Statistical Analysis Plan for the Post-Operative Nausea and Vomiting (PONV) Prophylaxis Domain

UPMC REMAP: Perioperative Medicine (Periop)

TABLE OF CONTENTS

1 SAP VERSION

The version is in this document's header and on the cover page.

1.1 Version history

Version 1: Finalized on 7 March 2024

2 SAP AUTHORS

Alison Althans, University of Pittsburgh

Lindsay Berry, Berry Consultants, Austin, Texas, USA

Stephen Esper, University of Pittsburgh

Jennifer Holder-Murray, University of Pittsburgh

Liz Lorenzi, Berry Consultants, Austin, Texas, USA

3 INTRODUCTION

This plan details the statistical analyses of the PONV Prophylaxis domain as outlined in the Core Adaptive Design Report (ADR) and PONV Prophylaxis Domain ADR. The ADRs outline analysis conventions for the overall platform and the domain, specifically, related to the primary and key secondary endpoints in the platform. This plan is prespecified for the unblinding of the data for the optimal and supraoptimal prophylaxis interventions in the PONV domain. The authors of this SAP are blinded to individual data but have been informed by the trial monitoring committee that supraoptimal prophylaxis met the prespecified statistical trigger for superiority to optimal prophylaxis based on the PONV endpoint. Additionally, after the decision was made to close the PONV domain and prior to finalizing the SAP, investigators reviewed summary interim data including estimates of treatment effect within the PONV domain. The primary analysis for this SAP will be conducted once the last patient randomized has completed 30-days of follow-up on the primary endpoint.

4 DESIGN CONSIDERATIONS

Periop explores multiple treatment domains by randomizing patients within multiple domains simultaneously. The adaptive platform trial was designed to produce modular results for individual interventions or full domains upon reaching platform conclusions. Since a predefined statistical trigger for superiority was met, the results for the supraoptimal and optimal prophylaxis interventions will be unblinded and made public. This document prespecifies the analysis plan for this unblinding.

Periop is designed with Bayesian analyses for the primary and key secondary endpoints in the trial. For each endpoint, one overarching Bayesian model, prespecified in the Adaptive Design Reports (ADRs), drives all adaptations, statistical triggers, and result summaries across all domains in the platform. The primary and key secondary analysis model will be used to report the results for the interventions in the PONV domain.

Periop is a perpetual adaptive platform trial with no planned maximum sample size. Periop defines several statistical triggers within the trial that, at any analysis of the trial, would result in public disclosure and a declaration of a platform conclusion. The following statistical triggers were defined for the PONV Domain:

1. Intervention Superiority. If one intervention has a posterior probability of 99% or greater that it is the optimal intervention in the domain, then a declaration of superiority of that intervention would be made. This statistical trigger was applied to the HFD and PONV endpoints.

2. Equivalence. If there is a greater than 90% probability that the odds ratio comparing the two interventions fall between 1/1.2 (0.83) and 1.2, then equivalence of the two interventions will be declared. This statistical trigger was applied to the HFD and PONV endpoints.

The 99% threshold for superiority was selected to have good properties for potential trial sizes. For example, the one-sided type I error rate of any conclusion of superiority for any single intervention vs another is less than 5% up to a total sample size of 5000 patients accounting for multiple interim analyses (see main and domain-specific ADRs).

5 UNBLINDING

Periop evaluates treatments across multiple treatment domains, and there are currently two domains to which patients are randomized. At the unblinding of the PONV domain, there are other interventions to which patients have been randomized in other domains that will not be unblinded at this analysis (the Analgesia domain is ongoing). The predefined primary analysis model will adjust for randomization status in all domains. Thus, in this SAP, there are analyses planned that will be conducted by the Statistical Analysis Committee (SAC) using additional randomizations and unblinding of other randomizations. The SAC is unblinded to all interventions and domains as part of their role in implementing the adaptive design for Periop. Additionally, some secondary analyses in the SAP will be conducted with only knowledge of the PONV domain allocation status for patients. These analyses may be conducted by investigators who are blinded to information about other interventions and domains.

