STATISTICAL ANALYSIS PLAN

SVI

A Multicenter Randomized Non-inferiority Clinical Trial of Rectal Indomethacin Alone vs. Indomethacin & Prophylactic Pancreatic Stent Placement for Preventing Post-ERCP Pancreatitis in High-Risk Cases

NCT02476279

Prepared by
Valerie Durkalski, Ph.D., MPH
Lydia Foster, MS
SVI Unblinded Statisticians
and
Cathie Spino, DSc
SVI Blinded Statistician

SVI Statistical and Data Management Center
Department of Public Health Sciences
Medical University of South Carolina
Charleston, SC 29425

Version 1.5
January 3, 2020

TABLE OF CONTENTS

1. STATISTICAL ANALYSIS PLAN AND STATISTICAL REPORTS

This document provides the details of statistical analyses planned for the SVI Trial, including an interim analysis for futility. In addition, it discusses the statistical issues relevant to these analyses (e.g., sample data to be used, missing data).

The Statistical and Data Management Center (SDMC) generates two statistical reports – an open report to be distributed to the SVI Trial Executive Committee and the Data and Safety Monitoring Board (DSMB), and a closed report to be distributed only to the DSMB. The timing of these reports is determined in consultation with the DSMB. Reports will be sent from the SVI SDMC to the NIDDK Liaison two weeks in advance of the scheduled meeting.

Each report provides cumulative summary statistics on enrollment; subject status in the study; baseline characteristics; protocol violations; safety data, including AEs and SAEs, severity, expectedness (anticipated/unanticipated) and relatedness to the study treatment; and data management/quality information (e.g., timeliness and completeness of data entry by the Clinical Sites via the SVI Trial Website; number of data queries generated and resolved). The statistics are provided for the overall study as well as by clinical center when applicable in the open report. For the closed report only, the statistics are also provided by treatment group (A vs B). If a report coincides in timing with a planned interim analysis, the analysis results are appended to the report.

2. SYNOPSIS OF THE STUDY

The purpose of the SVI study is to determine whether or not rectal indomethacin has no important loss of efficacy as compared to the combination of rectal indomethacin and prophylactic pancreatic stent placement in patients undergoing high-risk ERCP who require pancreatic stent placement (PSP) for the sole purpose of pancreatitis prevention. The primary efficacy endpoint is defined as post-ERCP pancreatitis defined per consensus (Altanta) criteria. Another way of stating the trial's purpose is that the proportion of subjects with post-ERCP pancreatitis on rectal indomethacin alone is not more than that of the combination of rectal indomethacin and prophylactic PSP by more than a pre-specified absolute amount (i.e., the non-inferiority margin).

The recently conducted trial by the SVI PI showed that indomethacin in addition to PSP protects against post-ERCP pancreatitis as compared to pancreatic stent alone. Because the combination treatment is already proven to be highly efficacious (low post-ERCP pancreatitis rate), the superiority of indomethacin alone in terms of efficacy is not of main interest. Rather, indomethacin alone offers the appealing characteristics of lower costs and potential of avoiding the phenomenon of attempted but failed PSP as well as the complications associated with placing a stent; therefore, the research goal is to determine if the post-ERCP pancreatitis rate of indomethacin alone is no more than some clinically acceptable amount from the rate for the combination treatment. To answer this question, a blinded, two-armed non-inferiority trial is designed. In summary, eligible subjects will be randomized to either the combination treatment or indomethacin alone. Subjects will be randomized during the ERCP procedure after eligibility is confirmed, and receive indomethacin at the time of randomization. The primary efficacy endpoint of post-ERCP pancreatitis within 2 days from randomization will be assessed by an independent adjudication panel. The duration of enrollment into the trial is anticipated to be four years. The subject follow-up period is 30 days from randomization.

3. OUTCOME VARIABLES

3.1 Primary Efficacy Outcome

The primary outcome is a binary outcome variable measuring the post-ERCP pancreatitis event for each subject. Post-ERCP pancreatitis (PEP) is defined as: 1) New or increased pain in the upper abdomen and 2) amylase or lipase ≥ 3x the upper limit of normal 24 hours after the procedure and 3) hospitalization (or prolongation of existing hospitalization) for at least 2 days (at least the night of ERCP & following night). These events must occur within 2 days of randomization in order to be considered PEP.

Using these consensus guidelines as a diagnostic framework, three adjudicators will independently assess for the development of PEP based on review of medical records and site-provided AE narrative for each study subject hospitalized within 2 days after the ERCP. PEP will be declared if at least 2 of the 3 adjudicators determine that a subject met all three components of the consensus diagnosis.

