PROCEPT BioRobotics

Official Title: The WATER II Study
Statistical Analysis Plan

NCT Number: NCT0312350

Document Date: 09-Jan-2018

PROCEPT BioRobotics Corporation · 900 Island Drive, Suite 101 · Redwood City, California 94065
office: +1.650.232.7200 · Fax: +1. 650.232.5782 · info@PROCEPT-BioRobotics.com

The WATER II Study

Statistical Analysis Plan

Reference: TP0124A, WATER II Study Clinical Investigational Protocol

Date: January 9, 2018

Prepared by:

Daniel Chen — 1/9/18

VP of Regulatory and Quality Affair and Biostatistics — Date

Signature

Approved By:

B — 09Jan 2018

VP of Clinical Affairs — Date

Signature

TABLE OF CONTENTS

1 BACKGROUND

WATER II is a prospective multicenter trial of Aquablation using AQUABEAM system for men with symptomatic benign prostatic hyperplasia (BPH) and large prostates (80-150 g). See clinical investigational protocol TP0124 for details regarding eligibility, treatments and assessments. This document describes the study's statistical analysis plan (SAP).

This study is related to WATER (TP0038) in that the same device is used on men with the same condition. The major difference is that WATER II targets men with larger prostates and does not have a concurrent control group.

The purpose of WATER II is to determine safety and efficacy in the target population.

2 ELECTRONIC DATA CAPTURE

PROCEPT is using "iMedNet" (MedNet Solutions, Minnetonka, MN 55305) as the electronic data capture provider. iMedNet is a fully functional, 21 CFR 11-compliant web-based system for managing case report forms and downloading study-related data.

3 DATA ANALYSIS SOFTWARE

Statistical analysis will be done using R,* an open source data analysis package.

4 ANALYSIS COHORTS

The following analytic cohorts are defined.

4.1 Modified Intent-to-Treat (mITT) Cohort

The mITT population includes all enrolled subjects in whom the Aquablation handpiece is inserted into the penile urethra. A patient found at the time of the procedure to have a condition that results in study exclusion does not contribute to the mITT cohort. The mITT population is the primary analysis population for both the primary safety and effectiveness endpoints.

4.2 Per-protocol (PP) Cohort

The per-protocol (PP) population is all mITT subjects who:

1. meet critical study eligibility criteria;
2. have no significant protocol deviations that could affect the validity of data; and
3. have evaluable assessment for the endpoint of interest.

A significant protocol deviation means non-adherence on the part of the subject or investigator to clinically significant protocol-specific inclusion/exclusion criteria, primary objective variable criteria or critical study requirements that could affect the scientific validity of the observed data point or lead to bias. Missing data are not imputed for analyses with the PP cohort.

4.3 Safety Cohort

The safety analysis population includes all subjects in whom Aquablation is initiated. This cohort is used for most safety analyses.

* See https://cran.r-project.org/

5 STUDY ENDPOINTS

5.1 Primary Effectiveness Endpoint

The primary effectiveness endpoint is the IPSS change score from baseline to 3 months. IPSS is an accepted and commonly used measure for symptom severity in BPH.[1] IPSS varies from 0 (no symptoms) to 35 (maximal symptoms). The mean 3-month IPSS change score (abbreviated $dIPSS_3$) in WATER was 16 points (SD=7 points). WATER used a non-inferiority margin of 5 points; 5 points represents the minimum change for men with moderate-to-severe BPH that is detectable.[2] The current study's goal is therefore to show that the IPSS change score statistically exceeds 16-5 = 11 points. The study's hypotheses are therefore:

H_0: $dIPSS_3 \leq 11$
H_a: $dIPSS_3 > 11$

A one-way t test with the mITT cohort will be used for the calculation. With 100 patients and assuming 10% subjects lost to follow, 90 evaluable subjects with an underlying change of 16 points and a change score SD of 7.5 points yields a 99% power to detect an improvement of more than 11 points.

A secondary goal is to show similarity of change scores between WATER II and WATER. Data from WATER and WATER II will be combined. Regression analyses controlling for baseline IPSS score (a strong predictor of change scores) will calculate differences in change scores between WATER and WATER II subjects. WATER II will be determined to have similar dIPSS3 scores if the 95% confidence limits for the difference between WATER and WATER II Aquablation subjects exceeds -5 points. The study is not necessarily powered for this comparison.

