Medtronic **Statistical Analysis Plan**	
Clinical Investigation Plan Title	Evaluation of BIS™ and Levels of Sedation with the Common Inhalational Anesthetics in Healthy Volunteers (OLIVER)
Clinical Investigation Plan Identifier	MDT19053OLIVER
Clinical Investigation Plan Version	*Version 5.0 (06-JAN-2022)*
Sponsor/Local Sponsor	Medtronic, Minimally Invasive Therapies Group (MITG) Respiratory, Gastrointestinal & Informatics (RGI) 6135 Gunbarrel Avenue, Boulder, CO 80301 U.S.A.
Document Version	*Version 3.0 (08-SEP-2022)*

Table of Contents

1 Version History

Version	Summary of Changes	Author(s)/Title
1.0	• Initial Release	Alex Shih, PhD, Principal Biostatistician Zengri Wang, PhD, Statistics Director
2.0	• Minimal changes were made to reflect the updated version of CIP	Alex Shih, PhD, Principal Biostatistician
3.0	• Minimal changes were made to reflect the updated version of CIP	Alex Shih, PhD, Sr. Principal Biostatistician

2 List of Abbreviations and Definitions of Terms

Term	*Definition/Acronyms*
ADE	Adverse Device Effect
AE	Adverse Event
ASA	American Society of Anesthesiology
BIS™	Bispectral Index (BIS) technology monitoring uses processed EEG signals to measure sedation depth based on the level of consciousness
CBC	Complete Blood Count
CO_2	Carbon Dioxide. It can be measured with a capnograph, a device that measures the concentration of carbon dioxide from each inspired and expired breath. Gases are collected with a non-invasive side stream from the inhaled and exhaled gases of the subject. Capnograph outputs numeric values and waveforms of the fractioned concentration of CO_2 of each breath
CRF	Case Report Form. Forms where the clinical data are collected. eCRF is the electronic version of the CRF
EC	Ethics Committee
ECG	Electrocardiogram. A diagnostic tool that measures and records the electrical activity of the heart
ED	Effective dose
EDC	Electronic Data Capture. Electronic systems where the data are collected through the eCRF (Oracle Clinical). May also be referred to as RDC (Remote Data Capture).
EEG	Electroencephalogram
ET	End-tidal concentration

Term	*Definition/Acronyms*
EMR	Electronic Medical Record. Digital version of a patient's medical record within a single facility.
$EtCO_2$	End-tidal Carbon Dioxide. The value of exhaled carbon dioxide displayed by the capnograph device
GCP	Good Clinical Practice
ICF	Informed Consent Form
IRB	Institutional Review Board
ISF	Investigator Site File. Regulatory binder supplied by the sponsor
LMA	Laryngeal mask
LOC	Level of consciousness
MAC	Minimum alveolar concentration
MOAA/S	Modified Observer's Assessment of Alertness/Sedation
PI	Principal Investigator. The person responsible for overseeing the study and assuring study completion in compliance with applicable regulations.
PIC	Patient Interface Cable
PK/PD	Pharmacokinetics and pharmacodynamics
SpO2	A non-invasive spectroscopic estimate of arterial oxygen saturation measured transcutaneously by a pulse oximeter
SAP	Statistical Analysis Plan
SOP	Standard Operating Procedures
TCI	Target-controlled infusion
TIVA	Total Intravenous Anesthesia
UADE	Unanticipated Adverse Device Effect

3 Introduction

The Evaluation of BIS™ and Levels of Sedation in Health Volunteers (OLIVER) Study is a multi-center, prospective randomized study to collect data to evaluate the relationship between BIS™ and inhaled anesthetics without and with opioids. Healthy Volunteers will be sequentially assigned to one of the study groups of anesthetics regimens at each clinical site. The BIS™ bilateral sensor will be placed on the subject's forehead to study the effects of these drugs on the brain. Each volunteer will be given a dose range of sequences of the target drug to achieve targeted drug concentration. Five groups of anesthetics regimens will be studied as follows:

1. Sevoflurane alone
2. Sevoflurane with Remifentanil
3. Sevoflurane with Fentanyl
4. Desflurane alone
5. Isoflurane alone

At each steady-state step, and after at least 12 minutes to achieve equilibrium, the Modified Observer's Assessment of Alertness/Sedation (MOAA/S) scale will be used to perform sedation/loss of consciousness assessment. The loss of verbal response will be defined according to the MOAA/S score value < 2 (0-1) and the presence of verbal response at values ≥ 2 (2-5).

