



**LEOPARD: Looking at EVAR Outcomes by Primary
Analysis of Randomized Data**

STATISTICAL ANALYSIS PLAN (SAP)

Protocol CP-0011

Revised as of February 24, 2020

**A Multicenter, Observational, Post-Market, Real-World Study
to Assess Outcomes of Patients Treated with the
AFX® Endovascular AAA Delivery System
Compared to Other EVAR Devices for
Endovascular Abdominal Aortic Aneurysm Repair**

Contents

1	LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS.....	4
2	INTRODUCTION	6
2.1	ANALYSIS PLAN SCOPE	7
3	STUDY OBJECTIVES AND ENDPOINTS	8
3.1	STUDY OBJECTIVE.....	8
3.2	CLINICAL ENDPOINTS.....	8
3.2.1	Primary Endpoint.....	8
3.2.2	Secondary Endpoints.....	9
3.2.3	Procedural Outcomes	9
3.3	SUB-STUDY ENDPOINTS	10
4	STUDY DESIGN	10
4.1	GENERAL DESIGN	10
4.2	RANDOMIZATION AND BLINDING	10
4.3	SAMPLE SIZE DETERMINATION	11
4.4	SCHEDULE OF ASSESSMENTS	14
4.5	VISIT WINDOWS	14
5	ANALYSIS POPULATIONS.....	15
5.1	DISPOSITION CATEGORIES.....	15
5.2	ANALYSES POPULATIONS	16
5.2.1	Intent-To-Treat (ITT) and Modified Intent-To-Treat (mITT)	16
5.2.2	As-Treated (AT)	16
5.2.3	Per-Protocol (PP).....	17
6	STATISTICAL ANALYSIS	17
6.1	GENERAL CONSIDERATIONS	17
6.2	ADJUSTMENT OF COVARIATES	18
6.2.1	Clinical Endpoints	18
6.2.2	Multi-Center Data and Pooling	18

6.2.3	Missing Data and Sensitivity Analyses	19
6.2.4	Use of AT/PP Subsets	21
6.3	ANALYSES OF PRIMARY ENDPOINT	21
	Endpoint Analyses	21
7	PROTOCOL DEVIATIONS	24
7.1	MAJOR PROTOCOL DEVIATIONS	24
8	ENROLLMENT AND ACCOUNTABILITY	24
9	DEMOGRAPHICS AND BASELINE CHARACTERISTICS	25
9.1	DEMOGRAPHICS	25
9.2	CURRENT MEDICATIONS AND MEDICAL HISTORY.....	25
10	SECONDARY ENDPOINTS	27
11	MAJOR ADVERSE EVENTS (MAES) AT 30-DAYS AND 12-MONTHS	27
11.1	ARC POST 12-MONTHS, UP TO FIVE-YEARS	27
11.2	MORTALITY AND ANEURYSM-RELATED MORTALITY	27
11.3	ENDOLEAKS	28
11.4	SECONDARY PROCEDURES	28
11.5	DEVICE INTEGRITY	28
11.6	ADJUNCTIVE PROCEDURES (CONCOMITANT INTERVENTIONS) AT THE INDEX PROCEDURE	28
12	APPENDICES	29
12.1	POWER ANALYSIS FOR SUPERIORITY ENDPOINT IN PASS, ORIGINAL PROPOSAL	29
12.2	POWER ANALYSIS FOR NON-INFERIORITY ENDPOINT IN PASS, USING ORIGINAL ASSUMPTIONS.....	30
12.3	POWER ANALYSIS FOR NON-INFERIORITY ENDPOINT IN PASS, WITH SMALLER SAMPLE SIZE AND ORIGINAL ASSUMPTIONS.....	31
12.4	POWER ANALYSIS FOR NON-INFERIORITY ENDPOINT IN PASS, WITH SMALLER SAMPLE SIZE AND UPDATED ASSUMPTION OF ~2.5% DVANTAGE.....	31

1 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Table 1. Abbreviations and Terms

Abbreviation or Term	Definition
AAA	Abdominal Aortic Aneurysm
AC	Analysis Close Date
ADE	Adverse Device Effect
AE	Adverse Event
AICc	Corrected Akaike Information Criterion
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
AT	As-Treated
BMI	Body Mass Index
BP	Blood Pressure
BPM	Beats Per Minute
CC	Complete Case
CEC	Clinical Events Committee
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disorder
CRA	Clinical Research Associate
CRF	Case Report Form
CSR	Clinical Study Report
CT	Computerized Tomography
CTCAE	Common Terminology Criteria for Adverse Events
CV	Cardiovascular
DE	Data Extract Date
DMC	Data Monitoring Committee
DSMB	Data Safety Monitoring Board
EC	Enrollment Close Date

Table 1. Abbreviations and Terms

Abbreviation or Term	Definition
ECG	12-lead Electrocardiogram
Echo	Echocardiogram
eCRF	Electronic case report form
eGFR	estimated Glomerular Filtration Rate
EVAR	Endovascular Aneurysm Repair
FDA	Food and Drug Administration
HIPAA	Health Insurance Portability and Accountability Act
ICER	Incremental Cost-Effectiveness Ratio
ICF / PIC	Informed Consent Form / Patient Informed Consent
ICH	International Conference on Harmonization
ICU	Intensive Care Unit
INB	Incremental Net Benefit
ITT	Intent-to-Treat
LOCF	Last Observation Carried Forward
MAE	Major Adverse Event
MAR	Missing at Random
MCAR	Missing Completely at Random
MCS	Mental Component Summary
MH	Medical History EDC Form
MI	Myocardial Infarction
mITT	Modified Intent-to-Treat
MMSE	Mini Mental State Examination
MNAR	Missing Not at Random
MRI	Magnetic Resonance Imaging
NI	Non-Inferiority
OSR	Open Surgical Repair

Table 1. Abbreviations and Terms

Abbreviation or Term	Definition
PCS	Physical Component Summary
PP	Per Protocol
QALY	Quality-Adjusted Life Year
SADE	Serious Adverse Device Event
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCRCT	Screening Computerized Tomography EDC Form
SD	Standard Deviation
SE	Standard Error
SOP	Standard Operating Procedure
SS	Sub-Study
SVS	Society for Vascular Surgery
TEAE	Treatment-Emergent Adverse Event
TIA	Transient Ischemic Attack
TLGs	Tables, Data Listings, and Graphs

2 INTRODUCTION

The purpose of the LEOPARD study is to evaluate the outcomes of contemporary EVAR in a real-world population using two classes of EVAR devices (anatomic fixation and proximal fixation). The study intends to provide new, high-quality prospective EVAR data to the scientific community.