6 INTERVENTIONS

There are two interventions within the PONV Prophylaxis Domain:

- A1: Optimal prophylaxis
- A2: Supraoptimal prophylaxis

Patients were randomized with fixed, equal randomization (1:1) between the two interventions. Response-adaptive randomization was not utilized in this domain.

7 ANALYSIS POPULATIONS

Analyses will be conducted in the following analysis populations:

1. Periop modified intention-to-treat (mITT): This population consists of all patients randomized to at least one intervention that have undergone the intended surgery. This analysis

population includes domains/interventions that are blinded to investigators, and analyses of this population will be performed by the SAC.

2. PONV mITT: This population consists of all patients randomized within the PONV domain that have undergone the intended surgery. Analyses in this population are restricted to unblinded domain/intervention assignments and may be conducted by investigators blinded to these ongoing groups.

Each of these analysis populations will include only patients randomized on or before the closure of enrollment in the PONV domain, December 14, 2023.

8 ENDPOINTS

The following end points will be analyzed, displayed graphically, and summarized through descriptive statistics.

1. **Hospital Free Days** (HFD)
   a. The primary endpoint for the platform is hospital free days at 30 days after the surgical encounter. HFD is an ordinal composite outcome incorporating death and duration of hospital stay up to 30 days. For patients that survive to 30 days, HFD is calculated by subtracting the total number of days hospitalized from 30, where the number of days hospitalized includes the index hospitalization and any subsequent readmissions up to day 30. Death within 30 days after surgery is recorded as the worst possible outcome in the ordinal scale. The possible values of HFD are, from worst to best: Death (labeled with –1), 0, 1, …, 30 (zero days hospitalized).

2. **Post-operative nausea and vomiting (PONV)**
   a. PONV is the key secondary endpoint for this domain. PONV is a dichotomous endpoint for whether the patient receives 1 or more doses of an antiemetic medication within 24 hours of surgery completion.

3. **30-day mortality**
   a. A dichotomous endpoint of death within 30 days of surgery. This endpoint corresponds to the –1 component of HFD.
   b. If fewer than 10 deaths have been observed in either of the two treatment groups, this endpoint will be summarized descriptively, and no analyses will be performed.

9 DESCRIPTIVE SUMMARIES

1. Continuous variables will be summarized using the following descriptive statistics: n, mean, standard deviation, median, interquartile range, minimum, and maximum.
2. Ordinal variables will be summarized by the cumulative frequency of each outcome. The 25th, 50th, and 75th percentiles will be summarized. All ordinal endpoints will be plotted using stacked bar plots and cumulative probability plots.
3. Dichotomous variables will be summarized by the number and proportion in each category.
4. Time-to-event variables will be summarized via Kaplan-Meier estimates and the observed quantiles (2.5th, 10th, 25th, 50th, 75th, 90th, and 97.5th percentiles, as available). Positive clinical event outcomes will be plotted as the cumulative rate of event, and negative events will be plotted as the cumulative rate of event-free.

10 BASELINE CHARACTERISTICS

The following demographics will be summarized by randomized assignment in the PONV mITT population: age, sex, body mass index, race/ethnicity, risk prediction, risk value, surgery type, preoperative ASA health score, modified Apfel score, hospital location, ERAS PowerPlan name, procedure grouping, MIS/Open/Converted, total procedure time, preexisting conditions (atrial fibrillation, CAD/MI, CHF, CKD/Acute chronic renal failure, COPD, depression, diabetes, GERD, hypertension), and tobacco use.

11 COMPLIANCE

The compliance will be summarized descriptively with the proportion of patients that received each medication and the number of medications received, for each randomized arm.

12 ANALYTIC APPROACH

Each inferential analysis will be done using a Bayesian model. A summary of the analyses methods is provided below. The analysis approach is based on the predefined primary analysis model in the platform ADR documents.

