A Randomized Trial of Vaginal Surgery for Uterovaginal Prolapse: Vaginal Hysterectomy With Native Tissue Vault Suspension vs. Mesh Hysteropexy Suspension (SUPeR)

NCT01802281

Statistical Analysis Plan (SAP) for 3 Year Outcome

December 15, 2018

STATISTICAL ANALYSIS PLAN

PFDN Protocol Number 24P01:

A Randomized Trial of Vaginal Surgery for Uterovaginal Prolapse: Vaginal Hysterectomy with Native Tissue Vault Suspension vs. Mesh Hysteropexy Suspension

Short name: Study of **U**terine **P**rolapse Procedur**e**s- **R**andomized Trial "SUPeR"

SAP VERSION: Version 1.0

SAP DATE: December 15, 2018

SPONSOR: NICHD

PREPARED BY: RTI International.
3040 Cornwallis Rd
Research Triangle Park, NC 27709-2104

AUTHOR (S): Dennis Wallace
Ryan Whitworth
Scott Grey

Contents

LIST OF ABBREVIATIONS

AAS	Activities Assessment Scale
ACOG	American Congress of Obstetricians and Gynecologists
AE	Adverse event
BIS	Body Image Scale
BMI	Body mass index
BRCA	Breast cancer gene
CRADI	Colorectal-Anal Distress Inventory
CRAIQ	Colorectal-Anal Impact Questionnaire
DCC	Data coordinating center
DSMB	Data and safety monitoring board
FDA	U. S. Food and Drug Administration
GEE	Generalized estimating equation
HRQOL	Health-related quality of life
ICER	Incremental cost-effectiveness ratio
IRB	Institutional Review Boars
ITT	Intention to treat
LOT	Life Orientation Test
MAR	Missing at random
MCAR	Missing completely at random
MCS	Mental Component Scale of the SF12
NICHD	National Institute of Child Health and Human Development
OAB	Overactive bladder syndrome
PCS	Physical Component Scale of the SF12
PFDI-SF	Pelvic Floor Distress Inventory – Short form
PFDN	Pelvic Floor Disorders Network
PFIQ-SF	Pelvic Floor Impact Questionnaire – Short form
PGI-I	Patient Global Impression of Improvement
PGSC	Patient Global Symptom Control
PISQ-IR	Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire – Revised version
POPDI	Pelvic Organ Prolapse Distress Inventory
POPIQ	Prolapse Impact Questionnaire
PVR	Post-void residual
QALY	Quality-adjusted life year
QOL	Quality of life
SAE	Serious adverse event
SAF	Safety Population

SAP	Statistical analysis plan
SF-12	12 Question Short Form Health Survey
TVL	Total vaginal length
UDI	Urinary Distress Inventory
USLS	Uterosacral ligament suspension

1 BACKGROUND AND PROTOCOL HISTORY

The apex of the vagina (either cervix or the vaginal cuff) is thought to be the keystone of pelvic organ support. Loss of apical support is usually present in women with prolapse that extends beyond the hymen. At least half of the observed variation in anterior compartment support may be explained by apical support. Adequate support for the vaginal apex is thought to be an essential component of a durable surgical repair for women with advanced prolapse. Because of the significant contribution of the apex to anterior vaginal support, surgical correction of the anterior and posterior walls may fail unless the apex is adequately supported. A better term for the most common prolapse conditions would be "uterovaginal" or "cuff-vaginal" prolapse. Apical suspension procedures can broadly be separated into those performed transvaginally and those performed abdominally. Abdominal procedures are performed via laparotomy or using conventional laparoscopic or robotically assisted-laparoscopic techniques. While various abdominal and vaginal approaches exist, national data suggest that 80-90% of prolapse surgery is performed vaginally. Transvaginal mesh systems were introduced to improve native tissue vaginal repairs without entry into the peritoneal cavity. FDA data released July, 2011 noted that in 2010 approximately 300,000 women underwent surgical procedures to repair pelvic organ prolapse (POP); 1/3 used mesh, and 75% of mesh procedures were done transvaginally. This study is specifically addressing transvaginal surgical repair of uterovaginal prolapse.

The mini-protocol was approved by the PFDN Steering Committee on April 27, 2012 and the protocol was initially approved by the PFDN Steering Committee on July 20, 2012. Version 1.0 of the protocol was approved on December 17, 2012, after review by the DSMB, FDA and the Steering Committee. The protocol was first amended on May 13, 2013 to change the acronym of the study to SUPeR. The protocol was also amended on July 22, 2016. The purpose of this amendment was to clarify the length of participation of individual participants. Specifically, the amendment clarified that all participants should be followed every 6 months for 60 months (5 years). Language was also added to the statistical section of the protocol to a description of the secondary survival analysis that will generate estimates of treatment success at 48 and 60 months. The SUPeR protocol was amended on March 7, 2018. The purpose of the amendment was to explicitly state that all subjects should be unmasked by a study coordinator at the completion of their 60-month study visit.

A single formal interim analysis for efficacy was planned for this study when the last participant enrolled has completed the 24-month follow-up exam. The results of this analysis was reviewed by the PFDN Data and Safety Monitoring Board for recommendations about study continuation and the study continued to completion.

2 PURPOSE OF THE ANALYSES

This statistical analysis plan (SAP) contains detailed information about statistical analyses to be performed to assess the efficacy and safety of vaginal hysterectomy with native tissue vault suspension and mesh hysteropexy suspension in women experiencing uterovaginal prolapse. The results of these analyses will be included in the primary manuscript and a series of pre-planned secondary manuscripts and be provided to Boston Scientific Corporation for a report on the efficacy and safety of mesh-associated surgery required by the FDA under a 522 letter request. Added exploratory analyses may be performed to support further manuscript development. These analyses will not require an update to the SAP.

3 STUDY OBJECTIVES AND OUTCOMES

3.1 Study Objectives

3.1.1 Primary Objectives

The primary purpose of this randomized clinical trial is to compare the effectiveness and safety of two transvaginal apical suspension strategies for uterovaginal prolapse: a mesh augmented hysteropexy vs. vaginal hysterectomy and uterosacral ligament suspension. The primary aim for this trial is to determine whether treatment success in women with symptomatic uterovaginal prolapse undergoing transvaginal mesh augmented hysteropexy differs in women undergoing vaginal hysterectomy and native tissue cuff suspension at time points through 3 years. This study will test the null hypothesis that treatment success will not differ in women with symptomatic uterovaginal prolapse undergoing vaginal surgery with a synthetic mesh hysteropexy compared to women undergoing vaginal hysterectomy with native tissue vaginal suspension against the alternative hypothesis that success does differ for the treatment regimens. Operationally, the hypothesis will be tested using a two-sided log-rank test to test for a difference in the risk of failure across the two treatment arms at a 0.05 level of significance.

3.1.2 Secondary Objectives

The secondary aims for the SUPeR study, which will be addressed through specific planned secondary analyses and associated hypothesis tests and treatment effect estimates, are:

1. Secondary Efficacy Outcomes: To compare detailed anatomic and comprehensive functional outcomes (including prolapse, urinary, sexual, bowel and health related quality of life (HRQOL) in both groups.

2. Safety: To measure and compare safety, adverse events (including mesh erosion and exposure), pain, and need for subsequent procedures in both groups.

3. Predictors of poor outcomes: To determine if advanced prolapse, age, obesity, smoking, menopausal status, estrogens, previous prolapse surgery, and physical activity levels, alone or in combination, predict higher treatment failure.

4. Cost- effectiveness: To compare the cost effectiveness of the two surgical approaches and relate the difference in cost of care between the two groups to differences in health utilities and health–related quality of life.

5. Body image: To describe changes in body image as measured by a validated scale, the Body Image Scale (BIS). in a group of women undergoing mesh augmented hysteropexy or vaginal hysterectomy and to evaluate whether or not changes in sexual function are associated with changes in body image.

3.2 Outcomes

The **primary outcome** for the study is treatment failure at any point after the participant leaves the operating room; note that per protocol requirements, the patient cannot leave the operating room

as a treatment failure. A participant will be considered a treatment failure if any ONE of the following criteria is met:

1) Report of bothersome vaginal bulge symptoms (see definition below), or
2) Re-treatment for prolapse (surgery or pessary), or
3) Any prolapse measure (Ba, C, Bp) is beyond the hymen (i.e. >0 cm)

Bothersome vaginal bulge symptoms = positive response to Question 3 of the PFDI-20: Do you usually have a bulge or something falling out that you can see or feel in your vaginal area? AND any degree of bother. This single question has been identified to most accurately and reliably identify those women with POP. An affirmative answer to this question was 96% sensitive (95%CI 92-100) and 79% specific (95%CI 77-92) for prolapse beyond the hymen. The 1-week test-retest reliability was good (kappa .84).

Participants not considered a treatment failure for the primary outcome will be considered a treatment success.