3.2 Secondary Efficacy Outcome

This study is designed to test the primary hypothesis. However, it also offers the opportunity to conduct analyses to evaluate important additional outcomes. The secondary outcome for the SVI trial is the proportion of subjects who experience moderate-severe post-ERCP pancreatitis events. Severity is defined per consensus criteria:

Mild - pancreatitis that results in hospitalization (or prolongation of existing hospitalization) for ≤3 days.

Moderate - pancreatitis that results in hospitalization (or prolongation of existing hospitalization) for 4-10 days.

Severe - pancreatitis that results in hospitalization (or prolongation of existing hospitalization) for > 10 days, or leads to the development of pancreatic necrosis or pseudocyst, or requires additional endoscopic, percutaneous, or surgical intervention.

This outcome will also be adjudicated by the independent panel.

4. NON-INFERIORITY MARGIN AND SAMPLE SIZE DETERMINATION FOR THE PRIMARY OBJECTIVE

4.1 Non-inferiority Margin

The non-inferiority margin is set at an absolute margin of 0.05; i.e., indomethacin alone will be considered non-inferior to the combination if it is shown that the upper limit of the two-sided 95% confidence interval of the absolute difference between the two PEP proportions (indomethacin – combination) is less than 5%. The value of this margin is based on a combination of statistical and clinical judgment and was chosen to ensure that the overall PEP proportion of the new treatment (indomethacin alone) demonstrates a clinically unimportant difference from the active comparator arm (the combination of stent and indomethacin) as well as a clinically relevant superiority over a putative placebo (i.e., stent alone). From a statistical perspective, the margin should retain at least 50% of the superiority of the combination of stent and indomethacin (the active control in the trial) when compared to stent alone.[1,2] Our recent indomethacin RCT revealed that the risk-adjusted proportion of subjects with post-ERCP pancreatitis

in the combination of indomethacin and stent arm was 9.5% and that of stent alone was 15.7%.[3,4] Unadjusted rates were 9.7% and 16.1%, respectively. Based on these results, the reduction of the combination arm on PEP as compared to stent alone is 6.4%. Therefore, taking a fraction of this value gives a non-inferiority margin (δ) of 3.2% (δ = .5*6.4). This margin is a fraction of the effect of combination compared to stent alone as assessed in the previous trial.

Independent to the statistical approach, a questionnaire was circulated to clinical stakeholders regarding how much better the combination would have to be in preventing PEP as compared to the indomethacin arm alone to justify continuing the use of the combination therapy. Seven of 11 respondents said 10% more and the remaining 4 said at least 5% more effective. Based on both the statistical and clinical information, the pre-specified margin was set at 5%. The clinical investigators judged that a difference in treatment effect of 5% or greater constitutes an important difference between PEP proportions for the two treatment arms indicating that a strategy of indomethacin alone is inappropriate to adopt due to the higher PEP proportion. Although the statistical guidance suggested a lower margin (3.2%), that small of a margin was considered impractical as it would require a sample size of approximately 1600 per arm if we want to maintain the proposed 85% power. In addition, it should be noted that the determining factor for claiming non-inferiority is the upper limit of a 95% confidence interval. If the rate in the PSP+RI arm is truly 9.7%, then the rate in the RI arm cannot be higher than 11.4% in order to claim non-inferiority as the two-sided 95% confidence interval for this difference is (-1.4%, 4.9%), assuming a sample size of 715 per arm. Assuming 1090 per arm, the rate in the RI arm cannot be higher than 12.0% (95% CI: -0.3%, 4.9%). Reducing the margin lower than 5% would have little gain on the clinical interpretation of the results and would result in an impractical increase in the sample size. As an example, assume a margin of 3% and a PSP+RI PEP rate of 9.7%, the trial would need 1700 per arm to show non-inferiority which would translate into an RI PEP rate of 10.6% and a confidence interval of -1.1%, 2.9%.

4.2 Sample Size Estimation

This study is a non-inferiority trial designed to determine if indomethacin alone is non-inferior to the combination arm (i.e., that the absolute difference between arms (indomethacin – combination) in the proportion of subjects with PEP is less than 5%). Sample size is estimated using a 95% confidence interval approach focusing on the upper confidence limit for a difference in proportions via simulation using nQuery.[5]

$$(\pi_I - \pi_C) \pm 1.96\sqrt{\frac{\pi_I(1-\pi_I)}{n_I} + \frac{\pi_C(1-\pi_C)}{n_C}}$$

where π_C represents the theoretical proportion of subjects with post-ERCP pancreatitis due to treatment with the combination, and π_I represents the theoretical proportion of subjects with PEP due to treatment with indomethacin alone.

