A Multicenter, Partially-Masked, Randomized, Controlled Study of Medical Therapy vs. Medical Therapy plus Optic Nerve Sheath Fenestration vs. Medical Therapy plus Stereotactic Ventriculoperitoneal Cerebrospinal Fluid Shunting in Subjects with Idiopathic Intracranial Hypertension and Moderate to Severe Visual Loss

Surgical Idiopathic Intracranial Hypertension Treatment Trial ("SIGHT")

STATISTICAL ANALYSIS PLAN

Version 1.0
Protocol Version 3.2

January 30, 2019

Revision History

The following table outlines changes made to the Statistical Analysis Plan.

Version Number	Protocol Version	Author	Approver	Effective Date	Study Stage	Revision Description
1.0	3.2	Nicole Foster	Mike McDermott	03-04-2019	No analyses done	Original Version

Author: _____

Approver: _____

Study Lead: _____

List of Abbreviations

Abbreviation	Definition
CBC	Complete Blood Count
CSF	Cerebrospinal Fluid
D	Diopter
dB	Decibel
DSMC	Data and Safety Monitoring Committee
ETDRS	Early Treatment Diabetic Retinopathy Study
GCL	Ganglion Cell Layer
GEE	Generalized Estimating Equations
HIT-6	Headache Impact Test
IIH	Idiopathic Intracranial Hypertension
ITT	Intention to Treat
MAR	Missing at Random
MCS	Mental Component Summary
MedDRA	Medical Dictionary for Regulatory Activities
mm Hg	Millimeters of Mercury
NEI	National Eye Institute
ONSF	Optic Nerve Sheath Fenestration
PCS	Physical Component Summary
PMD	Perimetric Mean Deviation
PRC	Photography Reading Center
RNFL	Retinal Nerve Fiber Layer
SF-36v2	Short Form Health Survey
VFQ-25	Visual Function Questionnaire
VFRC	Visual Field Reading Center
VPS	Sterotactic Ventriculo-Peritoneal CSF Shunting

1.0 Study Overview

The study is a randomized trial of adults (18-<64 years old) with idiopathic intracranial hypertension and moderate to severe visual loss without substantial recent treatment who are randomly assigned to (1) medical therapy, (2) medical therapy plus Optic Nerve Sheath Fenestration (ONSF), or (3) medical therapy plus Sterotactic Ventriculo-Peritoneal CSF Shunting (VPS).

The primary study objective is to determine whether the efficacy of VPS with medical therapy is superior to medical therapy alone or ONSF with medical therapy in reducing or reversing visual loss in subjects with idiopathic intracranial hypertension and moderate to severe visual loss. A secondary objective is to compare the treatment groups with respect to the time from randomization to treatment failure over up to 3 years.

The study protocol consists of:

- Randomized Trial
  - Screening visit
  - Baseline visit
  - Randomization
  - Safety Visits (timing dependent on treatment group)
  - Visit 1 (4 weeks from randomization)
  - Visit 2 (8 weeks from randomization)
  - Phone call 1 (12 weeks from randomization)
  - Visit 3 (16 weeks from randomization)
  - Phone call 2 (20 weeks from randomization)
  - Visit 4 (Primary Outcome; 26 weeks from randomization)
- Treatment Failure Identification Phase
  - Phone call (39 weeks from randomization)
  - Visit 5 (52 weeks from randomization)
  - Visit 6 (104 weeks from randomization)
  - Visit 7 (Final visit; 156 weeks from randomization).

Unscheduled visits may occur if early termination of treatment is necessary (testing will be the same as the primary outcome visit).

This document describes the analyses that will be performed.

2.0 Sample Size

In this clinical trial, 180 subjects with newly diagnosed IIH and moderate to severe visual loss (-27 dB ≤ PMD ≤ -6 dB) will be randomly assigned to receive either medical therapy, ONSF + medical therapy, or VPS + medical therapy (60 per group). This sample size should provide high power (~90%) to detect a difference in mean change of visual field between groups when the

true differences between any 2 of 3 groups is 4.5 dB with a two-tailed Bonferroni-adjusted significance level of 1.7%, allowing for an anticipated 10% drop-out. The rationale for the choice of a clinically significant treatment group difference of 4.5 dB is explained in the last paragraph of section 2.0.