A number of secondary outcome measures will be used to support the analyses for the secondary aims listed earlier. Outcomes that will be used in the planned analyses associated with each of the secondary aims include:

1. Anatomical measures of treatment efficacy in the two treatment arms obtained at 6-month intervals after surgery:
    a. Mean and median POPQ point (Ba, Bp, C) location measures postoperatively in the two treatment arms.
    b. Proportion of participants in each group with C > -1/2 TVL
2. Functional measures of treatment efficacy in the two treatment arms as obtained from the measurement at 6 weeks as well as those at 6 months post-surgery and every subsequent 6-month time period through 5 years post surgery:
    a. Participant impression of overall prolapse improvement at 6-month post-operative intervals as measured by Patients Global Impression of Improvement (PGI-I),
    b. Mean overall prolapse symptoms at 6-month post-operative intervals based on POPDI-6 scores.
    c. Duration of postoperative catheterization.
    d. Urinary function measured at 6-month post-operative time points using:
        i. Mean UDI-6 scores
        ii. Hunskaar Incontinence Severity Index
        iii. Risk of de novo voiding dysfunction
        iv. Risk of de novo incontinence
    e. Sexual functional measures obtained at 6-month post-operative time points including:
        i. PISQ-IR scale measures
        ii. Risk of de novo dyspareunia
    f. Bowel function obtained at 6-month post-operative time points using CRADI-8 scores

g. Quality of Life (QO)L measures obtained at 6-month post-operative time points including:
    i. General SF-12 physical component and mental component scales
    ii. Pelvic QOL as measured by PFIQ total score and three subscales
    iii. Functional activity as measured by the Activity Assessment Scale

3. Safety and tolerability of the two treatment regimens:
    a. Intraoperative safety measured in each of the treatment arms including:
        i. Operative time
        ii. Estimated blood loss
        iii. Proportion of participants with blood transfusion
        iv. Intra- and post-operative complications categorized using a modification of the Dindo Classification.
    b. Risk of adverse events on the two treatment arms as measured by proportion of participants with any of the following events:
        i. Mesh related complications: mesh exposure in the vagina or mesh erosion into another organ; note that level of complication will be characterized into based on the following response classification schema: (a) None or non-surgical medical intervention only; (b) Minor or intra-office surgical intervention; (c) Outpatient surgery; (d) Inpatient surgery
        ii. Rates of pain captured from the modified Surgical Pain Scale, pain medication use, and location of pain with Pain Mapping Instrument.
        iii. Pelvic infection
        iv. Risk of perioperative infections
        v. Risk of urinary tract infections
        vi. Vaginal infections with flora uncommon to the vaginal canal
        vii. Risk of any of De novo vaginal bleeding, atypical vaginal discharge, fistula formation, neuromuscular problems (including groin and leg pain)
        viii. Need for subsequent procedures- Any surgical or non-surgical treatment for pelvic floor disorders (including urinary incontinence, voiding dysfunction, defecatory dysfunction or fecal incontinence, recurrent prolapse, and dyspareunia/pelvic pain). Any subsequent uterine or cervical office or Operating Room procedure in hysteropexy group (e.g., cervical biopsy, LEEP, hysteroscopy, D and C, hysterectomy).
        ix. Subsequent uterine or cervical pathology
        x. Risk of vaginal scarring defined as: De novo vaginal scar requiring medical or surgical intervention, or adversely affecting quality of life.
        xi. Risk of Vaginal shortening, de novo dyspareunia, and
        xii. Risk of worsening dyspareunia with AE survey instrument.

c. Risk of important complications in each of the two arms at any point post-operatively. Important complications are defined as:

   i. Any Grade IIIb or greater Dindo complication, which will also include any intervention under a regional anesthetic. These concurrent or subsequent Operating Room interventions include but are not limited to: mesh removal, ureteral repair, abscess drainage, revision of vaginal stricture, operative hysteroscopy, or hysterectomy.

   ii. New onset (de novo) dyspareunia preventing vaginal intercourse

d. Intractable pelvic pain – defined as daily pelvic pain after the 6-week postoperative visit which significantly affects the participant's quality of life requiring ongoing management or is refractory to medical and physical therapy.

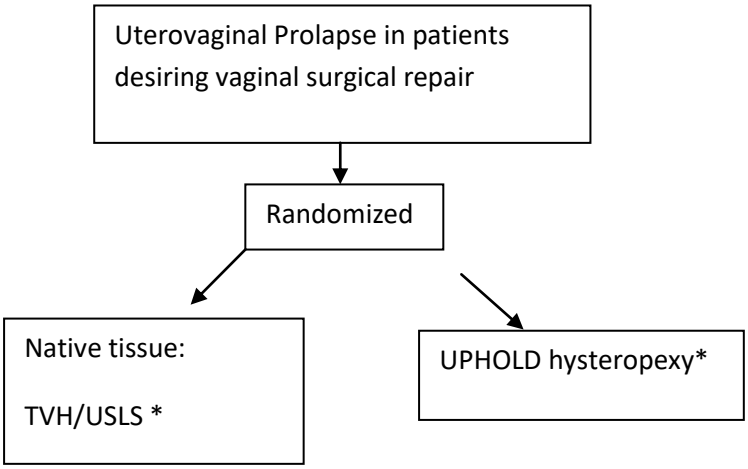
4. Key predictors of poor treatment outcomes defined as the effect of advanced prolapse, age, obesity, smoking, menopausal status, estrogens, primary vs. recurrent prolapse, and physical and functional activity as measured by the Functional Activity Scale on higher risk of treatment failure.
5. Cost-effectiveness of the two treatment regimens
   a. Clinical costs associated with participant's use of medical and non-medical resources related to urologic or gynecologic conditions will be collected during the follow up period. Direct and indirect costs of the treatment of apical pelvic organ prolapse with native tissue surgical repair or transvaginal mesh repair and women's preference for health states for improvement in pelvic organ prolapse will be estimated.
   b. Preference-based utility index determined from SF-6D.
6. Body image as measured by the *Body Image Scale* at 6-month post-operative intervals.

4 STUDY METHODS

4.1 Overall Study Design and Plan

The study is a multi-center, randomized, surgical trial of women with symptomatic uterovaginal prolapse desiring vaginal surgical treatment. The purpose of this study is to compare a vaginal hysterectomy with native tissue apical repairs with a non-trocar mesh hysteropexy repair. Eligible subjects will be randomized 1:1 to receive either vaginal hysterectomy with native tissue apical repairs or a non-trocar mesh hysteropexy repair for treatment of their uterovaginal prolapse. Each of the participants will be followed until 36 months after the last participant is randomized meaning that participants are expected to be followed for a period of 36 to 60 months to assess treatment outcome.

A study schematic is shown below:



4.2 Study Population

The study population is expected to comprise 180 participants (who were enrolled, consented, and randomized) as defined by the following eligibility criteria:

Inclusion Criteria:

1) Women aged 21 or older who have completed child -bearing
2) Prolapse beyond the hymen (defined as Ba, Bp, or C > 0 cm)
3) Uterine descent into at least the lower half of the vagina (defined as point C> -TVL/2))
4) Bulge symptoms as indicated on question 3 of the PFDI-20 form relating to 'sensation of bulging' or 'something falling out'
5) Desires vaginal surgical treatment for uterovaginal prolapse
6) Available for up to 60 month follow-up
7) Amenorrhea for the past 12 months from either menopause or endometrial ablation
8) Not pregnant, not at risk for pregnancy or agree to contraception if at risk for pregnancy (only applicable to the rare endometrial ablation patient)
9) Eligible for no cervical cancer screening for at least 3 years[40,41]

Exclusion Criteria:

1) Previous synthetic material (placed vaginally or abdominally) to augment POP repair
2) Known previous uterosacral or sacrospinous uterine suspension
3) Known adverse reaction to synthetic mesh or biological grafts; these complications include but are not limited to erosion, fistula, or abscess
4) Chronic pelvic pain
5) Pelvic radiation
6) Cervical elongation- defined as an expectation that the C point would be Stage 2 or greater postoperatively if a hysteropexy was performed. (Note: cervical shortening or trachelectomy is not an allowed intraoperative procedure within the hysteropexy treatment group).
7) Women at increased risk of cervical dysplasia requiring cervical cancer screening more often than every 3 years (e.g. HIV+ status, immunosuppression because of transplant related medications, Diethylstilbestrol (DES) exposure in utero, or previous treatment for cervical intraepithelial neoplasia (CIN)2, CIN3, or cancer)

8) Uterine abnormalities (symptomatic uterine fibroids, polyps, endometrial hyperplasia, endometrial cancer, or any uterine disease that precluded prolapse repair with uterine preservation in the opinion of the surgeon
9) Indication for ovarian removal (adnexal mass, BRCA 1/2 positivity, family history of ovarian cancer)
10) Current condition of amenorrhea caused by exogenous sex steroids or hypothalamic conditions.

4.3 Study Arm Assignment and Randomization

After consent is obtained from the participant and eligibility is determined, the participant will be randomized to one of the two treatment arms using a phone-based randomization system after the participant is in the operating room. The patient will be randomized to one of two apical procedures: either a vaginal hysterectomy with USLS or mesh strap vaginal hysteropexy (UPHOLD procedure). Other native tissue vaginal wall prolapse repairs (e.g. anterior colporrhaphy, posterior colporrhaphy, perineorrhaphy) will be allowed as needed.

Randomization (1:1 to the two treatment arms) will be performed using permuted blocks, with a block size that is known only to the DCC and will be stratified by site.

4.4 Masking and Data Lock

4.4.1 General Masking Procedures

The study surgeon is providing clinical care to enrolled participants, thus masking the surgeon to treatment allocation or participant symptoms is not practical or feasible, other than the allocation concealment prior to surgical randomization. The study surgeon will not be preforming the anatomic outcome assessments. It is our intent that when feasible and ethical, all outcomes assessors and the participant will be masked to the treatment allocation. All participants will be asked to remain masked to their treatment group for the duration of the study, although we recognize that unintentional unmasking by the participant may occur. Current ACOG recommendations allow women age 30 years or older with known recent negative cervical cytology and negative HPV testing to be screened no sooner than 3 years.40 Current US Preventative Services Task Force , American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology recommendations allow no screening for age <21, every 3 year screening for ages 21-29, every 5 year screening ages 30-65, and no screening for women older than 65. Therefore, all participants in this study should be able to go 3 years without screening and the overwhelming majority of participants in this study will be >30 and are allowed to have every 5 year screening. Unmasking is permitted for the rare amenorrheic woman with a uterus aged 18-30 eligible for this study who in the opinion of the study surgeon needs screening during the study. We recognize that this masking is for cervical cancer screening and women may still have regular exams. In the event the patient is about to have a pelvic exam or imaging study of her pelvis we will request the patient to remind her provider to not tell her about the status of the absence or presence of her uterus. Participants will be encouraged to see the study team for any gynecologic problems or evaluations during the course of the study. At every 6-month visit the participants will be queried if they are still blinded, and if not, what caused the unblinding and what group they think they are in? This will be done in a manner that reaffirms that blinding is preferred for the duration of the study. After the study is completed (maximum 60 months) participants will be queried as to what randomization arm they think they received and will then be notified of their uterine and mesh status. To minimize biases, subjective and most objective outcomes will be obtained by study nurses or coordinators masked to the procedure. POPQ measures will be obtained by co-investigators or different study nurses who will not be blinded to the surgical procedure because they will either see or not see a

cervix, but would be less biased than the operating surgeon towards over-reporting the anatomic surgical outcome.

