Evaluation of Implantable Tibial Neuromodulation Pivotal Study (TITAN 2)
Statistical Analysis Plan Version 4.0
09-Aug-2023
NCT05226286

Medtronic

Statistical Analysis Plan

Clinical Investigation Plan Title	Evaluation of Implantable Tibial Neuromodulation (TITAN 2) Pivotal Study
Clinical Investigation Plan Identifier	MDT20061
Clinical Investigation Plan Version	V4.0
Statistical Analysis Plan Version Date	09-AUG-2023
Sponsor/Local Sponsor	Medtronic, Inc. Neuromodulation 7000 Central Ave NE Minneapolis, MN 55432 USA

Table of Contents

1. Version History

Version	Summary of Changes	Author(s)/Title
1.0	• Not Applicable, New Document	[redacted] Distinguished Statistician

2. List of Abbreviations and Definitions of Terms

Abbreviation	Definition
ADE	Adverse Device Effect
AE	Adverse Event
BMI	Body Mass Index
CEC	Clinical Events Committee
CIP	Clinical Investigation Plan
CRF	Case Report Form
DD	Device Deficiency
eCRF	Electronic Case Report Form
FAS	Full Analysis Set
FCS	Fully Conditional Specification
FDA	Food and Drug Administration
HRQL	Health Related Quality of Life
ITNM	Implantable Tibial Neuromodulation
ITT	Intent-to-treat
IUSS	Indevus Urgency Severity Scale
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
[redacted]	[redacted]
OAB	Overactive Bladder
OAB-q	Overactive Bladder Quality of Life
PASS	Power Analysis & Sample Size (software)
[redacted]	[redacted]
PMA	Premarket Approval
PTNS	Percutaneous Tibial Nerve Stimulation
PTNM	Percutaneous Tibial Neuromodulation
QoL	Quality of Life
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SNM	Sacral Neuromodulation
TNM	Tibial Neuromodulation
UF	Urinary Frequency
UI	Urinary Incontinence
UPS	Urgency Perception Scale
USADE	Unanticipated Serious Adverse Device Effect
UU	Urinary Urgency
UUI	Urinary Urge Incontinence

3. Introduction

The medical condition of overactive bladder (OAB) is recognized as a symptom syndrome characterized by the symptoms of urinary urgency (UU), with or without urinary urge incontinence (UUI), usually with urgency frequency (UF) and nocturia (waking at night one or more times to void). Tibial neuromodulation (TNM) devices are designed to deliver electrical pulses that stimulate the tibial nerve that runs posterior to the tibia and extends to the sacral nerve plexus. These pulses are thought to relieve OAB symptoms by modulating communication between the bladder and central nervous system. In order to eliminate the barrier of frequent clinic visits for percutaneous tibial neuromodulation (PTNM) sessions while retaining the effectiveness and positive safety profile of tibial neuromodulation, implantable tibial neuromodulation (ITNM) devices have been proposed as an alternative to PTNM.

The Titan 2 pivotal study is intended to assess the safety and effectiveness of the Medtronic TNM system for the treatment of overactive bladder via a single arm, multicenter, prospective design. This statistical analysis plan (SAP) was created using the clinical investigation plan (CIP) version identified on the cover page and defines the analyses that will be included in the premarket approval (PMA) and final clinical study reports. Study progress reports will generally follow the methods of this SAP but will contain a small subset of planned analyses. Importantly, the progress reports will not include effectiveness results until after the PMA submission has been completed, as no interim assessments of effectiveness are planned for this study. Revisions to the SAP may be required if protocol changes impact the statistical analysis for the study, or if updates to the analyses are needed. In cases where the SAP is not updated, any changes to the planned analyses will be documented in the associated report. The study objectives are taken directly from the CIP.

4. Study Objectives

4.1 Primary Objective

To demonstrate that the percentage of subjects considered a UUI responder after 6 months of tibial neuromodulation exceeds a performance goal of 40%.

4.2 Primary Endpoint

Proportion of TNM subjects experiencing a reduction of 50% or more in daily urinary urge incontinence (UUI) episodes (UUI responder rate) at 6 months after device implant.

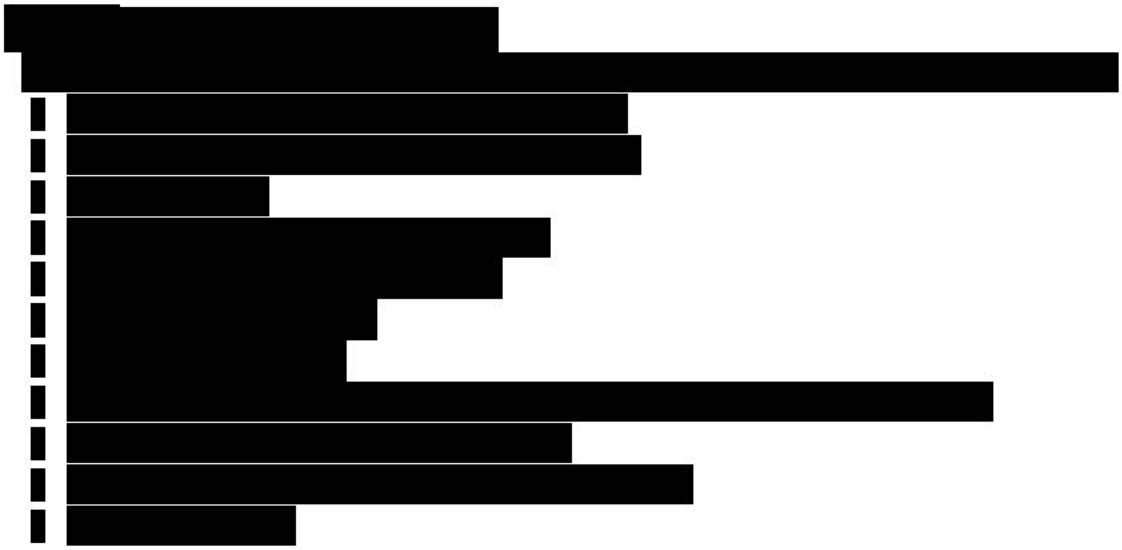
4.3 Secondary Objectives

- To demonstrate an improvement from baseline to 6 months in number of UUI episodes in subjects with UUI at baseline
- To demonstrate an improvement from baseline to 6 months in number of UF episodes in subjects with UF at baseline