Based on results from our prior randomized controlled trial, the proportion of subjects with PEP in the combination of indomethacin and stent is estimated at 9.5%.[3,4] This trial compared the PEP rate in patients at elevated risk for post-ERCP pancreatitis who

were randomized to rectal indomethacin versus those randomized to placebo. A secondary analysis of the trial compared the combination of pancreatic stent and indomethacin to that of stent alone. Post-ERCP pancreatitis occurred in 9.7% of those who received the combination (n=247) and in 16.1% who received stent alone (n=249). Risk-adjusted proportions were 9.5% and 11.5%, respectively. We chose to use the higher proportion for the active control (9.7%) for a more conservative approach to sample size estimation.

Based on the above information, the study is powered to assure 85% likelihood of identifying less than a 5% difference in PEP rates between the two treatment groups. For sample size estimation for a non-inferiority design, we set the event rate to be equal between the two treatment arms and the upper limit of the confidence interval at 5% (NI margin). The maximum sample size required for randomization is 1300 subjects (650 per treatment group). The ITT principle is applied to the primary analysis, and therefore, to safeguard against an approximate 5% drop-in/out and missing data in the two treatment groups, we inflate the sample size by a factor of 1.1 which is derived from $1/(1-R)^2$, where R is the proportion of non-adherence (Friedman et al, 1998, p108). Thus a total of 1,430 subjects will need to be enrolled and randomized.

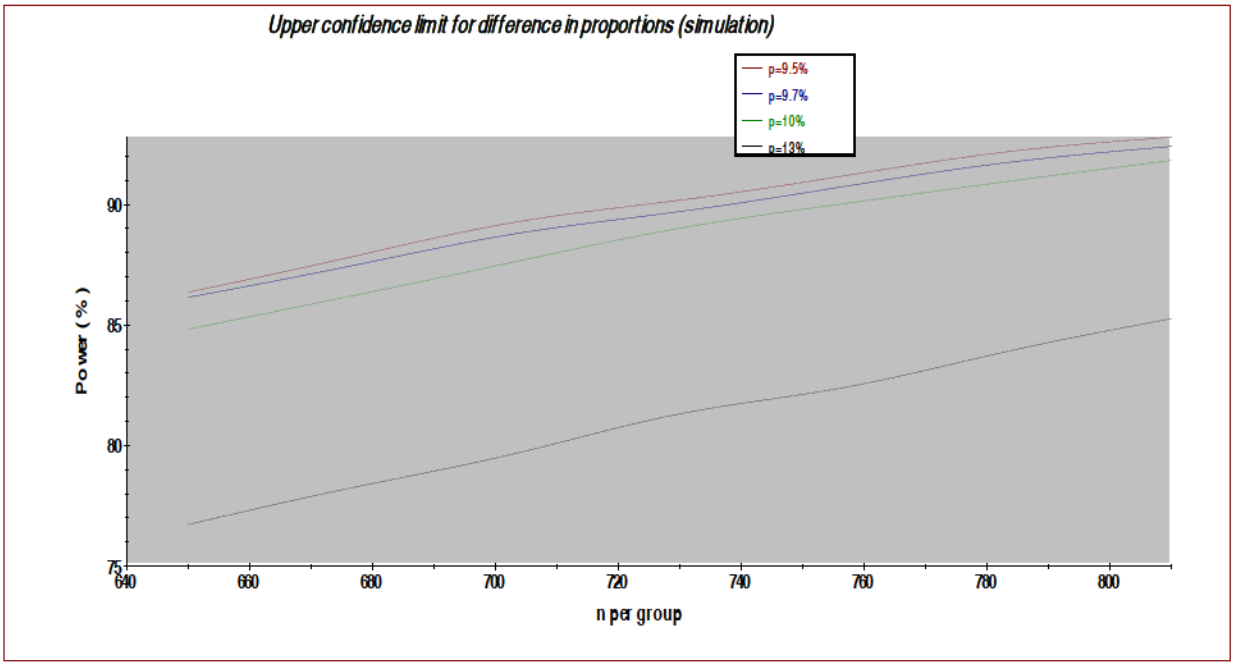
The sample size was increased to a maximum of 2,180 in order to attain 85% power. This increase was due to an observed higher event rate at the time of the first interim analysis. The DSMB recommendation was based on the review of the blinded sample size re-estimation and the conditional power analysis presented at the time of the first interim analysis in September 2017.

