Adalimumab in Juvenile Idiopathic Arthritis-associated Uveitis Stopping Trial (ADJUST)

Statistical Analysis Plan

Confidential

Version 5, February 14, 2024
Trial Registration Number: NCT 03816397

Prepared by: Ben Arnold, PhD, MPH, Travis Porco, PhD, MPH
F.I. Proctor Foundation
University of California, San Francisco
Coordinating Center
Email: ben.arnold@ucsf.edu

Table of Contents

1 Introduction

This Statistical Analysis Plan (SAP) describes the planned analyses for the ADalimumab in Juvenile idiopathic arthritis-associated Uveitis Stopping Trial (ADJUST). It includes specifications for the statistical analyses and the tables to be prepared for the interim analyses and final Clinical Study Report.

This SAP conforms to the guidelines for Statistical Analysis Plans released by the Society for Clinical Trials and is in accordance with recently published "Guidelines for the Content of Statistical Analysis Plans in Clinical Trials" in JAMA.[1] The background for this SAP can be found in the ADJUST Manual of Operations (MOP).

The planned analyses in this SAP will be included in future manuscripts. Exploratory analyses not expressly identified here may still be performed but will be clearly documented as such in the final Clinical Study Report.

The final SAP is subject to the approval of an appointed Data and Safety Monitoring Committee.

2 Investigational Plan

2.1 Study Design

This study is a multicenter, randomized, double-masked, Phase IV superiority clinical trial to evaluate the efficacy of adalimumab withdrawal in patients with controlled JIA-associated uveitis.

2.2 Study Population

Eligible patients will have controlled JIA-associated uveitis or chronic anterior uveitis (CAU) on adalimumab with no other suspected etiology of uveitis. Complete inclusion and exclusion criteria are given in ADJUST Manual of Operations Section 3.3.[2]

2.3 Specific Aims

2.3.1 Specific Aim 1

Primary Outcome of the trial. The primary outcome for Specific Aim 1 will be time to treatment failure, with censoring at 12 months. Treatment failure occurs when recurrence of ocular inflammation is observed and/or the recurrence of joint inflammation is persistent and severe enough to necessitate unmasking to manage the arthritis recurrence. The recurrence of ocular inflammation is defined as one or more of the features in Table 1 in at least one eye:

Table 1: Criteria for Treatment Failure by Uveitis Recurrence
(at least one of the following in at least one eye)

Parameter	Definition
Anterior chamber cells*	≥2-step increase** at two separate visits ≥7 days apart >0.5+ cell observed for ≥28 days ≥3+ cell observed at a single visit
Vitreous haze***	>0.5+ haze at a single visit
Choroid & retina	Active retinal / choroidal lesions or macular edema (CST >2 standard deviations above normal thickness or cysts in 1mm central subfield) at a single visit

*SUN critera[3], **increase is based on inflammation at baseline; ***NEI vitreous haze grading scale[4]

Secondary Objectives

- To compare the proportion of people exhibiting treatment failure between the randomization groups during the first 6 months and during the first 12 months of study duration.
- To compare the cost-effectiveness of adalimumab therapy between the randomization groups during 12 months.

Secondary Objectives with Alternative Outcomes

We propose to compare longitudinal values of specific variables between the two randomization groups. The following outcomes will be compared:

- Best Corrected Visual Acuity (BCVA) score,
- Childhood Health Assessment Questionnaire score (CHAQ)[5] in patients ages 5+,
- EuroQol 5 Dimension Youth Survey (EQ-5D-Y)[6] score in patients ages 4+,
- Children's Visual Functioning Questionnaire (CVFQ)[7] score in children aged 3 to 7 years,
- Effects of Youngster's Eyesight on Quality of Life (EYE-Q)[8] score in children ages 5+,
- Juvenile Arthritis Disease Activity Score (JADAS)[9] values: JADAS-10 and JADAS-27,
- Estimates of anti-adalimumab antibody (ADA) levels.

We also propose to compare proportions between the two randomization groups. These proportions correspond to the time until treatment failure or censoring time of 6 months as well as until treatment failure or censoring time of 12 months. Specifically:

- The proportion of patients who exhibit an arthritis flare, as defined in the Manual of Operations Section 6.4.1.[2]
- The proportion of patients who achieve an improvement in JADAS-10 as defined by Horneff et al.[9] The criteria for a significant improvement is as follows:
    - For patients with a baseline JADAS-10 value ≤15, a decrease of 4.
    - For patients with a baseline JADAS-10 value >15 and ≤25, a decrease of 10.
    - For patients with a baseline JADAS-10 value >25, a decrease of 17.
- The proportion of patients exhibiting macular edema.

2.3.2 Specific Aim 2

Primary Outcome: The principal outcome for Specific Aim 2 will be occurrence of

treatment failure (as defined in Section 2.3.1) in the first 6 months and in the first 12 months.

Secondary Objectives

- To determine the predictors of uveitis recurrence (treatment failure by ocular inflammation recurrence as defined in Table 1)
- To determine the predictors of time to treatment failure (as defined in Section 2.3.1) with censoring at 6 months and with censoring at 12 months,
- To determine the predictors of BCVA at 6 and 12 months,
- To determine the predictors of arthritis flare occurrence (as defined in MOP Section 6.4.1) in the first 6 months and in the first 12 months,
- In patients ages 5+, to determine the predictors of CHAQ score at 6 and 12 months.

Supplemental Objectives

Additional exploratory analyses will be conducted to determine predictors of additional variables:

- Number of missed treatment doses during the first 6 months, and during 12 months,
- CVFQ score in children aged 3 to 7 years at 6 and 12 months,
- EYE-Q score in children ages 5+ at 6 and 12 months,
- JADAS-10 and JADAS-27 scores at 6 and 12 months,
- Achievement of improvement in JADAS-10 (MOP Section 6.4.1) over the first 6 months, and over the first 12 months,
- Occurrence of macular edema during the first 6 months and during the first 12 months.

2.3.3 Specific Aim 3

Primary Outcome: The primary outcome for Specific Aim 3 is achievement of corticosteroid-sparing controlled ocular inflammation at months 6 and 12 following enrollment. Corticosteroid-sparing controlled ocular inflammation is defined by meeting all of the criteria in Table 2 in both eyes:

Table 2: Criteria for Corticosteroid-sparing Controlled Ocular Inflammation
(both eyes have to meet all criteria)

Parameter	Definition
Anterior chamber cells*	≤0.5+ cell
Vitreous haze**	≤0.5+ haze
Choroid & retina	No active retinal / choroidal lesions or macular edema (CST >2 standard deviations above normal thickness or cysts in 1mm central subfield)
Prednisolone acetate 1%	≤2 drops/day

*SUN criteria[3]; **NEI vitreous haze grading scale[4]

Secondary Objectives

- To compare BCVA at 6 months and 12 months between randomization groups,
- To compare the proportion of people exhibiting an arthritis flare (as defined in MOP Section 6.4.1) between randomization groups during the first 6 months and during the first 12 months,
- For children ages 5+, to compare the CHAQ score at 6 months and at 12 months between randomization groups.

Supplemental Objectives

- To determine the predictors of regaining control of uveitis (Table 2) at 12 months in patients randomized to placebo who have a uveitis flare (Treatment Failure by ocular inflammation recurrence as defined in Table 1),
- To compare the number of missed treatment doses over 6 months and over 12 months between randomization groups.

We propose to compare final values of specific variables between the two randomization groups. Final values correspond to the value at 6 months as well as at 12 months. We will compare the following variables:

- EuroQol 5 Dimension Youth Survey (EQ-5D-Y)[6] score in children ages 4+,
- Children's Visual Functioning Questionnaire (CVFQ)[7] score in children aged 3 to 7 years,
- Effects on Youngster's Eyesight on Quality of Life (EYE-Q)[8] score in children

ages 5+,
- Juvenile Arthritis Disease Activity Score[10] values: JADAS-10 and JADAS-27,
- Estimates of anti-adalimumab antibody (ADA) levels.

We also propose to compare proportions between the two randomization groups. These proportions correspond to the time until 6 months as well as until 12 months. The proportions of the following variables will be compared:
- The proportion of patients who achieve an improvement in JADAS-10 scores
- The proportion exhibiting macular edema
- The proportion of people who discontinued treatment due to intolerability
- The proportion of people who discontinued treatment due to safety

2.4 Randomization

2.4.1 Stratification

We propose stratification by country and use of antimetabolite medication. Within each stratum, assignments will be conducted using a randomly permuted block randomization scheme.

Patients will be enrolled in three countries: United States, United Kingdom and Australia. Patients will be recruited from fifteen to twenty study sites, with the possibility for sites to be added to meet the *fixed* sample size requirements. Patients will be randomized in a 1:1 ratio to two interventions: adalimumab or matched placebo. The treatment protocols are specified in the ADJUST Manual of Operations.

Stratification will be done by antimetabolite drug use based on a belief by the study team that concomitant antimetabolite drugs may have a differing impact on treatment outcomes of uveitis patients using TNF-α inhibitors.[11] Similarly, variations in the baseline demographics within each country may impact treatment outcomes with adalimimuab.

2.4.2 Randomization list

Lists of sequential randomization assignments will be prepared for each country. The randomization lists consist of a unique identifier for each patient, together with the assignments to treatment groups. The assignment of patient ID numbers and treatment randomization will be performed at enrollment. Research pharmacists and emergency contacts will be the only study personnel with access to treatment assignments for currently enrolled patients. These study personnel will not have access to treatment assignments for sites outside their country. The research pharmacist will perform each assignment for their site, and will not know treatment assignments prior to enrolling each new patient. At the time of enrollment, the research pharmacist will log into the online

randomization module and obtain the next assignment from the sequential list. If the site does not have a research pharmacist, the site will identify someone not otherwise involved in the study to perform this step.

2.4.3 Block randomization

This study will use randomly permuted block randomization to help to mask the treatment assignment. Blocks of size 2 and 4 will be used, with probabilities 0.7 and 0.3, respectively. The study will generate a separate set of block allocations for each stratum (country and use of antimetabolite medication at enrollment). The statistics package R will be used to generate a random permutation of assignment orders. Note that the same algorithm will be used in all permutation tests.

