STATISTICAL ANALYSIS PLAN

Percutaneous deep vein arterialization for the treatment of late-stage critical limb ischemia
The PROMISE II Trial

Protocol Version:	LF-CA-PR-3, Rev. 2
Sponsor:	LimFlow, Inc.
Statisticians	Andrew Mugglin Roseann White

February 2, 2022

STATISTICAL ANALYSIS PLAN

TABLE OF CONTENTS

ABBREVIATIONS AND ACRONYMS

Acronym	Description
ABI	Ankle-Brachial Index
AE	Adverse Event
AFS	Amputation Free Survival
BMI	Body Mass Index
BTK	Below the Knee
CEC	Clinical Events Committee
CI	Confidence Interval
CLTI	Chronic Limb Threatening Ischemia
CRF	Case Report Form
CRO	Contract Research Organization
CTO	Chronic total occlusion
DMC	Data Monitoring Committee
IFU	Instructions for Use
IRB	Investigational Review Board
ITT	Intention-to-Treat
KM	Kaplan-Meier
LCL	Lower Confidence Limit
LTFU	Lost to follow-up
MAE	Major Adverse Event
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intention-to-Treat
OPC	Objective Performance Criterion
PAD	Peripheral Arterial Disease
PG	Performance Goal
PP	Per Protocol
PT	Preferred Term
PTA	Percutaneous Transluminal Angioplasty
PRO	Patient-reported Outcomes
QVA	Quantitative Vascular Angiography
SAE	Serious Adverse Event
SADE	Serious Adverse Device Effect
SAP	Statistical Analysis Plan
SD	Standard Deviation
SFA	Superficial Femoral Artery
SOC	System Organ Class
SSR	Sample Size Re-estimation
TLF	Tables, Listings, and Figures
TLR	Target Lesion Revascularization
TVR	Target Vessel Revascularization
UADE	Unanticipated Adverse Device Effect
UCL	Upper confidence limit
YCRG	Yale Clinical Research Group

STATISTICAL ANALYSIS PLAN

1 PROTOCOL SYNOPSIS

Principal Investigator	Daniel Clair, MD Prisma Health Midlands 9 Richland Medical Park Drive, Columbia, SC 29203, USA Mehdi Shishehbor, DO, MPH, PhD University Hospitals Cleveland Medical Center, 11100 Euclid Avenue Cleveland, OH 44106
Sponsor	LimFlow, Inc. 2934 Scott Boulevard Santa Clara, CA 95054 Telephone: (888) 478-7705 Facsimile: (408) 898-1459 Email: PROMISE@limflow.com
Independent Review Committee, **Clinical Events Committee,** **Data Monitoring Committee,** **Wound Core Lab,** **Statistical Analysis**	Syntactx 4 World Trade Center; 44th Floor 150 Greenwich St. New York, NY 10006 Phone: 212-228-9000 Fax: 646-375-3183
Unblinded Statistician	Andrew Mugglin, Paradigm Statistics
Blinded Statistician	Roseann White, Independent Contractor
Study Title	Percutaneous Deep Vein Arterialization for the Treatment of Late-Stage Chronic Limb-Threatening Ischemia (The PROMISE II Trial)
Investigational Device	LimFlow System™
Indication for Use	The LimFlow System is indicated to treat chronic limb-threatening ischemia by creating an arteriovenous (AV) connection in the below-the-knee vasculature for treatment of chronic limb-threatening ischemia (CLTI).
Intended Use	The LimFlow System is intended for endovascular, minimally invasive procedures in patients with a clinical diagnosis of chronic limb-threatening ischemia and who have been determined to have no surgical or endovascular treatment option (i.e., “no option”).
Study Objective	The objective of this US pivotal trial is to investigate the safety and effectiveness of The LimFlow System for creating an AV connection in

	the Below The Knee (BTK) vascular system using an endovascular, minimally invasive approach to arterialize the pedal veins to treat Chronic Limb Threatening Ischemia (CLTI) in subjects ineligible for conventional endovascular or surgical limb salvage procedures.
Study Population	The study population is comprised of subjects who are confirmed as "no option" (ineligible for conventional endovascular or surgical limb salvage procedures) by the Independent Review Committee (IRC). "No option" is defined as either a) absence of a usable pedal artery target (endovascular or surgical approach), or b) the presence of a pedal artery target with absence of a viable single-segment vein in either lower extremity or either arm that could be used for autogenous vein conduit.
Study Design	A prospective, single-arm, multi-center pivotal study designed to confirm the safety and effectiveness of The LimFlow System. The study will consist of a minimum of 60 and up to 105 subjects. This study utilizes a Bayesian Goldilocks adaptive design for sample size determination.
Number of Sites	Up to 25 sites in the United States, United Kingdom, and Japan
Primary Endpoint	Amputation Free Survival (AFS) defined as freedom from major amputation or death at 6 months, compared to a literature-based performance goal. **Major Amputation:** above-ankle amputation of the index limb *or* **Death:** all-cause mortality
Secondary Endpoints	**Primary Patency:** Defined as the absence of occlusion of the endovascular intervention maintained with no additional or secondary surgical or endovascular procedures, at 30 days and 6 months. **Primary Assisted Patency:** Defined as the absence of occlusion of the endovascular intervention maintained with the use of additional or secondary surgical or endovascular procedures, as long as occlusion of the primary treated site has not occurred, at 30 days and 6 months. **Secondary Patency:** Defined as the absence of occlusion of the endovascular intervention that is maintained with the use of additional or secondary surgical or endovascular procedures after occlusion occurs, at 30 days and 6 months. **Limb Salvage:** Defined as the percentage of subjects with freedom from above-ankle amputation of the index limb, evaluated at 30 days, 3 and 6 months. **Change in Rutherford Classification:** Defined as a change of one

class or greater, as evaluated at 30 days, 3 and 6 months.

Technical Success: Defined as the successful creation of an arteriovenous fistula in the desired limb location with immediate morphological success.

Procedural Success: Defined as the combination of technical success, and absence of all-cause death, above-ankle amputation or clinically driven major re-intervention[1] of the stent-graft at 30 days.

Target Wound Healing: Defined as complete healing of the patient's target wound as evaluated at 30 days, 3, 6, 9 months, and 1 year.

All Wound Healing: Defined as complete healing of the patient's wounds as evaluated at 30 days, 3, 6, 9 months, and 1 year.

All Wound Area Reduction: Defined as reduction in area of the patient's wounds as evaluated at 30 days, 3, 6, 9 months, and 1 year.

Freedom from Contrast-Induced Nephropathy: Defined as subjects without acute (within 72 hours after intravenous contrast administration) impairment of renal function, measured as an absolute ≥0.5 mg/dL (44 µmol/L) increase compared to baseline SCr value.

Procedure Time: Defined as the time of the first puncture (venous or arterial) to when the last catheter is removed.

Radiation Exposure: Defined as patient radiation exposure (in milligray) during the procedure.

Contrast Volume: Defined as the total volume of contrast media (in milliliters) given during the procedure.

Inclusion Criteria

1) Subject must be > 21 and < 95 years of age

2) Clinical diagnosis of critical limb ischemia, defined as the following clinical assessments: previous angiogram or hemodynamic evidence of severely diminished arterial inflow of the index limb (e.g., ABI ≤ 0.39, TP / TcPO2 < 30 mm Hg) and:

   a. Rutherford Classification 5, ischemic ulceration OR

[1] The creation of a new surgical bypass, the use of thrombectomy or thrombolysis (i.e., procedures done in the setting of lost primary-assisted patency), or major surgical revision such as a jump graft or an interposition graft performed for occlusion of the stent graft.

b. Rutherford Classification 6, ischemic gangrene

3) Subject has been assessed by the Principal Investigator, reviewed by the Independent Review Committee (IRC), and determined that no conventional distal bypass, surgical or endovascular therapy for limb salvage is feasible due to either a) absence of a usable pedal artery target (endovascular or surgical approach), or b) the presence of a pedal artery target with absence of a viable single-segment vein in either lower extremity or either arm that could be used for autogenous vein conduit.