4.3 Sample Size Re-estimation Plan

We recognize that sample size estimation is based on assumptions and if the control (i.e., combination arm) PEP rate is higher than 9.7% then we may begin to see a decrease in power. We have looked at scenarios to examine the impact of a higher PEP rate than expected. Below is a power curve table assuming a control PEP rate of 9.5%, 9.7%, 10% and as high as 13% (the upper limit of the point estimate of the combination arm in the prior indomethacin trial). With a sample size of 650 per arm, the power drops from approximately 85% to 76% if the event proportion is as high as 13%. However, we anticipate that this high of an event proportion is unlikely and consider an event rate of 9.7 or 10% to be possible. To reduce the likelihood of an underpowered study due to incorrect sample size assumptions, a blinded sample size re-estimation will be conducted at the time of the first planned futility analysis (but prior to the futility analysis). The approach will follow Friede et al. methods using Blackwelder's sample size formula.[6] Because the trial's power is not substantially impacted if the overall observed rate is 9.7 or 10%, it is suggested that the recommendation to increase the original sample size only be considered if the observed rate is 13% or higher. The threshold of 13% was chosen based on the data from the prior indomethacin trial that showed a point estimate of the combination arm to be 9.7% (unadjusted) with a 95% confidence interval of (6.0%, 13.4%). Ultimately it is the DSMB's decision to recommend an increase in the total sample size and this decision should take into account the above proposed plan as well as the safety profile. Administratively, at the time of the planned interim analysis the unblinded statistician will provide the sample size review and include this information in the Closed DSMB report. This report will include the safety data (adverse event

information by treatment arm), data quality (protocol deviations by treatment arm) and the conditional power estimate.

The study team has no intention of decreasing sample size if the PEP rate is lower than 9.7%. The sample size re-estimation plan is for the sole purpose of avoiding an underpowered trial due to a higher than anticipated event rate. It is not for interim testing of a treatment effect and therefore will be conducting in a blinded manner. Combining sample size re-estimation with the proposed futility analysis will improve two aspects of the design – misspecification of the overall event rate (N re-estimation) and misspecification of the treatment effect (futility assessment) – while preserving the overall study error rates. In addition, the futility assessment will help protect from unnecessarily continuing the trial even if the re-estimation suggests a larger overall sample size.



5. DEFINITION OF TARGET POPULATION AND STUDY SAMPLES

5.1 Target Population

The targeted patient population for this trial is patients undergoing high-risk ERCP who require pancreatic stent placement for the sole purpose of pancreatitis prevention.

5.2 Intent-to-Treat Sample

As the primary analysis, all efficacy and safety outcome measures will be analyzed under the intent-to-treat principle (ITT). Under this principle, the evaluable sample includes all subjects who are randomized (regardless of what treatment was actually

received). Each subject will be analyzed according to the treatment group to which they were randomly assigned at the time of randomization.

It should be noted that the role of the ITT population in a non-inferiority trial is not equivalent to that of a traditional superiority trial.[7] The traditional role tends to dilute the treatment effect in favor of no difference between groups; however in the non-inferiority setting this dilution may favor the non-inferiority hypothesis.[8] Therefore the definition of ITT for a non-inferiority trial needs to take into account the impact of missing data on the primary outcome as well as that of non-compliance. For the SVI Trial, non-compliance of the indomethacin alone arm is anticipated to be rare due to the mode of delivery (rectal) and the one-time nature of this intervention. Non-compliance of the combination arm may occur if the endoscopist cannot place the stent however we anticipate this to be minimal. Both of these events are anticipated to be rare (no more than 5%). Missing primary outcome data is also anticipated to be rare with the expectation of minimal loss to follow up due to our brief 30-day follow-up period. Scenarios in which it may occur are if the subject cannot be reached and/or hospital records cannot be obtained. In the event of either of these scenarios, the primary outcome will be imputed (refer below to Handling of Missing Data Section).

5.3 Safety Analysis Sample

All randomized subjects are included in the safety analysis sample.

5.4 Per Protocol Sample

In addition to the defined ITT analysis sample, a per protocol sample is defined as a subset of the ITT sample. This sample will be used for secondary sensitivity analyses of the primary and secondary outcomes. The per protocol sample will include all randomized subjects that do not have the following protocol deviations:

- Eligibility violation
- Treatment crossover
- Missing primary outcome

6. RANDOMIZATION

A web-based central randomization system will be developed by the SDMC and installed on the WebDCU™ SVI study website using a minimal sufficient balance randomization scheme with a 1:1 allocation ratio.[9] Baseline site imbalance will be controlled at the minimal sufficient level so that there will not be serious treatment imbalance within site. The randomization algorithm will be fully detailed and documented in the SVI Randomization Plan and Validation Documents. These documents will be developed prior to the enrollment initiation and stored in a secure location at the SDMC. The documents will be archived with the study database at the end of the trial. Subjects, study coordinators and the adjudicators will all be blind as to which treatment was received. Subjects will not be told of their actual treatment until the Trial completion unless emergency unblinding is required. This information will be accessible in the event of an emergency via an emergency unblinding procedure (see MoP for further details).