The expected duration of each subject's participation is approximately 6 hours for both enrollment and procedure. The Enrollment Visit should take approximately 2 hours to complete, and the Study Visit will be approximately 4 hours per subject. Each subject will be contacted by phone within 48 hours of the study participation. Subjects will be continuously monitored throughout the study.

This document provides a detailed description of the statistical methods and procedures to be implemented during the analysis of the study.

This statistical analysis plan (SAP) is based on Version 5.0 of the Protocol dated on 06-JAN-2022.

4 Study Objectives

4.1 Primary Objective:

To determine BIS_{50} (BIS™ value at which 50% of subjects will be unresponsive at a given drug concentration).

4.2 Secondary Objective:

To determine BIS_{95} (BIS™ value at which 95% of subjects will be unresponsive at a given drug concentration).

To determine Prediction Probability score (P_K) for correctly predicting if the subject was responsive or unresponsive.

5 Investigation Plan

This is a multi-center, prospective randomized, two-stage design study to collect data to evaluate the relationship between BIS™ and inhaled anesthetics with opioids. Subjects will be enrolled in the study once all eligibility requirements for the study have been met. Subjects who give informed consent for the protocol in order to undergo screening for eligibility are not considered enrolled and should not be enrolled until the screening is completed. All subjects are healthy volunteers.

After the subjects are enrolled, they will be sequentially assigned to one of the study groups of anesthetics regimens at each clinical site. The BIS™ bilateral sensor will be placed on the subject's forehead to study the effects of these drugs on the brain. Each volunteer will be given a dose range of sequences of the target drug to achieve targeted drug concentration. Five groups of anesthetics regimens will be studied as follows:

- Sevoflurane alone
- Sevoflurane with Remifentanil
- Sevoflurane with Fentanyl
- Desflurane alone
- Isoflurane alone

At each steady-state step, and after at least 12 minutes to achieve equilibrium, the Modified Observer's Assessment of Alertness/Sedation (MOAA/S) scale will be used to perform sedation/loss of consciousness assessment. An anesthesiologist authorized to administer sedation drugs will be responsible for administering any procedural drugs for the sedation and monitoring of subject safety and physical state. All other personnel involved with the safety monitoring of subjects must be trained in and familiar with the management of recovery of sedated patients. The study site should have a setting that is fully equipped for the monitoring and support of the respiratory and cardiovascular function. Subjects will be continuously monitored throughout the study.

A two-stage design is used in this study. For each regimen group, 10 subjects will be enrolled in stage 1, and upon evaluation, up to 20 additional subjects will be enrolled in stage 2. With 5 regimen groups, a maximum of 150 subjects will be enrolled. An interim review will take place at the end of stage 1 (50 subjects, 10 subjects per regimen group). Both the primary and secondary objectives will be evaluated during the interim review, and one or more groups may be dropped upon interim results for safety concern (unacceptable safety events per clinical judgment).

Based on the number of sites and enrollment rates, study duration is expected to be up to approximately 9 months. The expected duration of each subject's participation is approximately 6 hours for both

enrollment and procedure. The Enrollment Visit should take approximately 2 hours to complete, and the Study Visit will be approximately 4 hours per subject. Each subject will be contacted by phone within 48 hours of the study participation. The Schedule of Events, **Table 1**, summarizes the intervals and data collection procedures.

Table 1: Schedule of Events

Study Tasks	Pre-screening No visit	Enrollment Visit 1	Execution Visit 2		Follow up Phone call
			Prior to Procedure	Procedure	
Eligibility Assessments					
Online or phone call pre-screening survey	x				
Informed Consent [1]		x			
Demographics		x			
Medical History		x			
Physical Exam		x			
Pulmonary function test		x			
Single 12-lead ECG		x			
Urine sample for the presence of cotinine		x	x		
Urine pregnancy test (Female)		x	x		
Complete blood count			x		
Inclusion/exclusion assessment		x	x		
Vital monitoring		x	x	x	
Concomitant Medication		x	x		
Urine drug screen and alcohol breathalyzer			x		
Safety Monitoring				x	
Sensors application				x	
The procedure with drug administration				x	
MOAA/S assessment				x	
Safety Assessments and Compensation					
Adverse Event Assessment [2]		x	x	x	x
Device Deficiency [3]				x	
Participant stipend [4]		x		x	

Formatted: Not Highlight

6 Determination of Sample Size

A two-stage adaptive design is used in this study. For each regimen group, 10 subjects will be enrolled in stage 1, and upon evaluation, up to 20 additional subjects will be enrolled in stage 2. With 5 regimen groups, a maximum of 150 subjects will be expected.