Missing data for dIPSS3 will be imputed using the following rules:

- Predicted value from baseline IPSS: In WATER, dIPSS3 was strongly and statistically significantly related to baseline IPSS (Figure 1). This finding is not surprising for maximally effective procedures that reduce scores to close to maximal (i.e., a ceiling effect). A linear regression model using WATER II available data will be used to predict missing change scores.
- Low value: A change score of 0 points
- High value: a change of 5 less than the maximal possible improvement, which represents change scores from WATER (see also Figure 1)

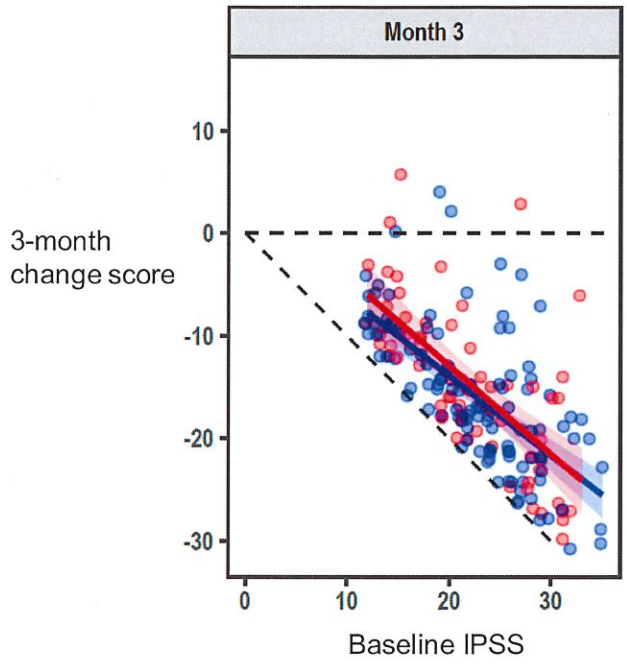

Figure 1. Relationship between baseline IPSS and 3-month change score in WATER. Blue = Aquablation, red = TURP.

5.2 Primary Safety Endpoint

The study's primary safety endpoint is the proportion of subjects with adverse events rated as probably or definitely related to the study procedure classified as Clavien-Dindo Grade 2 or higher or any Grade 1 event resulting in persistent disability (e.g. incontinence, ejaculatory disorder, or erectile dysfunction) evidenced through 3 months post treatment. Note that the Clavien-Dindo classification scheme is for grading postoperative complications, not events that reflect lack of treatment efficacy. The endpoint is adjudicated by the study's clinical events committee (CEC).

The study's primary endpoint hypothesis is a one-way statistical test (alpha=.025 for interpretation):

H_0: Mean (safety rate at 3 months) $\geq$ 65%
H_a: Mean (safety rate at 3 months) $<$ 65%

An exact binomial test (binom.test in R) will be used to determine the P-value for the above comparison. Power calculations showed that the primary safety endpoint with 90 evaluable subjects has an 80% power to detect a safety rate of less than 65%.

The primary safety endpoint is calculated using the mITT population.

6 SECONDARY ENDPOINT ASSESSMENTS

The following secondary endpoints are calculated.

6.1 Procedure Time

Length of procedure time as defined as AQUABEAM handpiece insertion to insertion of indwelling catheter. Mean and 95% confidence intervals will be calculated. There is no statistical hypothesis for this endpoint; it is considered informational.

6.2 Ejaculatory or Erectile Dysfunction

Proportion of subjects who were sexually active at baseline and experienced either ejaculatory or erectile dysfunction (per protocol definitions) at 3 months as adjudicated by the CEC as possibly, probably or definitely related to the study procedure. An exact binomial confidence interval will be calculated using binom.test or CIbinomial; there is no statistical hypothesis. Because ejaculatory and erectile dysfunction after prostate surgery are likely to have different underlying causes, similar calculations will be performed for each event type. Because there can be other causes (apart from surgery), further calculations will exclude subjects in whom the CEC determines that the event is unrelated (i.e., judged as neither possibly, probably or definitely related).

6.3 Change in Qmax

Change from baseline to 3 months in Qmax, a uroflow measurement. Calculations will be performed with and without patients with a catheter at the time of assessment. The "with" analysis will assume a Qmax of 0 for any patient with a urinary catheter in place at the time testing. Mean and confidence intervals will be calculated using a t test (two-tailed, alpha=.05) against a null hypothesis of no change and a historical semi-quantitative comparison will be made with WATER study data or other studies.

6.4 Change in PVR

Change from baseline to 3 months in PVR. Calculations will be similar to those for Qmax, i.e., calculations with and without patients who are catheterized. For a patient with a urinary catheter in place at the time testing, the PVR value will be assessed as not applicable. Mean and confidence intervals will be calculated using a t test (two-tailed, alpha=.05) against a null hypothesis of no change and a similar semi-quantitative comparison will be made to WATER and/or other studies.

6.5 Change in PSA

Change from baseline to 6 months in PSA. Mean and confidence intervals will be calculated using a t test (two-tailed, alpha=.05) against a null hypothesis of no change.

6.6 Change in MSHQ-EjD

Mean change from baseline in MSHQ-EjD score at 3 months. Preservation of ejaculatory function is defined as a change non-inferior to -4 points. (Note that there is no published minimally clinically important difference for MSHQ-EjD.) A t test with 95% confidence intervals will be used to evaluate this hypothesis. If non-inferiority is concluded, a secondary test will calculate whether the change score represents statistical superiority.