7. BLINDING

The study is conducted in a single blind manner with double blind outcome assessment. The ERCP team will be unblinded to treatment assignment, however the study subject, clinicians involved in the post-procedure care of the subject, and outcome assessor will be blind to the treatment assignment. Treatment allocation is concealed and subjects will not be told of their actual treatment until the trial completion unless emergency unblinding is required.

The SDMC's unblinded staff will produce two identical sealed envelopes that contain identification of treatment codes. Prior to initiation of the trial, one envelope of (2) is given to the NIDDK Liaison to the DSMB. Another is maintained in a locked file cabinet at the SDMC in its limited access central file room.

The DSMB is partially unblinded for the closed reports (data reported by A and B only). However, if it so wishes, it may be completely unblinded at any time during the trial. If the DSMB wishes to be unblinded on a particular subject only, the NIDDK Liaison to the DSMB should email the request to the unblinded SDMC biostatistician.

8. MISSING DATA

Under the ITT principle, all subjects who are randomized are included in the analysis. Although every attempt will be made to prevent incomplete data, a certain amount of missing data is inevitable due to losses to follow-up or withdrawn consents.

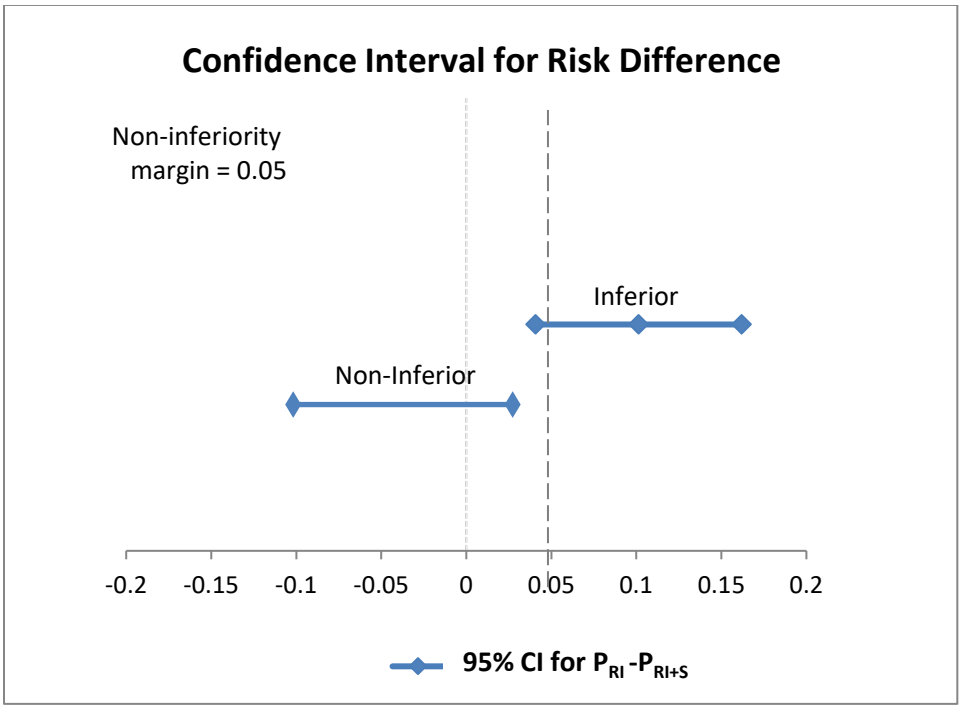
A thorough analysis of variables, reasons and patterns of missing data will be conducted. Based on the previously conducted indomethacin trial, we anticipate no more than a 5% non-adherence rate. The primary outcome will be determined from hospital records. If these data remain missing despite the extensive efforts mandated in the manual of operations to ensure complete subject follow-up, then the outcome will be imputed.

At the time of each planned analysis (interim and final), the unblinded statistician will report the amount of missing primary outcome data. If there is 5% or less missing primary outcome data, then all missing cases (regardless of treatment arm) will be considered worse case scenario (i.e., having a PEP event). If there is more than 5% missing data, then multiple imputation using SAS PROC MI and MIANALYZE will be used to impute the primary outcome. We chose 5% as our threshold for missing data due to the fact that the performance of the various imputation methods has negligible differences when the amount of missing data is small. In summary, 5 imputed data sets will be generated using PROC MI. The imputation model will include the primary analysis variables (PEP event, treatment) plus age and gender. Each imputed data set will be analyzed according to the specified primary analysis (Section 10.1.2) and MIANALYZE will be used to combine the results from the multiple imputed data sets to obtain a single set of parameter estimates. The multiple imputation method assumes missing at random (MAR) which means that the probability of missing outcome data can depend on the observed values of the individual but not on the missing values of the individual. Although we anticipate minimal missing data, sensitivity analyses will be conducted to assess the impact of any bias due to missing data. The final primary analysis will be re-run using complete case analysis (only those with outcome). If the treatment effect is robust, we expect these sensitivity analyses to yield similar inferences, particularly if the missing data are minimal (~5%). Any discrepancies between the sensitivity analyses and the primary analysis results will be investigated to understand the reason for the discrepancy.

