# Statistical Analysis Plan

| | |
|---|---|
| **Clinical Investigation Plan Title** | SPYRAL PIVOTAL – SPYRAL HTN-OFF MED |
| **Clinical Investigation Plan Version** | 10.0 |
| **Document Version** | 3.8 |
| **Document Date** | 06-JAN-2021 |
| **NCT #** | NCT02439749 |

# Table of Contents

# 1. Version History

| Version | Summary of Changes | Author(s)/Title |
|---|---|---|
| 1.0 | • Initial OFF MED Feasibility Study | Martin Fahy, Senior Principal Statistician |
| 2.0 | • OFF MED Pivotal Study | Martin Fahy, Senior Principal Statistician |
| 2.1 | • Update Interim Analysis Section | Martin Fahy, Senior Principal Statistician |
| 2.2 | • Update interim analysis stopping rules to stop only for expected success only if both efficacy endpoints meet the efficacy stopping criteria. Update the primary safety endpoint analysis section. | Martin Fahy, Senior Principal Statistician |
| 3.0 | • Changes to Bayesian design removing test for expected success.<br>• Changes to discount function formulation | Martin Fahy, Senior Principal Statistician |
| 3.1 | • Per protocol population updated to exclude subjects who did not receive their randomized treatment<br>• Primary endpoint treatment effect descriptions updated to be consistent that negative effect favors RDN<br>• Stochastic comparison in section 9.9.4.1 updated<br>• Prior distribution and hyper-parameter definitions in section 9.9.4.1 updated<br>• Section 9.4 updated to include an analysis to impute missing outcome data<br>• Subgroup analyses in section 9.11 updated<br>• Appendices IIa, IIb, V and VI updated to include unscheduled visits corresponding to repeat ABPM visits | Martin Fahy, Senior Principal Statistician |
| 3.2 | • Updated language to reflect "approximately 210" subjects included in the first interim analysis due to variable attrition rate<br>• Changed "We reject $H_0$ if the probability is greater than or equal to 97.5%" to "We reject $H_0$ if the probability is greater than 97.5%" in section 9.9.3 | Martin Fahy, Senior Principal Statistician |

| | | |
|---|---|---|
| 3.3 | • Updated hyperparameter value for $b_\beta$ to be 1e10, changed from 1e5 in section 9.9.4.1 | Martin Fahy, Senior Principal Statistician |
| 3.4 | • As-Treated Population analysis included in 9.1.3.4<br>• Section 9.9.4.1 updated to include definitions for $y_i$ and $x_i$ in model.<br>• Remove 5% attrition sentence from primary safety endpoint analysis in section 9.9.1.1.<br>• Clarify that the primary safety analysis will be performed using " The first 253 subjects with evaluable safety data from the sources in Table 2 will be used to perform the primary safety endpoint analysis." in section 9.9.1.1. | Martin Fahy, Senior Principal Statistician |
| 3.5 | • Added "mean centered" to section 9.9.4.1. for the definition of the $x_i$ model parameter | Martin Fahy, Senior Principal Statistician |
| **3.6** | • COVID-19 analyses section 9.12 added<br>• Medication index analyses | Martin Fahy, Senior Principal Statistician |
| **3.7** | • Appendices describing the algorithms for selecting office BP, 24-hour BP, lab values and drug testing values removed from SAP, and will only be included in study table specificaitons.<br>• Section 9.12 updated to include COVID-19 related protocol deviation tables | Martin Fahy, Senior Principal Statistician |
| **3.8** | • Validation requirements added in section 10.<br>• SAP template version added to footer | Martin Fahy, Senior Principal Statistician |

# 2. List of Abbreviations and Definitions of Terms

| Abbreviation | Definition |
|---|---|
| ABPM | Ambulatory Blood Pressure Monitoring |
| AE | Adverse Event |
| ANCOVA | Analysis of Covariance |
| BP | Blood Pressure |
| CIP | Clinical Investigation Plan |
| DBP | Diastolic Blood Pressure |
| DSMB | Data Safety Monitoring Board |
| eGFR | estimated Glomerular Filtration Rate |
| MAE | Major Adverse Events |
| OBP | Office Blood Pressure |
| SAP | Statistical Analysis Plan |
| SBP | Systolic Blood Pressure |
| UCL | Upper Confidence Limit |

# 3. Introduction

This document outlines the detailed statistical methods to be implemented for the data collected within the scope of the Medtronic Vascular SPYRAL PIVOTAL – SPYRAL HTN-OFF MED Study: Global Clinical Study of Renal Denervation with the Symplicity Spyral™ Multi-Electrode Renal Denervation System in Patients with Uncontrolled Hypertension in the Absence of Antihypertensive Medications. The purpose of this study is to obtain an assessment of the efficacy and safety of renal denervation in the absence of antihypertensive medications. Specifically, the SAP has the following purpose: to prospectively outline the types of analyses and presentations of data that will form the basis for conclusions to be reached that will answer the trial objectives outlined in the protocol, and to explain in detail how the data will be handled and analyzed, adhering to commonly accepted standards and practices of biostatistical analysis

in the medical device industry. Results obtained from the analyses outlined in this document will be the basis of the Clinical Study Report for this trial.

# 4. Study Objectives

The objective of this study is to evaluate safety and blood pressure response after renal denervation in patients with uncontrolled hypertension compared to a sham-controlled population, in the absence of antihypertensive medications. In this study, "Uncontrolled hypertension" means an office systolic blood pressure (SBP) ≥150 mmHg and <180 mmHg, an office DBP ≥90 mmHg and a 24-hour Ambulatory Blood Pressure Monitoring (ABPM) average SBP ≥140 mmHg to <170 mmHg measured at Screening Visit 2. Data obtained without the confounding presence of antihypertensive medications will be used to confirm the effect of renal denervation on elevated blood pressure.

# 5. Study Endpoints

## 5.1. Primary Endpoints

There are two primary endpoints in this study (one safety and one efficacy). The study will be considered successful if both the primary safety and efficacy endpoint hypotheses are met.

**Powered Primary Safety Endpoint**

Incidence of Major Adverse Events (MAE), defined as a composite of the following events, through one-month post-randomization (6 months for new renal artery stenosis)

- All-cause mortality
- End Stage Renal Disease (ESRD)
- Significant embolic event resulting in end-organ damage
- Renal artery perforation requiring intervention
- Renal artery dissection requiring intervention
- Vascular complications
- Hospitalization for hypertensive crisis not related to confirmed non-adherence with medications or the protocol
- New renal artery stenosis > 70%, confirmed by angiography and as determined by the angiographic core laboratory

**Powered Primary Efficacy Endpoint**

- Baseline adjusted change (using Analysis of Covariance) in systolic blood pressure (SBP) from baseline (Screening Visit 2) to 3 months post-procedure as measured by 24-hour Ambulatory Blood Pressure Monitoring (ABPM).

## 5.2. Secondary Endpoints

**Powered Secondary Efficacy Endpoint**

- Baseline adjusted change (using Analysis of Covariance) in office systolic blood pressure from baseline (Screening Visit 2) to 3 months post-procedure.

**Secondary Safety Endpoints**

Acute/Procedural Safety Secondary Endpoints – Compared Between Groups at 1 Month Post-Procedure:

- Significant embolic event resulting in end-organ damage
- Renal artery perforation requiring intervention
- Renal artery dissection requiring intervention
- Vascular complications
- End-Stage Renal Disease
- ≥40% decline in eGFR
- New Myocardial Infarction
- New Stroke
- Renal artery re-intervention
- Major bleeding according to TIMI definition (i.e. intracranial hemorrhage, ≥5g/dl decrease in hemoglobin concentration, a ≥15% absolute decrease in hematocrit, or death due to bleeding within 7 days of the procedure)
- Increase in serum creatinine > 50% from Screening Visit 2
- New renal artery stenosis > 70%, confirmed by angiography and as determined by the angiographic core laboratory
- Hospitalization for hypertensive crisis not related to confirmed non-adherence with medications or the protocol

Chronic Safety Secondary Endpoints – Compared Between Groups at 3, 6, 12, 24, and 36 Months Post-Randomization:

- Composite Safety Endpoint, defined as a composite of the following events:
    - All-cause mortality
    - End-Stage Renal Disease (ESRD)
    - Significant embolic event resulting in end-organ damage
    - Renal artery perforation requiring intervention
    - Renal artery dissection requiring intervention
    - Vascular complications
    - Hospitalization for hypertensive crisis not related to confirmed non-adherence with medications and/or the protocol.
    - New renal artery stenosis > 70%, confirmed by angiography and as determined by the angiographic core laboratory
- All-cause mortality
- End-Stage Renal Disease
- ≥40% decline in eGFR
- New Myocardial Infarction
- New Stroke

- Renal artery re-intervention
- Major bleeding according to TIMI definition (i.e. intracranial hemorrhage, ≥5g/dl decrease in hemoglobin concentration, a ≥15% absolute decrease in hematocrit, or death due to bleeding within 7 days of the procedure)
- Increase in serum creatinine > 50% from Screening Visit 2
- New Renal artery stenosis >70%, confirmed by angiography and as determined by the angiographic core laboratory
- Hospitalization for hypertensive crisis not related to confirmed non-adherence with medications or the protocol

**Secondary Efficacy Endpoints**

- Change in systolic blood pressure (SBP) from baseline (Screening Visit 2) as measured by 24-hour Ambulatory Blood Pressure Monitoring (ABPM) at 3, 6, 12, 24 and 36 months post-procedure.
- Change in office systolic blood pressure from baseline (Screening Visit 2) at 1, 3, 6, 12, 24 and 36 months post-procedure.
- Incidence of achieving target office systolic blood pressure (SBP <140 mmHg) at 1, 3, 6, 12, 24 and 36 months post-procedure.
- Change in office diastolic blood pressure from baseline (Screening Visit 2) at 1, 3, 6, 12, 24 and 36 months post-procedure.
- Change in diastolic blood pressure from baseline (Screening Visit 2) as measured by 24-hour Ambulatory Blood Pressure Monitoring (ABPM) at 3, 6, 12, 24 and 36 months post-procedure.