4) Proximally, the Target In-flow Artery at the cross-over point must fall within the recommended vessel diameter ranges for the LimFlow stent graft by visual estimation.

5) Prior stent(s) to the infrainguinal arteries (e.g., iliac, SFA, and popliteal) are allowed.

6) Planned minor amputation (e.g., partial toe, ray or proximal foot/transmetatarsal) of target extremity within 30 days after the index procedure is allowed.

7) Subject is willing and able to sign the informed consent form.

8) Subject is enrolled in an acceptable wound care network and has an adequate support network to ensure that subject complies with medication regimen and follow-up study visits.

9) Prior to enrollment (7-day window), women of childbearing potential must have a negative pregnancy test.

10) The primary wound is stable (e.g., not rapidly deteriorating and/or showing signs of healing).

11) Stable glycemic control, HbA1C < 10% (<269mg/dL)

12) Subjects requiring dialysis may be included, provided they meet all these requirements:

    a. On dialysis for > 6 months
    b. Autologous arteriovenous (AV) fistulaaccess used for hemodialysis
    c. Serum albumin > 30 g/liter BMI > 20

Exclusion Criteria

Subjects will be excluded from participating in this study if they meet the following criteria before initiation of the endovascular procedure:

1) Concomitant hepatic insufficiency, thrombophlebitis in the target limb, or non-treatable coagulation disorder within the past 90 days.

2) Active immunodeficiency disorder or receiving immunosuppressant therapy.

3) Prior peripheral arterial bypass procedure above or below the knee, which would inhibit proximal inflow to the stent graft or interventional revascularization procedure within 30 days.

4) Previous major amputation of the target limb or presence of a wound requiring a free flap, or absence of adequate viable tissue.

5) Life expectancy less than 12 months.

6) Documented myocardial infarction or stroke within the previous 90 days.

7) Active infection (e.g., fever, elevated WBC count >20.0 x 109/L, and/or positive blood culture) at the time of the index procedure that may preclude insertion of a prosthesis or require major amputation (e.g., osteomyelitis proximal to metatarsals).

8) Known or suspected allergies or contraindications to aspirin or P2Y12 inhibitors, heparin, stainless steel, nitinol or contrast agent that cannot be adequately pre-treated.

9) Subject is taking anti-coagulants, which in the opinion of the investigator, interferes with the ability to participate in the study (i.e., intermittent interruption of therapy for a procedure may compromise subject's safety).

10) Lower extremity vascular disease that may inhibit the procedure and/or jeopardize wound healing (e.g., vasculitis, Buerger's disease, significant edema in the target limb, deep venous thrombus in the target vein, hyperpigmentation, or medial ulceration above the ankle).

11) Significant acute or chronic kidney disease with a serum creatinine of > 2.5 mg/dl in subjects not undergoing dialysis.

12) Severe heart failure (e.g., NYHA Class IV), which in the opinion of the investigator may compromise the ability to safely undergo a

percutaneous procedure.

13) Any significant concurrent medical, psychological, or social condition, which may interfere with the subject's optimal participation in the study, in the opinion of the investigator.

14) The subject is participating in another investigational drug or device study that has not completed the primary endpoint or that clinically interferes with the endpoints of this study.

15) Subject is unwilling, unable, or unlikely for cognitive or social reasons to comply with the protocol or follow-up requirements.

Duration of Study	Each enrolled subject will be followed for up to 3 years (± 4 weeks) post-procedure. The primary endpoint assessments will be completed by 6 months (± 2 weeks), and secondary endpoint assessments will be completed by 12 months (+/- 1 month) post-procedure.
Follow-Up Schedule	Measures of safety and effectiveness will be assessed through hospital discharge, at 14 (+/- 3) days, 30 (+/- 7) days, at 2, 3, 6, and 9 months (+/- 2 weeks), at one year (+/- 4 weeks), two years (+/- 4 weeks), and three years (+/- 4 weeks). The study will be completed and closed after all subjects have either met a primary endpoint or completed the 3-year follow-up requirements.
Independent Review:	An Independent Review Committee (IRC), a Clinical Events Committee (CEC), a Data Monitoring Committee (DMC), and a Wound Core Lab will provide independent oversight per their respective charters.

2 SCOPE

This statistical analysis plan covers the analyses to be performed after enrolled patients reach the primary endpoint according to PROMISE II protocol (LF-CA-PR-3).

3 STATISTICAL OBJECTIVES

3.1 Primary Objective

The primary statistical objective is to determine if the test device meets its performance goal, which is to establish that the amputation free survival rate at 6 months post-index procedure is statistically higher than 0.54.

3.2 Secondary Objectives

The secondary objectives consist of determining the test device's point estimate and its precision for the following endpoints:

- **Primary Patency:** Defined as the absence of occlusion of the endovascular intervention maintained with no additional or secondary surgical or endovascular procedures, at 30 days and 6 months.
- **Primary Assisted Patency:** Defined as the absence of occlusion of the endovascular intervention maintained with using additional or secondary surgical or endovascular procedures, as long as occlusion of the primary treated site has not occurred, at 30 days and 6 months.

- **Secondary Patency**: Defined as the absence of occlusion of the endovascular intervention maintained with using additional or secondary surgical or endovascular procedures after occlusion occurs, at 30 days and 6 months.
- **Limb Salvage:** Defined as the percentage of subjects with freedom from above-ankle amputation of the index limb, evaluated at 30 days, 3 and 6 months.
- **Change in Rutherford Classification:** Defined as a change of one class or greater, as evaluated at 30 days, 3 and 6 months.
- **Technical Success:** Defined as successful placement of the arterial and venous catheters in the desired location in the limb, ability to place the stent-grafts, completion of the endovascular procedure, and immediate morphological success.
- **Procedural Success:** Defined as the combination of technical success, and absence of all-cause death, above-ankle amputation or clinically driven major re-intervention of the stent-graft at 30 days.
- **Target Wound Healing:** Defined as complete healing of the patient's target wound as evaluated at 30 days, 3, 6, 9 months, and 1 year.
- **All Wound Healing:** Defined as complete healing of the patient's wounds as evaluated at 30 days, 3, 6, 9 months, and 1 year.
- **Freedom from Contrast-Induced Nephropathy:** Defined as subjects without acute (within 72 hours after intravenous contrast administration) impairment of renal function, measured as an absolute ≥0.5 mg/dL (44 µmol/L) increase compared to baseline SCr value.
- **Procedure Time:** Defined as the time of the first puncture (venous or arterial) to when the last catheter is removed.
- **Radiation Exposure:** Defined as patient radiation exposure (in milligray) during the procedure.
- **Contrast Volume:** Defined as the total volume of contrast media (in milliliters) given during the procedure

4 STUDY DESIGN OVERVIEW

This design utilizes a Bayesian Goldilocks[2] adaptive design for sample size determination. The pivotal cohort will consist of a minimum of 60 subjects and a maximum of 105 subjects. When N=60 subjects have been enrolled into this cohort, enrollment may stop based on a Bayesian predictive probability calculation of whether the study is likely to pass its primary endpoint hypothesis test. If enrollment does not stop, 15 more subjects will be enrolled (N=75 total in the pivotal cohort), and the process will be repeated. This may continue up to a maximum of N=105 subjects in the pivotal cohort. Once enrollment stops, 6 months of follow-up will ensue, at the end of which the one and only primary endpoint hypothesis test will occur.

The study was originally designed to enroll a maximum of 120 subjects rather than 105. After the N=90 but before the N=105 interim sample size assessments had occurred, it was determined that logistical and business constraints were forcing a mid-study revision of the maximum sample size to N=105. This SAP pre-specifies this change and discusses the statistical impact in Sections 6.6 and 6.7.