2.4.4 Unique patient identifiers

Unique patient IDs will be generated. The first two characters will be numbers. The next character is a check sum character, which will be a single letter. The last three characters will be sequential digits beginning at 001. An example is 02B025; all identifiers have exactly six characters.

2.4.5 Random number generation

The sequence of random numbers depends on the choice of a numerical seed. This seed will be chosen by the principal statistician and used to create the randomization assignments, and concealed. The procedure in Porco, Stoller, Keenan et al. 2015 will be used.[12]

2.4.6 Provision of randomization list

Only after a patient has been enrolled will REDCap (Research Electronic Data Capture)[13] allow an intervention assignment.

2.4.7 Summary of disposition of randomization list

The research pharmacist and emergency contacts will receive access to the REDCap randomization module for their country. An encrypted electronic copy will be preserved

on cloud storage.

2.5 Masking

The ophthalmologists, rheumatologists, visual acuity (VA) examiners, optical coherence tomography (OCT) operators, clinic coordinators, refractionists, and patients and their caregivers will be masked to drug assignment. A placebo will facilitate masking. Full procedures for how masking will be maintained is found in the Manual of Operations.

2.6 Baseline reporting

2.6.1 Demographics and patient history

All demographic and history variables (age, sex, race, ethnicity, etc) determined at enrollment will be summarized by counts and percentages tabulated by treatment assignment and by site. We will similarly aggregate and summarize patient characteristics by country. Such comparisons are not hypothesis tests, but will be provided to the DSMC as descriptive information.

2.6.2 Prior and concurrent medication

We will present the topical corticosteroid baseline doses, the antimetabolite baseline doses, adalimumab doses and other baseline medications by randomization arm, study site and country. Prior medications for JIA or JIA-associated uveitis that are not curently being taken will also be presented.

2.6.3 Baseline comorbidities and history

Clinical variables at baseline will be presented by randomization arm, study site, and country. These clinical variables include baseline BCVA, baseline intraocular pressure, baseline anterior chamber cells (0 or 0.5+), baseline anterior chamber flare, as well as baseline JADAS-10 and -27 scores. Other baseline variables that will be summarized include age at diagnosis of JIA, age at diagnosis of uveitis, duration of adalimumab treatment, duration of uveitis inactity, arthritis subtype, and ADA positivity. We will present the median value for each continuous clinical variable and frequency for categorical variables, stratifying by randomization group, site, and country.

3 Statistical Considerations

3.1 Analysis

3.1.1 Specific Aim 1

Primary Analysis

The prespecified primary analysis for Specific Aim 1 is Cox proportional hazards regression. The outcome variable is time to treatment failure, as defined in Section 2.3.1.

- The primary covariate of interest is the randomization group.
- Time to treatment failure is measured in the number of days since patient randomization.
- The parameter of interest is the hazard ratio of treatment failure comparing placebo (stopping) to adalimumab. We will estimate the hazard ratio and its 95% confidence interval using a fixed effect for treatment assignment in the Cox proportional hazards model.
- Hypothesis testing will be conducted by Monte Carlo permutation testing (see below for details).
- Tests will be two-sided, at prespecified type I error probability alpha=0.05.
- Country and antimetabolite use will be included in the model as additional fixed effects, reflecting the stratified randomization.

Monte Carlo permutation testing. Hypothesis testing will be conducted using Monte Carlo permutation testing of the log hazard ratio coefficient estimated using the pre-specified primary analysis Cox proportional hazard model, based on 100000 replications (throughout this Statistical Analysis Plan). If the decision boundary lies within two standard deviations of the Monte Carlo estimate of the P-value, we will conduct one million replications and report the final value.[14]

Country. Any country with fewer than 5 enrollees will be aggregated with a similar country using a prespecified plan. Provided that we have fewer than 5 enrollees in Australia, we will aggregate them with the UK. Provided that we have fewer than 5 enrollees in the UK, we will aggregate them with Australia. If two countries experience fewer than 5 enrollees each, the US, the UK and Australia will be aggregated together.

Subsidiary Analyses of the Primary Model
Subsidiary analyses, or methodological sensitivity analyses, are conducted to assess the degree to which the choice of analytic method may have influenced the reporting of

the findings. Specifically, we mean: (a) choice of test statistic, (b) choice of estimation method, and (c) inclusion of covariates reflecting treatment site.

Such analyses will always be sharply distinguished from the primary prespecified analysis. Such analyses do not address new scientific questions. All such sensitivity analyses are designed to ensure that the findings of our trial have not been unduly influenced by the choice of statistical model.

- Heterogeneity between sites in this multicenter trial will be assessed in conformity with the ICH guidelines which state: "Marked heterogeneity may be identified by graphical display of the results of individual centres or by analytical methods, such as a significance test of the treatment-by-centre interaction. When using such a statistical significance test, it is important to recognise that this generally has low power in a trial designed to detect the main effect of treatment."[15]

  Rationale. Assessment of the role of country and site in this multicenter trial is complicated by the fact that we anticipate many sites, each contributing a relatively small number of enrollees (as indicated above). Moreover, some countries may contribute few enrollees altogether. Heterogeneity between countries and sites within countries may arise for many reasons, and we note that the characterization of such heterogeneity is not the scientific purpose of the trial. Our goal is to provide insight into whether or not such heterogeneity may have affected our conclusions.

  We note that examination of the collection of possible models below will be conducted by an analyst masked to the identity of each group. Each of these models will be fit, and all models that we fit will be reported to the DSMC and made available in supplemental material.

  Procedures. Summaries of treatment effect stratified by country will be provided. We will tabulate and report times to treatment failure by country.

  Although the exact number of sites has not been fixed, it is estimated that the study will enroll at approximately 15 to 20 sites (rendering site as a fixed effect is undesirable). The number of sites is expected to vary based on enrollment needs—changes in the number of enrollment sites will be based on the number of enrolled patients *alone* (i.e., completely masked to outcome). We emphasize that the trial is designed with a *fixed* sample size.

  Alternative models with interaction: tests of heterogeneity of effect. Consistency of effect between countries will be investigated. We propose to examine **homogeneity of treatment effect across countries** (by fitting survival models including country as a fixed effect along with treatment-country *interactions*). For the primary model and the aforementioned additional analyses of the primary

model, each will be repeated with an interaction term for country. All such analyses will be made available.

- As a sensitivity analysis, we will exclude treatment failures due to persistent and severe arthritis recurrence and assess the primary model with the recurrence of ocular inflammation only.
- We will examine treatment effect modification by disease type (JIA-associated uveitis vs. CAU) by fitting survival models including disease type as a fixed effect along with a treatment by disease type interaction. A further sensitivity analysis excluding individuals with CAU will be conducted. We acknowledge that these additional analyses may have low power since the trial is powered for main effects (not interaction) and excluding CAU patients will reduce the sample size. These analyses will also be made available.
- **Per-protocol.** We will conduct and report a per-protocol analysis excluding patients who missed 20% or more of their study medication.
- Our design yields failure times at discrete study visits. Statistically, we consider the failure time to occur at the first visit at which it is documented. A sensitivity analysis will be conducted in which we **impute the failure time** at the midpoint of the interval beginning with the last non-failure visit and the first failure visit.

Planned Secondary Analyses

Rationale. The proposed secondary analyses are intended to provide additional scientific information not present in the primary outcome and its analysis. These are of the following types: (1) use of the same outcome variable as the primary outcome, but at (a) different time points or (b) modeled by different covariates, and (2) use of different outcome variables.

- **The time to treatment failure, with censoring at 6 months,** a time to event outcome, will be analyzed using the template in the Primary Analysis above. This is the same outcome as for the primary outcome variable, except with observations restricted to 6 months.
- **The proportion of people with treatment failure over 6 months and 12 months**, a binary outcome, will be modeled using log binomial models to report relative risks. We will use the randomization group as a predictor. Clustered log-binomial regression will then be used to estimate the same regression model.
- **A decision tree model-based cost-effectiveness analysis will be conducted to assess adalimumab therapy.** The cost-effectiveness analysis will be conducted separately for the US, UK, and Australia. Healthcare costs will be gathered from appropriate US, UK, and Australia sources under the perspective of a third-party payer. In the US, costs will be estimated from the Medicare Physician Reimbursement Rates and the Red Book average wholesale prices. In the UK, costs will be estimated from the National Health System (NHS) reference costs published by the Department of Health. In Australia, costs will be estimated from the Medicare Benefits Scheme (MBS) reference costs. Each item will be multiplied by the unit cost associated with the item. We will consider the following

costs: (1) doctor visits, (2) procedural costs (e.g., OCT scans, laboratory tests, etc.), and (3) medications. Total costs per patients will be summarized in 2019 USD. Protocol-induced costs will not be included in the total patient costs as these are imposed by the trial protocol. We will use the following two measures of effectiveness: (1) time to inflammation recurrence during the 12-month study period and (2) gain in quality-adjusted life years (QALYs) over 12 months. Time to inflammation, or time to treatment failure, is defined in Section 2.3. The EQ-5D-Y from baseline, month 6, and month 12 will be used to calculate utility values ranging from 0 (equivalent to being dead) to 1 (perfect health). We will calculate the number of QALYs gained for each patient by estimating the area under the curve (AUC) using the three utility value estimates (e.g., baseline, month 6, and month 12).

We will use the TreeAge Pro software to develop our decision tree model. Our model will contain one decision node to stop or continue adalimumab. To calculate the incremental cost effectiveness ratio (ICER) we will divide the mean difference in patient costs between stopping and continuing adalimumab by the mean difference in effects.[16,17] As costs will be different for our three enrollment countries we will create separate decision tree models. The ICERs will be interpreted as the following: (1) the costs per inflammation-free years and (2) the costs per QALY gained. Non-parametric bootstrap statistical methods will be used to calculate confidence intervals and derive a cost-effectiveness acceptability curve.

Planned Secondary Analyses with Alternative Outcomes

The following additional analyses provide further information about the disposition of study subjects. In reporting, these will be sharply distinguished from the primary outcome. We propose to report 95% confidence intervals for estimates pertinent to the analyses described below. Note, these are of less scientific importance than the primary outcome analyses, but are included because of scientific interest.