- To demonstrate an improvement from baseline to 6 months in urinary urgency using the Urgency Perception Scale (UPS)
- To demonstrate an improvement from baseline to 6 months in quality of life (QoL) using the Overactive Bladder Symptom Quality of Life Questionnaire (OAB-q)

4.3.1 Secondary Endpoints

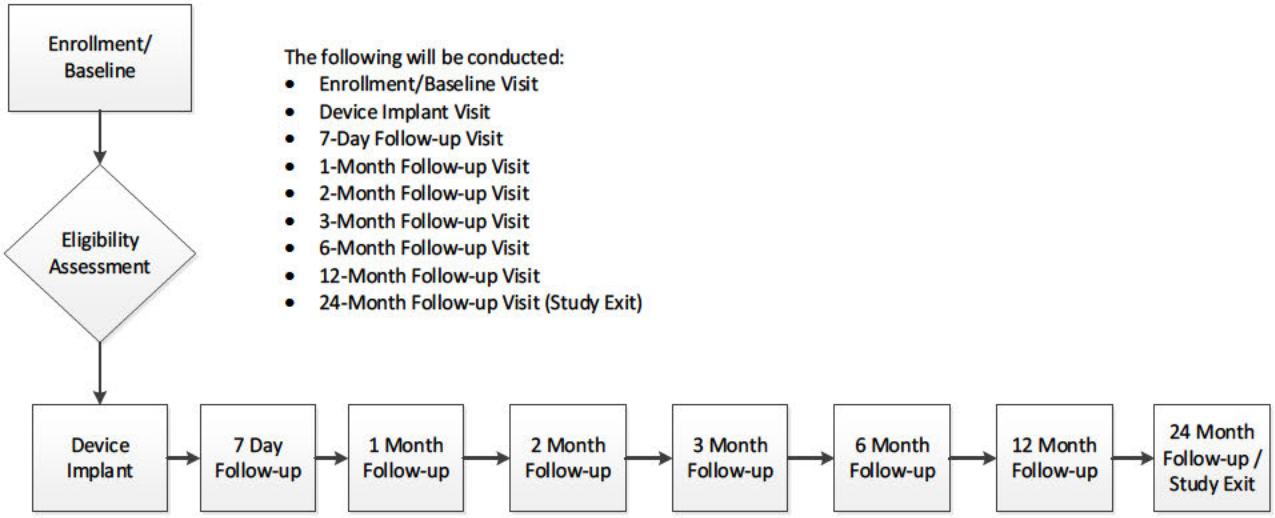
- Change in UUI episodes at 6 months compared to baseline in subjects with UUI at baseline
- Change in daily UF episodes at 6 months compared to baseline in subjects with UF at baseline
- Change in urinary urgency assessed through the UPS at 6 months compared to baseline
- Change in OAB-q HRQL (health related quality of life) Total Score at 6 months compared to baseline



5. Investigation Plan

This is a prospective, multicenter, single-arm pivotal study to assess safety and effectiveness for the implantable TNM device in subjects with overactive bladder. Subjects who meet all inclusion/exclusion criteria will be implanted with the implantable TNM device. A full list of eligibility criteria can be found in the CIP. Study visits are shown in Figure 1.

Figure 1: Study Visits



6. Determination of Sample Size

Sample size determination was made using PASS 2019. The sample size for the primary objective was estimated using a binomial distribution for a one-sided alpha=0.025 test to show a 15% increase for the proportion compared to a performance goal of 40%. Assuming the alternative hypothesis of ≥55% of subjects who are considered an UUI responder, a minimum of 121 implant attempted subjects achieves at least 90% power to reject a performance goal of 40%. The primary objective will have at least 88% power for any alternative hypothesis that is at least 15% larger than a performance goal. Secondary objectives are also expected to be adequately powered based on this sample size.

Assuming a ~40% attrition rate before implant, up to 200 subjects will be consented to achieve at least 121 subjects with implant attempted. Enrollment will be closed once the number of completed implants reaches 121; all enrolled subjects who meet eligibility criteria may be implanted (not to exceed 130 total implants). Implant attempted subjects who exit from the study will not be replaced.

7. Statistical Methods

7.1 Study Subjects

7.1.1 Disposition of Subjects

Subject disposition will be summarized in a flow diagram by study visit. A summary of subjects included in each analysis set will be provided. Discontinuation will be summarized by visit and discontinuation reasons.

7.1.2 Clinical Investigation Plan (CIP) Deviations

A study deviation is defined as an event within a study that did not occur according to the CIP. The frequency of CIP deviations will be summarized by type of deviation. A summary of deviation type by clinical study center will also be included.

7.1.3 Analysis Sets

All Enrolled analysis set: Includes subjects who completed the informed consent process.

Full Analysis Set (FAS): Includes subjects with an attempted implant. If all implant procedures are successful, this will be equivalent to the All Implanted Analysis set. In the case that all subjects have successful implant procedures, the name 'All Implanted Analysis set' will be used. This is the primary analysis set for both safety and effectiveness .

All Implanted Analysis set: Includes subjects who are implanted.

Complete Case analysis sets: Includes subjects who are implanted and provide outcome measures at both baseline and follow-up visits. This set is defined for each outcome measure at each follow-up visit.