LEOPARD was powered to evaluate both non-inferiority and superiority of the anatomical fixation group as measured by Aneurysm Related Complications (ARC). The methodology to analyze the primary endpoint was provided in the original protocol CP-0011 Rev. 1.0. The protocol originally planned for enrollment of 804 subjects in order to provide

sufficient power for testing both the non-inferiority and superiority hypotheses. However, an unplanned analysis was performed at the request of regulatory authorities as the most contemporary source of clinical data available for AFX/AFX2 was the LEOPARD study. This analysis revealed that while the AFX group was trending towards a small advantage, the magnitude of this difference was too small to support superiority with the originally planned sample size. As a result of this observation, an informed recommendation was made to discontinue enrollment to power the superiority analysis. However, non-inferiority testing was sufficiently powered with the previously enrolled subjects. It is noted that the protocol allowed for premature suspension of enrollment if it became apparent that the study could not fulfill its aims. The endpoint analysis will proceed as originally declared in the protocol but with a smaller sample size. Additionally, all imaging-related findings will be evaluated using site-reported information instead of a core-lab due to resource limitations. This SAP will define how these updates will modify the original planned approach.

2.1 ANALYSIS PLAN SCOPE

This Statistical Analysis Plan (SAP) describes the planned analyses for the LEOPARD study. All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association ¹ and the Royal Statistical Society ², for statistical practice. It will serve as the main guidance for Biostatisticians and Statistical Programmers involved in data analysis.

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. If deviations from this analysis plan are made, they will clearly be labeled as such in the CSR. As mentioned prior, the SAP will modify the approach originally provided in the

¹ American Statistical Association. (2018) Ethical Guidelines for Statistical Practice. Prepared by the Committee on Professional Ethics, Apr 14 2018. <http://www.amstat.org/about/ethicalguidelines.cfm>

² Royal Statistical Society. (2014) The Royal Statistical Society: Code of Conduct. <http://www.rss.org.uk/Images/PDF/join-us/RSS-Code-of-Conduct-2014.pdf>

protocol CP-0011 Rev. 1.0. to account for the lack of available corelab data for the endpoint.

Post-hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post-hoc, or unplanned, exploratory analyses performed will be clearly identified as such in the final CSR. In some cases, particularly with respect to healthcare economic analyses, final model(s) may depend on the nature of the collected data and therefore it might not be possible to rigorously define each prospective model.

The reader of this SAP is encouraged to also read the clinical protocol, and other identified documents, for details on the planned conduct of this study.

3 STUDY OBJECTIVES AND ENDPOINTS

3.1 STUDY OBJECTIVE

The objective of the LEOPARD post-market study is to compare the Endologix AFX/AFX2 Endovascular AAA System using anatomical fixation, with other approved endovascular systems that use proximal fixation. Comparisons will be made using clinically relevant endpoints.

3.2 CLINICAL ENDPOINTS

3.2.1 PRIMARY ENDPOINT

The primary endpoint of the study is one-year survival in absence of Aneurysm Related Complications (ARC) at 1 year. ARC is a composite of the following:

- Peri-Operative Death (\leq 30 days)
- Aneurysm Rupture
- Conversion to Open Surgical Repair

- Post-operative Endoleaks
- Migration ($\geq 10\text{mm}$)
- Aneurysm Enlargement ($\geq 5\text{mm}$)
- Endograft Limb Occlusions
- Reinterventions for device- or aneurysm-related complications

3.2.2 SECONDARY ENDPOINTS

Occurrences of the following various adverse events and other device-based outcomes will be collected:

- Major Adverse Events (MAEs)
- ARC and components of ARC after 12-months up to 5-years
- Aneurysm-Related Mortality
- Endoleaks, Classified by Type
- Loss of Neck Apposition
- AAA-Related Secondary Procedures
- Loss of Device Integrity
- Adjunctive procedures deemed necessary during implant

Adverse event outcomes will be presented within the following time frames: within 30 days (early), 31 to 365 days (late), and at 1 to 5 years. Results will be reported for total patients impacted.

3.2.3 PROCEDURAL OUTCOMES

The following index procedure outcomes will be summarized. All evaluations will be descriptive in nature, no hypothesis testing will be performed.

- Total Procedure Time (time from first incision to closure)
- Duration of Endovascular Access (time from catheter introduction to catheter removal)

- Duration of Anesthesia
- Fluoroscopy Time
- Total Radiation Exposure
- Contrast Volume
- Time in ICU

3.3 SUB-STUDY ENDPOINTS

The protocol originally called for a sub-study to assess quality-of-life and health-care economic endpoints. The former sub-study is no longer planned for, as data was not collected. The economic sub-study will be addressed in separate documentation.

4 STUDY DESIGN

4.1 GENERAL DESIGN

The LEOPARD study is a prospective, randomized, multi-center study, intended to compare outcomes of the anatomically stabilized AFX/AFX2 Endovascular AAA System to a reference group composed of proximally-fixated EVAR devices. Each investigator selected one comparator device of their choice before enrolling their first patient and this device served as the comparator device for that Investigator throughout the course of enrollment. It is considered a major deviation for an investigator to use a comparator device different from the one they selected prior to enrolling their first subject.

4.2 RANDOMIZATION AND BLINDING

CP-0011 implemented a 1:1 randomization between AFX/AFX2 and a generic “Comparator”. Each investigator choose between the Medtronic Endurant, Cook Zenith, and Gore Excluder EVAR systems prior to enrolling their first subject. The

chosen EVAR system was logged into the EDC system and served as the comparator device for that investigator for the duration of the clinical study.