Summary of Quality of Life (QOL) Measures (EQ5D and SF36)

## 5.3. Additional analyses

The following additional analyses will be conducted:

- Antihypertensive medication usage throughout the study, including escape subjects prior to 3-months and subjects reintroduced to medications after 3 months.
- Additional procedural characteristics e.g. treatment duration, frequency of distal renal artery treatment, ablations per vessel, location of ablations, number of ablations per patient and other characteristics will be analyzed to assess their impact on blood pressure.

# 6. Investigation Plan

The SPYRAL PIVOTAL – SPYRAL HTN-OFF MED study is a multi-center, international, prospective, single blinded, randomized, interventional, sham-controlled study. In order to demonstrate that catheter-based renal denervation using the Symplicity Spyral catheter and the Symplicity G3 generator is an effective and safe treatment for hypertension in the absence of antihypertensive medications, study subjects will be randomized to the Denervation or Control group in a 1:1 fashion. Subjects will be studied in the absence of antihypertensive medications to assess the impact of renal denervation on systolic blood pressure in the absence of medication. Patients with hypertension will be enrolled in accordance with the inclusion

and exclusion criteria specified in the protocol. Approximately 1800 subjects will be screened in order to randomize up to 433 subjects, which includes 80 subjects to be used as an informative prior in a Bayesian analysis. Subjects will be enrolled at up to 50 study centers in the United States, Canada, Japan, Australia and countries where CE mark applies. Additional geographies may be added. Enrollment is expected to take approximately 33 months. Subjects may participate in the study from the time of signing consent until completion of three years of follow-up after procedure. Control subjects may be offered renal denervation therapy (crossover) after their 6 month follow-up visit and will be followed-up for two years.

# 7. Randomization and Blinding

Randomization will be stratified by study center at a 1:1 ratio to:

- Denervation group (RDN): Subjects remain blinded and are treated with the renal denervation procedure.
- Control group: Subjects remain blinded and remain on the catheterization lab table for at least 20 minutes prior to introducer sheath removal.

Investigational sites will access randomization allocation via a password-protected system that can only be accessed by those approved by the study sponsor.

All study staff and necessary hospital personnel will be instructed that subjects are not to be informed of their randomization assignments and appropriate measures should be taken to minimize the risk of premature unblinding.

The Investigator performing the catheterization lab procedures and his/her designated study staff will be blinded to a subject's randomization group up until the angiography is completed and inclusion/exclusion confirmed. However, investigators performing study follow-up visits and the subject's referring/managing physicians will not be proactively informed of a subject's treatment assignment to minimize potential bias in the subject's care decisions. However, to specifically minimize potential bias in the measurement of Office BP and ABPM, each investigational site will specify several designated "blinded" members of their study staff that will not be informed of the subject's group assignments and will be responsible for performing the office blood pressure measurements, conducting ABPM preparation and printing results upon a patient completing the ABPM. Prior to unblinding, the blinding effectiveness will be assessed by asking blinded study staff which group they believe the subject was randomized to.

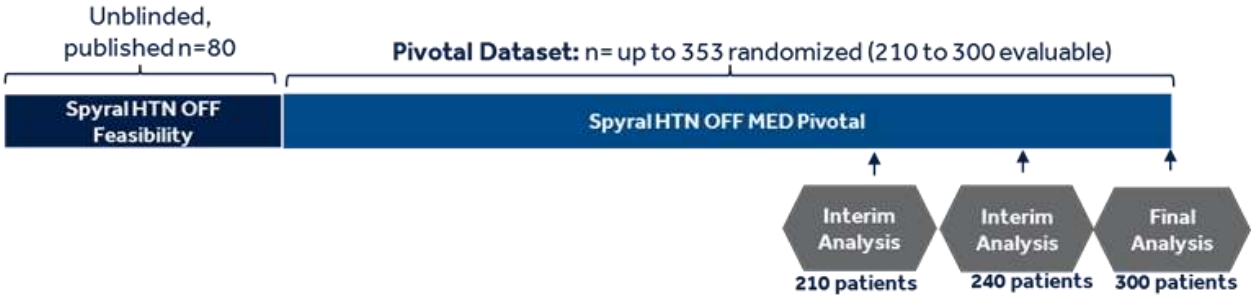
Subjects will be blinded during the renal angiogram by a combination of conscious sedation, sensory isolation (e.g., blindfold and music), and lack of familiarity to the procedural details and duration (i.e., subjects will not know the difference between the renal angiography procedure alone and the renal angiography and denervation procedure). Subjects will continue to be blinded by only interacting with blinded site personnel through the 6 month follow-up visit post-procedure. Blinding effectiveness will be assessed by asking the subject which group they believe they were randomized to. All subjects will be unblinded after the completion of their required 6-month follow-up testing.

# 8. Determination of Sample Size

This study will be conducted as an adaptive Bayesian trial with an informative prior. A Bayesian power prior approach [3,4] in conjunction with a discount function will be used to incorporate the prior data. The discount function reduces the strength of the prior data if disagreements are observed with the pivotal data.

The prior data consists of the first consecutively randomized 80 patients in the SPYRAL HTN-OFF MED study, the results from these 80 subjects have been analyzed and published [6]. The weight of the prior data will be adjusted using a discount function, which scales from 0 to 1, according to the similarity of the prior and pivotal data. This discount function adjusts the amount of weight the prior receives. This prevents the use of an informative prior where exchangeability issues are present (e.g., the prior and pivotal data are quite different). This discount function approach was proposed by the Medical Device Innovative Consortium (MDIC) working group and is a collaborative effort between FDA and industry through the MDIC [1,2]. If the analyses show a high level of agreement for pivotal data compared to the prior, the prior will be weighted at or near 100%. If the pivotal data perform worse than or much better than the prior, then the prior will receive very little or zero weight. The Bayesian adaptive design is set up to enroll patients until a sufficient sample size is achieved to have high probability of meeting the endpoint.

The sample size of the study will vary from 210 to 300 subjects with 3-month follow-up due to the adaptive nature of the trial. This will require approximately 247 to 353 randomized subjects to account for a 15% attrition at 3-months. Based on this attrition rate, the interim analyses will take place after the 210$^{th}$ and 240$^{th}$ subjects have completed 3-month follow-up, with a maximum study size of 300 subjects if the study does not stop at either interim look. The actual rate of attrition will dictate the available sample size for analyses. At each interim analysis, enrollment may be stopped for efficacy or futility.




Simulations of trial design operating characteristics performed to demonstrate control of type I error & power are presented in section 9.9.6 for the primary and secondary efficacy endpoints.

# 9. Statistical Methods

## 9.1. Study Subjects

### 9.1.1. Disposition of Subjects

A subject disposition table will be provided for each follow-up visit containing the following information:

- The number of subjects who died or withdrew prior to each follow-up
- The number of subjects eligible for each follow-up visit
- The number of subjects completing each follow-up visit within the protocol specified window
- The number of subjects completing each follow-up outside the protocol specified window
- The number of subjects who did not complete their follow-up

### 9.1.2. Clinical Investigation Plan (CIP) Deviations

A study deviation is an event where the investigator or investigational site personnel did not conduct the clinical study according to the Clinical Investigation Plan or Clinical Study Agreement. The investigator is not allowed to deviate from the above mentioned documents except with prior approval and under emergency circumstances. All deviations shall be documented and explained, regardless the reason for the deviation. Medtronic will assess the significance of all deviations and evaluate the need to amend the Clinical Investigation Plan or to early terminate the investigation, in accordance with Medtronic SOPs.

### 9.1.3. Analysis Sets

#### 9.1.3.1. Intention-To-Treat (ITT) Population

This is the primary analysis set for efficacy and safety evaluation and consists of all randomized subjects, analyzed according to their randomized treatment. Subjects who meet the anti-hypertensive medication escape criteria (OSBP>180 mmHg or safety reasons) will be analyzed using Last Observation Carried Forward (LOCF) for their blood pressure measurements out to 3 months. Safety outcomes, and office and ambulatory blood pressure outcomes at each follow-up visit will be presented for this population.