- **The BCVA**, a continuous outcome, will be modeled using a linear regression model. Analyses will be conducted with BCVA represented as logMAR. A linear regression model for the actual BCVA at month 6 and 12 will be fitted, adjusting for baseline BCVA. The model will include fixed effects for country and baseline antimetabolite use, following the stratified randomization. The following will serve as a template for analyzing other continuous variables. Only eyes affected with uveitis will be included in all the vision analyses.

  The rationale for further longitudinal analysis is that patients who fail or experience flares may exhibit decreased visual acuity. Such inferences may be complicated by the fact that following failure and the consequent cessation of trial protocol, pertinent values of the visual acuity are no longer available. Analysis of available data only is expected to yield biased inference, and we propose to avoid

conditioning on post-randomization covariates.

We propose imputation of post-failure BCVA by *last observation carried forward* (LOCF), followed by pooled regression (clustered robust standard errors on individual). We propose to do so only when BCVA is available at the time (study visit) at which treatment failure is declared. We will use all BCVA measurements until either treatment failure or completion of the trial and report the coefficient for randomization group, adjusting for baseline antimetabolite use and country.

We also propose to descriptively report mean, median BCVA (together with standard deviation and interquartile range) of patients at each visit. Actual BCVA will be reported in a decriptive analysis.

This collection of procedures serves as a template for all continuous outcomes for which we assess final values (below).

In Aim 3, these regression models will be examined on an intent to treat basis, so time/outcomes after treatment failure are included in the analyses.

The following regression analyses will be conducted following the templates above: statistical predictor will be randomization group, and we include fixed effects for baseline antimetabolite and country. A separate model will include country by treatment interaction. Treatment effects will be only reported by stratum when heterogeneity or interactions are found.

Final values of the following continuous outcome variables will be modeled following the template found above in this Planned Secondary Analyses with Alternative Outcomes section:

- **The CHAQ score**
    - For this quality-of-life measurement and the three specified immediately below, a separate analysis comparing the patient's baseline value with their value at the time of time failure, 6 and 12 months will be considered.
- **The EQ-5D-Y score**
- **The CVFQ score**
- **The EYE-Q score**
- **The JADAS-10 and JADAS-27 scores**
- **Estimated ADA levels**

Final values of the following binary outcome variables will be modeled with log binomial regression following the similar template in the Planned Subsidiary Analyses of the Primary Model section:

- **The proportion of patients who exhibit one or more arthritis flares**
- **The proportion of patients who achieve an improvement in JADAS-10**
- **The proportion exhibiting macular edema**

Finally, we propose to measure patients' perceived treatment allocation to assess the effectiveness of patient masking to their allocation. We will measure patients' perceived treatment group before unmasking at either the time of treatment failure or at 12 months (the primary endpoint) if the patient does not experience treatment failure. We test for differences between groups using a Fisher's exact test for differences in a categorical outcome (adalimumab, placebo, don't know) stratified by whether the patient had treatment failure. Stratification is necessary to prevent bias in the event that the treatment group influences treatment failure, which in turn influences the patient's guess.

Analysis of secondary outcomes following the interim: Given the DSMC's decision on February 14, 2024 to stop enrollment based on interim efficacy, the investigator team decided to publish the interim analysis of the primary outcome immediately, without waiting for enrolled patients to complete their 12-month follow-up. To provide timely information for key secondary outcomes at the interim point in the trial given incomplete follow-up of longitudinal outcomes on many patients, we propose the following analyses that use all available information at the time of the interim. We propose to include longitudinal measures of four secondary outcomes measured through 12 months, treatment failure or censoring (whichever occurs first): BCVA, JADAS scores, adalimumab drug levels, and adalimumab antibody levels. The analysis of each outcome will follow the same general method. A longitudinal, linear mixed model will be fit, including fixed effects for stratification variables (antimetabolite use, country), treatment assignment, and the baseline measure of the outcome. A random effect for patient will account for repeated outcome measures. This is identical to the planned analyses above, but includes all available longitudinal measures (not only those at the primary endpoint or censoring at 6- or 12-months). These new analyses were specified in February 2024, after knowledge of the primary analysis and interim stopping.

As a robustness check for these secondary analyses, we will estimate a joint model that combines the likelihood from the treatment failure survival model with the likelihood of the longitudinal mixed model for the secondary outcomes.[18,19] A joint model of the likelihood enables unbiased estimation of the treatment effect on the longitudinal outcomes while accounting for potentially informative missingness due to treatment failure, where longitudinal measures might be missing not at random (MNAR). We propose to fit a joint model using a Bayesian Monte-Carlo Markov Chain (MCMC) approach implemented in the JMbayes2 R package.[20] We will report results of the joint model with a particular focus on the effect of treatment on differences in the longitudinal outcome, accounting for the treatment failure process. We note that since JADAS scores, drug levels, and antibody levels are measured infrequently in the trial: at 3 months (JADAS only), 6 months, 12 months, and time of treatment failure, that it might be difficult or impossible to fit longitudinal models that include a random effect for slope (required by currently implemented joint modeling software, such as JMbayes2). We will assess MCMC parameter convergence using standard diagnostics (trace and density plots of model parameters, Gelman-Rubin convergence statistics) and will report any

failure to converge.

3.1.2 Specific Aim 2

Primary Analysis

As emphasized earlier, the data available in this clinical trial provide an opportunity to assess the capacity of selected variables in predicting a uveitis flare in both treatment groups. Exploratory analyses will be conducted as follows.

Descriptive tabulation and plotting of all outcome variables with respect to each potential explanatory variable will be reported.

Univariate log binomial regression of the occurrence of a uveitis flare during the first 12 months will be examined. The following predictors (i.e. regressors) will be examined:

- baseline blood serum levels of myeloid-related protein MRP8/14 (calprotectin),
- baseline serum levels of erythrocye sedimentation rate (ESR),
- baseline serum levels of C reactive protein,
- baseline ADA levels,
- baseline anterior chamber cell level (Grade 0 or 0.5),
- randomization group,
- age,
- sex,
- race,
- Tanner stage for pubertal stage,
- baseline antimetabolite use,
- duration of control of uveitis (MOP Section 2.4.1) before baseline,
- length of time taking adalimumab before baseline,
- country

We next propose multivariate analysis and model selection using the standard LASSO (least absolute shrinkage and selection operator[21]). The LASSO penalty will be estimated using cross-validation with at least 1000 replications and a random 20% of the data used as the test set (the remaining being a training set). Variables selected by the LASSO procedure will then be fit to the full data set using ordinary logistic regression (unpenalized) as well as log binomial regression (glm, log link). The results will be compared to model selection based on backwards stepwise regression (R package `stepAIC`). We will also compare the results with general elastic net regression, where using the ridge penalty (in addition to the LASSO penalty) improves estimates in case of multicollinearity among predictors. Two-way interaction terms will be explored, with particular emphasis on interactions with randomization group. (We will constrain model selection to only include an interaction term between two variables when the variables themselves are included.) Additional software packages or methods may be applied.

Results of exploratory (hypothesis generating) procedures will always be reported as such. Significance P-values following model selection procedures are dubious.

Similar methods will be used to explore predictors for the following outcomes:

- occurrence of uveitis flare in the first 6 months
- occurrence of arthritis flare in the first 6 months, and the first 12 months
- improvement of JADAS-10 score at 6 months, and at 12 months
- occurrence of macular edema in the first 6 months, and at 12 months

We will also conduct exploratory analysis for the number of missed treatment doses during the first 6 months, and during the first 12 months. The analysis will follow the same template as above, except that negative binomial regression will be used instead of log binomial and logistic regression.

Continuous outcomes will be examined longitudinally. Above, we emphasized modeling of final values (at 12 months for treatment successes, and at the time of failure otherwise). The following are also of interest:

- The 6 and 12-month values for all patients, including patients who experienced a treatment failure (i.e., including time following reinstatement of adalimumab if in the placebo group or following salvage therapy in the adalimumab group)
- All longitudinal observations, including BCVA imputed with LOCF post treatment failure

For continuous outcomes, plots of each measurement along with the outcome status will be prepared for each patient. Plots will include data for patients after any treatment failure, in a different color. We plan to separately examine patients using antimetabolites at baseline and those who were not. We will conduct linear mixed effects regression for these longitudinal outcomes, using time dependent covariates. The continuous variables of interest are:

- BCVA
- CHAQ score
- EYE-Q score
- JADAS-10 and JADAS-27 scores

Application of LASSO procedures to linear mixed models will be conducted with the R package `glmmlasso`.

Finally, a similar exploratory analysis will be used for the time to treatment failure (as defined earlier), again using LASSO regression with the same set of covariates. LASSO model selection will be implemented using the R package `coxnet`. We note that other model selection procedures (such as backwards stepwise regression) can be considered, provided uncorrected significance P-values are not reported after model selection.

3.1.3 Specific Aim 3

Primary analysis. The outcome variable is the proportion of people achieving corticosteroid-sparing controlled ocular inflammation at month 12 of the study. We propose log binomial regression, with randomization group as a covariate and fixed effects for baseline antimetabolite use and country. The analysis will follow the same template as in Specific Aim 1 for analysis of binary outcomes (including assessment of homogeneity by country). A linear binomial regression, with the identity link, will be used to calculate risk differences between groups. Significance will be assessed with a two-sided alpha of 0.05. The distinction between the analyses in Aims 1 and 3 are that in Aim 1, outcomes are reported at 6 and 12 months, censoring at treatment failure, while in Aim 3, outcomes are reported by intent-to-treat at 6 and 12 months, regardless of treatment changes in the event of treatment failure. We propose to use Firth bias correction in this log binomial regression, and additional information is provided below.