The 1:1 randomization was implemented via an unstratified 1:1 randomization list loaded into the EDC system – site-specific randomization lists were not created, and prospective stratifying factors were not taken into account. Balance in the 1:1 randomization was expected to be seen across all sites. This randomization list was generated using permuted blocks and the block size (4) and was blinded to the investigators, sites, and subjects. Investigators and study subjects were also blinded to treatment assignment, prior to treatment assignment being generated via the EDC system. Both investigators and study subjects are unblinded to treatment assignment after randomization.

4.3 SAMPLE SIZE DETERMINATION

This study was powered with respect to the primary clinical endpoint, which was an assessment of the proportion of subjects having freedom from ARC at 1-year. It is important to note that this assessment compared the AFX/AFX2 Endovascular AAA System to the pooled set of comparator devices. The primary clinical endpoint was designed to be sequentially evaluated, first for non-inferiority and then for superiority if the null hypothesis associated with non-inferiority is rejected. This “as-good-as-or-better” design³ will be done with no Type I error rate penalty⁴.

As the sample size is driven by the hypothesis with the highest performance thresholds, the superiority hypothesis determined LEOPARD’s original sample size. The assumptions for this hypothesis are seen in **Table 2**.

³ Julious, Steven. *Sample Sizes for Clinical Trials*. Boca Raton: Chapman & Hall/CRC. 2010.

⁴ U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). March 2010. *Guidance for Industry: Non-Inferiority Clinical Trials*. Silver Spring, Maryland.

Table 2. Superiority Sample Size Calculation Assumptions

Item	Assumed Value(s)
Statistical Test	One-sided 95% lower confidence interval (or equivalently, 90% two-sided interval) for the difference in proportions of ARC-free subjects
Hypothesis Formulation	$H_0: p_{AFX} - p_{COMPARATOR} \leq 0$ $H_A: p_{AFX} - p_{COMPARATOR} > 0$ Alternatively, the hypothesis may be written: $H_0: p_{AFX} \leq p_{COMPARATOR}$ $H_A: p_{AFX} > p_{COMPARATOR}$
Type I Error Rate	5% (i.e., $\alpha=0.05$)
Power	80% (i.e., $\beta=0.20$)
AFX Response Rate	86%
Comparator Response Rate	79%
Dropout Rate *	10%

*"Dropout" subjects for these purposes consist of (1) lost to follow-up, (2) withdrawal of informed consent, and (3) death on or after Day 30. Subjects dying prior to Day 30 contribute to the ARC endpoint as failures.

Based on these assumptions, a sample size of 804 subjects is required in order to achieve 724 subjects at Day 365, which yields the desired power superiority test characteristics. These sample size calculations were also performed with PASS version 12.0.2. Power calculations output for the original superiority test can be found in **Section 11.1**, below.

Given the sample size required for the superiority evaluation, the power of the non-inferiority test can be estimated based on the following assumptions:

Table 3. Non-Inferiority Sample Size Calculation Assumptions

Item	Assumed Value(s)
Statistical Test	One-sided 95% lower confidence interval (equivalent to a two-sided 90% CI) for the difference in proportions of ARC-free subjects
Hypothesis Formulation	$H_0: p_{AFX} - p_{COMPARATOR} \leq -8\%$ $H_A: p_{AFX} - p_{COMPARATOR} > -8\%$ Alternatively, the hypothesis may be written: $H_0: p_{AFX} \leq p_{COMPARATOR} - 8\%$ $H_A: p_{AFX} > p_{COMPARATOR} - 8\%$
Type I Error Rate	5% (i.e., $\alpha=0.05$)
AFX Response Rate	86%
Comparator Response Rate	79%
Margin	8%
Dropout Rate *	10%

*"Dropout" subjects for these purposes consist of (1) lost to follow-up, (2) withdrawal of informed consent, and (3) death on or after Day 30. Subjects dying prior to Day 30 contribute to the ARC endpoint as failures.

Based on these assumptions and the sample size required by the superiority test, the power of the non-inferiority test is estimated to be >99%. Sample size calculations were performed with PASS version 12.0.2 and the output is found in **Section 12.2**, below.

A simulation study using SAS/STAT® software was conducted to determine the power associated with the proposed two-stage hypothesis testing paradigm⁵. For the purposes

⁵ The output was generated using SAS/STAT software, version 9.4, of the SAS System for Windows. Copyright © 2010 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute, Inc., Cary, NC, USA.

of this simulation, the assumptions presented in **Table 3** were implemented and N=10,000 simulations of 724 subjects each were generated, in which half the subjects were assumed to receive AFX/AFX2 Endovascular AAA System and half the subjects were assumed to receive Comparator AAA EVAR. The RANBIN() function was used to generate the results for each subject; a fixed seed of 1,103 was used for the sequence of 10,000 Endologix AFX subjects, and a fixed seed of 1,104 was used for the sequence of 10,000 comparator subjects. Upon successful rejection of the non-inferiority null hypothesis, the superiority null hypothesis was tested. The sequence of null hypotheses was correctly rejected N=7985 times, showing that the original design has approximately 80% power to correctly reject both false null hypotheses under the assumptions in **Table 2** and **Table 3**.

The power of the non-inferiority test was re-evaluated using the new sample size and assumptions in **Table 3**. With a total of 455 subjects enrolled, 410 are assumed to reach the 1-year endpoint after 10% dropout. This sample size results in 99% power for the test (refer to **Section 12.3** below). It may also be reasonable to update the relative advantage assumption using the trends seen between the two groups. Updating both the sample size and the assumptions to an approximate 2.5% advantage for AFX/AFX2 Endovascular AAA System, results in a power of 90.6%. The power output from PASS is noted in **Section 12.4**, below.

4.4 SCHEDULE OF ASSESSMENTS

After discharge, subjects should have scheduled follow-up visits at 1-month, 6-months, 1-year, and annually through 5 years. This follow-up schedule may be modified to meet standard of care requirements at individual sites. Section 5.7 of the protocol (CP-0011) documents the suggested schedule of assessments.