#### 9.1.3.2. Modified Intention-To-Treat (ITT) Population

All randomized subjects, analyzed according to their randomized treatment. Subjects who meet the anti-hypertensive medication escape criteria (OSBP>180 mmHg or safety reasons) will be excluded from this population. Office and ambulatory blood pressure outcomes out to 3 months will be presented for this population.

#### 9.1.3.3. Per Protocol Population

All randomized subjects, meeting the following criteria:

1. Subjects showing medication compliance in blood and/or urine (via drug testing data) at screening visit 2 (SV2) and 3-months.
2. Exclude subjects with protocol deviation code 101 (consent not obtained).
3. Exclude subjects who do not meet the following Inclusion/Exclusion criteria.

- Inclusion: Individual has an office systolic blood pressure (SBP) ≥150 mmHg and <180 mmHg and an office DBP ≥ 90 mmHg measured at Screening Visit 2, according to the guidelines in Appendix L7 of the study protocol.
- Inclusion: Individual has a 24-hour ABPM average SBP ≥140 and <170 mmHg measured at Screening Visit 2, according to guidelines in Appendix L7 of the study protocol.
- Exclusion: Individual has undergone prior renal denervation.
- Exclusion: Individual has renal artery anatomy that is ineligible for treatment.
- Exclusion: Individual has one or more of the following conditions: stable or unstable angina within 3 months of enrollment, myocardial infarction within 3 months of enrollment; heart failure cerebrovascular accident or transient ischemic attack or atrial fibrillation at any time.

4. Exclude subjects meeting the Anti-Hypertensive Medication Escape Criteria (OSBP >180 mmHg or safety reasons).
5. Exclude subjects who did not receive the treatment they were randomized to.

Office and ambulatory blood pressure outcomes out to 3 months will be presented for this population.

#### 9.1.3.4. As Treated Population

All randomized subjects, analyzed according to the actual treatment received. Subjects randomized to RDN who do not get treated will be analyzed in the control arm. Subjects who meet the anti-hypertensive medication escape criteria (OSBP>180 mmHg or safety reasons) will be analyzed using Last Observation Carried Forward (LOCF) for their blood pressure measurements out to 3 months. Office and ambulatory blood pressure outcomes out to 3 months will be presented for this population.

### 9.1.4. Crossover Procedures

Control subjects may crossover and receive renal denervation therapy at or after their 6 month follow-up visit. For the subjects who had already completed their 6 month visit, the decision to crossover must take place at their next in person visit. All subjects will have 30 days from the Crossover visit (6, 12, 24, 36 month follow-up or Unscheduled visit) to undergo the crossover procedure. Subjects who are more than 30 days from their 6, 12, 24 or 36 month, must complete a crossover Baseline visit prior to crossing over. Control subjects who are coming in for their scheduled 6, 12, 24 and 36 month follow-up visit do not need to come in for a crossover Baseline visit if all the required crossover Baseline procedures occurred, including lipid panel, hs-crp and uric acid blood tests. To crossover, the subject cannot meet any of the anatomical and glomerular filtration rate (eGFR) exclusion criteria, and cannot be pregnant, nursing or planning to become pregnant during the course of the study follow-up. Crossover subjects will undergo follow-up visits at 1, 3, and 6 months (± 14 days) and annually at 12 and 24 months (± 30 days) post-procedure.

## 9.2. General Methodology

Descriptive statistics of continuous outcomes will be presented by treatment group and include sample size, mean, median, standard deviation, minimum and maximum. For categorical outcomes, the number and percentage of subjects in each category will be presented by treatment group. Statistical

comparisons between treatment groups will be made using the independent samples t-test for continuous outcomes and Fisher's exact test for categorical outcomes. Paired tests will be used to compare changes from baseline to follow-up within each treatment group. All statistical analyses will be performed using SAS for Windows (version 9.2 or higher) or other widely accepted statistical or graphical software. Patient data listings and tabular and graphical presentations of results will be provided. Unless otherwise specified, a two-sided 0.05 level of significance will be used to declare treatment groups significantly different.

## 9.3. Poolability Analyses

### 9.3.1. Poolability of study centers

The following analyses will be performed to evaluate the poolability of data from different study centers. If the resulting tests are significant at the 0.15 level, further exploratory analysis will be attempted to identify covariates that may explain differences. Otherwise, the data will be considered to be poolable across study centers.

- A logistic regression will be conducted, with Major Adverse Events (MAE) as the dependent variable, and treatment, study center, treatment * study centers as independent variables. If the interaction term is not significant at 0.15 level, then the treatment effect in the primary safety endpoint is considered consistent among the sites.
- A linear regression will be conducted, with change in systolic blood pressure from baseline to 3 months as measured by 24-hour systolic Ambulatory Blood Pressure Monitoring (ABPM) as the dependent variable, and baseline systolic ABPM, treatment, study center, treatment * study center as independent variables. If the interaction term is not significant at 0.15 level, then the treatment effect in the primary effectiveness endpoint is considered consistent among the study centers.

### 9.3.2. Poolability of US and Canada

The following analyses will be performed to evaluate the poolability of US and Canadian sites. If the resulting tests are significant at the 0.15 level, further exploratory analysis will be attempted to identify covariates that may explain differences. Otherwise, the data will be considered to be poolable between these regions.

- A logistic regression will be conducted, with MAE as the dependent variable, and treatment, US/Canada, treatment * US/Canada as independent variables. If the interaction term is not significant at 0.15 level, then the treatment effect in the primary safety endpoint is considered consistent between US and Canadian regions.
- A linear regression will be conducted, with change in systolic blood pressure from baseline to 3 months as measured by 24-hour systolic Ambulatory Blood Pressure Monitoring (ABPM) as the dependent variable, and baseline systolic ABPM, treatment, US/Canada , treatment * US/Canada as independent variables. If the interaction term is not significant at 0.15 level, then the treatment effect in the primary effectiveness endpoint is considered consistent between US and Canadian regions.

### 9.3.3. Poolability of North America (US and Canada)/Rest of World (ROW)

The following analyses will be performed to evaluate the poolability of data from North America (NA) and Rest of World (ROW) sites. If the resulting tests are significant at the 0.15 level, further exploratory analysis will be attempted to identify covariates that may explain differences. Otherwise, the data will be considered to be poolable across regions.

- A logistic regression will be conducted, with MAE as the dependent variable, and treatment, NA/ROW, treatment * NA/ROW as independent variables. If the interaction term is not significant at 0.15 level, then the treatment effect in the primary safety endpoint is considered consistent between NA and ROW subgroups.
- A linear regression will be conducted, with change in systolic blood pressure from baseline to 3 months as measured by 24-hour systolic Ambulatory Blood Pressure Monitoring (ABPM) as the dependent variable, and baseline systolic ABPM, treatment, NA/ROW, treatment * NA/ROW as independent variables. If the interaction term is not significant at 0.15 level, then the treatment effect in the primary effectiveness endpoint is considered consistent between NA and ROW subgroups.

## 9.4. Handling of Missing Data and Dropouts

Every effort will be made to minimize missing data for the primary and secondary powered efficacy endpoints. A secondary analysis will be performed for both efficacy endpoints where missing outcome data are imputed using SAS PROC MI. Missing 3-month outcomes will be imputed using baseline SBP, treatment group, age, gender and BMI. One hundred imputed datasets will be generated, and a pooled estimate of the treatment effect will be generated using SAS PROC MIANALYZE.

## 9.5. Adjustments for Multiple Comparisons

The primary safety and efficacy endpoints are independently powered and no adjustments for multiple comparisons will be made.

## 9.6. Demographic and Other Baseline Characteristics

Baseline variables will be tabulated. Categorical variables, including binary variables, will be reported by giving the number and percentage of patients in each category. Continuous variables will be reported by presenting the number of values, mean, standard deviation, median, minimum, and maximum value for each. No imputation will be performed for missing data unless otherwise stated.

## 9.7. Treatment Characteristics

Renal denervation treatment measures such as number of ablation attempts and number of generator codes will be summarized separately for each kidney, and for combined kidneys. Anti-hypertensive medication use will also be summarized at baseline and at each follow-up.

## 9.8. Interim Analyses

Interim analyses will be conducted and reviewed by the DSMB, along with an independent organization that will be performing the Bayesian analyses. Medtronic personnel will not have access to any unblinded results prior to the primary endpoint analyses. The interim analyses will take place at N=210 and N=240 subjects with 3-month follow-up, with a maximum study size of N=300 subjects if the study does not stop at either interim look. At each interim analysis, enrollment may be stopped for efficacy or expected futility. Both the primary and secondary effectiveness endpoints will be evaluated during these interim looks and enrollment will only stop at an interim analysis if both endpoints meet the following stopping criteria.