Analysis following the interim. Given the DSMC's decision on February 14, 2024 to stop enrollment but continue follow-up, a Cox proportional hazards regression for time to regaining control after treatment failure will be conducted as not all patients had reached month 12 at the interim. Two outcome variables for regaining control will be part of the planned analyses: time to initial control and time to sustained control. Time to initial control is defined as the first instance of meeting control criteria (as defined in Section 2.3.3) following the declaration of treatment failure. Time to sustained control is defined as the first instance of meeting control whereafter control is maintained through their month 12 visit, as it is possible that patients who experience failure will regain control, but then flare again. At the interim, participants who had experienced treatment failure will be included in the time to initial control analysis. Only patients who experienced treatment failure and had completed their month 12 visit by interim will be included in the time to sustained control analysis. When 12 month follow up is complete for all patients, all patients can be included in the pre-specified sustained control analysis.

Planned Secondary Analyses
Continuous outcome variables and binary outcome variables will be examined on an intent-to-treat basis in a similar way as indicated in Specific Aim 1. The following are planned analyses:

- **The BCVA at 6 and 12 months**, a continuous outcome,
- **The proportion of people exhibiting an arthritis flare during the first 6 months and during the first 12 months**, a binary outcome
- **The CHAQ score at 6 and 12 months**, a continuous outcome.

Planned Subsidiary Analyses
We propose to conduct the primary analysis and planned secondary analyses (as described above) using the subgroup of people who were randomized to discontinue adalimumab, subsequently had treatment failure, and then restarted adalimumab. We

will also conduct exploratory analyses to find predictors of reestablishing control of inflammation. The exploratory procedures described in Specific Aim 2 will be used (including descriptive tabulations, graphical exploration, and model selection using LASSO or other procedures). As before, exploratory analyses will always be labeled as such, and all analyses conducted will be reported.

Using procedures analogous to those in Specific Aim 1, we will also examine on an intent to treat basis:

- the number of missed treatment doses over 6 months and over 12 months (using negative binomial regression),
- the EQ-5D-Y score at 6 months and at 12 months (linear regression),
- the CVFQ score at 6 months and at 12 months (linear regression),
- the EYE-Q score at 6 months and at 12 months (linear regression),
- the JADAS-10 and JADAS-27 scores at 6 months and at 12 months (linear regression),
- the proportion showing improvement in JADAS-10 at 6 months and at 12 months (log binomial regression),
- the proportion exhibiting macular edema over 6 months and over 12 months (log binomial regression),
- ADA levels at 6 months and at 12 months (linear regression),
- the proportion discontinuing treatment due to intolerability by 6 months and by 12 months,
- the proportion discontinuing treatment due to safety by 6 months and by 12 months.

For all binomial regressions, we will report the results of other binomial regressions (such as logistic regression), if requested. Such results will always be sharply distinguished from the primary prespecified analyses.

3.2 Transformations and model adequacy

3.2.1 Primary Analysis

We will assess the adequacy of the proportional hazards assumption using Schoenfeld residual plots, but supplement this with the Gill-Schumacher procedure and other graphical methods. Violations of the proportional hazards assumption will be reported, and will lead us to report additional analyses (the prespecified primary analysis will always be reported). We propose to report accelerated failure time models using the

same covariates, in the event of failure of the proportional hazards assumption.

3.2.2 Model validation and sensitivity

In all cases, standard statistical procedures will always be followed to ensure that no evidence indicates a violation of the assumptions underlying the statistical models used.

- We will assess the proportional hazards assumption for all Cox regressions using the methods given above (though we note that our primary hypothesis test is based on permutation).
- Also, we will examine Hosmer-Lemeshow goodness-of-fit tests for continuous predictors in binomial regressions. Linear models will always be assessed using residual plots (residuals vs. predicted values, and QQ plots), together with tests for normality (Anderson-Darling and Shapiro-Wilk procedures). Goodness of fit tests may be underpowered when they are needed the most (in small sample size studies). We do not propose formal thresholds for goodness-of-fit, but will add all such assessments to an analytic log, and make them available to the DSMC and statistical reviewers. Assessments of overdispersion will be conducted for count models.
- In the event of convergence difficulties in log binomial models we propose to use Poisson procedures with robust variance estimators. We propose to report the results of logistic regression in the event that all such procedures are unsuccessful. All such procedures will be reported.[22]
- Jackknife influence estimates will be used in all analyses; single observations that could change the conclusions will always be reported; supplemental analyses dropping highly influential points will be reported at the discretion of the DSMC or statistical reviewers. We note that the presence of highly influential points does not invalidate the conclusions of a study, but such assessments contribute to the overall understanding of the trial. In our view, reporting the results of analyses dropping individual observations based on statistical outlier/influence tests *alone* is a dubious practice, in the absence of additional information about those outliers.
- Analyses in which our primary interest is in final outcomes (at either the 6 or 12 month timepoint for individuals who did not change treatment, or the time of failure for those who failed) will still be repeated using all available data (at all time points).

Failure of the modeling assumptions (such as normality) will result in conducting additional analyses. First, for continuous outcome variables, we will undertake normalizing or variance-stabilizing transformations of the outcome variable (such as power transformations). Second, we note that we are proposing use of permutation-based procedures for statistical testing whenever possible. Third, the use of bootstrap procedures, when applicable, will be considered in estimation of standard errors.

3.2.3 Missing covariates

Analyses using missing covariates will not apply to the primary analysis. In the primary analysis all individuals contribute observation time until censoring, and no covariates are expected to be missing. However, supplemental analyses of other outcomes, as well as the exploratory analyses in Aim 2, may necessitate treatment of missing data.

A complete case analysis will be used for all secondary analyses. If covariate data are missing for a large proportion of participants (>15%), we will also include a sensitivity analysis that uses regression-based multiple imputation (based on an assumption of missingness at random) to assess the impact of missing data on secondary outcomes of interest. We will use the `mice` package in R (Multiple Imputation with Chained Equations) with 100 replications to best impute the missing values. These sensitivity analyses may be used to determine the threshold value where imputation would change the overall findings (i.e., directly address the assumption of missingness at random). We note that Aim 2 is exploratory and hypothesis-generating.

Note that complete case analyses are, in general, biased, while imputation methods require an untestable assumption of missingness at random (or specific models of the nonrandom nature of missingness).

3.3 Sample Size Estimation

3.3.1 Primary Calculation

A sample size of 118 subjects (59 in each group) provides 88% power to detect a hazard ratio of 2.0 for the time to treatment failure comparing the group randomized to discontinue adalimumab to the group randomized to continue using adalimumab, assuming a median time until treatment failure of 10 weeks in the group that discontinues adalimumab (Arm 1) and of 20 weeks in the group that continues on adalimumab (Arm 2), an equal allocation between groups, and a 10% total loss to follow-up.

For sample size planning, we use the approximate formula given in Friedman et al (2010) for the number in each group:

$$N = \frac{\left(Z_\alpha + Z_\beta\right)^2 (\phi(\lambda_C) + \phi(\lambda_I))}{(\lambda_I - \lambda_C)^2}$$

Where λ denotes the rate in the intervention arm (λ_I) and control arm (λ_c), $\phi(\lambda) = \lambda^2 \,/\, 1 - e^{-\lambda T}$ at censoring time T, and Z_α, Z_β are quantiles from the standard normal distribution at power (1-β) and significance level α.

In a retrospective study by our group of JIA-associated uveitis patients on immunomodulatory therapies (antimetabolites, cyclosporine, TNF-α inhibitors, or combination therapy), 37% of patients attempted to stop treatment after achieving corticosteroid-sparing control of inflammation.[23] Of patients who stopped TNF-α inhibitors, nearly 50% experienced a recurrence of inflammation within 100 days after stopping treatment. Approximately 80% of patients had a recurrence of ocular inflammation within one year after stopping treatment.

Furthermore, a retrospective study of uveitis patients on infliximab showed that patients with JIA-associated uveitis who attempted to stop anti-TNF-α therapy after achieving control of inflammation experienced recurrence of inflammation after a median of 76 days.[24] Similarly, in a small retrospective study outside our group, 6 pediatric uveitis patients flared between 3 and 7 months after discontinuing adalimumab.[25] Based on these data, we can expect to observe a majority of treatment failures within the first 6 months and nearly all treatment failures by Month 12 among patients randomized to discontinue adalimumab treatment. It is important to note that stopping infliximab treatment involves a gradual taper of the dose over the course of several weeks. In both studies, time to recurrence of inflammation was measured from the start of the taper rather than the point at which patients had completely stopped the medication. Because discontinuing adalimumab treatment does not involve a taper in our trial, we expect the time to recurrence of inflammation to be shorter as compared to infliximab. Therefore, we anticipate time to treatment failure of 10 weeks in patients randomized to discontinue adalimumab in ADJUST.

Table 3. Previous time-to-event uveitis studies

Author	Population	Median Time to Uveitis Relapse
Acharya et al. (2018)	JIA-associated uveitis patients who discontinued TNF-α (n = 11)	100 days
Shakoor, Esterberg, & Acharya (2014)	JIA-associated uveitis patients who discontinued infliximab (n = 18)	76 days

Castiblanco, Meese, & Foster (2016)	Uveitis pediatric patients who discontinued adalimumab	3 to 7 months (range)

In a clinical trial of JIA-associated uveitis patients with active inflammation on a stable dose of antimetabolite, 114 patients were randomized to adalimumab or placebo biweekly subcutaneous injections and followed for up to 18 months.[26] The mean duration of sustained inactive disease was 179.3 ±16.9 days in the adalimumab treatment group (vs. 14.5±23.9 days in the placebo group). An open label long-term follow-up study of non-infectious adult uveitis patients treated with adalimumab demonstrated that, among patients with inactive disease who entered the study, over 90% of patients still had inactive disease at week 54.[27] At 78 weeks, 74% still had inactive disease.[28] Given the available evidence on disease recurrence for patients on long-term adalimumab treatment, we can expect a lower recurrence rate and longer time to treatment failure in patients continuing on adalimumab compared to those discontinuing treatment. While these studies suggest that most uveitis patients on adalimumab remain inflammation-free greater than one year, we have conservatively estimated 20 weeks to be the median time to treatment failure among patients randomized to continue adalimumab treatment.

Thus, 118 patients total (59 per group), provides 88% power to detect a hazard ratio of 2.0 for the time to treatment failure, assuming a 10% loss to follow-up. We assume an alpha of 0.05 (two sided). Adjusting for country in our analyses is assumed to give us even greater power. A power table is provided in Table 4 below, varying the median time until treatment failure (in weeks).

