# STATISTICAL ANALYSIS PLAN

# Three-Arm Apical Suspension Trial for Post-Hysterectomy Vault Prolapse: Prospective Randomized Trial Involving Sacral Colpopexy, Transvaginal Mesh and Native Tissue Apical Repair

Short Title: **A**pical **S**uspension Repair for Vault **P**rolapse **I**n a Three-Arm **R**andomiz**e**d Trial Design "ASPIRe"

| | |
|---|---|
| **SAP VERSION:** | Version 2.0 |
| **SAP DATE:** | February 20, 2023 |
| | |
| **PROTOCOL VERSION:** | Version 6.0 |
| **PROTOCOL DATE:** | February 12, 2021 |
| | |
| **SPONSOR:** | NICHD |
| **TRIAL REGISTRATION #:** | 27P01 |
| | |
| **PRINCIPAL INVESTIGATOR:** | Shawn Menefee |
| | |
| **PREPARED BY:** | RTI International. |
| | 3040 Cornwallis Rd |
| | Research Triangle Park, NC 27709-2104 |
| **AUTHOR (S):** | Ryan Whitworth, Sonia Thomas, Amaanti Sridhar |

# Contents

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ACOG | American Congress of Obstetricians and Gynecologists |
| AE | Adverse event |
| BIPOP | Body Image Pelvic Organ Prolapse |
| BMI | Body mass index |
| BPPS | Body part pain score |
| CI | Confidence interval |
| CRADI | Colorectal Anal Distress Inventory |
| CRAIQ | Colorectal Anal Impact Questionnaire |
| DCC | Data coordinating center |
| DSMB | Data and safety monitoring board |
| DRS | Decision Regret Scale |
| FAS | Functional Activity Assessment Scale |
| FASt | Frailty ASPIRE substudy |
| FDA | U. S. Food and Drug Administration |
| FI | Fecal incontinence |
| HRQOL | Health-related quality of life |
| ICER | Incremental cost-effectiveness ratio |
| IRB | Institutional Review Boars |
| ITT | Intention to treat |
| MAR | Missing at random |
| MCAR | Missing completely at random |
| MCS | Mental Component Scale of the SF12 |
| MedDRA | Medical dictionary for regulatory activities |
| NICHD | National Institute of Child Health and Human Development |
| NTR | Native Tissue Repair for apical prolapse |
| OAB | Overactive bladder syndrome |
| PCS | Physical Component Scale of the SF12 |
| PFDI | Pelvic Floor Distress Inventory |
| PFIQ | Pelvic Floor Impact Questionnaire |
| PISQ-IR | Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire – Revised version |
| POPDI | Pelvic Organ Prolapse Distress Inventory |
| POPIQ | Prolapse Impact Questionnaire |
| POP-Q | Pelvic Organ Prolapse Quantification |
| PPAR | Patient Perspective in AE Reporting substudy |
| PVR | Post-void residual |

| | |
|---|---|
| QALY | Quality-adjusted life year |
| QOL | Quality of life |
| SAE | Serious adverse event |
| SAF | Safety Population |
| SAP | Statistical analysis plan |
| SC | Sacral Colpopexy |
| SDS | Satisfaction with Decision Scale |
| SF-12 | 12 Question Short Form Health Survey |
| SOC | MedDRA System Organ Class |
| SPS | Surgical pain scale |
| SSLS | Sacrospinous Ligament Suspension |
| SUI | Stress urinary incontinence |
| TVL | Total vaginal length |
| TVM | Transvaginal Mesh |
| UDI | Urinary Distress Inventory |
| UI | Urinary incontinence |
| UIQ | Urinary Impact Questionnaire |
| USLS | Uterosacral ligament suspension |
| UTI | Urinary tract infection |
| UUI | Urge urinary incontinence |

# 1    BACKGROUND AND PROTOCOL HISTORY

Pelvic organ prolapse (POP) is a common condition in women with more than 225,000 surgeries performed annually. Post-hysterectomy vaginal vault (apical) prolapse can be managed surgically with various techniques and approaches ranging from obliterative procedures to reconstructive procedures performed with the placement of grafts and mesh (apical transvaginal mesh repairs, sacral colpopexy) or with the patient's own tissue (native tissue repairs). While traditional native tissue vaginal repairs remain the most common approach to surgical correction of POP, there are some estimates that one out of three POP surgeries in 2010 used mesh with three out of four mesh procedures being completed transvaginally. The use of transvaginal mesh for POP repairs markedly increased until the FDA warning in 2011 described "serious complications associated with transvaginal placement of surgical mesh in repair of pelvic organ prolapse and stress urinary incontinence." While sacral colpopexy (SC) has long been considered the gold standard treatment of apical prolapse, it is the least commonly performed procedure and requires levels of skill beyond traditional or grafted vaginal surgery. While many surgeons favor specific surgical approaches for vaginal vault prolapse, most surgeons choose the approach based on prior training and experience rather than scientific evidence to guide their decision-making.

The purpose of this study is to compare abdominal sacral colpopexy to transvaginal mesh and sacral colpopexy and transvaginal mesh repair to vaginal native tissue repairs for apical prolapse. At this point in time, there is no strong evidence that either uterosacral or sacrospinous ligament suspension is superior as a vaginal native tissue repair; thus, investigators may use their preference for either of these native tissue suspensions. For the apical TVM arm of the study, a specific non-trocar mesh procedure will be studied, namely the Uphold LITE procedure which will be performed according to manufacturer recommendations. For abdominal sacral colpopexy, all current techniques including open, laparoscopic, and robotic have similar objective success and any approach will be allowed for SC arm. The sacral colpopexy will be performed in a similar fashion by all approaches.

The concept was approved by the PFDN Steering Committee on October 11, 2013, mini-protocol on January 17, 2014, and the full protocol was initially approved on April 25, 2014.  Version 1.0 of the protocol was approved on November 24, 2015, after review by the DSMB and the Steering Committee.

- The protocol was first amended (Version 2.0) on January 27, 2016, to add Elevate as one of the allowed types of transvaginal mesh and update text to be consistent with an FDA order in January 2016 which reclassified transvaginal mesh devices from class II (moderate-risk) to class III (high-risk).
- A new version, 3.0, of the Protocol was issued on December 5, 2016, to remove the Elevate device as a mesh surgery option because the company manufacturing the kits halted production in March 2016 and went out of business, add University of Texas at Southwestern as a clinical site, and remove the Functional Activity Scale, PFDI-20, and BIPOP assessments from the 6 week time point.
- Version 4.0 (December 27, 2018) updated the target enrollment to randomized <u>and treated</u> participants, added a step-down test of superiority of TVM to SC if TVM is found non-inferior to SC, and defined the primary analysis populations (MITT for superiority, Per protocol for non-inferiority).

The protocol was updated to Version 5.0 on April 18, 2019, in response to FDA action. On April 16, 2019, the FDA issued a notification that required all manufactures of transvaginal mesh for prolapse repair to stop selling and distributing the product. Following the FDA notification, the Steering Committee along with the NICHD and DSMB felt that it would be unreasonable to continue with

enrollment of patients into the TVM arm of the trial. Thus, the TVM arm was halted at this time. All participants including those who received TVM were continued to be followed as recommended by the FDA and PFDN Steering Committee. Participants who were randomized to receive TVM were notified of the FDA statement and recommendation for follow up.

Given this notification, the Pelvic Floor Disorders Network (PFDN) Steering Committee, in consultation with the NICHD and DSMB, elected to halt the Transvaginal Mesh Repair (TVM) arm of this study and to close the study to enrolment on April 18, 2019. No new participants were consented after April 18, 2019, and no further TVM surgeries were performed, only the Sacral Colpopexy (SC) or the Transvaginal Native Tissue Repair (NT) (two groups instead of three). This decision required changes to the randomization plan, the protocol, and the consent. The protocol amendment indicated, at the time, there were 18 remaining consented or randomized participants who were awaiting surgery. Remaining patients that were randomized to TVM were to be re-randomized to SC and NT in a ratio of 1:1. A planned sensitivity analysis to explore the impact of this protocol amendment is described in Section 9.1.

No formal interim analysis for efficacy was planned for this study. The primary analysis will occur after the final participant enrolled completes her 36 month follow up visit.

Protocol Version 6.0 (December 2020) added the eASPIRE long term follow-up, which is not covered in this SAP

## 2    PURPOSE OF THE ANALYSES

This statistical analysis plan (SAP) contains detailed information about statistical analyses to be performed to assess efficacy and safety across three methods of vault suspension. The results of these analyses will be included in the primary manuscript and a series of pre-planned secondary manuscripts. 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 three-arm randomized clinical trial is to determine if apical transvaginal mesh placement is non-inferior to sacral colpopexy for anatomic correction of post-hysterectomy vaginal vault prolapse and to determine if mesh reinforced repairs performed by abdominal or vaginal approach are superior to native tissue vaginal repair. The primary aims for this trial are:

1. To determine if Apical Transvaginal Mesh is non-inferior to Sacral Colpopexy for anatomic correction of post-hysterectomy vaginal vault prolapse at time points through 3 years, using a non-inferiority hazard ratio margin of 1.93.

   1a  In the case where Apical Transvaginal Mesh is shown to be statistically significantly non-inferior to Sacral Colpopexy for anatomic correction of post-hysterectomy vaginal vault prolapse at time points through 3 years, to determine if Apical

Transvaginal Mesh is superior to Sacral Colpopexy for anatomic correction of post-hysterectomy vaginal vault prolapse at time points through 3 years

2. To determine if Sacral Colpopexy is superior to Native Tissue Repair for anatomic correction of post-hysterectomy vaginal vault prolapse at time points through 3 years

3. To determine if Apical Transvaginal Mesh is superior to Native Tissue Repair for anatomic correction of post-hysterectomy vaginal vault prolapse at time points through 3 years.

This study will test the null hypotheses that treatment failure will not differ between vaginally and abdominally placed mesh for vaginal vault prolapse, and mesh repairs (regardless of route of implantation) will be superior to native tissue apical suspension.

### 3.1.2 Secondary Objectives

The secondary aims for the ASPIRe 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) across three methods of vault suspension.

2. Safety: To measure and compare safety, adverse events (including mesh erosion and exposure), pain, and need for subsequent procedures across three methods of vault suspension.

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. Body image: To describe changes in body image as measured by a validated scale, the Body Image Pelvic Organ Prolapse Scale (BIPOP). in a group of women undergoing apical repair with and without mesh and to evaluate whether or not changes in sexual function are associated with changes in body image.

5. Preference Evaluation: To assess patient and surgeon reasoning for declining participation in the trial. This will be determined by the CONSORT diagram for enrollment.

6. Cost- effectiveness: To compare the cost across three methods of vault suspension.

7. Global Composite Outcome: To evaluate the development of a valid and reliable Global Composite Outcome that balances adverse events and patient-centered outcomes to anatomic definitions of failure and success.

8. Patient Perspective in AE Reporting (PPAR): To evaluate the patient's perspective about adverse events and their role in patient decision-making outcomes. The aims of PPAR include comparing patient versus surgeon rankings of complication grade, outcome, expectedness and seriousness, to estimate the association between patient rankings of AEs

with decision-making and quality of life outcomes, and to determine if their perspective about AEs changes over time.

9. Frailty ASPIRe Study (FASt): To determine the impact of preoperative frailty and mobility on surgical treatment outcomes and postoperative complications of older women following surgical correction of apical pelvic organ prolapse (POP).

## 3.2 Outcomes

The **primary outcome** for the study is treatment failure at any point greater than 6 weeks 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. .

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 in the primary ASPIRE manuscript 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 location measures postoperatively in the two treatment arms.

    b. Proportion of participants in each group with C > -2/3 TVL

    c. Mean Median values of the maximum extent of prolapse (defined as leading edge of prolapse-Ba, C, Bp)

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, and the overall PFDI total score .

    c. Urinary function measured at 6-month post-operative time points using:

        i. Mean UDI-6 scores

        ii. Risk of de novo voiding dysfunction

      iii.   Risk of de novo incontinence (stress, urge, and mixed)

  d.  Sexual functional measures obtained at 6-month post-operative time points including:

      i.   PISQ-IR scale measures

      ii.   Risk of de novo dyspareunia

  e.  Bowel function obtained at 6-month post-operative time points using CRADI-8 scores

  f.  Quality of Life (QOL) 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 (prolapse: POPIQ, urinary: UIQ, bowel: CRAIQ)

      iii.   Functional activity as measured by the Functional Activity Assessment Scale (FAS)

  g.  Regret/Satisfaction:

      i.   Regret with Decision Regret Scale (DRS-PFD)

      ii.   Satisfaction with Decision Scale (SDS-PFD)

3.  Safety and tolerability of the two treatment regimens:

  a.  Intraoperative safety measured in each of the treatment arms including:

      i.   Mean 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 (SPS), pain medication use (during 24 hours prior to completion of the SPS Assessment), and Body Part Pain Scales (BPPS) assessment.

      iii.   Pelvic infection

          1.  Risk of perioperative infections

          2.  Risk of urinary tract infections

          3.  Vaginal infections with flora uncommon to the vaginal canal

> iv.  Risk of any of De novo vaginal bleeding, atypical vaginal discharge, fistula formation, neuromuscular problems (including groin and leg pain)
>
> v.  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).
>
> vi.  Risk of vaginal scarring defined as: De novo vaginal scar requiring medical or surgical intervention, or adversely affecting quality of life.
>
> vii.  Risk of Vaginal shortening, de novo dyspareunia, and
>
> viii.  Risk of worsening dyspareunia with AE survey instrument.

The following study outcomes are to be presented in secondary manuscripts, each with its own SAP. Definitions and analyses for these outcomes are not specified in this SAP, which covers analyses to be presented in the primary results manuscript:

4.  Key predictors of poor treatment outcomes defined as the effect of advanced prolapse, age, obesity, smoking, estrogens, primary vs. recurrent prolapse on higher risk of treatment failure.

5.  Body image as measured by the *BIPOP Scale* at 6-month post-operative intervals.

6.  Cost-effectiveness of the three 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.

7.  Global Composite Outcome

8.  Patient Perspectives in Adverse Event Reporting (PPAR) Study, which includes patient reporting of frequency and severity levels of selected AEs and complications compared against physician assessments from the same events.

9.  Frailty ASPIRe Study (FASt), which includes measurements of mini-cog, Katz ADL, functional comorbidity index, Functional Assessment Scale, and timed up and go.

# 4   STUDY METHODS

## 4.1 Overall Study Design and Plan

The study is a multi-center randomized, surgical trial of women with symptomatic post-hysterectomy apical (cuff) prolapse desiring surgical treatment. This study will compare the three available surgical treatments performed in usual practice. The purpose of this study is to compare two commonly performed mesh apical repairs (sacral colpopexy vs. Apical Transvaginal Mesh) and vaginal native

tissue apical repairs. The primary outcome is measured over time (up to 60 months) using a survival analysis approach.

Participants are followed in ASPIRe until the last participant reaches 36 months of follow-up, so at study completion, participants will have variable follow-up between 36 and 60 months.

A study schematic is shown below:
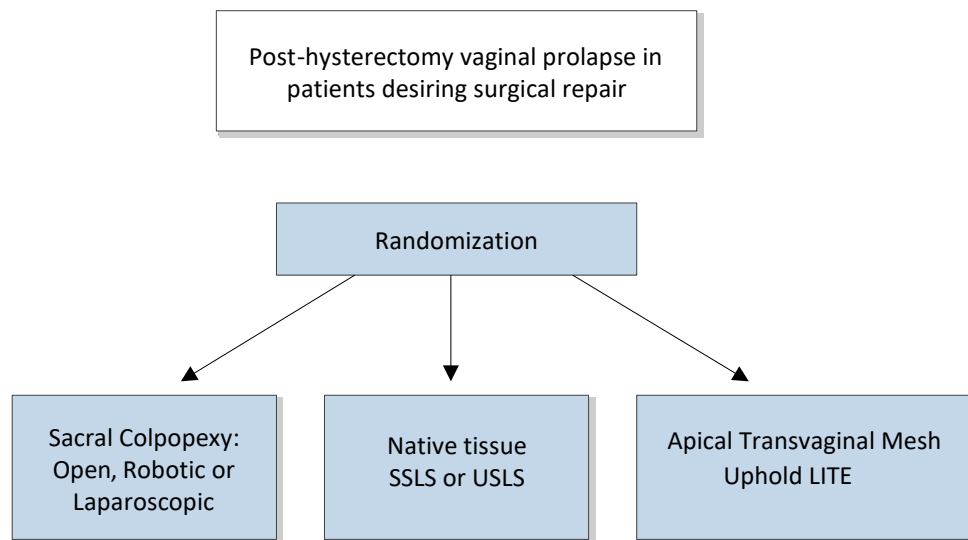


*All groups may receive concomitant anterior and/or posterior repairs and full-length mid-urethral slings as indicated, per the discretion of the surgeon.

## 4.2 Study Population

The study population was planned to be comprised of 363 participants (who were enrolled, consented, randomized and treated) as defined by the following eligibility criteria:

The following paragraph comes from the DSMB Open Report's Protocol Synopsis section and is consistent with the protocol.

