ischaemic stroke using a Delphi consensus process. Dev Med Child Neurol. 2017 Jul;59(7):713-8.

2. Fullerton HJ, Stence N, Hills NK, Jiang B, Amlie-Lefond C, Bernard TJ, Friedman NR, Ichord R, Mackay MT, Rafay MF, Chabrier S, Steinlin M, Elkind MSV, deVeber GA, Wintermark M, Investigators V. Focal Cerebral Arteriopathy of Childhood: Novel severity score and natural history. Stroke. 2018 Nov;49(11):2590-6.

3. Steinlin M, Bigi S, Stojanovski B, Gajera J, Regenyi M, El-Koussy M, Mackay MT. Focal Cerebral Arteriopathy: Do Steroids Improve Outcome? Stroke. 2017 Jul 21;48(9):2375-82.

4. Elbers J, Armstrong D, Yau I, Benseler S. Vascular Imaging Outcomes of Childhood Primary Angiitis of the Central Nervous System. Pediatr Neurol. 2016 Oct;63:53-9.

5. Lambertsen KL, Finsen B, Clausen BH. Post-stroke inflammation-target or tool for therapy? Acta Neuropathol. 2018 Nov 27.

6. Agrawal N, Johnston SC, Wu YW, Sidney S, Fullerton HJ. Imaging data reveal a higher pediatric stroke incidence than prior US estimates. Stroke. 2009 Sep 17;40(11):3415-21.

7. Wintermark M, Hills NK, DeVeber GA, Barkovich AJ, Bernard TJ, Friedman NR, Mackay MT, Kirton A, Zhu G, Leiva-Salinas C, Hou Q, Fullerton HJ. Clinical and Imaging Characteristics of Arteriopathy Subtypes in Children with Arterial Ischemic Stroke: Results of the VIPS Study. AJNR Am J Neuroradiol. 2017 Nov;38(11):2172-9.

8. Avins AL. Can unequal be more fair? Ethics, subject allocation, and randomised clinical trials. J Med Ethics. 1998 Dec;24(6):401-8.

9. Miller F, Zohar S, Stallard N, Madan J, Posch M, Hee SW, Pearce M, Vagero M, Day S. Approaches to sample size calculation for clinical trials in rare diseases. Pharm Stat. 2018 May;17(3):214-30.

10. Hansson L, Hedner T, Dahlof B. Prospective randomized open blinded end-point (PROBE) study. A novel design for intervention trials. Prospective Randomized Open Blinded End-Point. Blood Press. 1992 Aug;1(2):113-9.

11. Wintermark M, Hills NK, deVeber GA, Barkovich AJ, Elkind MS, Sear K, Zhu G, Leiva-Salinas C, Hou Q, Dowling MM, Bernard TJ, Friedman NR, Ichord RN, Fullerton HJ. Arteriopathy diagnosis in childhood arterial ischemic stroke: results of the vascular effects of infection in pediatric stroke study. Stroke. 2014 Dec;45(12):3597-605.

12. Robert C, Casella G. Monte Carlo statistical methods: Springer Science & Business Media; 2004.

13. Posch M, Bauer P. Adaptive Two Stage Designs and the Conditional Error Function. Biometrical Journal. 1999;41(6):689-96.

14. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. Biometrics. 1999 Dec;55(4):1286-90.

15. Muller HH, Schafer H. Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. Biometrics. 2001 Sep;57(3):886-91.

16. Liu Q, Proschan MA, Pledger GW. A Unified Theory of Two-Stage Adaptive Designs. Journal of the American Statistical Association. 2002 2002/12/01;97(460):1034-41.