Based on the dose-response relationship (logistic regression) with an allowable error of ±15% for BIS_{50} and coefficient of variation (CV) of 25% at the alpha level of 0.05, the sample size will provide sufficient power (>80%) to evaluate the performance of anesthetic agents. An interim review will take place at the end of stage 1 (50 subjects, 10 subjects per regimen group).

In terms of hypothesis testing, let BIS_{50} be unit 1. With 25% coefficient of variation, the standard deviation (σ) is calculated as $BIS_{50} \times CV = 0.25$. To detect 15% difference (allowable error), the target hypothesis can be written as:

$$H_0: \mu = 1 \quad vs. \quad H_a: \mu = 1.15.$$

With two-sided alpha of 0.05 and power of 80%, the required sample size is 24 (PASS – Tests for One Mean). Of note that the statistical approach has been published in the literature (see reference: Kopman 2011).

By further considering the interim review and adaptiveness of design, we expect at least 80% data information will be retained from Stage 1, leading to a minimum *effective sample size* of 10*0.8+20 = 28 for the final analysis. Adding 2 more subjects to have 30 subjects in each regimen will ensure adequate data to evaluate the dose-response relationship.

7 Statistical Methods

7.1 Study Subjects

7.1.1 Disposition of Subjects

Subject disposition will be summarized with frequency tables.

7.1.2 Clinical Investigation Plan (CIP) Deviations

A study deviation is defined as an event when the investigator or site personnel did not conduct the study according to the protocol or the clinical trial agreement.

Study deviations must be reported, regardless of whether medically justifiable, pre-approved by the study leader, or taken to protect the subject in an emergency.

7.1.3 Analysis Sets

Full Analysis Set (FAS): All randomized subjects will be included in the analysis. For those subjects who do not complete a follow-up call, the analysis will include all data up to the last collected point or follow-up call. Subjects who completed the procedure phase will not be replaced. If the subject discontinues prior to randomization or during the procedure phase, she/he will be replaced, but will count toward the total sample size if they were randomized.

7.2 General Methodology

All statistical analyses will be performed using Statistical Analysis System (SAS) for Windows (version 9.3 or higher, SAS Institute Inc. Cary, NC) or other widely accepted statistical or graphical software. In general, data for all study subjects will be presented. Individual data will be presented in subject listings.

Descriptive statistics will be used to present the data and to summarize study outcomes. Discrete variables will be presented using frequency distributions and cross tabulations. Continuous variables will be summarized by presenting the number of observations (N), mean, standard deviation, median, minimum, and maximum values.

For safety assessments, adverse events will be summarized using frequency counts and percentages. Any unexpected adverse events will be reported and discussed.

7.2.1 BIS™ Value Analysis

Study subjects will be sequentially assigned to one of the target groups of anesthetics regimens while having BIS™ VISTA Monitoring System, the 4 Channel BISx, and BIS™ bilateral sensor placed on their forehead during these regimens to study the effects of these drugs on the brain. Each subject will be given a dose range of sequences of the target drug to achieve targeted drug concentration.

The modified Observer's Assessment of Alertness/Sedation (MOAA/S) scale will be used to perform sedation/ loss of consciousness assessment. Any subject who responds to any verbal command with MOAA/S assessment score of 2, 3, 4, and 5 will be considered as responsive. Subjects with MOAA/S assessment score of 0 and 1 will be considered unresponsive. Tetanic Electrical Stimulation (TES) will be employed with an MOAA/S assessment score of ≤2 for the first two subjects in each group and <2 for all other subjects.

The BIS™ is a variable derived from the electroencephalograph (EEG) that has been reported to have the ability to measure the hypnotic component of the anesthetic state. It is a dimensionless number from 0 to 100, and decreasing values indicate more sedation and hypnosis.

The BIS™ index value collected from the left side of the brain will be used for primary analysis to reflect the typical operation of the device from this study. The BIS™ index value collected from the right side of the brain will be used for internal research purposes.

The logistic model (simplified Sigmoid E_{max} model) will be used to analyze the relationship between the BIS™ index and loss of responsiveness through the probability of response curves. The values of BIS™ at which 50% (BIS_{50}) and 95% (BIS_{95}) of subjects are unresponsive, and their 95% confidential intervals will be derived. The systematic variance between groups will be evaluated.