1. The first interim analysis takes place when the first 210 subjects have 3-month follow-up data available (requiring approximately 247 randomized subjects to account for attrition). The Bayesian efficacy analysis will be performed and P[suc] will be calculated, where P[suc] is the probability of accepting the alternative efficacy endpoint hypotheses, $P\left(\mu < 0 \middle| \boldsymbol{y}, \boldsymbol{y_0}, \widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)\right)$, and is defined in detail in section 9.9.3.
    a. If P[suc] > 0.975 for both the primary and secondary efficacy endpoints, then the study has met the efficacy hypotheses and enrollment will be stopped. Any additional subjects that have been enrolled before the decision is made to stop for efficacy will be pooled with the existing subjects and analyzed as a secondary cohort.
    b. We calculate the probability of futility based on the maximum study size of 300 subjects which requires us to impute the outcomes for subjects who have not yet been enrolled (see sections 9.8.2 and 9.8.3 below for details). If the posterior probability of futility from this calculation is < 0.05 for both the primary and secondary efficacy endpoints, then the study will have met the futility boundary and enrollment will be stopped. Any additional subjects that have been enrolled before the decision is made to stop for futility will be pooled with the existing subjects and analyzed as a secondary cohort.
    c. If the stopping rules in a and b above are not met for either the primary or secondary efficacy endpoints, then we continue enrolling subjects to the second interim analysis.
2. If we don't stop for efficacy or futility at the first interim analysis then enrollment will continue until the second interim analysis when 3-month follow-up data is available for the first 240 subjects (requiring approximately 282 randomized subjects to account for attrition). The Bayesian efficacy analysis will be performed and P[suc] will be calculated.
    a. If P[suc] > 0.975 for both the primary and secondary efficacy endpoints, then the study has met the efficacy hypotheses and enrollment will be stopped. Any additional subjects that have been enrolled before the decision is made to stop for efficacy will be pooled with the existing subjects and analyzed as a secondary cohort.
    b. We calculate the probability of futility based on the maximum study size of 300 subjects which requires us to impute the outcomes for subjects who have not yet been enrolled (see sections 9.8.2 and 9.8.3 below for details). If the posterior probability of futility from this calculation is < 0.05 for both the primary and secondary efficacy endpoints, then the study will have met the futility boundary and enrollment will be stopped. Any additional

subjects that have been enrolled before the decision is made to stop for futility will be pooled with the existing subjects and analyzed as a secondary cohort.

c. If the stopping rules in a and b above are not met for either the primary or secondary efficacy endpoints, then we continue enrolling subjects to the final analysis.

3. If we don't stop for efficacy or futility at the second interim analysis, then enrollment will continue until the maximum study size of 300 subjects with 3-month follow-up data (requiring approximately 353 randomized subjects to account for attrition). The Bayesian efficacy analysis will be performed and P[suc] will be calculated.
   a. If P[suc] > 0.975 for the primary efficacy endpoint, then we have met the primary efficacy hypothesis.
   b. If P[suc] > 0.975 for the secondary efficacy endpoint, then we have met the secondary efficacy hypothesis.

### 9.8.1. Mathematical forms for success:

$$P\left(\mu < 0 \middle| \boldsymbol{y}, \boldsymbol{y_0}, \widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)\right) > 0.975$$

where $\boldsymbol{y}$ and $\boldsymbol{y_0}$ represent the pivotal data and prior data, respectively, the notation $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ is used to denote that the estimate of the discounting parameter $\widehat{\alpha_0}$, which depends on the pivotal data, prior data, and the Weibull shape and scale parameters, and $\mu = \mu_t - \mu_c$ represents the baseline-adjusted treatment effect of BP change comparing RDN and control groups. See section 9.9.3 for more details.

### 9.8.2. Mathematical forms for futility:

$$\frac{\sum_1^{N_{rep}} I\left(P\left(\mu < 0 \middle| \boldsymbol{y_{imp}}\right) > 0.975\right)}{N_{rep}} < 0.05$$

Where $\boldsymbol{y_{imp}}$ is a single completed dataset after imputation, $N_{rep}$ is the number of imputation simulations done and $\mu = \mu_t - \mu_c$ represents the baseline-adjusted treatment effect of BP change comparing RDN and control groups where $\mu_t$ and $\mu_c$ are the baseline adjusted BP changes in the RDN and control groups respectively.

### 9.8.3. Futility imputation procedure

For the futility calculation, we calculate the probability of futility based on the maximum study size of 300 subjects which means we have three types of subjects to consider:

1. Subjects who have reached their 3-month endpoint.
2. Subjects who are enrolled and have baseline blood pressure but have not reached their 3-month endpoint.
3. Subjects who have not been enrolled.

Imputation procedures will be used to impute the blood pressure change for subjects of type 2 and type 3 above.

For type 2 subjects we use the following procedure:

2.1 Construct the posterior predictive distribution for blood pressure change using type 1 subjects, incorporating the prior data.
2.2 Simulate samples from the predictive distribution to impute the missing values of blood pressure change for type 2 subjects, conditional on observed baseline blood pressure.

For type 3 subjects, we do the following:

3.1 Construct the posterior predictive distribution for blood pressure change using type 1 subjects, including the prior data.
3.2 Simulate baseline blood pressure for type 3 patients.
3.3 Simulate blood pressure change from the predictive distribution.
3.4 Impute blood pressure change for type 3 subjects, conditional on simulated baseline blood pressure.

We then combine the type 1 subjects with the imputed subjects from steps 2.2 and 3.4 into a single dataset and construct the endpoint using the BayesDP function. This procedure will be repeated for many datasets ($N_{rep} > 1000$) and we calculate the number of times the alternative hypothesis is accepted. If the proportion of times the alternative hypothesis is accepted is less than 5%, stop enrollment due to futility.

The bayesDP package, version 1.3.2, is available on the Comprehensive R Archive Network (CRAN) [https://CRAN.R-project.org/].

## 9.9. Evaluation of Objectives

The study will be considered successful if we meet both the primary safety and efficacy hypotheses.

### 9.9.1. Primary Safety Endpoint

The primary safety endpoint of the study is the incidence of Major Adverse Events (MAE), defined as a composite of the following events through one month post-randomization (6 months for new renal artery stenosis):

- All-cause mortality
- End-stage Renal Disease (ESRD)
- Significant embolic event resulting in end-organ damage
- Renal artery perforation requiring intervention
- Renal artery dissection requiring intervention
- Vascular complications
- Hospitalization for hypertensive crisis not related to confirmed non-adherence with medications or the protocol
- New renal artery stenosis >70%

The primary safety analysis will be performed using the ITT population defined in 9.1.3.1.

#### 9.9.1.1. Primary Safety Endpoint Analysis

Medtronic is using a performance goal approach to power the primary safety endpoint.

The safety performance goal for the Major Adverse Event (MAE) rate was developed based on review of and comparison to event rates of other renal interventions. The review of renal intervention literature reported event rates of 3.6 to 17.2%. The reported events differed among the studies; however, for a subset of these studies, we could estimate rates for a composite of events similar to our protocol's MAE composite (see Table 1 below). The major adverse event rate from these studies was 7.1%.

**Table 1: MAE Rates of Literature Reported Studies**

| | MAE Rate |
|---|---|
| **ROCHA[1]** | 4.8% |
| **ASTRAL[2]** | 10.1% |
| **Bax[3]** | 17.2% |
| **Van Jaarsveld[4]** | 3.6% |
| **Laird[5]** | 8.0% |
| **Coral[6]** | 5.1% |
| **Bradaric[7]** | 5.7% |
| **Jaff[8]** | 6.9% |
| **Bersin[9]** | 9.8% |
| **Overall Average (weighted by study size)** | **7.1%** |

[1] Rocha-Singh K, Jaff MR, Rosenfield K, ASPIRE-2 Trial Investigators. Evaluation of the safety and effectiveness of renal artery stenting after unsuccessful balloon angioplasty: the ASPIRE-2 study. J Am Coll Cardiol. 2005 Sep 6;46(5):776-83.

[2] Wheatley K, Ives N, Gray R, et al. Revascularization versus medical therapy for renal-artery stenosis. N Engl J Med 2009; 361:1953–1962.

[3] Bax L, Woittiez AJ, Kouwenberg HJ, et al. Stent placement in patients with atherosclerotic renal artery stenosis and impaired renal function: a randomized trial. Ann Intern Med 2009; 150:840–841.

[4] Van Jaarsveld, et al. The effect of balloon angioplasty on hypertension in atherosclerotic reanl-artery stenosis. N Engl J Med 2000; 342: 1007-14.

[5] Laird, et al. Safety and efficacy of renal artery stenting following suboptimal renal angioplasty for de novo and restenotic ostial lesions: results from a nonrandomized, prospective multicenter registry. J Vasc Interv Radiol 2010; 21: 627-637.

[6] Cooper CJ, Murphy TP, Cutlip DE, et al. Stenting and Medical Therapy for Atherosclerotic Renal-Artery Stenosis. NEJM 2014; 370(1): 13-22.

[7] Bradaric, C. et al. Drug-eluting stents versus bare-metal stents for the prevention of restenosis in patients with renovascular disease.

[8] Jaff MR, Bates M, Sullivan T, et al. Significant reduction in systolic blood pressure following renal artery stenting in patients with uncontrolled hypertension: results from the HERCULES trial. Catheter Cardiovasc Interv. 2012 Sep 1;80(3):343-50.

[9] Bersin RM, Ansel G, Rizzo A, et al. Nine-Month Results of the REFORM Study: A Prospective, Single-Arm, Multicenter Clinical Study of the Safety and Effectiveness of the Formula Balloon-Expandable Stent for Treatment of Renal Artery Stenosis. Catheterization and Cardiovascular Interventions 82:266–273 (2013).