Inclusion Criteria:
1. Women aged 21 or older
2. Prior total hysterectomy (no cervix present)
3. Prolapse beyond the hymen (defined as Ba, Bp, or C > 0 cm)
4. Vaginal cuff descent into at least the lower two thirds of the vagina (defined as point C> -2/3 TVL)
5. Bothersome bulge symptoms as indicated on question 3 of the PFDI-20 form relating to 'sensation of bulging' or 'something falling out'
6. Desires surgical treatment for post-hysterectomy vaginal prolapse
7. Available for up to 60 month follow-up

Exclusion Criteria:
1. Previous synthetic material or biologic grafts (placed vaginally or abdominally) to augment POP repair including anterior, posterior and/or apical compartments

2. Known previous formal SSLS performed for either uterovaginal or post-hysterectomy vaginal vault prolapse
3. Known adverse reaction to synthetic mesh or biological grafts; these complications include but are not limited to erosion, fistula, or abscess
4. Unresolved chronic pelvic pain-active
5. Prior abdominal or pelvic radiation
6. Contraindication to any of the index surgical procedures
    a. Known Horseshoe Kidney or Pelvic Mass Overlying the sacrum
    b. Active diverticular abscess or diverticulitis
    c. Shortened vaginal length (<6 cm TVL)

## 4.3 Study Arm Assignment and Randomization

After eligibility is determined and consent is obtained, randomization will occur. Ideally, the time from randomization to surgery will be no longer than 6 weeks (42 days). This time interval will be tracked to determine if delay from randomization to surgery ultimately impacts entry into study represented by those patients undergoing the index surgical procedure. If more than 6 weeks transpires between randomization and surgery, a protocol deviation will be completed. If another visit (i.e. preoperative visit) occurs after randomization, the surgeon should continue to mask the patient to the details of the randomized apical procedure until after completion of the surgery. Randomization in the operating room is not feasible because of the unique equipment to perform laparoscopic and robotic sacral colpopexy. Often robotic rooms are in high demand and need to be reserved and utilized when scheduled.

The randomization will occur in a ratio of 1:1:1 for each treatment arm with an equal chance of being randomized into each treatment group. While the potential existed to perform a 2:2:1 (SC:TVM:NT) randomization scheme given the prior reported evidence using composite outcomes, the working group believes that this scheme would significantly increase patient confusion in counseling for the study and may impact investigator equipoise. In addition, this randomization schema would result in unequal group size, making evaluation of adverse events more difficult and potentially unbalanced between groups.

Randomization will be performed using permuted blocks with a block size that is known only to the DCC and will be stratified by site and age category (<65 and ≥65). For each participant, the web-based system will determine the treatment allocation from a static randomization table developed by the study randomization statistician prior to the start of the study. Only the study statistician and randomization system programmer will have access to the randomization table to minimize the risk of selection bias.

On April 16, 2019, the FDA ordered the removal of transvaginal mesh products including Uphold LITE from the market. The PFDN Steering Committee, with approval from NICHD and the DSMB, halted study enrollment for new patients and halted randomization into the TVM arm. At the time, there were 344 completed surgeries and 18 remaining consented or randomized participants who were awaiting surgery. Remaining patients that were randomized to TVM will be re-randomized to receive one of the two remaining arms (SC and NT). This randomization will occur in a ratio of 1:1

Beginning after April 18, 2019, the pending participants who have not undergone surgery will be notified that they will not be randomized into the TVM arm given that this arm is closed at this time. These participants will remain masked to their surgical randomization (SC and NT) until after their surgical repair.

All remaining participants were randomized by the original electronic randomization system. If a participant was randomized to TVM, then they were manually re-randomized to one of the other two treatment arms (SC or NT) in a ratio of 1:1 using a single non-stratified randomization list with permuted blocks of size 2 and 4. The new assignment is communicated to the unmasked site coordinator and updated in the Medidata RAVE Data Management System. This process including Quality Control (QC) procedures was documented in a Note to File dated April 23, 2019.

### 4.4 Masking and Data Lock

#### 4.4.1   General Masking Procedures

The participant will remain masked until after surgical repair. Those participants receiving sacral colpopexy will have abdominal incisions, making masking for this technique impossible. Because masking is impractical for the sacral colpopexy arm, the unmasking of the two vaginal arms will occur postoperatively. Given that the primary outcome is based primarily on outcome measures obtained by a masked examiner, unmasking of the participants should not bias the outcomes.

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.

To minimize biases, follow-up POP-Q measures and complications identified via a physical exam will be obtained by co-investigators or study nurses who are masked to the treatment group. Given the masked follow-up, assessment of efficacy outcomes will not occur until the 6-month visit and participants who all have had prior hysterectomy present with abdominal incisions should not risk unmasking. The masked examiner should not inquire about surgical type and the participant should be instructed not to discuss with the masked examiner. Any participant's concerns with details and specifics of the surgery should be forwarded to the participant's surgeon. We realize that unmasking may occur. If masking occurs, the site should complete a protocol deviation.

| Masking | Sacral Colpopexy Intervention | Vaginal Mesh Intervention | Native Tissue Intervention |
|---|---|---|---|
| **Participant** | Preop only | Preop only | Preop only |
| **Study Coordinator or Study Nurse** | No | No | No |
| **Telephone Interviewer (if applicable)** | No | No | No |
| **Study Surgeon** | No | No | No |
| **Anatomic Evaluator** | Yes | Yes | Yes |

#### 4.4.2   Database Lock

The ASPIRe (Medidata) database will be locked approximately 6 to 8 weeks after the last subject completes her 3-year study visit. The last surgery was performed on June 10, 2019, and that participant's final ASPIRe visit (3 year) is estimated to occur around June 2022. The details of the database lock are outlined in the separate database lock plan.

The E-ASPIRe (REDCap) database is maintained separately from the ASPIRe database and is outside the scope of this SAP.

### 4.5 Study Flow Chart of Assessments and Evaluations

The study flow chart of visits, assessments and evaluations is in Table 3.

Table 3 Study Flow Chart of Assessments and Evaluations

| MEASURE | TIMELINE OF MEASURES | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SCREEN | BASE LINE | PERI OP | 6 WK | 6M | 12 M | 18 M | 24 M | 30 M | 36 M | 42 M | 48 M | 54 M | 60 M |
| Demographics | X | | | | | | | | | | | | | |
| Medical History | X | | | | | | | | | | | | | |
| Physical Exam (height, weight and PVR) | | X | | | | | | | | | | | | |
| Surgeon's Report and Hospitalization | | | X | | | | | | | | | | | |
| POP-Q for Inclusion Criteria Unmasked Staff | | X | | | | | | | | | | | | |
| Study Surgery Status | | | X | | | | | | | | | | | |
| 6 Week Unmasked Evaluator Assessment (includes Complications Assessment) | | | | X | | | | | | | | | | |
| 6 Week Postoperative Recovery Assessment | | | | X | | | | | | | | | | |
| 6-60 Month Masked Evaluator Assessment (includes POP-Q and Complications Assessment) | | | | | X | X | X | X | X | X | X | X | X | X |
| Retreatment and Complications Assessments Unmasked (6-60 month visit form) | | | | | X | X | X | X | X | X | X | X | X | X |
| AE Review | | | X | X | X | X | X | X | X | X | X | X | X | X |
| Exams for Mesh Exposure | | | | X | X | X | X | X | X | X | X | X | X | X |
| Non-Pain Med Collection | | X | X | X | X | X | X | X | X | X | X | X | X | X |

| MEASURE | SCREEN | BASE LINE | PERI OP | 6 WK | 6M | 12 M | 18 M | 24 M | 30 M | 36 M | 42 M | 48 M | 54 M | 60 M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | TIMELINE OF MEASURES | | | | | | |
| Catheterization Follow-Up* | | | | Q 2 weeks if need for continuous catheterization post 6-Week Visit | | | | | | | | | | |
| Functional Activity Scale | | X | | | X | X | | X | | X | | X | | X |
| Surgical Pain Scale | | X | | X | X | X | | X | | X | | X | | X |
| Body Part Pain Score | | X | | X | X | X | | X | | X | | X | | X |
| PFDI-20 (includes POPDI-6, CRADI-8, UDI-6) | X | | | | X | X | | X | | X | | X | | X |
| PFDI-20 Question 3 Only | | | | | | | X | | X | | X | | X | |
| PFIQ-7 | | X | | | X | X | | X | | X | | X | | X |
| PGI-I | | | | | X | X | | X | | X | | X | | X |
| PISQ-IR | | X | | | X | X | | X | | X | | X | | X |
| BIPOP | | X | | | X | X | | X | | X | | X | | X |
| SF-12 | | X | | | X | X | | X | | X | | X | | X |
| DRS-PFD/SDS-PFD | | | | | | X | | X | | X | | | | |

*If 6 Week Post-Operative Visit indicates that participant requires continued catheterization, complete this every 2 weeks until catheterization ends

## 5       ANALYSIS POPULATIONS

Because randomization occurred prior to scheduling the surgery, some randomized patients did not receive surgery. Also, during surgery some patients were determined by the surgeon to be inappropriate candidates to receive the randomized surgery due to anatomic or other contraindications discovered at the time of the surgery and undeterminable prior to surgery. In these cases, the surgeon selected amongst the other two study surgeries based on physician preference.

Evaluations for superiority will follow a modified intent-to-treat (MITT) approach, and evaluations of non-inferiority will follow a per-protocol approach, which is the customary conservative approach for non-inferiority hypotheses.

To define the MITT population, the intent-to-treat population (all randomized participants) is modified to remove patients who were randomized but never received surgery.

The per-protocol population is the subset of the modified intent-to-treat population excluding all major protocol violations related to study treatment, which might be receipt of a treatment other than the randomized treatment and might be another type of violation.

In addition, given that some patients received a treatment selected by the physician when their randomized treatment was contra-indicated, the MITT and Per-protocol populations were further refined so that treatment switches that were conservative against the null hypothesis are included in the population, and treatment switches that might make the comparison biased against the null are excluded (see below).

Safety data will be evaluated for the safety population that will exclude randomized participants who discontinued the study without having surgery. Participants who received an alternate surgery (treatment switch) after an intraoperative complication or medical contraindication or received a type of surgery other than one of the 3 study treatments will be presented by their randomized (index) surgery. <u>Participants whose treatment was switched prior to surgery or very early during surgery with no initial intraoperative complication caused by the randomized (index) surgery or medical contraindication for the selected surgery (i.e. participant refusal or accidental switch) will be presented by their switched final procedure they received in safety displays</u>. Such participants will be identified by study PI review prior to database lock.

Specifically, populations are defined as follows, supported by the chart (**Table 3**) evaluating the impact of each type of treatment switching on the analysis.

Populations were defined considering all possible types of treatment switching in a blinded fashion without knowing if particular treatment switch scenarios occurred. As such, although all scenarios were considered, planned sensitivity analyses are specified a priori to be conducted only if there are at least 6 participants whose inclusion or values will be impacted by implementing that analysis.  Six was selected prior to data unblinding since if a treatment group had 6 such issues, 6 is 5% of the planned 120 participants per treatment arm. In addition, if the MITT-narrow population is identified as only being 2 or fewer patients smaller than the MITT population, then the benefit of applying the MITT-narrow population definition is a priori considered to be outweighed by the benefit of instead using the MITT population in place of the MITT-narrow population in order to achieve results that are substantially easier to present and interpret. After database lock and unblinding, but prior to conducting any efficacy analyses, the population sizes will be evaluated and any such evaluations that

deemed not to be done or to use a different analysis population based on these sample size rules will be documented. Two was selected prior to data unblinding as a very small number unlikely to impact analysis at a level that would warrant the complexities of having a separate analysis population.

**MITT:** all randomized patients, excluding patients who discontinued the study without having surgery. Treatment switchers are included in the treatment to which they were randomized.

**MITT- narrow**: all randomized patients, excluding patients who discontinued the study without having surgery and patients randomized to receive SC but received TM or vice-versa (randomized to TM but received SC). This is a subset of MITT. Treatment switchers are included in the treatment to which they were randomized.

**Per Protocol**: all randomized patients excluding all major protocol violations related to treatment including patients who discontinued the study without having surgery, all patients with treatment-related protocol violations, and all patients who did not receive their randomized treatment. This is a subset of the MITT-narrow.

**Aim 1 Per Protocol Sensitivity (A1PPS)**: all randomized patients excluding all major protocol violations related to treatment including patients who discontinued the study without having surgery, all patients with treatment-related protocol violations, and all patients who did not receive their randomized treatment, EXCEPT that TM to NTR switchers are NOT excluded. This is an expansion of the Per-protocol population.

**Safety**: all randomized patients, excluding patients who discontinued the study without having surgery. Treatment switchers are included in the treatment they were randomized or **received, based** on medical review of the intra-operative condition or event that led to the switch prior to database lock. This is the same participants as the MITT population.

A Venn diagram of the efficacy populations is as follows:
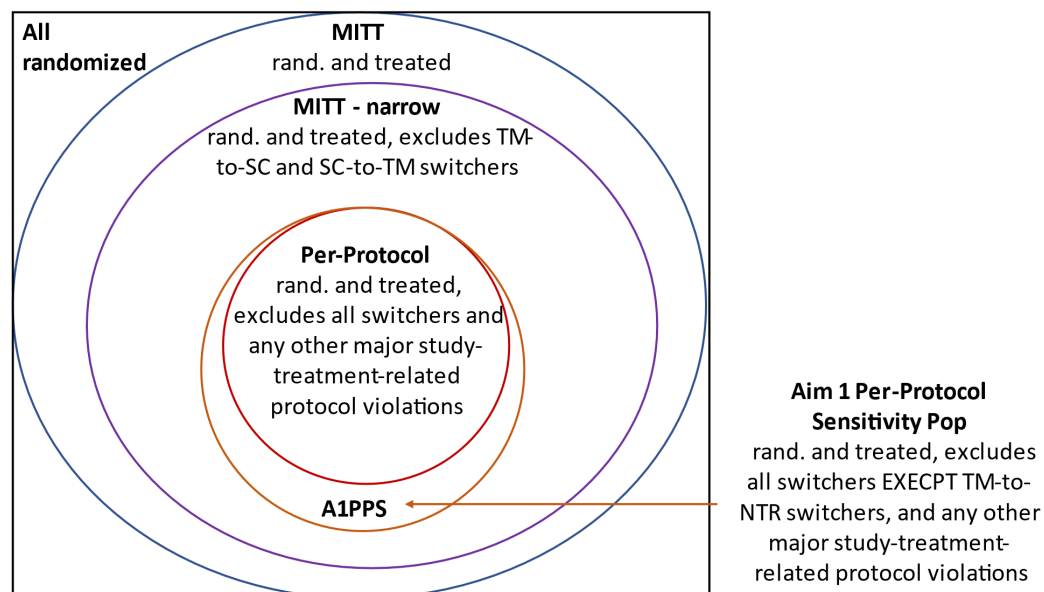
**Analysis Populations used for Analysis of the primary outcome (time to treatment failure), including sensitivity analyses**

Populations to be used in the analyses for the primary and secondary efficacy outcomes are summarized in Table 2:

**Table 2 Summary of Analysis Populations**

| Type | Primary Efficacy Outcome: Time to surgical failure | | | | All Secondary Efficacy Outcomes |
|---|---|---|---|---|---|
| | Aim 1<br><br>Non-inferiority of TVM to SC | Aim 1a<br><br>Superiority of TVM to SC | Aim 2<br><br>Superiority of TVM to NTR | Aim 3<br><br>Superiority of SC to NTR | |
| Primary | Per-Protocol | MITT | MITT-narrow | MITT-narrow | MITT |
| Sensitivity | Aim 1 PP Sensitivity Pop | -- | -- | -- | -- |
| Supportive | MITT | Per-Protocol | MITT and Per-Protocol | MITT and Per-Protocol | Per-Protocol |

Aim 1 (non-inferiority of TM to SC):

- The primary analysis will be conducted on the **per-protocol** population.

- A supportive secondary analysis will be conducted on the **MITT** population.
- Sensitivity analyses of the primary analysis will add patients to the per-protocol population who were randomized to TM but who (A) were never treated (imputed as failures with time to failure = 6 months), or (b) were switched during surgery from TM to NTR (**A1PPS** population). Each sensitivity analysis will only be completed if the additional patients add up to at least 5 patients.

Aim 1a (If TM is found to be non-inferior to SC, then a step-down comparison will evaluate superiority of TM to SC):

- The primary analysis will be conducted on the **MITT** population.
- A supportive secondary analysis will be conducted on the **Per-Protocol** population.
- A sensitivity of the primary analysis will add patients to the MITT population who were randomized to SC or TM but who were never treated (imputed as failures with time to failure = 6 months). Sensitivity analysis will only be completed if the additional patients add up to at least 5 patients.