[2] Broglio KR, Connor JT, Berry SM., "Not too big, not too small: a goldilocks approach to sample size selection," *J Biopharm Stat.* 2014;24(3):685-705.

5 ANALYSIS POPULATIONS

5.1 Modified Intention-To-Treat Population

mITT is defined as all subjects where a LimFlow device was introduced into the patient, regardless of technical or procedural success or major protocol deviation. This is the primary population for analysis.

5.2 Per Protocol Population

The Per-protocol (PP) population is defined as all subjects who met all inclusion/exclusion criteria and had a successful procedure. This is a secondary population for analysis.

5.3 Untreated Population

This population consists of all consented subjects who met all inclusion and exclusion criteria but are not included in the mITT.population.

6 PRIMARY ENDPOINT ANALYSIS

The primary endpoint is the Amputation-Free Survival Rate (AFS) at 6 months (180 days). The primary endpoint will be evaluated against a performance goal of 0.54. See Appendix 1 (Section 10) for justification of this performance goal.

6.1 Statistical Hypotheses for the Primary Endpoint

The null and alternative hypotheses for the primary endpoint are:

$$H_0: \phi \leq 0.54$$

$$H_a: \phi > 0.54$$

where ϕ is the proportion of subjects that are amputation-free at 180 days (i.e., AFS at 180 days).

6.2 Statistical Analysis

The statistical methods are Bayesian, and ϕ is assigned a non-informative Beta(1,1) prior distribution. Given a set X of binary observations (amputation-free at 180 days, not amputation-free at 180 days, with no missing data), the posterior distribution $f(\phi \mid X)$ also follows a Beta distribution, by conjugacy.

The alternative hypothesis will be accepted (and the null hypothesis rejected) if the posterior probability of the alternative hypothesis exceeds a threshold ψ = 0.977, where ψ is chosen to control the type I error rate of the design at a level ≤ 0.025.

Because some missing data are expected, Bayesian multiple imputation will be employed in the principal analysis. More detail on this imputation is given in Sections 6.4, 6.5, and 11.

6.3 Sample Size Considerations

Although the analysis methods are Bayesian, the following frequentist scenarios are used to guide the determination of minimum and maximum sample size.

- According to PASS 13 [3], when using α = 0.025 (one-sided) in an exact binomial test, achieving 90% power to test the hypotheses in Section 6.1 when the true AFS = 0.70 would require a sample size of 101 subjects. Allowing for 10% missing data would require a sample size of N=113 subjects
- If the true AFS = 0.68, a sample size of 101 subjects would achieve 81% power. Allowing for 10% missing data would require a sample size of N=113 subjects

[3] PASS 13 Power Analysis and Sample Size Software (2014). NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/pass

- If the true AFS = 0.75, 90% power requires a sample size of N=56 subjects. Allowing for 10% missing data would require a sample size of N=63 subjects.

There is considerable uncertainty in the true AFS rate ϕ. Consequently, it is desirable to utilize a study design that achieves at least 80% power when AFS ≥ 0.68 and 90% power when AFS ≥ 0.70, while allowing early stopping of enrollment if the AFS rate proves to be very high (e.g., 0.75) or very low (≤ 0.54). Therefore, a study design is selected wherein the sample size is adaptively determined, with a minimum value of 60 and a maximum value of 105.

6.4 Sampling Plan (Interim and Final Analyses)

This study utilizes a Bayesian adaptive Goldilocks[4] design, in which several interim analyses are conducted for the purpose of assessing sample size adequacy, while the one and only analysis that tests the alternative hypothesis occurs once enrollment has stopped and all enrolled subjects have had the opportunity to be followed for 6 months (180 days). The schedule of analyses is as follows.

The originally planned set of possible sample sizes was $\{N_i = 60, 75, 90, 105, 120\}$. At the i^{th} interim analysis (when N_i subjects have been enrolled, i = 1,2,3,4), two predictive probabilities were to be calculated:

- PP_{win} = Pr(Eventual Win | observed and predicted outcomes, $N=N_i$ is the final sample size)
- PP_{fut} = Pr(Eventual Win | observed and predicted outcomes, N=120 is the final sample size)

Stopping of Accrual was to be based on the following criteria:

- Criterion for stopping accrual for expected success: PP_{win} exceeds a suitably high threshold W_i.
- Criterion for stopping accrual for futility: PP_{fut} is less than a suitably low threshold F

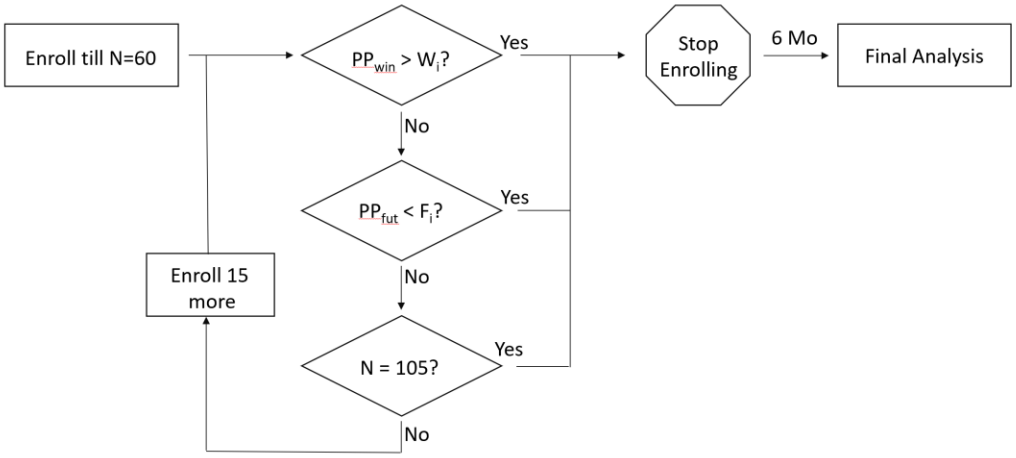
In either case, follow-up would continue until all enrolled subjects have had the opportunity to be followed for 6 months, and the final analysis (i.e., test of whether $\phi > 0.54$) would occur. If neither of these conditions holds at the i^{th} interim analysis, enrollment would continue to the next larger sample size, subject to a maximum of N=120 subjects.

The revision to a maximum of N=105 subjects requires only a change that $\{N_i = 60, 75, 90, 105\}$ and interim analyses for i = {1,2,3}. Since several of the previously planned interim sample size adequacy assessments have already occurred as of the time of the revision, the definition of PP_{fut} is not changing. I.e., PP_{fut} is still the probability of an eventual win with N=120 subjects. Even though there is no possibility that N=120 will actually be enrolled, PP_{fut} defined this way can still serve as a valid quantity for use in stopping enrollment for futility.

Figure 1 displays the modified design as a flowchart.

[4] Broglio KR, Connor JT, Berry SM., "Not too big, not too small: a goldilocks approach to sample size selection," *J Biopharm Stat*. 2014;24(3):685-705.

Figure 1 Flowchart of Sampling Plan



The predictive probability thresholds W_i and F_i are shown in **Table 1**. With the revision, the row for interim analysis 4 has been deleted, but all other quantities remain unchanged from the original design.

Table 1 Predictive Probability Thresholds for Stopping Accrual at Interim Analyses

Interim Analysis	Sample size (N_i)	Threshold for Stopping Accrual for Promising Trend (Wi)	Threshold for Stopping Accrual for Futile Trend (F_i)
1	60	0.98	0.05
2	75	0.95	0.10
3	90	0.90	0.15

To compute the predictive probabilities PP_{win} and PP_{fut} at each interim analysis, any subject with a missing 6-month outcome will have that outcome imputed via a Bayesian predictive probability model that incorporates observed follow-up to date. The set of imputed outcomes X^* is then combined with the set of observed 6-month outcomes X_{obs} to become a "completed" dataset, with a resulting posterior probability $P(\phi > 0.54 \mid X^*, X_{obs})$ that is easily calculated from the conjugate beta posterior distribution. This process is then repeated to obtain many completed datasets. The proportion of these completed datasets that result in $P(\phi > 0.54 \mid X^*, X_{obs}) > \Psi$ is the quantity PP_{win} defined above.