The logistic model is as follows:

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}}$$

where *P* is the probability of "unresponsive", *X* is the explanatory variable (BIS™ value or anesthetic agent concentration on the log scale), and *a* & *b* are model parameters.

For each regimen group, one can estimate the BIS™ value and the effect-site concentration at which 50% subjects are unresponsive, i.e., BIS_{50} and EC_{50} respectively. The corresponding 95% confidence intervals and standard errors for the estimates will be derived. The same evaluation will be performed for BIS_{95}, at which 95% subjects are unresponsive. The systematic variance between groups will be evaluated accordingly.

In addition, the Prediction Probability score (P_k) for correctly predicting if the subject was responsive or unresponsive will be estimated and reported (See details in reference: Smith 1996).

The interim review will be performed at the end of stage 1 when all subjects (50 subjects, 10 subjects in each regimen group) have completed the evaluation. BIS™ values will be estimated using the logistic model (Hill equation). One or more groups may be dropped upon interim results for safety concern (unacceptable safety events per clinical judgment). The pre-specific precision for BI™S value estimation is ±15%. For the final analysis, data from stage 1 and stage 2 will be combined by further taking to account poolability across sites (see details in the next section) and the adaptive nature of the 2-stage design (see details in the Appendix).

7.3 Center Pooling

This is a multi-center clinical study, with standardization of subject enrollment, data entry and Adverse Event reporting. All investigational sites will follow the requirements of a common protocol, data collection procedures and forms. Upon justifications, data from all sites will be pooled for analysis.

Analyses to justify pooling will include the following:

Data Pooling Across Sites

Poolability across sites will be assessed using logistic regression model on the primary outcome. The model will include "Responsive: Yes/No" as response variable, and BIS, Site, BIS*Site interaction as exploratory variables. Sites with fewer than five subjects will be combined based on the order from smallest to largest. A P-value of 0.15 or less is considered statistically significant for the evaluation (FDA preferred).

If the sites are found to be significantly heterogeneous with respect to the primary outcome, additional analyses will be conducted to assess differences between sites in baseline and procedural variables that might explain differences in the primary outcome.

Data Pooling of Stage 1 and Stage 2

The study consists of two stages: first 10 subjects (stage 1), and then followed by 20 additional subjects (stage 2). Although no differences will be expected across the two stages, statistical testing will be performed to ensure the 2-stage data can be pooled for analysis. The testing will be based on the logistic regression model with binary "Responsive: Yes/No" as response variable, BIS, Stage indicator (X), BIS*X interaction as exploratory variables.

In case significant differences are present across the two stages, simply pooling data may not be valid. In this case, an alternative strategy will be used for statistical evaluation. Stage 1 data will be utilized as prior, which will be integrated into stage 2 data using the Bayesian Data Integration. More specifically, power prior and discounting function will be used to accommodate the difference across the two stages. Technical details can be found in the Appendix.

7.4 Handling of Missing, Unused, and Spurious Data and Dropouts

The primary analysis will be based on all evaluable data. Handling of dropouts and missing data will depend on their frequency and the nature of the outcome measure.

For those subjects who do not complete a follow-up call, the analysis will include all data up to the last collected point or follow-up call. Subjects who completed the procedure phase will not be replaced. If the subject discontinues prior to randomization or during the procedure phase, she/he will be replaced, but will count toward the total sample size if they were randomized.

7.5 Adjustments for Multiple Comparisons

No multiplicity adjustments will be considered in this study. As aforementioned, poolability of data from stage 1 and 2 will tested and justified. Dose-response relationship will be evaluated for each of the study regimens. Between-group comparisons will be summarized with point estimates and 95% confidence intervals. P-values will be provided for reference purpose only.

7.6 Demographic and Other Baseline Characteristics

Subject demographics, medical history will be summarized using descriptive statistics for continuous variables (N, mean, standard deviation, number of observations, minimum and maximum) and frequency tables for discrete variables.

Baseline demographic data will be summarized and reported in a table entitled Summary of Subject Demographics. This table summarizes the subject population with respect to age at entry into the study, gender. Age will be reported in years, and will be summarized using descriptive statistics: N, arithmetic mean, standard deviation, median, and range (i.e., minimum and maximum values). Subjects with missing data that cannot be resolved prior to database lock will not be included in the tabulation and excluded from the summary statistics; Gender will be summarized using counts and percentages. In addition to the reported values, unknown or unreported values will also be reported. The supportive data for the demographics table will be presented in a listing entitled Subject Demographics.