Aims 2 and 3 (superiority of SC to NTR and superiority of TM to NTR):

- The primary analysis will be conducted on the **MITT-narrow** population.
- A supportive secondary analysis will be conducted on the **Per-Protocol** population, and also on the **MITT** population.
- A sensitivity of the primary analysis will add patients to the MITT-narrow population who were randomized to SC or TM but were never treated (imputed as failures with time to failure = 6 months). Sensitivity analysis will only be completed if the additional patients add up to at least 5 patients in either the SC or the TM treatment group.


All secondary efficacy evaluations will be analyzed on the **MITT** population (primary) and also on the **Per-Protocol** population (supportive).


Safety Evaluations:

- All safety analyses will be conducted on the safety population *(with selected switchers presented <u>as treated</u>)*.

**Table 3. Definition of analysis populations for patient groups based on status of treatment received**

| Comparison<br><br>Population: | Aim 1:<br>**Primary Non-Inferiority:**<br>**TM Non-Infer. to SC**<br><br>**Per-Protocol (PP)** | Aim 1A:<br>**Step-down Superiority**<br>**SC vs TM (2-sided)**<br><br>**Modified ITT (MITT)** | Aims 2 and 3:<br>**Primary Superiority:**<br>**SC vs NTR,**<br>**TM vs NTR (2-sided)**<br><br>**Modified ITT (MITT) - Narrow** |
|---|---|---|---|
| **1**. Randomized and treated with no major protocol violations | INCLUDE | INCLUDE | INCLUDE |
| **2**. Randomized and treated, but the assigned study treatment was **substantially modified or not completed, or there was some other major protocol violation related to the STUDY treatment** (However, there was NO switch to one of the other 2 study treatments) | EXCLUDE<br><br>Standard guideline is to exclude modifications and violations which could muddy true treatment differences.<br><br>Secondary Supportive analysis:<br><br>INCLUDE in MITT analysis (as planned is a relevant secondary question) | INCLUDE<br><br><br><br><br>Secondary Supportive analysis:<br><br>EXCLUDE in Per-Protocol analysis | INCLUDE<br><br><br><br><br>Secondary Supportive analysis:<br><br>EXCLUDE in Per-Protocol analysis |
| **3**. Randomized and withdrew pre-surgery: women with no surgery and no follow-up info | EXCLUDE<br><br>Sensitivity analysis: | EXCLUDE<br><br>Sensitivity analysis: | EXCLUDE<br><br>Sensitivity analysis: |

| | | | |
|---|---|---|---|
| For a <u>sensitivity analysis</u>, to be conservative, we should assume they did not get surgery for a reason related to the type of surgery (Biased) and impute a worst outcome. This sensitivity analysis will only be done if there are at least 6 participants in this category in a treatment group. | Do a sensitivity analysis adding <u>TM cases only</u> to the primary PP analysis as <u>failure at 6 months</u> to test if TM is NI to SC cases – this stacks the deck in favor of SC. | Do a sensitivity analysis adding <u>SC and TM cases only</u> to the primary MITT analysis, as failures. This does not stack the deck in favor of either treatment group. | Do a sensitivity analysis adding <u>SC and TM cases only</u> to the primary MITT-narrow analysis, as <u>failure at 6 months</u> - this stacks the deck in favor of NTR (presumed to be the worst treatment) |
| **Randomized but received AN ALTERNATE protocol treatment at MD discretion** (cases 4-6): | | | |
| **4.** NTR switched to TM or SC<br><br>This should never happen, as NTR should never be contra-indicated. Switching would be BIASED as other 2 treatments are expected to be superior | N/A (exclude from PP pop definition) | N/A (include in MITT definition) | INCLUDE<br><br>This is assumed to be conservative as switch is expected to improve outcome of NTR and thus make planned superiority comparison tougher<br><br>Secondary Supportive analysis:<br>EXCLUDE from a PP analysis |
| **5.** SC or TM switched to NTR<br><br>Although the protocol did not specify a plan of what to do if SC or TM could not be done, including switchers in the ITT defines the primary hypothesis to compare "start out planning | EXCLUDE<br><br>Major protocol violation, it is assumed this would make the outcome worse for the switcher and muddy the comparison. | INCLUDE<br><br>This is intent-to-treat, or ICH E9R1 "treatment policy" and compares the treatments as they were planned to be. | INCLUDE<br><br>This is assumed to be conservative as switch is expected to worsen outcome of SC/TM and thus make planned superiority comparison tougher. |

| | | | |
|---|---|---|---|
| to do SC" versus "start out planning to do TM" Note that NT is expected to have worse outcome. | <u>Sensitivity analyses:</u><br><br>1: Do a sensitivity analysis INCLUDING these cases in the primary PP analysis <u>for TM group only (excluding the SC cases only)</u>, to see if it is NI to per-protocol SC cases – this stacks the deck in favor of SC. This sensitivity analysis will only be done if there are at least 6 participants in this category to add back.<br><br><u>Secondary Supportive analysis:</u> Include both groups in MITT analysis. (as-planned is a relevant secondary question) | <u>Secondary supportive analysis:</u><br><br>EXCLUDE in a PP analysis | <u>Secondary Supportive analysis:</u><br><br>EXCLUDE both from a PP analysis |
| **6.** SC switched to TM<br><br>or TM switched to SC<br><br><br>The protocol did not specify a plan of what to do if SC or TM could not be done. Since the switching to another treatment was surgeon preference, then the outcome for switchers must be considered biased, and primary analysis is focused on "as-treated". | EXCLUDE<br><br>This is a straightforward major protocol violation for the non-inferiority analysis. Since the primary goal of the trial is to compare TM to SC, switching muddies that comparison<br><br><br>Secondary Supportive analysis: Include both groups in MITT analysis. As-planned is a relevant secondary evaluation | INCLUDE<br><br>This is conservative as switch makes superiority comparison tougher<br><br><br><u>Secondary supportive analysis:</u> EXCLUDE in a Per Protocol analysis | EXCLUDE<br><br>Although SC and TM MIGHT be similar, they also might not, and so mixing up the groups would not be appropriate for understandable estimates of treatment effect vs NTR.  Therefore, these cases are excluded from the MITT-narrow in primary comparison with NTR<br><br><br><u>Secondary supportive analysis:</u><br><br>INCLUDE these cases for an "as-planned" analysis using the MITT population |

# 6 SAMPLE SIZE DETERMINATION

Sample size calculations were generated under the assumptions that the three study arms are 1) mesh augmented sacral colpopexy, 2) transvaginal mesh repair, and 3) native tissue repair for vault prolapse. Further assumption was that the study would have an overall Type I error rate of 0.05, that a randomization ratio of 1:1:1 is preferred for feasibility reasons, and that the Type I error (alpha) would be distributed among the 3 following three hypotheses related to the 3 treatment arms in a way that optimizes sample size and at the same time is scientifically defensible:

- H1: transvaginal mesh is noninferior to mesh augmented sacral colpopexy [spend alpha of 0.03 using a one-sided noninferiority test]

- H2: mesh augmented sacral colpopexy is superior to native tissue repair [spend alpha of 0.01 using a two-sided test of superiority]

- H3: transvaginal mesh repair is superior to native tissue repair [spend alpha of 0.01 using a two-sided test of superiority]

The sample size estimates also assume that the primary analyses for each of the three hypothesis tests will be based on a survival analysis model with a 2-year recruitment period and a 3-year follow-up period after the last participant is randomized and that loss to follow-up on each arm will be no more than 5% per year, that the 2-year success rates for mesh-augmented sacral colpopexy and transvaginal mesh repair will be 80% and that the 2-year success rate for native tissue will be 60%. These 2-year success rates represent hazards of 0.1116 and 0.2554, respectively, under the assumption that the failures follow an exponential survival model. The non-inferiority margin was selected to show that transvaginal mesh success rate at 2 years is no worse than a 15% difference from sacral colpopexy, or 65%. This success rate represents a hazard of 0.2154 and the noninferiority margin corresponds to a hazard ratio of 1.93.

Under the assumptions outlined above, the hypothesis test that drives the sample size is the test on noninferiority of transvaginal mesh to mesh-augmented sacrocolpopexy. A sample size of 121 participants per arm will provide **85%** power to demonstrate noninferiority under the assumptions outlined above. Assuming that 121 participants will also be randomized and treated to native tissue repair, the overall sample size of 363 participants will provide greater than **93%** power to demonstrate that each of the mesh augmented arms is superior to native tissue repair under the assumptions outlined above.

After April 18, 2019, the study was closed to new surgeries of transvaginal mesh. The remaining 18 consented or randomized patients were randomized to receive either sacral colpopexy or native tissue repair in a 1:1 ratio. This modification is expected to have minimal to no impact on the planned statistical power, given that the recruitment period had been extended from 2 years to 36 months (thereby increasing power), and the final sample size of the TVM group will be approximately 95% of the planned size, while the other two study arms are expected to meet or even slightly exceed the planned sample size.

# 7    STATISTICAL / ANALYTICAL ISSUES

## 7.1 General Statistical Methods

All statistical methods will be based on the International Conference on Harmonisation (ICH) E9 document "Statistical Principles for Clinical Trials" unless otherwise requested by journal. All statistical computations will be performed, and data summaries will be created using SAS 9.4 or higher. If additional statistical packages are required, these will be identified in the presentation of results.

In summary tables of continuous or ordinal numeric variables, the mean, median, 25th and 75th percentiles (Interquartile range IQR), confidence interval (CI), standard deviation (SD) and standard error (SE) will be presented to one more decimal place than the original data.

In summary tables of categorical variables, counts and percentages will be used. The denominator for each percentage will be the number of subjects within the population treatment group unless otherwise specified.

All hypothesis testing will be carried out at the 5% (2-sided) significance level and confidence intervals will be 95% unless otherwise specified.

P-values will be rounded to three decimal places. P-values less than 0.001 will be reported as < 0.001 and greater than 0.999 will be reported as 0.999.

## 7.2 Adjustments for Covariates

Indicator variables for the study stratification factors of site (see site pooling in multisite studies, below) and age category (<65 and ≥65) will be included as covariates in most efficacy analyses performed for this study.

## 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. Sensitivity analyses addressing the impact of missing data will be specifically specified.

**Impact of COVID 19** Some of the follow-up period of this study coincided with the COVID pandemic (March 2020 – July 2022).  At varying times across sites during the pandemic, patients were unable or unwilling to attend in-person visits, or clinic site personnel were unable to conduct in-patient visits. Accommodations were made allowing for remote phone contact visits rather than in-person visits. All such visits were noted as a COVID-related protocol violation.  As a result, there are more study visits missed than ordinarily expected in pre-pandemic, and also a percentage of study visits done remotely in which the subjective assessments of the primary outcome were recorded, but the anatomical POP_Q exam of prolapse was not able to be performed.  An effort was made by all sites to have participants complete a final POP-Q exam prior to study completion whenever possible. Assessments of the impact of missingness from the COVID pandemic on the primary outcome and sensitivity analyses are specified in applicable sections below.

## 7.4 Interim Analyses and Data Monitoring

Given the unique nature of a three-arm randomized surgical trial, the Steering Committee initially considered the development of stopping rules for this protocol. The working group also considered the performance of a pilot study, but the disadvantage of enrollment of eligible patients that would not be used in the analysis was believed to offer minimal advantages compared to the development of an ancillary recruitment plan for slow recruitment versus futility/stopping rules for minimal recruitment. After careful consideration, the Steering Committee elected not to have formal stopping rules but to allow the DSMB and Steering Committee to monitor recruitment on a regular basis (similar to other PFDN studies). Thus, enrollment will initially be assessed 6 months after initiating the study. Based on prior PFDN studies, a goal of 20% of total enrollment will be set. If this goal is not obtained, additional non-network sites will be considered from a pool of sites pre-approved prior to this time.

The working group also considered the possibility of having one or more formal interim analyses that would provide for early stopping for a demonstrated efficacy benefit but rejected that option for two reasons. First, the study is designed to enroll only 121 participants per treatment arm, and the group was concerned that stopping the study with fewer participants than 121 per arm, even with relatively small p-values, might limit the impact of any findings on clinical practice because of the small sample sizes. Second, the hypothesis tests among the 3 arms, with a combination of superiority and non-inferiority hypotheses, will be complicated to explain to the clinical community if all 3 arms enroll to study completion. Stopping one of the arms early and testing two of the 3 hypotheses with a less than full sample and taking the other two arms to completion would complicate the explanation of the study even further. Given the small likelihood of stopping a study early for 2 of the 3 hypotheses, the complications of the approach appear to outweigh any potential benefits.

## 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. 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)
- Visit window outliers
    - o 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 masked data.
- Classification of complications ***and*** adverse events based on indicator variable definitions from the SAP. The summary will include a count and percentage for each of the

complications listed in manuscript table shells as well as a table that shows the percentage of individuals having different numbers of complications. The summary will also include the distribution of Dindo scores, and the listing of complications that generated the Dindo scores.

- After completion of manual review and prior to final database lock, queries will be issued as follows:
  o Change all exposure or erosion complications to explicitly state "for/from prolapse surgery" or "for/from sling" (case insensitive) as appropriate.
  o Assure that all exposure or erosions reported on the AE log were also recorded on a complications table as indicated in the preceding bullet.

## 7.6 Multisite Studies

This study is conducted in nine sites in the Pelvic Floor Disorders Network. For this multisite study, randomization of study participants was stratified within site. Consequently, for all model-based primary and secondary analyses, site will be included as a fixed effect in the models.

Pooling of sites is defined as grouping the subjects from one or more sites together and considering these subjects as representing a single site for the purpose of summary and analysis. The purpose of pooling is to ensure that each site contains a sufficient number of subjects for sample size requirements of statistical analysis methods that consider site as a factor.

For all statistical analyses, sites with fewer than 25 randomized patients will be pooled based on geographical location. Specifically, UCSD (n=17) and sub-site Kaiser-Downey (n=11) will be pooled with Kaiser- San Diego (n= 61), New Mexico (n=1) will be pooled with UTSW (N=26), and Cleveland Clinic (n=3) will be pooled with University of Pittsburg (N=25), resulting in 7 pooled sites.

As a supplemental analysis associated with the primary outcome, we will examine descriptively whether the treatment effect varies across pooled 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 are 3 formal primary hypothesis tests and 1 formal step-down test for this study, for the primary outcome of time to treatment failure. The overall Type-1 error (alpha) is maintained at 0.05 by distributing the alpha among the 3 hypotheses as follows:

- H1: transvaginal mesh is noninferior to mesh augmented sacral colpopexy [spend alpha of **0.03** using a one-sided noninferiority test and **an upper bound 1-sided 97% CI**, which is equivalent to the upper bound of a 2-sided 94% CI.]

- H1a (tested only if H1 null hypothesis of inferiority is rejected): transvaginal mesh is superior to mesh augmented sacral colpopexy [spend alpha of **0.03** using a two-sided test of superiority and a **97% CI**]

- H2: mesh augmented sacral colpopexy is superior to native tissue repair [spend alpha of **0.01** using a two-sided test of superiority and a **99% CI**]

- H3: transvaginal mesh repair is superior to native tissue repair [spend alpha of **0.01** using a two-sided test of superiority and a **99% CI**]

All statistical tests for secondary, and exploratory analyses will be conducted at a 5% type I error rate (two-sided) and no adjustments for multiplicity will be made. P-values and confidence intervals are provided for descriptive purposes only and are not to be evaluated relative to statistical significance.

### 7.8 Assessment Windows

Ideally, the time from randomization to surgery will be no longer than 6 weeks (42 days). If more than 6 weeks transpires between randomization and surgery, a protocol deviation will be completed. Similarly, the time from consent to surgery should be no longer than 4 months. If more than 4 months transpires between randomization and surgery, a protocol deviation will be completed. Regardless of timing, screening and baseline assessments were not repeated.

If surgery is not scheduled within 6 months after randomization due to patient-specific reasons unrelated to the masked treatment assignment, the patient may be discontinued from the trial.

The 6-week post-op visit had a -1 week, +2 week visit window. All other visits were completed at 6-month intervals following the surgery 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 Consort 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.

### 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. Treatment groups will not be compared statistically

# 9      EFFICACY ANALYSES

## 9.1 Overview of Efficacy Analyses Methods

All analyses of the primary efficacy outcome will be performed using populations as specified in Section 5. Analyses of all secondary efficacy outcomes will be performed using the MITT population unless otherwise specified (see section 5, analysis populations). All efficacy variables will be listed by subject within study center and assessment time.  The data will be summarized by treatment group. Additional details are provided in the specific sections below.

## 9.2 Assessment of Non-Inferiority and superiority of TVM to SC

Non-inferiority of TVM to SC will be assessed only for the primary outcome of time to treatment failure.  All secondary and exploratory assessments will be evaluated descriptively with a standard 90% confidence interval.

For the primary outcome, non-inferiority is to be assessed using a 1-sided $p=0.03$ test of the hazard ratio relative to the 1.93 non-inferiority margin. This will be evaluated via a p-value for the 1-sided test that the TVM to SC hazard ratio is not more than 1.93, corresponding to the upper bound of a 2-sided 94% CI being less than the non-inferiority margin of 1.93. The purpose of the non-inferiority test is to assess if we are 97% confident that the TVM hazard is not more (larger) than SC hazard by an amount of 1.93.  If the upper bound of the CI is less than 1.93, (corresponding to the 1-sided p-value < 0.03): then the null hypothesis is rejected, and non-inferiority is assumed. The lower bound of the CI is not evaluated.