9. EFFICACY ANALYSIS

9.1 Primary Outcome Variable Analysis

The analysis of the primary outcome is the difference ($p_I - p_C$) between the two treatment arms in the observed proportion of patients experiencing PEP. This difference will not be adjusted for covariates. However as a secondary analysis of the primary outcome, prognostic variables will be examined as described in detail below. Non-inferiority will be assessed on both the intention-to-treat and per protocol populations (defined above) at the one-sided 2.5% significance level by using the upper bound of the two-sided 95% confidence interval for the difference in PEP event rates. Non-inferiority can be claimed if both analysis populations show non-inferiority (i.e., the two-sided confidence interval is less than 5%).



If (and only if) non-inferiority is claimed, an analysis for superiority will be conducted using a one-sided two-sample test for independent binomial proportions. This test will be conducted at a 0.025 significance level.

Due to a protocol amendment that resulted in a change in the timing of the indomethacin administration (from end of procedure to time of randomization), we have included a sensitivity analysis of the primary outcome. The above analysis will be repeated only including those randomized under the new protocol where indomethacin should be administered at time of randomization. We also will examine the PEP rates in the population that was treated with indomethacin post-

ERCP (initial protocol) as compared to those treated at time of randomization.

9.2 Interim Analysis Plan

Two interim analyses for futility using conditional power will be conducted when approximately one-third (N~472) of the original sample size (N=1430) and one-half (N~1090) of the revised sample size (N=2180) have been evaluated for the primary outcome and all of the outcomes to be used in the analysis are adjudicated. This timeline can be altered based on the input from the DSMB. The goal of the interim analysis plan is to determine whether to stop the trial early because it is unlikely to show non-inferiority at the final analysis. A conditional power will be calculated to assess the probability of observing non-inferiority at the final analysis conditional on the observed data and assumptions on the PEP event rates for the remainder of the trial.[10] Conditional power will be calculated under two different assumptions: 1) the assumption that the PEP rates in the two treatment arms for the remainder of the trial will be the same as those hypothesized (p_C=9.7%; p_I=9.7%); and, 2) the assumption that the PEP rates observed at the interim analysis will be maintained for the remainder of the trial. The probability of the observed events rates continuing for the remainder of the trial will be provided to the DSMB. Specifically, we will calculate the probability of observing the results at the interim look (or more extreme values) assuming that the hypothesized event rates are true. If these probabilities are low, then it may be more appropriate to use the conditional power results under the assumption of the observed rates. Conditional power will be calculated for three analysis samples: the ITT samples (see section 5.2), the per protocol sample (see section 5.4), and the ITT time sample. The ITT time sample includes subjects randomized under protocol version 3 or later, where indomethacin should be administered at the time of randomization.

A formal, non-binding stopping guideline is provided to the DSMB in order to preserve the type II error rate (claiming inferiority when in fact non-inferiority is present). If the conditional power falls below 20% then the DSMB may consider recommending to stop the trial due to futility. A formal stopping rule for overwhelming 'non-inferiority' is not provided as the study team would only like to stop the study early for overwhelming 'superiority'. If the conditional power (as defined above) under either assumption is 100% at either interim assessment than the statistical team will conduct a one-sided binomial test of two independent proportions to determine if statistical significance for superiority is met. Superiority is based on this one-sided test and defined as the event proportion being statistically lower (p<0.025) in the RI arm as compared to the combination arm. If met, the DSMB may consider recommending to stop the trial due to overwhelming efficacy (superiority). A simulation study was conducted to examine the operating characteristics based on various conditional power thresholds using a control event proportion of 9.7% and 13% at the interim looks. These results guided our decision to only test for superiority if the conditional power is 100%.

First Interim Analysis (n=238 per arm)

Control arm PEP rate ≈ 9.7% (n=23)		
Subjects with PEP in Experimental Arm	Conditional Power	One-Sided Chi-Sq p-value

Control arm PEP rate ≈ 13% (n=31)		
Subjects with PEP in Experimental Arm	Conditional Power	One-Sided Chi-Sq p-value

N	Current PEP Rate	Assuming current PEP rate	Assuming 9.7% PEP rate	
11	4.62%	100%	100%	0.0164*
12	5.04%	100%	100%	0.0267
15	6.30%	100%	99%	0.0880
20	8.40%	99%	97%	0.3157
25	10.50%	79%	90%	0.3804
29	12.18%	28%	79%	0.1890
30	12.61%	19%**	76%	0.1539
35	14.71%	1%**	54%	0.0463
40	16.81%	0%**	32%	0.0107
43	18.07%	0%**	21%	0.0040
44	18.49%	0%**	18%**	0.0028