Masking	Uterovaginal Prolapse Intervention
Participant	Yes
Study coordinator or study nurse	Yes
Telephone interviewer (if applicable)	Yes
Study surgeon	No
Anatomic Evaluator#	No#

to maximize masking, the anatomic evaluator not be the study coordinator or study nurse who should remain masked. The anatomic evaluator should be a Co-investigator, fellow, or other qualified nurse who did not perform the surgery.

4.4.2 Database Lock

Because the primary analyses are scheduled at the point that all study participants have completed their 36-month follow-up at a time that many participants are still in long-term follow-up, database lock will be accomplished in two phases. Because the primary analyses utilize a survival analysis approach, the first lock will include visits through 36 months on all study participants plus those 42-, 48-, 54-, and 60-month visits that occur on or before February 16, 2018. That date was selected in a masked fashion by the protocol team to facilitate efficient closure of appropriate files. Any visit-specific forms that meet these criteria will be locked during the first lock. However, forms that are log-based rather than visit-specific (e.g., protocol deviation form, AE form, and concomitant medication form) will not be locked as those forms will have accumulating data. Rather a snapshot of the data will be taken at some date after February 16, 2018, and any events that occurred prior to that date will be archived in a separate analysis database. The remainder of the study data will be locked and corresponding analyses will be completed once all participants have completed their 60-month follow-up period.

4.5 Study Flow Chart of Assessments and Evaluations

Measure	Base-line	Peri-op	6 wks	6 mo	12 mo	18 mo	24 mo	30 mo	36 mo	42 mo	48 mo	54 mo	60 mo
Demographic Info	X												
Medical History	X												
Operative and Perioperative review		X											
Postoperative recovery			X										
POPQ	X			X	X	X	X	X	X	X	X	X	X
AE review		X	X	X	X	X	X	X	X	X	X	X	X
Exam for mesh exposure			X	X	X	X	X	X	X	X	X	X	X
PFDI – Question 3	X		X	X	X	X	X	X	X	X	X	X	X
Functional Activity Scale	X		X	X	X	X	X		X		X		X
Surgical Pain Scale	X		X	X	X	X	X		X		X		X
PFDI- 20 (includes POPDI -6, CRADI-8, UDI-6)	X		X	X	X	X	X		X		X		X
PFIQ	X			X	X	X	X		X		X		X
PGI-_I				X	X	X	X		X		X		X
ISI	X		X	X	X	X	X		X		X		X
PISQ-IR	X			X	X	X	X		X		X		X
BIS	X		X	X	X	X	X		X		X		X
SF-12	X			X	X	X	X		X		X		X

5 ANALYSIS POPULATIONS

Intention to Treat (ITT) Population

The primary analysis population and the population for all secondary analyses will be the modified intention to treat population, which includes all randomized subjects. All subjects will be assigned to the arm to which they were randomized irrespective of treatment received.

Safety (SAF) Population

The safety population will comprise all subjects who received any study treatment grouped by actual treatment received, irrespective of treatment dose received.

6 SAMPLE SIZE DETERMINATION

This study is designed to compare the relative effectiveness of two transvaginal apical suspension strategies for uterovaginal prolapse: a mesh augmented hysteropexy vs. vaginal hysterectomy and uterosacral ligament suspension. Power and sample size calculations were generated to determine the sample size needed to test for treatment difference favoring the mesh augmented strategy (i.e., a superiority trial) across the study arms for a variety of assumptions about effect size and study follow-up time. For all power analyses, we assumed that failure in both arms follows an exponential survival model and that the native tissue repair has a success rate of 80% at 24 months. This 80% calculation is based on an assumed 85% anatomic success rate for uterosacral ligament suspension (Table 2) and an additional 5% failure rate based on symptom failure. In the only published Uphold series there is a 98 % 12 month anatomic success rate.32 Although no 2-year outcome data are available for the UPHOLD procedure, under an assumption of a constant hazard, the 1-year success rate of 98% yields an estimated 2-year anatomic success rate of 96%; combining this estimate with a similar 5% rate for symptom failure yields a 91% composite (anatomic and symptom) success rate. All analyses also assumed that statistical tests would be conducted with a Type I error rate of 0.05 with no adjustment for multiple comparisons. Sample size estimates assumed that the 2-year success rate in the mesh-augmented arm is in the range of 90% to 93% at 24 months (note that this represents a hazard ratio in the range of 0.33 to 0.47 under the assumed exponential survival model), that the enrollment time for the study is 2 years, the loss-to-follow-up on both arms is no more than 5% per year, and the total study duration from last participant enrolled to stopping the study for final analysis is 36 months with one interim analyses for efficacy when the last participant reaches the 24-month follow-up. Table 6 provides estimates of the power for the different assumptions for total sample sizes in the range of 160 to 300 participants.

Total Sample Size	**Power as a Function of 2-Year Effect Size—with base 2-Year Success of 80% and 36-Month Follow-up after last enrollment and a 2-year Enrollment Period**			
	∆= 0.10 HR=0.472	**Δ=0.11 HR=0.423**	**Δ=0.12 HR=0.374**	**Δ=0.13 HR=0.325**
160	0.87	0.93	0.96	0.99
180	0.89	0.95	0.98	0.99
200	0.92	0.96	0.99	>0.99
220	0.94	0.98	0.99	>0.99
240	0.96	0.98	0.99	>0.99
260	0.95	0.99	0.99	>0.99
280	0.98	0.99	>0.99	>0.99
300	0.98	0.99	>0.99	>0.99

Based on these calculations, a total of 180 participants will be randomized at a 1:1 ratio to the two treatment arms. This sample size will provide a power of 0.89 to detect an additive difference of 10% in the 2-year success rate (i.e. a hazard ration 0f 0.472) and a power of 0.99 to detect additive differences of 11% or greater in the 2-year success rate (hazard ratios of 0.423 or less).

7 STATISTICAL / ANALYTICAL ISSUES

7.1 General Rules

All statistical computations will be performed and data summaries will be created using SAS 9.2 or higher. If additional statistical packages are required, these will be discussed in the study report. For summaries of study data, categorical measures will be summarized in tables listing the frequency and the percentage of subjects in each study arm; continuous data will be summarized by presenting mean, standard deviation, median and range; and ordinal data will be summarized by only presenting median and range.

7.2 Adjustments for Covariates

Indicator variables for the study stratification of site will be included as covariates in most efficacy analyses performed for this study (details in section 9). Additionally, demographic and baseline characteristics for subjects and clinicians will be compared between study arms using analysis of covariance techniques for continuous measures, Mantel-Haenszel mean score test using standardized midrank scores for ordinal measures, and Cochran Mantel-Haenszel chi-square tests for general association for categorical measures. If sample sizes allow, these analyses will control for the study stratification factors. If these analyses suggest that substantial differences exist among arms, the use as covariates of these parameters on which the arms differ will be explored in secondary exploratory analyses of the efficacy data.

7.3 Handling of Dropouts and Missing Data

Standard procedures will be used to ensure that data are as complete and accurate as possible. The study was designed to obtain as much follow-up data as possible on all randomized subjects. In analyses, a full accounting will be made for all data items. Generally, missing data will initially be treated as randomly missing (either missing at random (MAR) or missing completely at random (MCAR) as appropriate for the analytic approach) with no data imputation. Because we anticipate that the amount of missing data will be minimal, we will use straightforward multiple imputation procedures based on baseline covariates and treatment group to assess sensitivity to the random missingness assumption for both primary and secondary analyses.

7.4 Interim Analyses and Data Monitoring

A single planned interim analysis is planned with the last randomized participant reaches the 24-month follow-up exam. The purpose of the interim analysis is to stop for benefit. Stopping boundaries will generated using Lan-DeMets alpha spending functions with the O'Brien-Fleming type approach. The alpha level for the interim analysis is expected to be about 0.003 and to allow for this interim look, the final test will be conducted at the 0.047 level of significance.

The DSMB will also review enrollment data and safety data at 3-month intervals and will have the responsibility of recommending to the Director of NICHD that the study be stopped for safety or futility. Each of these issues is addressed in the paragraphs below.

The interim safety analyses conducted by the DSMB will be conducted based on a review of any study deaths (note that the number of deaths in the population under study is expected to be minimal) and important complications as defined in Section 3.2. The important complication rate will be

calculated at 6 month intervals for both groups; the protocol committee considers that a true difference of more than 15 percentage points in the complication rates for the two surgical procedures to represent an important difference between groups. Stopping the study will be considered by the DSMB during their six-month reviews if they find compelling evidence based on point and interval estimates of the important complication rates in the two study arms that the true difference is 15 percentage points or greater. The DSMB is charged with determining the level of evidence that they consider compelling for stopping the study for safety reasons.

While the study is designed to demonstrate the superiority of either of the two treatment arms in comparison to the other, the comparison of the efficacy and safety outcomes for the two treatment arms has important public health implications even if a statistical difference in efficacy is not demonstrated. Consequently, the futility analyses will focus on failure to adequately enroll the study rather than on inability to show a difference in efficacy between the arms. The study is designed to enroll 180 participants over a 24 month period or 7.5 participants per month across the 8 participating sites. The DSMB will evaluate enrollment at 6-month intervals starting after the first site has achieved IRB approval. If at these 6-month reviews, the DSMB finds that the study is not enrolling at a rate of at least 5 participants per month (which would allow the study to enroll over a 3-year period) they will consider recommending stopping the study for futility.