Table 4: Total Sample Size Based on Median Time to Treatment Failure for Stopping vs. Continuing Adalimumab

		80% Power	90% Power
Stopping Treatment Versus Continuing Treatment	8 weeks vs 16 weeks	92	124
	10 weeks vs 20 weeks	96	126
	12 weeks vs 24 weeks	98	130

3.3.2 Power for Subgroup Analyses and Other Analyses

Our primary analysis is powered at 88% and most secondary analyses are reported at 80% power.

3.3.2.1 Specific Aim 1

Planned Secondary Analyses

- **Time to treatment failure, with censoring at 6 months.** We anticipate 80% power to detect a hazard ratio of 2.0 between randomization groups, based on 118 subjects (we anticipate same loss to follow up of 10%).
- **The proportion of people with uveitis flares over 6 months.** For sample size planning, we assume a test of proportions between the two randomization groups. Based on prior research, we expect 70% of those who discontinue adalimumab to experience flare by 6 months.[24] We anticipate 80% power to detect a difference of 28% between the groups, based on 118 subjects. Note that we anticipate fewer flares in the continuation group, but the test will be two-sided at alpha of 5%.
- **The proportion of people with uveitis flares over 12 months.** Based on prior research, we expect about 80% of those who discontinue adalimumab will have a uveitis flare by 12 months.[24] With our total sample size of 118 and the assumed 10% loss to follow-up by 12 months, with 80% power, we anticipate a power to detect a difference of 28% in the flare rate between the groups.

Planned Secondary Analyses with Alternative Outcomes

- **Longitudinal BCVA.** For sample size assessment, we assume a T-test comparing change scores between the two randomization groups. We assume a standard deviation of 0.59, based on prior research.[29] The total sample size of 118 and the assumed 10% loss to follow-up by 12 months provides 80% power to detect a difference of 0.32 logMar.
- **The CHAQ score.** For sample size planning, we assume a T-test comparing change scores between the two randomization groups. We shall assume a standard deviation of 23 based on prior research.[30] Thus with the total sample size of 118 and the assumed 10% loss to follow-up by 12 months, with 80% power we can detect a difference of approximately 13 points.
- **The EQ-5D-Y score.** For sample size planning, we asssume a T-test comparing change scores between randomization groups. Assuming a standard deviation of 17.3,[31] the total sample size of 118 and assumed 10% loss to follow-up by 12 months provides over 80% power to detect a mean difference of 7.9 points.
- **The CVFQ score.** For sample size planning, we assume a T-test comparing change scores between the randomization groups. Assuming a standard deviation of 1.70 based on a validation study of the CVFQ,[32] a total sample size of 118 and assumed 10% loss to follow-up by 12 months provides 80% power to detect a mean difference of 0.93 points.
- **The EYE-Q score.** For sample size planning, we assume a T-test comparing change scores between the randomization groups. Assuming a standard deviation of 0.75,[8] the total sample size of 118 and assumed 10% loss to follow-up by 12 months provides approximately 80% power to detect a mean difference of 0.41 points.

- **The JADAS-10 score.** As before, we use a T-test comparing change scores between the two groups for sample size planning. Assuming a standard deviation of 8.23,[10] the total sample size of 118 and assumed 10% loss to follow up by 12 months provides approximately 80% power to detect a mean difference of 4.5 points.
- **The JADAS-27 score.** Similarly, assuming a standard deviation of 8.16,[10] the sample size of 118 and assumed 10% loss to follow-up by 12 months provides approximately 80% power to detect a mean difference of 4.5 points.
- **The ADA levels.** Assuming a standard deviation in ADA levels of 10,[33] the total sample size of 118 (and, the 10% assumed loss to follow-up by 12 months) suggest approximately 80% power to detect a difference of 5.5 in ADA serum levels in μg/mL.
- **The proportion of people with arthritis flares.** We assume (for sample size assessment) a test of proportions between the two randomization groups. Based on prior research, we expect 1% of those who continue adalimumab to have a flare by 6 months.[26] We anticipate 80% power to detect a difference of approximately 16% in the proportion of individuals experiencing arthritis flares among those who discontinue adalimumab (assuming a total sample size of 118 and 10% loss to follow-up by 12 months).
- **The proportion of patients who achieve an improvement in JADAS-10.** Assuming that 85% of those continuing on adalimumab will have an improvement in JADAS-10 score from baseline,[34] a total sample size of 118 (and 10% assumed loss to follow-up) provides approximately 80% power to detect a difference of at least 26% in the proportion of patients achieving an improvement between treatment groups.
- **The proportion of people with macular edema.** We anticipate at most 1% of the individuals continuing treatment to develop macular edema, given its absence at baseline. We anticipate approximately 80% power to detect a difference of 15% (in absolute terms) in macular edema in the discontinuation arm (given 118 subjects).

The analyses for censoring at 6 months are similar to 12 months, with small decreases in detectable effect sizes due to assuming slightly shorter follow-up by month 12.

3.3.2.2 Specific Aim 2

Specific Aim 2 is exploratory, and is hypothesis generating. However, we present selected power analyses to illustrate the sufficiency of the sample size of the trial—fixed in Specific Aim 1—to inform regression models of the exploratory outcomes in Specific Aim 2.

For modeling the effect of the potentially predictive biomarker MRP8/14, we assume a

median split as a guide to power. Assuming approximately 50 uveitis flares among our patient population, we anticipate 80% power to detect a difference in MRP8/14 levels of 140 ng/ml (assuming a standard deviation of 250 ng/ml, based on previous studies).[35,36]

For modeling the number of missed study treatment doses in the first 12 months, we anticipate over 80% power to detect a difference of 2.8 missed doses between the discontinuation and continuation groups (assuming 118 patients enrolled, and a standard deviation of 5).

For the time to failure outcome within 12 months, we anticipate 88% power to detect a hazard ratio of 2.0 between patients taking antimetabolites at baseline and those not using antimetabolites, assuming approximately equal numbers in each group (assuming 10% loss to follow-up).

3.3.2.3 Specific Aim 3

Primary Analysis

The proportion of people who achieved corticosteroid-sparing controlled ocular inflammation (Table 2) by month 6 of the study based on treatment intervention. For sample size planning, we assume a test of proportions between the two randomization groups. Based on prior research, we assume that 30% of those who discontinue adalimumab will have controlled ocular inflammation by month 6.[24] We anticipate over 80% power to detect a 29% difference between the two groups, based on 118 subjects.

The proportion of people who achieved corticosteroid-sparing controlled ocular inflammation by month 12 of the study based on treatment intervention. For sample size planning, we assume a test of proportions between the two randomization groups. Based on prior research, we assume that 24% of those who discontinue adalimumab will have controlled inflammation by month 12.[24] We anticipate approximately 80% power to detect a difference of 28% between the groups.

Secondary Analyses
Power analyses for the continuous outcomes in this Aim are similar to those in Specific Aim 1.

3.4 Missing Data and loss to follow-up

Values of the primary study endpoint (time to treatment failure) can be analyzed when the individual is lost to follow-up, treating such individuals as having been censored after their last visit.

We will compare baseline characteristics of individuals lost to follow-up with those who continued in the trial, tabulating by age, sex, study site, and other covariates available at baseline. Significant predictors of loss to follow-up will be then included in models of the outcome variables and reported.

As suggested above, we will use a complete case analysis under the assumption of missingness completely at random for all secondary analyses. However, if values of covariates are missing for more than 15% of participants, then a sensitivity analysis will be conducted where missing covariates will be imputed using regression-based multiple imputation, using the `mice` package in R. Missing outcome data will be analyzed as last observation carried forward, and by sensitivity analysis in which we assess how extreme missing values would have to be in order to change the conclusions of the trial.

3.5 Multiple comparisons

No multiple comparison corrections will be made. All analyses following the primary analysis for each of the three Aims will be reported as descriptive and will be presented with 95% confidence intervals.

3.6 Interim Monitoring

The study will be monitored by a Data Safety Monitoring Committee (DSMC) appointed by the National Eye Institute. There will be in-person meetings in Year 01 and Year 05, and teleconference calls every 6 months in between, or as deemed necessary.

The DSMC will receive fullly unmasked information at any time. Such reports will be prepared by the principal statistician following the instructions of the DSMC.

3.7 Stopping Guidance and Interim Analysis

Timing of the interim analysis. The Data and Safety Monitoring Committee will meet to review the interim efficacy data when 34 treatment failures have occurred. This represents approximately 40% of the anticipated 84 events expected in the trial based on the assumed median failure times used in the primary analysis sample size calculation (median 10 weeks in the stopping group, and 20 weeks in the adalimumab group). Timing the interim at 40% of events was chosen as a trade-off between having sufficient events to make interim efficacy, futility and safety analyses informative, and

timelineness with respect to the trial informing patient care in the event of stopping under one of the guidelines.

The DSMC will make one of the following recommendations:

- Continue the trial without modifications
- Continue the trial with study modifications
- Stop enrollment in the trial because of efficacy
- Stop enrollment in the trial because of futility
- Stop enrollment in the trial because of safety concerns

The DSMC will take into consideration (a) safety, (b) efficacy, (c) clinical importance, and (d) validity to make a final decision regarding stopping guidance implementation. If patient safety is not at risk, then consideration of secondary outcomes may overweigh the need for immediate reporting of the primary outcomes.

Efficacy. An unmasked interim analysis will be conducted when 34 treatment failures have occurred (the primary endpoint) to determine whether or not there is sufficient evidence to answer the primary aim and justify stopping additional enrollments.

We propose a group sequential boundary for judging the statistical significance of the primary outcome, allowing for the interim analysis. We propose a Kim-DeMets flexible alpha spending approach, implemented in the gsDesign package in R, with power function $\alpha^*(t^*)^{\theta}$, where θ =1.75 (θ chosen so that the two-sided *P*-value to stop the trial for efficacy is exactly 0.01 when 40% of events have ocurred).[37] The detectable HR at the interim analysis is 2.4 at the original, 88% power. A nominal two-sided P-value of 0.044 at the final analysis will account for the proposed interim analysis under a group sequential design. The Appendix includes details of the parameters and output from R's gsDesign package.