As Treated analysis sets: Includes only those complete case subjects who received stimulation in the period prior to the visit. This set is defined for each outcome measure at each follow-up visit. Subjects will be excluded from the As Treated analysis sets if they meet 1 or more of the following:

- If the subject received zero therapy sessions in the 2 weeks prior to the visit
- If the subject experienced a consecutive 4 week period with zero therapy sessions in the 3 months prior to the visit.

Note: In the case of missing daily usage data, all missing usage days will be considered as days where the subject received a therapy session as expected per their current program (i.e., missing usage data will not contribute to meeting the 2 criteria above).

The study requires subjects to have diagnosed UUI to be implanted. For analysis purposes, Urinary Frequency (UF) subjects are those who have ≥10 urinary frequency episodes per day at baseline and all subjects will be considered to have urinary urgency since all eligible subjects have UUI.

Some supportive, secondary objective, [redacted] may use different definitions or subsets of the analysis sets listed above (e.g., changed assumptions for those subjects considered to have either UUI or UF).

7.2 General Methodology

The study will be considered successful when the primary objective is met. To support an indication for OAB, device therapy effectiveness will be demonstrated for the symptoms associated with OAB, including UUI, UF, and urgency.

For each of the objectives, the available data will be summarized, and missing data will be discussed. The main analysis of the study objectives will be intention-to-treat, meaning that the analysis will use the Full Analysis Set (FAS). The FAS includes all subjects, using the intention-to-treat (ITT) principle, who have an attempted device implant. Additional information on missing data plans are found in section 7.4.

Analyses of all data will be summarized with standard descriptive summary statistics, including counts and percentages for categorical variables, and mean, standard deviation, median, minimum, and maximum for continuous variables. The primary objective will be tested as described in section 7.9.1.1. If the primary objective is passed, the secondary objectives will each be tested as described in section 7.9.2 following the Hochberg multiple testing strategy. All p-values will be reported as 2-sided tests. Additionally, 95% Confidence intervals will be supplied for selected measures.

Safety data will be summarized as the count of events, count of subjects, and percentage of subjects who experienced the event. Subject disposition will be illustrated in a flow diagram. Subject visits will be tabulated and compliance to the visit schedule and visit windows will be summarized. Attrition will be identified and summarized, including the number and causes of death if any occur in the study.

Data analysis will be performed by Medtronic-employed statisticians or designees. A validated statistical software package (e.g., SAS version 9.4 or higher) will be used to analyze the study results. PASS 2019 sample size calculation software was utilized for determination of sample sizes and associated statistical power calculations.

7.2.1 Justification of study design

The single-arm nature of the study design has a limitation of no randomized concurrent control group. However, due to therapy sensations delivered by TNM, no blinded study design is possible. Without blinding, expectation differences may result in widespread placebo effect differences between the active and control therapies, which may introduce group differences that are not due to treatment. Instead, the use of a performance goal defines a specific relevant effect that must be overcome to identify a successful therapy.

7.2.2 Diary Data Analysis

Several objectives [redacted] are based on the analysis of voiding diary data. The diary will be collected for 3 days prior to or after each study visit. The center will assign each diary to the corresponding visit in the eCRF. In some cases, the diary might be completed for slightly less or more than 3 days (e.g., for 68 or 75 rather than the expected 72 hours). All diaries which are greater than 48

hours will be used for analysis. For UI, UUI, and UF, the total number of episodes will be determined for each subject's diary collection period at each visit. UUI episodes are those leaks with urgency greater than none, urinary incontinence (UI) episodes are all leaks, and UF are all voiding episodes. The actual total length of the diary for each visit will be determined by using the subject-specific diary start and end time reported by the center. The number of episodes will be divided by the total length of the diary for that subject at the visit and reported as a daily (24 hour) rate (i.e., 'per day'). The change will be determined by subtracting the baseline value per day from the follow-up value per day. Percentage change will be determined by dividing the change in episodes per day by the baseline value. Negative values for change and percentage change represent an improvement. Responder rate for UUI is defined as the proportion of TNM subjects experiencing a reduction of 50% or more in daily UUI episodes. Responder rate for UF is defined as the proportion of TNM subjects experiencing a reduction of 50% or more in daily UF episodes or a return of voiding to <8 voids per day. Responder rate for UI is defined as the proportion of TNM subjects experiencing a reduction of 50% or more in daily UI episodes. Responder rate for UU is defined as the proportion of TNM subjects with an improvement of at least 1 point on the UPS. OAB responder rate is defined as the proportion of TNM subjects who are considered either a UUI, UF, or UU responder.

[redacted]

7.3 Center Pooling

The investigators of this study will conduct the study according to this protocol and use the study-specific eCRFs to collect study data. The study center personnel will be trained prior to the study initiation at each study center. Periodic study monitoring by Medtronic will ensure compliance with

protocol requirements. Therefore, there is no a priori provision to exclude any study centers from the analysis. The data from all study centers will be pooled for analysis. To reduce the possibility of atypical results from a study center overly influencing the combined results, no more than 24 subjects will be implanted at each study center. The per-study center enrollment cap may be increased upon Sponsor approval.

As an exploratory analysis to assess if treatment effect differs by study center, Freeman and Halton's extension of Fisher's exact test will be conducted using UUI responder as the response variable. This analysis will be based on the Complete Case analysis set for the primary objective. If the p-value approaches statistical significance (defined as ≤0.15), the percentage of subjects who are considered a UUI responder will be presented by center. If center(s) causing the significance are identified, variables relating to patient characteristics and other factors will be analyzed to try to identify why this center is showing a different treatment effect.

To increase the power of the test for treatment-by-center interaction, all study centers that are too small (fewer than four implanted subjects) to be reliably analyzed alone or to have a noticeable impact on overall study results will be combined into one "virtual" study center. All other study centers with four or more implanted subjects will stand alone in the analysis. If the virtual study center is larger than 30 implanted subjects (~25% of the required sample size), the virtual study center will be split into 2 virtual study centers defined by ordering the study centers alphabetically and splitting the study centers as near the midpoint as possible. In the case that the midpoint is within an individual study center, that study center will be included with the study centers from earlier in the alphabet.