4.5 VISIT WINDOWS

Because subjects may obtain imaging outside of the designated clinical assessment windows, it is necessary to implement data analysis windows that ensures all of the

important imaging data is evaluated at the relevant visits. **Table 4** presents the analysis windows that will be used in this study:

Table 4. Visit Windows

Visit	Target Day	Clinical Follow-up Window (days)	Imaging Analytic Window (days)
Index Procedure	N/A	0	0
1-Month	30	15 – 45	15-90
6-Months	182	152 – 212	91-304
1-Year	365	305 – 425	305-639*
Year 2	730	640 – 820	640-1004
Year 3	1095	1005 – 1185	1005-1369
Year 4	1460	1370 – 1550	1370-1734
Year 5	1825	1735 – 1915	1735-2130

* Per the protocol’s discussion regarding the primary endpoint, the 1-year endpoint will utilize imaging driven observations within 365 +/- 60 days (day 305-425).

5 ANALYSIS POPULATIONS

5.1 DISPOSITION CATEGORIES

A patient is considered a Study Subject once they sign the Informed Consent. They are given a subject number and following this, are randomized to a group. Per the Protocol, a study subject is considered Enrolled upon insertion of an EVAR delivery system. A study subject may be considered Discontinued for the following reasons:

- Lost-To-Follow-Up
- Withdrew Consent
- Explantation
- Death

- Other (e.g., discontinued due to physician discretion)

A study subject is considered to be a Completer if they do not discontinue. If interim administrative analyses of the data are undertaken, the subject dispositions will be presented in a compliance table. Once all subjects have either completed or discontinued, the number of active subjects will be zero (0).

5.2 ANALYSES POPULATIONS

5.2.1 INTENT-TO-TREAT (ITT) AND MODIFIED INTENT-TO-TREAT (MITT)

The ITT analysis population consists of all randomized subjects. While the intent of the study is to implant all randomized subjects, it is possible that some randomized subjects will not undergo implantation due to unforeseen issues. If this were to occur in a substantial number of subjects, it is possible that the results may be biased towards non-inferiority as the groups would become increasingly similar. Thus a mITT population is defined, which consists of all ITT subjects that subsequently underwent EVAR. All endpoints in this study, unless otherwise specifically noted, will be evaluated against the mITT analysis population. Subjects in the ITT analysis population are assumed to have been given their randomized EVAR treatment, regardless of the treatment actually given.

5.2.2 AS-TREATED (AT)

The AT analyses population consists of all subjects actually undergoing the Index Procedure and receiving an EVAR device. Subjects in the AT analyses population are analyzed according to the EVAR treatment actually given. As the ITT/mITT analysis may not be considered conservative, analysis of the AT population approach supports the ITT/mITT analysis by providing evidence of performance among groups consisting of the actual devices implanted. This can be useful to evaluate

the endpoint in case some subjects inadvertently receive a device other than what they were randomized to.

5.2.3 PER-PROTOCOL (PP)

The PP analysis population consists of all ITT subjects having no major protocol deviations that impact the primary endpoints. Note that this population may substantially overlap with the AT population. Major protocol deviations are defined in **Section 7**, below.

6 STATISTICAL ANALYSIS

6.1 GENERAL CONSIDERATIONS

Descriptive summaries will generally be provided for baseline characteristics such as demographics, medical history, and vascular dimensions. Quantitative variables (such as demographics and vascular variables) will be described using the following summary descriptive statistics: number of non-missing values (subjects), mean, standard deviation, median, first and third quartiles, and minimum and maximum values. Comparisons between sites or treatment groups for variables of this sort will be performed by 2-sided F-tests. Qualitative parameters (such as medical history) will be described using frequencies and percentages. Percentages will be calculated on the number of non-missing observations. In all cases, the number of missing values will be specified. Confidence intervals for relevant binomial variables will be calculated using exact methods. Comparisons will be made with Fisher's Exact Test for 2x2 tables, and the extension of the test (Fisher-Freeman-Halton) when more than 2 categories exist. Unless otherwise specified, the exact form of each algorithm used will be the default of SAS version 9.4 or later⁶.

⁶ SAS/BASE and SAS/STAT software, version 9.4 or greater, of the SAS System for Windows. Copyright © 2010 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute, Inc., Cary, NC, USA.

In general, analyses will be stratified by treatment groups for both the MITT and AT analysis populations. P-values for testing treatment difference on endpoints other than primary efficacy may be provided; when done, it will be for supportive purposes. All formal statistical tests outside of the primary endpoint will be two-sided unless otherwise specifically noted. Type I error rates will be controlled at $\alpha = 0.05$ on a per-hypothesis-per-analysis basis unless otherwise noted. The study has been powered for the primary endpoint only. No adjustment for multiple comparisons will be made.

6.2 ADJUSTMENT OF COVARIATES

6.2.1 CLINICAL ENDPOINTS

The primary and secondary clinical endpoints (see **Sections 3.2, 3.2.2, and 3.2.3** above) in this study are evaluated between ‘Endologix AFX’ and a pooled ‘Comparator’, where ‘Comparator’ consists of EVAR devices manufactured by Medtronic, Gore, or Cook.

In the case of further explorations of the primary and secondary clinical endpoints, either treatment group (Comparator vs. Sponsor) or manufacturer will be given priority and always included as predictor variables regardless of degrees-of-freedom consideration.

6.2.2 MULTI-CENTER DATA AND POOLING

It is prospectively planned to produce a pooled clinical efficacy data set across sites for the primary endpoint. This will be done prior to evaluating for heterogeneity of treatment effects across sites with respect to ARC at Day 365. There are a large number of planned sites (up to 80) and it is expected that some sites may recruit as few as one (1) subject. In order to perform a treatment-by-site interaction analysis, it is desirable to collapse the total number of considered sites to prevent SITE from consuming too many degrees of freedom. Therefore, pseudo-sites will be utilized.