If non-inferiority is assumed, then a further step-down test of superiority of TVM to SC is completed by a 2-sided test with $p=0.03$, corresponding to a 2-sided 97% confidence interval. Note that while non-inferiority is evaluated with a protocol-specified 1-sided test, superiority is evaluated with a 2-sided test.

### 9.3 Efficacy Variables

| Variable | CRF Source | Type | Timepoint(s) | Definition |
|---|---|---|---|---|
| **Primary Efficacy Outcome** | | | | |
| Time to surgical failure | CRF-13, CRF-12, CRF-14/37 | Continuous | 6 months through 60 months | Among subjects who experience a surgical failure (see definition below) starting at the 6 month post-operative visit through their last attended ASPIRE visit , the time to surgical failure is defined as the number of months (i.e. months as intended by the visit schedule) from the surgical intervention to first post-operative visit with surgical failure.<br><br>Surgical failure is defined at each attended post-operative visit if any ONE of the following criteria is met:<br><br>- Retreatment failure is defined as retreatment for prolapse via additional surgery or pessary as reported on coordinator follow-up CRF-13<br><br>- Anatomic failure is defined as any prolapse beyond the hymen (i.e. POPQ points Ba, Bp, or C > 0.0 cm) as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination<br><br>- Symptomatic failure is defined as experiencing bothersome vaginal bulge symptoms as reported on CRF-14/37. Bothersome vaginal bulge symptoms are defined as a positive response to Question 3 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 greater than "Not at all" (i.e. "Somewhat", "Moderately", or "Quite a bit") |
| | | | | Participants not meeting the criteria for surgical failure at the attended post-operative visit will be considered not a surgical failure, regardless of missingness of the 3 components of the surgical outcome (i.e. retreatment failure, anatomic failure, and symptomatic failure). |
| | | | | The time to surgical failure will be missing for all participants who are considered not a surgical failure at each of the attended post-operative visits 6 months through their last attended ASPIRE visit scheduled in 6-month intervals up to 60 months. |
| | | | | Note that since ASPIRE followed participants until the last participant completed their 36 month visit, ASPIRE study completers have a varying amount of follow-up time from 36 to 60 months. |
| | | | | For survival analysis, participants with no failure by their last attended study visit are censored, and with follow-up duration equal to the number of months (i.e. months as intended by the visit schedule) corresponding to the last attended visit. |
| | | | | Note that due to COVID, there are more missed visits than would have been expected pre-pandemic, as well as remotely-attended visits with missing POP-Q exams. The calculation of the primary outcome is unchanged. |

| Secondary Efficacy Outcomes | | | | |
|---|---|---|---|---|
| Indicator of surgical failure at post-operative visit under the permanent failure state assumption | CRF-13, CRF-12, CRF-14/37 | Binary | 1-, 2-, 3- , 4-, 5-years | ***At each timepoint***, surgical failure under the permanent failure state assumption is defined if any one of the three criteria (retreatment, anatomic, or symptomatic) is met at the timepoint's corresponding attended post-operative visit ***or*** at any prior attended post-operative visits (i.e. 6-month post-operative visits 6 months up to 60 months):<br><br>Participants not meeting the criteria for surgical failure at the timepoint's corresponding post-operative visit (if attended) and at all prior attended post-operative visits will be considered not a surgical failure, regardless of missingness of the 3 components of the surgical outcome at any of the attended post-operative visits through the timepoint. Participants who discontinued prior to the 6-month post-operative visit will be missing the permanent failure state surgical outcome at all timepoints.<br><br>By this definition, a subject can be defined as a surgical failure at timepoints corresponding to missed post-operative visits or post-operative visits occurring after subject's discontinuation from the study IF the subject had a surgical failure prior to the missed visit or discontinuation. |
| Initial failure type | CRF-13, CRF-12, CRF-14/37 | Nominal Categorical (4 levels) | 1-, 2-, 3- , 4-, 5-years | At each timepoint, subjects who experience a surgical failure under the permanent failure state assumption, failure type will be defined for the initially occurring surgical failure.<br><br>Each initial surgical failure will be categorized into the following 4 types: "retreatment failure", "anatomic |

| | | | | failure", "symptomatic failure", and "anatomic and symptomatic failure" where retreatment failure corresponds to retreatment for prolapse, symptomatic failure corresponds to bothersome vaginal bulge symptoms, and anatomic failure corresponds to prolapse beyond the hymen as described in the surgical failure under the permanent failure state assumption definition above. Retreatment failure type is prioritized above the anatomic and/or symptomatic failure types. |
|---|---|---|---|---|
| Indicator of surgical failure at post-operative visit under the transient failure state assumption | CRF-13, CRF-12, CRF-14/37 | Binary | 1-, 2-, 3- , 4-, 5-years | At each timepoint, surgical failure under the transient failure state assumption is defined if any one of the three criteria (retreatment failure at that visit or any prior visit, anatomic failure at that visit, or symptomatic failure at that visit) is met at that visit. |
| | | | | Participants not meeting the criteria for surgical failure at the timepoint will be considered not a surgical failure, regardless of missingness of the 3 components of the surgical outcome at the timepoint's corresponding attended post-operative visit. |
| | | | | Participants who have discontinued from the study by timepoint's corresponding post-operative visit will be missing the transient failure state surgical outcome at the timepoint. |
| | | | | Participants who missed the post-operative visit corresponding to the timepoint <u>without a retreatment failure at any prior attended post-operative visit</u> will also be missing the transient failure state surgical outcome at the timepoint. |

| | | | | By this definition, a subject can be defined as a surgical failure at timepoints where the corresponding post-operative visit is missed IF the subject is still participating in the study and had a surgical failure due to retreatment failure at any prior attended post-operative visit (i.e. post-operative visits 6 months up to 60 months). |
|---|---|---|---|---|
| At-visit failure type | CRF-13, CRF-12, CRF-14/37 | Nominal Categorical (4 levels) | 1-, 2-, 3- , 4-, 5-years | At each timepoint, subjects who experience a surgical failure under the transient failure state assumption, failure type will be defined for the surgical failure occurring at the timepoint.<br><br>Each surgical failure will be categorized into the following 4 types: "retreatment failure", "anatomic failure", "symptomatic failure", and "anatomic and symptomatic failure" where retreatment failure corresponds to retreatment for prolapse, symptomatic failure corresponds to bothersome vaginal bulge symptoms, and anatomic failure corresponds to prolapse beyond the hymen as described in the surgical failure under the transient failure state assumption definition above. Retreatment failure type is prioritized above the anatomic and/or symptomatic failure types. |
| Indicator of retreatment (surgery or pessary) for prolapse | CRF-13 | Binary | 6-months through 60 months | Through the timepoint of follow-up, based on retreatment for pelvic organ prolapse as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years, the indicator is defined as follows:<br><br>1 = Yes, if **surgery or pessary** is reported for questions A3a or A3b respectively at any attended post-operative visit |

| | | | | |
|---|---|---|---|---|
| | | | | 0 = No, if subject attended at least one post-operative visit 6 months or after and reported no additional treatment (surgery or pessary) for pelvic organ prolapse was performed for question A1 across all attended post-operative visits. Subjects with missing responses to question A1 are included as long as the subject has at least one non-missing response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| | | | | *Note: Prior to and upon database lock, any other free text responses recorded on CRF-13 Question B5 "Since surgery, did you have a reoperation* (P)?" specifying retreatment for POP will be assessed, and if any, included in this definition at the discretion of the principal investigator. |
| Indicator of surgical retreatment for prolapse | CRF-13 | Binary | 6-months through 60 months | Through the timepoint of follow-up, based on retreatment for pelvic organ prolapse as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years, the indicator is defined as follows: |
| | | | | 1 = Yes, if **surgery** is reported for question A3a at any attended post-operative visit |
| | | | | 0 = No, if subject attended at least one post-operative visit 6 months or after and reported no additional surgery treatment for pelvic organ prolapse was performed for question A1 or subject reported additional treatment for question A1 and the additional treatment |

| | | | | |
|---|---|---|---|---|
| | | | | was not surgery as reported for question A3a across all attended post-operative visits. Subjects with missing responses to question A1 are included as long as the subject has at least one non-missing response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Indicator of pessary-only retreatment for prolapse | CRF-13 | Binary | 6-months through 60 months | Through the timepoint of follow-up, based on retreatment for pelvic organ prolapse as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years, the indicator is defined as follows:<br><br>1 = Yes, if **only pessary** is reported for question A3b at any attended post-operative visit. Subjects with surgical retreatment and pessary retreatment for prolapse as reported for questions A3a and A3b respectively are *not* included here.<br><br>0 = No, else if subject attended at least one post-operative visit 6 months or after and reported no additional pessary treatment or both surgery AND pessary treatment) for pelvic organ prolapse was performed for question A1 or subject reported additional treatment for question A1 and the additional treatment was not pessary as reported for question A3b across all attended post-operative visits. Subjects with surgical retreatment and pessary retreatment for prolapse as reported for questions A3a and A3b respectively *are* |

| | | | | |
|---|---|---|---|---|
| | | | | included here. Subjects with missing responses to question A1 are included as long as the subject has at least one non-missing response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Indicator of retreatment (surgical or non-surgical) for urinary incontinence | CRF-13 | Binary | 6-months through 60 months | Through the timepoint of follow-up, based on retreatment for urinary incontinence as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years, the indicator is defined as follows:<br><br>1 = Yes, if **surgery, medications, pessary for incontinence, supervised pelvic muscle exercises, time voiding & fluid management, periurethral injection, Botox injection, e-stim, or other specified** is reported for the A5 question group at any attended post-operative visit<br><br>0 = No, else if subject attended at least one post-operative visit 6 months or after and reported no additional treatment for urinary incontinence was performed for question A4 across all attended post-operative visits. Subjects with missing responses to question A4 are included as long as the subject has at least one non-missing response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |

| Anatomic POPQ point Ba measurement | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60-months | At each timepoint, POPQ point Ba as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit. |
|---|---|---|---|---|
| Anatomic POPQ point Bp measurement | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60-months | At each timepoint, POPQ point Bp as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit. |
| Anatomic POPQ point C measurement | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60-months | At each timepoint, POPQ point C as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit. |
| Anatomic POPQ point GH measurement | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60-months | At each timepoint, POPQ point GH as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit. |
| Anatomic POPQ point Aa measurement | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60-months | At each timepoint, POPQ point Aa as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit. |
| Anatomic POPQ point Ap measurement | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60-months | At each timepoint, POPQ point Ap as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit. |
| Anatomic POPQ point Pb measurement | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60-months | At each timepoint, POPQ point Pb as reported on the evaluator (blinded) follow-up CRF-12 during POPQ |

| | | | | examination at the timepoint's corresponding attended post-operative visit. |
|---|---|---|---|---|
| Anatomic total vaginal length (TVL) measurement | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60- months | At each timepoint, total vaginal length (TVL) as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit. |
| Indicator of anatomic POPQ point C measurement > -2/3 of anatomic total vaginal length (TVL) measurement ("Apical Descent") | CRF-12 | Binary | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60- months | At each timepoint, based on POPQ point C and total vaginal length (TVL) as reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit, the indicator is defined as follows: <br><br> 1 = Yes, if POPQ point C > -2/3 x TVL, both POPQ point C and TVL are non-missing <br><br> 0 = No, else if POPQ point C <= -2/3 x TVL, both POPQ point C and TVL are non-missing <br><br> . = Missing, otherwise |
| Maximum extent of prolapse: leading edge of Ba, Bp, C, | CRF-12 | Continuous | 6-, 12-, 18-, 24-, 30-, 36-, 42-, 48-, 54-, 60- months | At each timepoint, POPQ maximum value of points Ba, Bp, or C reported on the evaluator (blinded) follow-up CRF-12 during POPQ examination at the timepoint's corresponding attended post-operative visit. |
| Change from baseline in PFDI-20 UDI subscale | PFDI-20 | Continuous | 6-, 12-, 24-, 36-, 48, 60- months | The PFDI-20 UDI subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2006). Specifically, average the responses for questions Q15 through Q20 (0=no, 1=yes: not at all, 2=yes: somewhat, 3=yes: moderately, 4=yes: quite a bit) and multiply by 25. If there are missing |

| | | | | |
|---|---|---|---|---|
| | | | | responses, then the score is the average of the non-missing responses multiplied by 25. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in PFDI-20 CRADI subscale | PFDI-20 | Continuous | 6-, 12-, 24-, 36-, 48, 60-months | The PFDI-20 CRADI subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2006). Specifically, average the responses for questions Q7 through Q14 (0=no, 1=yes: not at all, 2=yes: somewhat, 3=yes: moderately, 4=yes: quite a bit) and multiply by 25. If there are missing responses, then the score is the average of the non-missing responses multiplied by 25. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in PFDI-20 POPDI subscale | PFDI-20 | Continuous | 6-, 12-, 24-, 36-, 48, 60-months | The PFDI-20 POPDI subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2006). Specifically, average the responses for questions Q1 through Q6 (0=no, 1=yes: not |

| | | | | at all, 2=yes: somewhat, 3=yes: moderately, 4=yes: quite a bit) and multiply by 25. If there are missing responses, then the score is the average of the non-missing responses multiplied by 25. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
|---|---|---|---|---|
| Change from baseline in PFDI-20 Global Score | PFDI-20 | Continuous | 6-, 12-, 24-, 36-, 48, 60-months | The PFDI-20 Global Score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2006). Specifically, sum the scores from the UDI, CRADI, and POPDI subscales. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in PFIQ-7 UIQ subscale | PFIQ-7 | Continuous | 6-, 12-, 24-, 36-, 48, 60-months | The PFIQ-7 UIQ subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2006). Specifically, average the responses for questions Q1a through Q7a (0=not at all, ... 3=quite a bit) and multiply by 100/3. If there are missing responses, then the score is the average of the non-missing responses multiplied by 100/3. The outcome will then be computed as the difference in score at each |

| | | | | timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
|---|---|---|---|---|
| Change from baseline in PFIQ-7 CRAIQ subscale | PFIQ-7 | Continuous | 6-, 12-, 24-, 36-, 48, 60-months | The PFIQ-7 CRAIQ subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2006). Specifically, average the responses for questions Q1b through Q7b (0=not at all, ..., 3=quite a bit) and multiply by 100/3. If there are missing responses, then the score is the average of the non-missing responses multiplied by 100/3. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in PFIQ-7 POPIQ subscale | PFIQ-7 | Continuous | 6-, 12-, 24-, 36-, 48, 60-months | The PFIQ-7 POPIQ subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2006). Specifically, average the responses for questions Q1c through Q7c (0=not at all, …, 3=quite a bit) and multiply by 100/3. If there are missing responses, then the score is the average of the non-missing responses multiplied by 100/3. The outcome will then be computed as the difference in score at each |

| | | | | |
|---|---|---|---|---|
| | | | | timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in PFIQ-7 Global Score | PFIQ-7 | Continuous | 6-, 12-, 24-, 36-, 48, 60-months | The PFIQ-7 Global Score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2006). Specifically, sum the scores from the UIQ, CRAIQ, and POPIQ subscales. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Indicator of improvement based on Patient Global Impression of Improvement (PGII) | PGII | Binary | 6-, 12-, 24-, 36-, 48, 60-months | At each timepoint, based on PGII as reported on the PGII at the timepoint's corresponding attended post-operative visit, the indicator is defined as follows: 1 = Yes, if improvement of 1=very much better or 2=much better are reported on the PGII 0 = No, else if improvement of 3=a little better, …, 7=very much worse are reported on the PGII . = Missing, otherwise |
| Change from baseline in PISQ-IR sexually active – average score | PISQ-IR | Continuous | 6-, 12-, 24-, 36-, 48, 60-months | The PISQ-IR sexually active – average score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using |

| | | | | standard scoring algorithms for the IUGA method (Constantine 2017). Specifically, for subjects with a sexual partner (i.e. Q12: 1=yes) sum the scores for questions Q7, Q8a-c, Q9, Q10, Q11, Q13, Q14a-b, Q15, Q16, Q17, Q18, Q19a-c, Q20a-d using reverse scores for Q8b-c, Q9, Q11, Q14a-b, Q16, Q17, Q18, and Q19a-c (Q7, Q8a-c, Q9, Q11: 1=never, ..., 5=[almost] always; Q10: 1=much less intense, ..., 5=much more intense; Q13: 1=all of the time, ..., 4=hardly ever/rarely; Q14a-b: 1=very positive, ..., 4=very negative; Q15: 1=never, ..., 5=always; Q16: 1=daily, ...,5=never; Q17: 1=very high, ..., 5=very low or none at all; Q18: 1=not at all, ..., 4=a lot; Q19a-c: 1=satisfied, ..., 5=dissatisfied; Q20a-d: 1=strongly agree, ..., 4=strongly disagree). If there are more than 10 missing responses, then a total score is not calculated. Similarly, for subjects without a sexual partner (i.e. Q12: 2=no) sum the scores as described above excluding Q13 and Q14a-b from the sum as these questions are not collected among subjects without a sexual partner. If there are more than 9 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 each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active. |
| --- | --- | --- | --- | --- |

