Statistical Analysis Plan (SAP)

Evaluating the Clinical Efficacy of Thulio vs. Holmium Laser Enucleation of the Prostate

Personnel

This statistical analysis plan was authored by Kamilla Abdurakhmanov, MS (biostatistician in the Department of Quantitative Sciences [QHS] at Cleveland Clinic) and approved by Olivia Hogue, PhD, MPH (biostatistics faculty in QHS), and Smita De, MD, PhD (Principal Investigator).

Biostatistician (Author)

Name: Kamilla Abdurakhmanov, MS

Signature: Kamilla A.

Date: 08/26/2025

Biostatistician Reviewer

Name: Olivia Hogue, PhD, MPH

Signature:

Date: 8-26-2025

Principal Investigator

Name: Smita De, MD, PhD

Signature:

Date: 8/28/25

Abbreviations and Definitions

BPH	Benign Prostatic Hyperplasia
IPSS	International Prostate Symptom Score
QoL	Quality of Life
PVR	Post-void residual
SHIM	Sexual Health in Men Score

Table of Contents

1. Introduction

This statistical analysis plan (SAP) describes the planned statistical methods to be used during the reporting and analysis of data collected under the Thulio vs. Holmium Laser Enucleation of the Prostate clinical trial. Any revisions to the protocol that impact the planned analyses may require updates to the SAP.

2. Study Objectives

2.1 Primary Objective

To evaluate the efficacy of the new Thulio laser by determining non-inferiority in hemostasis timing between the new laser and the standard Holmium laser among patients undergoing laser enucleation of the prostate for the surgical treatment of benign prostatic hyperplasia (BPH).

2.2 Secondary Objective

To compare the treatment efficacy of the new laser to the standard laser by enucleation time, active laser time, morcellation time, change in hemoglobin, applied total laser energy, enucleation efficiency (g/min), IPSS (International Prostate Symptom Score), QoL (Quality of Life), uroflowmetry/PVR (post-void residual), duration of catheterization, estimated blood loss, Sexual Health in Men Score (SHIM), 30 day complications, fiber burn back, surgeon estimation of collateral tissue injury, surgeon estimation of overall performance, volume of tissue removed, length of stay.

3. Study Design

This study is a prospective, single-blinded, randomized, parallel non-inferiority trial that will be performed at Cleveland Clinic in Ohio.

3.1 Randomization

A total of 50 study subjects will be allocated in a 1:1 ratio to either receive the new laser procedure or the standard laser procedure. Stratified block randomization with block sizes of 4 will be implemented using R software by a single biostatistician at Cleveland Clinic. Subjects will be stratified by prostate size (≤100 grams, >100 grams) where subjects within each prostate size group will fill one block with 2 subjects receiving the new laser and 2 receiving the standard, resulting in 12 blocks of 4 for a total of 48 patients. 6 blocks will be in each prostate size group. With a total sample of 50 subjects and block sizes of 4, the two remaining subjects will need to be randomly assigned using simple randomization. If both subjects fall into the same prostate size strata, they will be randomized using simple randomization. If the 2 subjects fall into different strata, they will be independently randomized using simple randomization. Details on sample size considerations are in section 4 of this SAP. An R program will generate uniform random variables that correspond to a fixed sequence of group assignments in a block ahead of time, which are then used to order the groups from the smallest to largest variable within the block. This method will ensure the intervention groups are highly balanced over time and the subject allocation is not easily predictable. The sequence will be in the Research Electronic Data Capture (Redcap) randomization module. Investigators will be able to write the patient's prostate size group and receive their allocation on the day of the surgery. An interim analysis will be conducted after 25 subjects have been recruited to determine if the groups are balanced by prostate size, or if more subjects need to be recruited for one prostate size strata. Imbalance between the groups on other baseline characteristics will be assessed using standardized mean difference.

3.2 Blinding

Study subjects will be blinded to the their randomized group assignment from the time of randomization to the completion of follow-ups. The surgical team cannot be blinded due to the nature of the study with laser devices appearing different. However, the surgical team will not know the randomization until the day of the surgery. The blinding will be broken once the study participants complete their follow-ups. Prior to this, the only permissible reason for unblinding will be to protect the safety or welfare of the subject as determined by the investigator, or at the request of the subject.

4. Sample Size

4.1 Design and general considerations

In a previous study comparing Moses 2.0 with non-Moses technology for HoLEP, hemostasis time in both studied groups was positively skewed with means that exceeded the medians (Nevo et al., 2021). Based on the distributional summaries, it appears that a few extreme measures may be to blame. In that study, the mean hemostasis time was 46% lower for the Moses 2.0 group, and the median was 27% lower than the reference group. For the current sample size estimation, we assumed that the hemostasis time followed a log-normal distribution, with a conservative estimate of the coefficient of variation (CV) of 0.66, which is slightly larger than the CV in the Moses 2.0 group in the referenced study (CV = 0.57). For purposes of generalizability, we considered scenarios where the new device would be up to 30%, 20%, and 10% faster than the standard, and where the two devices would be equal in their timing. Regions of non-inferiority between 20% and 35% were considered. Table 1 summarizes the minimum total sample size required to detect non-inferiority under these set of assumed differences and non-inferiority regions. Calculations assume 80% power and one-sided t-tests for lognormal data with a CV of 0.66 and alpha of 0.05. Power and sample size calculations were performed R software (version 4.5.1).

Table 1. Total Sample size per group to detect non-inferiority in hemostasis time.

	New Laser Performance Relative to Standard			
Non-inferiority Region (%)	**30% Faster**	**20% Faster**	**10% Faster**	**Equal**
20	34	56	110	272
25	30	48	86	182
30	26	40	68	132
35	24	36	56	102

One-sided t-test for lognormal data with 80% power and alpha = 0.05.