7.5 Masked Data Review

A masked data review is planned prior to the data lock once all study participants reach the 36-month visit. All eligibility, protocol deviation, and visit window data will be reviewed by the protocol team in a masked fashion. Details of the results from the masked review will be documented in a SAP addendum once the review is complete. Specific items to be addressed in the masked data review are:

- Reason for study withdrawal and missed visits (adverse event, death, patient moved, lack of efficacy, etc. to determine whether the event was likely to be missing at random or informatively missing).
- Characterization of individual failures (by time and reason for failure) to ascertain failure time for the primary outcome and failure across time span for secondary analyses related to changing status of anatomical prolapse failure and bulge symptoms.
- Review of deviations (excluding missed/incomplete visits/assessments)
- Summary of interoperative complications from C3 in Form 9. The summary will include a count and percentage for each of the complications listed in the table as well as a table that shows the percentage of individuals have different numbers of complications. The summary will also include the distribution of Dindo scores, and the listing of complications that generated the Dindo scores.
- Duration of catheterization (categorical measure of none, less than 1 week, 1 to 2 weeks, 2 to 4 weeks and greater than 4 weeks).
- Visit windows
    - All baseline data will be used regardless of time of collection so long as the data were collected prior to treatment initiation. If more than one assessment of an outcome is reported within a visit window, the earliest assessment will be used. If no assessments are available for a visit (e.g., assessments were completed but none within study window), then an out of window assessment will be used for that visit so long as the out of window assessment does not also fall into a window for a different study visit. Rules for classifying visit membership for out of window assessments will be further determined as part of a review by the protocol team using masked data.
- UTI classification

7.6 Multicenter Studies

For this multicenter study, randomization of study participants was stratified within center. Consequently, for all model-based primary and secondary analyses, center will be included as a fixed effect in the models. As an ancillary analysis associated with the primary outcome we will examine descriptively whether the treatment effect varies across sites; however, no other analyses will assess site differences in treatment effect because sample sizes are inadequate to support evaluation of site-level effects.

7.7 Multiple Comparisons and Multiplicity

There is only one formal hypothesis test for this study. As such, a statistical test will be conducted at a 5% type I error rate (two-sided) for the primary efficacy measure, and no adjustments for multiplicity other than control for the single interim analysis will be made. All analyses of secondary outcomes are exploratory in nature; therefore, p-values and confidence intervals are provided for descriptive purposes only. All p-values provided for any baseline and demographic characteristics and safety parameters will be for descriptive purposes only. As such, unless otherwise specified, p-values presented will be on a per analysis basis, with no further control for multiple tests.

7.8 Examination of Subgroups

No a priori subgroup analyses were defined in the protocol.

7.9 Assessment Windows

Baseline assessments were to be completed no longer than 3 months prior to surgery with assessments repeated if participant surgery is delayed for over 3 months. All other visits were completed at 6-month intervals with a ± six-week window around the visit. Coordinators attempted to complete all follow-up visits, even if they couldn't be completed within window. For both primary and secondary analysis, decisions about how to treat out-of-window visits will be made during the masked data review prior to unmasking data.

8 STUDY SUBJECT CHARACTERIZATION

8.1 Participant Disposition

Participant eligibility status will be summarized and listed by study arm and overall disposition of study participants will be described using a standard cohort diagram. The number of subjects randomized; completing or discontinuing from study therapy; completing each 6-month follow-up visit will be summarized by study arm. Reasons for study treatment discontinuation and study withdrawal will be listed.

8.2 Study Treatment Exposure And Compliance

Because of the surgical nature of the intervention, treatment exposure and compliance are not anticipated to be an issue for this study.

8.3 Protocol Deviations

Protocol deviations are identified via automated checks of the clinical database and reported by site study coordinators in the study data management system. Protocol deviations will be listed by site with information such as type of deviation, time of occurrence, and reason. Incidence rate of protocol deviations will also be summarized overall and for each protocol deviation category by site.

- Incidence rate of protocol deviations will be calculated as: number of deviations divided by the number of subject months at the site

8.4 Demographic and Baseline Characteristics

Demographic and baseline clinical characteristics for the study participants will be summarized by study arm using the general analysis rules describe above. Variables of interest include: age (years), parity, gravidity, race and ethnicity, marital status, education level (classified as binary variable as having some college or greater or no college education), health insurance status (private only, Medicare/Medicaid only, combination of both), smoking status (never, previous, current), prior prolapse surgery, BMI, and baseline levels of all QOL measures.

9 EFFICACY ANALYSES

9.1 Overview of Efficacy Analyses Methods

All efficacy analyses will be performed using the ITT population unless otherwise specified. All efficacy variables will be listed by subject within study center and assessment time. The data will be summarized by treatment group.

The primary efficacy outcome measure is a binary measure of treatment failure/success as determined at 6-month time intervals post-intervention, and the primary analytic approach will utilize a survival model to evaluate differences in time to treatment failure between the two treatment arms. Generally, all secondary efficacy variables will be collected longitudinally across this study at six-month. Consequently, secondary analyses will be conducted using appropriate models for these correlated data collected across time. Specifically, continuous outcome variables will be analyzed using linear mixed models and binary outcome variables will be analyzed using robust Poisson regression models with a log link. Most models will include terms for treatment, time, and the treatment by time interaction and for the stratification variables included in the study design (site). Additional details are provided in the specific sections below.

9.2 Efficacy Variables

Variable	Type	Definition
Primary Outcome		
Treatment failure/success obtained at 6-month intervals for a period of 3 to 5 years after surgery	Binary	Treatment failure will be defined as the occurrence of any ONE of the following criteria: (1) report of bothersome vaginal bulge symptoms (see definition below); (2) re-treatment for prolapse (surgery or pessary); or (3) any prolapse measure (Ba, C, Bp) is beyond the hymen (i.e. >0 cm). Bothersome vaginal bulge symptoms are defined as positive response to Question 3 of the PFDI-20: Do you usually have a bulge or something falling out that you can see or feel in your vaginal area? AND any degree of bother.
Secondary Efficacy Outcomes		
Anatomical measures of key pelvic floor metrics (Ba, Bp, and C) obtained at 6-month intervals post-operatively	Continuous	Data measurements obtained directly from the POPQ exam.
Anatomical measure of cervical or cuff location relative to total vaginal length	Binary	Obtained by comparing the cervical or cuff location (Point C) to the total vaginal length (TVL) from the POPQ exam obtained at 6-month post-operative intervals. A participant will be defined as positive at any time point if C > -1/2 TVL.
Prolapse symptoms at 6-month post-operative levels as defined by	Binary	The PGI-I is a single question form collected every 6 months that asks for a response to the question: “Check the number that best describes how your post-operative condition is now, compared with how it was before you had the surgery.” The response is a 7-level Likert scale from 1 Very much better to 7 Very much worse. If the participant response is either 1 (very much better), 2 (much better), and 3 (better) then the indicator will be coded as Yes. If the participant response is 4 or greater, the indicator will be coded as No. If the PGI-I is missing for a visit then the indicator will be coded as missing for that visit.

Variable	Type	Definition
Prolapse symptoms at 6-month post-operative levels as defined by POPDI-6 scores from the PFDI-SF	Continuous	The PFDI-SF Pelvic Organ Prolapse Distress Inventory (POPDI) scale will be computed at 6 month intervals using standard scoring algorithms described by Barber (2006) to create a measure scaled 0 to 100 at each time point based in data collected on the PFDI. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.
Duration of post-operative catheterization	Continuous	Time to removal of the catheter post-operatively obtained directly from CRF data.
Urinary functional distress as measured PFDI-SF Urinary Distress Inventory scale at 6-month post-operative intervals	Continuous	The PFDI-SF UDI scale will be computed at 6 month post-operative intervals using standard scoring algorithms described by Barber (2006) to create a measure scaled 0 to 100 at each time point. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.
Incontinence severity at 6-month post-operative intervals as measures by the Hunskaar Incontinence Severity Index	Ordinal	The ISI scale will be computed at 6-month post operative intervals using the standardized scoring algorithm to generate an ordinal scale (no incontinence or incontinence levels of slight, moderate, severe, or very severe) as described by Sandvik (2000).
Risk of de novo voiding dysfunction defined at 6-month post-operative intervals (evaluated in aggregate fashion)	Binary	Defined as any incidence of a specific voiding dysfunction based on patient-supplied AE information. Rather than reporting data at six-month intervals, results will be aggregated across the follow-up period based on decision at masked review.

Variable	Type	Definition
Risk of de novo incontinence defined at 6-month post-operative intervals (evaluated in aggregate fashion)	Binary	Defined as any incident case of new or worsening stress urinary incontinence or new or worsening urgency urinary incontinence as reported on either the Patient-reported AE forms or on the complication form obtained at six-month intervals. Based on masked data review, results will be aggregated across the reporting period rather than evaluated at six-month intervals
PISQ-IR not sexually active – partner related subscale score change from baseline	Continuous	The PISQ-IR not sexually active – partner related subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q2a, Q2b (1=strongly agree,… 4=strongly disagree). If there is more than 1 missing response then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are not sexually active.
PISQ-IR not sexually active – condition specific subscale score change from baseline	Continuous	The PISQ-IR not sexually active – condition specific subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q2c, Q2d, Q2e (1=strongly agree,… 4=strongly disagree). If there is more than 1 missing response then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are not sexually active.

Variable	Type	Definition
PISQ-IR not sexually active – global quality subscale score change from baseline	Continuous	The PISQ-IR not sexually active – global quality subscale score will be at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q4a, Q4b, Q5a, and Q6 using reverse scores for all but Q5a (Q4a and Q4b are likert scales of 1 to 5; Q5a: 1=strongly agree,… 4=strongly disagree; Q6: 1=Not at all, …, 4=a lot). If there are more than 2 missing responses then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are not sexually active.
PISQ-IR not sexually active – condition impact subscale score change from baseline	Continuous	The PISQ-IR not sexually active – condition subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q3, Q5b, Q5c using reverse scores for Q3 (Q3: 1=Not at all,…,4=a lot; Q5b, Q5c: 1=strongly agree,… 4=strongly disagree). If there is more than 1 missing response then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are not sexually active.
PISQ-IR sexually active – arousal, orgasm subscale score change from baseline	Continuous	The PISQ-IR sexually active – arousal, orgasm subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q7, Q8a, Q10, and Q11 using reverse scores for Q11 (Q7, Q8a, Q11: 1=never, …,5=[almost] always; Q10: 1=much less intense,…5=much more intense; check box response to Q11 = 1). If there are more than 2 missing responses then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.