Pooling will be performed by generally combining the numerically closest sites when sorted by site number, pooling previously created pseudo-sites into a site (if an event exists among them), and adding any remaining sites to an existing pooled site. Sites with 6 or fewer patients will first be sorted by site number. Starting with the lowest site number, sites will be combined into a pseudo-site not to exceed the mean of sites with counts >6. If sites with ≤ 6 patients remain, these will be combined into a second pseudo-site, and so forth. If there are sites left at the end of the process that cannot be combined into a pseudo-site of greater than 6 patients, these sites will be added to the last pseudo-site even if the resulting size is greater than the average of the remaining sites. To determine if there is a similar response across the study sites for each response, the primary endpoint will be tested for homogeneity by an extension of the Fisher's exact test (Fisher-Freeman-Halton). If the p-value for study site is less than 0.15 for any response, differences between study sites will be assumed. Site poolability will not be formally evaluated for any other endpoints. Ad-hoc attempts to find an explanation for the heterogeneity will be undertaken.

6.2.3 MISSING DATA AND SENSITIVITY ANALYSES

The central strategy to prevent missing data is to minimize its occurrence through careful design and execution of the study. It is the intent of this study to collect data as stated in the eCRFs without any missing values. It is noted that with modern EDC systems, there is greatly diminished likelihood of missing data for any events or outcomes for a clinical study subject. However, the potential for missing data remains, so analyses will be conducted to evaluate the robustness of the study results, accounting for missing observations. The primary endpoint will be analyzed based on

completed cases (i.e., those who have evaluable information for the endpoint), and thus approaches that evaluate missing data will help frame conclusions around the endpoint.

6.2.3.1 HANDLING OF DROPOUTS

It is known that the statistical methods available for handling missing data rely on assumptions that cannot be verified⁷.

Multiple imputation will be implemented in order to assess the endpoint, and will be limited to the clinical primary efficacy analysis. The clinical primary efficacy endpoint is a composite of several components as described in **Section 3.2** above.

Imputation methods for sensitivity will include the hot-deck procedure. Values will be assigned by randomly sampling subjects (with replacement) having similar baseline ASA classification scores and outcomes at 1-year (also known as “Hot Deck Imputation within classes.”) This will be performed ten times with ten different seeds (47744, 04603, 44522, 62783 39347, 72310, 41460, 31052, 40814, and 94297)⁸. The composite endpoint outcome will be imputed, not the individual components that make up the composite. The results of the ten imputations will be presented in a summary of the corresponding 95% confidence intervals for the difference between groups, and/or a pooled z-score by the method described in Rubin⁹ and available in SAS PROC MIANALYZE.

⁷ Committee for Medicinal Products for Human Use, European Medicines Agency. Guideline on Missing Data in Confirmatory Clinical Trials. July 2, 2010.

⁸ Selected from a pseudo-random number table in Steele R and Torrie J, p. 429 row 22. Principles and Procedures of Statistics 1960. McGraw-Hill, New York.

⁹ Rubin D, Little R. Statistical Analysis with Missing Data. 2002. John Wiley and Sons, New York.

6.2.4 USE OF AT/PP SUBSETS

The mITT and AT analysis populations will be applied to every clinical endpoint. The PP analysis population shall be applied to specific clinical endpoints as indicated in the relevant discussions below and will serve to demonstrate how the ideal subject fared in this trial. It is possible that the PP and AT populations will be equivalent. Results obtained from analysis of endpoints with the PP analysis population shall be considered supportive.

6.2.4.1 NON-INFERIORITY STUDY CONSIDERATIONS

It is known that ITT analyses are potentially less conservative when used exclusively to evaluate the clinical primary endpoint in non-inferiority studies (4). For this reason, the AT analysis population will also be applied to not only the clinical primary endpoint but also the clinical secondary endpoints. Also, rejection of the null hypothesis of inferiority will lead to a step-into-superiority analysis with no Type I error rate penalty (see **Section 4.3**).

6.3 ANALYSES OF PRIMARY ENDPOINT

ENDPOINT ANALYSES

The primary endpoint of the study is freedom from Aneurysm-Related Complications (ARC) at one year. ARC is a composite of the following:

- Peri-Operative Death (< 30 days)
- Rupture
- Conversion to Open Repair
- Post-Operative Endoleak
- Migration ($\geq 10\text{mm}$)
- Aneurysm Enlargement ($\geq 5\text{mm}$)

- Endograft Limb Occlusions
- Reinterventions for device- or aneurysm-related complications

The original study sample size was estimated under the intent of comparing the lower 95% confidence interval boundary to the non-inferiority margin and zero, in order to evaluate the non-inferiority and superiority hypotheses respectively. A p-value is superfluous when considering the one-sided confidence interval, though often is presented in non-inferiority studies. A p-value may be calculated by the un-pooled z-test (Wald). Power calculations were performed in PASS, and it was found that alternative tests (Farrington-Manning likelihood score, un-pooled z-test for two independent proportions) provided nearly equivalent sample sizes under the original assumptions. The un-pooled Z-test (Wald) test statistic for non-inferiority is given as:

$$Z_W = \frac{(\hat{\pi}_1 - \hat{\pi}_2) + \Delta}{\sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}}$$

where Δ represents the non-inferiority margin; if $\Delta=0$, this test statistic is equivalent to the un-pooled Z-test (Wald) for superiority. It is known that the Wald statistic for superiority offers poor performance in the presence of small sample sizes. It has been shown that Type I error rates are inflated in the presence of small sample sizes for non-inferiority as well¹⁰. Given the large sample sizes associated with the primary endpoint the Wald statistic for non-inferiority and superiority comparison of proportions is expected to perform adequately in the ITT analysis population, assuming there is minimal censoring. Furthermore, the Central Limit Theorem provides support for the assumption of normality behind the Z-test. To calculate

¹⁰ Kawasaki, Y., Zhang, F., and Miyaoka, E. (2010) "Comparisons of Test Statistics for Non Inferiority Test for the Difference between Two Independent Binomial Proportions", *Am. J. Biostat.*, vol. 1, no. 1, pps. 23-31.

the one-sided 95% confidence interval for the difference of the two groups for the primary endpoint in an exact fashion, the default PROC FREQ algorithm in SAS will be utilized. It is noted that SAS documentation states that the “method fm=score” statement provides exact confidence limits based on the Farrington-Manning likelihood score. This can be considered for additional information, with the expectation that differences in methodologies are minimal.

