Core Adaptive Design Report

UPMC REMAP: Perioperative Medicine (Periop)

TABLE OF CONTENTS

ABBREVIATIONS

ADR: Adaptive Design Report
HFD: Hospital Free Days
ITT: Intention to Treat
MCMC: Markov Chain Monte Carlo
mITT: Modified Intention to Treat
OR: Odds Ratio
PP: Per Protocol
RAR: Response Adaptive Randomization

1. DOCUMENT VERSION

The version of the Core Adaptive Design Report is in this document's header and on the cover page.

1.1. Version History

Version 1: 3 April 2023
Version 2: 14 June 2023

- Changed analysis population to be modified intention to treat. This will include all patients who have undergone surgery and analyzing them by the interventions to which they were randomized.
- Modified first interim timing to be at least three months from the initiation of the trial
- Updated definition of time buckets. The time buckets are defined from the most recent date of randomization with a known-30 day HFD outcome.
- Clarified burn-in period for a new intervention prior to RAR beginning. The burn-in period will be at least 100 patients randomized to the new intervention and with 30-day HFD outcomes observed.
- Included a specific proportion of missingness for a covariate (10%) for when a missing category will be defined.

2. DESIGN OVERVIEW

UPMC REMAP Periop is a platform trial evaluating multiple interventions simultaneously within one overarching infrastructure. The platform consists of multiple domains (sets of competing interventions with a common clinical mode) that run simultaneously, and domains may be opened/closed during the duration of the platform. Each domain contains two or more interventions that are compared against one another. Participants in the platform can be randomized to exactly one intervention per domain for which they are eligible and may be randomized to multiple domains simultaneously based on their eligibility and the availability of domains by enrollment site. Each domain may have adaptive features such as adding/removing interventions based on prespecified triggers and/or response adaptive randomization. Adaptive analyses will be conducted regularly during the conduct of the platform, and pre-specified statistical triggers will be evaluated at each analysis. If a statistical trigger is met at an adaptive analysis, there would be a resulting platform conclusion and potential adaptation of the design. The platform may have modular results in which results of one domain are publicly disclosed while other domains are still enrolling.

The platform is built prospectively to have a flexible and modular structure. The flexible aspects of the design are planned and part of the protocol. In contrast to trials with one or few interventions, this platform is designed generically for a flexible number of domains/interventions with the

possibility that the design evolves as the science evolves. This Adaptive Design Report (ADR) describes the generic elements of the statistical design which apply to all domains in the platform. Each individual domain will have a domain-specific ADR that defines domain-specific design features including additional key endpoint(s), adaptations, statistical modeling/triggers, and operating characteristics. The status of the platform at any point in time will be included in a separate Current State of the Statistical Models document.

2.1. Primary Endpoint

The primary endpoint for the platform is hospital free days (HFD) at 30 days after the surgical encounter. HFD is the total number of days free of the hospital within 30 days after surgery. This is calculated by subtracting the total number of days hospitalized from 30. The number of days hospitalized includes the index hospitalization and any days readmitted within 30 days after surgery. Death within 30 days after surgery is recorded as the worst possible outcome of –1 HFD. The possible values of HFD are, from worst to best, –1 (death), 0, 1, …, 30 (zero days hospitalized).

2.2. Domain-Specific Key Secondary Endpoints

Domains may specify domain-specific key secondary endpoints which may be used to evaluate statistical triggers within the domain. Details on domain-specific key secondary endpoints including statistical modeling and conclusions will be summarized within the domain-specific ADR.

2.3. Analysis Populations

The primary analysis set will consist of all participants that are randomized to at least one intervention within at least one domain and have undergone the intended surgery. The primary analysis set will apply a modified intention to treat (ITT) principle by analyzing all participants who have undergone surgery by the interventions to which they were randomized. All participants in the perpetual trial will become a part of the accruing trial dataset and remain in the primary analysis set at adaptive analyses while the trial is running. Domain-specific key secondary endpoints may be analyzed in analysis populations that are restricted to specific domains.