| Change from baseline in Functional Activity Assessment Scale (FAS) overall score | AAS | Continuous | 6-, 12-, 24-, 36-, 48, 60- months | The Functional Activities Assessment Scale (FAS) score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (McCarthy 2005). Specifically, sum the scores for questions Q1, …, Q13 (1=no difficult, …, 5=not able to do it; responses of 6=did not do it for other reasons will not be included in the sum). If there are more than 6 missing responses, then a total score is not calculated. To handle missing values, the raw score is obtained by dividing the sum by the number of items answered. The final score is obtained by scaling the raw score to range from 0 to 100 using the following formula: |
|---|---|---|---|---|
| | | | | $(65 – \text{raw score})/52 \times 100$ |
| | | | | Note: 65 is the maximum raw score, and 52 is the range of possible scores (65-13=52). |
| | | | | The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in SF-12 physical component PCS Subscale | SF-12 | Continuous | 6-, 12-, 24-, 36-, 48, 60- months | At baseline and at each timepoint, based on physical component items as reported on the SF-12 at the timepoint's corresponding attended post-operative visit, the scoring of the physical component subscale is defined in **Attachment 1.** |

| | | | | The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
|---|---|---|---|---|
| Change from baseline in SF-12 mental component MCS Subscale | SF-12 | Continuous | 6-, 12-, 24-, 36-, 48, 60- months | At baseline and at each timepoint, based on mental component items as reported on the SF-12 at the timepoint's corresponding attended post-operative visit, the scoring of the mental component subscale is defined in *Attachment 1*. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Indicator of new or worsening stress urinary incontinence (SUI) | CRF-12, CRF-13, CRF-96 | Binary | surgery through 3 years | Through the timepoint of follow-up, based on the Adverse Event Log (CRF-96) or specific complications as reported on the masked post-operative CRF-12 and coordinator follow-up CRF-13 collected at 6-weeks and in 6-month intervals through 3 years, the indicator is defined as follows: 1 = Yes, if **new or worsening SUI** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or "Stress urinary incontinence"" in the term reported on the AE Log |

| | | | | |
|---|---|---|---|---|
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| | | | | *Note: Prior to and upon database lock, any stress urinary incontinence events identified by the MedDRA coded preferred term "stress urinary incontinence" ) with start date after surgery collected ONLY on the AE log and not reported on any of the complications tables will be assessed, and if any, included in this definition at the discretion of the principal investigator. |
| Indicator of new or worsening urgency urinary incontinence (UUI) | CRF-12, CRF-13, CRF-96 | Binary | surgery through 3 years | Through the timepoint of follow-up, based on the Adverse Event Log (CRF-96) or specific complications as reported on the masked post-operative CRF-12 and coordinator follow-up CRF-13 collected at 6-weeks and in 6-month intervals through 3 years, the indicator is defined as follows:<br><br>1 = Yes, if **new or worsening UUI** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or "Urge incontinence" in the term reported in the AE log. |

| | | | | |
|---|---|---|---|---|
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise<br><br>*Note: Prior to and upon database lock, any new or worsening urgency urinary incontinence events identified by the MedDRA coded preferred term "urgency urinary incontinence" ) collected ONLY on the AE log with start date after surgery and not reported on any of the complications tables will be assessed, and if any, included in this definition at the discretion of the principal investigator. |
| Indicator of new or worsening fecal incontinence (FI) | CRF-12, CRF-13, CRF-96 | Binary | surgery through 3 years | Through the timepoint of follow-up, based on the Adverse Event Log (CRF-96) or specific complications as reported on the masked post-operative CRF-12 and coordinator follow-up CRF-13 collected at 6-weeks and in 6-month intervals through 3 years, the indicator is defined as follows:<br><br>1 = Yes, if **new or worsening FI** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or "Fecal incontinence" in the term reported in the AE log. |

| | | | | |
|---|---|---|---|---|
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise<br><br>*Note: Prior to and upon database lock, any new or worsening fecal incontinence events identified by the MedDRA coded preferred term "fecal incontinence" ) with start date after surgery collected ONLY on the AE log and not reported on any of the complications tables will be assessed, and if any, included in this definition at the discretion of the principal investigator. |
| Indicator of sexually active denovo dyspareunia** | PISQ-IR | Binary | surgery through 3 years | At each timepoint among women who were sexually active (PISQ-IR Q1: 1=yes) at baseline and at the timepoint's corresponding attended post-operative visit, based on the response to the **PISQ-IR question Q11 "How often do you feel pain during sexual intercourse?"** at baseline and the timepoint's corresponding attended post-operative visit, the indicator is defined as follows:<br><br>1 = Yes, if a response of 1=never or 2=rarely is reported at baseline and a response of 3=sometimes, 4=usually, or 5=always is |

| | | | | |
|---|---|---|---|---|
| | | | | reported at the timepoint's corresponding attended post-operative visit<br><br>0 = No, else if there is a non-missing response reported at baseline and at the timepoint's corresponding attended post-operative visit<br><br>. = Missing, otherwise |
| Indicator of not sexually active denovo dyspareunia** | PISQ-IR | Binary | surgery through 3 years | At each timepoint among women who were not sexually active (PISQ-IR Q1: 2=no) at baseline and at the timepoint's corresponding attended post-operative visit, based on the response to the **PISQ-IR question Q2e asking if pain is a reason for not being sexually active** at baseline and the timepoint's corresponding attended post-operative visit, the indicator is defined as follows:<br><br>1 = Yes, if a response of 3=somewhat disagree or 4=strongly disagree is reported at baseline and a response of 1=strongly agree or 2=somewhat agree is reported at the timepoint's corresponding attended post-operative visit<br><br>0 = No, else if there is a non-missing response reported at baseline and at the timepoint's corresponding attended post-operative visit<br><br>. = Missing, otherwise |
| Indicator of dyspareunia | CRF-13, CRF-96 | Binary | surgery through 3 years | Through the timepoint of follow-up, based on the Adverse Event Log (CRF-96) or specific complications as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 3 years, the indicator is defined as follows: |

| | | | | |
|---|---|---|---|---|
| | | | | 1 = Yes, if **denovo dyspareunia, worsening dyspareunia, or dyspareunia preventing intercourse** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or "Dyspareunia" in the term reported in the AE log.<br><br>0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise<br><br>*Note: Prior to and upon database lock, any dyspareunia events identified by the MedDRA coded preferred term "dyspareunia" ) with start date after surgery collected ONLY on the AE log and not reported on any of the complications tables will be assessed, and if any, included in this definition at the discretion of the principal investigator. |
| Indicator of new or worsening voiding dysfunction | CRF-12, CRF-13, CRF-96 | Binary | surgery through 3 years | Through the timepoint of follow-up, based on the Adverse Event Log (CRF-96) or specific complications as reported on the masked post-operative CRF-12 and coordinator follow-up CRF-13 collected at 6-weeks and |

| | | | | in 6-month intervals through 3 years, the indicator is defined as follows: |
|---|---|---|---|---|
| | | | | 1 = Yes, if **new or worsening difficulty emptying of bladder** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or "Urinary Retention" in the term reported in the AE log.. Masked review identified free-text other specify complication fields containing **"incomplete bladder emptying", "urinary retention"** (case insensitive) on any of the complication tables as described above or on CRF-12 collected at 6-weeks and in 6-month intervals through 3 years and 1-year intervals thereafter. |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| | | | | *Note: Prior to and upon database lock, any voiding dysfunction events identified by the MedDRA coded preferred term "urinary retention" collected ONLY on the AE log with start date after surgery and not reported on any of the complications tables will be assessed, and if |

| | | | | any, included in this definition at the discretion of the principal investigator. |
|---|---|---|---|---|
| Regret with Decision Scale (DRS-PFD) | CRF-25 | Continuous | 12-, 24-, and 36-months | 5-item questionnaire with a 5-point response scale. Scoring consists of reversing the scores of the 2 negatively phrased items, then taking the mean of the 5 items (range, 1-5). |
| Satisfaction with Decision Scale (SDS-PFD) | CRF-25 | Continuous | 12-, 24-, and 36-months | 6-item questionnaire with a 5-point response scale. Scoring consists of taking the mean of the 6 items (range, 1-5). |

## 9.4 Primary Analysis Methods

The primary analysis is conducted on a modified intent-to-treat population for tests of superiority and the Per Protocol population for tests of non-inferiority. Because there were some treatment switchers during surgery, some switchers are removed from populations to maintain conservative and interpretable evaluations. See section 5. for definitions of populations, and details about which specific population is primary and supportive for the 4 primary hypotheses.

### 9.4.1 Primary Efficacy Outcome

The primary efficacy outcome is the time to surgical failure and is evaluated using all the data collected through the last operated subject's 3-year visit (primary endpoint). Since enrollment spanned 3 years, the duration of follow-up for some subjects included in the primary analysis is 5 years (60 months). Specifically, among subjects who experience a surgical failure, time to surgical failure is defined as the number of months (i.e. 6, 12, 18, 24, 30, 36, 42, 48, 54, or 60 months as intended by the 6-month interval visit schedule through a maximum of 5 years) from surgical intervention to first surgical failure at any post-operative visit 6 months through a maximum of 5 years after surgery.

Though surgical failure status is recorded at the 6-weeks post-operative visit, this data is not analyzed as these measures are collected too close to the date of surgical intervention. Instead, all subjects who attended the 6-week post-operative visit (i.e. 1.5 months) will be considered as not surgical failures at this visit. Subjects without a surgical failure at the end of follow-up are censored at their last attended visit.

A participant will be considered a surgical failure at the post-surgical visit if any ONE of the following criteria is met:

- Retreatment failure: retreatment for prolapse (surgery or pessary)
- Anatomic failure: prolapse beyond the hymen
- Symptomatic failure: bothersome vaginal bulge symptoms

Based on the definition of the primary outcome, the primary analytic approach will utilize a survival model to evaluate differences in time to surgical failure between the surgical arms.

Supporting numbers for survival analysis will be presented, including the number censored in each surgical arm in each period (from model-based approach described below) as well as the median, 25th, and 75th percentile for follow-up time in months for each surgical arm.

### 9.4.2 Primary Efficacy Analysis: Model-Based Survival Analysis

This model-based approach will serve as the primary analytic method for testing the comparative efficacy of the three strategies for uterovaginal prolapse as it can control for study design and potentially confounding variables. To appropriately account for the interval censored and right censored nature of the primary outcome data, and potential violation of the proportional hazards assumption across the entire follow-up time, while controlling for study design and potentially confounding variables in assessing intervention effects, a piecewise exponential (PWE) survival model will be employed. The PWE survival model is a model in which the time scale is divided into $K$ periods and the underlying hazard function, $h(t)$, is assumed to be constant within each period such that $h(t) = \lambda_k \exp(\beta \mathbf{X})$, where $\lambda_k$ is the hazard function for period $k$, and $\mathbf{X}$ is a 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 underlying hazard is constant within each period.

Among subjects who experience a surgical failure, the time to surgical failure (primary outcome) will be interval censored (i.e. lower bound, upper bound where both bounds are non-missing and lower bound < upper bound). Subjects who do not experience a surgical failure will be right censored (i.e. non-missing lower bound and missing upper bound) at their last attended visit. Specifically, the censoring interval for implementation in interval censored analysis will be defined as follows:

- lower bound = number of months (i.e. months as intended the by visit schedule) from the surgical intervention to the *last attended visit where the subject was considered to be not a surgical failure* across all post-operative visits starting at the 6 month visit. As such, subjects who were surgical failures at the 6-month post-operative visit will have a lower bound of 1.5 months corresponding to the 6-week post-operative visit. Based on this definition, the lower bound for the primary endpoint of 3 to a maximum of 5 years can take any of the following values 1.5, 6, 12, 18, 24, 30, 36, 42, 48, 54, or 60 months.

- upper bound = number of months (i.e. months as intended by the visit schedule) from the surgical intervention to *first post-operative visit with surgical failure* across all attended post-operative visits starting at the 6 month visit.  As such, subjects who do not experience a surgical failure through the endpoint, will have a missing upper bound and will be right censored at their lower bound. Based on this definition, the upper bound for the primary endpoint of 3 to a maximum of 5 years can take any of the following values 6, 12, 18, 24, 30, 36, 42, 48, 54, or 60 months.


**Piecewise Exponential (PWE) Survival Model**

A constant underlying hazard assumption in each of the periods will be assumed. For the primary outcome analysis through the primary endpoint of 3 to a maximum of 5 years, an interval censored proportional hazard model will be fitted with an assumed underlying piecewise exponential hazard with five constant-hazard periods: [0, 12], (12, 24], (24, 36], (36, 48] and (48, 60] and controlled for randomized surgical intervention, pooled site and age category (<65 and ≥65) consistent with study randomization design. If the model does not converge, the choice of period cut points will be evaluated with a plot of the NPMLE survival function (see Nonparametric Survival Analysis for Interval-Censored Data below) and periods will be combined as needed for model convergence while retaining clinical relevance. The PWE survival model will be fit using the SAS ICPHREG procedure.

From this single model, point estimates of the overall adjusted hazard ratios across all time periods along with confidence intervals and p-values corresponding to the 4 a priori specified hypothesis tests comparing risk of surgical failure across the surgical arms are obtained. The log-log transformation will be used to calculate confidence intervals. Specifically, the first test will compare the overall adjusted hazard ratio across all time periods of the transvaginal mesh arm to the mesh augmented sacrocolpopexy arm against the hypothesized non-inferiority margin of 1.93 evaluated at a one-sided significance level of α = 0.03. The p-value will test if the hazard ratio is not more than 1.93. A corresponding upper 1-sided 97% confidence interval will be provided. The second test is for superiority of transvaginal mesh to mesh augmented sacrocolpopexy arm against the hypothesized null ratio of 1.0 evaluated at a two-sided significance level of α = 0.03, with corresponding 97% 2-sided confidence interval. While this test will be provided, it is a step-down of the first test, and will be evaluated for statistical significance only if the first test rejects the inferiority null hypothesis. The other two hypothesis tests of superiority will compare the overall adjusted hazard ratio across all time periods of transvaginal mesh and mesh augmented sacrocolpopexy, respectively, to native tissue

repair against the hypothesized null ratio of 1.0 each evaluated at a two-sided significance level of $\alpha = 0.01$, and corresponding 2-sided 99% confidence interval. Note that pairwise comparisons will be completed within the model that contains all 3 treatment groups.

In addition, model estimates and 95% confidence intervals for the failure probability in each surgical group will be obtained at years 1, 2, 3, 4, and 5. Model-adjusted hazard rates (estimate and 95% confidence intervals) for each period will be obtained. The hazard rates will be reported as incidence density per person-year for each of the periods. For all estimates and contrasts, pooled sites and age category categories ($<65$ and $\geq 65$) will be given equal weight.

Descriptive statistics for the analysis will also include the number of participants by treatment group censored within each time interval, the median (IQR) follow-up time by treatment group, and the model hazard rate (incidence density per person year) by treatment group within each time interval.

The general structure of the SAS code that will be used for the piecewise model is summarized below.

```
PROC ICPHREG PLOTS(CL TIMERANGE=(0,60))=(survival) ALPHA=&alpha. NLOPTIONS(MAXIT=500);
CLASS treatment;
MODEL (LowerboundTIME, UpperboundTIME)= treatment Age_group poolsite1 poolsite2 poolsite3 poolsite4 poolsite5 poolsite6  /
          BASE=piecewiseexponential(INTERVALS=(12,24,36,48))/ HAZSCALE=HAZARD;
HAZARDRATIO "HR" treatment / DIFF=REFERENCE;
TEST treatment / HTYPE=3;
BASELINE OUT=work.pwe_surv_haz_estimates COVARIATES=work.pwe_covs TIMELIST=0 to 60 by 6 SURVIVAL=_ALL_
          HAZARD=_ALL_ / ROWID=treatment CLTYPE=LOGLOG;
```

### Test of Proportionality Assumption

This piecewise exponential model assumes proportional hazards for the treatment groups over time, in particular that the baseline hazard is constant within each period. The proportional hazards assumption will be assessed by fitting the same primary outcome model with a treatment by time interaction and tested via the likelihood ratio test at the 0.05 significance level. The treatment by time interaction term will be derived as the product of the reference cell coded surgical intervention variables and time where time is derived transforming censoring intervals to a single time measure as follows:

- time = lower bound of the censoring interval, if subject is right censored (i.e. lower bound is non-missing and upper bound is missing)

- time = midpoint of the censoring interval, else if subject is interval censored (i.e. lower bound < upper bound, both non-missing) or not censored (i.e. i.e. lower bound = upper bound, both non-missing)

If the proportionality hazards assumption is violated, additional graphic analyses will be generated to help understand how the proportionality assumption was violated. A figure will be generated of the fitted PWE model estimates and 95% confidence bands for failure probability overlaid with failure probability estimates obtained by nonparametric methods for interval censored data stratified by surgical intervention (see Nonparametric Survival Analysis for Interval-Censored Data below). Through the same nonparametric survival analysis method for interval censored data, a figure will be generated of the hazard functions for the three surgical arms using smoothed Epanechnikov kernel estimation. These two figures will be studied closely to identify approximate time interval(s) where the hazard ratios are distinctly different. If this departure can be characterized as quantitative rather than qualitative using the terminology used by Gail and Simon to describe interactions, then this finding indicates that the relative risk is of a different magnitude at different time points, but that

relative risk doesn't change direction at any time point. Therefore, describing the surgical effect as an average effect over the full time period is a reasonable and clinically meaningful approach. However, if the hazard functions are not only different, but substantially cross then this departure is characterized as qualitative and alternative parametric models that better fit the relatively complex hazard structures will be explored and compared to the pre-specified primary model. If this is the case, an addendum to the SAP will be written describing the new model-based efficacy analysis methods.