The primary clinical endpoint is an assessment of the proportion of subjects having freedom from ARC (see Section 3.2.1 above for a definition) at 1-year. For components of ARC that require imaging, it is noted that the protocol considers an imaging window (365 ± 60 days) for the purposes of including pertinent information. Components of ARC that do not require imaging for evaluation will be considered until post-op day 365. Each Discontinued subject is censored for the purposes of the endpoint unless they have already experienced an ARC event. The AFX/AFX2 Endovascular AAA System study arm will be compared to the pooled set of comparator devices (Medtronic Endurant, Cook Zenith, and Gore Excluder EVAR systems). This will be done using a one-sided 95% lower confidence interval (or equivalently, 90% two-sided interval) for the difference in proportions of ARC-free subjects at 1-year. The primary clinical endpoint is designed to be sequentially evaluated, first from a non-inferiority assessment and secondly from a superiority assessment if the null hypothesis associated with non-inferiority is rejected. This will be done with no Type I error rate penalty (4).

Analysis of the primary efficacy endpoint will proceed with un-imputed, and imputed, mITT and AT data. The mITT data will be the primary dataset for making claims, while the other approaches will serve as a check on the reasonableness of the mITT results since it is known the mITT assessment of non-inferiority can be anti-conservative. If strong differences are found between the results, ad-hoc investigations will be undertaken to determine the cause.

7 PROTOCOL DEVIATIONS

Protocol deviations will be evaluated as per the clinical database and additionally as tracked by the operational study team members. Protocol deviations will be tabulated and a listing will be provided.

7.1 MAJOR PROTOCOL DEVIATIONS

The following are Major Protocol Deviations:

- An Investigator applies a device other than the one the subject was randomized to
- A study subject is found to have violated at least one Inclusion or Exclusion criteria
- The patient received a device that violated the anatomical IFU indications which could impact outcomes

8 ENROLLMENT AND ACCOUNTABILITY

Subject accountability by site through 5 years will be tabulated. This will include the number of subjects randomized, enrolled, number who died, number who discontinued or withdrew (subject self-withdrawal, site withdrawal of the subject, or converted to open repair/explanted), and completed through year 5. The reason for discontinuation/withdrawal will be reported. Compliance tables (similar to **Table 4** below) will be generated for the subjects across 5 years of follow-up. These will be provided for the overall study, and AT comparator device and AFX device groups. Corelab compliance will also be evaluated, if available.

Table 4. Patient Visit Compliance

Visit	Eligible for Follow-Up	Subjects with Follow-Up	Overdue (Past)	Missed Visit	In Window, Follow-Up Pending	Not due for next visit	Site Performed Imaging	Corelab Reviewed Imaging	Events Occurring Before Next Interval			
									Withdrawal LTF/	Died	Conversion	Conversion + Died
Operative	n	NA	NA	NA	NA	NA	NA	NA	x	x	x	x

Table 4. Patient Visit Compliance

Visit	Eligible for Follow-Up	Subjects with Follow-Up	Overdue (Past)	Missed Visit	In Window, Follow-Up Pending	Not due for next visit	Site Performed Imaging	Corelab Reviewed Imaging	Events Occurring Before Next Interval			
									Withdrawal LTF/	Died	Conversion	Conversion + Died
1-Month	n	x (y%)	x	x	x	x	x (y%)	x (y%)	x	x	x	x
6-Months	n	x (y%)	x	x	x	x	x (y%)	x (y%)	x	x	x	x
1-Year	n	x (y%)	x	x	x	x	x (y%)	x (y%)	x	x	x	x
2-Years	n	x (y%)	x	x	x	x	x (y%)	x (y%)	x	x	x	x
3-Years	n	x (y%)	x	x	x	x	x (y%)	x (y%)	x	x	x	x
4-Years	n	x (y%)	x	x	x	x	x (y%)	x (y%)	x	x	x	x
5-Years	n	x (y%)	x	x	x	x	x (y%)	x (y%)	x	x	x	x

9 DEMOGRAPHICS AND BASELINE CHARACTERISTICS

9.1 DEMOGRAPHICS

Demographics and baseline characteristics will be descriptively summarized for the comparator and sponsor groups. These include:

- Age (years)
- Gender
- Race (% Caucasian)
- Height (cm)
- Weight (kg)
- Calculated Body Mass Index (BMI)
- ASA Class
- Blood Pressure at Index Hospitalization (mmHg)
- Serum Creatinine (mg/dL)
- Calculated eGFR (mL/min/m²)
- Hemoglobin (g/dL)
- Medication Types

9.2 CURRENT MEDICATIONS AND MEDICAL HISTORY

Descriptive summary statistics of these data will be completed for all subjects in the ITT and AT populations, by treatment groups. Listings will also be provided.

Summary statistics will include the number and percentage of subjects taking:

- Aspirin (ASA)
- Non-aspirin anti-platelets
- Anti-coagulants
 - Among those taking anti-coagulants, use of Coumadin/Vitamin K antagonists
- Anti-hypertensives
 - Among those on anti-hypertensives, use of calcium channel blockers or ace-inhibitors
- Statins, and/or
- Analgesics

In a similar fashion, the number and percentage of subjects having the following medical histories will be presented:

- Angina
- Aortic Valve Repair or Replacement
- Arrhythmia
- Cancer
- Cerebrovascular accident
- Chronic Obstructive Pulmonary Disease
- Coagulopathy or uncontrolled bleeding disorder
- Congestive Heart Failure
- Coronary Artery Disease
- Diabetes Mellitus
- Family history of AAA
- Gastrointestinal Abnormality
- Heart Valve Disease
- History of Abdominal Surgery
- History of CABG
- History of smoking
- Hypercholesterolemia
- Hyperlipidemia
- Hypertension
- Liver Disease
- Myocardial Infarction
- Pacemaker or implantable cardioverter-defibrillator
- Paraparesis
- Paraplegia
- Percutaneous Coronary Intervention
- Peripheral Arterial Occlusive Disease
- Peripheral Vascular Disease
- Renal Insufficiency
- TIA
- Thoracic Aortic Aneurysm

10 SECONDARY ENDPOINTS

All secondary endpoints will be assessed for each treatment group using ITT and AT populations. No imputation will be performed, unless otherwise noted. These secondary endpoints will be presented descriptively, and with Kaplan-Meier estimates when noted. Non-inferiority and superiority testing are not performed for the secondary endpoints.