3. PRIMARY STATISTICAL ANALYSIS MODEL

Inferences in this trial are based on a Bayesian statistical model, which models the posterior probability of HFD outcomes for each of the interventions within each domain. The model

incorporates the empirical outcomes that accumulate during the trial in terms of the observed HFD outcomes and prior knowledge in the form of a prior distribution. The model is designed to evolve throughout the duration of the platform. If new domains/interventions are added, then new parameters are added to the statistical model. The modeling principles described in this section are used in the evolving form of the statistical model.

The primary analysis model of HFD is a Bayesian cumulative logistic (proportional odds) model. The model incorporates effects for all interventions across all domains within a single model. Each domain in the platform contributes a set of additive treatment effects for the domain interventions. The form of the additive effects for each domain depends on the design features of the specific domain and are defined in each domain-specific ADR. In this document, we provide default formulations for additive effects for each domain which may be modified by each domain-specific ADR. Unless otherwise noted, it is assumed that there are no interactions between interventions across domains.

3.1. Primary Model Formulation

Let $Y_i \in \{-1, \dots, 30\}$ denote the HFD outcome for participant i. We use $\gamma_{i,j}$ to denote the probability participant i has an HFD outcome of j or lower. For $j = -1, \dots, 29$, the cumulative logistic model formulation is:

$$\text{logit}(\gamma_{ij}) = \boldsymbol{s_i^T \alpha_j} + \boldsymbol{z_i^T \beta} + \boldsymbol{x_i^T \theta}, \qquad (1)$$

where $\text{logit}(\gamma_{ij}) = \log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right)$ is the log odds of γ_{ij}. The vector $\boldsymbol{s_i} = (s_{i1}, \dots, s_{iS})$ includes indicators of the surgery type received by participant i. The intercept parameters $\boldsymbol{\alpha_j}$ determine the cumulative probabilities of HFD outcomes by surgery type for the reference participant group (all $\boldsymbol{z_i}$ and $\boldsymbol{x_i}$ equal to zero). The variables $\boldsymbol{z_i} = (z_{i1}, \dots, z_{ip})$ represent any covariates included in the model, and the vector $\boldsymbol{x}_i = (x_{i1}, \dots, x_{iK})$ consists of indicators of intervention assignment for each of the K interventions included in the model. If participant i was randomized to intervention t at baseline, then $x_{it} = 1$; otherwise $x_{it} = 0$. The parameter $\boldsymbol{\theta}$ is the *K*-dimensional vector of treatment effects for interventions $1, \dots, K$. Because the $\boldsymbol{\theta}$ vector is common across all j levels, the treatment effects are proportional effects. Each DSA will specify the full additive model within that domain for modelling the treatment effects for each intervention. Since the model is parameterized in terms of

the cumulative probability of negative events (γ_{ij} is the probability of an HFD outcome of j or worse), positive values of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ correspond with a higher probability of bad HFD outcomes and negative values correspond with a lower probability of bad HFD outcomes. Model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ will also be exponentiated and presented as odds ratios for one group of patients relative to the reference group.

The statistical model adjusts for the variation in HFD outcomes by age, sex, hospital site, surgery type, time epoch, randomization within each domain, and baseline preoperative health class. The prior distributions for each model parameter are described in the sections below.

3.2. HFD Distribution by Surgery Type

Separate distributions of HFD are estimated by surgery type. For surgery type $s \in 1, \ldots, S$, the prior distribution for the $\alpha_{j,s}$ parameters is a logistic transformation of a Dirichlet distribution:

$$\boldsymbol{\pi_s} \sim Dirichlet(\boldsymbol{p_{0,s}}),$$

$$\alpha_{j,s} = logit\left(\sum_{l=1}^{j} \pi_{l,s}\right)$$

With the Dirichlet concentration parameters ($\boldsymbol{p_{0,s}}$) pre-specified based on the observed rates of HFD outcomes in pre-trial data. The total weight of the Dirichlet prior concentration parameter is equivalent to 1 participant worth of information.