### 9.4.3   Sensitivity Efficacy Analyses of Primary Outcome

Outlined below are various sensitivity analyses that will be conducted on the primary outcome to evaluate the robustness of the primary findings. The details of each of the planned sensitivity analyses are described below.

**Nonparametric Survival Analysis for Interval-Censored Data**

This analysis will employ nonparametric maximum likelihood estimation (NPMLE) to estimate survival functions for interval censored data (censoring intervals defined as described above) stratified by randomized surgical intervention 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) and multiple imputation is used to account for the uncertainty of ranking overlapping intervals across participants caused by missed visits. 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). Pairwise comparisons of the surgical arms corresponding to the 3 a priori specified hypotheses testing for superiority as described above will be performed based on the generalized log-rank test statistic and the corresponding variance-covariance matrix (Huang, Lee, and Yu, 2008). The log-log transformation will be used to calculate confidence intervals. The nonparametric survival analysis methods detailed above will be carried out using the ICLIFETEST procedure in SAS. 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 primary model-based efficacy analysis.

In addition, this analysis will serve as a way to check the assumed PWE model assumptions, particularly the proportional hazard assumption and specification of the constant underlying hazard across periods. The Kaplan Meier survival curves will be graphically evaluated to evaluate the choice of periods specified in the PWE model, should the PWE model not converge. The Epanechnikov kernel-smoothed hazard rates from the non-parametric interval-censored Kaplan-Meier analysis will be generated and graphically assessed for violation of the proportional hazard assumption.

The general structure of the SAS code that will be used is summarized below.

```
PROC ICLIFETEST PLOTS=ALL METHOD=EMICM CONFTYPE=LOGLOG IMPUTE(SEED=rand_seed.) ALPHA=&alpha;
STRATA treatment;
TIME (LowerboundTIME, UpperboundTIME);
```

**Descriptive Cross-Sectional Analysis**

This descriptive cross-sectional analysis will serve as additional support of the primary efficacy analysis methods. Each surgical failure will be categorized into the following 4 types: "retreatment failure", "anatomic failure", "symptomatic failure", and "anatomic and symptomatic failure" where retreatment failure corresponds to retreatment for prolapse, symptomatic failure corresponds to bothersome vaginal bulge symptoms, and anatomic failure corresponds to prolapse beyond the hymen all as described above. Retreatment failure type is prioritized above the anatomic and/or symptomatic failure types. The category "anatomic and symptomatic failure" is defined when bothersome bulge and prolapse beyond the hymen were first reported at the same study visit.

This analysis will compare the proportions of surgical failures and obtain point estimates and 95% confidence intervals for pairwise risk differences between each of the mesh repairs versus the native tissue repair as well as between the two mesh repairs as well as the 2 df overall test at each year 1, 2, and 3 based on Mantel-Haenszel estimates for the common risk difference stratified by pooled site and age category (<65 and ≥65) (event rates permitting) with Wald-type confidence intervals with the estimate of the variance of the risk difference based on Sato (Sato 1984). Descriptive proportions of surgical failure types at each year will be provided. This analysis will evaluate surgical failure by visit based on surgical failure under the permanent failure state assumption and transient failure state assumption by initial failure type and at-visit failure type respectively.

Under the permanent failure state assumption, the number of the participants at each visit includes all participants who were still participating in the study or had a failure outcome prior to withdrawal. The denominator for the cumulative failure at each visit includes all participants who were still participating in the study and attended the visit or had any failure outcome prior to their withdrawal/missed visit. Under the permanent failure state assumption, the failure type corresponding to the first failure time will be provided, where the 4 failure types are mutually exclusive, with retreatment failure prioritized above the anatomic and/or symptomatic failure types.

Under the transient failure state assumption, the number of participants at each visit includes all participants who were still participating in the study. The denominator for the transient failures at each visit includes all participants who were still participating in the study and attended the visit or who missed the visit but had a retreatment failure prior to their missed visit. The failure type corresponding to the failure at each time point will be provided, where the 4 failure types are mutually exclusive, with retreatment failure prioritized above the anatomic and/or symptomatic failure types and assumed to be a permanent stated at all subsequent visits irrespective of the participant's attendance at the visit as long as the person was still participating in the study. Anatomic and symptomatic failure components are considered to be transient conditions, and with outcomes at each time point based specifically in the measurements obtained at that time point.

A figure will be created to depict the timeline of relapse for each participant, with participants each indicated by a line corresponding to their duration in the study. The line will show the time of censoring for non-relapsed participants, the time of failure for participants that remain a failure, and the multiple times that a participant was a failure for participants with transient failure symptoms. Due to missed visits caused by COVID, it will also indicate visits that were missed.

**Sensitivity Analysis examining impact of FDA order**

A sensitivity of the primary analysis treatment group comparisons on the primary analysis population will be performed to assess any potential impact from the protocol amendment in response to the April 16, 209 FDA order removing transvaginal mesh from the market on primary efficacy. The analysis will be performed on the subset of participants treated before April 18, 2019. Patients excluded from the sensitivity analysis were told they would not be getting TVM prior to surgery, yet

each of these patients did consent to the 3-arm trial and will be treated under the revised protocol so they are not protocol violations. Results of the sensitivity analysis will be compared to those obtained from the primary efficacy analyses. Any important differences will be reported.

### Sensitivity Analysis on Supportive Populations

A sensitivity analysis of the primary outcome through the primary endpoint of 3 to a maximum of 5 years on the supportive population (as defined in section 5) for each of the 4 primary hypothesis tests will be conducted via the previously described PWE survival model using the same model parameterizations and covariate adjustments as the primary efficacy analysis. The supporting nonparametric survival model (described above) will also be fitted.

### Sensitivity Analysis without adjustment for site, or age category

A sensitivity analysis of the primary outcome through 3 to a maximum of 5 years on the primary analysis population will be conducted via the previously described PWE survival model, but only controlling for the randomized surgical intervention, with no adjustment for pooled site, or age category (all other model specifications will remain the same as the primary efficacy analysis).

### Sensitivity Analysis evaluating impact of missed visits due to COVID pandemic

A sensitivity analysis of the primary outcome through the primary endpoint of 3 to a maximum of 5 years on the primary analysis population will be conducted via the previously described PWE survival model, with an additional covariate controlling for the year of the participant enrollment, and also testing for the interaction between enrollment cohort and overall treatment effect. Since enrollment was over 3 years, and the last patients were enrolled in 2018, the year 1 cohort had impact of 2020 and 2021 pandemic at the end of their follow-up, and the last participants enrolled had this impact in earlier years of their enrollment.

### Sensitivity Analysis treating site as a random effect

A sensitivity analysis of the primary outcome through the primary endpoint of 3 to a maximum of 5 years on the primary analysis population will be conducted to assess impact of treating site as a random rather than fixed effect. Due to limitations of the methodology and SAS procedures for interval censored data, there is no direct option to model site as a random effect. Therefore, the sensitivity assessment will be completed as follows: as done for the primary SUPeR 3-year *JAMA* publication: First, we will fit the previously described PWE survival model, but instead of interval censoring we will right censor the failures at the midpoint of the interval (all other model specifications will remain the same as the primary efficacy analysis). We will confirm that this model produces essentially the same hazard ratio point and interval estimates and p-value as the primary interval censored model. Next, we will fit a second version of this same model as a frailty model moving site from a fixed effect to a random effect. We expect that the point estimates from the random site frailty model will be similar to the primary model, and the confidence interval to be slightly wider, resulting in a slightly larger p-value.

## 9.5 Secondary Efficacy Analysis Methods

### 9.5.1 Secondary Efficacy Outcomes

All secondary anatomical and functional efficacy outcomes are collected longitudinally across the study. All analyses for secondary outcomes 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.

For all secondary outcomes, point estimates and 95% confidence intervals for pairwise comparisons between each of the mesh repairs versus the native tissue repair as well as between the two mesh repairs will be obtained along with a p-value for the overall test assessing if any of the treatment effects are different. Treatment groups will be compared in the MITT population for all secondary efficacy outcomes.

Unless otherwise specified, all other secondary efficacy outcomes collected across multiple visits, analyses will include data only collected through 3 years (visits occurring after 3 years are excluded) using appropriate adjusted repeated measures models that account for subject correlations across visits. If the data corresponding to a secondary efficacy outcome is also collected at baseline prior to surgical intervention, these outcomes will be analyzed as change from baseline. Models will be adjusted for randomized surgical intervention, time in months as intended by the visit schedule, intervention by time interaction, baseline value, pooled site and age category consistent with study randomization design.

Treatment groups will be compared for any secondary efficacy outcomes evaluated in an aggregate fashion (e.g. proportion of participants with an event occurrence through 3 years), using a chi-square test or Fishers exact test for rare events. Additional details are provided in the specific sections below.

### 9.5.2 General Approach

As outlined in the above sections, the secondary outcome variables comprise a combination of binary, ordinal, and continuous measures that are repeatedly collected at post-operative visits. These outcome measures will be summarized in tabular fashion by treatment arm over the first 3 years of follow-up period. We will use model-based approaches appropriate for the variable type to compare the outcomes for the three treatment arms across time: linear mixed models for repeated measures (MMRM) for continuous variables and logistic generalized mixed models for repeated measures for binary efficacy outcomes. The few ordinal outcomes will either be analyzed in the same manner as continuous scores or otherwise dichotomized.

Because the primary scientific interest for the study is in the trajectories of the outcomes across time, a longitudinal approach incorporating **all post-operative time points through 3 years** will be employed for all repeated continuous assessment. Specifically, the unit of analysis will be the participant visit, where visits are defined as in Section 4.5 starting 6 months post operatively unless otherwise specified, and the models will incorporate treatment group, time, treatment group by time interaction, as well as pooled site and age category, the randomization stratification variables, and baseline value as fixed effects. Time (or visit) will be treated as a categorical variable to incorporate the expected non-linearity of time effects. The correlation structure between visits for a participant will be fit as an autoregressive AR1.

The general structure of the SAS code that will be used for the linear mixed model (using the POPDI-6 as an example) is summarized below.

Proc Mixed data=ASPIRe;
WHERE visit le 36; *** Constrains analysis to first 36 months of follow-up;

```
Class treatment visit pooledcenter agegroup participant;
Model POPDI-6_change = POPDI-6_baseline treatment visit treatment*visit pooledcenter agegroup/ DDFM=KR;
Repeated visit / type =AR1 subject=participant R RCORR;
Lsmeans treatment*visit/ pdiff CL;*** adjusted mean and adjusted mean difference estimates **;
Contrast 'POPDI change at Month 36: TVM vs SC' treatment 0 -1 1 treatment*month 0 0 0 0 0 0
                                                                    0 0 0 0 0 -1
                                                                    0 0 0 0 0 1; ;
Contrast 'POPDI change at Month 36: TVM vs NTR' treatment -1 0 1 treatment*month 0 0 0 0 0 -1
                                                                    0 0 0 0 0 0
                                                                    0 0 0 0 0 1 ;
Contrast 'POPDI change at Month 36: SC vs NTR' treatment -1 1 0 treatment*month 0 0 0 0 0 -1
                                                                    0 0 0 0 0 1
                                                                    0 0 0 0 0 0 ;
CONTRAST '2 df trt grp comparison at Month 36:  treatment -1 1 0 treatment*month 0 0 0 0 0 -1
                                                                    0 0 0 0 0 1
                                                                    0 0 0 0 0 0 ,
                                    treatment -1 0 1 treatment*month 0 0 0 0 0 -1
                                                                    0 0 0 0 0 0
                                                                    0 0 0 0 0 1 ;
```

As suggested by the SAS code, the treatment groups will be tested using the F-test from the Contrast statements, which test whether mean change in POPDI at 36 months differs for each of the pairwise comparisons of the surgical arms of interest, as well as the overall 2 degree of freedom test comparing the 3 treatments at 36 months. The LSMEANS statement will provide by-visit estimates of the post-operative POPDI6 change 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 pairwise comparisons of the surgical arms of interest and will provide descriptive measures of differences between the pairwise comparisons of the surgical arms of interest at each study visit.

If the repeated outcome is binary rather than continuous (such as Apical Descent C > -2/3 x TVL at each visit), a generalized linear mixed model will be used in the analysis. The SAS procedure GLIMMIX will be used with a binomial distribution and an identity link to generate comparable comparisons between the treatment arms using a risk difference rather than a mean difference.

```
Proc GLIMMIX data=ASPIRe;
WHERE visit le 36; *** Constrains analysis to first 36 months of follow-up;
NLOPTIONS TECHNIQUE=nrridg;
Class treatment visit pooledcenter agegroup participant;
Model dependent_variable = treatment visit treatment*visit pooledcenter agegroup/ DIST=BIN, Link=Identity DDFM=KR;
Random visit / type =AR1 subject=participant RSIDE;
Lsmeans treatment*visit/ DIFF CL; *** adjusted event rate estimates and adjusted risk difference estimates **
*** code assumes treatment group takes values NTR, SC, TVM ***;
ESTIMATE 'risk difference at Month 36: TVM vs SC' treatment 0 -1 1 treatment*month 0 0 0 0 0 0
                                                                    0 0 0 0 0 -1
                                                                    0 0 0 0 0 1 / CL;
ESTIMATE 'risk difference at Month 36: TVM vs NTR' treatment -1 0 1 treatment*month 0 0 0 0 0 -1
                                                                    0 0 0 0 0 0
                                                                    0 0 0 0 0 1 / CL;
ESTIMATE 'risk difference at Month 36: SC vs NTR' treatment -1 1 0 treatment*month 0 0 0 0 0 -1
                                                                    0 0 0 0 0 1
                                                                    0 0 0 0 0 0 / CL;
CONTRAST '2 df trt grp comparison at Month 36:  treatment -1 1 0 treatment*month 0 0 0 0 0 -1
                                                                    0 0 0 0 0 1
                                                                    0 0 0 0 0 0 ,
                                    treatment -1 0 1 treatment*month 0 0 0 0 0 -1
                                                                    0 0 0 0 0 0
                                                                    0 0 0 0 0 1 ;
```

If this does not converge, we will change to a simpler compound symmetric covariance structure. If the model does not converge, then we will simplify further by removing pooled site and then age category from the model. If this model does not converge, we will conduct an unadjusted chi-square analysis for risk differences using the 3-year timepoint data, as described for binary outcomes below

Treatment groups will be compared for any secondary efficacy outcomes evaluated in an aggregate fashion as opposed to by visit (e.g. proportion of participants with an event occurrence through the final visit), using an unadjusted chi-squared tests or Fisher's exact tests for rare events with expected cell counts < 5. Risk differences with 95% CIs are based on Mantel-Haenszel estimates for the risk difference with Wald-type CIs. For events with expected cell counts < 5, the exact risk difference and 95% CI are obtained by exact methods based on the score statistic based on Chan and Zhang (ref Chan, Zhang 1995). Where specified, Binomial 95% CIs for proportions are based on the exact Clopper-Pearson method (ref Clopper, Pearson).

In SAS, using PROC FREQ:

The site and age group adjusted mantel-Haenszel chi-square test and risk difference with CIs using Wald-type variance under the null hypothesis as specified by Sato are obtained by:

    table poolediste*agegrp*trtgrp*outcomevar / riskdiff (COMMON) (METHOD=WALD EQUAL VAR=NULL ) CMH;

The unadjusted chi-square test and risk difference with CIs using Wald-type variance under the null hypothesis as specified by Sato are obtained by:

    table trtgrp*outcomevar / riskdiff (METHOD=WALD EQUAL VAR=NULL) CHISQ CMH;

The unadjusted exact test and score risk difference CIs are obtained by:

    table trtgrp*outcomevar / FISHER;    exact riskdiff (METHOD=SCORE);

# 10 SAFETY ANALYSES

## 10.1 Overview of Safety Analysis Methods

All safety and tolerability analyses will be performed on the safety population, consisting of all randomized and treated patients regardless of inclusion/exclusion violations, and summarized by the treatment actually received.