The primary outcome variable in this trial will be the change from baseline to the first of Month 6 or time of treatment failure in PMD in an eligible eye. The sample size considerations initially focus on data from the best eligible eye, since most subjects are expected to contribute only one eye to the primary analysis, but addition of the other eligible eye is also considered below. In 2012, the IIH Study Group performed a retrospective chart review of consecutive newly diagnosed patients that met the modified Dandy criteria for IIH at 30 of the 41 participating sites. Data on PMD were available from 91 patients at two time points, before and after intervention, with the median follow-up time being 6.0 months (interquartile range 4.1 to 7.0 months). Patients received either medical treatment (n = 43), ONSF (n = 24), or VPS (n = 24). The mean (± standard deviation) changes in PMD (in dB) in the best eligible eye over the follow-up period were 5.5 ± 6.9 in the medical group, 2.0 ± 8.9 in the ONSF group, and 8.2 ± 8.0 in the VPS group; overall these values were 5.3 ± 8.0. The distribution of these changes was slightly more peaked than would be expected for a normal distribution, as illustrated in Figure 1a and 1b.

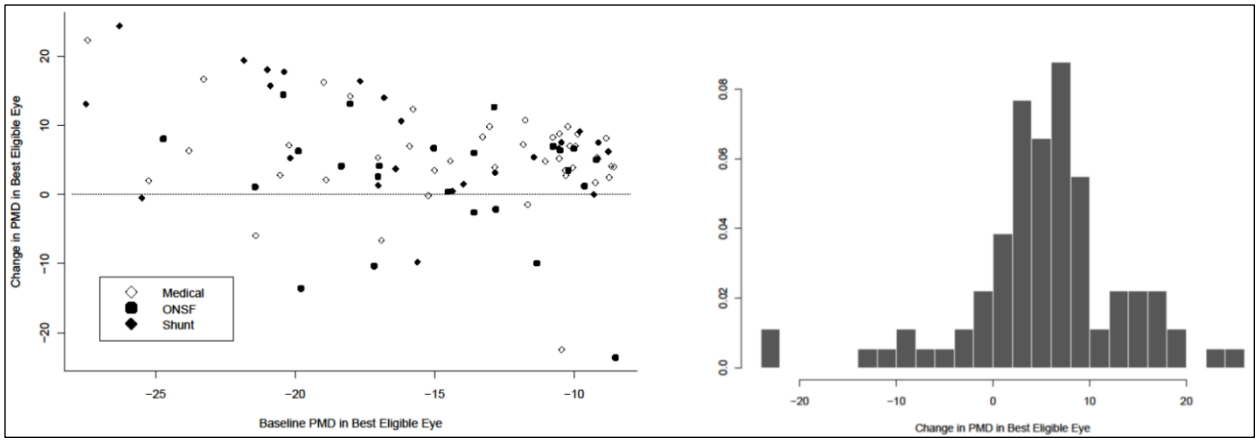

Figure 1a and 1b. Distribution of Changes in PMD in the Best Eligible Eye

The figure on the left (a) shows the distribution of changes in PMD in the best eligible eye according to baseline PMD and intervention. Most of the patients demonstrated improvement in PMD. The figure on the right (b) shows a histogram of the changes in PMD in the best eligible eye regardless of intervention. The distribution appears to be characterized by a high concentration of changes between 0 and 10 PMD and is more peaked than normal.

Given the cross-sectional nature of the preliminary data, sample size determination based on the use of GEE for the primary analysis is very similar to that based on the use of an analysis of covariance model. If an analysis of covariance model is fit to the preliminary data, with change in PMD as the outcome variable and treatment group and baseline PMD as the independent variables, the standard deviation of the residuals is 6.3 dB. The differences between treatment groups in adjusted mean response are quite small in this analysis: -0.09 dB difference between

the ONSF and medical groups, and 1.13 dB difference between the VPS and medical groups. Also, a GEE analysis produces identical adjusted group means and slightly different estimated standard errors than those from the analysis of covariance model.

It may be noted that the standard deviation of the residuals in the above analysis of covariance model is substantially less than the standard deviations in the individual treatment groups, reflecting adjustment for the baseline value of PMD in the analysis of covariance model and the fairly strong correlation between the baseline and final PMD values ($r = 0.52$). When considering only subjects who were followed for at least 5.5 months ($n = 56$), the correlation between the baseline and final PMD values was actually somewhat higher ($r = 0.62$) and the residual standard deviation in the analysis of covariance model was smaller (6.0 dB).