To calculate PP_{fut}, the dataset is augmented with $(120 - N_i)$ yet-to-be-enrolled subjects (who have 0 days of follow-up and no event), and outcomes for all subjects with a missing 6-month outcome will have that outcome similarly imputed via a Bayesian predictive probability model that incorporates observed follow-up to date. By combining imputed and observed 6-month outcomes, a "completed" dataset X^* is obtained. This process is then repeated to obtain many completed datasets, and the proportion of these completed datasets that result in $P(\phi > 0.54 \mid X^*, X_{obs}) > \Psi$ is the quantity PP_{fut} defined above.

When follow-up is complete, the final posterior distribution is computed via Bayesian multiple imputation. The imputation will use the same predictive model that is used at the interim analyses. Each "completed" dataset implies a value for $P(\phi > 0.54 \mid X^*, X_{obs})$; averaging these probabilities over the many completed datasets integrates out the uncertainty in the imputation model and produces the final posterior probability $P(\phi > 0.54 \mid$

X_{obs}). (See Appendix 2 (Section 11) for mathematical details.) This final posterior probability is then compared to the threshold ψ to determine whether the alternative hypothesis of Section 6.1 should be accepted.

6.4.1 Special Considerations for COVID-19

It is possible that COVID-19 will impact the incidence of AFS. To obtain data that adequately assesses this impact, subject accrual will not actually be stopped for futility unless both of the following criteria are met:

I. $PP_{fut} < F_i$, using AFS as the endpoint
II. $PP_{fut} < F_i$, using a modified definition of AFS as the endpoint. In this modified definition, deaths documented to be caused by COVID-19 will be excluded from the definition of AFS, and subjects who die of COVID-19 will be censored on their death dates.

If both of the above conditions are met, then accrual will stop for futility. However, if Condition I is met but Condition II is not met, then accrual of subjects will continue to the next larger sample size, subject to the maximum of N=105.

At the conclusion of the trial, analyses will be presented in two ways:

1. Analyses resulting from the dataset obtained when strictly following Condition I. This analysis will exclude all data from subjects accrued after Condition I was met and will likely indicate failure to meet the performance goal. This is the principal analysis.
2. Analyses resulting from all accrued data. This analysis will include data from subjects accrued after Condition I was met. This is a sensitivity analysis. Other planned sensitivity analyses are presented in Section 9.1**Error! Reference source not found.**.

6.5 Predictive Model

This section describes the prediction model that is used to predict the 6-month outcomes for subjects who have not yet reached 6 months. This model enables computation of predictive probabilities that form the basis for the sample size decisions in the Goldilocks design; it also serves as the “imputer’s model” in the Bayesian multiple imputation that is used in the final inference.

Outcomes at 6 months are binary and are denoted "E" (for having experienced a primary outcome Event) or "N" (for not having experienced an event). At any analysis, it is expected that some subjects will not have completed 6 months, and the 6-month outcomes for those subjects will be imputed based on a statistical model, as described below.

At the time of analysis, subjects will fall into one of three categories:

A. Status E or N has been observed at 180 days. For example, subject is known to be event-free at 180 days (N), or subject has had a primary event prior to 180 days and at least 180 days has elapsed since that subject entered the analysis cohort (E). (Note: an event at exactly 180 days will count as an event.)

B. Subject is known to have had an event at time < 180 days, but the subject has not yet had the opportunity to be followed for 180 days. For example, subject dies at Day 100 but the data cutoff date for analysis is only 25 days later, before the subject has the opportunity to be followed for 180 days.

C. Subject is censored without known event at time < 180 days.

Subjects in Group A have observed 180-day event status and need no imputation. Subjects in Group B will have their 180-day status set to E (with probability 1). Subjects in Group C will have their 180-day outcome imputed to be E with a probability that is subject-specific and is based on how long that subject has been followed and also on the event-free survival experience of other subjects, calculated via a Bayesian piecewise exponential survival model, as follows.

In this imputation model, the hazard function h(τ) is piecewise constant over the intervals [0, 30 days], (30 days, 60 days], (60 days, 90 days], and (90 days, 180 days]. Specifically, let each subject's time to event be modeled such that the survivor function is

$$S(t) = e^{-\int_0^t h(\tau)d\tau}, \text{ where}$$

$$h(\tau) = \begin{cases} \lambda_1, & \tau \in [0, 30\ days] \\ \lambda_2, & \tau \in (30, 60\ days] \\ \lambda_3, & \tau \in (60, 90\ days] \\ \lambda_4, & \tau \in (90, 180\ days] \end{cases}$$

and let $\Lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$. We assign each λ_i (i = 1, 2, 3, 4) an independent vague Gamma(0.001, 0.001) prior distribution (i.e., mean = 1 and variance = 1000), and because of the well-known Poisson-Gamma conjugacy, the posterior distribution f(λ_i | data) is a Gamma(0.001 + D_i, 0.001 + T_i) distribution, where D_i is the number of events that occurs in interval *i* and T_i is the total amount of follow-up over all subjects in interval *i* (in days). Moreover, because these are non-overlapping intervals, the posterior distributions of λ_1, λ_2, λ_3, and λ_4 are independent. Samples from the posterior distributions of λ_1, λ_2, λ_3, and λ_4 can be drawn directly, without the need for Markov chain Monte Carlo techniques.

For a subject j followed for Q_j < 180 days without a known event (i.e, censored at Q_j < 180 days), the probability θ_j that this subject has an event by Day 180 can be computed as

$$\begin{aligned} \theta_j &= \text{Prob}[T \le 180 \mid T > Q_j] \\ &= \text{Prob}[Q_j < T \le 180] \,/\, \text{Prob}[T > Q_j] \\ &= (\, S(Q_j) - S(180)\,) \,/\, S(Q_j) \end{aligned}$$

where S(t) = exp(-H(t)) and H(t) is the cumulative hazard function

$$H(t) = \int_0^t h(\tau)d\tau = \begin{cases} \lambda_1 t, & t \in [0, 30]\ days \\ 30\lambda_1 + (t-30)\lambda_2, & t \in (30, 60]\ days \\ 30\lambda_1 + (60-30)\lambda_2 + (t-60)\lambda_3, & t \in (60, 90]\ days \\ 30\lambda_1 + (60-30)\lambda_2 + (90-60)\lambda_3 + (t-90)\lambda_4, & t \in (90, 180]\ days \end{cases}$$

The distribution of θ_j can thus be easily computed from the piecewise constant form of h(τ) to be a function of Λ and Q_j. This can be easily and quickly accomplished multiple times by first drawing samples from the gamma posterior distributions of λ_1, λ_2, λ_3, and λ_4 and subsequently transforming these into samples from the subject-specific distribution of θ_j for subject j. The missing 180-day binary event status for subject j is then imputed multiple times from a Bernoulli distribution with parameter θ_j (once for each sampled value of θ_j).

6.6 Operating Characteristics of the Design

The design described above has been subjected to extensive simulation in order to evaluate its anticipated performance in a variety of scenarios. In the simulation, the proportion of simulated trials that result in accepting the alternative hypothesis of Section 6.1 is the estimated power when ϕ > 0.54 and the estimated type I error rate when $\phi \le 0.54$. The estimated operating characteristics are based on 10,000 simulated trials per scenario, with predictions for each subject with unobserved final outcome (for each of the 10,000 trials) based on 10,000 draws from the relevant posterior predictive probability distribution.