7.4 Handling of Missing, Unused, and Spurious Data and Dropouts

Primary and secondary objectives will be analyzed using the FAS. The FAS includes all subjects, using the intention-to-treat principle, who have an attempted device implant. All attempts will be made to minimize missing data. For diary-based metrics, missing is defined as any diary which has a total length of less than 48 hours, as defined in section 7.2.2. The OAB-q may be missing if the HRQL Total cannot be calculated; scoring details are found in section 7.9.2.4. If 5% of the data or fewer are missing for the analysis, no imputation will be used, and the data will be analyzed using the Complete Case analysis set (Buhi 2008). Otherwise, if missing data are observed, imputation will be used for the primary and secondary objectives.

The primary objective imputation will be done using multiple imputation (MI). Predictive mean matching multiple imputation will be used for each variable. First, multiple imputation will be applied to impute missing values in average daily UUI episodes. Unless model errors dictate otherwise, each model will consider Baseline and Month 1, Month 2, Month 3, and Month 6 variables for the UUI episodes variable and demographic variables of age, gender, history of pregnancy, and BMI. The fully conditional specification (FCS) method within SAS (version 9.4 or higher) with 30 repetitions and 100 burn-in iterations will be used for imputation. No limits will be defined for imputed values. Following

imputation, the 6-month outcome data will be converted to the appropriate endpoint (i.e., change from baseline/UUI responder). The estimate and variance of the mean proportion of subjects considered a UUI responder will be determined for each imputation dataset. For the primary objective, the mean and variance for the proportion will be determined using MI analysis methods and the objective will be tested using a Z-test comparing the observed responder rate to the primary objective performance goal of 40%.

For the UUI, UF, and OAB-Q secondary objectives, the following imputation steps will be taken:

- For secondary objectives related to UUI episodes, the imputation datasets created for the primary objective will be used. For OAB-q HRQL and UF episodes, a new set of imputation datasets will be created following the same methods as the primary objective but replacing UUI episodes with OAB-q HRQL or UF episodes. For OAB-q HRQL, the imputation model contains HRQL scores only at those visits where the assessment was collected (Baseline, Month 3, Month 6). Constraints will be set so that the imputed OAB-q HRQL are restricted to values ranging from 0-100. If the MI procedure is unable to successfully impute values within that range, the constraints will be applied after imputation, such that any values less than 0 will be set to 0, and any values greater than 100 will be set to 100.
- After imputation, the normality for the outcome metric will be checked through Shapiro-Wilk test for each imputation dataset. If p-value>0.05 for at least 15 imputed datasets (half of the total), it will be considered that the data are normally distributed. In this case, the mean and standard error of the outcome metric will be calculated within each imputed dataset, and then combined using MI analysis methods. The mean and standard error of the outcome metric will be reported, and the objective will be tested with a t-test.
- If the distribution of outcome metric is non-normal (p-values ≤0.05 in more than 15 of the imputed datasets), MI will not be used. Instead, last observations carried forward (LOCF) will be applied, with the imputed dataset tested using the Wilcoxon signed-rank test.

For the secondary objective related to urgency using the UPS, the imputation datasets will be created using multiple imputation (MI). Logistic regression multiple imputation will be used for each variable which accounts for the ordinal nature of the UPS. First, multiple imputation will be applied to impute missing values in UPS. Unless model errors dictate otherwise, each model will consider Baseline, Month 3, and Month 6 variables for the UPS variable and demographic variables of age, gender, history of pregnancy, and BMI. The fully conditional specification (FCS) method within SAS (version 9.4 or higher) with 30 repetitions and 100 burn-in iterations will be used for imputation. Imputed values will be required to be 1, 2, or 3 to match the original scale. After imputation, the mean and standard error of the change from Baseline to Month 6 in UPS will be calculated within each imputed dataset, and then combined using MI analysis methods. The mean and standard error will be reported, and the objective will be tested with a t-test.

Sensitivity analyses will be completed for the primary and secondary objectives including subjects who received a TNM implant (“All Implanted Analysis”), using only subjects with outcome data at baseline

and follow-up ("Complete Case Analysis"), and using only those complete case subjects who received stimulation ("As Treated Analysis").

[redacted] Comprehensive descriptions of subject accountability and missing data will be provided.

7.5 Adjustments for Multiple Comparisons

The primary objective is the percentage of subjects considered a UUI responder at 6 months. Four secondary objectives will also be assessed, all of which will be assessed 6 months after implantation of tibial neuromodulation.

To support an indication for OAB, device effectiveness will be demonstrated for all the symptoms associated with OAB, including UUI, UF, and urgency. UUI therapy effectiveness will be demonstrated if the primary objective proportion of subjects with UUI response is greater than the performance goal (40%). UF therapy effectiveness will be demonstrated if the secondary UF objective demonstrates an improvement from baseline to 6 months in number of UF episodes. Urgency therapy effectiveness will be demonstrated if the secondary urgency objective demonstrates an improvement from baseline to 6 months in UU.

The p-values for the primary and secondary objectives will be calculated according to the methods outlined in their respective analysis sections. The p-values will be reported unadjusted and two-sided; however, declarations of statistical significance will follow the procedures laid out in this section.

The primary objective will be tested separately from the secondary objectives. If the one-sided primary objective p-value is <0.025, then the null hypothesis will be rejected, and the test will be declared statistically significant.