Assuming a standard deviation of 6.5 dB, and a Bonferroni-adjusted two-tailed significance level of 1.7%, in order to detect a group difference assuming a true difference of 4.5 dB with 88% power, a sample size of 54 subjects per group is required. The sample size will be inflated to 60 subjects per group (180 total) to accommodate an anticipated 10% rate of subject withdrawal/dropout.

The assumed standard deviation of 6.5 dB was selected based on the calculated standard deviation of the residuals from the above analysis of covariance model (6.3 dB) under the following assumptions: 1) the distribution of baseline PMD in the study cohort will be similar to that observed in the pilot study, and 2) the higher standard deviation observed in the pilot study surgical groups can be attributed to the increased probability of those with worse PMD at baseline to receive surgery (participants with worse PMD at baseline had higher variability of change in PMD)[a]. The power remains above 80% even if the assumed standard deviation is as large as 7.1 dB.

The inclusion of the worst eligible eye is expected to increase power, but the increase is expected to be small (approximately 3%) because the preliminary data from the chart review indicate that only ~40% of subjects will contribute a second eligible eye to the analysis and the correlation between the outcomes in the two eyes is quite high (0.82 in our sample). This assumes that the group difference will be comparable in the best and worst eligible eyes, as we anticipate.

The chosen effect size of 4.5 dB is based on the following rationale. Visual field defects are similar in IIH and glaucoma and data on the relationship between vision-specific quality of life (as measured by the NEI-VFQ-25) and PMD from the better seeing eye are available from 213 subjects with glaucoma in the Los Angeles Latino Eye Study. A regression analysis yielded the finding that a change of 1 dB in PMD corresponded to an approximately one-unit score change on the NEI-VFQ-25 composite score. In another study, Suner et al. used data from two clinical trials in neovascular age-related macular degeneration and anchor-based methods to estimate the change in NEI-VFQ-25 composite score that corresponds to a change of ≥ 15 letters (~ 3 lines) in visual acuity, a value generally accepted as clinically significant. They concluded that a change of 4-6 points on the NEI-VFQ-25 composite score should be considered clinically significant. Taken together, the results of these two investigations suggest that a change in PMD of 4.5 dB

corresponds to a change in NEI-VFQ-25 score that would be considered to be of minimal clinical significance.

[a]The DSMC calculated an estimated standard deviation of change in PMD of 6.9 dB based on the assumption that the standard deviations in the pilot data were larger for participants who received surgery, and that the increased variability was due to the surgical intervention itself. Since the surgical groups were under-represented in the pilot data, the estimated variances in these groups were appropriately up-weighted in the calculation of the pooled standard deviation, resulting in the estimate of 6.9 dB. The assumptions underlying the original calculation (6.5 dB) and the DSMC calculation (6.9 dB) are both plausible. The DSMC agreed to move forward with the 6.5 dB standard deviation.

3.0 Analysis

3.1 Intention-to-Treat Principle

The primary statistical analyses for this trial will be performed according to the intention-to-treat principle and will include all randomized subjects and eligible eyes. An eye is defined as eligible if the following hold at the baseline visit:

- Inclusion criteria
  - Visual field loss meeting the following based on two full threshold 24-2 size V tests reviewed by the Visual Field Reading Center (VFRC)
    - PMD from -6 dB to -27 dB
    - Reproducible visual loss present on automated perimetry including no more than 15% false positives
  - Visual acuity better than 20/200 (39 or more letters correct)
- Exclusion criteria
  - Intraocular pressure currently >28 mm Hg or >30 mm Hg at any time in the past
  - Refractive error of more than -6.00 or more than +6.00 sphere or more than 3.00 cylinder with the following exceptions:
    - Eyes with more than 6.00D of myopia but less than 8.00D of myopia are eligible if:
      - There are no abnormalities on ophthalmoscopy or fundus photos related to myopia that are associated with visual loss (such as staphyloma, retinal thinning in the posterior pole, or more than mild optic disc tilt), and
      - The individual will wear a contact lens for all perimetry examinations with the appropriate correction
    - Eyes with more than 6.00D of hyperopia but less than 8.00D of hyperopia are eligible if:
      - There is an unambiguous characteristic halo of peripapillary edema as opposed to features of a small crowded disc or other hyperopic change related to visual loss determined by the Site Investigator or the PRC Director (or his designate), and
      - The individual will wear a contact lens for all perimetry examinations with the appropriate correction (which can be corrected for perimetry or the patient's own contact lens with over correction by lens at perimeter)
  - Other disorders causing visual loss except for refractive error and amblyopia, including cells in the vitreous or iritis
  - Large optic disc drusen on exam or in previous history (small drusen of the disc can occur with longstanding papilledema and are allowed if not so numerous that investigator determines they are contributing to vision loss).