The performance goal is set to be 7.1%, which is the meta-analysis rate from historical trials in Table 1.

The primary safety null and alternative hypotheses are:

$H_0$: $\pi \geq 7.1\%$ vs.

$H_a$: $\pi < 7.1\%$

where $\pi$ is the MAE rate for patients undergoing renal denervation. Under the assumption that the true rate is 3.5%, and using a one-sided 0.05 level of significance, an evaluable sample size of 253 renal denervation patients yields 80% power to reject the null hypothesis in favor of the alternative. The exact binomial test was used for the sample size calculation for the primary safety endpoint hypothesis.

In other words, the primary safety endpoint hypotheses is designed to show whether the true MAE rate is lower than 7.1%. Compared to the literature reported event rates for renal intervention, we believe that these thresholds are appropriate for demonstrating safety of the device given the expected performance rates of similar renal intervention trials, particularly when balanced with the expected blood pressure reductions.

Medtronic proposes multiple sources of study patients as shown in Table 2 below to ensure 253 patients treated with the Symplicity Spyral catheter (including branch treatment) are available for analysis. The first 253 subjects with evaluable safety data from the sources in Table 2 will be used to perform the primary safety endpoint analysis.

With a sample size of 253 and a one-sided significance level of 0.05, a maximum of 11 subjects with MAE will enable us to meet the safety primary endpoint, resulting in an event rate of 4.3% with a one-sided 95% upper confidence bound of 7.09% using the exact binomial method.

**Table 2: Study Sources of Patients for Primary Safety Endpoint Data**

| **Study** |
|---|
| SPYRAL HTN-OFF MED prior data (First 80 Subjects)<br>Randomized 1:1 to RDN:CONTROL |
| SPYRAL PIVOTAL – SPYRAL HTN-OFF MED<br>Randomized 1:1 to RDN:CONTROL |
| SPYRAL HTN-ON MED First 106<br>Randomized 1:1 to RDN:CONTROL |
| SPYRAL HTN ON MED Extension<br>Randomized 2:1 to RDN:CONTROL |
| SPYRAL HTN-OFF MED Crossovers (from prior and pivotal) |
| SPYRAL HTN-ON MED Crossovers (from first 106 and Extension subjects) |

### 9.9.2. Secondary Safety Objectives

The following secondary safety endpoints will be assessed:

- Acute/Procedural Safety Secondary Endpoints – Compared Between Groups at 1 Month Post-Procedure:

- Significant embolic event resulting in end-organ damage
- Renal artery perforation requiring intervention
- Renal artery dissection requiring intervention
- Vascular complications
- End-Stage Renal Disease
- ≥40% decline in eGFR
- New myocardial infarction
- New stroke
- Renal artery re-intervention
- Major bleeding according to TIMI definition (i.e. intracranial hemorrhage, ≥5g/dl decrease in hemoglobin concentration, a ≥15% absolute decrease in hematocrit, or death due to bleeding within 7 days of the procedure)
- Increase in serum creatinine >50% from Screening Visit 2
- New renal artery stenosis >70%, confirmed by angiography and as determined by the angiographic core laboratory
- Hospitalization for hypertensive crisis not related to confirmed non-adherence with medications or the protocol

- Chronic Safety Secondary Endpoints – Compared Between Groups at 3, 6, 12, 24, and 36 Months Post- Randomization:
    - Composite Safety Endpoint, defined as a composite of the following events:
        - All-cause mortality
        - End-Stage Renal Disease (ESRD)
        - Significant embolic event resulting in end-organ damage
        - Renal artery perforation requiring intervention
        - Renal artery dissection requiring intervention
        - Vascular complications
        - Hospitalization for hypertensive crisis not related to confirmed non-adherence with medications and/or the protocol
        - New renal artery stenosis >70%, confirmed by angiography and as determined by the angiographic core laboratory
    - ≥ 40% decline in eGFR
    - Increase in serum creatinine >50% from Screening Visit 2
    - New myocardial infarction
    - New stroke
    - Renal artery re-intervention
    - Major bleeding according to TIMI definition (i.e. intracranial hemorrhage, ≥5g/dl decrease in hemoglobin concentration, a ≥15% absolute decrease in hematocrit, or death due to bleeding within 7 days of the procedure)
    - Hospitalization for hypertensive crisis not related to confirmed non-adherence with medication and/or protocol.
    - Summary of health related Quality of Life (HRQoL) analysis based on reporting measures using accepted QoL instruments (EQ5D and SF36).

All the safety endpoints will be adjudicated by the Clinical Events Committee (CEC). The following algorithm will be used to evaluate the safety event rates: The denominator will include all subjects who either had a CEC adjudicated event prior to the time of interest (180 days for 6 months events, for example), or had last contact date that is beyond the lower window of the follow up (166 days for 6 month events, for example). The numerator will include all subjects who had CEC adjudicated events up to the time of interest (180 days for 6 months events, for example).

The secondary safety endpoints, out to 6-months follow-up, will be compared between treatment groups using Fisher's exact test. Two-sided 95% confidence intervals of the difference between treatment groups will also be presented.

After 6-months follow-up, control subjects may crossover (undergo renal denervation), and secondary safety endpoints will be summarized by group (RDN, Crossovers, Non-Crossovers). RDN vs. Crossover vs. Non-Crossover groups will be compared out to 36 months post-procedure for the RDN and Non-Crossover groups and out to 24 months for the Crossover group using chi-square tests for categorical data and ANOVA for continuous data.

The secondary safety analyses will be performed using the ITT population defined in section 9.1.3.1.

#### 9.9.2.1. Renal Artery Stenosis Evaluation at 12 Months

With an expected rate of 3.1% for renal artery stenosis at 12 months [7], a sample size of 100 subjects will provide a 95% confidence interval of approximately (0.6%, 8.5%) using the exact method (calculated using an event rate of 3/100=3%).

Descriptive statistics of this endpoint at 12 months will be provided using counts, percentages and the 95% confidence interval.

### 9.9.3. Primary Efficacy Endpoint

The primary efficacy endpoint of the study is the baseline adjusted (analysis of covariance/ANCOVA) change in SBP from baseline (screening visit 2) to 3-months post-procedure as measured by 24-hour Ambulatory Blood Pressure Monitoring (ABPM).

In the context of an ANCOVA linear regression model, $\mu = \mu_t - \mu_c$ represents the baseline-adjusted treatment effect of BP change comparing RDN and control groups where $\mu_t$ and $\mu_c$ are the baseline adjusted BP changes in the RDN and control groups respectively. Let $\boldsymbol{y} = \{\boldsymbol{y_t}, \boldsymbol{y_c}\}$ and $\boldsymbol{y_0} = \{\boldsymbol{y_{0t}}, \boldsymbol{y_{0c}}\}$ represent the pivotal data and prior data respectively, where $t =$ RDN and $c =$ control group. Let the hypotheses for the study be the following:

$H_0: \mu = 0$

$H_a: \mu < 0$

We reject $H_0$ if the probability is greater than 97.5%, i.e.

$$P\left(\mu < 0 \middle| \boldsymbol{y}, \boldsymbol{y_0}, \widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)\right) > 0.975$$

where the notation $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ is used to denote that the estimate of $\widehat{\alpha_0}$ depends on the pivotal data, prior data, and the Weibull shape and scale parameters. In conjunction with a pre-specified decision rule controlling the prior data weight, the estimate of $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ represents a measure of similarity between pivotal and prior data. Alternatively, in the absence of $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$, i.e., $P(\mu < 0 | \boldsymbol{y}, \boldsymbol{y_0})$, full weight would be given to the prior data.

### 9.9.4. Primary Efficacy Endpoint Analysis

The power prior discount function approach is used to estimate $\mu$, and determine $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$, the strength of the prior data used to estimate $\mu$. $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ ranges from 0 to 1, where 1 means that 100% of the prior data is used and 0 means that no prior data is used. Before beginning the study, an initial value is chosen for $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$, call this value $\alpha_{max}$. This $\alpha_{max}$ value is the maximum strength the prior data can receive. We intend to use the same enrollment criteria for the prior and pivotal studies, and therefore believe that a value of $\alpha_{max} = 1$ is appropriate.

At interim looks and at the final analysis, we analyze the data using the power prior discount function method, this method will discount $\alpha_{max}$ to an appropriate value $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ where $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k) \leq \alpha_{max}$. This discounting is based on the discount function which is discussed in detail in the next section.

Under the adaptive procedure, if the pivotal data diverges from the prior data at an interim look, the discount function will discount the strength of the prior data, thus requiring continued enrollment to maintain power to achieve the endpoint. Alternatively, if the prior and pivotal data agree, there will be a smaller penalty from the discount function, thus fewer prospective patients would be needed to maintain power, and enrollment may stop early.

The ITT population defined in section 9.1.3.1. will be used as the primary analysis population for this endpoint. Secondary effectiveness analyses will also be performed using the modified ITT, per-protocol and as treated populations defined in sections 9.1.3.2., 9.1.3.3. and 9.1.3.4.