Futility. Early discontinuation due to the unlikeliness of significant findings conditional on interim results may be considered, based on the original sample size considerations. For evaluating futility, we propose discontinuation for futility if the conditional power to detect a hazard ratio of 2.0 for placebo/adalimumab (the original design effect) is below 20% at the interim analysis.[38] Conditional power will be derived by simulation of the unobserved future patients according to the original alternative hypothesis. Our proposed futility guidance is one-sided, and is subordinate to efficacy analyses.

Harm. Stopping for harm will be done at the judgment of the DSMC. Harm will be assessed using adverse events, and especially serious adverse events. Ethical considerations require careful consideration and judgment by the DSMC.

All subjects who provide informed consent will be accounted for in this study. We will present the number lost to follow-up and the number of protocol deviations by arm.

If the proposed stopping criteria are met from an efficacy standpoint and there are no safety concerns, enrollment will stop and patients who have already enrolled will be followed through their final visits following the current study protocol. Should they experience a treatment failure, they will be unmasked per protocol and can continue to receive study-provided medication through Month 12. Their outcomes are valuable to overall results (primary and secondary analyses). The trial's main result provides information which informs decision-making regarding the risks of stopping adalimumab. Patients who are enrolled have already made the decision that they are comfortable being randomized to stopping therapy, so they are already past the point of making that decision. Even if unmasked and on placebo, it would not be expected that such a patient would resume therapy if they did not experience a recurrence of inflammation.

Unmasking active patients at such an interim point compromises patient and personnel efforts already initiated towards the trial, as well as their study data. Enrolled patients will remain in the study. Enrollment, however, shall stop in an effort to disseminate study results sooner so that patients considering discontinuation of adalimumab can make better informed decisions about future treatment. The DSMC may reconsider the decision regarding the protocol for patients who have not completed their follow-up if/when stopping guideline criteria are met.

For the Statistical Analysis Plan, adverse event reporting will be descriptive in nature. Serious adverse events (SAEs) are reported directly to the medical monitor within 24 hours of the time the study site learns of them, and the medical monitor will subsequently pass this information on to the DSMC Chair. The medical monitor will receive notification of the event, the timing of the event, a medical narrative from the study site, the site location, and the patient identification number. The statistician will report the study treatment assignment to the DSMC Chair if deemed necessary by the DSMC. If use of either intervention clearly results in an unacceptable increase in the risk of treatment failures, then the study will be stopped.

In the event of that the DSMC stops the trial early, for any reason, the final analysis will be conducted on the available data.

3.8 Final Analyses

All analyses will be completed when all patients complete their last visit at 12 months.

3.9 Software

The standard software program R (v 3.5 or higher, R Foundation fo Statistical Computing, Vienna, Austria) will be used for all descriptive and inferential analyses.

4 Analysis Populations

4.1 Summary

The following analysis populations are planned for this study:

- The **screening population**, which includes all patients who are screened for participation in the trial.
- The **safety population**, which includes all patients who received the intervention.
- The **intent-to-treat efficacy population**, which includes all patients who are randomized.
- The **per-protocol efficacy population**, which includes all patients in the intent-to-treat efficacy population, excluding patients with any of the following: 1. Major protocol deviations, or 2. Noncompliance with study medications (more than 20% of the treatment missed).

4.2 Major protocol deviations

The incidence of deviations from the inclusion and exclusion criteria will be summarized using counts and percentages, and the treatment groups compared for the overall frequency of deviations using a 2xN Fisher's exact test. Similar deviations will be grouped into general categories of deviations for a more condensed summary. A listing of deviations by participant will also be produced. Any major deviations from the protocol will be listed and/or summarized, including, but not limited to, participants who:

- Never received treatment,
- Were subsequently found to be ineligible for the study,
- Never returned for a follow-up visit,
- Have follow-up visits outside the prescribed visit window.

The number and percentage of randomized participants actually receiving the study intervention and permanently discontinuing the study inteventions will be summarized. A summary of study participants randomized by site will also be provided. Randomization groups will be compared for the proportion and reason for study intervention discontinuation using the Fisher's Exact test. A summary of participant status at the end

of the study period will also be generated with categories including lost to follow-up.

5 Data Collection and Quality Assurance

5.1 Quality assurance and security

Data collection forms, training, security, and quality assurance are discussed in the Manual of Operations.[2]

5.2 Analysis sets

Data sets for analysis will be produced at the Proctor Coordinating Center by the Data Analysis Committee. Each will be a CSV file containing a single header line whose variable names match the REDCap database. Each column will be a different variable. Each row will be a different observation. Character strings will be used whenever possible. Missing values will be coded as NA. Checks will be made to ensure that each variable in the analysis set has in-range values.

A detailed codebook will be prepared, containing for each variable, (a) the form from which the variable is derived, (b) text of the question, (c) all possible values for the variable, and (d) summary statistics for the variable. Note that all codes and character strings that represent categorical factors will be clearly defined. Units for each continuous variable will be indicated for each variable. Each release of the analysis set will be accompanied by the corresponding version of the codebook.

5.3 Data monitoring reports

Data monitoring reports will be prepared based on analysis data sets. The monitoring reports will include recruitment reports for each site, compliance reports, retention reports and data quality reports. These will be reviewed at the central site on a monthly basis and communicated to the study sites on a monthly basis.

6 Human Subjects

6.1 Summary of final dispositions

All subjects who provide informed consent will be accounted for in this study. The frequency of subjects in each population will be presented. We will also present the frequency of subjects in each subgroup, the frequency of withdrawal and loss to follow-up and any major protocol violations.

6.2 Data and Safety Monitoring Committee

6.2.1 Scope

The DSMC will be appointed by the NEI. We propose that this committee consist of 5-7 individuals, and should include uveitis specialists, a rheumatologist, a trialist, an independent biostatistician, and a bioethicist. The NEI Program Official serves as a resource to the DSMC but is not on the DSMC, does not vote, attends the open sessions, and may attend the closed sessions when deemed necessary by the Chair of the DSMC. The committee will meet in person at least once in Year 01 and 05. Other meetings may be arranged. All study protocols will be subject to review and approval by the Institutional Review Board at UCSF (central IRB for US sites), the South Central Berkshire Research Ethics Committee, the ethics committee at Melbourne, and by the DSMC.

6.2.2 Meetings

All in-person and teleconference meetings of the DSMC and study personnel will consist of (a) “open” sessions, which may be attended as needed by masked study personnel, and (b) “closed” sessions attended by only unmasked study personnel and (c) “closed” sessions attended only by the DSMC personnel. Care will be taken so that no intervention assignments, or data which would allow intervention assignments to be determined will be revealed during the open sessions.

The DSMC will be unmasked. Closed reports will tabulate baseline covariates, adverse events, and outcomes by treatment assignment and study site. Written closed reports will always use the labels Treatment A and Treatment B for increased information security. However, the DSMC will know which drug corresponds to which label.

The Coordinating Center’s biostatistician will determine the database closure dates for each report in advance; archival copies of the (a) main REDCap database, and (b) study analysis file as they exist at the time of each report will be maintained.

Interim reports for the DSMC will be prepared by the Proctor Coordinating Center. These reports will include (a) recuitment overall, and by study site, (b) compliance and (c) retention. The reports will also list study outcomes and all adverse outcomes, including deaths.

All reports will be sent using secure email to the members of the DSMC at least one week prior to each meeting.

Each printed (hard copy) interim report will be labeled clearly as confidential, bound so that the contents are not visible from the outside, and labeled with the name of each person authorized to receive it. Reports will be kept in possession of the primary biostatistician and research data analyst and only delivered in person or by encrypted email; reports not delivered due to absences are to be shredded. In addition, redacted versions of the interim reports will be prepared which contain no masked study information, and which are suitable for restricted distribution to other personnel on an as-needed basis. All hard copies will be destroyed at the end of each meeting, except for a copy to be kept in a locked file cabinet accessible only to the research data analyst.

6.2.3 Decisions

The DSMC will not make decisions about the trial, but rather advise the NEI and the Executive Committee of ADJUST as to whether a protocol should continue as scheduled or undergo a modification (details specified in the DSMC charter, section 5.5). The DSMC will make recommendations with the benefit of prespecified decision guidelines. Interim stopping guidelines for efficacy, futility, and safety are described above, in section 3.7 of the SAP. The DSMC will monitor safety data throughout the trial and will have the full support of the unmasked team at the trial's Data Coordinating Center to provide additional details on patient status or outcomes should that be needed to inform a recommendation to NEI and ADJUST Executive Committee.

7 Safety and tolerability

The analysis of safety in this study will include summaries of the following:

- Exposure
- Adverse events
    - Non-serious and serious ocular and non-ocular adverse events
    - Adverse events leading to withdrawal from the study, and from the drug
    - Any deaths

7.1 Exposure

Individuals are assumed to have exposure to the intervention corresponding to the group to which they were randomized.

7.2 Adverse Events

7.2.1 Individual Events

Adverse event reporting procedures are described fully in Protocol Section 8.3.[2] Non-serious adverse events are described in Protocol Section 8.3.1.[2] Serious non-ocular or ocular adverse events (which must be repoorted within 24 hours and which require a narrative form) are described in Protocol Section 8.3.2.[2] Adverse events will be reported in all presentations and publications.

Safety-related events that occurred at or before treatment failure with censoring at 6 months and then 12 months (two different groups of safety-related events) will be tabulated and reported. Descriptive tables of the number and frequency of adverse events will be broken down by treatment group, age, sex, antimetabolite use and known comorbidities. We will report total adverse events and serious adverse events, cross-tabulated by whether the adverse events were anticipated or unanticipated and by whether or not the adverse event led to discontinuation of medication.

We do not propose statistical evaluation of adverse events, since the study is not powered for such assessments.