6.7 Change in IIEF-5

Mean change from baseline in IIEF-5 (aka SHIM, part of IIEF-15) at 3 months. Preservation of erectile function, meaning a change score non-inferior to -6 points. An approach similar to that listed for MSHQ-EjD will be used.

6.8 Reintervention Rate

Re-intervention rate at 3 months. Re-intervention is defined as the need for additional prostate tissue resection following the index procedure. An exact binomial confidence interval (binom.test) will be calculated. No statistical hypothesis is proposed.

6.9 Urinary Catheter Discontinuation

The proportion of participants who were using a urinary catheter at any time 45 days prior to enrollment who are not using a catheter any time 45 days prior to the 3-month visit will be assessed. No statistical hypothesis is proposed.

7 ADDITIONAL ENDPOINT ASSESSMENTS

Additional exploratory statistical analysis will be done using standard statistical methods. Analyses of interest are:

1. Proportion/rate of all Clavien-Dindo classifications at month 3.

2. Changes in the following ratings across all time points: IPSS, IPSS-QOL, Incontinence Severity Index (ISI), IIEF-15, MSHQ-EjD, pelvic pain intensity score, dysuria frequency and intensity score, and uroflow measures. For this and other repeated assessments, repeated measures analysis of variance may be used to calculate change from baseline. Semi-quantitative comparisons will be made to results from WATER.

3. Relationship between baseline IPSS and IPSS change scores.

4. Duration of bladder catheterization.

5. Whether discharged home with a bladder catheter.

6. Change in hemoglobin from preoperative to postoperative. Additional analysis will be performed to see if prostate size, anesthesia type, number of Aquablation passes, and intravesical median lobes were predictors of bleeding loss.

7. Reoperation rate at 1 year for lack of efficacy. (This endpoint excludes reoperation for complications.)

8. Change in proportion of subjects using BPH medications (e.g., alpha blockers, alpha-reductase inhibitors).

9. Prostate size reduction from baseline to 3 months as measured TRUS. A t test will be used for the comparison.

10. Relationship between prostate size and procedure or resection times. Linear or other models will be used to explore relationships.

11. Proportion of subjects with device- or procedure-related adverse event.

12. Proportion of subjects in whom re-catheterization was needed between discharge after the index procedure and month 3. Re-catheterization is defined as the need to place a urinary catheter in the bladder for symptoms related to BPH (primarily inability to urinate). This excludes re-catheterization for study purposes or for purposes unrelated to LUTS.

13. Amount of irrigation fluid used intraoperatively (liters). A t test or Wilcoxon's test will be used to compare to values from WATER.

8 SUBGROUP ANALYSES

Subgroup analyses of primary safety and effectiveness endpoints, as well as selected secondary endpoints, will be performed for the following subgroups:

- baseline IPSS scores of <20 vs. ≥20
- baseline prostate size of <100g vs. ≥100g
- Age <65 vs. ≥65 years at baseline

9 POOLING

Data will be pooled across sites when performing statistical analysis. The justification for pooling includes the following:

- Study sites will be following the same protocol
- Study sites use the same device system
- Study sites follow the same instructions for use
- Study subjects are enrolled using identical criteria across sites
- Frequent contact with sites and monitoring of study data

Potential heterogeneity of results will be examined. For the primary effectiveness endpoint, analysis of variance will be used to determine heterogeneity of effect sizes across site. Heterogeneity will be assumed to be present if the site × treatment interaction p-value is <.05. Additional models may incorporate potential predictors of IPSS change scores, including baseline IPSS scores, procedure-related variables (ablation time, resection time). If these variables do not adequately explain heterogeneity across sites, mixed models will be used that assume variation of change scores or treatment effect across sites.

For the primary safety endpoint, heterogeneity will be assessed using a Mantel-Haenzel odds ratio test or similar test. Heterogeneity will be assumed to be present if the site × treatment interaction p-value is <.05. If evidence of non-poolability is found, baseline and procedural variables found to be different between sites will serve as predictors in a logistic or linear regression that also includes as predictors the treatment assignment, site, and site-by-treatment interaction. If these variables do not adequately explain heterogeneity across sites, mixed models will be used that assume variation of event rates across sites. Mixed models treat some factors (e.g., site ID) as a random effect and are often used in this situation.

10 CITATIONS

1. Barry, M. J. *et al.* The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J. Urol.* **148,** 1549–1557; discussion 1564 (1992).
2. Barry, M. J. *et al.* Benign prostatic hyperplasia specific health status measures in clinical research: how much change in the American Urological Association symptom index and the benign prostatic hyperplasia impact index is perceptible to patients? *J. Urol.* **154,** 1770–1774 (1995).