N	Current PEP Rate	Assuming current PEP rate	Assuming 9.7% PEP rates	
17	7.14%	100%	100%	0.0165*
18	7.56%	100%	100%	0.0250
20	8.40%	100%	100%	0.0515
25	10.50%	100%	98%	0.1967
30	12.61%	91%	93%	0.4455
35	14.71%	42%	82%	0.2979
37	15.55%	22%	76%	0.2160
38	15.97%	15%**	72%	0.1811
40	16.81%	5%**	64%	0.1234
45	18.91%	0%**	41%	0.0399
50	21.01%	0%**	21%	0.0102
51	21.43%	0% **	18%**	0.0076

*Stop for superiority **Stop for futility

Second Interim Analysis (n=545 per arm)

Control arm PEP rate ≈ 9.7% (n=53)				
Subjects with PEP in Experimental Arm		Conditional Power		One-Sided Chi-Sq p-value
N	Current PEP Rate	Assuming current PEP rate	Assuming 9.7% PEP rate	
35	6.42%	100%	100%	0.0227*
36	6.61%	100%	100%	0.0300
40	7.34%	100%	100%	0.0793
50	9.17%	100%	100%	0.3780
54	9.91%	100%	100%	0.4595
55	10.09%	99%	99%	0.4197
60	11.01%	90%	98%	0.2434
70	12.84%	21%	83%	0.0518
71	13.03%	15%**	81%	0.0430
80	14.68%	0%**	46%	0.0062
87	15.96%	0%**	20%	0.0010
88	16.15%	0%**	17%**	0.0008

Control arm PEP rate ≈ 13% (n=71)				
Subjects with PEP in Experimental Arm		Conditional Power		One-Sided Chi-Sq p-value
N	Current PEP Rate	Assuming current PEP rate	Assuming 9.7% PEP rate	
50	9.17%	100%	100%	0.0214*
51	9.36%	100%	100%	0.0273
60	11.01%	100%	100%	0.1528
68	12.48%	100%	100%	0.3927
69	12.66%	99%	100%	0.4282
70	12.84%	99%	100%	0.464
71	13.03%	99%	100%	0.5
72	13.21%	98%	99%	0.4643
80	14.68%	66%	95%	0.215
87	15.96%	20%	82%	0.0843
88	16.15%	15%**	78%	0.0723
90	16.51%	9%**	72%	0.0524
100	18.35%	0%**	31%	0.0079
103	18.90%	0%**	21%	0.0041
104	19.08%	0%**	18%**	0.0032

*Stop for superiority **Stop for futility

The SDMC will be responsible for conducting these analyses and compiling the reports for the DSMB. Since several factors need to be taken into consideration before stopping a study, safety and study progress also will be taken into consideration by the DSMB and NIDDK in the decision to stop the study if the futility boundary is crossed.

9.3 Secondary Outcome Variable Analysis

Analysis of the secondary outcome, the proportion of subjects with moderate-severe PEP, will be analyzed in a similar manner as the primary outcome.

We also will examine the impact of pre-specified characteristics on the primary and secondary outcomes. Specifically, the goal of this analysis is to examine prognostic variables and their impact on the primary analysis. We also will explore subgroup populations for future hypothesis generation. The baseline variables that will be examined are: age (continuous), gender, race, body mass index (continuous), suspicion of sphincter of Oddi dysfunction, prior post-ERCP pancreatitis, history of recurrent pancreatitis, manometrically documented sphincter of Oddi dysfunction, difficult cannulation, pre-cut (access) sphincterotomy, pancreatic sphincterotomy, pancreatic acinarization, biliary sphincterotomy, pancreatic stent placement, double wire technique, make/model/length/caliber of pancreatic stent, trainee involvement, prophylactic stent characteristics, type of sphincter of Oddi dysfunction, and inpatient vs. outpatient status. The assessment of each of these variables will be conducted using a logistic regression model with the dichotomous primary outcome, PEP event. Treatment arm will be a main effect in the model and the above covariates examined. Each covariate will be assessed individually first with a model that includes interaction effect with the treatment. If a significant interaction is observed ($p < 0.10$), then exploratory subgroup analyses will be considered. For each model, we will be looking at the parameter estimate of the covariate to examine its association with outcome. We also will be examining the parameter estimate of the treatment in the presence of the covariate. A multivariable model will be created based on the results of the individual models (a combination of statistical and clinical significance will be used for model inclusion). The final multivariable model will be based on goodness of fit criteria. For this final model, the adjusted risk difference will be estimated with a 95% confidence interval. If the upper limit of this confidence interval is 5% or greater, then non-inferiority of the treatments cannot be concluded in the presence of the particular covariates. This analysis will be run on both the ITT and PP analysis populations.