## 10.2 Safety and Tolerability Variables

| Variable | CRF Source | Type | Timepoint(s) | Definition |
|---|---|---|---|---|
| **Secondary Safety and Tolerability Outcomes** | | | | |
| Indicator of prolapse mesh exposure in the anterior vagina | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up, the indicator is defined as follows:<br><br>1 = Yes, if **mesh exposure in vagina** is reported as evident (i.e. response of "AE", "SAE") and Anterior=Yes and Description of Management of Mesh Exposure indicates the mesh was "from/for Prolapse Surgery" (case insensitive)<br><br>on any of the complication tables as described above<br><br>0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Indicator of prolapse mesh exposure in the posterior vagina | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up, the indicator is defined as follows:<br><br>1 = Yes, if **mesh exposure in vagina** is reported as evident (i.e. response of "AE", "SAE") and Posterior=Yes and Description of Management of Mesh Exposure indicates the mesh was "from/for Prolapse Surgery" (case insensitive)<br><br>on any of the complication tables as described above<br><br>0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are |

| | | | | |
|---|---|---|---|---|
| | | | | included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| Indicator of prolapse mesh exposure in the distal (lower half/introitus) vagina | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up, the indicator is defined as follows: |
| | | | | 1 = Yes, if **mesh exposure in vagina** is reported as evident (i.e. response of "AE", "SAE") and distal (lower half/introitus)=Yes and Description of Management of Mesh Exposure indicates the mesh was "from/for Prolapse Surgery" (case insensitive) |
| | | | | on any of the complication tables as described above |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| Indicator of prolapse mesh exposure in the apex of vagina | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up, the indicator is defined as follows: |
| | | | | 1 = Yes, if **mesh exposure in vagina** is reported as evident (i.e. response of "AE", "SAE") and apex=Yes and Description of Management of Mesh Exposure indicates the mesh was "from/for Prolapse Surgery" (case insensitive) |
| | | | | on any of the complication tables as described above |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are |

| | | | | |
|---|---|---|---|---|
| | | | | included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). <br><br> . = Missing, otherwise |
| Indicator of prolapse mesh erosion (exposure into adjacent organs) | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up among subjects who received the hysteropexy surgical intervention as reported on the surgeon's report CRF-8 (i.e. response to question A8 is "Uphold Lite"), based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention and based on the specific complications as reported on the evaluator (blinded) follow-up CRF-10/12 collected at 6-weeks and in 6-month intervals through 5 years, the indicator is defined as follows: <br><br> 1 = Yes, if **mesh erosion into viscera (urethra, bladder, ureter, rectum, or bowel locations)** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above <br><br> 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). <br><br> . = Missing, otherwise |
| Indicator of midurethral sling (MUS)mesh exposure in the anterior vagina | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up, the indicator is defined as follows: <br><br> 1 = Yes, if **mesh exposure in vagina** is reported as evident (i.e. response of "AE", "SAE") and Anterior=Yes and Description of Management of Mesh Exposure indicates the mesh was "from/for sling" (case insensitive) |

| | | | | |
|---|---|---|---|---|
| | | | | on any of the complication tables as described above<br><br>0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Indicator of midurethral sling (MUS)mesh exposure in the posterior vagina | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up, the indicator is defined as follows:<br><br>1 = Yes, if **mesh exposure in vagina** is reported as evident (i.e. response of "AE", "SAE") and Posterior=Yes and Description of Management of Mesh Exposure indicates the mesh was "from/for sling" (case insensitive)<br><br>on any of the complication tables as described above<br><br>0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Indicator of midurethral sling (MUS)mesh exposure in the distal (lower half/introitus) vagina | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up, the indicator is defined as follows:<br><br>1 = Yes, if **mesh exposure in vagina** is reported as evident (i.e. response of "AE", "SAE") and distal (lower half/introitus)=Yes and Description of Management of Mesh Exposure indicates the mesh was "from/for sling" (case insensitive) |

| | | | | on any of the complication tables as described above |
|---|---|---|---|---|
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| Indicator of midurethral sling (MUS)mesh exposure in the apex of vagina | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up, the indicator is defined as follows: |
| | | | | 1 = Yes, if **mesh exposure in vagina** is reported as evident (i.e. response of "AE", "SAE") and apex=Yes and Description of Management of Mesh Exposure indicates the mesh was "from/for sling" (case insensitive) |
| | | | | on any of the complication tables as described above |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| Indicator of midurethral sling (MUS) mesh erosion (exposure into adjacent organs) | CRF-8, CRF-10/12, AE log | Binary | Post-surgery through last visit | Through the timepoint of follow-up among subjects who received transobturator or retropubic midurethral sling as reported on the surgeon's report CRF-8 (i.e. questions E1 or E2 is "Yes" for performed), based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention and based on the specific complications as reported on the evaluator |

| | | | | (blinded) follow-up CRF-10/12 collected at 6-weeks and in 6-month intervals through 5 years, the indicator is defined as follows: |
|---|---|---|---|---|
| | | | | 1 = Yes, if **mesh erosion into viscera (other specified location containing "sling")** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **AE verbatim term containing "mus mesh erosion"** (case insensitive) |
| | | | | If subject did NOT receive hysteropexy surgical intervention as reported on the surgeon's report CRF-8 (i.e. response to question A8 is "Uphold Lite"), then **mesh erosion into viscera (urethra, bladder, ureter, rectum, bowel, or other specified locations)** reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above are also included as there are no other sources of mesh. |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 weeks or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| | | | | *Note: See above masked review note specified for outcome Indicator of hysteropexy mesh exposure. |
| Change from baseline in surgical pain scale SPS rest subscale score** | SPS | Continuous | 6-weeks thru last visit | The SPS rest subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2012). Specifically, score the response for question Q1 (0=no pain sensation, …, 10=most intense pain imaginable). If the |

| | | | | |
|---|---|---|---|---|
| | | | | question is missing a response, then the score is not calculated. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in surgical pain scale SPS normal activities subscale score** | SPS | Continuous | 6-weeks thru last visit | The SPS normal activities subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2012). Specifically, score the response for question Q2 (0=no pain sensation, …, 10=most intense pain imaginable). If the question is missing a response, then the score is not calculated. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in surgical pain scale SPS exercise subscale score** | SPS | Continuous | 6-weeks thru last visit | The SPS exercise subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2012). Specifically, score the response for question Q3 (0=no pain sensation, …, 10=most intense pain imaginable). If the question is missing a response, then the score is not calculated. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Change from baseline in surgical pain scale SPS worst pain subscale score** | SPS | Continuous | 6-weeks thru last visit | The SPS worst pain subscale score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (Barber 2012). Specifically, score the response for question Q4 |

| | | | | (0=not bad at all, …, 10=most intense bad feeling imaginable). If the question is missing a response, then the score is not calculated. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
|---|---|---|---|---|
| Change from baseline in body part pain score BPPS score** | SPS | Continuous | 6-weeks thru last visit | The BPPS score will be computed at baseline and at each timepoint's corresponding attended post-operative visit during which assessments are administered and collected using standard scoring algorithms (TOMUS study). Specifically, average the responses for questions Q1 through Q7 (0=no or yes: no pain, …, 10=yes: most intense). If there are missing responses, then the score is the average of the non-missing responses. The outcome will then be computed as the difference in score at each timepoint's corresponding attended post-operative visit and the score at baseline. If data for the assessment are missing at a timepoint's corresponding attended post-operative visit, the outcome variable will be coded as missing. |
| Maximum modified Dindo score across all complications

And indicator for the Maximum modified Dindo score ≥ **grade III** | CRF-8, CRF-9, CRF-10/12, CRF-11, CRF-13 | Ordinal Categorical (11 levels) | 6-weeks thru last visit | Through the timepoint of follow-up, based on the single Dindo score for the most severe complication as reported on the surgeon's report (unblinded) CRF-8, hospitalization CRF-9, evaluator (blinded) follow-up CRF-10/12 collected during surgery, during hospitalization, and at 6-weeks and in 6-month intervals through 5 years respectively, and as reported on the coordinator post-operative CRF-11, and coordinator follow-up CRF-13 collected at 6-weeks, and in 6-month intervals through 5 years respectively, the **maximum of the Dindo scores** (0.0=no complications, 1.0=grade I, 2.1=grade IIa, 2.2=grade IIabx, 2.3=grade IIb, 3.1=grade IIIo, 3.2=grade IIIa, 3.3=grade IIIb, 4.1=grade IVa, 4.2=grade IVb, 5.0=grade V) is taken **excluding Dindo scores reported for granulation tissue, suture exposure, and/or suture erosion complications reported ≤ 12 weeks from surgical intervention (i.e. surgery and 6-week visits)** on any of the complication tables as described above. |

| | | | | If a subject reported no complications, then their maximum Dindo score will be assigned a value of 0.0=no complication. A missing maximum Dindo score is assigned to subjects who discontinued prior to their surgical intervention visit. |
|---|---|---|---|---|
| | | | | Then, the indicator for **maximum Dindo score ≥ grade III** is defined as follows: |
| | | | | 1 = Yes, if the **maximum of the Dindo scores** as described above is **≥ grade III** (i.e. score ≥ 3.1). |
| | | | | 0 = No, else if the **maximum of the Dindo scores** as described above is **< grade III** (i.e. score < 3.1). |
| | | | | . = Missing, otherwise |
| | | | | *Note: Prior to final database lock, a listing of all reported Dindo scores grade III or higher will be generated along with all complications reported on the corresponding form(s) and visit(s) and manually reviewed internally to ensure that the open-text specified complication corresponding to the reported Dindo score is reported as evident (i.e. response of "AE", "SAE") on the complication table(s). In addition, a masked data review by the AE Adjudication Committee will be performed to ensure that all reported Dindo scores of grade III and above are reasonable for the specified complication. |
| Indicator of excessive granulation tissue after 12 weeks | CRF-10, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the evaluator (blinded) follow-up CRF-10 collected in 6-month intervals through 5 years, the indicator is defined as follows: |
| | | | | 1 = Yes, if **granulation tissue** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the |

| | | | | |
|---|---|---|---|---|
| | | | | **MedDRA coded preferred term "excessive granulation tissue"** (case insensitive) |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| Indicator of suture exposure after 12 weeks | CRF-10, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the evaluator (blinded) follow-up CRF-10 collected in 6-month intervals through 5 years, the indicator is defined as follows: |
| | | | | 1 = Yes, if **suture exposure (anterior, posterior, introitus, or apex locations)** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **AE verbatim term containing "suture exposure"** (case insensitive) |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |

| Indicator of suture erosion after 12 weeks | CRF-13, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years, the indicator is defined as follows:<br><br>1 = Yes, if **suture erosion (urethra, bladder, ureter, rectum, bowel, or other specified location)** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **AE verbatim term containing "suture erosion"** (case insensitive)<br><br>0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
|---|---|---|---|---|
| Indicator of pelvic pain after 12 weeks | CRF-13, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years respectively, the indicator is defined as follows:<br><br>1 = Yes, if **pelvic pain or daily pelvic pain** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **MedDRA coded preferred term "pelvic pain"** (case insensitive) |

| | | | | |
|---|---|---|---|---|
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Indicator of daily pelvic pain after 12 weeks | CRF-13, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years respectively, the indicator is defined as follows:<br><br>1 = Yes, if **daily pelvic pain** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **MedDRA coded term "pelvic pain"** (case insensitive) and the **AE verbatim term containing "daily pelvic pain"** (case insensitive)<br><br>0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Indicator of pelvic infection/abscess | CRF-9, CRF-11, | Binary | surgery through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention and based on the specific |

| | CRF-13, AE log | | | complications as reported on the hospitalization CRF-9, coordinator post-operative CRF-11, and coordinator follow-up CRF-13 collected during hospitalization, at 6-weeks, and in 6-month intervals through 5 years respectively, the indicator is defined as follows: |
|---|---|---|---|---|
| | | | | 1 = Yes, if **pelvic infection/abscess** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **AE verbatim terms containing "pelvic infection", "pelvic abscess"** (case insensitive) |
| | | | | 0 = No, else if subject attended at least one post-operative visit immediately following surgical intervention or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| Indicator of urinary tract infections (UTI) | CRF-9, CRF-11, CRF-13, AE log | Binary | surgery through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention and based on the specific complications as reported on the hospitalization CRF-9, coordinator post-operative CRF-11, and coordinator follow-up CRF-13 collected during hospitalization, at 6-weeks, and in 6-month intervals through 5 years respectively, the indicator is defined as follows: |
| | | | | 1 = Yes, if **lower urinary tract infection** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **MedDRA coded preferred term "urinary tract infection"** (case insensitive) |
| | | | | 0 = No, else if subject attended at least one post-operative visit immediately following surgical intervention or after and |

| | | | | |
|---|---|---|---|---|
| | | | | complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Indicator of vaginal infection or atypical vaginal discharge after 12 weeks | CRF-13, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years respectively, the indicator is defined as follows:<br><br>1 = Yes, if **vaginal infection or atypical/vaginal discharge** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **MedDRA coded preferred terms "vaginal infection", "vulvovaginal mycotic infection", "bacterial vaginosis"** (case insensitive) **and other adverse events as determined during masked review**. Masked review identified no further terms.<br><br>0 = No, else if subject attended at least one post-operative visit immediately following surgical intervention or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise<br><br>*Note: During masked review prior to final database lock, a listing of all adverse events on the AE log will be generated with AE verbatim terms mapped to a MedDRA coded preferred term and system organ class and |

| | | | | manually reviewed by a clinical investigator to identify all AEs that may be a vaginal infection or atypical vaginal discharge. These AEs will be identified either by their MedDRA coded preferred term(s) or the verbatim term(s). If needed, a detailed listing of the manually identified AEs will be generated with details from the AE log including onset date, end date, grade, attribution, serious indicator, outcome, and comments and manually reviewed by a clinical investigator to identify all AEs that are a vaginal infection or an atypical vaginal discharge. |
|---|---|---|---|---|
| Indicator of vaginal bleeding after 12 weeks | CRF-13, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the coordinator follow-up CRF-13 collected in 6-month intervals through 5 years respectively, the indicator is defined as follows: <br><br> 1 = Yes, if **denovo vaginal bleeding** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **MedDRA coded preferred term "vaginal haemorrhage"** (case insensitive) <br><br> 0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). <br><br> . = Missing, otherwise |
| Indicator of fistula formation | CRF-9, CRF-11, CRF-13, AE log | Binary | surgery through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention and based on the specific complications as reported on the hospitalization CRF-9, coordinator post-operative CRF-11, and coordinator follow-up CRF-13 collected during |

| | | | | hospitalization, at 6-weeks, and in 6-month intervals through 5 years respectively, the indicator is defined as follows: |
|---|---|---|---|---|
| | | | | 1 = Yes, if **vesicovaginal fistula, rectovaginal fistula, ureterovaginal fistula, urethrovaginal fistula, enterovaginal/colovaginal fistula, or enterovesical fistula** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **AE verbatim terms containing "vesicovaginal fistula", "rectovaginal fistula", "ureterovaginal fistula", "urethrovaginal fistula", "enterovaginal/colovaginal fistula", "enterovesical fistula"** (case insensitive) |
| | | | | 0 = No, else if subject attended at least one post-operative visit immediately following surgical intervention or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| Indicator of neuromuscular disorder | CRF-8, CRF-9, CRF-11, CRF-13, AE log, (CRF-10) | Binary | surgery through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention and based on the specific complications as reported on the surgeon's report (unblinded) CRF-8, hospitalization CRF-9, coordinator post-operative CRF-11, and coordinator follow-up CRF-13 collected during surgery, during hospitalization, at 6-weeks, and in 6-month intervals through 5 years respectively, the indicator is defined as follows: |
| | | | | 1 = Yes, if **neuromuscular disorder** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **MedDRA coded preferred terms "neuromyopathy",** |

| | | | | |
|---|---|---|---|---|
| | | | | **"sciatica", "spondylitic myelopathy"** (case insensitive) **and other adverse events as determined during masked review.** Masked review identified free-text other specify complication fields containing **"sciatic pain"** (case insensitive) on any of the complication tables as described above or on CRF-10 collected at 6-weeks and in 6-month intervals through 5 years.<br><br>0 = No, else if subject attended at least one post-operative visit immediately following surgical intervention or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise<br><br>*Note: During masked review prior to final database lock, a listing of all adverse events on the AE log will be generated with AE verbatim terms mapped to a MedDRA coded preferred term and system organ class and manually reviewed by a clinical investigator to identify all AEs that may be a neuromuscular disorder. These AEs will be identified either by their MedDRA coded preferred term(s) or the verbatim term(s). If needed, a detailed listing of the manually identified AEs will be generated with details from the AE log including onset date, end date, grade, attribution, serious indicator, outcome, and comments and manually reviewed by a clinical investigator to identify all AEs that are a neuromuscular disorder. |
| Indicator of vaginal scarring after 12 weeks | CRF-10, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the evaluator (blinded) follow-up CRF-10 collected at 6-weeks and in 6-month intervals through 5 years, the indicator is defined as follows: |