7.7 Interim Review

The interim review will be performed at the end of stage 1 when all subjects have completed the evaluation. BIS™ values will be estimated using logistic regression (Hill equation) based on primary outcome "Responsive: Yes/No". One or more regimens may be dropped upon interim results for safety concern (unacceptable safety events per clinical judgment).

7.8 Evaluation of Objectives

Th objective of this study is to investigate the relationship between BIS™ and anesthetic agents across a wide range of hypnotic states and to provide evidence to support BIS™ performance in use with these anesthetic agents.

Primary Objective:

To determine BIS_{50} (BIS™ value at which 50% of subjects will be unresponsive at given drug concentrations).

Secondary Objective:

To determine BIS_{95} (BIS™ value at which 95% of subjects will be unresponsive at given drug concentrations).

The logistic model (simplified Emax model) will be used to estimate the BIS™ values, along with corresponding 95% confidence intervals and standard errors.

Prediction Probability score (P_K) will be calculated following the details in the Smith's paper (Smith 1996). All technical details were provided in the Appendix in the paper. SAS procedure PROC LOGSTIC will be used to perform the calculation.

Differences in subject characteristics will be analyzed using t-test (continuous variables including age, weight and height) or Chi-square test (categorical variables including gender). The correlation coefficient for the relationship between sedation score and BIS™ value will be calculated for each group using a linear regression model. Additionally, BIS™ values and haemodynamic variables will be analyzed within the groups using analysis of variance (ANOVA) for repeated measures.

7.9 Safety Evaluation

Safety evaluation will be based on all enrolled subjects in this study. Adverse events will be summarized using frequency counts and percentages. Descriptive statistics will be provided by event type, severity, and relationship to study procedures and devices.

Unanticipated device effect (UADE), Serious Adverse Device Effect (SADE) and Unanticipated Serious Adverse Device Effect (USADE) will also be summarized by frequency table. Individual listings of adverse

events, including event type, start date, duration, severity, and device-relatedness, will be provided as appropriate.

7.10 Health Outcomes Analyses

This section is not applicable to this study.

7.11 Changes to Planned Analysis

This section is not applicable to this study.

8 Validation Requirements

Output will be validated by level I or II validation.

Level I: The peer reviewer independently programs output and then compares the output with that generated by the original Statistical Programmer.

Level II: The peer reviewer reviews the code; where appropriate, performs manual calculations or simple programming checks to verify the output.

9 Statistical Appendices

I. Bayesian Data Integration with Power Prior

A. Power Prior

Consider statistical data integration of two sources: historical data and current data. Simply pooling the data may not yield valid scientific inferences due to potential variations across data sources. The variations may or may not be observed and ignoring the variations will likely lead to biased estimates. For valid inferences, appropriate statistical methods and justifications will be needed. The Bayesian power prior is a statistical method often used to achieve the data integration. The detailed formulation is as follows.

Let the data for the current study be denoted by D and the corresponding likelihood function be denoted as $L(\theta|D)$, where θ is a vector of parameters of interest. Suppose we have the historical data D_0 from a previous study. Let $L(\theta|D_0)$ denote the likelihood function for the historical data D_0. Here, $L(\theta|D)$ and $L(\theta|D_0)$ are general likelihood functions for the corresponding statistical model to be used for inferences.

The basic formulation of the power prior can be written as:

$$\pi(\theta|D_0, \alpha_0) \propto L(\theta|D_0)^{\alpha_0}\pi_0(\theta),$$

where $0 \le \alpha_0 \le 1$ is a *scalar* parameter and $\pi_0(\theta)$ is the *initial prior* for the parameter θ before the historical data D_0 is observed. In many applications, $\pi_0(\theta)$ is taken to be an improper prior. With the power prior, the corresponding posterior distribution of θ is given by

$$\pi(\theta|D, D_0, \alpha_0) \propto L(\theta|D)\, L(\theta|D_0)^{\alpha_0}\pi_0(\theta).$$