7.2.2 Pooled adverse events

Adverse events will be analyzed according to six main categories:

- Proportion of subjects with any ocular adverse event
- Proportion of subjects with any ocular serious adverse event
- Proportion of subjects with any systemic adverse event
- Proportion of subjects with any systemic serious adverse event
- Proportion of subjects with any laboratory adverse events
- Proportion of subjects with any laboratory serious adverse events

Note, hypothesis tests based on adverse events may be less than useful because the trial is not powered for these comparisons. These will only be provided to the DSMC if requested.

8 Reporting conventions

- All tables, data listings and figures will be presented in landscape orientation
- Direct annotation of figures will be preferred to legends. All figures with more than one variable or item will contain direct annotation
- Color will be used in figures only when needed to enhance clarity of communication

9 Appendix

All computations will be performed using the R statistical software (R Core Team 2021).[39] Specification of the random number seed and pseudorandom number algorithm determines the entire randomization assignment. Accordingly, the random number seed will be kept confidential.

10 References

1. Gamble C, Krishan A, Stocken D, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA*. Published online 2017. doi:10.1001/jama.2017.18556

2. Acharya NR. *Adalimumab in Juvenile Idiopathic Arthrisis-Associated Uveitis Stopping Trial (ADJUST)*.; 2019.

3. Jabs DA, Nussenblatt RB, Rosenbaum JT, Standardization of Uveitis Nomenclature (SUN) Working Group. Standardization of Uveitis Nomenclature for Reporting Clinical Data. Results of the First International Workshop. *Am J Ophthalmol*. 2005;140(3):509-516. doi:https://doi.org/10.1016/j.ajo.2005.03.057

4. Nussenblatt RB, Palestine AG, Chan CC, Roberge F. Standardizatlon of Vitreal inflammatory Activity in Intermediate and Posterior Uveitis. *Ophthalmology*. 1985;92(4):467-471. doi:10.1016/S0161-6420(85)34001-0

5. Singh G, Athreya BH, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. *Arthritis Rheum*. 1994;37(12):1761-1769. doi:10.1002/art.1780371209

6. Ravens-Sieberer U, Wille N, Badia X, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res*. 2010;19(6):887-897. doi:10.1007/s11136-010-9649-x

7. Felius J, Stager DR, Berry PM, et al. Development of an instrument to assess vision-related quality of life in young children. *Am J Ophthalmol*. 2004;138(3):362-372. doi:10.1016/j.ajo.2004.05.010

8. Angeles-Han ST, Griffin KW, Harrison MJ, et al. Development of a vision-related quality of life instrument for children ages 8-18 years for use in juvenile idiopathic arthritis-associated uveitis. *Arthritis Care Res*. Published online 2011. doi:10.1002/acr.20524

9. Consolaro A, Ruperto N, Bazso A, et al. Development and validation of a composite disease activity score for juvenile idiopathic arthritis. *Arthritis Rheum*. 2009;61(5):658-666. doi:10.1002/art.24516

10. Mcerlane F, Beresford MW, Baildam EM, et al. Validity of a three-variable Juvenile Arthritis Disease Activity Score in children with new-onset juvenile idiopathic arthritis. Published online 2012. doi:10.1136/annrheumdis

11. Martel JN, Esterberg E, Nagpal A, Acharya NR. Infliximab and adalimumab for uveitis. *Ocul Immunol Inflamm*. 2012;20(1):18-26. doi:10.3109/09273948.2011.633205

12. Porco TC, Stoller NE, Keenan JD, Bailey RL, Lietman TM. Public key cryptography for quality assurance in randomization for clinical trials. *Contemp Clin Trials*. 2015;42:167-168. doi:10.1016/j.cct.2015.03.016

13. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010

14. Rigdon J, Hudgens M. Randomization inference for treatment effects on a binary outcome. *Stat Med*. 2015;34(6):924-935. doi:doi:10.1002/sim.6384

15. International Conference on Harmonisation E9 Expert Working Group. Statistical Principles for Clinical Trials E9. *Stat Med*. 1999;18(15):1905-1942.

16. Gray A, Clarke PM, Wolstenholme JL, Wordsworth S. *Applied Methods of Cost-Effectiveness Analysis in Health Care*. Oxford University Press; 2011.

17. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programme. Third Edition*. Oxford: Oxford University Press; 2005.

18. Tsiatis A, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Stat Sin*. 2004;14:793-818.

19. Rizopoulos Dimitris. *Joint Models for Longitudinal and Time-to-Event Data : With Applications in R*. 1st ed. CRC Press; 2012. doi:10.1201/b12208

20. Rizopoulos D, Papageorgiou G, Afonso PM. JMbayes2: Extended Joint Models for Longitudinal and Time-to-Event Data. Published online June 26, 2023. Accessed May 29, 2024. https://cran.r-project.org/web/packages/JMbayes2/index.html

21. Hastie T, Tibshirani R, Wainwright M. Statistical Learning with Sparsity: The Lasso and Generalizations. *Crc*. Published online 2015:362. doi:10.1201/b18401-1

22. Williamson T, Eliasziw M, Fick GH. Log-binomial models: Exploring failed convergence. *Emerg Themes Epidemiol*. 2013;10(1):1-10. doi:10.1186/1742-7622-10-14

23. Acharya NR, Patel S, Homayounfar G, et al. Relapse of Juvenile Idiopathic Arthritis Associated Uveitis After Discontinuation of Immunomodulatory Therapy. *Ocul*

Immunol Inflamm. Published online 2018.

24. Shakoor A, Esterberg E, Acharya NR. Recurrence of uveitis after discontinuation of infliximab. *Ocul Immunol Inflamm*. 2014;22(2):96-101. doi:10.3109/09273948.2013.812222

25. Castiblanco C, Meese H, Foster CS. Treatment of pediatric uveitis with adalimumab: The MERSI experience. *J AAPOS*. 2016;20(2):145-147. doi:10.1016/j.jaapos.2015.12.006

26. Ramanan A V., Dick AD, Jones AP, et al. Adalimumab plus Methotrexate for Uveitis in Juvenile Idiopathic Arthritis. *N Engl J Med*. 2017;376(17):1637-1646. doi:10.1056/NEJMoa1614160

27. Suhler EB, Jaffe GJ, Nguyen QD, et al. Long–Term Safety and Efficacy of Adalimumab in Patients with Non-Infectious Intermediate, Posterior, or Panuveitis in an Ongoing Open-Label Study. In: *American College of Rheumatology*. ; 2016.

28. Suhler EB, Adán A, Brézin AP, et al. Safety and Efficacy of Adalimumab in Patients with Noninfectious Uveitis in an Ongoing Open-Label Study: VISUAL III. *Ophthalmology*. 2018;125(7):1075-1087. doi:10.1016/j.ophtha.2017.12.039

29. Díaz-Llopis M, Salom D, Garcia-De-Vicuña C, et al. Treatment of refractory uveitis with adalimumab: A prospective multicenter study of 131 patients. *Ophthalmology*. 2012;119(8):1575-1581. doi:10.1016/j.ophtha.2012.02.018

30. Gumus, Duygu; Basaran, Ozge; Cakar, Nilgun; Unco, Nermin; Celikel Acar B. Evaluation of Childhood Health Assessment Questionnaire in Juvenile Idiopathic Arthritis: A Single Center Experience From Turkey. *Arch Rheumatol*. 2015;30(1):57-62. doi:10.5606/ArchRheumatol.2015.5130

31. Wu XY, Ohinmaa A, Johnson JA, Veugelers PJ. Assessment of children's own health status using visual analogue scale and descriptive system of the EQ-5D-Y: linkage between two systems. *Qual Life Res*. 2014;23(2):393-402. doi:10.1007/s11136-013-0479-5

32. Birch EE, Cheng CS, Felius J. Validity and reliability of the Children's Visual Function Questionnaire (CVFQ). *J AAPOS*. 2007;11(5):473-479. doi:10.1016/j.jaapos.2007.02.011

33. Skrabl-Baumgartner A, Erwa W, Muntean W, Jahnel J. Anti-adalimumab antibodies in juvenile idiopathic arthritis: Frequent association with loss of response. *Scand J Rheumatol*. 2015;44(5):359-362. doi:10.3109/03009742.2015.1022213

34. Horneff G, Becker I. Definition of improvement in juvenile idiopathic arthritis using

the juvenile arthritis disease activity score. *Rheumatol Oxf Engl*. 2014;53(7):1229-1234. doi:10.1093/rheumatology/ket470

35. Foell D, Wulffraat N, Wedderburn LR, et al. Methotrexate withdrawal at 6 vs 12 months in juvenile idiopathic arthritis in remission a randomized clinical trial. *JAMA*. 2010;303(13):1266-1273. doi:10.1001/jama.2010.375

36. Anink J a, Van Suijlekom-Smit LWA a, Otten MH a, et al. MRP8/14 serum levels as a predictor of response to starting and stopping anti-TNF treatment in juvenile idiopathic arthritis. *Arthritis Res Ther*. Published online 2015. doi:10.1186/s13075-015-0723-1

37. Kim K, DeMets DL. Design and analysis of group sequential tests based on the type I error spending rate function. *Biometrika*. 1987;74(1):149-154. doi:10.1093/biomet/74.1.149

38. Lachin JM. A review of methods for futility stopping based on conditional power. *Stat Med*. 2005;24(18):2747-2764. doi:10.1002/sim.2151

39. R Core Team. R: A language and environment for statistical computing. Published online 2021. https://www.R-project.org/

11 Abbreviations and Acronyms

ADA anti-adalimumab antibody
BCVA best corrected visual acuity
CHAQ Childhood Health Assessment Questionnaire (MOP Section 6.6.1)
CVFQ Children's Visual Functioning Questionnaire (MOP Section 6.6.2)
DSMC Data and Safety Monitoring Committee
ESR Erythrocyte sedimentation rate
EYE-Q The Effects of Youngster's Eyesight on Quality of Life (MOP Section 6.6.3)
EQ-5D-Y EuroQol 5 Dimension Youth Survey (MOP Section 6.6.4)
ICH International Council for Harmonisation
JADAS Juvenile Arthritis Disease Activity Score (MOP Section 6.4.1)
JIA Juvenile idiopathic arthritis
MRP 8/14 Myeloid-related protein 8/14 calcprotein
NEI National Eye Institute
SAP Statistical Analysis Plan