| | | | | |
|---|---|---|---|---|
| | | | | 1 = Yes, if **vaginal scarring** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **MedDRA coded preferred term "vulvovaginal injury"** (case insensitive) and the **AE verbatim term containing "scar"** (case insensitive) |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not Evident" response (assuming low missing response rate among attended visits). |
| | | | | . = Missing, otherwise |
| Indicator of vaginal shortening after 12 weeks | CRF-10, AE log | Binary | after 12 weeks through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended adverse events as reported on the AE log with an onset start date **12 weeks after the date of surgical intervention** and based on the specific complications as reported on the evaluator (blinded) follow-up CRF-10 collected at 6-weeks and in 6-month intervals through 5 years, the indicator is defined as follows: |
| | | | | 1 = Yes, if **vaginal shortening** is reported as evident (i.e. response of "AE", "SAE") on any of the complication tables as described above or on the AE log identified by the **AE verbatim terms containing "vaginal shortening"** (case insensitive) |
| | | | | 0 = No, else if subject attended at least one post-operative visit 6 months or after and complication(s) of interest specified above are reported as "Not Evident" across all attended post-operative visits. Subjects with missing responses are included as long as the subject has at least one "Not |

| | | | | |
|---|---|---|---|---|
| | | | | Evident" response (assuming low missing response rate among attended visits).<br><br>. = Missing, otherwise |
| Reportable adverse events by system organ class (MedDRA coding) | AE log | Nominal Categorical | surgery through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended **reportable** adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention, the adverse events will be coded following the MedDRA dictionary.<br><br>Reportable adverse events are defined in the MOP as follows:<br><br>- During the First Six Weeks Following Surgery: All AEs and SAEs of grade II or higher will be collected.<br><br>- Between Six Weeks and Six Months: All AEs and SAEs that are deaths, require a hospitalization or an emergency room visit, and all AEs or SAEs of grade II or higher that are at least possibly related to the pelvis or surgery, in the opinion of the investigator, will be collected.<br><br>- At All Remaining Visits: Only deaths and AEs or SAEs of grade II or higher that are at least possibly related to the pelvis or surgery, in the opinion of the investigator, will be collected.<br><br>*Note: Events reported on the AE log corresponding to the following functional efficacy outcomes will be excluded: new or worsening stress urinary incontinence (**MedDRA coded preferred term "stress urinary incontinence"**), new or worsening urgency urinary incontinence (**MedDRA coded preferred term "urge incontinence"**), new or worsening fecal incontinence (**MedDRA coded preferred term "faecal incontinence"**), dyspareunia (**MedDRA coded preferred term "dyspareunia"**), and new or worsening voiding dysfunction (**MedDRA coded preferred term "urinary retention"**) (case insensitive). |
| Reportable serious adverse events by system organ class (MedDRA coding) | AE log | Nominal Categorical | surgery through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended **reportable** adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention where the |

| | | | | adverse events will be coded following the MedDRA dictionary and classified as a serious adverse event as defined in the MOP as any **reportable** adverse event that:<br><br>- results in death<br><br>- is life-threatening (places the subject at immediate risk of death from the event as it occurred)<br><br>- results in inpatient hospitalization or prolongation of existing hospitalization<br><br>- results in a persistent or significant disability/incapacity<br><br>- is another medically important condition - based upon appropriate medical judgment, may jeopardize the subject's health and may require medical or surgical intervention to prevent one of the other outcomes listed in this definition (examples of such events include allergic bronchospasm requiring intensive treatment in the emergency room or at home, blood dyscrasias or convulsions that do not result in inpatient hospitalization, or the development of drug dependency or drug abuse)<br><br>- medical or surgical intervention was necessary to preclude permanent impairment of a body function, or prevent permanent damage to a body structure, either situation suspected to be due to the use of a medical product (Devices)<br><br>A life-threatening AE is defined as any AE that in the view of the investigator places the study subject at immediate risk of death. It does not include an AE that might have caused death had it occurred in a more severe form.<br><br>A disability is defined as a substantial disruption of a person's ability to conduct normal life functions. |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| | | | | *Note: Events reported on the AE log corresponding to the following functional efficacy outcomes will be excluded: new or worsening stress urinary incontinence (**MedDRA coded preferred term "stress urinary incontinence"**), new or worsening urgency urinary incontinence (**MedDRA coded preferred term "urge incontinence"**), new or worsening fecal incontinence (**MedDRA coded preferred term "faecal incontinence"**), dyspareunia (**MedDRA coded preferred term "dyspareunia"**), and new or worsening voiding dysfunction (**MedDRA coded preferred term "urinary retention"**) (case insensitive). |
| Indicator of death outcome | AE log | Binary | surgery through last visit | Through the timepoint of follow-up, based on the systematic collection of open-ended **reportable** adverse events as reported on the AE log with an onset start date on or after the date of surgical intervention where the adverse events will be coded following the MedDRA dictionary, the indicator is defined as follows: <br><br> 1 = Yes, if the outcome of death is reported on the AE log as described above <br><br> 0 = No, else if subject attended the surgical intervention visit <br><br> . = Missing, otherwise |

### 10.3      Secondary Safety/Tolerability Outcomes

In general, data for all secondary safety/tolerability outcomes are collected longitudinally across the study on complication tables at each visit and systematic collection of open-ended adverse events reported on the AE log. Secondary safety/tolerability outcomes are generally reported as a binary outcome evaluated in an aggregate fashion (e.g. through 5 years).

Treatment groups will be compared for binary safety outcomes using unadjusted chi-squared tests or Fisher's exact tests for rare events with expected cell counts < 5.. Risk differences with 95% CIs are based on Mantel-Haenszel estimates for the risk difference with Wald-type CIs. For events with expected cell counts < 5, the exact risk difference and 95% CI are obtained by exact methods based on the score statistic based on Chan and Zhang (ref Chan, Zhang 1995). Where specified, Binomial 95% CIs for proportions are based on the exact Clopper-Pearson method (ref Clopper, Pearson). These are the same methods as used for aggregate binary efficacy outcomes.

Maximum dindo score will be analyzed as an ordinal outcome with an unadjusted Mantel-Haenszel chi-squared mean score test with standardized midranks scoring (MODRIDIT option in SAS).

For any continuous or binary safety outcomes measured and analyzed as repeated measures, the same Mixed models as specified for efficacy assessments will be used. Continuous or ordinal safety outcomes assessed at a single timepoint, mean differences, 95% CIs, and $P$ values are unadjusted and based on $t$ tests for normally distributed continuous outcomes, and $P$ values for non-normal continuous outcomes (summarized with medians) are from Wilcoxon Rank Sum tests.

### 10.4      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.

MedDRA SOC and preferred term for SAEs, treatment-related SAEs and SAEs with an outcome of death will be summarized by treatment arm.

### 10.5      Other Safety Outcomes

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

## 11    ANALYSIS OF OTHER OUTCOMES

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

## 12    UPDATES TO ORIGINAL STATISTICAL ANALYSIS PLAN

With Version 2 of this SAP (February 20, 2023), the following clarification and addition were made to the original version.

Clarification to Section 9.4.2 Test of Proportionality Assumption:

Due to limitation of the SAS procedure and the use of the ID statement, the likelihood ratio test was conducted comparing nested semi-parametric interval censored models analogous to the primary efficacy analysis methods (but further relaxing the piecewise constant hazard assumption to be unspecified--no parametric assumptions made on the baseline hazard only that it is proportional).

Addition to Section 9.4.3 Sensitivity Efficacy Analyses of Primary Outcome

**Post-hoc Sensitivity Analysis excluding data after the 3 year visit**

A sensitivity analysis of the primary outcome through the primary endpoint excluding data after the 3 year visit on the primary analysis population will be conducted to evaluate the proportional hazards assumption during the first 3 years. Failure probability, hazard rate, and hazard ratio estimates and their corresponding confidence intervals and p-values are from an interval-censored proportional hazard model with an assumed underlying piecewise exponential hazard with three constant-hazard periods: [0, 12], (12, 24], (24, 36]. Excluding follow-up years 4 and 5 removed 14 events (2 events in NTR, 5 events in SC, and 7 events in TVM). The proportional hazards assumption is not violated in the MITT population (p=0.155).

## 13    REFERENCES

Chan IS, Zhang Z. Test-based exact confidence intervals for the difference of two binomial proportions. *Biometrics*. 1999;55(4):1202-1209. doi:10.1111/j.0006-341X.1999.01202.x

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial, *Biometrika*, 1934 26(4): 404–413, https://doi.org/10.1093/biomet/26.4.404

Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics*. 1985 Jun;41(2):361-72.

Huang J, Lee C, Yu Q. A generalized log-rank test for interval-censored failure time data via multiple imputation. *Statistics in Medicine*. 2008;27:3217–3226

Nager CW, Visco AG, Richter HE, et al. Effect of Vaginal Mesh Hysteropexy vs Vaginal Hysterectomy with Uterosacral Ligament Suspension on Treatment Failure in Women with Uterovaginal Prolapse: A Randomized Clinical Trial. *JAMA.* 2019;322(11):1054–1065. doi:10.1001/jama.2019.12812

Sato T. On the variance estimator for the Mantel-Haenszel risk difference. *Biometrics*. 1989; 45:1323-1324.

Sun J. A non-parametric test for interval-censored failure time data with application to AIDS studies. *Stat Med*. 1996;15(13):1387-1395. doi:10.1002/(SICI)1097-0258(19960715)15: 13<1387::AID-SIM268>3.0.CO;2-R

Sun, J. Nonparametric Test for Doubly Interval-Censored Failure Time Data. *Lifetime Data Anal* 7, 363–375 (2001). https://doi.org/10.1023/A:1012596731505

Wellner JA, Zhan Y. A Hybrid Algorithm for Computation of the Nonparametric Maximum Likelihood Estimator from Censored Data, *Journal of the American Statistical Association*,1997 92:439, 945-959, DOI: 10.1080/01621459.1997.10474049

## 14    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.  Specific display shells will be generated and reviewed by the protocol team prior to the initiation of analyses.

Potential list of displays for the primary manuscript are as follows: Table and figure shells are in a separate document.

| Tables – in manuscript |
| --- |
| Table 1 Baseline Characteristics in the Modified Intent-to-Treat Population |

Table X Time until Surgical Treatment Failure for MITT, PP, & MITT Narrow (3 separate tables – perhaps put some or all in online supplement, depending on results)

Table 2 Secondary Efficacy, Safety/Tolerability and Masking Outcomes in the Modified Intent-to-Treat Population

Table 3 Secondary Efficacy, Safety/Tolerability and Masking Outcomes

**Figures – in manuscript**

Figure 1 Consort Diagram

Figure 2 Failure Probability for the Primary Outcome Comparing Native Tissue, Sacral Colpopexy, and Transvaginal Mesh in the Modified Intent-to-Treat Population (4-panel figure showing (a) 3-line survival plot with no confidence intervals of testing, (B-D): each of the pairwise tests with confidence intervals and showing modeling results. Panel for TM vs SC will either be of non-inferiority, or if that is accepted, with be for superiority. Pairwise plots may need to use different applicable populations than the 3-way summary plot.

Figure 3 Distribution of the Composite Surgical Failure Outcome in the Modified Intent-to-Treat Population ("ASSORT diagram")

**Online Supplement**

eFigures 1-3. Diagram of each of the 3 surgical procedures

eFigure 4. Pelvic Organ Prolapse Quantification (POPQ) Schematic

eTable 1 Analysis Populations and Visit Completion – All Randomized Participants

eTable XX follow-up durations and COVID timing by calendar time of enrollment

eTable 1A Data Collection/Completeness of Primary Efficacy by Randomized Treatment Group

eTable 1B Primary Outcome (First Event Incorporated) Completeness by Randomized Treatment Group

eFigure Y. Failure Probability for the Primary Outcome in the Per-Protocol Population

eFigure Z. Failure Probability for the Primary Outcome in the MITT-Narrow

eFigure 9. Distribution of Surgical Failure Outcome among Participants with Intermittent Failures in the Modified Intent-to-Treat Population

eFigure 4. Assumption Check: Smoothed Hazard Rates for the Primary Outcome

eFigure 5. Assumption check: Estimated Log Cumulative Hazard by Log-Time for the Primary Outcome

eTable X. Sensitivity Analysis: Failure Probability for the Primary Outcome in primary populations with further reduction of patients enrolled after FDA order to stop mesh.

eTable X. Sensitivity Analysis: Unadjusted Failure Probability for the Primary Outcome

eTable X. Sensitivity Analysis: Failure Probability for the Primary Outcome adjusting for enrollment cohort

eTable X. Sensitivity Analysis: Hazard Ratio for the Primary Outcome with Random Site Effect

eTable 2 Concomitant procedures in the Modified Intent-to-Treat Population

eTable 3 Perioperative outcomes in the Modified Intent-to-Treat Population

eTable 4. Sexual Function and Dyspareunia Results from PISQ-IR a in the Modified Intent-to-Treat Population by Randomized Intervention

eTable 5. Adverse Events by System Organ Class and Preferred Term in Safety Population by Received Surgical Intervention

eTable 6. Serious Adverse Events by System Organ Class and Preferred Term in Safety Population by Received Surgical Intervention

# 15    ATTACHMENTS

## 15.1    Attachment 1: Scoring of the SF-12 PCS and MCS Subscales

The scoring of the SF-12 QOL assessment was found at https://drhays.dgsom.ucla.edu/files/view/docs/programs-utilities/sf12v1_short.sas.txt. This attachment provides a summary of the calculations for both the Physical Component Summary (PCS) and Mental Component Summary (MCS) subscales of the SF-12.

The SF-12 is comprised of 12 questions. Q1, Q8, and Q12 use five-number scales; Q4 through Q7 use yes/no scales; and Q 2 and Q3 have three-number scales; and Q9 through Q11 use six-number scales. For Q1 and Q8, values are assigned to each response as follows: 100=excellent, 75=very good, 50=good, 25=fair, 0=poor. For Q12, values are assigned to each response as follows: 100=none of the time, 75=a little of the time, 50=some of the time, 25=most of the time, 0=all of the time. Q2 and Q3 receive values as follows: 0=yes, limited a lot; 50=yes, limited a little; 100=no, not limited at all. For Q4-Q7, values are assigned as follows: 0=yes and 100=no. For Q9 and Q10, values are assigned to each response as follows: 100=all of the time, 80=most of the time, 60=a good bit of the time, 40=some of the time, 20=a little of the time, 0=none of the time. For Q11, values are assigned to each response as follows: 0=all of the time, 20=most of the time, 40=a good bit of the time, 60=some of the time, 80=a little of the time, 100=none of the time.

With values assigned to each question, the raw PCS and MCS scores are then calculated as a weighted sum based on the assigned value for each question. The weights are provided in Table A-1 and Table A-2. The PCS and MCS subscales are then calculated by multiplying these sums by 10 and adding a constant to each raw score (PCS=56.57706, MCS=60.75781).

**Table A-1. Sum Weights for Question Values – PCS**

| Question Number | Assigned Value | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 20-25 | 40 | 50 | 60 | 75-80 | 100 |

| Question Number | 0 | 20-25 | 40 | 50 | 60 | 75-80 | 100 |
|---|---|---|---|---|---|---|---|
| 1 | -8.37399 | -5.56461 | | -3.02396 | | -1.31872 | 0 |
| 2 | -7.23216 | | | -3.45555 | | | 0 |
| 3 | -6.24397 | | | -2.73557 | | | 0 |
| 4 | -4.61617 | | | | | | 0 |
| 5 | -5.51747 | | | | | | 0 |
| 6 | 3.04365 | | | | | | 0 |
| 7 | 2.32091 | | | | | | 0 |
| 8 | -11.25544 | -8.38063 | | -6.50522 | | -3.80130 | 0 |
| 9 | 3.46638 | 2.90426 | 2.37241 | | 1.36689 | 0.66514 | 0 |
| 10 | -2.44706 | -2.02168 | -1.61850 | | -1.14387 | -0.42251 | 0 |
| 11 | 4.61446 | 3.41593 | 2.34247 | | 1.28044 | 0.41188 | 0 |
| 12 | -0.33682 | -0.94342 | | -0.18043 | | 0.11038 | 0 |

**Table A-2. Sum Weights for Question Values – MCS**

| Question Number | Assigned Value | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 20-25 | 40 | 50 | 60 | 75-80 | 100 |
| 1 | -1.71175 | -0.16891 | | 0.03482 | | -0.06064 | 0 |
| 2 | 3.93115 | | | 1.86840 | | | 0 |
| 3 | 2.68282 | | | 1.43103 | | | 0 |
| 4 | 1.44060 | | | | | | 0 |
| 5 | 1.66968 | | | | | | 0 |
| 6 | -6.82672 | | | | | | 0 |
| 7 | -5.69921 | | | | | | 0 |
| 8 | 1.48619 | 1.76691 | | 1.49384 | | 0.90384 | 0 |
| 9 | -10.19085 | -7.92717 | -6.31121 | | -4.09842 | -1.94949 | 0 |
| 10 | -6.02409 | -4.88962 | -3.29805 | | -1.65178 | -0.92057 | 0 |
| 11 | -16.15395 | -10.77911 | -8.09914 | | -4.59055 | -1.95934 | 0 |
| 12 | -6.29724 | -8.26066 | | -5.63286 | | -3.13896 | 0 |