From the above formula, we see that α_0 weights the historical data relative to the likelihood of the current data, and thus the parameter α_0 controls the influence of the historical data on $L(\theta|D)$. The parameter α_0 can be interpreted as a precision parameter for the historical data. Since D_0 is historical data, it is unnatural in practical applications to weight the historical data more than the current data; thus it is scientifically sound to restrict the range of α_0 to be between 0 and 1, that is, $0 \le \alpha_0 \le 1$. Setting $\alpha_0 = 1$ corresponds to the update of $\pi_0(\theta)$ using the Bayes theorem. That is, with $\alpha_0 = 1$ corresponds to the posterior distribution of θ based on the historical data D_0. When $\alpha_0 = 0$, the prior does not depend on the historical data; in this case, $\pi(\theta|D_0, \alpha_0 = 0) = \pi_0(\theta)$. Thus, $\alpha_0 = 0$ is equivalent to a prior specification without incorporation of historical data. Therefore, the approach can be viewed as a generalization of the usual Bayesian update of $\pi_0(\theta)$. The parameter α_0 allows to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the previous and current studies.

Since the power prior is essentially a likelihood function raised to a power, it shares all the properties that likelihood functions have, and therefore has several advantages over other non-likelihood-based priors. Due to its validity and ease of interpretation, the power prior approach has been widely adopted in practical applications and is commonly seen in medical device regulatory submissions.

B. Primary Efficacy Endpoint

In this study, the primary efficacy endpoint is the BIS_{50}, defined as the BIS™ value at which 50% of subjects are unresponsive at given drug concentrations. The BIS™ is a variable derived from the electroencephalograph (EEG) that has been reported to have the ability to measure the hypnotic component of the anesthetic state. It is a dimensionless number from 0 to 100, and decreasing values indicate more sedation and hypnosis. The MOAA/S scales and BIS™ value will be assessed and recorded when the subject is awake and at different drug concentrations.

The dose-response relationship will be estimated using the logistic model. The model is written as follows:

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}}$$

where *P* is the probability of "unresponsive", *X* is the explanatory variable (BIS™ value or anesthetic agent concentration on the log scale), and *a* & *b* are corresponding model parameters.

In the context of the dose-response model, let θ represents the target BIS™ value on the log scale. Let $D = \{y_1\}$ and $D_0 = \{y_0\}$ represent the current data and prior data respectively. The parameter estimation will be carried out based on the posterior distribution (upon normalization):

$$\pi(\theta|D, D_0, \widehat{\alpha_0}) \propto L(\theta|D)\, L(\theta|D_0)^{\widehat{\alpha_0}} \pi_0(\theta)$$

where the notation $\widehat{\alpha_0}$ is used to denote that the estimate of α_0 based on the observed data. Of note that the estimate depends on the current data (D) and prior data (D_0). In conjunction with a pre-specified decision rule controlling the prior data weight, the estimate of $\widehat{\alpha_0}$ represents a measure of similarity between the current data and prior data. Alternatively, in the absence of $\widehat{\alpha_0}$, i.e., full weight would be given to the prior data.

C. Statistical Analysis

The Bayesian power prior with discounting function will be used to estimate the target parameter and determine $\widehat{\alpha_0}$, the strength of the prior data used to support the current data. Of note $\widehat{\alpha_0}$ ranges from 0 to 1, where 1 means that 100% of the prior data is used and 0 means that no prior data is used. Before beginning the current study, an initial value is chosen for $\widehat{\alpha_0}$, and let's call this value α_{max}. This α_{max} value is the maximum strength that the prior data can receive. For this study, we intend to use the same protocol and enrollment criteria for the prior and current studies, and therefore believe that a value of $\alpha_{max} = 1$ is appropriate.

The study data will be analyzed using the power prior method along with discounting function. The method will discount α_{max} to an appropriate value $\widehat{\alpha_0}$ where $\widehat{\alpha_0} \leq \alpha_{max}$. This discounting scale is based on the discounting function which will be discussed in detail in the next section.

Under the adaptive procedure, if the current study data diverges from the prior data, the discounting function will discount the strength of the prior data. Alternatively, if the prior and current study data agree, there will be no or minimal penalty from the discounting function, thus a large amount of data information would be integrated into current data for statistical inferences.

The primary analysis set consists of all evaluable subjects and will be used for the primary analysis. The per-protocol set will be used for supportive analysis.

D. Discounting Function Estimation Method

The method of power prior discounting function is comprised of four steps: **Compare, Discount, Combine, and Estimate.** Details in each step are presented below.

Compare:

We start by stochastically comparing current data vs prior data as follows.