Variable	Type	Definition
PISQ-IR sexually active – condition specific subscale score change from baseline	Continuous	The PISQ-IR sexually active – condition specific subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q8b, Q8c, Q9 using reverse scores for all (Q8b, Q8c, Q9: 1=never, …,5=[almost] always). If there is more than 1 missing response then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.
PISQ-IR sexually active – partner related subscale score change from baseline	Continuous	The PISQ-IR sexually active – partner related subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q13, Q14a, Q14b using reverse scores for Q14a and Q14b (Q13: 1=all of the time,…4=hardly ever/rarely; Q14a, Q14b: 1=very positive, …,4=very negative). If there is more than 1 missing response then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active and have a sexual partner.
PISQ-IR sexually active – desire subscale score change from baseline	Continuous	The PISQ-IR sexually active – desire subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q15, Q16, Q17 using reverse scores for Q16 and Q17 (Q15: 1=Never,…5=Always; Q16: 1=Daily,…,5=Never; Q17: 1=Very high,…, 5=Very low or none at all). If there is more than 1 missing response then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.

Variable	Type	Definition
PISQ-IR sexually active – condition impact subscale score change from baseline	Continuous	The PISQ-IR sexually active – condition impact subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q18, Q20b-d using reverse scores for Q18 (Q18: 1=not at all,…4=a lot; Q20b-d: 1=strongly agree,…,4=strongly disagree). If there are more than 2 missing responses then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.
PISQ-IR sexually active – global quality rating subscale score change from baseline	Continuous	The PISQ-IR sexually active – global quality rating subscale score will be computed at 6-month intervals using standard scoring algorithms. Specifically, sum the scores for questions Q19a-Q19c, Q20a using reverse scores for Q19a-Q19c (Q19a-c: 1=satisfied,…5=dissatisfied; Q20a: 1=strongly agree,…,4=strongly disagree). If there are more than 2 missing responses then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at Month 1 and the score at baseline. If data at any month are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.
Risk of de novo dyspareunia defined at 6-month post-operative intervals (aggregated across post-surgical reporting period)	Binary	Defined via two different mechanisms: (a) as an event of dyspareunia as listed on AE form or the 6-month complication evaluation form; or (b) assessed via reporting on the PISQ-IR instrument. Note that because on masked review women moved in and out of the dyspareunia category after surgery, the decision was made to evaluate all incidents of dyspareunia irrespective of de novo status.
Bowel function as measured by the PFDI-SF Colorectal-Anal Distress Inventory scale at 6-month post-operative intervals	Continuous	The PFDI-SF Colorectal-Anal Distress Inventory (CRADI) scale will be computed at 6 month post-operative intervals using standard scoring algorithms described by Barber (2006) to create a measure scaled 0 to 100 at each time point. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.

Variable	Type	Definition
Physical component summary scale from SF-12 (PCS-12) at 6-month post-operative intervals	Continuous	PCS-12 will be computed at 6 month post-operative intervals using standard algorithms from the SF-12 developer. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.
Mental component summary scale from SF-12 (MCS-12) at 6-month post-operative intervals	Continuous	MCS-12 will be computed at 6 month post-operative intervals using standard algorithms from the SF-12 developer. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.
PFIQ-SF Urinary Impact Questionnaire (UIQ-7) at 6-month post-operative intervals	Continuous	The PFIQ-SF Urinary Impact Questionnaire (UIQ-7) score will be computed at 6 month post-operative intervals using standard scoring algorithms described by Barber (2006) to create a measure scaled 0 to 100 at each time point. If data at any month are missing, the outcome variable will be coded as missing. All longitudinal analyses or time-specific analyses will use all available data.
PFIQ-SF Colorectal-Anal Impact Questionnaire (CRAIQ-7) at 6-month post-operative intervals	Continuous	The PFIQ-SF Colorectal-Anal Impact Questionnaire (CRAIQ-7) score will be computed at 6 month post-operative intervals using standard scoring algorithms described by Barber (2006) to create a measure scaled 0 to 100 at each time point. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.
PFIQ-SF Pelvic Organ Prolapse Impact Questionnaire (POPIQ-7) at 6-month post-operative intervals	Continuous	PFIQ-SF Pelvic Organ Prolapse Impact Questionnaire (POPIQ-7) score will be computed at 6 month post-operative intervals using standard scoring algorithms described by Barber (2006) to create a measure scaled 0 to 100 at each time point. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.
PFIQ-SF Total Score at 6-month post-operative intervals	Continuous	PFIQ-SF Total Score will be computed at 6 month post-operative intervals by summing the 3 subscales defined above to create a total score with a possible 300 points. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.

Variable	Type	Definition
Functional activity levels measured at 6-month post-operative intervals using the Activities Assessment Scale (AAS)	Continuous	The Activities Assessment Scale (AAS) score will be computed at 6 month post-operative intervals using standard scoring algorithms described by McCarthy (2005) to create a measure scaled 0 to 100 at each time point. If data at any month are missing, the outcome variable will be coded as missing for that month. All longitudinal analyses or time-specific analyses will use all available data.

9.3 Primary Analysis Methods

The primary efficacy outcome measure is a binary measure of treatment failure/success as determined at 6-month time intervals post-intervention. Surgical failure rates will be compared graphically across the two treatment arms with Kaplan-Meier type plots, with survival curves computed using the product-limit estimator appropriately extended for interval estimation. While the original protocol indicated that a log-rank test would be use for the primary analysis, additional data that have come to light since the development of the protocol led to a masked decision to modify the primary analysis approach as specified in detail below.

This analysis contained several limitations that became apparent upon the studies initiation that makes revision necessary. First, the survival analysis methods initially proposed were developed for right censored data, while the outcome data for the bulge symptoms and anatomic failure is collected at the six 6-month follow-up visits, making the data a combination of interval censored and right censored. Second, results from long term follow-up of the OPTIMAL trial showed that the proportional hazards assumption is not likely hold over the 60-month follow-up time. Third, recent research has shown that it's possible for patients with uterovaginal prolapse to move in and out of "treatment failure" as defined above, such that these repeated failures are ignored in the defined analysis. The analysis proposed are selected to address these limitations.

The first two sections below describe a modified analytic approach that treat participants in a manner consistent with the original analysis design in that study participants that meet the criteria for treatment failure a remain treatment failures until they are lost to follow-up or the study ends. The proposed model-based and nonparametric analytic methods address interval censoring and non-proportional hazards in the assessment of the study's primary hypothesis. The model-based approach will serve as the primary analytic method for testing the efficacy of the two strategies for uterovaginal prolapse as it can control for study design and potentially confounding variables, while the nonparametric method will serve a confirmatory role, particularly in the masked analysis as detailed below. The third section focuses on modeling recurrent events to better understand the impact of the interventions over time. The analytic methods for estimating survival curves will first be conducted in a blinded fashion on all subjects, ignoring treatment assignment as a part of a masked analysis, then comparing subjects randomized their interventions in an unmasked analysis.

Model-based analysis:

To appropriately account for the interval censored nature of the follow-up data, where the proportional hazards assumption is unlikely to hold, and control for study design and potentially confounding variables in assessing intervention effects, a piecewise exponential (PWE) survival model with mixed effects will be employed. The PWE survival model is a model in which the time scale is divided into K periods and the baseline hazard function, $h(t)$, is assumed to be constant within each period such that $h(t) = \lambda_k \exp(\beta \mathbf{X})$, where λ_k is the hazard function for period k, $\mathbf{X}$ matrix of explanatory variables with vector β coefficients (Breslow 1974). This piecewise approach effectively eliminates the need to specify the distribution of the hazard function and satisfies the proportional hazards assumption if the baseline hazard is constant within each period. The PWE survival model has been shown to be equivalent to a Poisson regression model if the baseline hazard function is constant as a function of time [i.e. $h(t) = \lambda(t) = \lambda$] (Laird & Olivier, 1981). This means that the PWE survival model can be fitted with a generalized linear mixed model (GLMM) by taking survival data consisting of observed survival time t_i for the i-th subject with a treatment failure indicator d_i within the k-th period, so that t_{ik} denotes survival time and d_{ik} the occurrence of treatment failure for the i-th subject in period k, creating a number of pseudo-observations, one for each combination of subject and period. A PWE survival model can then be fit by handling treatment failure indicators as if they were Poisson observations with means $\mu_{ik} = \lambda_{ik} t_{ik}$, where λ_{ik} is the hazard for the i-th subject in period

k. Also needed is an offset variable denoting a logarithm of the time at risk during each period (Crowther et al., 2012). It is important to note that PWE survival models do not assume that the d_{ik} have independent Poisson distributions as a subject who has treatment failure in period *k(i)*, must have had no failure in all prior periods. Rather, each *i*-th subject across period covariance is modeled as a random effect in the GLMM.

The GLMM of the PWE survival model will be is specified with (1) a Poisson distribution, (2) an offset variable denoting the logarithm of the exposure time during a given period, and (3) subject-specific random intercepts to incorporate within-subject homogeneity, using the SAS GLIMMIX procedure in SAS STAT version 14.2 (SAS Institute Inc. 2016). The study's follow-up time will be divided into three periods: first, before and up to 12-months; second, 13 months to 30 months; and third, 31 months and beyond. The choice cut points will be evaluated during the masked analysis with a plot of the NPMLE survival function (described below) and adjusted if necessary. As it is possible for subjects to have a right censored treatment failure at the beginning of a period, resulting in an exposure time of zero and a offset variable that is undefined, subjects experiencing a treatment failure at the beginning of a given period will be assigned an exposure time of 0.5 months and an offset variable of log(0.5). The unmasked analysis will use the GLMM specified above to generate an overall test of the failure hazard ratio across the two interventions using a two-sided hypothesis test with an overall type I error rate of 0.05, controlling for the design variables of site the subject was recruited into. Also, the point and interval estimates of 12-, 24- and 36-month treatment failure rates on each intervention arm as well as hazard ratios between the two arms will be provided. Additional models will be used to evaluate potentially confounding variables of subject age and prior prolapse surgery.