3.3. Time (Era) Effects

The date of randomization for each participant will be adjusted for in the primary analysis model. Sequential 13-week time eras will be defined from the most recent date of randomization with a known 30-day HFD outcome backwards in time to the start of randomization. The most recent time period is the 13-week period preceding the most recent randomization with a known 30-day HFD outcome and is considered the reference group in the model. Effects for each previous era are estimated relative to the most recent era with a first-order normal dynamic linear model (NDLM). The first-order NDLM is sequentially defined moving backwards in time,

$$\beta_t \sim N(\beta_{t-1}, \tau^2_{time}), t = 1, \ldots, T,$$

With the following hyperprior on the variance parameter:

$$\tau^2_{time} \sim IG(0.25, 0.1).$$

The NDLM model for the eras smooths the estimate of each era over the course of the trial. The parameter τ^2_{time} is the variance component that controls the amount of smoothing from one era to the next. This parameter is estimated by the data with a hyper-prior distribution. The prior distribution is an Inverse-Gamma distribution with weight of 1 observation worth of data that the era effects have small changes, 0.10, from one era to the next.

3.4. Age Effects

Age will be categorized into three groups: 1) Age 40 or lower, 2) Age 41 to 60, and 3) Age 61 or higher. The middle age group, 41 to 60, will be the reference group. Effects will be estimated for the remaining groups with independent N(0,1) priors.

3.5. Sex Effects

Sex will be adjusted for in the HFD model. Males will be the reference group, and an effect will be estimated for females relative to males with a N(0,1) prior.

3.6. Hospital Effects

The model includes an adjustment for HFD outcomes by hospital. At the first adaptive analysis, the largest hospital is selected as the reference group, and all other hospital effects are estimated relative to the reference. For model stability, the reference hospital may be changed during the trial depending on enrollment. All hospital effects are given independent N(0,1) prior distributions.

3.7. Preoperative Health Score

The model includes an adjustment for the baseline American Society of Anaesthesiologists' (ASA) score of each participant. The ASA score is a subjective assessment of a participant's overall health that spans from level I (completely healthy) to level V (not expected to survive 24 hours). The reference group for the covariate effects will be Level I, and effects will be estimated for every category relative to the reference group. Each effect will be estimated with independent N(0,1) prior distributions. If there are any ASA score categories with zero patients, the effects for those categories will be fixed to zero.

3.8. Domain Randomization Effects

The model includes a covariate for whether each participant was randomized within each domain in the model. For each domain, an indicator is defined based on whether the participant was randomized within the domain. If the participant was not randomized within a domain (based on eligibility of the participant or availability of the domain), the indicator variable is set to 0. Each effect for randomization within a domain is estimated with independent N(0,1) priors.

3.9. Intervention Effects

The prior for each intervention effect is defined in the domain-specific ADR. In this section, we describe default priors for intervention effects that may be modified for each domain. Within each domain, one intervention, typically the control/standard of care if applicable, will be the reference intervention with treatment effect fixed to zero. The effect of all remaining interventions in the domain will be estimated relative to the reference intervention. The default prior for each intervention effect is an independent N(0,1) prior.

4. STATISTICAL QUANTITIES

The posterior distributions of the model parameters are calculated using a Markov Chain Monte Carlo (MCMC) algorithm. The MCMC algorithm allows for the generation of a large number (e.g., 100,000) of draws from the joint posterior distributions of unknown model parameters. Posterior quantities are calculated based on the posterior samples and used to implement the adaptive features of the design including the evaluation of statistical triggers and allocation updates. In this section, the general term of a “larger treatment effect” means a treatment effect indicating benefit of the relevant treatment. Figure 1 shows a visual representation of the posterior probability quantities on a hypothetical treatment effect distribution for one intervention relative to a comparator. The vertical lines in each panel represent hypothetical pre-specified thresholds for each trigger.

4.1. Posterior probability an intervention is optimal in a domain

The posterior probability that an intervention is optimal within a domain is calculated as the proportion of posterior draws in which the intervention has the largest (most beneficial) treatment effect across all interventions within the domain. This calculation is restricted to *active* interventions in the domain, i.e., interventions that are still actively randomizing.

4.2. Posterior probability of superiority to a comparator intervention

The posterior probability that one intervention is superior to another intervention within the domain is calculated as the proportion of posterior samples in which the intervention has a larger treatment effect than the comparator intervention.