Every effort will be made to retain subjects in this study, to promote adherence to the study protocol, and to collect all data at every visit. If a subject cannot tolerate study medication or refuses to receive the study intervention, we will continue to follow and evaluate that subject if he/she is willing. If a subject drops out, attempts will be made to bring the subject back for a final evaluation. Compliance with trial procedures, drop-outs/drop-ins, and reasons for subject withdrawal will be carefully tracked throughout the study.

3.2 Analysis of the Primary Outcome Variable

3.2.1 Definition of Primary Outcome

The primary outcome variable will be the change from baseline to first of Month 6 or time of treatment failure in PMD in an eligible eye, with data from all eligible eyes included in the primary analysis (refer to the table below for a description of eyes included in the primary analysis).

Eyes at baseline	Eyes included in primary analysis
No eligible eyes→ subject not eligible	None (subject not eligible)
One eligible eye and non-eligible eye too good	One eye that was eligible at baseline, regardless of what happened to either eye during study
One eligible eye and non-eligible eye too bad	One eye that was eligible at baseline, regardless of what happened to either eye during study
Two eligible eyes	Both eyes that were eligible at baseline, regardless of what happened to either eye during study

Treatment of subjects who have reached criteria for treatment failure will be at the discretion of the Site Investigator, and this treatment may yield a different outcome than randomized treatment. For this reason, the primary outcome variable for subjects who reach criteria for treatment failure prior to Month 6 will be the PMD measured at the time of treatment failure for purposes of the primary analysis. It is anticipated that no more than 15% of subjects will reach criteria for treatment failure prior to Month 6.

PMD (baseline or outcome) will be the average PMD from the two reliable visual field examinations (i.e. meeting Visual Field Reading Center criteria) performed at the visit; if only one reliable result is obtained on the visual field examination, the one reliable result will be analyzed; if a reliable result is not obtained on the visual field examinations, the PMD will be considered missing.

Outcome PMD (excluding PMD at time of treatment failure) from visual field testing performed out of the window of 168-242 days from randomization will be considered missing for the primary analysis. Note this means a subject can be up to two weeks early and up to a full 60 days late, this latter choice was made with clinician input that they would rather have a measurement than impute.

3.2.2 Primary Statistical Model

The primary statistical analysis will involve fitting an analysis of covariance model using generalized estimating equations (GEE) with treatment group as the factor of interest and baseline PMD as a covariate. These analyses will accommodate correlation among the within-subject responses between the two eyes; an exchangeable working correlation structure will be used. Standard errors for the model parameters will be estimated using the robust "sandwich" estimators. The model also does not rely on the assumption of normality.

This model will be used to determine Bonferroni-adjusted confidence intervals for the three pair-wise differences among the adjusted treatment group mean responses (treatment effects); likewise, tests will be performed to compare the adjusted treatment group means using a Bonferroni-adjusted two-tailed significance level. An overall confidence coefficient of 98.3% and corresponding significance level of 1.7% for each comparison will be maintained, but as discussed in section 3.7.2 below, the confidence coefficient for interval estimation and significance level for hypothesis testing will be adjusted for the interim analysis for efficacy.

3.2.3 Adjustment for Baseline Characteristics

Baseline characteristics of subjects will be summarized overall and by treatment group. Formal statistical comparisons between treatment groups will not be performed; instead, potential confounding will be explored by examination of the magnitudes of treatment group differences in the distributions of visual acuity, papilledema grade, and duration of IIH. Other variables for which there are baseline differences among treatment groups that could be associated with the outcome also will be evaluated for a potential confounding effect. Continuous variables will be described using means, standard deviations, medians, quartiles, and ranges, and categorical variables will be described using percentages.

If clinically important differences are found between the groups at baseline, the primary outcome analyses will be repeated after statistically adjusting for these differences. These analyses will be considered secondary, however.

3.2.4 Investigation of Treatment by Covariate Interactions

We will investigate the interaction between treatment group and selected baseline covariates (age, race/ethnicity, PMD (average of measures from two reliable visual field examinations), papilledema grade, RNFL thickness, total retinal thickness, optic nerve head volume, visual acuity, presence of transient visual obscurations, and the symptom of constant visual loss) separately by adding the appropriate main effect and interaction terms to the primary statistical model and testing for significance of the interaction.