The Hochberg method for the multiplicity adjustment will be used to test the four secondary objective hypotheses, thereby maintaining an overall two-sided type I error rate at 0.05 for these objectives. Each p-value for the four secondary objectives will be calculated. If all p-values are less than 0.05 (two sided), all will be declared significant. If not, the p-values will be ordered from largest to smallest (identified as p1-p4 in this section). If p1 is greater than 0.05, the remaining p-values (p2, p3, p4) will be compared against an alpha of 0.025 (0.05/2). If the three remaining p-values are all less than 0.025, they will all be declared significant. If p2 is greater than 0.025, the remaining p-values (p3, p4) will be compared against an alpha of 0.017 (0.05/3). If p3 and p4 are both less than 0.017, they will both be declared significant. If p3 is greater than 0.017, p4 will be compared against an alpha of 0.013 (0.05/4). If p4 is less than 0.013, it will be declared significant.

The multiplicity adjustment will be performed for evaluations of the secondary endpoints only. No multiplicity adjustment will be performed for supporting, sensitivity, or subset analyses [redacted]

7.6 Demographic and Other Baseline Characteristics

Demographics and other baseline characteristics summarized will include:

- Age, sex, race, and ethnicity
- Years since UUI diagnosis
- Secondary diagnoses
- Body Mass Index (BMI)
- Medical history
- Steroid use and smoking status
- OAB treatment history
- Baseline UUI episodes per day (for those subjects with UUI diagnosis), baseline voids per day (for those subjects with UF diagnosis), quality of life (OAB-q HRQL score), average degree of urgency for all UI and UF episodes, and number of subjects with each urgency level on the UPS

7.7 Treatment Characteristics

Treatment measures summarized will include stimulation programming measures, such as programmed amplitude, pulse width, and frequency (Hz), program use, and a summary of usage data. Measures of delivered treatment will be summarized with standard descriptive statistics including means, standard deviations, medians, minimum, and maximum for each follow-up visit.

7.8 Interim Analyses

No interim assessment of effectiveness will be completed. Safety stopping rules are documented in the CIP.

Study progress reports will be submitted annually. Analysis for the purposes of supporting pre-market approval applications will occur once all subjects have completed the 6-month follow-up visit and will include the primary and secondary objectives and detailed safety summaries. Additional measures will be addressed based on available data at time of the pre-market approval application. A final report will be prepared once all data collection has ended and all subjects have completed the 24-month follow-up visit or have been exited. The inferential analysis for the primary and secondary objectives will not be updated for the final report.

7.9 Evaluation of Objectives

7.9.1 Primary Objective

7.9.1.1 Primary Objective – UUI Responder Rate

To demonstrate that the percentage of subjects who are considered a UUI responder after 6 months of tibial neuromodulation exceeds a performance goal of 40%.

Hypothesis: The percentage of subjects who are considered a UUI responder 6 months after tibial neuromodulation implantation is greater than 40%.

H_0: $p \leq 40\%$

H_A: $p > 40\%$

Where p = the percentage of subjects who are UUI responders

Endpoint Definition and Derivation: A 3-day urinary voiding diary will be used to collect bladder symptoms, including UUI episodes. The primary objective endpoint is the proportion of tibial neuromodulation subjects considered a UUI responder at 6 months after device implant. A subject is considered a UUI responder if they meet the UUI responder criteria as defined in the voiding diary analysis section 7.2.2.

Performance Requirements: The null hypothesis will be rejected if the one-sided p-value is less than 0.025. Only the two-sided p-value will be reported.

Rationale for Performance Criteria: The proposed study design includes a primary objective performance goal of 40%. The performance goal defines an active historical control for this study and was defined to position the TNM outcome between the expected performance for medications and the more complex SNM treatment.

Sample Size Justification: The sample size justification for the primary objective is provided in section 6.

Analysis Methods: The analysis for UUI responder rate will be summarized at 6 months after implant. If 5% of the data or fewer are missing for the primary analysis, no imputation will be used, and the objective will be tested using a binomial test for the Complete Case analysis set. Otherwise, missing data will be imputed for the Full Analysis Set as specified in section 7.4, and the objective will be tested using a Z-test. The two-sided 95% confidence interval will also be provided.

Determination of Subjects/Data for Analysis: The primary objective analysis will include all subjects from the Full Analysis Set (i.e., all implant attempted subjects) as defined in section 7.1.3.

Supportive and Sensitivity Analyses: There are no requirements or expectations for the results of the supporting and sensitivity analyses. They are meant to investigate the robustness of the study results and/or to provide further details on the therapy effects.

Additional supporting analyses to further understand the primary objective will be completed through analyses of the change in UUI (secondary objective) and percentage change in UUI (additional measure).

To assess the robustness of the primary objective results under the influence of missing data, sensitivity and subgroup analyses will be conducted using alternative missing data methods and by including different subsets of subjects. The first sensitivity analysis will be completed including subjects from the Primary Objective Complete Case analysis set. This analysis set will be used for a second sensitivity analysis which will be completed including only subjects who received stimulation ('As treated analysis') at the 6-month follow-up visit. An analysis will also be completed to evaluate UUI responder rate differences between different stimulation programs, comparing stimulation programs only if there are an adequate number of subjects in each stimulation program to ensure a stable estimate of response. A tipping point sensitivity analysis will also be completed. A final sensitivity analysis will be conducted on the Primary Objective Complete Case analysis set assigning response failure for the following subjects: those who took any dose of OAB medication in the 2 weeks prior to the diary collection period, and those who received an advanced therapy at any time from implant to the 6-month follow-up visit.

Complete Case analysis set subgroup analyses will be done for the following subject groups: sex, age (divided by median age; divided by age 65), race, and ethnicity. Subjects on the demographic cut point will be added to the adjacent group containing fewer subjects. Any subgroup with less than 10 subjects will not be reported.

As an exploratory analysis to assess if treatment effect differs by study center, a supporting analysis will be completed, as described in 7.3. This analysis will test whether the treatment effect is homogeneous between study centers for the UUI responder rate at 6 months.