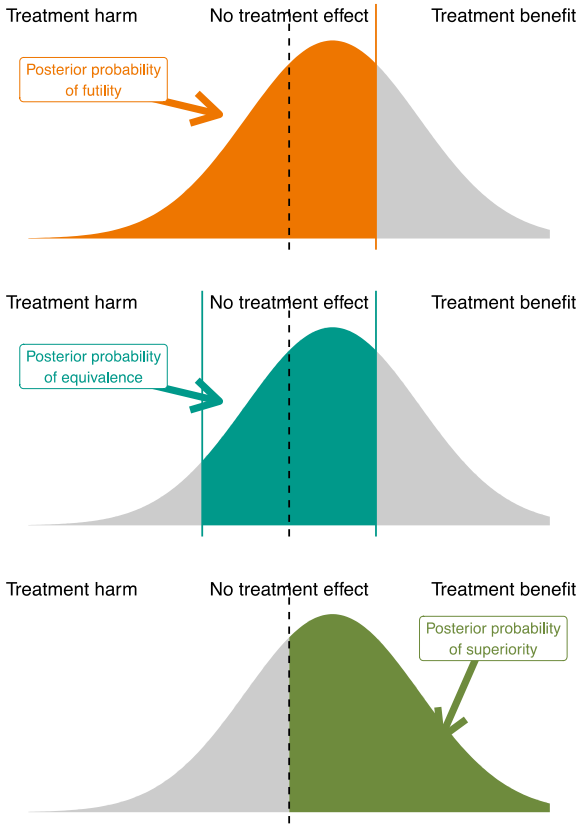
4.3. Posterior probability of equivalence

The posterior probability that one intervention is equivalent to another intervention within the domain is calculated as the proportion of posterior samples in which the intervention treatment effect is within a pre-specified threshold from the comparator intervention treatment effect. The equivalence threshold will be pre-specified in each domain-specific ADR.

4.4. Posterior probability of futility

The posterior probability of futility for one intervention compared to another intervention is calculated as the proportions of posterior samples in which the treatment effect is no larger than a pre-specified futility threshold. The futility probability is equivalent to a one-sided version of the equivalence trigger.

Figure 1. Visualization of posterior probability calculations on a hypothetical treatment effect posterior distribution



5. COMMON DESIGN FEATURES

5.1. Adaptive Analyses

This platform is designed to be perpetual with no designated end or sample size cap. The goals of the trial are to both treat participants effectively while also investigating the relative benefit of different interventions. Adaptive analyses will be conducted throughout the trial process. The first adaptive analysis will be conducted at least three months after the initiation of the trial. After the first adaptive analysis, they will be planned to be repeated quarterly for the remainder of the trial. A regular time point (e.g., first of the month) will be selected to trigger the running of adaptive analyses. Adaptive analyses may be skipped if enrollment is slow and will be communicated to the statistical analysis committee in advance of the adaptive analysis. At the time of an adaptive analysis, the HFD analysis will only use data for participants that have complete 30-day HFD endpoints at the time of data extraction and the opportunity to complete 30 days. The analysis model for domain-specific secondary analysis models will use data from participants with the

opportunity to complete the domain-specific endpoint. The analysis model run at the adaptive analysis will be used to evaluate the pre-defined statistical triggers for each domain and to update allocation proportions if RAR is specified in the domain.

5.2. Allocation

Participants will be randomized to one intervention within each domain for which they are eligible. To be randomized within a domain, a participant must be eligible for the domain and at least two interventions within the domain. Details of the allocation proportions will be described in each domain-specific ADR. Allocation to interventions in a domain may be fixed over time or may be adaptively adjusted based on the accruing efficacy data.

5.3. Response Adaptive Randomization

Some domains may use response adaptive randomization (RAR) to adaptively adjust the allocation proportion to interventions based on the accruing efficacy data. The data on the primary endpoint (HFD) will determine the randomization proportions for each intervention within the domain.