Since the power to detect potentially meaningful interactions will be limited, the magnitudes of mean responses to treatment in the relevant subgroups will be examined. The observation of clinically important subgroup differences in mean treatment response will serve as hypothesis generation for possible future studies designed to address specifically the issue of differential therapeutic response. Although these analyses are purely exploratory, those involving papilledema grade, PMD, visual acuity, and race/ethnicity will be given higher priority.

Interpretation of subgroup analyses should be viewed with caution, irrespective of whether the overall analysis demonstrates a significant treatment group difference. Extreme caution is particularly needed in the absence of such an overall difference.

3.2.5 Verification of Model Assumptions

The underlying assumptions of the statistical model to be used in the primary analysis will be thoroughly checked (e.g., linearity) through assessment of scatterplots of change in PMD versus baseline PMD within treatment group, and remedial measures (e.g., inclusion of baseline PMD categorized according to randomization strata) may be taken if serious violations of these assumptions are detected.

3.2.6 Treatment of Missing Data

Multiple imputation will be used to deal with missing data. This will be applied using a regression-based imputation model. For subjects with complete data up to a particular visit, a linear regression model (identity link) will be fit that includes the outcome at that visit as the dependent variable and outcomes at previous visits and treatment group as independent variables. Separate models will be similarly constructed for each visit (Weeks 4, 8, 16, and 26). Using these regression models, a missing value for a subject at a particular visit will be imputed as a draw from the predictive distribution given the outcomes at previous visits (some possibly imputed) and treatment group. This will be done sequentially starting with the Week 4 visit and ending with the Week 26 (Month 6) visit. This process will be repeated 100 times, resulting in 100 complete analysis data sets. The analyses (Section 3.2.2) will be performed separately for each of the 100 complete analysis data sets, and the results will be combined into one multiple imputation inference (estimated treatment effect and associated confidence interval and p-value) using Rubin's rules. This approach is appropriate for data sets that have a monotone missing

data pattern. If the data set does not precisely have this pattern, the monotone data augmentation method using Markov-Chain Monte-Carlo as proposed by Li (1988) and Liu (1993) will be used to impute the small amount of missing data that is required to make the missing data pattern monotone before applying the multiple imputation algorithm described above. This approach should accommodate missing data in an appropriate way under the missing at random (MAR) assumption.

3.2.7 Sensitivity Analysis

The primary analysis will be completed using the ITT principle described in *Section 3.1*. The primary analysis will be repeated with the following considerations as a secondary analysis:

- According to treatment received (whether or not this was the randomly assigned treatment)
- Excluding participants without outcome PMD (Month 6 or exam at early treatment failure)
- Exclude participants who took less than 80% of tolerated dosage of medication
- Exclude participants with major protocol violations such as: receipt of additional treatment outside of treatment assignment prior to first of 6-month visit or visit at which treatment failure declared, or major eligibility deviation that could impact the ability to assess the primary outcome.

The identification of subjects to be excluded from these analyses will be determined before the masking is broken (i.e., before data analysis).

3.3 Analysis of Secondary Outcome Variables for Efficacy

3.3.1 Definition of Secondary Outcomes

The following are secondary outcomes for efficacy:

- Change from baseline PMD to Month 6 PMD
  - PMD defined as described in *Section 3.2.1*
  - Month 6 PMD value will be used even for participants with treatment failure prior to Month 6
- Change from baseline to Month 6 CSF opening pressure measurement by lumbar puncture
- Change from baseline to Month 6 in papilledema grade
  - Papilledema grade determined by reading center and grade determined by study center will be considered separate outcomes and will be analyzed separately
- Change from baseline to Month 6 (pre-lumbar puncture) in RNFL thickness
- Change from baseline to Month 6 (pre-lumbar puncture) in total retinal thickness
- Change from baseline to Month 6 (pre-lumbar puncture) in optic nerve head volume
- Change from baseline to Month 6 (pre-lumbar puncture) in ganglion cell layer thickness