UUI responder rate will be assessed over time through 24 months as part of the additional measures.

7.9.2 Secondary Objective(s)

The following secondary objectives have specific hypotheses to be tested. After the primary objective is met, these four secondary objectives will be tested as described in 7.5.

7.9.2.1 Secondary Objective – UUI Change

To demonstrate an improvement from baseline to 6 months in number of UUI episodes in subjects with UUI at baseline

Hypothesis: The change from baseline in UUI episodes is not 0.

$H_0: \mu = 0$

H_A: $\mu \neq 0$

Where μ = the change in UUI episodes

Endpoint Definition and Derivation: A 3-day urinary voiding diary will be used to collect bladder symptoms, including UUI episodes. Calculation of change in UUI is defined in the voiding diary analysis section 7.2.2. A negative change in UUI episodes represents an improvement.

Performance Requirements: The null hypothesis will be rejected if the two-sided p-value is determined to be significant, as described in section 7.5.

Sample Size Justification: Data from published studies or studies presented at conferences on TNM outcomes at 3, 6, and 9 months were used as starting points to estimate expected reduction in UUI episodes as well as expected patient-to-patient variability. Assumption on the expected mean difference (-2.0 episodes per day) and standard deviation (3.5 episodes per day) was used to assess statistical power. Presuming a two-sided, one-sample t-test with alphas ranging from 0.013 to 0.05, power is calculated to be no less than 99% for any of the scenarios examined with a sample size of 121 subjects. Additionally, the minimum detectable change in mean episodes was calculated to be as low as -1.0.

Analysis Methods: The analysis for change in UUI will be summarized for baseline, at 6 months after implant, and for the change from baseline. If 5% of the data or fewer are missing for the analysis, no imputation will be used, and the objective will be tested using a paired t-test following the multiple testing strategy described in section 7.5. If more than 5% of the data are missing, missing data will be imputed as specified in section 7.4. If non-normality of the data is detected via a significant ($p<0.05$) Shapiro-Wilk test, a Wilcoxon signed-rank test will replace the t-test.

Determination of Subjects/Data for Analysis: The change in UUI analysis will include all subjects from the Full Analysis Set (i.e., all implant attempted subjects) as defined in section 7.1.3.

Supportive and Sensitivity Analyses: There are no requirements or expectations for the results of the supporting and sensitivity analyses.

Additional supporting analyses to further understand the UUI secondary objective will be completed through analyses of the percentage change in UUI (additional measure).

To assess the robustness of the objective results under the influence of missing data, sensitivity analyses will be conducted using alternative missing data methods. The first sensitivity analysis will be completed including subjects from the Secondary Objective (UUI Change) Complete Case analysis set. This analysis set will be used for a second sensitivity analysis which will be completed including only subjects who received stimulation ('As treated analysis') at the 6-month follow-up visit. A final complete case sensitivity analysis will be performed comparing the change in UUI episodes to a performance goal of 1 episode/day.

UUI change from baseline will be assessed over time through 24-months as part of the additional measures.

7.9.2.2 Secondary Objective – UF Change

To demonstrate an improvement from baseline to 6 months in number of UF episodes in subjects with UF at baseline

Hypothesis: The change from baseline in UF episodes is not 0.

H_0: $\mu = 0$

H_A: $\mu \neq 0$

Where μ = the change in UF episodes

Endpoint Definition and Derivation: A 3-day urinary voiding diary will be used to collect bladder symptoms, including UF episodes. UF episodes include all urinary voids. Calculation of change in UF is defined in the voiding diary analysis section 7.2.2. A negative change in UF episodes represents an improvement.

Performance Requirements: The null hypothesis will be rejected if the two-sided p-value is determined to be significant, as described in section 7.5.

Sample Size Justification: Data from published studies on TNM outcomes at 6 months were used as a starting point to estimate expected reduction in UF episodes as well as expected patient-to-patient variability. Assumption on the expected mean difference (-2.0 episodes per day) and standard deviations (3.0 episodes per day) were used to assess statistical power. Presuming a two-sided, one-sample t-test with alphas ranging from 0.013 to 0.05, power is calculated to be no less than 98% for any of the scenarios examined with a sample size of 51 subjects (as only 42% of the total 121 subjects in the study are expected to have UF with 10 episodes per day at baseline). Additionally, the minimum detectable difference in mean daily episodes was calculated to be as low as 1.4.

Rationale for Clinical Relevance: No published data are available defining minimum clinically meaningful improvement for the change in UF episodes. In the absence of published definitions, assessment of effect size can be used to determine the level of effect. As published by Cohen (1988)[27], an effect size of 0.2 indicates a small effect, 0.5 a modest effect, and 0.8 a large effect. A modest effect size would correspond to a change of -1.5 episodes per day. The assumed mean ± standard deviation difference of 2.0 ± 3.0 UF episodes per day (i.e., an effect size = 0.67) for implantable tibial therapy is considered a medium to large effect.

Analysis Methods: The analysis for change in UF will be summarized for baseline, at 6 months after implant, and for the change from baseline. If 5% of the data or fewer are missing for the analysis, no imputation will be used, and the objective will be tested using a paired t-test following the multiple testing strategy described in section 7.5. If more than 5% of the data are missing, missing data will be

imputed as specified in section 7.4. If non-normality of the data is detected via a significant ($p<0.05$) Shapiro-Wilk test, a Wilcoxon signed-rank test will replace the t-test.

Determination of Subjects/Data for Analysis: The change in UF analysis will include all subjects from the Full Analysis Set (i.e., all implant attempted subjects) who have UF as defined in section 7.1.3.

Supportive and Sensitivity Analyses: There are no requirements or expectations for the results of the supporting and sensitivity analyses.