Nonparametric analysis:

This analysis will employ nonparametric maximum likelihood estimation (NPMLE) to estimate survival functions in a similar manner as the Kaplan-Meier estimator for right censored data, but also employs a combination of the EM and iterative convex minorant (ICM) algorithms to estimate the survival function in non-overlapping intervals (Wellner and Zhan, 1997). Standard errors of the survival curves and the covariance matrix for the generalized rank statistic will be generated with the multiple imputation method developed by Sun (2001). The overall test for the difference in the two interventions survival curves will be conducted with the generalized log-rank test (Huang, Lee, and Yu, 2008). In addition, point estimates and 95% confidence limits will be estimated at 12-, 24- and 36-month intervals. The nonparametric survival analysis methods detailed above will be carried out using the ICLIFETEST procedure. The advantage of this approach is that it makes no assumptions about the distribution of the interval centered data and permits estimated survival curves to be non-monotonic (no proportional hazards assumption), although it is unable to adjust for study design and potential confounders. This analysis will primarily be used to confirm the results of the model-based analysis, particularly the specification of the piecewise survival function in the masked analysis.

Analysis of multiple treatment failures:

Data from other trials that emerged after this study was designed have shown that patients with uterovaginal prolapse can move in and out of the state of "treatment failure" after surgical repair, and that the standard survival methods employed above ignore these repeated events; as such the survival approach may underestimate the impact of the interventions in the treatment of uterovaginal prolapse. This concern is specifically relevant for the treatment failure criteria anatomic failure and reported bothersome vaginal bulge symptoms, which are measured repeatedly at each of the follow-ups while retreatment with surgery or pessary result in a permanent failure. Below are two analytic approaches that focus on answering two questions that may be of practical interest to clinicians: (1) is there a difference in the prevalence of subjects meeting failure criteria (anatomic failure, bothersome

symptoms or retreatment) at 12-, 24- and 36- month follow-up times between the interventions; and (2) is there a difference in the rates of anatomic failure and bothersome bulge symptoms over follow-up time between the two interventions.

For the first question, at 12-, 24- and 36-month follow-up times, the proportion of subjects meeting specific failure criteria (anatomic failure, bothersome bulge symptoms, retreatment, no failure) will be compared across intervention groups using the Fisher-Freeman-Halton's exact test for r x c tables as proportion of subjects receiving retreatment will likely be small (Lydersen et al. 2007). If a subject meets more than one failure criteria, she will be classified as follows: retreatment and any combination of anatomic failure and bothersome bulge symptoms will be classified as retreatment; no retreatment but both anatomic failure and bothersome bulge symptoms will be classified as anatomic failure. Additionally, within each follow-up time, the odds of subjects being classified as treatment failures at that follow-up time (not being retained as failures if they met failure criteria at an earlier follow-up) will be compared across treatment groups with logistic regression models that will control for site, age and prior prolapse surgery.

Rates of anatomic failure and bothersome bulge symptoms over the follow-up time will be modeled using a GLMM (a shared frailty model), fitted by Gaussian quadrature in SAS NLMIXED procedure, where a piecewise baseline hazard function will again be employed to generate a parametric model that eliminates the need to specify the distribution of the hazard function and satisfies the proportional hazards assumption if the baseline hazard is constant within each period (Liu & Huang, 2008). The model will be specified with a period indicator variable for each follow-up; variables for treatment failure and for the length of time in each follow-up period; study design (site) indicator variables; an intervention indicator; and as having a normal distribution for the subject random intercept.

Note that this model uses each 6-month follow-up interval as a period in the piecewise model. This is primarily to assure that within each period, there is only one treatment failure/success indicator. It is possible to combine multiple intervals into periods when treatment failure become sparse. This can be done in the masked analysis. The unmasked analysis will generate an overall test of the failure hazard ratio across the two interventions using a two-sided hypothesis test with an overall type I error rate of 0.05, controlling for the design variables of site the subject was recruited into. Also, the point and interval estimates of 12-, 24- and 36-month treatment failure rates on each intervention arm as well as hazard ratios between the two arms will be provided. Additional models will be used to evaluate potentially confounding variables of subject age and prior prolapse surgery.

All analyses will be implemented using SAS Version 9.3 or later.

9.4 Secondary Efficacy Analysis Methods

9.4.1 General Approach

As outlined in the above sections, the secondary outcome variables comprise a combination of binary, ordinal, and continuous measures that are collected at 6-month intervals through post-operative Month 24 and at either 6- or 12-month intervals thereafter. These outcome measures will be summarized in both tabular and graphical fashion by treatment arm over the full follow-up period, although the focus of the analysis will be on the first 36 months of the post-operative period. We will then use model-based approaches appropriate for the variable type (linear mixed models for continuous variables and GEE extensions of robust Poisson regression models for binary outcomes) to compare the outcomes for the two treatment arms across time.

Because the primary scientific interest for the study is in the trajectories of various the outcomes across time and because information about the patterns of these trajectories is limited, a longitudinal approach incorporating all time points will generally be employed. Specifically, the unit of analysis will be the participant visit, where visits are defined as in Section 4.5 starting 6 months post operatively, and the models will incorporate treatment group, time, and treatment group by time, as well as site, the randomization stratification variable, as fixed effects in the model with time (or visit) will be treated as a categorical variable to incorporate the expected non-linearity of time effects. Participant will be included as a random effect in the linear mixed model with an assumed compound symmetric covariance structure across the post-operative observation period and this same compound symmetric structure will be assumed in the GEE model with observations clustered within participant. The general structure of the SAS code that will be used for the mixed model (using the POPDI-6 as an example) is summarized below.

```
Proc Mixed data=SUPeR;
WHERE visit le 36; *** Constrains analysis to first 36 months of follow-up;
Class treatment visit center participant;
Model POPDI-6=treatment visit treatment*visit center/solution;
Random participant;
Lsmeans treatment*visit/ pdiff CI;
Contrast 'POPDI Outcome' treatment 7 -7 treatment*month 1 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 -1;
```

As suggested by the SAS code, the primary hypothesis will be tested using the F-test from the Contrast statement, which tests whether mean POPDI across 36 months differs for the two treatment arms. Note that the LSMEANS statement will provide by visit estimates of the post-operative POPDI6 score for each treatment arm from Month 6 through Month 36. These estimates will allow a description of the difference in trajectory of treatment effect in the two arms and will provide descriptive measures of differences between the two arms at each study visit. If the outcome is binary rather than continuous, the SAS procedure GENMOD will be used with a Poisson distributional assumption to generate comparable comparisons between the treatment arms in which the difference is assessed using a relative risk rather than mean difference metric.

9.4.2 Comparison of Anatomical Measures of Treatment Efficacy

One component of the first secondary efficacy aim is to compare anatomical measures of treatment efficacy across the two arms. For the 3 continuous measures (Ba, Bp, and C), the linear model described above will be used to generate mean estimates of each of these 3 metrics for the two treatment arms at each of the post-operative visits. For the binary measure of C > ½ TLV, large-sample chi-square tests will be used to compare incidence at different time points.

9.4.3 Comparison of Functional Measures of Treatment Efficacy

A second component of the secondary efficacy aim for this study is to compare multiple functional and quality of life for the two treatment regimens. As outlined in Sections 3.2 and 9.2, outcomes associated with this aim will include a number of continuous (e.g., scales from the PFDI-SF, PFIQ-SF, PISQ-R and SF-12 scales) and binary indicators (global improvement and de novo risk of voiding dysfunction, incontinence, dyspareunia). While the analyses for these secondary outcomes will produce point and interval estimates as well as p-values from hypotheses, all analyses are considered to be descriptive rather than inferential, so p-values will be interpreted as measures of evidence of descriptive differences in the treatment arms, not formal inferences.

Analyses for the continuous outcome measures associated with this efficacy aim will be analogous to those described above. For these continuous secondary outcomes, additional models that account for covariates found to be potentially confounded with treatment in the preliminary descriptive analysis

(specifically any baseline or clinical values found to be associated with treatment at the 0.10 level) will be fit to ensure that these potentially confounding effects don't change treatment effect conclusions.

Analyses for the binary outcome measures associated with this efficacy aim will also be analogous to those described for the primary analyses in that they account for the longitudinal structure of the data, but they will be modified to account for the binary nature of the outcome measure and for any differences in outcome timing. Analyses will be conducted using a robust Poisson regression model as implemented through a generalized linear model with an assumed Poisson distribution and log link. The model structure will be comparable to that described above with each model including fixed effects treatment group, visit (as a categorical measure), treatment by visit interaction and, site. Participant will be handled as a clustering variable with an assumed compound symmetric covariance structure within participant. However, p-values and interval estimates of effect sizes will be obtained using GEE extensions of the generalized linear model using the Liang and Zeger sandwich estimator. For these outcomes, additional models that account for covariates found to be potentially confounded with treatment in the preliminary descriptive analysis (specifically any baseline or clinical values found to be associated with treatment at the 0.10 level) will be fit to ensure that these potentially confounding effects don't change treatment effect conclusions.

9.4.4 Predictors of Poor Treatment Effect

Two approaches will be utilized to evaluate predictors of poor treatment effect. First a binary outcome of treatment failure over a 36-month post-operative period will be created. Contingency tables and univariate logistic regression models will be used to determine whether this binary outcome is related to the potential predictors of advanced prolapse, age, obesity, smoking, menopausal status, estrogens, primary vs. recurrent prolapse, and physical and functional activity. Any covariates that are related to the outcome measure at a p-value of 0.10 will be utilized in a multivariable logistic regression model to develop an overall predictive model for treatment failure.

The second approach will involve fitting a multivariable complementary log-log model comparable to the model described in the Section 9.3 that includes the covariates described above with any covariates with a p-value of 0.10 or greater retained in the final model.

9.4.5 Cost-effectiveness of Treatment Regimens

The cost-effectiveness analysis will be conducted from a payer perspective and will be expressed as incremental cost required to produce one additional unit of quality-adjusted life year (QALY). Data on each participant's use of medical and non-medical resources related to urologic or gynecologic conditions will be collected during the follow up period. Direct and indirect costs of the treatment of apical pelvic organ prolapse with native tissue surgical repair or transvaginal mesh repair and women's preference for health states for improvement in pelvic organ prolapse will be estimated.

We plan to capture incremental health care resource use related to study interventions and complications and other prolapse management (such as pessary use or additional surgery). Costs will be estimated using the resource costing method where medical service use from each study case report form is monetized by multiplying the number of units of each medical service by the average unit cost of this service in dollars. This method allows a consistent capture of resource use when costs are incurred across multiple health systems or payers. Detailed case report forms, that include the procedures performed (e.g. surgical interventions) and clinical events (e.g. complications, readmissions) will be completed by the study coordinator at study visits. Data from three resource types (physician visits, hospital procedures and admissions, and emergency room visits) will be collected. Cost for each medical service use will be assigned based on national Medicare

reimbursement rates, as indicated in the table below. Additionally, we will obtain detailed billing records for a limited number of procedures and hospitalizations in selected study sites (e.g. prolonged admission to the ICU or readmission to the hospital for a surgical complication).