- Change from baseline to Month 6 (pre-lumbar puncture) in optic nerve canal shape
- Change from baseline to Month 6 in ETDRS visual acuity score
- Change from baseline to Month 6 in SF-36v2 scores
    - Change will be calculated separately for each of the eight health domains and the physical component summary (PCS) and mental component summary (MCS) scores
    - Scores determined via SF-36v2 scoring software
- Change from baseline to Month 6 in VFQ-25 score
    - Score determined as an average of sub-scale scores as described in the VFQ-25 manual (..\..\References\Participant Questionnaires\VFQ25_Manual.pdf)
        - Scoring of missing items will follow the scoring instrument
- Change from baseline to Month 6 in VFQ-25 Neuro-Ophthalmic Supplement score
    - Score determined as described in the 10-item Neuro-Ophthalmic instructions (..\..\References\Participant Questionnaires\Prior to Executed Agreement\10ItemNeuroOpthalVFQ25.pdf)
        - Scoring of missing items will follow the scoring instrument
- Change from baseline to Month 6 in HIT-6
    - Score determined as a sum of scores for each question (Never = 6, Rarely = 8, Sometimes = 10, Very Often = 11, Always = 13; total score range 36 to 78 with higher score indicating greater life impact)
    - Surveys with $\geq$1 missing responses will be given an overall null score for the visit
- Change in headache severity from baseline to Month 6 as reported on the Headache Assessment form
    - Analyzed as a continuous value with the following categories assigned for display:
        - Mild = score 1-4 | Moderate = score 5-7 | Severe = score 8-10
- Month 6 VFRC determination of whether the visual field examination has improved, remained the same, or worsened
    - All available visual field examinations will be analyzed. For the initial secondary outcome, the Month 6 VFRC rating will be used; a supplemental secondary outcome will use the VFRC rating at time of treatment failure for subjects experiencing treatment failure prior to Month 6.
- Change in weight from baseline to Month 6

Data from all eligible eyes will be included in the analysis. With the exception of CSF opening pressure and VFRC determination of visual field status, the outcomes also will be examined with change from baseline to Months 12, 24, 36.

3.3.2 Secondary Statistical Models

Treatment effect on secondary outcome variables for efficacy that are continuous will be analyzed using the same methods described in *Section 3.2.2* for the primary outcome variable, except that a working correlation structure will not be needed for variables that are not eye-specific. Variables to be analyzed in this manner are:

- Change in PMD at Month 6
- Change in CSF opening pressure (does not require working correlation structure)
- Change in papilledema grade (both reading center and study center outcomes)
- Change in RNFL thickness
- Change in total retinal thickness
- Change in optic nerve head volume
- Change in ganglion cell layer thickness
- Change in optic nerve canal shape
- Change in ETDRS visual acuity score
- Change in SF-36 scores (8 subscale domains, PCS and MCS; does not require working correlation structure)
- Change in VFQ-25 score (does not require working correlation structure)
- Change in VFQ-25 Neuro-Ophthalmic Supplement score (does not require working correlation structure)
- Change in HIT-6 (does not require working correlation structure)
- Change in headache severity as reported on Headache Assessment form (does not require working correlation structure)
- Change in weight

The treatment effect on the ordinal outcome of Month 6 visual field examination status by the VFRC (improved, no change, worse) will be assessed via a logistic regression model (proportional odds model). A likelihood ratio test will be performed for significance of the adjusted treatment group odds ratios representing pair-wise treatment group comparisons; 98.3% confidence intervals will be constructed for these odds ratios.

All models will include treatment group as the factor of interest and the baseline value of the outcome variable as a covariate (to replace baseline PMD adjustment in the primary for continuous outcomes; Month 6 visual field examination status by VFRC will retain inclusion of baseline PMD).

3.3.3 Adjustment for Baseline Covariates

Assessment for and implementation of adjustment for baseline covariates will follow methods described in *Section 3.2.3*.

3.3.4 Investigation of Treatment by Covariate Interactions

Investigation of interaction between treatment group and selected baseline covariates will follow methods described in *Section 3.2.4*.

3.3.5 Verification of Model Assumptions

Methods described in *Section 3.2.5* will be used to check underlying assumptions of the statistical model for each outcome.

3.3.6 Treatment of Missing Data

Methods described in *Section 3.2.6* will be used to impute missing data at Month 6 for continuous outcomes.

Missing values of Month 6 visual field examination status will not be imputed.

3.4 Compliance Outcomes

Data concerning compliance with acetazolamide (pill counts, serum bicarbonate levels) and surgical therapy will be summarized by treatment group and visit. Subjects with two eligible eyes will have this information summarized by eye as well as by treatment group and visit.

3.5 Analysis of Safety and Tolerability Outcomes

3.5.1 Adverse Events

All reportable adverse events will be tabulated by treatment group in a listing of each reported Medical Dictionary for Regulatory Activities (MedDRA) term and summarized over each MedDRA System Organ Class. Details will be provided in a listing of each event.