To assess the robustness of the objective results under the influence of missing data, sensitivity analyses will be conducted using alternative missing data methods. The first sensitivity analysis will be completed including subjects from the Secondary Objective (UF Change) Complete Case analysis set. This analysis set will be used for a second sensitivity analysis which will be completed including only subjects who received stimulation ('As treated analysis') at the 6-month follow-up visit. A final complete case sensitivity analysis will be performed comparing the change in UF episodes to a performance goal of 1 episode/day. A Complete Case sensitivity analysis will be provided assessing the UF change from baseline using different definitions of UF (UF defined as ≥8 and UF defined as ≥12).

UF change from baseline will be assessed over time through 24 months as part of the additional measures.

7.9.2.3 Secondary Objective – UPS Change

To demonstrate an improvement from baseline to 6 months in urgency using the Urgency Perception Scale (UPS)

Hypothesis: The change from baseline in UPS is not 0.

H_0: $\mu = 0$

H_A: $\mu \neq 0$

Where μ = the change in UPS score

Endpoint Definition and Derivation: The UPS will be used to collect OAB urgency. The UPS was developed to assess urgency associated with overactive bladder (OAB). The UPS is a single question assessed at each visit and has 3 options for responses: 1, 'I am usually not able to hold urine'; 2, 'I am usually able to hold urine until I reach the toilet if I go immediately'; and 3, 'I am usually able to finish what I am doing before going to the toilet'. Change from baseline is calculated by subtracting the baseline value from the follow-up value. Therefore, the change can range from -2 to 2 (with positive change indicating an improvement)[23].

Performance Requirements: The null hypothesis will be rejected if the two-sided p-value is determined to be significant, as described in section 7.5.

Sample Size Justification: Internal data from TNM outcomes were used as a starting point to estimate expected change in UPS as well as patient-to-patient variability in that change. Assumptions on the expected mean difference (0.50) and standard deviation (0.80) were used to assess statistical power. Presuming a two-sided, one-sample t-test with alphas ranging from 0.013 to 0.05, power is calculated to be at least 99% for any of the scenarios examined with a sample size of 121 subjects. Additionally, the minimum detectable change in UPS was calculated to be as low as 0.24 points.

Rationale for Clinical Relevance: No published data are available defining minimum clinically meaningful improvement for UPS. In the absence of published definitions, assessment of effect size can be used to determine the level of effect. As published by Cohen (1988)[27], an effect size of 0.2 indicates a small effect, 0.5 a modest effect, and 0.8 a large effect. A modest effect size would correspond to a change of 0.4 in the mean UPS. The assumed mean ± standard deviation change of 0.5 ± 0.8 in the UPS (i.e., an effect size = 0.63) for tibial therapy is considered a medium to large effect.

Analysis Methods: The analysis for UPS will be summarized for baseline, at 6 months after implant, and for the change from baseline. If 5% of the data or fewer are missing for the analysis, no imputation will be used, and the objective will be tested using a paired t-test following the multiple testing strategy described in section 7.5. If more than 5% of the data are missing, missing data will be imputed as specified in section 7.4. If non-normality of the data is detected via a significant ($p<0.05$) Shapiro-Wilk test, a Wilcoxon signed-rank test will replace the t-test.

Determination of Subjects/Data for Analysis: The UPS analysis will include all subjects from the Full Analysis Set (i.e., all implant attempted subjects) as defined in section 7.1.3 as all subjects are expected to have urgency.

Supportive and Sensitivity Analyses: There are no requirements or expectations for the results of the supporting and sensitivity analyses.

To assess the robustness of the objective results under the influence of missing data, sensitivity analyses will be conducted using alternative missing data methods. The first sensitivity analysis will be completed including subjects from the Secondary Objective (UPS) Complete Case analysis set. This analysis set will be used for a second sensitivity analysis which will be completed including only subjects who received stimulation ('As treated analysis') at the 6-month follow-up visit. A third sensitivity analysis will be performed excluding those subjects who report limited problems with urgency, as reported by a baseline score of 3 on the UPS. A final complete case sensitivity analysis will be performed comparing the change in urgency to a performance goal of 0.4.

A supportive analysis will be completed summarizing the percentage of subjects in each category at each visit. The change in UPS will be further categorized as improvement (1-2 point improvement), no change, and worsening (1-2 point worsening), with the number and percentage of subjects summarized for each category.

To test the change in UPS using an alternative statistical analysis method, a supportive analysis will be completed recategorizing the values at each visit as into "not able to hold urine" (response level 1) and "able to hold urine" (levels 2/3). Change from baseline to Month 6 will be tested using Fisher's exact test for the 2x2 contingency table.

UPS change from baseline will be assessed over time through 24 months as part of the additional measures.

7.9.2.4 Secondary Objective – OAB-q Change

To demonstrate an improvement from baseline to 6 months in quality of life (QoL) using the Overactive Bladder Symptom Quality of Life Questionnaire (OAB-q)

Hypothesis: The change from baseline in OAB-q HRQL is not 0.

H_0: $\mu = 0$

H_A: $\mu \neq 0$

Where μ = the change in OAB-q HRQL total score

Endpoint Definition and Derivation: The OAB-q will be used to collect quality of life. The OAB-q was developed to assess symptom bother and the impact of overactive bladder (OAB) on health-related quality of life (HRQL). The Health-Related Quality of Life (HRQL) total score is calculated by summing the four subscale scores (coping, concern, sleep, social). For each subscale, if <50% of the items are missing, the subscale is retained with the mean of the items present used to impute the missing items in that subscale. If ≥50% of items in any subscale are missing, the HRQL Total will be missing. A positive change in OAB-q HRQL score represents an improvement.

Performance Requirements: The null hypothesis will be rejected if the two-sided p-value is determined to be significant, as described in section 7.5.

Sample Size Justification:

Data from a published study on TNM outcomes at 6 months were used as a starting point to estimate expected change in OAB-q HRQL score as well as patient-to-patient variability in that change.