For the multivariable model, it is possible that several of the baseline characteristics will be eligible for the model. In this case, we will derive a risk score based on the methods used in the previous indomethacin trial.[3] Individual patient risk scores will be determined by assigning one point for each major inclusion criteria and 0.5 points for each minor inclusion criteria (see below table). Similar to the above methods, the derived risk score will be examined for inclusion in the final multivariable model. In addition to the risk score, we will use a propensity score approach to derive a baseline variable that describes the baseline risk. The results of each risk score approach will be compared to provide helpful information for future trials.

Major Risk Criteria (1pt each)	Minor Risk Criteria (0.5pts each)
Suspicion of SOD	Age < 50
Prior post-ERCP pancreatitis	Female gender
Difficult cannulation	History of recurrent pancreatitis
Pre-cut sphincterotomy	
Pancreatic sphincterotomy	

Several procedures have been incorporated into the study design (i.e., procedure manual, training, protocol violation monitoring, outcome adjudication) to reduce center effects; however, these effects should not be ignored for this trial. As a secondary analysis of the study outcomes, a two-sided confidence interval for the difference in the PEP event rate between the two treatment arms will be calculated using the Newcombe method, stratifying by site.[11] Non-inferiority will be concluded if the upper limit of the confidence interval is less than 5%. Similar to the assessment of potential prognostic variables, the results of this secondary analysis will be taken into account when interpreting the final results of the trial.

9.4 Exploratory Analyses

At the end of the study, study investigators may wish to explore other relationships between the treatment and outcomes and/or covariates. Because the number of subgroup analyses could be large and may be a combination of non-inferiority and superiority hypotheses, all subgroup analyses will be conducted using a two-tailed significance level of 0.01. The Publication Committee of the SVI Trial will review proof of concept papers with analysis plan submitted by any investigators wishing to do so before any further analyses are conducted by the study statisticians.

10. SAFETY ANALYSES

All adverse events and serious adverse events will be summarized by AE code in terms of frequency of the event, number of subjects having the event, severity, expectedness (anticipated/unanticipated) and relatedness to the study treatment. In addition to the continual monitoring of adverse events by the safety monitor and DSMB and the planned statistical monitoring for safety (described in the SVI Safety Plan), final analyses of specified safety outcomes will be conducted. The proportion of subjects experiencing serious adverse events will be compared between the two treatment arms using Fisher's exact test. All tests for safety will use a two-tailed significance level of 0.05. See the Safety Monitoring Plan for further details about safety monitoring.

11. REFERENCES

1. Committee for Medicinal Products for Human Use (CHMP) Guideline on the Choice of the Non-Inferiority Margin. *Statistics in Medicine* 2006, 25:1628-1638.

2. ICH Harmonized Tripartite Guideline, "Choice of Control Group and Related Issues in Clinical Trials (E10)" 2000.

3. Elmunzer BJ, Scheiman JM, Lehman GA, et al.; U.S. Cooperative for Outcomes Research in Endoscopy (USCORE). A randomized trial of rectal indomethacin to prevent post-ERCP pancreatitis. N Engl J Med 2012;366:1414-22.

4. Elmunzer BJ, Higgins PDR, Saini SD, et al. Does rectal indomethacin eliminate the need for prophylactic pancreatic stent placement in patients undergoing high-risk ERCP? Am J Gastroenterol. 2013;108:410-5.

5. Elashoff J. (2007) nQuery Advisor, Version 7.0 User's Guide. Los Angeles, CA.

6. Friede T, Mitchell C, Muller-Velten G. Blinded Sample Size Reestimation in Non-inferiority Trials with Binary Endpoints. *Biometrical Journal* 2007; 49(6); 903-916.

7. Friedman LM, Furberg CD, DeMets DL. Fundamentals of Clinical Trials, 3rd ed. Springer-Verlag, New York. 1998 (p.108).

8. Brittain E and Lin D. A Comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. *Statistics in Medicine* 2004, 24(1): 1-10.

9. Zhao W, Hill MD, Palesch Y. Minimal sufficient balance – a new strategy to balance baseline covariates and preserve randomness of treatment allocation. *Stat Methods Med Res.* 2012 Jan 26.

10. Bratton DJ, Williams HC, Kahan BC, Phillips PP, Nunn AJ. When inferiority meet non-inferiority: implications for interim analyses. *Clinical Trials* 2012; 9:605-609.

11. Yan X, Gang S. Stratified Wilson and Newcombe Confidence Intervals for Multiple Binomial Proportions. *Statistics in Biopharmaceutical Research* 2012; 2:3, 329-335.