Service	**Price Weight**
Physician visit	Medicare reimbursement
Surgical intervention and admission	Medicare reimbursement
Complication hospitalization - routine	Medicare reimbursement
Complication hospitalization – significant	Billing record – actual amount paid
ER – routine complication	Medicare reimbursement
ER – significant complication	Billing record – actual amount paid
Subsequent surgery	Medicare reimbursement

The SF-6D preference-based utility index algorithm derived from the SF-12 instrument53 will be used to calculate each participant's utility index at baseline and various follow up time points based on her responses to the SF-12 questionnaire. The SF-6D focuses on seven of the eight health domains covered by the SF-12: physical functioning, role participation (combined role-physical and role-emotional), social functioning, bodily pain, mental health, and vitality.). This instrument has been previously used in women with urinary incontinence54. These data will be used to compare change in QALYs between the two treatment groups. We are choosing to use a general scale to calculate change in utilities (rather than condition-specific) to allow for comparison of cost-effectiveness results with other interventions and diseases. Because the follow up period for participants spans at least three years, costs and QALYs in the second year and third year of follow up will be discounted using a 3% discount rate/year.

Differential mean costs and differential mean QALYs between the two treatment groups will be estimated using multiple regression analysis. Specifically, a generalized linear model with appropriate link function (e.g., log-link) and response probability distribution (e.g., gamma distribution) will be used to analyze costs due to the potential skewness and heteroscedasticity of medical expenditure data, while an ordinary least squares regression will be used for analyzing QALY data. The models will account for treatment group, study site and stratification factors, as well as other characteristics of the participants that are found to differ significantly between the traditional vaginal hysterectomy with native tissue vault suspension and mesh hysteropexy suspension groups. When estimating QALYs, we will also adjust for participants' baseline utility scores to account for potential imbalance in baseline utility between the two treatment groups.55

We will calculate the incremental cost-effectiveness ratio (ICER), which is the differential mean costs divided by the differential mean QALYs between the two groups, to assess the additional costs associated with each additional QALY gained. Our base case analysis will be conducted based on participants with complete data. Sensitivity analysis will be conducted to include participants with incomplete data using the multiple imputation method. Non-parametric bootstrapping resampling technique will be used to derive the 95% confidence interval for the ICER. In addition, cost-effectiveness acceptability curve (CEAC) will be generated to illustrate the likelihood that one treatment is more cost-effective than the other with various ceiling cost-effectiveness ratios.

In the case that a statistically significant difference in changes in utilities (as measured by SF-6D) between the treatment groups is not detected, we plan to conduct supplemental analyses using alternative outcome measures, such as incremental cost per treatment success, incremental cost per POP HRQOL, or incremental cost per satisfaction.

The cost-effectiveness evaluations will be conducted as within-trial comparisons. A decision analytic model will also be developed from trial data to evaluate the trajectory of the cost-effectiveness ratio over a lifetime; assuming an average life expectancy, given the average age of participants at the time of the intervention.

9.4.6 Changes in Body Image

Total BIS scores will be calculated as described by Hopwood et al, and changes in BIS scores will be correlated with changes in PISQ R scores, the leading edge of prolapse, and POPDI scores using Pearson's and Spearman correlation coefficients. If measures are correlated, we will further describe changes in individual items of the BIS scale before and after treatment for prolapse, to determine which portions of the BIS scale explain the overall change in BIS scores, and evaluate the associations between BIS item scores and PISQ IR item scores. General linear model analyses will be used to evaluate how well the combination of BIS scores, demographic and experimental variables explain changes in PISQ R scores.

Between randomization group comparisons will also be made using linear models for total and item scores controlling for randomization strata. We will also utilize extensions of this model to evaluate if any differences in BIS scores are either modified by or confounded with differences in either efficacy or side effects.

10 SAFETY ANALYSES

10.1 Overview of Safety Analysis Methods

All safety analyses will be performed using the SAF population unless otherwise specified. Study arm summaries will be constructed across all study sites. Descriptive p-values comparing the study arms will be provided on most safety table summaries and will be obtained using Fisher exact tests if the occurrence of such events is rare and Cochrane-Mantel-Haenszel tests for more common events.

10.2 Adverse Events

As described in Section 3.2, four types of measures will be utilized to characterize participant safety and tolerability measures. First, as a part of the inter-operative and peri-operative data collection forms, information will be collected on surgical complications that characterize the risk and tolerability of the procedure itself. Second, participants will be asked about specific targeted complications on both the post-operative visit and a standard study evaluation collected at 6-month intervals (with those complications automatically populating the AE form. Third, participants will be asked through a response to open-ended questions to report any post-surgery adverse events to the study coordinator or physician at the 6-week visit and at the subsequent follow-up visits at 6-month intervals. If a serious adverse event, particularly a hospitalization, occurs, the participants will be asked to report that immediately to the study coordinator or physician. All adverse events are captured by the study coordinator on the adverse event logs. Finally, specific pain information is being collected as a part of all follow-up visits to characterize differences in post-operative pain beyond 6 weeks.

The group of measures collected inter-operatively and peri-operatively includes those in the table below. These measures include continuous measures as well binary and ordinal measures. For the

continuous measures, summaries will be presented by treatment arm that include mean levels and standard deviations and linear models with treatment as the primary predictor, controlling for site will be used to test for differences by treatment arm. For the binary and ordinal measures, summaries will included counts and percentages associated with each level of the binary or ordinal measure and Cochran-Mantel-Haenszel statistics and associated p-value with analyses stratified by site will be used to characterize differences in the treatment arms.

The AEs collected through the standard AE form will be summarized using standard contingency tables by treatment arm. These summaries will be provided by individual treatment-emergent AEs, AEs by body system, and AEs by subject. In addition to these general descriptions, specific comparison of risk and differences across the treatment arms will be evaluated for the outcomes listed in the table below.

Summary of Safety Variables		
Inter-operative and peri-operative measures		
Variable	**Type**	**Definition**
Operative time	Continuous	Obtained directly from peri-operative form
Blood loss	Continuous	Obtained directly from peri-operative form
Blood transfusion	Binary	Obtained directly from peri-operative form
Complications	Ordinal	All complications will be collected and scored on a five-point ordinal scale using the Dindo classification scheme (Dindo, 2004). Each participant will then be classified according to the most severe complication
Routinely Collected Adverse Events		
Mesh-related Complications	Ordinal	Mesh related complications: mesh exposure in the vagina or mesh erosion into another organ; note that level of complication will be characterized into based on the following response classification schema: (a) None or non-surgical medical intervention only; (b) Minor or intra-office surgical intervention; (c) Outpatient surgery; (d) Inpatient surgery
Pelvic infection	Binary	Any incident of pelvic infection will be captured as a Yes response (AE or SAE) to Question 13 in the complications section of the Hospitalization form or Question 14 of the follow-up evaluation form. It will also be captured based on open-ended AE responses with comments indicative of pelvic infection.
Perioperative infection	Binary	A perioperative infection will be captured as a Yes response (AE or SAE) to one or more of Questions 13, 14, 22, 23, 24, or 25 of the Hospitalization form or Questions 15, 16, 17, 18, 19, or 10 of the Post-operative visit form.

UTI	Binary	A UTI will be captured as a Yes response (AE or SAE) to Question 23 of the Hospitalization form, Question 18 of the Post-operative visit form, or Question 20 or the 6- to 60-Month Study follow-up form or by an event classified as a Urinary Tract Infection based on Preferred terms in the AE form.
Vaginal infections	Binary	A Vaginal Infection will be captured as a Yes response (AE or SAE) to Question 22 of the Hospitalization form, Question 17 of the Post-operative visit form, or Question 19 or the 6- to 60- Month Study follow-up form or by an event classified as a vaginal infection based on either the preferred term or comments for an event on the AE form.
Vaginal shortening	Binary	Vaginal shortening will be determined by an event classified as a vaginal shortening based on either the preferred term or comments for an event on the AE form.
De novo vaginal bleeding	Binary	Based on masked data review this event was modified from De novo vaginal bleeding to any incident of vaginal bleeding because of the small number of events. It will be captured as a Yes response (AE or SAE) to Question 17 on the 6- to 60- Month Study follow-up form or by an event classified as a vaginal bleeding based on either the preferred term or comments for an event on the AE form.
Atypical vaginal discharge	Binary	Atypical vaginal discharge will be captured as a Yes response (AE or SAE) to Question 16 on the 6- to 60-Month Study follow-up form or by an event classified as atypical vaginal discharge based on either the preferred term or comments for an event on the AE form.
Fistula formation	Binary	Fistula formation will be captured as a Yes response (AE or SAE) to Question 16 of the Hospitalization form, Question 8 of the Post-operative visit form, or Question 7 or the 6- to 60- Month Study follow-up form or by an event classified as a fistula based on either the preferred term or comments for an event on the AE form
Neuromuscular problems	Binary	A neuromuscular problem will be captured as a Yes response (AE or SAE) to Question 12 of the Hospitalization form, Question 14 of the Post-operative visit form, or Question 13 or the 6- to 60- Month Study follow-up form or by an event classified as a neuromuscular disorder based on either the preferred term or comments for an event on the AE form
Need for subsequent procedures	Binary	Need for subsequent procedures will be classified as Yes if the responses on Section E of the Post-operative visit form or Section D of the 6-to-60 Month Evaluation form indicated that the participant had a procedure performed in

		the office in response to a complication from the surgery or the participant had a reoperation operating room for complication related to surgery, recurrent POP, SUI, or any gynecologic, vaginal, pelvic, or reproductive organ condition.
Uterine or cervical pathology	Binary	Classified as Yes based on masked review of AE preferred terms.
Vaginal scaring	Binary	Vaginal scarring will be determined by an event classified as a vaginal scarring based on either the preferred term or comments for an event on the AE form.
Dyspareunia	Binary	Defined via two different mechanisms: (a) as an event of dyspareunia as listed on AE form or the 6-month complication evaluation form; or (b) assessed via reporting on the PISQ-IR instrument. Note that because on masked review women moved in and out of the dyspareunia category after surgery, the decision was made to evaluate all incidents of dyspareunia irrespective of de novo status.