In addition, the following will be tabulated by treatment group. When applicable, events will be tabulated by eye within person:

- Number of adverse events per subject
- Number of subjects with at least one event
- Number of subjects with at least one event per person-month (person-month based on subject duration in the study)
- Number of serious adverse events per subject
- Number of subjects with at least one serious adverse event
- Number of hospitalizations and reasons for the hospitalization (overall and per subject, separately)
- Number of adverse events thought by investigator to be related to study drug and frequency of each type of such event
- Number of subjects who stopped the intervention in response to an adverse event

For binary variables (occurrence of at least one event, occurrence of at least one serious adverse event, occurrence of at least one hospitalization, occurrence of at least one adverse event related to treatment, occurrence of stopping the intervention in response to adverse event), Fisher exact tests will be used to compare treatment groups. For counts (number of adverse events per subject, number of serious adverse events per subject, number of hospitalizations per subject), groups will be compared using Poisson regression with inclusion of the logarithm of study follow-up duration for the participant as an offset term; a negative binomial model will be considered if overdispersion is present, and a zero-inflated negative binomial model will be used if there appear to be excessive zero counts.

For each adverse event, the treatment groups will be compared in a pair-wise fashion regarding the occurrence of at least one event using Fisher's exact tests; numbers of individual events will also be described. The comparisons will be repeated excluding all mild symptoms. Similar analyses will be performed after grouping adverse events by body system using Medical Dictionary for Regulatory Activities (MedDRA) coding. All subjects will be included in these analyses.

3.5.2 Laboratory Test Results and Vital Signs

Continuous measures of safety such as laboratory test results (CBC with platelet count, electrolytes, potassium, bicarbonate, and liver function tests) and vital signs and anthropometric measures (blood pressure, weight, and waist circumference) will be analyzed descriptively through summary statistics and tabulations.

Proportions of subjects with particular laboratory test abnormalities will be compared between the treatment groups in a pair-wise fashion using Fisher’s exact tests.

3.6 Analysis of Long-Term Follow-Up Data

The following outcome will be examined in analysis of long-term follow-up data:

- Time from randomization to treatment failure

3.6.1 Analysis Model

Treatment group differences with respect to long-term outcomes will be assessed by fitting a Cox proportional hazards regression model with treatment group as the factor of interest and baseline PMD in the best eligible eye as a covariate. Due to the discrete time data, Efron approximation will be used to deal with tied event times. The model will be used to determine Bonferroni-adjusted 98.3% confidence intervals for the adjusted hazard ratios for the three pair-wise treatment group comparisons. Likewise, likelihood ratio tests will be performed for the significance of these hazard ratios using a Bonferroni-adjusted two-tailed significance level of 1.7%. Kaplan-Meier curves will be used to describe the cumulative probability of treatment failure over time in each treatment group.

For subjects who do not experience the outcome, event times will be censored at the last subject contact at which the subject was determined to not have experienced the outcome (e.g., at time of premature withdrawal from the trial or at the final trial visit).

3.6.2 Adjustment for Baseline Covariates

Adjustment for additional baseline covariates (in addition to baseline PMD and papilledema grade in best eligible eye) may be considered depending on the comparability of the treatment groups at baseline (refer to *Section 3.2.3*).

3.6.3 Investigation of Treatment by Covariate Interactions

Examination of interactions between treatment group and selected baseline covariates will be performed using the Cox proportional hazards model following the guidelines established in *Section 3.2.4*.

3.6.4 Verification of Model Assumptions

The underlying assumptions of the Cox proportional hazards model will be checked and a thorough analysis of the martingale residuals and other diagnostics will be performed. Remedial measures (covariate transformation) will be taken if serious violations of these assumptions are detected.

The proportional hazards assumption will be assessed graphically by plotting $\log(-\log(\hat{S}(t))$ vs. log(time) for each of the treatment groups, and by examining plots of smoothed Schoenfeld residuals. The assumption also will be examined by dividing the time scale into 6-month periods and estimating the treatment group hazard ratios separately in each of these periods through the use of time-dependent covariates. The period length of 6 months may be adjusted prior to unmasking, based on the observed distribution of event times, if relatively few events occur during 6-month periods. Treatment group comparisons will be described in this manner if the proportional hazards assumption appears to be seriously violated.

The assumption of independence between censoring time and the (unobserved) event time will be assessed by performing a sensitivity analysis in which subjects with event times that are censored prior to their scheduled end of follow-up will be treated as having experienced the event a short time (one week) after censoring.