#### 9.9.4.1. Discount Function Estimation Method

The power prior discount function method is comprised of four steps: **Compare, Discount, Combine, and Estimate.**

**Compare:**

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

For each treatment group, we separately fit the model to the combined prior and pivotal data:

$$y_i = \tilde{\beta}_0 + \tilde{\beta}_1 I(i \in \text{prior}) + \tilde{\beta}_2 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \tau^2),$$

where $I(i \in \text{prior}) = 1$ if the subject is from the prior dataset, and 0 otherwise, $y_i$ is the BP change for the $i$th observation and $x_i$ is the mean centered baseline BP for the $i$th observation. With flat priors on each parameter, we estimate the posterior probability that $\tilde{\beta}_1 > 0$ by first computing, using Monte Carlo sampling

$$p^* = P[\tilde{\beta}_1 > 0 \mid \boldsymbol{y}, \boldsymbol{y_0}].$$

Having calculated this separately for both the RDN ($p_t^*$) and control groups ($p_c^*$), they are transformed to $p_t$ and $p_c$ using

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

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

**Discount:**

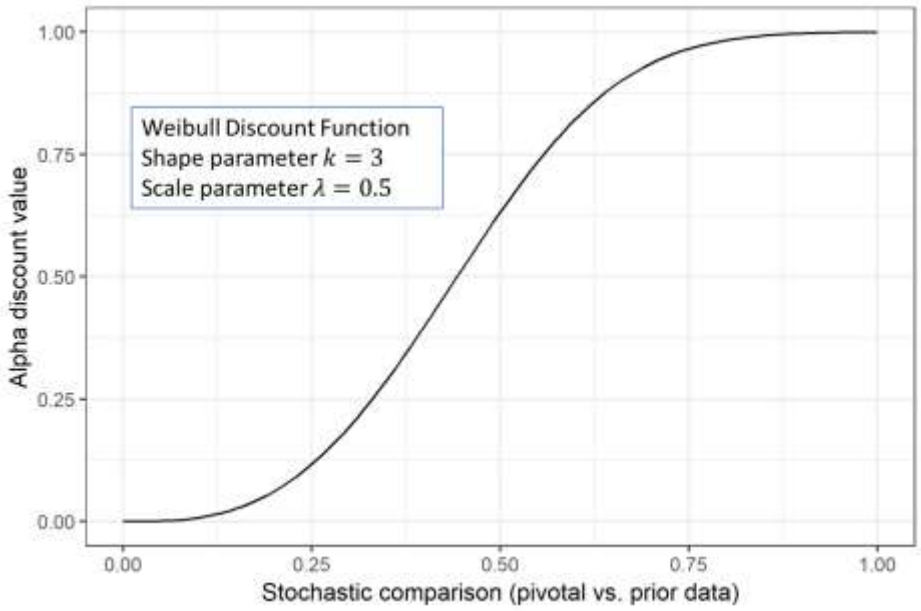
We discount $\alpha_{max}$ based on the value of $p_t$ and $p_c$ from the **Compare** step and the discount function $F(p)$,

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

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

$$F(p) = 1 - e^{-\left(\frac{p}{\lambda}\right)^k}$$

For this study, we will be using a shape parameter of $k = 3$ and a scale parameter of $\lambda = 0.5$ (illustrated below). Note that we will use the same Weibull function parameters for both RDN and control groups, but $p_t$ and $p_c$ will have different values from the **Compare** step.




**Combine:**

Using the power prior method and $\widehat{\alpha_0}$ we can combine the prior and pivotal data together using Bayesian techniques to construct the posterior distribution for $\mu$ as follows. We first begin with a hierarchical linear regression model of the form:

$$\begin{aligned}
y_i \mid \mu_t, \mu_c, \beta, \sigma^2 &\sim N(\mu_t I(i \in t) + \mu_c I(i \in c) + x_i\beta,\ \sigma^2),\ i = 1, \ldots, n, \\
\mu_t \mid \mu_{0t}, \sigma_{0t}^2, \widehat{\alpha_{0t}} &\sim N(\mu_{0t},\ \sigma_{0t}^2/\widehat{\alpha_{0t}}), \\
\mu_c \mid \mu_{0c}, \sigma_{0c}^2, \widehat{\alpha_{0c}} &\sim N(\mu_{0c},\ \sigma_{0c}^2/\widehat{\alpha_{0c}}), \\
\beta &\sim N\left(a_\beta,\ b_\beta^2\right), \\
\pi(\sigma^2) &\propto \sigma^{-2},
\end{aligned}$$

where $N(\cdot)$ denotes a normal distribution, $I(\cdot)$ is the indicator function where $I(i \in t)$ indicates that observation $i$ is in the RDN group and $I(i \in c)$ indicates observation $i$ is in the control group, $y_i$ is the BP change for the $i$th observation, $x_i$ is the mean centered baseline BP value for the $i$th observation, and

each of $\mu_{0t}$, $\mu_{0c}$, $\sigma^2_{0t}$, $\sigma^2_{0c}$ are hyperparameters estimated from the historical data, and $a_\beta = 0$ and $b_\beta = 10^{10}$.[10]

To carry out estimation in a computationally efficient manner, we rely on a reparameterization derived in Gelman [4]. First, construction vectors and matrices: $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{x}_t$ and $\boldsymbol{x}_c$ vectors of binary treatment control indicators taking values 0 and 1, $\boldsymbol{x}_\beta$ the vector of baseline values, $\boldsymbol{X} = \left(\boldsymbol{x}_t \mid \boldsymbol{x}_c \mid \boldsymbol{x}_\beta\right)$ the $n \times 3$ design matrix, $\boldsymbol{\theta} = (\mu_t, \mu_c, \beta)^T$ parameters of interest to be estimation. Then, we can write the model as

$$\boldsymbol{y}_* \sim N(\boldsymbol{X}_*\boldsymbol{\theta}, \boldsymbol{\Sigma}_*),$$

where $\boldsymbol{y}_* = \left(\boldsymbol{y}^T, \mu_{0t}, \mu_{0c}, a_\beta\right)^T$, $\boldsymbol{X}_* = (\boldsymbol{X}^T, \boldsymbol{I}_3)^T$, $\boldsymbol{I}_3$ is a $3 \times 3$ identity matrix, and

$$\boldsymbol{\Sigma}_* = \begin{pmatrix} \sigma^2 \boldsymbol{I}_n & 0 \\ 0 & \boldsymbol{\Sigma}_\theta \end{pmatrix},$$

where

$$\boldsymbol{\Sigma}_\theta = \begin{pmatrix} \sigma^2_{0t}/\widehat{\alpha_{0t}} & 0 & 0 \\ 0 & \sigma^2_{0c}/\widehat{\alpha_{0c}} & 0 \\ 0 & 0 & b^2_\beta \end{pmatrix}.$$

The posterior mean of $\boldsymbol{\theta}$ is found via least squares as

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}_*^T \boldsymbol{\Sigma}_*^{-1} \boldsymbol{X}_*)^{-1} \boldsymbol{X}_*^T \boldsymbol{\Sigma}_*^{-1} \boldsymbol{y}_*,$$

and the posterior variance of $\boldsymbol{\theta}$ is

$$\boldsymbol{V}_\theta = (\boldsymbol{X}_*^T \boldsymbol{\Sigma}_*^{-1} \boldsymbol{X}_*)^{-1}.$$

Thus, the posterior distribution of $\boldsymbol{\theta}$ is $\boldsymbol{\theta} \mid \boldsymbol{y}_*, \boldsymbol{\Sigma}_* \sim N\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{V}_\theta\right)$. Both $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{V}_\theta$ are composed of an unknown $\sigma^2$. The marginal posterior distribution of $\sigma^2$ is

$$q(\sigma^2 \mid \boldsymbol{y}) \propto \pi(\sigma^2) |\sigma^2 \boldsymbol{I}_n|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}_* - \boldsymbol{X}_*\widehat{\boldsymbol{\theta}}\right)^T \boldsymbol{\Sigma}_*^{-1} \left(\boldsymbol{y}_* - \boldsymbol{X}_*\widehat{\boldsymbol{\theta}}\right)\right\},$$

which does not have a known distributional form. Here, $\pi(\sigma^2)$ is the prior distribution of $\sigma^2$. Thus, to draw samples from the posterior distribution of $\sigma^2$, we rely on a grid search to sample values of $\sigma^2$.

We proceed by drawing samples from $\sigma^2 \sim q(\sigma^2|\boldsymbol{y})$. Using these estimates, we input them into the posterior distribution of $\boldsymbol{\theta}$. By repeating this process a large number of times, we will be drawing posterior samples from $\boldsymbol{\theta}$ which will account for the uncertainty in both $\sigma^2$ as well as appropriately weighting the prior data based on $\widehat{\alpha_{0t}}$ and $\widehat{\alpha_{0c}}$.

We can than construct the following contrast of interest from the drawn samples:

$$\mu = \mu_t - \mu_c.$$

[10] In the R package bayesDP version 1.3.2, the bdplm function parameter 'prior_covariate_sd' is described as "The prior standard deviation(s) of the covariate effect(s). Default value is 1e4." Upon inspection of the code, it appears the standard deviation is internally scaled by 1e6, thus yielding a more diffuse prior for $\beta$. Therefore, the actual value of $b_\beta$ used in the analysis is 1e10.