12.1 Primary Analysis of Primary Endpoint

The primary analysis model is a Bayesian cumulative logistic model for the ordinal primary endpoint (see Core ADR). The model estimates odds ratio effects for interventions relative to the reference intervention in the domain, where an OR<1 indicates better outcomes. The full details of the model

are predefined in the ADRs and the Current State of The Statistical Model document. The model adjusts for variation in outcomes by the following:

- Surgery type
- Age category (≤40, 41-60 (reference), 61+)
- Sex
- Hospital
- Pre-operative ASA score (level I (completely healthy) to level V (not expected to survive 24 hours)
- Time; 13-week buckets of time working backwards from the last enrolled patient.
- Each intervention within each domain

The analyses in this SAP use the following conventions:

- If there are small or empty counts in the HFD outcome, these categories may be collapsed with adjacent levels of the ordinal scale to improve model stability.
- All time buckets with <5 patients complete may be combined with the more recent neighboring bucket.
- If there are other covariate groups with <5 patients complete, the covariate may be collapsed to avoid issues with model stability.

1.1.1. Proportional Odds Assumption

The primary analysis model assumes a proportional effect of treatment across the scale of the ordinal outcome. To assess the robustness of the results to this assumption, a dichotomous model is fit to cumulative dichotomizations of ordinal outcome and the treatment effect odds ratio for each dichotomous break is presented. For HFD, odds ratios will be presented for each of the following dichotomous breaks: less than or equal to –1, 7, 14, 21, 22, 23, 24, 25, 26, 27, 28, and 29, respectively. No statistical test of proportional odds is conducted.

12.2 Analytic Approach for Key Secondary Endpoint

The key secondary endpoint is analyzed with a Bayesian logistic regression model (see ADRs and Current State of the Statistical Model). The logistic regression model adjusts for age category, sex, hospital, baseline preoperative health class, surgery type, and randomization within each domain. Covariates are specified in the same way as the HFD analysis described above. For factors that also appear in the HFD model, the same priors will be used in the PONV model (see Current State document). Priors on other parameters are provided in the PONV ADR.

12.3 Analytic Approach for Other Dichotomous Endpoints

A Bayesian logistic regression model will be used for each additional dichotomous outcome. The model will always specify the "event" as the negative outcome and be parameterized so that an odds-ratio <1 implies benefit to patients. The model is the standard logistic link function model:

$$log\left(\frac{\pi}{1-\pi}\right) = \alpha + [factors]$$

References will be made to the factors in the model and their prior distribution. Many of these factors will be the same as the primary analysis model, with the same priors, as the parameters have similar interpretation. If not otherwise specified, the prior distribution for the main effect is $\alpha \sim N(0, 1.82^2)$ (similar to a uniform prior on the probability scale).

12.4 Markov Chain Monte Carlo (MCMC) Model Stability

The Bayesian models have many parameters and there may be risk of poor model stability, including convergence and mixing behavior of the MCMC sampler. These instabilities may be based on sparse data on the outcome or covariates. The statisticians running the model may make changes that do not affect the overall interpretation but provide reliable model diagnostics and scientific rigor. Any alterations will be noted.

12.5 Model Outputs

For the ordinal and dichotomous endpoints, the odds-ratios will be summarized. Unless noted otherwise, all models will be parameterized so that an odds-ratio less than 1 indicates clinical benefit. The standard model outputs for each treatment effect will be the posterior median, 95% credible intervals, posterior mean, and posterior standard deviation. All credible intervals will be equal-tailed intervals, so 95% intervals will range from the 2.5th to 97.5th quantile of the posterior distribution.

For each inferential model, a posterior probability that one arm is superior will be provided for each comparison between arms. This posterior probability has been identified as the primary analysis metric between arms.

12.6 Exploratory Analyses

Exploratory analyses not identified in this SAP may be performed to support the clinical understanding of results. Any post-hoc or unplanned analyses that are reported will be clearly identified and will not be evaluated for statistical significance using formal hypothesis tests. Post-hoc exploratory analyses may use the following methods (or similar Bayesian versions of each analysis method):

- Ordinal endpoints will be compared using a cumulative proportional odds model with summaries of the odds-ratio and 95% confidence intervals.
- Time-to-Event analyses will utilize a Cox proportional hazards model (or Bayesian piecewise exponential model), summarizing the hazard ratios and 95% confidence intervals.
- Continuous endpoints will compare means with 95% confidence intervals based on two-sample t-test procedures.
- Dichotomous proportions will be compared using logistic regression summarizing the odds-ratio and 95% confidence intervals.