Table 2 displays the operating characteristics and summarizes the performance of the adaptive sample size determination algorithm under various assumptions about the amputation-free rate (ϕ). Column headings are:

- ϕ^* -– The amputation-free survival rate (at 180 days) for which data were simulated
- **Power** — The power of the study. I.e., the proportion (out of 10,000 simulated trials per row) that resulted in a determination that $\phi > 0.54$.
- **Average N** — the average sample size achieved (for the pivotal cohort) in each row (of the 10,000 simulated trials per row)
- **N=60 (Futile, Promise)** — the proportion of the 10,000 trials that stopped accrual at N=60, by reason
  - **Futile** — study success is unlikely even with the maximum sample size because $PP_{fut} < F_i$
  - **Promise** — study success is likely to be established with this sample size because $PP_{win} > W_i$
- **N=75 (Futile, Promise)** — analogous information about the N=75 sample size analysis
- **N=90 (Futile, Promise)** — analogous information about the N=90 sample size analysis
- **N=105 (Max)** — The proportion of the 10,000 trials that did not stop accrual until the maximum sample size was reached (N=105)

Table 2 Operating Characteristics of the Revised Design

ϕ^*	Power	Average N	At N=60 (Futile, Promise)	At N=75 (Futile, Promise)	At N=90 (Futile, Promise)	At N=105 (Max)
0.54	0.0239	74.8	0.4647	0.2317	0.1379	0.1556
			0.0019	0.0033	0.0049	
0.65	0.5892	90.4	0.0529	0.0594	0.0594	0.4926
			0.0911	0.1159	0.1287	
0.66	0.6682	89.9	0.0355	0.0422	0.0478	0.4753
			0.1181	0.1336	0.1475	
0.67	0.7419	88.4	0.0260	0.0314	0.0374	0.4334
			0.1486	0.1578	0.1654	
0.68	0.8019	87.0	0.0208	0.0228	0.0235	0.3952
			0.1726	0.1865	0.1786	
0.69	0.8547	85.1	0.0112	0.0158	0.0171	0.3537
			0.2190	0.2049	0.1783	
0.70	0.9051	82.5	0.0093	0.0099	0.0133	0.2888
			0.2654	0.2287	0.1846	
0.75	0.9915	71.4	0.0005	0.0004	0.0011	0.0746
			0.5219	0.2699	0.1316	

In Table 2, the observed type I error rate (when $\phi^* = 0.54$) is 0.0239 and the average achieved sample size is 74.8. In this scenario, the trial stopped accruing at N=60 46.66% of the time (46.47% for futility, 0.19% for promising trend), at N=75 23.50% of the time (23.17% for futility, 0.33% for promising trend), at N=90 14.28% of the time (13.79% for futility, 0.49% for promising trend), and at N=105 15.56% of the time.

When $\phi^* = 0.70$, the estimated power is 0.9051 and the average achieved sample size is 82.5. In this scenario, the trial stopped accruing at N=60 27.47% of the time (0.93% for futility, 26.54% for promising trend), at N=75 23.86% of the time (0.99% for futility, 22,87% for promising trend), at N=90 19.79% of the time (1.33% for futility, 18.46% for promising trend), and at N=105 28.88% of the time. Overall, the design performs as intended, with type I error rate ≤ 0.025, high power when $\phi^* \geq 0.68$, and smaller sample sizes when the simulated amputation-free survival rate ϕ is particularly high (e.g., $\phi^* = 0.75$) or particularly low (e.g., $\phi^* = 0.54$).

6.6.1 *Comparison to Operating Characteristics of Original Design*

Table 3 displays operating characteristics of the original design (with a maximum of N=120 subjects). In comparing Table 2 and Table 3, it is clear that the stopping probabilities at N=60, 75, and 90 have not changed, and the probability of stopping at N=105 in Table 2 is the sum of the probabilities of stopping at N=105 and N=120 in Table 3. For example, in the ϕ* = 0.54 row, Table 2 shows a probability = 0.1556 of stopping enrollment at N=105, which is the sum of 0.0688 + 0.0058 + 0.0810 (entries in the N=105 and N=120 columns in Table 3). The Average N in Table 3 tends to be slightly smaller, as is the Power. The observed Type I error rate of 0.0239 remains < 0.025 and not meaningfully different from the value 0.0244 seen in Table 3.

Table 3 Operating Characteristics of the Original Design

Expected Accrual Rate (5/mo)			Probability for Stopping Enrollment (by Reason)				
ϕ^*	Power	Average N	At N=60 (Futile, Promise)	At N=75 (Futile, Promise)	At N=90 (Futile, Promise)	At N=105 (Futile, Promise)	At N=120 (Max)
0.54	0.0244	76.0	0.4647	0.2317	0.1379	0.0688	0.0810
			0.0019	0.0033	0.0049	0.0058	
0.65	0.6336	95.2	0.0529	0.0594	0.0594	0.0466	0.3166
			0.0911	0.1159	0.1287	0.1294	
0.66	0.7130	94.3	0.0355	0.0422	0.0478	0.0357	0.2967
			0.1181	0.1336	0.1475	0.1429	
0.67	0.7810	92.3	0.0260	0.0314	0.0374	0.0238	0.2614
			0.1486	0.1578	0.1654	0.1482	
0.68	0.8423	90.3	0.0208	0.0228	0.0235	0.0180	0.2222
			0.1726	0.1865	0.1786	0.1550	
0.69	0.8879	87.9	0.0112	0.0158	0.0171	0.0118	0.1890
			0.2190	0.2049	0.1783	0.1529	
0.70	0.9253	84.6	0.0093	0.0099	0.0133	0.0080	0.1384
			0.2654	0.2287	0.1846	0.1424	
0.75	0.9945	71.7	0.0005	0.0004	0.0011	0.0001	0.0227
			0.5219	0.2699	0.1316	0.0518	

6.7 Simulation Details and Sensitivity

In order to generate data for the study of operating characteristics, several assumptions about the data mechanism had to be made. These assumptions are intrinsic to the generation of artificial data for the simulation exercise but are not part of the analysis that is conducted on any data (real or simulated). This section describes these assumptions and also the sensitivity of the design to these assumptions.

Timing and Rates of Events

Simulated event data are generated via a time-to-event mechanism using a piecewise exponential model. The default scenario uses a piecewise-constant hazard defined as

$$h(\tau) = \begin{cases} \lambda_1, & \tau \in [0, 30\ days] \\ \lambda_2, & \tau \in (30, 60\ days] \\ \lambda_3, & \tau \in (60, 90\ days] \\ \lambda_4, & \tau \in (90, 180\ days] \end{cases}$$

where $\lambda_1 = 0.005417298$, $\lambda_2 = 0.00647187$, $\lambda_3 = 0.004045362$, and $\lambda_4 = 0.0015350038$. Let the vector $\Lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$. These values are thought to reasonably represent what will occur in this study in that they induce event-free survival rates of S(30) = 0.85, S(60) = 0.70, S(90) = 0.62, and S(180) = 0.54. In order to generate data with varying values of S(180), we scale the vector Λ by a constant k, noting that

$$S(180) = \exp(-1 \times k \times\{ \lambda_1(30–0) + \lambda_2(60–30) + \lambda_3(90–60) + \lambda_4(180\text{-}90)\}) = e^{-k} \times 0.54$$

and for any desired value of S(180), such as 0.54, 0.65, 0.70, 0.75, etc., we can algebraically solve for k and then generate time-to-event data according to a piecewise exponential distribution with this new vector of parameters $k\Lambda$. In this manner, time-to-event data are generated for every desired ϕ^* value in the tables of operating characteristics (Table 2).