Use of this contrast leads to the univariate distribution of interest concerning the mean BP change difference between the test and control groups.

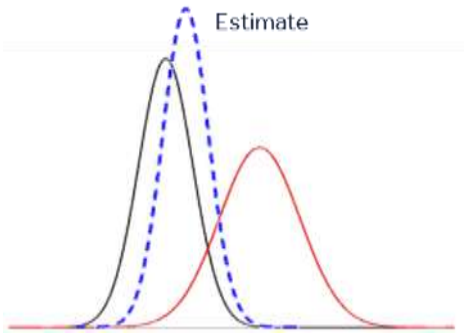
**Estimate:**

The posterior distribution from the combined prior and pivotal data is used to estimate the posterior probability

$$P(\mu < 0 | \boldsymbol{y}, \boldsymbol{y_0}, \widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)\,) \qquad (1)$$

where the notation $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ is used to denote that the estimate of $\widehat{\alpha_0}$ depends on the pivotal data, prior data, and the Weibull shape and scale parameters. In conjunction with a prespecified decision rule controlling the prior data weight, the estimate of $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ represents a measure of similarity between pivotal and prior data. Alternatively, in the absence of $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$, i.e., $P(\mu < 0 | \boldsymbol{y}, \boldsymbol{y_0})$, full weight would be given to the prior data.

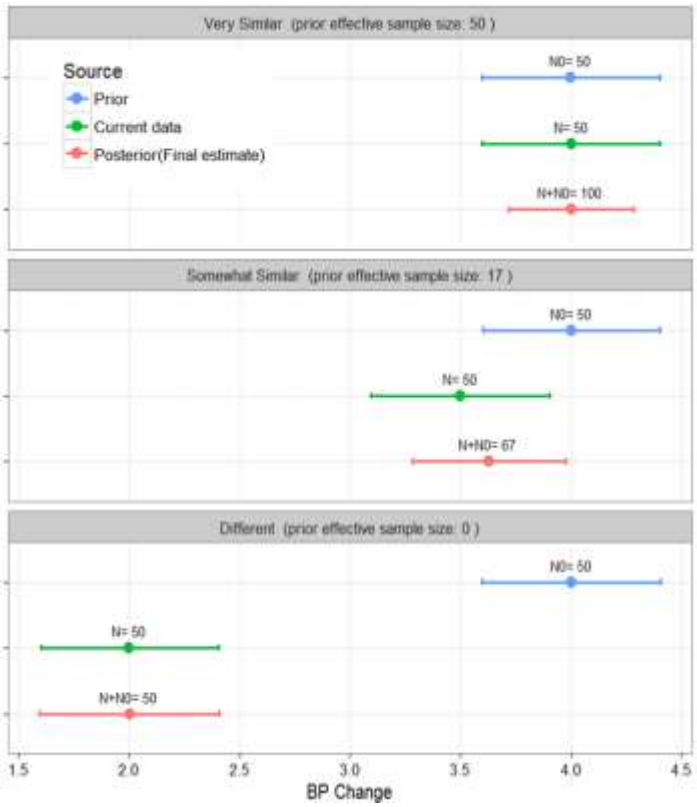
The blue dashed line in the figure below is an illustrative example of the estimate from pivotal data (black) and prior data (red).




The analysis in (1) is performed at all interim looks and at the final analysis.

#### 9.9.4.2. Illustration of Discount Function Scenarios

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




The panels in this figure can be interpreted as follows:

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

### 9.9.5. Secondary Efficacy Objectives

#### 9.9.5.1. Powered Secondary Efficacy Endpoint

The secondary powered efficacy endpoint is the baseline adjusted change in office SBP from baseline (screening visit 2) to 3-months post procedure, compared between treatment groups.

In the context of an ANCOVA linear regression model, $\mu = \mu_t - \mu_c$ represents the baseline-adjusted treatment effect of BP change comparing test and control groups where $\mu_t$ and $\mu_c$ are the baseline adjusted BP changes in the RDN and control groups respectively. Let $\boldsymbol{y} = \{\boldsymbol{y_t}, \boldsymbol{y_c}\}$ and $\boldsymbol{y_0} = \{\boldsymbol{y_{0t}}, \boldsymbol{y_{0c}}\}$ represent the pivotal data and prior data respectively, where the subscripts $t =$ RDN group and $c =$ control group. Let the hypotheses for the study be the following:

$H_0: \mu = 0$

$H_a: \mu < 0$

We reject $H_0$ if the probability is greater than 97.5%, i.e.

$$P\left(\mu < 0 \middle| y, y_0, \widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)\right) > 0.975$$

where the notation $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ is used to denote that the estimate of $\widehat{\alpha_0}$ depends on the pivotal data, prior data, and the Weibull shape and scale parameters. In conjunction with a pre-specified decision rule controlling the prior data weight, the estimate of $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$ represents a measure of similarity between pivotal and prior data. Alternatively, in the absence of $\widehat{\alpha_0}(\boldsymbol{y}, \boldsymbol{y_0}, \lambda, k)$, i.e., $P(\mu < 0 | \boldsymbol{y}, \boldsymbol{y_0})$, full weight would be given to the prior data.

The same statistical methods as outlined in section 9.9.4 will be used to analyze the powered secondary endpoint. The ITT population defined in section 9.1.3.1. will be used as the primary analysis population for this endpoint. Secondary effectiveness analyses will also be performed using the modified ITT, per-protocol and as treated populations defined in sections 9.1.3.2., 9.1.3.3. and 9.1.3.4.

### 9.9.6. Simulation of Primary and Secondary Efficacy Endpoint Operating Characteristics

As outlined in section 7.8 enrollment will be stopped at an interim analysis for efficacy or futility only if both the primary and secondary efficacy endpoints meet the stopping criteria. Simulations were performed to assess operating characteristics for the primary and secondary efficacy endpoints. We used 8000 trial simulations to estimate the power and 15000 simulations to estimate the type I error. Tables 3 and 4 summarize the assumptions that were made for the primary and secondary efficacy endpoint simulations, and table 5 summarizes the operating characteristics for the efficacy evaluation. The overall power for the efficacy evaluation from table 5 is 94%, with a one-sided type I error rate of 0.029 for 24-hour SBP and 0.026 for Office SBP.

**Table 3: Simulation Assumptions for Primary Efficacy Endpoint**

| | |
|---|---|
| Enrollment Rate | 10 Subjects / Month |
| Prior Baseline Adjusted RDN Group Mean/SE | -5.30 / 1.65 mmHg |
| Prior RDN Group N | 35 |
| Prior Baseline Adjusted Control Group Mean/SE | -0.74 / 1.62 mmHg |
| Prior Control Group N | 36 |
| Maximum Prior Patients | 35 RDN + 36 Control = 71 |
| Pivotal Study Expected Treatment Difference | -4.0 mmHg |
| Pivotal Study RDN Group Mean/SD | -4.74 / 12 mmHg |
| Pivotal Study Control Group Mean/SD | -0.74 / 12 mmHg |

| | |
|---|---|
| Weibull Discount Function Parameters | Shape: $k = 3$, Scale: $\lambda = 0.5$ |

**Table 4: Simulation Assumptions for Secondary Efficacy Endpoint**

| | |
|---|---|
| Enrollment Rate | 10 Subjects / Month |
| Prior Baseline Adjusted RDN Group Mean/SE | -9.69 / 2.20 mmHg |
| Prior RDN Group N | 37 |
| Prior Baseline Adjusted Control Group Mean/SE | -2.54 / 2.09 mmHg |
| Prior Control Group N | 41 |
| Maximum Prior Patients | 37 + 41 = 78 |
| Pivotal Study Expected Treatment Difference | -6.5 mmHg |
| Pivotal Study RDN Group Mean/SD | -9.04 / 16 mmHg |
| Pivotal Study Control Group Mean/SD | -2.54 / 16 mmHg |
| Weibull Discount Function Parameters | Shape: $k = 3$, Scale: $\lambda = 0.5$ |

**Table 5: Operating Characteristics for Primary and Secondary Efficacy Endpoints**

| | |
|---|---|
| Trial Success Rate (Power) | 94% |
| Type I Error (one-sided) | 24-Hr SBP: 0.029<br>Office SBP: 0.026 |
| First Interim Look N | N=210 |
| Power at First Interim Look | 83% |
| Second Interim Look N | N=240 |
| Power at Second Interim Look | 89% |
| Maximum Study Size | N=300 |
| % of Simulations that Stop for Futility | 0.05% |

### 9.9.7. Secondary Analysis of Primary and Secondary Efficacy Endpoints

According to section 7.8, the first and second interim analyses takes place when the first 210 and 240 subjects have 3-month follow-up data available respectively. At each interim analysis, the Bayesian efficacy analysis will be performed, and a decision can be made to stop the study for futility or efficacy.

If the study stops for efficacy or futility at either the first or second interim analysis, then any additional subjects that have been enrolled before the decision to stop has been made will not be part of the primary endpoint analysis, but instead will be pooled with the existing subjects and analyzed as a secondary cohort. The same analysis outlined in section 9.9.4 and 9.9.5 will be used on this pooled cohort.