17. Thall PF, Russell KE. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. Biometrics. 1998 Mar;54(1):251-64.

18. Yin G, Li Y, Ji Y. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. Biometrics. 2006 Sep;62(3):777-84.

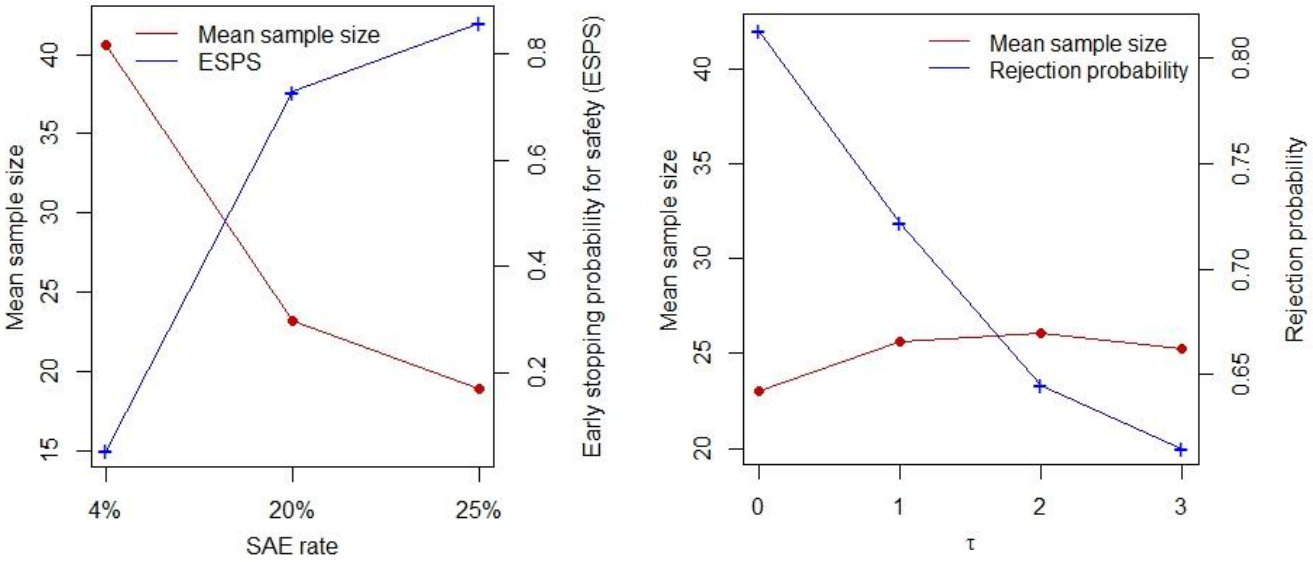
19. Guo B, Yuan Y. Bayesian Phase I/II Biomarker-Based Dose Finding for Precision Medicine With Molecularly Targeted Agents. Journal of the American Statistical Association. 2017 2017/04/03;112(518):508-20.

20. Simon R. Optimal two-stage designs for phase II clinical trials. Controlled clinical trials. 1989 Mar;10(1):1-10.

21. Bryant J, Day R. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. Biometrics. 1995 Dec;51(4):1372-83.

Figure:

Figure 1. *Plot of sensitivity analysis for serious adverse event (SAE) rate (left panel) and variability of site effect (right panel) when the difference of FCASS measured between corticosteroids and control was -1.6 (i.e., the fourth scenario).*



Tables:

Table 1. *Results for FOCAS trial design under five scenarios.*

Scenario (Δ_C, Δ_E)	Rejection probability		Mean sample size		Early stopping probability	
	FOCAS	Conventional	FOCAS	Conventional	superiority	futility
(3.8, 4.0)	0.141	0.147	35.62	42	0.123	0.136
(3.8, 3.8)	0.195	0.2	35.15	42	0.168	0.108
(3.8, 2.8)	0.589	0.582	28.68	42	0.504	0.027
(3.8, 2.2)	0.813	0.799	23.03	42	0.721	0.009
(3.8, 1.8)	0.912	0.896	19.40	42	0.842	0.004

Table 2. *Sensitivity analysis results for FOCAS trial design under non-normal error distribution.*

Error distribution	Scenario (Δ_C, Δ_E)	Rejection probability	Mean sample size	Early stopping probability for	
				superiority	futility
Fat-tailed (t distribution)	(3.8, 4.0)	0.051	39.89	0.043	0.048
	(3.8, 3.8)	0.084	39.69	0.067	0.032
	(3.8, 2.8)	0.650	29.13	0.538	0.004
	(3.8, 2.2)	0.920	19.02	0.858	0.002
	(3.8, 1.8)	0.975	14.75	0.946	0.001
Flat-tailed (uniform)	(3.8, 4.0)	0.144	35.39	0.126	0.142
	(3.8, 3.8)	0.194	35.03	0.168	0.114
	(3.8, 2.8)	0.590	28.71	0.505	0.028
	(3.8, 2.2)	0.809	23.21	0.721	0.011
	(3.8, 1.8)	0.909	19.70	0.836	0.005

Online Supplemental Material: Appendix

A. Prior specification.

A simple linear model is used to model the efficacy outcome Y_{ij} with the site effect $\alpha_j, j = 1, \dots, J$. For computational simplicity, we take $\xi = \sigma\tau$ and $\eta_g = \sigma_g s_g$ for $g = E$ or C. This choice brings a normal-inverse-gamma conjugate prior described by

$$\alpha_j|\sigma^2 \sim N(a, \sigma^2\tau^2)$$

$$\theta_{G_{ij}=g}|\sigma^2 \sim N\left(\mu_g, \sigma^2 s_g^2\right), \quad g = E \text{ or } C$$

$$\sigma^2 \sim IG(w_1, w_1/w_2),$$

where $a, \tau^2, \mu_g, s_g^2, w_1$, and w_2 are hyperparameters. Since all FCASS measurements obtained from multiple sites will be scored centrally, it is reasonable to assume small variability for the site effect. So, we assign the site effect parameter α_j a normal prior distribution with $a = 0$ and $\tau^2 = 0.1$. Conditional prior distributions of treatment effect θ_E or θ_C given σ^2 are specified by historical data and the regularized vague prior approach using the fact that any change in an input variable within two standard deviation below the mean and two standard deviation above the mean most likely results in the difference of response variable. This approach would provide more reliable inference for a rare disease by assigning prior distributions which are vague enough to cover the plausible values of the parameter. Based on the VIPS data, we obtained $\mu_C = 3.8$ for prior mean and $\sigma = 3$. Also, the FCASS change from baseline to 1 month ranges between 0 and 9. Thus, $s_E = s_C = 1$ are specified under the equal variance assumption. The prior distribution of σ^2 is the inverse gamma distribution with shape parameter w_1 and rate parameter w_1/w_2. The prior mean of $1/\sigma^2$ is w_2 and the prior effective sample size is w_1, which is an intuitive measure for the amount of information provided by the prior (Morita et al., 2008). The prior effective sample size is generally small to ensure that the prior does not