3.7 Interim Analysis

3.7.1 Interim Analysis for Safety

Interim analyses of safety data will be performed periodically throughout the trial. While the safety of subjects will be the primary concern of the DSMC, it is difficult to formulate precise

stopping guidelines that would cover all of the possible situations that might arise. Adverse events, particularly serious adverse events and surgical complications, will have to be considered carefully by the DSMC in terms of treatment group imbalances and severity. Events of particular concern include the following: death, absence of light perception, hypokalemia (from Lasix use), surgery-associated visual loss (from either VPS or ONSF), fenestration failure and orbital infection (from ONSF), and shunt failure, infection, seizures, and subdural hematoma (from VPS). If potential safety concerns are identified, the DSMC may require review of visual field data in order to evaluate the risk-benefit of continuing the trial as planned or modifying (or halting) the trial.

3.7.2 Interim Analysis for Efficacy

We will perform a single interim analysis for efficacy based on the primary outcome efficacy variable. This will be performed after ~50% of the subjects have completed (or were scheduled to have completed, based on their randomization date) their Month 6 visit and will only include data from these subjects. Given that recruitment of the 180 subjects will take place over 3 years, assuming that recruitment is uniform over time, it is anticipated that slightly fewer than 70% of the subjects will be enrolled at the time of the interim analysis.

The analysis will involve pair-wise comparisons among the treatment groups with respect to the primary outcome efficacy variable; the significance level used for each comparison will be determined by an O'Brien-Fleming α-spending function for a two-group comparison divided by 3 (Bonferroni correction). In this case, the boundaries will be $Z = 3.394$ for the interim analysis and $Z = 2.400$ for the final analysis. Assuming that there are 30 subjects per group and that the standard deviation is 6.5 dB at the interim analysis, the boundary will be crossed if a group difference exceeds approximately 5.7 dB. This monitoring procedure will have a negligible impact on the overall Type I error probability: the significance level at the final analysis corresponding to $Z = 2.400$ is $\alpha = 0.0164$. Point and interval estimates of treatment effects for the primary analysis, as well as reported p-values, will be adjusted for the interim analysis. The bias-adjusted mean will be used for point estimation and confidence intervals and p-values based on the MLE (sample mean) ordering of the sample space defined by the group sequential design will be reported.

The efficacy boundary will be considered to be non-binding. We believe that it may be prudent to halt or modify the trial only if (1) two of the treatment groups are each shown to be superior to the third (in which case the third group may be dropped) or (2) one of the treatment groups is shown to be superior to each of the other two (in which case the trial may be halted). Of course, the relative safety profiles of the treatments would have to factor into these considerations as well.

3.8 Additional Tabulations

A flow chart will track all participants who were enrolled, indicating the reason for any who did not complete the study and providing details on randomized subjects determined to be ineligible. A summary of safety visits, phone calls, and information on completion of questionnaires will be compiled.

Tabulations will include:

- Demographic and clinical characteristics at enrollment by treatment group
- Eligible eye characteristics by treatment group
- Non-eligible eye characteristics by treatment group
- Overall number of treatment failures by treatment group
- Overall number of surgical malfunctions by treatment group (ONSF and VPS groups)
- Duration (days) from enrollment to randomization
- Duration (days) from randomization to surgery (ONSF and VPS groups)

Specific Protocol Tabulations by Treatment Group

<u>6-Month RCT</u>

- Receipt of surgery other than that randomly assigned
- Adherence to medical management and diet by visit
- Daily dosage of acetazolamide and furosemide prescribed at each visit
- Medical therapy at each visit
- ONSF completion status (ONSF group)
- VPS completion status (VPS group)
- Surgical complications (separately for ONSF and VPS groups)
- Reoperations (ONSF)
- Shunt revisions (VPS)
- Headache by visit
- Weight by visit
- Visual field at follow-up (eligible eyes and non-eligible eyes, separately)
- Visual acuity at follow-up (eligible eyes and non-eligible eyes, separately)
- Papilledema grading at follow-up (eligible eyes and non-eligible eyes, separately)
- RNFL and macula GCL at follow-up (eligible eyes and non-eligible eyes, separately)
- Protocol deviations

<u>Treatment Failure Identification Phase (Post 6-Month RCT)</u>

- Adherence to medical management and diet by visit
- Daily dosage of acetazolamide and furosemide prescribed at each visit
- Medical therapy at each visit
- Protocol deviations.