12.7 Missing Data

Missing outcome data is expected to be minimal, and no imputation of missing outcome data is planned for analyses. If any patients are missing covariate values, these will be imputed based on the modal value at a given site.

13 SPECIFIC PROSPECTIVE ANALYSES

The specific prospective analyses are summarized in the table and described in detail below.

#	Status	Population	Endpoint	Other
1	Primary	Periop mITT	HFD	Includes all platform interventions.
2	Key Secondary	Periop mITT	PONV	Includes all platform interventions. Same model as HFD without time effects.
3	Primary (Component)	Periop mITT	30-day mortality	Includes all platform interventions. Same model as HFD.
4	Sensitivity	Periop mITT	Dichotomized HFD	A logistic regression will be run for several dichotomizations of HFD as a robustness check.
5	Secondary	PONV mITT	HFD	PONV interventions only.
6	Secondary	PONV mITT	PONV	PONV interventions only.
7	Secondary	PONV mITT	30-day mortality	PONV interventions only.
8	Secondary	PONV mITT	HFD	Remove time effects.
9	Subgroup	PONV mITT	HFD	Differential effect by surgery subtype
10	Subgroup	PONV mITT	PONV	Differential effect by surgery subtype
11	Subgroup	PONV mITT	HFD	Differential effect by intrathecal morphine (received/did not receive)

12	Subgroup	PONV mITT	PONV	Differential effect by intrathecal morphine (received/did not receive)
13	Subgroup	PONV mITT	HFD	Differential effect by whether received dexamethasone
14	Subgroup	PONV mITT	PONV	Differential effect by whether received dexamethasone
15	Subgroup	PONV mITT	HFD	Differential effect by Apfel score (Level 0-4)
16	Subgroup	PONV mITT	PONV	Differential effect by Apfel score (Level 0-4)
17	Subgroup	PONV mITT	PONV	Differential effect by 2v3v4v5 (2, 3 are optimal with/without dex) (4,5 are supraoptimal with/without dex)
18	Subgroup	PONV mITT	HFD	Differential effect by 2v3v4v5
19	Sensitivity	PONV mITT	HFD	No covariate adjustment, keep intercepts by surgery type
20	Sensitivity	PONV mITT	PONV	No covariate adjustment
21	Sensitivity	PONV mITT	HFD	Prior on treatment effect is N(0, 10)
22	Sensitivity	PONV mITT	PONV	Prior on treatment effect is N(0, 10)

13.1 Reporting of Analysis Results

For each analysis model, the following summaries will be reported when applicable:

Parameter	Posterior Summary			
	Median	95% Credible Interval	Mean	Standard Deviation
Age ≤40				
Age 41-60				
Age 61+				
Female (vs male)				
Time Bucket t (for t in 1, …, T)				
Hospital h (for h in 1, …, H)				
Surgery type s (for s in 1, …, S)				
Pre-Surgery ASA Score (I-V)				
Supraoptimal prophylaxis (vs optimal prophylaxis)				
Main effect of subgroup				
Treatment effect by subgroup				

For each analysis model, the following comparisons will be made, when applicable:

- Supraoptimal prophylaxis will be compared to optimal prophylaxis for superiority. The posterior probability that the OR<1 will be used to define superiority. In subgroup analyses, this probability will be provided by subgroup.
- Supraoptimal prophylaxis will be compared to optimal prophylaxis for equivalence. A 90% or greater probability of an odds ratio between 1/1.2 (0.83) and 1.2 will be used as a statistical trigger for equivalence. In subgroup models, this probability will be provided by subgroup.
- For the sensitivity analysis assessing the proportional odds assumption, the simvastatin OR will be summarized for each dichotomization of HFD.