For each study group, we separately fit the model to the combined prior and current data:

$$H(p_i) = \beta_0 + \beta_1 I(i \in \text{prior}) + \beta_2 d_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2),$$

where H is the inverse logit function, p_i is the probability of target response, d_i is the dose concentration (log scale), and the indication function $I(i \in \text{prior}) = 1$ if the subject is from the prior data, and 0 otherwise. With a flat prior on each parameter, we estimate the posterior probability that $\beta_1 > 0$ by first computing the parameter, using Monte Carlo sampling

$$p^* = P[\beta_1 > 0 \mid D, D_0].$$

Having calculated the parameter for each group, we use the transformation:

$$p = \begin{cases} 2p^*, & p^* \leq 0.5 \\ 2(1-p^*), & p^* > 0.5 \end{cases}$$

Now, under this transformation, if p is close to 0, there is a high probability that the current data and prior data come from different populations and discounting should be applied to reduce the influence of the prior. On the other hand, if p is close to 1, there is a high probability that the current data and prior data are from similar populations and minimal discounting should be applied.

Discount:

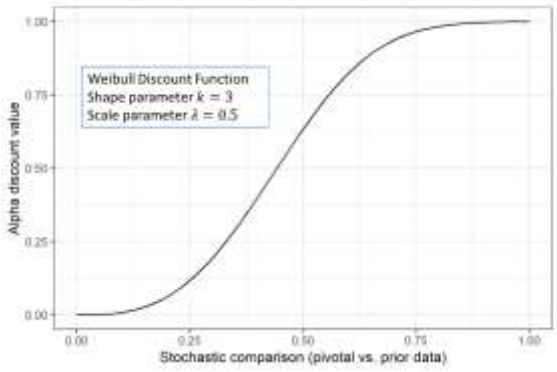
Let discount α_{max} based on the value of p from the **Compare** step and the discounting function $F(p)$,

$$\widehat{\alpha_0} = \alpha_{max} F(p),$$

where $F(p)$ is a function between 0 and 1. A two-sided Weibull function will be utilized as follows:

$$F(p) = 1 - e^{-(p/\lambda)^k}.$$

For this study, we will be using a shape parameter of $k = 3$ and a scale parameter of $\lambda = 0.5$ (illustrated below).



Combine:

Using the power prior method and discounting factor $\widehat{\alpha_0}$, one can combine the prior and current data together using the Bayesian method to construct the posterior distribution for the parameters of interest.

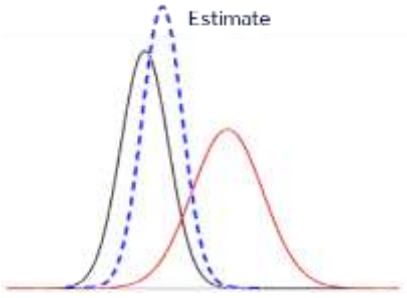
Estimate:

The posterior distribution from the combined prior and current data is used to make statistical inferences:

$$\hat{\pi}(\theta|D, D_0, \widehat{\alpha_0}) \propto L(\theta|D)\, L(\theta|D_0)^{\widehat{\alpha_0}} \pi_0(\theta)$$

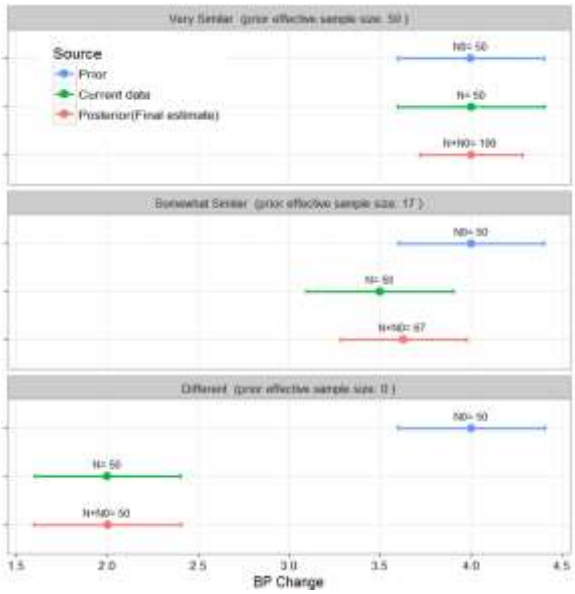
where the notation $\widehat{\alpha_0}$ is used to denote that the estimate of α_0 depends on the current data, prior data, and the Weibull shape and scale parameters. In conjunction with a prespecified decision rule controlling the prior data weight, the estimate of $\widehat{\alpha_0}$ represents a measure of similarity between current and prior data. Alternatively, in the absence of $\widehat{\alpha_0}$, full weight would be given to the prior data.