The censoring mechanism applied to these simulated event times is induced by the accrual rates, the timing of analyses, and by a random dropout pattern. For every simulated subject, a study start time is assigned according to an assumed accrual pattern, and when an analysis occurs, such as the first sample size evaluation when 60 subjects have been enrolled, each simulated subject's event time is censored at that time (if an event has not previously occurred for that subject). In addition, a designated fraction of simulated subjects is randomly chosen to be lost to follow-up at random times that are uniformly distributed over their 180-day study period, and these subjects are censored at the minimum of the dropout time and the analysis time (assuming they have not yet had an event). For other analysis times (interim or final), censoring times are simply re-computed based on the new time of analysis, and some simulated subjects who did not have an event at the first sample size analysis may have events at a subsequent analysis.

Lost to Follow-up / Data Missingness

Simulations were conducted under the assumption of 10% lost to follow-up (LTFU) over the 6-month study period. With each simulated trial, each subject was designated to be "lost" with probability 0.10; thus, the exact number of subjects designated to be lost from one simulated trial to the next varied, though on average it was 10%. Furthermore, those subjects designated as "lost" were assumed to have been lost uniformly in time over the 6-month period. Subsequent to this, subjects were assigned event times. If the event time occurred after the LTFU time, that subject was considered censored at the LTFU time, whereas if the event time occurred before the LTFU time, that subject had an observed event at the event time.

Rate of Accrual

The speed at which subjects are recruited into the trial is an important assumption embedded within the simulations to determine operating characteristics. The expected rate of accrual into the analysis population is 5 subjects/month, after an initial ramp-up period for the first 7 months of 1, 2, 2, 3, 3, 4, and 4 subjects/month. This "expected" accrual rate was used in the generation of operating characteristics shown in Table 2.

If enrollment is faster or slower than this assumed rate, the amount of information available at the sample size analyses would be affected, and thus the operating characteristics could also be affected. Table 4 illustrates how the study design performs if accrual is slower or faster than this expectation. The alternate accrual rates are:

- Slower accrual assumes 3 subjects/month (steady state) after an initial ramp-up period for the first 4 months of 1, 1, 2, and 2 subjects/month.
- Faster accrual assumes 7 subjects/month (steady state) after an initial ramp-up period for the first 5 months of 2, 3, 4, 5, and 6 subjects/month.

The results in Table 4 demonstrate that the type I error rate is controlled at the level 0.025 as the rate of accrual varies. The observed values are 0.0260 (slower enrollment) and 0.0253 (faster enrollment). These two values (as well as the value 0.0239 in Table 2) are all within the range of what would be expected when the type I error

rate is 0.025 in a simulation of 10,000 repetitions: the upper bound of a 95% probability interval = 0.025 + 1.96×√(0.025 × 0.975)/10000) = 0.0281, and all three values are less than 0.0281.

As with the operating characteristics in Table 2, these estimated operating characteristics are also based on 10,000 simulated trials per scenario, with predictions for each subject with unobserved final outcome (for each of the 10,000 trials) based on 10,000 draws from the relevant predictive probability distribution.

Table 4 Operating Characteristics of the Revised Design with Varying Accrual Rates

ϕ^*	Power	Average N	At N=60 (Promise, Futile)	At N=75 (Promise, Futile)	At N=90 (Promise, Futile)	At N=105 (Max)
Slower Accrual Rate (3/month) – Null Scenario						
0.54	0.0260	73.6	0.4978	0.2245	0.1241	0.1403
			0.0044	0.0038	0.0051	
Faster Accrual Rate (7/month) – Null Scenario						
0.54	0.0253	76.7	0.4020	0.2503	0.1543	0.1810
			0.0038	0.0035	0.0051	

6.7.1 *Comparison to Operating Characteristics of Original Design*

Table 5 displays corresponding operating characteristics of the original design (with a maximum of N=120 subjects). In comparing Table 4 and Table 5, it is clear that the stopping probabilities at N=60, 75, and 90 have not changed, and the probability of stopping at N=105 in Table 4 is the sum of the probabilities of stopping at N=105 and N=120 in Table 5. For example, in the first row, Table 4 shows a probability = 0.1403 of stopping enrollment at N=105, which is the sum of 0.0617 + 0.0061 + 0.0725 (entries in the N=105 and N=120 columns in Table 5). The Average N in Table 5 tends to be slightly smaller. The observed power (i.e, Type I error rate) is not meaningfully different (0.0260 and 0.0253 in Table 4, 0.0260 and 0.0246 in Table 5).

Table 5 Operating Characteristics of the Original Design with Varying Accrual Rates

			Probability for Stopping Enrollment (by Reason)				
ϕ^*	Power	Average N	At N=60 (Promise, Futile)	At N=75 (Promise, Futile)	At N=90 (Promise, Futile)	At N=105 (Promise, Futile)	At N=120 (Max)
Slower Accrual Rate (3/month) – Null Scenario							
0.54	0.0260	74.7	0.4978	0.2245	0.1241	0.0617	0.0725
			0.0044	0.0038	0.0051	0.0061	
Faster Accrual Rate (7/month) – Null Scenario							
0.54	0.0246	78.2	0.4020	0.2503	0.1543	0.0790	0.0956
			0.0038	0.0035	0.0051	0.0064	

7 SECONDARY ENDPOINTS ANALYSIS

A listing of secondary endpoints is given in Section 3.2. Secondary endpoints will be evaluated with descriptive statistics, along with 95% confidence intervals computed using frequentist methods. No hypothesis tests will be conducted.

Methods:

- Secondary endpoints that are binary (e.g., technical success, freedom from contrast-induced nephropathy at 72 hours) will be summarized as a proportion, with a 95% confidence interval computed via an exact binomial calculation.

- Secondary endpoints that involve the occurrence (or not) of events at times ≥ 30 days will be summarized with Kaplan-Meier time-to-event methods. Confidence intervals will be computed on the log(survival) scale.
- Continuous endpoints (e.g., procedure time, radiation exposure, contrast volume) will be summarized with means, medians, SD, quartiles, min, and max. Confidence intervals will be computed assuming approximate normality.

8 SAFETY ANALYSES

Adverse events (AE) will be processed by MedDRA coding with System Organ Class (SOC) and Preferred Term (PT) by a medical monitor. They will be evaluated further for severe adverse event (SAE) categorization. Adverse events will be summarized by tabulating the number and percentages of enrolled subjects experiencing each event through milestone time points. Device- and procedure-related adverse events will be summarized separately.

9 OTHER ANALYSES

9.1 Sensitivity Analyses for the Primary Endpoint

At the conclusion of the study, several planned analyses will assess sensitivity to certain analysis and data features.

9.1.1 *Sensitivity to Missing Data*

All reasonable efforts will be undertaken to minimize missing data. However, some missingness is inevitable, and the study is designed with the expectation that there may be up to 10% missing primary data at 6 months. The reasons for missing data will be described in detail and evaluated for assessment of possible bias.

By design, the analysis of the primary endpoint multiply imputes 6-month outcomes for any subjects without measured 6-month outcomes, based on available follow-up and the imputation model previously described (Section 6.5). Several additional analyses are planned that will explore the sensitivity of the main conclusions, specifically:

- A completers-only analysis of binary outcomes will be conducted. This analysis will exclude subjects with missing final primary outcome, presenting the posterior probability that $\phi > 0.54$ based on those subjects whose status is known at Day 180. The prior for ϕ will be Beta(1,1).
- A tipping-point analysis of binary outcomes will be conducted. This analysis will systematically consider each subject with missing outcome at Day 180 as an Event or a Non-Event at 180 days, in all combinations. Posterior probabilities (that $\phi > 0.54$) will be presented for each combination. The prior for ϕ will be Beta(1,1).
- A multiple-imputation analysis similar to the primary analysis will be conducted, only this sensitivity analysis will use a piecewise exponential prediction model with 6 pieces (rather than 4), with the time axis from 0 to 180 days partitioned such that there are approximately equal numbers of events in each piece. Specifically, if a total of V events are observed within [0, 180 days], the k^{th} piece (k=1, …, 5) in the partition will end with a cutpoint equal to the smallest integer day D such that $D \geq k \times V/6$. The 6^{th} piece ends at 180 days.