4.2 Determination

The sample size of 50 was selected based on feasibility and previous similar studies comparing laser technologies (Nevo et al., 2021). If we assume that the new device will be 20% faster on hemostasis time, and that a 25% non-inferiority region is used, then with 48 total patients we will have 80% power to prove non-inferiority. These calculations do not incorporate loss to follow-up, which is expected to be small given that the primary outcome is captured intraoperatively. Two patients were added to the sample size to account for potential loss to follow up or withdrawal.

5. Statistical Analyses

All statistical analyses for this study will be performed using R software (version 4.5.1).

5.1 Descriptive Statistics

Descriptive statistics will be presented for all clinically relevant baseline demographic, medical history, and clinical characteristic variables. Patient and procedure related characteristics will be summarized overall and by group using frequencies with percentages for categorical factors, means with standard deviations for normally distributed continuous measures, and medians with quartiles for other continuous measures. Mann-Whitney-U tests for continuous variables and chi-squared tests for categorical variables will be used to compare the patient characteristics between the intervention groups at a significance level of 0.05. Imbalance between the groups on baseline characteristics will be assessed using standardized mean difference.

5.2 Missing Data

Due to the nature of this trial, premature discontinuation and loss-to-follow-up are not expected, and data missing at random are expected to be minimal. Permanent missing data may occur due to subject withdrawal. No missing data is expected to be imputed.

5.3 Subject Disposition

The number of subjects who are enrolled and complete clinical follow-up will be summarized at the final follow-up visit. The number of subjects who complete the study or exit early (withdrawal, loss-to-follow-up, etc.) will be summarized by reason.

5.4 Analysis of Study Endpoints

Analyses will be performed using all-treated and per-protocol analysis. Differences in all-treated and per-protocol populations are expected to be small because patients are likely to attend standard follow-up. The populations are defined as follows:

All-treated population: Any subject randomized into the study and had surgery will be included in the analysis of the primary outcomes. Any patient who drops out before having surgery cannot be included.

Protocol-compliant population: Any subject who was randomized, completed surgery, and presented for standard follow up will be included in all primary and secondary outcome analyses.

5.4.1 Primary Endpoint

The primary endpoint is hemostasis time. A one-sided t-test with a non-inferiority margin of 0.25 will be performed for the log-transformed mean hemostasis time of the new and standard laser groups, using a significance level of 0.05, and a confidence interval for the mean difference will be generated. The null and alternative hypotheses are as follows:

H_0: $\mu \geq \mu_R + M_{NI}$
H_1: $\mu < \mu_R + M_{NI}$
Where μ is the mean hemostasis time in the group receiving the new laser, μ_R is the mean hemostasis time in the group receiving the standard laser, and M_{NI} is the margin of non-inferiority. If the analysis needs to be adjusted for confounders, a multiple linear regression model will be used with possible transformations or interaction terms as necessary.

5.4.2 Secondary Endpoints

The secondary endpoints are enucleation time, active laser time, morcellation time, change in hemoglobin, applied total laser energy, enucleation efficacy (g/min), IPSS, QoL, uroflowmetry,

PVR, duration of catheterization, estimated blood loss, Sexual Health in Men Score (SHIM), 30 day complications, fiber burn back, surgeon estimation of collateral tissue injury, surgeon estimation of overall performance, volume of tissue removed, and length of stay. These endpoints will be evaluated for superiority using two-sided tests to determine if the mean differences of the secondary endpoints are significantly larger in the laser intervention group than the standard, for variables where having a higher means is considered superior. Enucleation efficacy, volume of tissue removed, and surgeon estimation of overall performance would need to have a higher mean among patients in the treatment group compared to the standard group to be considered superior. Therefore, the null and alternative hypotheses are as follows:

$H_0: \mu_T \leq \mu_C$
$H_1: \mu_T > \mu_C$
Where μ_T is the mean for the new laser group and μ_C is the mean for the standard laser group. The enucleation time, active laser time, morcellation time, applied total laser energy, length of stay, surgeon estimation of collateral tissue injury, estimated blood loss, fiber burn back, 30 day complications, and duration of catheterization would have lower means in the treatment group compared to the standard to be considered superior. For those variables, the null and alternative hypotheses are as follows:

$H_0: \mu_T \geq \mu_C$
$H_1: \mu_T < \mu_C$

Multiple linear regression models will be used if the analysis needs to be adjusted for confounders, with possible transformations or interaction terms as necessary. These models will also adjust for baseline levels for outcome measures taken before and after the procedure (IPSS, QoL, SHIM, uroflowmetry/PVR, hemoglobin). Distributional assumptions of the outcomes will be evaluated using Shapiro-Wilk tests and graphically (residual, scale-location, and Q-Q plots). If departures from normality are observed, transformations will be considered as a primary approach, with nonparametric testing being used if remedies through transformation fail. Mean differences with 95% confidence intervals will be estimated for all secondary endpoints.

6. References

Labes D, Schütz H, Lang B (2024). _PowerTOST: Power and Sample Size for (Bio)Equivalence Studies_. R package version 1.5-6, commit ba8989075b03e226af55a8600990020e68cd1255, <https://github.com/Detlew/PowerTOST>.

Nevo, A., Faraj, K. S., Cheney, S. M., Moore, J. P., Stern, K. L., Borofsky, M., ... & Humphreys, M. R. (2021). Holmium laser enucleation of the prostate using Moses 2.0 vs non-Moses: a randomised controlled trial. *BJU international*, *127*(5), 553-559.