### 9.9.8. Other Secondary Efficacy Endpoints

The following additional secondary efficacy endpoints will be assessed:

- Change in systolic blood pressure (SBP) from baseline (Screening Visit 2) as measured by 24-hour Ambulatory Blood Pressure Monitoring (ABPM) at 3, 6, 12, 24 and 36 months post-procedure.
- Change in office systolic blood pressure from baseline (Screening Visit 2) at 1, 3, 6, 12, 24 and 36 months post-procedure.
- Incidence of achieving target office systolic blood pressure (SBP <140 mmHg) at 1, 3, 6, 12, 24 and 36 months post-procedure.
- Change in office diastolic blood pressure from baseline (Screening Visit 2) at 1, 3, 6, 12, 24 and 36 months post-procedure.
- Change in diastolic blood pressure from baseline (Screening Visit 2) as measured by 24-hour Ambulatory Blood Pressure Monitoring (ABPM) at 3, 6, 12, 24 and 36 months post-procedure.

RDN vs. control groups will be compared out to 6-months post randomization, prior to the crossover procedure. Statistical comparisons will be performed using the independent samples t-test for continuous endpoints and Fisher's exact test for categorical endpoints. In addition, two-sided 95% confidence intervals of the difference between RDN and control groups will be presented. Changes in blood pressure measurements from baseline to follow-up within each treatment group will be assessed using paired t-tests. Two-sided 95% confidence intervals of the mean change from baseline will be presented for each treatment group. Analysis of Covariance (ANCOVA) models, adjusting the treatment effect for the baseline BP measurements will also be applied to all continuous secondary endpoints.

RDN vs. Crossover vs. Non-Crossover groups will be compared out to 36 months post-procedure for the RDN and Non-Crossover groups and out to 24 months for the Crossover group using chi-square tests for categorical data and ANOVA for continuous data.

Changes in blood pressure measurements from baseline to follow-up within each group will be assessed using paired *t*-tests. Two-sided 95% confidence intervals of the mean change from baseline will also be presented for each group.

The secondary efficacy analyses will be presented for all the study populations defined in section 9.1.3.

### 9.9.9. Additional Objectives

The following additional analyses will be conducted:

- Quality of Life (QOL) measures (EQ5D and SF36).
- Antihypertensive medication usage through 36 months.
- Additional procedural characteristics e.g. treatment duration, frequency of distal renal artery treatment, ablations per vessel, location of ablations, number of ablations per patient and other characteristics will be analyzed to assess their impact on blood pressure.
- Medication adherence will be assessed using results from drug testing. In addition, we will perform analyses to evaluate the effect of medication adherence on blood pressure change.
- Analyses looking at long term imaging will be performed.

RDN vs. control groups will be compared out to 6-months post randomization, prior to the crossover procedure. Statistical comparisons will be performed using the independent samples t-test for continuous endpoints and Fisher's exact test for categorical endpoints. In addition, two-sided 95% confidence intervals of the difference between RDN and control groups will be presented.

RDN vs. Crossover vs. Non-Crossover groups will be compared out to 36 months post-procedure for the RDN and Non-Crossover groups and out to 24 months for the Crossover group using chi-square tests for categorical data and ANOVA for continuous data.

Changes in blood pressure measurements from baseline to follow-up within each group will be assessed using paired t-tests. Two-sided 95% confidence intervals of the mean change from baseline will also be presented for each group.

These additional analyses will be presented for the ITT study population defined in section 9.1.3.1.

### 9.9.10. Medication Burden Analyses

Upon completion of a valid ABPM at the 3 month follow-up visit, subjects with an office SBP ≥ 140 mmHg will begin an antihypertensive medication regimen following protocol guidelines at the discretion of the study investigator. We will calculate the medication burden for each subject at 6, 12, 24 and 36-months using medication index scores that take into account the class, dose and frequency of each anti-hypertensive medication being taken.

- These scores will be compared between RDN and Control arms at the 6-month follow-up visit.
- At 12, 24 and 36 month follow-up visits, the scores will be compared between RDN, crossover and non-crossover arms.
- Multivariable predictors of medication index scores adjusting for treatment arm and BP measures will be performed.

## 9.10. Safety Evaluation

Adverse Event (AE) information will be collected by the site from subject enrollment (consent) through study termination. AEs will be followed until the event has resolved (in the case of permanent impairment, the event will be followed until it stabilizes, and the overall clinical outcome has been ascertained).

The Investigator will report any adverse events that may occur to the Sponsor, and will assess seriousness, relationship (to the device, procedure and renal denervation therapy where applicable), subsequent intervention required, resolution status and whether or not the adverse event resulted in the subject's discontinuation from the study. The Investigator will provide further information regarding adverse events as requested by the Sponsor.

## 9.11. Subgroup Analyses

Analysis will be carried out for the following subgroups to assess consistency of results.

- Gender
- Age at baseline <65 vs. ≥ 65 (years)
- BMI by tertiles ($kg/m^2$)
- Type 2 diabetics vs. non-diabetics
- Current smokers vs. non-smokers
- Baseline eGFR < 60 vs. ≥ 60 ($mL/min/1.73\ m^2$)
- Obstructive sleep apnea yes vs. no
- US vs. OUS Subjects
- US African American vs. US Non African American subjects
- European vs. Japanese vs. Australian subjects
- Baseline ambulatory SBP by tertiles and medians (mmHg)

- Baseline office SBP by tertiles and medians (mmHg)
- Baseline ambulatory heart rate by tertiles and medians (bpm)
- Baseline office heart rate by tertiles and medians (bpm)
- 24-Hour Pulse Pressure <60 vs. >= 60 mmHg (mmHg)
- Orthostatic Hypertension at baseline
- Orthostatic Tachycardia at baseline
- Plasma Renin Activity at baseline <0.65 vs. ≥0.65 (ng/mL/h)
- Aldosterone Renin Ratio at baseline by tertiles
- Aldosterone at baseline by tertiles (ng/dL)
- Tertile analysis by total number of ablations performed
- Tertile analysis by total number of ablations performed in branch vessels
- Tertile analysis by total number of ablations performed in main renal artery vessels
- Tertile analysis by total number of 45 second ablations performed
- Medication adherent vs. non-adherent subjects at screening visit 2 (SV2) and 3 months (from urine and serum tests)

### 9.12. COVID-19 Related Analyses

In accordance with FDA guidance document "FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic" [8], we will provide COVID-19 related protocol deviation tables and AE/SAE tables summarizing the site-reported adverse events attributed to COVID-19.

### 9.13. Changes to Planned Analysis

There are no changes to the planned analysis at this time.

## 10. Validation Requirements

Statistical programming for the analysis datasets, primary endpoints, secondary safety endpoints, and secondary effectiveness endpoints require Level 1 (independent) validation. Other objectives and sub-group analyses require Level 1 (independent) or Level 2 (Peer review) validation.

# 11. References

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

[2] T. Haddad, A. Himes, L. Thompson, T. Irony, R. Nair, " Incorporation of stochastic engineering models as prior information in Bayesian medical device trials", *Journal of Biopharmaceutical Statistics*, (2017) DOI: 10.1080/10543406.2017.1300907.

[3] J. G. Ibrahim and M.-H. Chen, "Power prior distributions for regression models," *Statistical Science,* (2000), 15(1) 46-60.

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

[5] Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P., Bayesian adaptive methods for clinical trials. CRC press, 2010.

[6] Townsend RR, Mahfoud F, Kandzari DE, Kario K, Pocock S, Weber MA, Ewen S, Tsioufis K, Tousoulis D, Sharp AS, Watkinson AF. Catheter-based renal denervation in patients with uncontrolled hypertension in the absence of antihypertensive medications (SPYRAL HTN-OFF MED): a randomised, sham-controlled, proof-of-concept trial. The Lancet. 2017 Aug 28.

[7] Sanders, Margreet F., et al. "Renal artery and parenchymal changes after renal denervation: assessment by magnetic resonance angiography." European radiology 27.9 (2017): 3934-3941

[8] FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic. Guidance for Industry, Investigators, and Institutional Review Boards. March 2020.

# 12. Statistical Appendices

## 12.1. Appendix I: Imputation of Missing Dates

### Imputation of Missing AE Onset Date

| Valid Portion | Missing Portion | Imputed Value for Missing Portion |
|---|---|---|
| Month, Year | Day | Set Day = first day of that month and year, then set the day = later of (New onset date, informed consent date). |
| Year | Day, Month | Set date = later of (January 1st of that year, informed consent date). |
| None | Day, Month, Year | Informed consent date. |

### Imputation of Missing Medication Start Date

| Valid Portion | Missing Portion | Imputed Value for missing Portion |
|---|---|---|
| Month, Year | Day | Set Day = first day of that month and year |
| Year | Day, Month | Set date = January 1st of that year |
| None | Day, Month, Year | SV2 date |

### Imputation of Missing Medication Stop Date

| Valid Portion | Missing Portion | Imputed Value for missing Portion |
|---|---|---|
| Month, Year | Day | Set Day = first day of that month and year, then set the day = later of (New date, SV2 date, start date). |
| Year | Day, Month | Set date = later of (January 1st of that year, SV2 date, start date). |