11 MAJOR ADVERSE EVENTS (MAES) AT 30-DAYS AND 12-MONTHS

The major adverse event composite and components will be presented after 30-days (early), 31 to 365 days (late), and at 12-months. Results will be presented for each treatment group using ITT and AT populations. Individual major adverse event components include (a) All-cause Mortality, (b) Bowel Ischemia, (c) Myocardial Infarction, (d) Paraplegia, (e) Renal Failure, (f) Respiratory Failure, (g) Stroke, and (h) Index Procedural Blood Loss $\geq 1,000\text{mL}$.

11.1 ARC POST 12-MONTHS, UP TO FIVE-YEARS

The Aneurysm-Related Complications analysis will be evaluated via construction of Kaplan-Meier freedom-from-event curves across the duration of the study. This does not serve as the statistical methodology that formally analyzes the endpoint, but rather provides supplemental information. Results will be presented for each treatment group using ITT and AT populations. Estimates will be provided at annual intervals. The number of subjects at risk and standard errors will be presented for these intervals. Superiority and non-inferiority testing will not be conducted at time points beyond 1-year. The overall ARC composite rate, and individual ARC components will be provided in tabular format.

11.2 MORTALITY AND ANEURYSM-RELATED MORTALITY

All-Cause Mortality (ACM) and Aneurysm-Related Mortality (ARM) will be analyzed

in a tabular descriptive fashion as well as by the Kaplan-Meier approach. Results will be presented for each treatment group using ITT and AT populations. These rates will be presented annually. The number and percentage of subjects with mortality (all-cause and aneurysm-related) will be presented within 30 days (early), 31 to 365 days (late), and annually (years 1 through 5).

11.3 ENDOLEAKS

The number of each Endoleak type (1A, 1B, II, III, IV, or unknown) will be tabulated across the imaging windows for each treatment group. A Kaplan-Meier analysis will be conducted for the various types of Endoleak. Interventions for Endoleak will be presented descriptively.

11.4 SECONDARY PROCEDURES

Device related Secondary Interventions (including those performed for resolution of Endoleaks, device thrombosis/occlusion, rupture, migration, fracture, kinking, infection, aneurysm sac expansion, and/or a device defect) will be presented within 30 days (early), 31 to 365 days (late), and at 1 to 5 years. The overall secondary procedure incidence and the individual component incidence will be provided. Kaplan-Meier time to event analyses will be presented for secondary procedures.

11.5 DEVICE INTEGRITY

Measures of device patency (stent stenosis and occlusion) and integrity (kinking and fracture) will be tabulated and presented descriptively.

11.6 ADJUNCTIVE PROCEDURES (CONCOMITANT INTERVENTIONS) AT THE INDEX PROCEDURE

Concomitant interventions will be tallied by treatment group and tabulated by type.

12 APPENDICES

12.1 POWER ANALYSIS FOR SUPERIORITY ENDPOINT IN PASS, ORIGINAL PROPOSAL

Two Independent Proportions (Superiority by a Margin) Power Analysis

Numeric Results of Tests Based on the Difference: P1 - P2

H0: P1 - P2 ≤ D0. H1: P1 - P2 = D1 > D0. Test Statistic: Z test (unpooled)

	Sample Size	Sample Size	Prop	Prop H0	Prop H1	Diff	Diff	Target	Actual	
	Grp 1	Grp 2	Grp 2 or Control	Grp 1 or Trtmnt	Grp 1 or Trtmnt	if H0	if H1	Alpha	Alpha	Beta
Power	N1	N2	P2	P1.0	P1.1	D0	D1	0.0500		0.1993
0.8007	362	362	0.7900	0.7900	0.8600	0.0000	0.0700			

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 100.

References

- Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
- Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference or Non-Unity Relative Risk.' Statistics in Medicine, Vol. 9, pages 1447-1454.
- Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
- Gart, John J. and Nam, Jun-mo. 1988. 'Approximate Interval Estimation of the Ratio in Binomial Parameters: A Review and Corrections for Skewness.' Biometrics, Volume 44, Issue 2, 323-338.
- Gart, John J. and Nam, Jun-mo. 1990. 'Approximate Interval Estimation of the Difference in Binomial Parameters: Correction for Skewness and Extension to Multiple Tables.' Biometrics, Volume 46, Issue 3, 637-643.
- Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
- Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
- Miettinen, O.S. and Nurminen, M. 1985. 'Comparative analysis of two rates.' Statistics in Medicine 4: 213-226.

Report Definitions

'H0' is an abbreviation for the NULL hypothesis. This is the hypothesis being evaluated by the statistical test.

'H1' is an abbreviation for the ALTERNATIVE hypothesis. This hypothesis gives the 'true' parameter values.

'Power' is the probability of rejecting a false null hypothesis. It should be close to one.

'N1 and N2' are the sizes of the samples drawn from the corresponding populations.

'P2' is the proportion for group two. This is the standard, reference, baseline, or control group.

'P1.0' is the proportion for group one (treatment group) assuming the null hypothesis (H0).

'P1.1' is the proportion for group one (treatment group) assuming the alternative hypothesis (H1).

'Target Alpha' is the probability of rejecting a true null hypothesis that was desired.

'Actual Alpha' is the value of alpha that is actually achieved.