12 Document Revision History

Version	Date	Summary of Changes, Justification, and Timing vis-à-vis key trial events (enrollment completion, interim analyses, unmasking, etc)
1	2019-08-16	Original Document
2	2020-09-02	Incorporated input from the trial's first DSMC meeting, and streamlined the document, including: • Changed the primary analysis to include fixed effects for randomization strata (no random effects). Harmonized the language throughout the SAP to reflect this change • Reduced the number of ancillary analyses in the additional analysis section for Aim 1. • Updated basic data summaries to be by site and country (rather than site only) since many sites will enroll fewer than 10 participants. • Added patient masking analysis (Section 3.1.1) • Clarified multiple imputation for secondary analyses (Sections 3.2.3 and 3.4)
3	2021-11-01	• Included CAU participants in the study population (Section 2.2) • Removed confusing language that mentioned a 'primary objectives,' and diistracted from the Primary Outcomes (Section 2.3.1-2.3.3) • Added an analysis to the statistical considerations to investigate effect modification between CAU and JIA-associated uveitis patients (Section 3.1.1) • Added the rationale for the stopping rules regarding efficacy (Section 3.7)
4	2023-06-09	• Removed ANA positivity requirement for CAU per Protocol 1.6.1 (Section 2.2) • Added a pre-specified interim analysis with stopping guidance (Section 3.7) • Added decision making schema regarding database closure dates for if/when stopping guideline criteria are met (Section 6.2) • Clarified that occurrence of vitreous haze, retinal or choroidal inflammation, or macular edema at a single visit constitutes treatment failure, in line with Protocol

		v1.6 (Section 2.3.1) • Removed age cap for the CHAQ, EYE-Q and EQ-5D-Y quality of life questionnaire analyses to include patients older than 18 years (Section 2.3)
5	2024-02-14	• These changes to the SAP were made following the interim analysis after unmasking. • Added detail to the primary analysis permutation test statistic (specified it is the log hazard ratio, with 100,000 permutations) and clarified estimation of the 95% confidence interval for the hazard ratio based on the Cox proportional hazards model (Section 3.1.1). • Removed a proposed (ancillary) Bayesian analysis of the trial after the trial was stopped at the interim analysis (Section 3.1.1). The investigator team determined there was insufficient time to conduct a survey of experts to construct a prior for the analysis, and the treatment effect was so much larger than anticipated it was determined there would be little added from a re-analysis of the trial from a Bayesian perspective. • Added longitudinal analyses of secondary outcomes (BCVA, JADAS, adalimumab drug levels, adalimumab antibody levels) at the interim analysis. This was added following the interim stopping of the trial to report key secondary outcomes at this time, before all patients had completed 12 months follow-up (Section 3.1.1). • Added analyses for time to regaining control after treatment failure. This was added following the interim analysis and unmasking, per the DSMC's recommendation to further elucidate post-failure patient outcomes in the presence of incomplete follow-up, before all patients completed 12 months of follow-up (Section 3.1.3)

13 Appendix

13.1 Estimates of interim efficacy calculations

Here, we have included the details of the interim efficacy calculations. Calculations were implemented using R's gsDesign package. They assume a single interim efficacy analysis takes place when 40% of the treatment failures (n=34 of an anticipated 83 total failures) occur. The calculations further assume that the interim analysis will spend 0.01 of the two-sided 0.05 alpha, using a Kim-DeMets power spending function.

```
>
> #--------------------------------
> # ADJUST trial
> # Interim analysis planning
> #
> # Ben Arnold. June 8, 2023
> #
> # the assumptions and approach
> # follow the vignette online at:
> # https://cran.r-
project.org/web/packages/gsDesign/vignettes/gsSurvBasicExamples.html
> #--------------------------------
>
> #--------------------------------
> # preamble
> #--------------------------------
> rm(list=ls())
> library(gsDesign)
>
>
> #--------------------------------
> # fixed design calculation
> #
> # Note: this calculation uses
> # a slightly different approach
> # from the original trial's sample
> # size calculation. All approaches
> # suggest the trial will require
> # 81-83 treatment failures to have
> # 88% power to detect a HR=2.0
> # The original calculation assumed
> # patients would be enrolled
> # and followed until exactly 48 weeks,
> # the timing of the primary endpoint. This
> # mimics the actual design.
```

```
> # gsDesign assumes that patients
> # will enrolled over a period (we assume 1 week)
> # and will be followed for a minimum of 48 weeks
> # but may continue to be followed past
> # that point. There is also a slightly difference
> # in how the package handles losses to follow-up
> # compared with the original calculation.
> # the slight difference in approach results in
> # the same total number of events required
> # with slightly different total sample sizes (118 vs 122)
> # however, this nuance would not change
> # the interim calculations and adjustment
> # which is based on number of events, not total patients
> # -------------------------------
>
> # Median Adalimumab time-to-event
> # assume time to event in STOP/placebo is 10
> # and time to event in continue is 20
> median_fail0 <- 20
> # Exponential dropout rate per unit of time (assume 10% per year)
> # exp(-0.0022*48 weeks ) = 0.9
> eta <- 0.0022
>
> # Hypothesized experimental/control hazard ratio (alternate hypothesis)
> # The original sample size calculation design effect is HR = 2. (or 0.5)
> hr <- 0.5
> # Null hazard ratio (1 for superiority, >1 for non-inferiority)
> hr0 <- 1
> # Type I error (1-sided)
> alpha <- .025
> # Type II error (1-power)
> # set at 12% (88% power, per original sample size calculation)
> beta <- 0.12
>
> # accrual assumptions
> # Enrollment period duration (only one)
> # assume all patients are enrolled simultaneously, over 1 week
> R <- c(1)
> # Follow-up duration of last patient enrolled
> minfup <- 48
> # Study duration (weeks) / 4 years = 192 weeks
> # if set to NULL, then nSurv solves for total study duration
> # required for the last patient to complete minfup (48 wks)
> T <- 49
>
> # Relative enrollment rates during above periods (constant)
> gamma <- c(1)
> # Randomization ratio, experimental/control
> ratio <- 1
>
```

```
> #-------------------------------
> # calculate the fixed design,
> # using gsDesign's nSurv()
> #-------------------------------
> d_fixed <- nSurv(
+   lambdaC = log(2) / median_fail0,
+   hr = hr,
+   hr0 = hr0,
+   eta = eta,
+   gamma = gamma,
+   R = R,
+   T = T,
+   minfup = minfup,
+   ratio = 1,
+   alpha = alpha,
+   beta = beta
+ )
>
> d_fixed
Fixed design, two-arm trial with time-to-event
outcome (Lachin and Foulkes, 1986).
Solving for:  Accrual rate
Hazard ratio                  H1/H0=0.5/1
Study duration:                   T=49
Accrual duration:                   1
Min. end-of-study follow-up: minfup=48
Expected events (total, H1):        81.0482
Expected sample size (total):       122.2344
Accrual rates:
    Stratum 1
0-1  122.2344
Control event rates (H1):
      Stratum 1
0-Inf    0.0347
Censoring rates:
      Stratum 1
0-Inf    0.0022
Power:                 100*(1-beta)=88%
Type I error (1-sided):   100*alpha=2.5%
Equal randomization:          ratio=1
>
>
>
> #-------------------------------
> # Allow for a single interim analysis
> #-------------------------------
>
> # Number of analyses (interim + final)
> k <- 2
> # Timing of interim analyses (k-1 increasing numbers >0 and <1).
```

```
> # Proportion of final events at each interim.
> timing <- c(0.4)
> # Efficacy bound spending function.
> # We use Lan-DeMets Power spending function .
> # No parameter required for this spending function.
> # with power parameter set to 1.75, which corresponds to
> # a nominal two-sided P-value of 0.01 at the interim with 40% of events
> sfu <-  sfPower # use sfLDOF approximating O'Brien-Fleming bound
> sfupar <- 1.75 # use NULL if O'Brien-Fleming
> # Futility bound spending function -- NOT used, set test.type = 1 below
> sfl <- sfLDOF
> # Futility bound spending parameter specification -- NONE set test.type = 1 below
> sflpar <- 0 # NULL
>
> # Generate design
> d_seq <- gsSurv(
+   test.type = 1,
+   k = k, timing = timing, R = R, gamma = gamma, eta = eta,
+   minfup = minfup, T = T, lambdaC = log(2) / median_fail0,
+   hr = hr, hr0 = hr0, beta = beta, alpha = alpha,
+   sfu = sfu, sfupar = sfupar, sfl = sfl, sflpar = sflpar
+ )
>
> d_seq
Time to event group sequential design with HR= 0.5
Equal randomization:          ratio=1
One-sided group sequential design with
88 % power and 2.5 % Type I Error.

  Analysis N   Z   Nominal p Spend
         1 34 2.57     0.005 0.005
         2 84 2.01     0.022 0.020
     Total                   0.0250

++ alpha spending:
 Kim-DeMets (power) spending function with rho = 1.75.

Boundary crossing probabilities and expected sample size
assume any cross stops the trial

Upper boundary (power or Type I Error)
          Analysis
   Theta      1      2 Total E{N}
  0.0000 0.0050 0.0200 0.025 82.9
  0.3482 0.2857 0.5943 0.880 68.9
             T        n   Events HR efficacy
IA 1  12.75148 125.3428 33.24370       0.410
Final 49.00000 125.3428 83.10925       0.643
Accrual rates:
    Stratum 1
```

```
0-1    125.34
Control event rates (H1):
      Stratum 1
0-Inf      0.03
Censoring rates:
      Stratum 1
0-Inf         0
> summary(d_seq)
[1] "One-sided group sequential design with 2 analyses, time-to-event outcome with
sample size 126 and 84 events required, 88 percent power, 2.5 percent (1-sided) Type I
error to detect a hazard ratio of 0.5. Enrollment and total study durations are
assumed to be 1 and 49 months, respectively. Efficacy bounds derived using a Kim-
DeMets (power) spending function with rho = 1.75."
>
>
> proc.time()
   user  system elapsed
  0.535   0.076   0.601
```