provide an inappropriate amount of information. Here, we take $w_1 = 1/2$. The value of w_2 is chosen by the historical data, i.e., $w_2 = 1/3^2$. Similarly, for the safety outcome, the prior mean of π_E is ν and the prior effective sample size is ϕ. We use default value of $\phi = 1$. From the physicians' expert knowledge, ν is chosen by the expected probability of serious adverse event for intervention 0.04. Thus, we have $\pi_E \sim Beta(0.04, 0.96)$.

B. Posterior distribution.

The FOCAS trial design uses a conjugate prior for the likelihood function for the observed data under the Bayesian framework. This provides the closed form of posterior distribution for parameters and makes our design simple and convenient for inference. Under the situation where the site effect and treatment effect are independent, we combine $\alpha_1, \dots, \alpha_J$, $\theta_E I(G_{ij} = E)$ and $\theta_C I(G_{ij} = C)$ into a vector and consider the corresponding design matrix $\boldsymbol{X}_t$ and covariance matrix $\boldsymbol{V}_t$. Let $c = w_1 + n/2$, $e = w_1/w_2 + (\boldsymbol{\mu}_t^T \boldsymbol{V}_t^{-1} \boldsymbol{\mu}_t + \boldsymbol{y}^T \boldsymbol{y} - \boldsymbol{\mu}_0^T \boldsymbol{V}_0^{-1} \boldsymbol{\mu}_0)/2$, where

$$\boldsymbol{\mu}_0 = (\boldsymbol{V}_t^{-1} + \boldsymbol{X}_t^T \boldsymbol{X}_t)^{-1} (\boldsymbol{V}_t^{-1} \boldsymbol{\mu}_t + \boldsymbol{X}_t^T \boldsymbol{y})$$

$$\boldsymbol{V}_0 = (\boldsymbol{V}_t^{-1} + \boldsymbol{X}_t^T \boldsymbol{X}_t)^{-1}$$

Then,

$$\{\alpha_1, \dots, \alpha_J, \theta_E I(G_{ij} = E), \theta_C I(G_{ij} = C)\}^T | \boldsymbol{y} \sim MVt_\nu(\boldsymbol{\mu}_0, (e/c)\boldsymbol{V}_0)$$

$$\sigma^2 | \boldsymbol{y} \sim IG(c, e)$$

$$\pi_{G_{ij}} | \boldsymbol{z} \sim Beta(\phi\nu + \sum_i z_{ij}, \phi(1-\nu) + \sum_i (1 - z_{ij})),$$

Where $MVt_{\nu}(\boldsymbol{A}, \boldsymbol{B})$ denotes a multivariate t distribution with degrees of freedom ν, location vector $\boldsymbol{A}$ and shape matrix $\boldsymbol{B}$, $\nu = 2c$ and z_{ij} denotes the number of observed serious adverse event.

C. Calibration of cutoffs for Bayesian sequential monitoring rule

The proposed Bayesian sequential monitoring rule involves three design parameters c_1, c_2 and c_3, where c_1 controls type I error rate, which occurs from superiority stopping when the null is true, and c_2 controls type II error rate, which occurs from futility stopping under the alternative. The value of c_3 relates to the overall stopping rate due to safety, which does not matter with type I or II error for hypothesis testing. Our proposed design is an adaptive design and it is not possible to analytically calculate the type I and II error rates. Rather than analytic errors, we commonly determine the empirical type I and II error rates from simulation. In this setting, we assign the initial values of c_1 and c_2, and perform simulation to calculate the empirical type I and II error rates. Based on our experiences, the reasonable initial values of c_1 and c_2 are $1 - \alpha$ and power, respectively. The survey results also suggests to search the values of c_1 and c_2 over the range >=80% and <17%, respectively. The suggestion will save several rounds of calibrations to obtain the target type I and II error rates as well as make sense with experts' experience and knowledge. If the type I error rate is lower/higher than the desirable level, we decrease/increase the value of c_1, and if the calculated type II error rate is lower/higher than the desirable level, we decrease/increase the value of c_2. We repeat this calibration process until the desirable type I and II error rate are obtained. Similarly, we vary the values of c_3 until the desirable early stopping probability for safety is obtained.

Reference for Online Material:

Morita, S., Thall, P. F., & Müller, P. (2012). Prior effective sample size in conditionally independent hierarchical models. Bayesian analysis (Online), 7 (3).