'Beta' is the probability of accepting a false H0. Beta = 1 - Power.

Summary Statements

Group sample sizes of 362 in group one and 362 in group two achieve 80% power to detect a

2 Musick • Irvine, CA 92618 • (949) 595-7200 • Fax (949) 595-7373

www.endologix.com

Page 29 of 31

difference between the group proportions of 0.0700. The group two proportion is 0.7900. The group one proportion is assumed to be 0.7900 under the null hypothesis and 0.8600 under the alternative hypothesis. The test statistic used is the one-sided Z test (unpooled). The significance level of the test was targeted at 0.0500. The significance level actually achieved by this design is NA.

12.2 POWER ANALYSIS FOR NON-INFERIORITY ENDPOINT IN PASS, USING ORIGINAL ASSUMPTIONS

Power Analysis of Non-Inferiority Tests of Two Independent Proportions
Numeric Results for Non-Inferiority Tests Based on the Difference: P1 - P2
H0: P1 - P2 ≤ D0. H1: P1 - P2 = D1 > D0. Test Statistic: Z test (unpooled)

	Sample Size	Sample Size	Non-Inf. Grp 2 Prop	Non-Inf. Grp 1 Prop	Actual Grp 1 Prop	Non-Inf. Margin Diff	Actual Margin Diff	Target Alpha	Actual Alpha	Beta
	Grp 1	Grp 2	P2	P1.0	P1.1	D0	D1			
Power	N1	N2								
0.9999	362	362	0.7900	0.7100	0.8600	-0.0800	0.0700	0.0500	0.04990	0.0001

Note: exact results based on the binomial were only calculated when both N1 and N2 were less than 2000.

References

- Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
- Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference or Non-Unity Relative Risk.' Statistics in Medicine, Vol. 9, pages 1447-1454.
- Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
- Gart, John J. and Nam, Jun-mo. 1988. 'Approximate Interval Estimation of the Ratio in Binomial Parameters: A Review and Corrections for Skewness.' Biometrics, Volume 44, Issue 2, 323-338.
- Gart, John J. and Nam, Jun-mo. 1990. 'Approximate Interval Estimation of the Difference in Binomial Parameters: Correction for Skewness and Extension to Multiple Tables.' Biometrics, Volume 46, Issue 3, 637-643.
- Lachin, John M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
- Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, Mass.
- Miettinen, O.S. and Nurminen, M. 1985. 'Comparative analysis of two rates.' Statistics in Medicine 4: 213-226.

Report Definitions

- 'Power' is the probability of rejecting a false null hypothesis.
- 'N1 and N2' are the sizes of the samples drawn from the corresponding groups.
- 'P2' is the response rate for group two which is the standard, reference, baseline, or control group.
- 'P1.0' is the smallest treatment-group response rate that still yields a non-inferiority conclusion.
- 'P1.1' is the treatment-group response rate at which the power is calculated.
- 'D0' is the non-inferiority margin. It is the difference P1-P2 assuming H0.
- 'D1' is the actual difference, P1-P2, at which the power is calculated.
- 'Target Alpha' is the probability of rejecting a true null hypothesis that was desired.
- 'Actual Alpha' is the value of alpha that is actually achieved. Actual Alpha is only shown when Exact Calculations are used (see the Options tab).
- 'Beta' is the probability of accepting a false H0. Beta = 1 - Power.

'Grp 1' refers to Group 1 which is the treatment or experimental group.
 'Grp 2' refers to Group 2 which is the reference, standard, or control group.
 'Non-Inf.' refers to a small distance from the reference proportion that is still considered non-inferior.
 'Actual' refers to the true value at which the power is computed.

Summary Statements

Sample sizes of 362 in group one and 362 in group two achieve 100% power to detect a non-inferiority margin difference between the group proportions of -0.0800. The reference group proportion is 0.7900. The treatment group proportion is assumed to be 0.7100 under the null hypothesis of inferiority. The power was computed for the case when the actual treatment group proportion is 0.8600. The test statistic used is the one-sided Z test (unpooled). The significance level of the test was targeted at 0.0500. The significance level actually achieved by this design is 0.0499.

12.3 POWER ANALYSIS FOR NON-INFERIORITY ENDPOINT IN PASS, WITH SMALLER SAMPLE SIZE AND ORIGINAL ASSUMPTIONS

Power Analysis of Non-Inferiority Tests of Two Independent Proportions

Numeric Results for Non-Inferiority Tests Based on the Difference: P1 - P2

H0: P1 - P2 ≤ D0. H1: P1 - P2 = D1 > D0. Test Statistic: Z test (unpooled)

	Sample Size Grp 1	Sample Size Grp 2	Non-Inf. Grp 2 Prop	Non-Inf. Grp 1 Prop	Actual Grp 1 Prop	Non-Inf. Margin Diff	Actual Margin Diff	Target Alpha	Actual Alpha	Beta
Power	N1	N2	P2	P1.0	P1.1	D0	D1			
0.9911	205	205	0.7900	0.7100	0.8600	-0.0800	0.0700	0.0500		0.0089

12.4 POWER ANALYSIS FOR NON-INFERIORITY ENDPOINT IN PASS, WITH SMALLER SAMPLE SIZE AND UPDATED ASSUMPTION OF ~2.5% DVANTAGE

Power Analysis of Non-Inferiority Tests of Two Independent Proportions

Numeric Results for Non-Inferiority Tests Based on the Difference: P1 - P2

H0: P1 - P2 ≤ D0. H1: P1 - P2 = D1 > D0. Test Statistic: Z test (unpooled)

	Sample Size Grp 1	Sample Size Grp 2	Non-Inf. Grp 2 Prop	Non-Inf. Grp 1 Prop	Actual Grp 1 Prop	Non-Inf. Margin Diff	Actual Margin Diff	Target Alpha	Actual Alpha	Beta
Power	N1	N2	P2	P1.0	P1.1	D0	D1			
0.9056	205	205	0.8350	0.7550	0.8600	-0.0800	0.0250	0.0500		0.0944