9.1.2 *Sensitivity to COVID-19 deaths*

By definition (Section 1), AFS includes deaths documented to be caused by COVID-19. The following additional analyses are planned to assess sensitivity of the study's conclusions to the handling of deaths related to COVID-19.

- An analysis wherein all deaths due to COVID-19 are censored on the date of death.
- A tipping point analysis in which COVID-19 deaths are included, or not, in the definition of AFS, in every combination. This analysis will use the methods of Section 6 (similar to the principal analysis) but will

conduct the imputation/prediction for subjects with missing final status as if they are similar to the cohort of subjects who did not die of COVID-19. In other words, if *k* subjects die of COVID-19, the parameters of the piecewise exponential model in Section 6.5 (the imputation model) will be estimated while removing the *k* subjects who died of COVID-19, but the final posterior probability will be computed by re-inserting the COVID deaths (in separate analyses) as follows:
    - *k* COVID deaths count as deaths at Day 180.
    - (*k*–1) COVID deaths count as deaths at Day 180, 1 counts as a non-AFS event at Day 180
    - (*k*–2) COVID deaths count as deaths at Day 180, 2 count as a non-AFS event at Day 180
    - Etc.
    - 0 COVID deaths count as deaths at Day 180, *k* count as non-AFS events at Day 180
- A frequentist competing risk survival model in which deaths due to COVID-19 are considered competing risks. Cause-specific cumulative incidence functions (CIF) will be estimated, and confidence intervals for the CIF at 180 days will be reported. The levels of confidence will be 95% (two-sided) and 97.7% (one-sided, analogous to the Bayesian posterior threshold $\psi = 0.977$).
- Other analyses suggested by patterns of observed data may be conducted.

9.2 Subgroup Analyses for the Primary Endpoint

Descriptive statistics and Bayesian posterior distribution summaries (e.g., mean, median, 95% credible interval) will be computed for the amputation-free rate parameter ϕ when computed separately according to each of the subgroups defined below. Pre-specified subgroups are:

- Sex (Male/Female)
- Dialysis status (Yes/No)
- Age (≤ 70, > 70)
- Diabetes (Type I, Type II, None)
- Race/Ethnicity
- Rutherford Classification

9.3 Demographic and Clinical Characteristics

Tables of baseline characteristics will be summarized using descriptive statistics. For continuous measures, the number of observations, mean, median, quartiles, minimum, and maximum will be reported for each measure. For categorical measures, the proportion, numerator, and denominator for each category will be reported.

Untreated subjects (see Section 5.3) will described with similar descriptive statistics and with a listing.

The proportion of subjects screened and enrolled will be tabulated by site, along with measures of compliance to follow-up visits.

9.4 Tabulation of Individual Participant Data

Patient data listings, including demographics, baseline characteristics, safety data, adverse events, procedural data, and endpoints, will be provided.

9.5 Heterogeneity

Heterogeneity across sites will be assessed as follows. Based on completers, a (frequentist) Cox proportional hazards regression will be conducted, with a term for site included in the model. If there is not a significant site effect ($p > 0.10$), data will be considered poolable across sites. If there is a significant site effect ($p < 0.10$), an investigation will be conducted into whether subject baseline characteristics explain this difference. Candidate baseline characteristics include those that (1) significantly predict AFS, (2) significantly differ across sites, or (3) are deemed to be clinically important. If a baseline characteristic (or set of characteristics) explains the difference (in that inclusion of the characteristic(s) in the proportional hazards regression model causes the site effect to become non-significant, $p > 0.10$), a summary Bayesian analysis of AFS adjusting for the characteristic(s) will be conducted. If baseline characteristics do not explain the difference, results will be

presented by site and in various groupings of sites as appropriate (e.g., perhaps removing any markedly different sites).

The summary Bayesian analysis will be a Bayesian piecewise exponential survival regression that adjusts for the baseline characteristic(s) described above. Summaries of the posterior distribution of the 180-day AFS rate (e.g., posterior mean/median, 95% credible interval, posterior probability that $\phi > 0.54$) will be presented.

In these analyses, small sites (contributing < 5 subjects) will be combined into one or more larger pseudo-sites of size ≥ 5 but ≤ 10. An analysis omitting small sites will also be presented. Descriptive statistics will be presented for all sites.

Heterogeneity analyses across the following factors (in addition to sites) will be conducted in a similar manner:

- US/OUS
- Sex (Male/Female)
- Dialysis status (Yes/No)
- Age (≤ 70, > 70)
- Diabetes (Type I, Type II, None)
- Race/Ethicity

9.6 Exploratory Multivariate Analysis

Potential contributing factors for AFS failures will be investigated as follows:

Using frequentist Cox proportional hazards regression models, the following variables will be investigated in univariate models:

- Sex (Male/Female)
- Dialysis status (Yes/No)
- Age (≤ 70, > 70)
- Diabetes (Type I, Type II, None)
- Race/Ethnicity
- Rutherford Classification
- Wound Area (continuous measure)

Any variable that is significantly related to AFS (at the 0.10 level) will be entered into a Cox multiple regression model. Variables no longer significant at the 0.05 level will be removed via backward elimination. The final model will be presented.

10 APPENDIX 1 – PERFORMANCE GOAL DERIVATION

A literature search was performed by the Yale Clinical Research Group (YCRG) according to a pre-specified protocol, and the search methods and results have been documented in a Literature Search Report ("Literature Search Report for the LimFlow System" Version 1.0 dated 19 December 2017). The literature search identified 36 publications meeting inclusion criteria that reported amputation-free survival (AFS), the composite of all-cause death or major (above-the-knee) amputation, all-cause mortality, major amputation, or wound healing in no-option CLTI subjects (patient with CLTI Rutherford score > 4 that are inoperable) who were receiving the standard of care (medication) or placebo.

Of the 36 publications, 27 publications reported the 6- or 12-month rate of amputation-free survival (or the composite of all-cause mortality or major amputation, which was converted into amputation-free survival) and contributed to the meta-analysis. Of these, 23 studies reported AFS at 6 months, and 19 reported AFS at 12 months.

In a review of the studies used for developing the performance goal, some noted that the studies that used Fontaine IV for the classification of critical limb ischemia had the highest AFS rates. Fontaine IV is a clinical assessment only, and though it is associated with Rutherford 5 and 6, it is less specific (see **Figure**).

The specificity of the Rutherford classification assures that the population enrolled in the studies will have the level of disease expected in the LimFlow clinical trial. In studies that used Fontaine, the lack of specificity and objective criteria can lead to a broader population being included that might be considered Rutherford 4 not included in the current study.

Figure 2: Comparison of Rutherford and Fontaine

Fontaine Grade	Symptoms	Grade	Category	Clinical Description	Objective Criteria
Stage I	Asymptomatic, incomplete blood vessel obstruction	0	0	Asymtomatic – no hemodinamically significant occlusive disease	Normal treadmill* or reactive hyperthermia test
			1	Mild claudication	
Stage II	Mild claudication pain in limb	I	2	Moderate claudication	Resting AP<60mmHg, ankle or metatarsal PVR flat or barely pulsatile; TP < 40mmHg
			3	Severe claudication	
Stage III	Rest pain, mostly in the feet	II	4	Ischemic rest pain	Complete treadmill exercise; AP after exercise >50 mmHg ≥20mmHg lower than resting value
Stage IV	Necrosis and/or gangrene of the limb	III	5	Minor tissue loss – nonhealing ulcer; focal gangrene with diffuse pedal ischemia	Resting AP<60mmHg, ankle or metatarsal PVR flat or barely pulsatile; TP < 40mmHg
			6	Major tissue loss – extending transmetatarsally; functional foot no longer salvageable	

Studies that only used the Fontaine classification were compared to those that only used the Rutherford classification using logistic regression. The SAS code was:

```
proc logistic data=limlos_adj;
class study status class;
freq num;
model status(event='no event')= class;
run;
```

where status is whether there was a death or major amputation event at 6 months or not, the class was the CLTI classification system used (R or F), and num is the variable that contains the number of subjects with or without an event.