10.3 Deaths and Serious Adverse Events

A serious adverse event (SAE) is defined as any event that occurs during the 'active' phase of treatment, or the follow-up periods, and either: (1) results in death, or (2) requires inpatient hospitalization or a prolongation of existing hospitalization, or (3) is a congenital anomaly/birth defect, or (4) results in persistent or significant disability / incapacity, or (5) is life-threatening, or (6) requires intervention to prevent one of the above outcomes.

SAEs will be listed and SAEs, treatment-related SAEs and SAEs with an outcome of death will be summarized by treatment arm and site if there are enough events to summarize. Separate displays listing and summarizing death will also be created.

10.4 Other Safety Outcomes

All AEs identified on the AE forms will be summarized by treatment arm using MedDRA-classified verbatim terms and system organ class with numbers and percentages summarized by treatment arm.

10.5 Concomitant Medications

Limited data on concomitant medications related to pelvic floor disorders as well as antibiotic information will be collected as a part of this protocol. These medications will be listed and the number of subjects receiving each medication will be summarized by each reported term as well as drug class separately for each treatment. No formal analysis of concomitant medications is planned.

11 PHARMACOKINETIC ANALYSES

No pharmacokinetic analyses are planned.

12 ANALYSIS OF OTHER OUTCOMES

No analyses of outcomes other than efficacy and safety/tolerability outcomes are planned.

13 REPORTING CONVENTIONS

Unless required otherwise by a journal, the following rules are standard:

- Moment statistics including mean and standard deviation will be reported at 1 more significant digit than the precision of the data.
- Order statistics including median, min and max will be reported to the same level of precision as the original observations. If any values are calculated out to have more significant digits, then the value should be rounded so that it is the same level of precision as the original data.
- Following SAS rules, the median will be reported as the average of the two middle numbers if the dataset contains even numbers.
- Test statistics including t and z test statistics will be reported to two decimal places.
- P-value will be reported to 3 decimal places if > 0.001. If it is less than 0.001 then report '<0.001'. Report p-values as 0.05 rather than .05.
- No preliminary rounding should be performed, rounding should only occur after analysis. To round, consider digit to right of last significant digit: if < 5 round down, if >=5 round up.

14 CHANGES TO THE ANALYSES PLANNED IN THE PROTOCOL

To be completed at a later date

15 REFERENCES

To be completed

16 LIST OF POTENTIAL DISPLAYS

Data displays may be added, deleted, rearranged or the structure may be modified after finalization of the SAP. Such changes require no amendment to the SAP as long as the change does not contradict the text of the SAP. Also note that specific display shells will be generated for each potential manuscript and reviewed by the protocol team prior to the initiation of analyses for that manuscript.

Tables
Participant Eligibility
Participant Disposition
Protocol Deviations
Demographic and Baseline Characteristics
Primary Efficacy Model Results
Secondary Efficacy Outcome Measures
Safety and Tolerability Summary
Adverse Events

Figures
Consort diagram of participant disposition
Duration of Treatment Success Kaplan Meier Curves
Data Listings
Subject Eligibility
Subject Disposition
Protocol Deviations
Adverse Events
Serious Adverse Events

17 ATTACHMENTS

Potential Study Displays

Table 1. Patient Disposition by Treatment Group

Variable	Treatment Group	
	Native Tissue Hysterectomy N=	Mesh Hysteropexy N=
RCT Consent	n	n
Screen Failures	n	n
Pre-Randomization Withdrawals	n	n
Screen Failures and Pre-Randomization Withdrawals/RCT Consent (%)	xx.x%	xx.x%
RCT Randomization	n	n
Total Post-Randomization Withdrawals	n	n
Completed Month 6	n	n
Completed Month 12	n	n
Completed Month 24	n	n
Completed Month 36	n	n
Completed Month 48	n	n
Completed Month 60	n	n

* p-value from Chi-square or Fisher's exact tests for treatment-group comparisons.

Table 2: Baseline characteristics of the two arms

Measure	TVH	Hysteropexy	P value
Age in years			
Gravity			
Parity			
BMI (Kg/m^2)			
Menopausal status (almost everyone will be postmenopausal so we can just report % postmenopausal)			
Marital status			
Education			
Health Insurance			
History of smoking			
Current smoker			
Past history of surgery for SUI			
Past History of POP surgery			
Other Pelvic surgery			
Duration of POP symptoms			
History of recurrent UTI			
Yes to any PMH question e3 *			
POPQ Ba			
POPQ C			
POPQ Bp			
PVR (mean, median, IQR, usually has a right tail skewed distribution)			
PFDI -20			
POPDI-6			
UDI 6			
CRADI-8			
PFIQ			
ISI			
BIS			
PISQ-IR			
SF-12			
Functional activities score			
Pain scale score			
Number of pain medications (Mean number)			

Table 3: Efficacy Outcome Components Overall by treatment

Failure reasons	TVH	Hysteropexy	P value
Retreatment (n,%)			
Prolapse past the hymen only (n,%)			
Bulge symptoms only (n,%)			
Prolapse and Bulge symptoms (n,%)			

Table 3b: Efficacy Outcome over time by treatment

		Treatment Group			
Characteristic	*Statistic/ Category*	*UPHOLD Procedure (N=88)*	*Vaginal hysterectomy and USLS (N=87)*	*Total (N=175)*	*P-value*
12 Month Visit Outcome	Not Failure				
	Anatomic Failure				
	Anatomic and sympt. Failure				
	Retreatment Failure				
	Symptomatic Failure				
24 Month Visit Outcome	Not Failure				
	Anatomic Failure				
	Anatomic and symp. Failure				
	Retreatment Failure				
	Symptomatic Failure				
36 Month Visit Outcome	Not Failure				
	Anatomic Failure				
	Anatomic and symp. Failure				
	Retreatment Failure				
	Symptomatic Failure				

Table 4: Perioperative Outcomes by treatment

		Treatment Group			
Characteristic	*Statistic/ Category*	*UPHOLD Procedure (N=88)*	*Vaginal hysterectomy and USLS (N=87)*	*Total (N=175)*	*P-value*
Operative time in minutes	N				
	Mean (SD)				
	Median				
	Quartiles 1 & 3				
	Range				
Estimated blood loss in ml	N				
	Mean (SD)				
	Median				
	Quartiles 1 & 3				
	Range				
Blood transfusion rate	No				
	Yes				
Number of postoperative tests	N				
	Mean (SD)				
	Median				
	Quartiles 1 & 3				
	Range				
Voiding spontaneously at discharge	No				
	Yes				
Days in hospital	N				
	Mean (SD)				
	Median				
	Quartiles 1 & 3				
	Range				
Return to OR	No				
	Yes				
ICU admission rate	No				
	Yes				
DIndo complications (hospital)	None				
	I				
	IIa				
	IIb				
	IIIb				
	IIIo				
	IVa				
Hemolytic agent use	AE				
	No				
Urethral kink	AE				
	No				
Device malfunction	SAE				
	AE				
	No				
Bladder injury	AE				
	No				
Vaginal laceration	AE				
	No				
Postoperative thrombosis	AE				

		Treatment Group			
Characteristic	*Statistic/ Category*	*UPHOLD Procedure (N=88)*	*Vaginal hysterectomy and USLS (N=87)*	*Total (N=175)*	*P-value*
	No				
Atelectasis	AE				
	No				
Acute MI	SAE				
	No				
Bronchiectasis	AE				
	No				
Respiratory failure	SAE				
	No				
Femoral nerve injury	AE				
	No				
Urinary retention	AE				
	No				
Bladder spasm	SAE				
	No				
Musculoskeletal pain	SAE				
	No				
Arrhythmia	AE				
	No				
Blood culture positive	AE				
	No				
Peritoneal perforation	AE				
	No				
Pneumonia	AE				
	No				
Pulmonary edema	AE				
	No				
Urinary tract infection	AE				
	No				
Wound infection	AE				
	No				
Post-procedural hemorrhage	SAE				
	No				

Table 5: Six-Week Safety and Complication Outcomes

		Treatment Group			
Characteristic	*Statistic/ Category*	*UPHOLD Procedure (N=88)*	*Vaginal hysterectomy and USLS (N=87)*	*Total (N=175)*	*P-value*
Excessive granulation tissue	AE				
	No				
Stress urinary incontinence	AE				
	No				
Urinary tract infection	AE				
	No				
Constipation	AE				
	No				
Urge incontinence	AE				
	No				
Faecal incontinence	AE				
	No				
Urinary incontinence	AE				
	No				
Vulvovaginal adhesion	AE				
	No				
Hypoaesthesia	No				
cPain in extremity	AE				
	No				
Dysuria	AE				
	No				
Vulvovagina lpain	AE				
	No				
Lichensclerosus	AE				
	No				
Colitis	SAE				
	No				
Vulvovaginal injury	AE				
	No				
Bartholin scyst	AE				
	No				
Suture related complication	AE				
	No				
Urinary retention	AE				
	No				
Vaginal hematoma	AE				
	No				
Valval hemorrhage	AE				
	No				
Colitis, ischemic	SAE				
	No				
Phlebitis	AE				
	No				
Rash	AE				
	No				
Vaginal infection	AE				
	No				
Medical device site reaction	AE				
	No				
Hematoma	SAE				
	No				

Table 6: Overall Summary of Adverse Events

	Native Tissue Hysterectomy (N=xx)	Mesh Hysteropexy (N=xx)	p-value [1]
Adverse Events			
Number of AEs			
Number of Subjects With			
Any AE			
AE by Study Relationship [2]			
Unrelated			
Possibly			
Probably			
Definitely			
o AE by Severity [3]			
Mild			
Moderate			
Severe			
Life-threatening			
Specific Events of Interest			
Hospitalization/ER Visit			
Serious Adverse Events			
Number of SAEs			
Number of Subjects With			
Any SAE			
Deaths			