For illustrative purposes, the blue dashed line in the figure below is an example of the integrated estimate from current data (black) and prior data (red).



E. Illustration of Discounting Function Scenarios

The Figure below shows how the discounting function operates with hypothetical data sets



The panels in this figure can be interpreted as follows:

- Top panel: The current data is very similar to the prior. The discounting function allows for full strength of the prior. The posterior (final estimate) is a balance between the prior and current study.
- Middle panel: The current data is similar to the prior. The discounting function penalty is moderate, resulting in a prior effective sample size of 17 out of a max of 50. Because the agreement is reasonable, the posterior (final estimate) is similar to both the prior and current study.
- Bottom panel: The current data shows lower performance than the prior. The discounting function produces a substantial penalty resulting in no weight to the prior. The posterior (final estimate) is essentially the same as the current study.

F. Simulation of Operating Characteristics

Extensive simulations were performed to assess operating characteristics for the statistical analysis. We used 10k trial simulations to estimate the power and 10k simulations to estimate the operating characteristics. The following

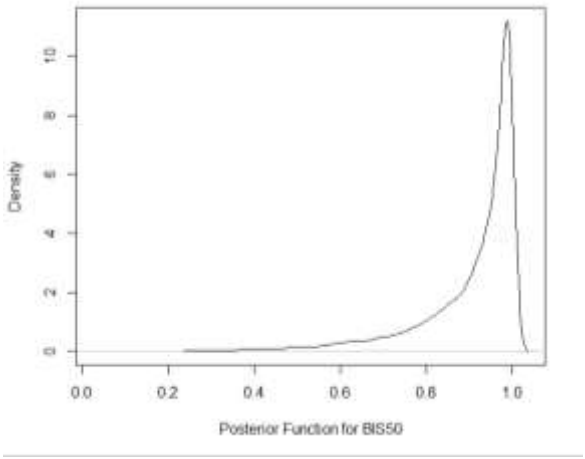
table summarizes the assumptions that were made for the primary endpoint simulations, and the figure summarizes the operating characteristics for the discounting parameter.

Table: Simulation Assumptions for Primary Endpoint Analysis

Assumptions	Values
Prior Subjects	10
Current Subjects	20
BIS_{50} Value (Mean)	60
Prior Data BIS_{50} Value (CV%)	25%
Prior Data BIS_{50} Value (SD)	1.25
Current Data BIS_{50} Value (CV%)	25%
BIS_{50} Value Ratio (Current vs Prior)	1.15
Weibull Discounting Function Parameters	shape: $k = 3$, scale: $\lambda = 0.5$

Simulation results are summarized in the following figure. As expected, there is good distributional overlapping between the prior and current data. The amount of prior data integrated to the current data is quantified by the discounting parameter, which has a mean value of 90.4% with corresponding 95% credible interval (53.1%, 99.9%). Overall, the variations between prior and current data have been taken into account and valid statistical inferences can be made.

Figure: Posterior distribution of the discounting parameter (10k simulations)



G. References

[1] http://mdic.org/computer-modeling/virtual-patients/

[2] Haddad A, Himes A, Thompson L, Irony T, Nair R. Incorporation of stochastic engineering models as prior information in Bayesian medical device trials. *J Biopharma Stat*. 2017. DOI: 10.1080/10543406.2017.1300907.

[3] Ibrahim JG, Chen MH. Power prior distributions for regression models. *StatisticalScience.* 2000;15(1):46-60.

[4] Ibrahim JG, Chen MH, Gwon Y, Chen F. The Power Prior: Theory and Applications. *Stat Med*. 2015; 34(28): 3724–3749.

[5] Gelman A, Carlin J, Stern H, D. Rubin D. Bayesian data analysis, Boca Raton: Chapman & Hall/ CRC, 2004.

[6] Berry SM, Carlin BP, Lee, JJ, Muller P. Bayesian adaptive methods for clinical trials. CRC press, 2010.

[7] Kopman AF, Lien CA, Naguib M. Neuromuscular dose-response studies: determining sample size. *British Journal of Anaesthesia*. 2011;106(2):194-198.

[8] Smith WD, Dutton RC, Smith NT. Measuring the Performance of Anesthetic Depth Indicators. *Anesthesiology.* 1996; 84:38-51.