The results of the logistic regression are shown in **Table 6**. The classification system effect has a low p-value (p <0.0001) and an odds ratio (Rutherford 5 or 6 versus Fontaine IV) of 2.891.

Table 6: Results of Logistic regression predicting AFS using CLTI classification method

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	0.5020	0.0766	42.9900	<.0001
class	**FIV**	1	0.5308	0.0766	48.0546	<.0001

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
class FIV vs R56	2.891	2.141	3.903

The 6-month AFS rates in studies shown in **Table 7** (reprint from Yale Clinical Research Group) that used the Fontaine classification were adjusted in two ways: a) the average of the percentage of Rutherford 5 or 6 in the other studies as reported in b) the odds ratio. The adjusted numbers were very similar between the two methods so for this document, the odds ratio adjusted numbers will be used. After the odds ratio adjustments,

the final AFS rate point estimate was 42.0% with an upper 95% confidence limit of 51.2%. However, in acknowledgement of FDAs concern around the performance goal and the previous discussions since 2017, LimFlow is not proposing to change the previously communicated performance goal based on the prior Yale work where the AFS rate point estimate was 43.7%, with an upper 95% confidence limit of 53.5%. Since the hypothesis test for this study is a test for superiority, the performance goal is set at the upper 95% confidence limit (UCL) of 53.5%, rounded up to 54%.

Table 7: Reprint of from YCRG:
Meta-analysis of 6-month AFS rate in no-option patients with Rutherford category 5 or 6 CLI

Study	N	Event-free survivors (n)	Observed AFS Rate	Included Rutherford Categories	Observed Proportion R4	Observed Proportion R 5/6	Imputed Proportion R 5/6	Adjusted AFS Rate
Brass et al 2006[3]	177	146	82.5%	4, 5, 6	NR	NR	66.9%	59.3%
Teraa et al. 2015[4]	79	66	83.5%	3, 4, 5, 6	31.6%	63.3%	NA	58.3%
Dubsky et al. 2013[5]	22	10	45.5%	4, 5, 6	NR	NR	66.9%	32.7%
Iafrati et al. 2016[6]	34	22	64.7%	5	0.0%	100.0%	NA	64.7%
Anghel et al. 2011[11]	14	3	21.4%	4,5	50.0%	50.0%	NA	13.5%
Li et al. 2013[12]	29	23	79.3%	4, 5, 6	NR	NR	66.9%	57.0%
Benoit et al. 2011[13]	14	9	64.3%	4,5	50.0%	50.0%	NA	40.4%
Gupta et al. 2013[14]	10	8	80.0%	4, 5, 6	20.0%	80.0%	NA	64.7%
Szabo et al 2013[17]	10	4	40.0%	4, 5, 6	NR	NR	66.9%	28.8%
Belch et al. 2011[18]	259	196	75.7%	4, 5, 6	NR	NR	66.9%	54.4%
Losordo et al. 2012[19]	12	8	66.7%	4,5	41.7%	58.3%	NA	44.7%
Nikol et al. 2008[20]	56	34	60.7%	4, 5, 6	NR	NR	66.9%	43.7%
Powell et al. 2012[21]	24	17	70.8%	4, 5, 6	NR	NR	66.9%	50.9%
Idei et al. 2011[22]	30	3	10.0%	4, 5, 6	27.0%	73.0%	NA	7.6%
Pignon et al. 2017[16]	19	14	73.7%	4,5	35.0%	65.0%	NA	52.1%
Wang et al. 2018[24]	36	28	77.8%	4,5	66.7%	33.3%	NA	43.5%
Faglia et al. 2010[25]	27	3	11.1%	4,5,6	37.0%	63.0%	NA	7.7%
Dalla Paola et al. 2019[26]	84	50	59.5%	4,5,6	NR	NR	66.9%	42.8%
Dubsky et al. 2019[27]	44	31	70.5%	4,5,6	NR	NR	66.9%	50.7%
Faglia et al. 2012[28]	12	3	25.0%	5.6	0.0%	100.0%	NA	25.0%
		Simple Average	58.1%				**Simple Average**	42.1%
		Weighted Average	68.3%				**Weighted Average**	49.2%
		Meta-Analytic Average	**58.6%**				**Meta-Analytic Average**	**42.0%**
		Min	10.0%				**Min**	7.6%
		Max	83.5%				**Max**	64.7%
		SD	24.1%				**SD**	17.7%

Meta-analytic estimate of the historical rate of AFS at 6 months in no-option R5/R6 CLTI patients is 42.0% (95% CI, 32.8-51.2).

11 APPENDIX 2 — BAYESIAN MULTIPLE IMPUTATION METHODOLOGY

Multiple imputation inference involves three distinct phases[5]:

1. The missing data are filled in M times to generate M complete datasets.
2. The M complete datasets are analyzed by using standard procedures
3. The results of the M complete datasets are combined for inference.

This is accomplished in this study as follows:

1. The missing data are filled in M=10,000 times via the predictive model described in Section 6.5. The multiple imputation literature sometimes calls this the "imputer's model," but in our Bayesian framework, this is simply the posterior predictive distribution of unobserved data, after integrating out the parameters of the piecewise exponential model described in Section 6.5. This results in M=10,000 completed datasets.

2. The M complete datasets consist of binary outcomes (Event, No Event at 180 days) and are analyzed using the "analyst's model," which in this case is the conjugate beta / binomial model. Letting X_{obs} denote the set of observed binary outcomes and X^*_m denote the m^{th} set of imputed binary outcomes, the posterior probability distribution $f(\phi \mid X_{obs}, X^*_m)$ is a beta distribution. For each of the m completed datasets (X_{obs}, X^*_m), the posterior probability $\text{Prob}(\phi > 0.54 \mid X_{obs}, X^*_m)$ is easily computed from the c.d.f. of the corresponding beta posterior distribution.

3. Combining results from these 10,000 completed datasets is accomplished by integrating over the posterior predictive distribution of the imputed values. Numerically, it is as simple as computing the average of the M=10,000 posterior probabilities from Step 2. In particular, since

$$f(\phi \mid X_{obs}) = \int f(\phi \mid X^*_m, X_{obs}) \cdot f(X^*_m \mid X_{obs})\, dX^*_m \ ,$$

by Monte Carlo integration,

$$f(\phi \mid X_{obs}) = \int f(\phi \mid X^*_m, X_{obs}) \cdot f(X^*_m \mid X_{obs})\, dX^*_m \approx 1/M \sum f(\phi \mid X^*_m, X_{obs}),$$

where the M realizations of X^*_m come from the posterior predictive distribution of $f(X^*_m \mid X_{obs})$. This is a linear combination of the M beta posterior densities. It follows that the c.d.f. $F(\phi \mid X_{obs}) \approx 1/M \sum F(\phi \mid X^*_m, X_{obs})$, and thus the posterior probability

$$P(\phi > 0.54 \mid X_{obs}) \approx 1/M \sum (1 - F(0.54 \mid X^*_m, X_{obs})),$$

which is the average of the M=10,000 posterior probabilities from Step 2.

A quantile q of the posterior distribution $f(\phi \mid X_{obs})$ can be computed by setting $q = 1/M \sum F(\phi \mid X^*_m, X_{obs})$ and numerically solving for ϕ using a root-finding algorithm. In this manner, a 95% Bayesian credible interval is computed from the 2.5^{th} and 97.5^{th} percentiles. A posterior median can be computed when q is set to 0.50. A posterior mean is computed as the average of the means of each of the 10,000 posterior beta distributions.

[5] Molenbeerghs G & Kenward MG (2007), *Missing Data in Clinical Studies*, Wiley & Sons, page 106.