When a domain is added to the platform, there will be an initial period (the burn-in) in which randomization is fixed and equal for each intervention in the domain. The initiation of RAR will be specific to the domain and may be the first adaptive analysis after the domain has been added to the platform or after certain platform conclusions are made. Once initiated, RAR will be used to update allocation for each intervention in the domain. The randomization for each participant is based on the probability that each intervention is optimal within the domain. For domain $d \in 1, \dots, D$, we use $O_d(j)$ to denote the probability that intervention $j \in 1, \dots, J$ is optimal in domain d. The probability a participant is randomized to intervention j in domain d is denoted $\rho_D(j)$ and calculated as:

$$\rho_D(j) = (1 - w)O_d(j) + (w)\left(\frac{1}{J}\right).$$

The allocation probability is a weighted average of the probability the intervention is optimal in the domain and equal allocation to all interventions. The weight, $w \in (0,1)$, is pre-defined within each domain-specific ADR. A weight of $w = 0$ would result in RAR proportions that are proportional to $O_d(j)$. A weight of $w = 1$ would simplify to fixed equal randomization within the domain. Typical values of w will be small values (0.1-0.25) depending on the number of interventions within the

domain. The choice of w will imply a minimum/maximum RAR probability for each intervention in the domain.

5.4. Introduction of New Interventions

If a domain is using RAR and a new intervention is added to the domain, the randomization to the new intervention will be "blocked" during a burn-in period to guarantee a minimum sample size before RAR is applied. If there are K interventions in the domain, a fixed allocation of 1/K will be used for the new intervention. The allocation proportion for other interventions will be re-normalized to sum to 1-1/K. The burn-in period for the new intervention will last until at least 100 participants have been allocated to the new intervention and have observed their 30-day HFD outcomes. At the next adaptive analysis, a full update to the RAR proportions will be implemented if applicable. Note that if RAR has not been initiated in the domain, the new intervention will continue with fixed randomization even after the burn-in period until the domain-specific rules are met to start RAR.

5.5. Missing Data

There is no imputation of missing HFD outcomes, and participants with missing HFD outcomes will be excluded from the modeling. Unknown covariate values may be imputed. For categorical variables, the most prevalent covariate value will be imputed. For continuous variables, the mean value will be imputed. If the amount of missingness for a covariate is substantial (>10%), a missing category may be defined.

6. STATISTICAL TRIGGERS

The Core ADR defines a set of default statistical triggers for the platform. Each domain-specific ADR will specify the set of statistical triggers that will be applied. At each adaptive analysis, the set of pre-defined statistical triggers will be evaluated. If a statistical trigger is met, design adaptations may occur including public disclosure of the results, removal of interventions, and/or closure of domains. The possible actions taken after each trigger are described below but may vary depending on the context of the individual domains and interventions being compared.

6.1. Intervention Superiority

At any adaptive analysis, if a single intervention has at least a 99% posterior probability of being the optimal intervention within a domain, then that intervention will be deemed as superior to all other interventions in that domain. If this trigger is met, it is expected that the domain is closed, future participants will receive the superior intervention, and a public disclosure of results is made.

6.2. Intervention Inferiority

At any adaptive analysis, if a single intervention has less than a 1% posterior probability of being the optimal intervention within a domain, then that intervention will be deemed as inferior to the other interventions in that domain. If this trigger is met, it is expected that the inferior arm will be dropped from randomization. Domains may implement specific rules that determine when an arm may drop for inferiority, including delaying the action until other conclusions have been met. Public reporting of the platform conclusion may occur depending on whether reporting would be unblinding to the effectiveness of ongoing interventions. If a domain has three or more interventions, the probability threshold of 1% may be adjusted based on the number of interventions in the domain.

6.3. Equivalence of Two Interventions

If two interventions have at least a 90% posterior probability of being equivalent, defined as relative treatment effects within a prespecified equivalence margin, then the two interventions will be deemed as equivalent. The equivalence margin depends on the endpoint and type of treatment effect parameter (odds ratio, difference in means, etc.). For HFD, the default equivalence margin is an odds ratio between 1/1.2 (0.83) and 1.2 for the comparison of two interventions. If the equivalence trigger is met, the action will be determined by the TSC and may include dropping of one of the two equivalent interventions, pooling of the equivalent interventions in the statistical model, or other appropriate actions.

6.4. Intervention Futility

At an adaptive analysis, an intervention may be declared as futile if the probability of a large treatment effect compared to another intervention is small. For example, for HFD, a futility trigger may be defined as a less than 5% probability that the odds ratio comparing two interventions falls below 0.9 (where OR<1 indicates benefit). Futility will only be applied in specific scenarios, for

example, comparing the performance of combination interventions to the single agent components. The futility margin and probability threshold will be prespecified within each domain-specific ADR.