Assumptions on the expected mean change (20 points) and standard deviation (30 points) were used to assess statistical power. Presuming a two-sided, one-sample t-test with alphas ranging from 0.013 to 0.05, power is calculated to be no less than 99% for any of the scenarios examined with a sample size of 121 subjects. Additionally, the minimum detectable change in OAB-q HRQL score was calculated to be as low as 9 points.

Analysis Methods: The analysis for OAB-q HRQL total score will be summarized for baseline, at 6 months after implant, and for the change from baseline. If 5% of the data or fewer are missing for the analysis, no imputation will be used, and the objective will be tested using a paired t-test following the multiple

testing strategy described in section 7.5. If more than 5% of the data are missing, missing data will be imputed as specified in section 7.4. If non-normality of the data is detected via a significant ($p<0.05$) Shapiro-Wilk test, a Wilcoxon signed-rank test will replace the t-test.

Determination of Subjects/Data for Analysis: The OAB-q total score analysis will include all subjects from the Full Analysis Set (i.e., all implant attempted subjects) as defined in section 7.1.3.

Supportive and Sensitivity Analyses: There are no requirements or expectations for the results of the supporting and sensitivity analyses.

To assess the robustness of the objective results under the influence of missing data, sensitivity analyses will be conducted using alternative missing data methods. The first sensitivity analysis will be completed including subjects from the Secondary Objective (OABq HRQL score) Complete Case analysis set. This analysis set will be used for a second sensitivity analysis which will be completed including only subjects who received stimulation ('As treated analysis') at the 6-month follow-up visit. A final complete case sensitivity analysis will be performed comparing the change in OAB-q HRQL to a performance goal of 10.

OAB-q HRQL total score change from baseline will be assessed over time through 24 months as part of the additional measures. In addition, the symptom bother score and OAB-q HRQL domains will be analyzed over time through 24 months. A supporting analysis will be completed for each visit by summarizing the number and percentage of subjects which have an improvement in the OAB-q HRQL score of at least 10 points.

[redacted]

[redacted]

[redacted]

To assess the robustness of the long-term results for the primary and secondary objective endpoint metrics (UUI responder, change in UUI episodes, change in UF episodes, change in urgency, and change in OAB-q HRQL) sensitivity analyses will be conducted using alternative missing data methods and subgroups of subjects. The first sensitivity analysis will be completed including all implanted subjects imputing data as described for the primary and secondary objectives extending the included outcome data through the last available visit. The Primary Objective Complete Case analysis set will be used for a second sensitivity analysis which will be completed including only subjects who received stimulation ('As treated analysis') in the period prior to the visit. A final sensitivity analysis will be conducted excluding all subjects who took OAB medication or received an advanced therapy between implant and the visit. In addition, the analyses for UUI responder rate, change in episodes, and percentage change in episodes will also be repeated for all urinary incontinence episodes (UI).

7.10 Safety Evaluation

All AEs and DDs will be collected throughout the study once the informed consent form is signed. AEs and DDs will be coded using the most recent version of Medical Dictionary for Regulatory Affairs (MedDRA). Severity will be assessed by the Investigator. Relatedness and seriousness of the events will be classified by the Investigator, CEC, and the Sponsor. Cases where the Investigator's classification does not match the CEC's classification will be displayed within the final report; however, the CEC's adjudication will be used for data analysis.

Adverse events that occur prior to implant will be summarized. Adverse events that occur on or after tibial neuromodulation implant date through study discontinuation will also be summarized; summaries specific to the timeframe from implant to the 6-month follow-up visit will be included as well as those that include all post-implant follow-up (i.e. though "24 months"). AE tables summarizing adverse device effects (ADE) will also be included. ADE are events deemed related to the neurostimulator, procedure, external device, study aids, or therapy. In addition to tables of overall ADE, summaries by the individual relatedness component will be summarized using the following categories: procedure-related, device-related (neurostimulator/external device), therapy-related, and study aids-related. Resolution (number of events/percentage) of ADE will be summarized using similar time groupings to other summaries. ADE will also be summarized grouping the events by Common Terminology Criteria for Adverse Events (CTCAE) grade.

The denominator for post-implant summaries will be the number of subjects in the All Implanted analysis set, while for pre-implant summaries it will include all enrolled subjects. Events occurring on date of implant will be considered post-implant events.

Summaries of serious adverse events (SAEs) will also be completed in similar ways to AE summaries. Unanticipated Serious Adverse Device Effects (USADE) will be described. Adverse events in subjects with implant attempted but without full implant will also be summarized as part of additional AE summary tables of the Full Analysis Set. Device-related deaths and system modifications will be summarized in detail.

Device deficiencies will be summarized by device type (i.e., communicator, handset, neurostimulator, wireless recharger/charging dock, external patient accessories, and sterile procedure accessories) through the 24 month visit.

Events will be summarized with number of events, number of subjects who experienced the event, and percentage of subjects who experienced one or more events.

7.11 Health Outcomes Analyses

No health outcomes analyses are planned for this study.

[redacted]

8. Validation Requirements

Statistical programming code that affects the result of the main analysis (e.g., not including sensitivity or supporting analyses) for the primary objective shall be validated using Level I validation. Programming code that affects the result of the main analysis for the secondary objectives shall be validated using at least Level II validation. In addition, those main statistical analyses that are planned for publication and have not been previously validated shall be validated using at least Level II validation. The CIP deviation summary shall be validated using at least Level III validation and the high-level adverse event summary shall be validated using at least Level II validation. Additional measures where a p-value or confidence interval has been generated may need to be validated using at least Level II validation.

9. References

Buhi ER, Goodson P, Neilands TB. Out of sight, not out of mind: strategies for handling missing data. *Am J Health Behav*. 2008;32(1):83-92. doi:10.5555/ajhb.2008.32.1.83

Cohen J. *Statistical Power Analysis for the Behavioural Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988