CONFIDENTIAL

Pelvic Floor Disorders Network

Protocol

Effects of Surgical Treatment Enhanced with Exercise for Mixed Urinary Incontinence (ESTEEM)

Short title: Combined treatment for mixed incontinence

Concept Proposal: Approved by Steering Committee 7/22/2011
Mini Protocol: Approved by Steering Committee 7/19/2012
Protocol: Approved by Steering Committee 10/11/2012

Revised: February 12, 2014

Protocol Chair:
Vivian W. Sung
Brown/Women and Infants Hospital
Division of Urogynecology
695 Eddy Street, lower level
Providence, RI 02906
vsung@wihri.org
(401) 274-1122 x 8740

Protocol Committee:
Brown/Women and Infants: Deborah L. Myers (Co-Chair)
Cleveland Clinic: Beri Ridgeway
Duke University Medical Center: Alison Weidner
University of Alabama at Birmingham: Holly Richter
University of California-San Diego: Emily Lukacz
University of New Mexico: Gena Dunivan
University of Pennsylvania: Ariana Smith, Diane Newman
University of Pittsburgh: Pamela Moalli, Diane Borello-France
RTI DCC: Marie Gantz
NIH/NICHD: Susan Meikle

ABBREVIATIONS

ABC	Anticholinergic versus Botox Comparison trial
ATLAS	Ambulatory Treatments for Leakage Associated with Stress Incontinence trial
BBUSQ	Birmingham Bowel Urinary Symptom Questionnaire
BD	Bladder diary
BE- DRI	Behavior Enhances Drug Reduction of Incontinence trial
BPTx	Behavioral/pelvic floor therapy
CDF	Cumulative distribution function
CST	Cough stress test
DCC	Data Coordinating Center
DO	Detrusor overactivity
DSMB	Data and Safety Monitoring Board
EQ-5D	European Quality of Life-5 Dimensions
HRQOL	Health related quality of life
IE	Incontinence episode
ICI	International Consultation on Incontinence
ICS	International Continence Society
IIQ	Incontinence Impact Questionnaire
IRB	Institutional Review Board
ITT	Intention-to-treat
IUGA	International Urogynecological Association
MESA	Medical, Epidemiologic, and Social Aspects of Aging
MID	Minimum important difference
MIMOSA	Mixed Incontinence: Medical or Surgical Approach trial
MSM	Medical Safety Monitor
MUI	Mixed urinary incontinence
MUS	Mid-urethral sling
OAB	Overactive bladder
OAB-q	Overactive Bladder Questionnaire
OAB-q-SS	Overactive Bladder Questionnaire-Symptom subscale
OAB-SAT-q	Overactive Bladder Questionnaire-Satisfaction with Treatment Questionnaire
OPTIMAL	Operations and Pelvic Muscle Training in the Management of Apical Support Loss trial
PFD	Pelvic floor disorder
PFDI	Pelvic Floor Disorder Inventory
PFDN	Pelvic Floor Disorders Network
PFMT	Pelvic floor muscle training
PGI-I	Patient Global Impression- Improvement
PGI-S	Patient Global Impression-Severity
PISQ	Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire
POPQ	Pelvic Organ Prolapse Quantification system
PVR	Postvoid residual
QoL	Quality of life
QUID	Questionnaire for Urinary Incontinence Diagnosis
RCT	Randomized controlled trial
RUBI	Refractory idiopathic urge incontinence and botulinum A injection trial
SAE	Serious adverse event
SD	Standard deviation
SISTEr	Stress Incontinence Surgical Treatment Efficacy Trial

SUI	Stress urinary incontinence
TOMUS	Trial of Mid-Urethral Slings
TOT	Transobturator tape sling
TVT	Tension-free vaginal tape sling
TVT-O	Tension-free vaginal tape obturator
UDE	Urodynamic evaluation
UDI	Urogenital Distress Inventory
UI	Urinary incontinence
UIE	Urinary incontinence episode
UITN	Urinary Incontinence Treatment Network
UUI	Urge urinary incontinence
ValUE	Value of Urodynamic Evaluation trial
VPFMC	Voluntary pelvic floor muscle contraction
3IQ	3 Incontinence Questions Assessment Tool

TABLE OF CONTENTS

1. Study Aims

Mixed urinary incontinence (MUI), defined as both stress urinary incontinence (SUI) and urge urinary incontinence (UUI), is a challenging condition and there are limited trials evaluating interventions that can optimize treatment outcomes. The overarching goal of this randomized trial is to estimate the effect of combined midurethral sling (MUS) and peri-operative behavioral/pelvic floor therapy (BPTx) compared to MUS alone on successful treatment of MUI symptoms in 472 women. Secondary objectives include estimating the effect of combined treatment compared to MUS on improving OAB and SUI outcomes separately, need for additional treatment, time to failure and identifying predictors of poor outcomes in this MUI population.

1.1. Primary Aim:

To assess whether combined MUS + peri-operative BPTx is superior to MUS alone for improving MUI symptoms at 1 year in women electing surgical treatment.

Primary Outcome: Change in severity of MUI symptoms at 1 year following MUS measured using the Urogenital Distress Inventory (UDI).[2]

Primary Null Hypothesis: There is no difference in the change in MUI symptoms between women receiving combined MUS+BPTx versus MUS alone at 1 year following MUS surgery.

Primary Alternative Hypothesis: Combined MUS+BPTx is superior to MUS alone for improving change in MUI symptoms at 1 year following MUS surgery.

1.2. Secondary Aims:

1. OAB symptom outcomes: To assess whether combined MUS+BPTx is superior to MUS alone for improving change in OAB symptoms at 1 year in women electing surgical treatment.
-OAB symptoms will be measured using UDI-irritative subscale scores

Secondary Alternative Hypothesis: Combined MUS+BPTx is superior to MUS alone for improving change in OAB symptoms in women with MUI at 1 year following MUS surgery.

2. SUI symptom outcomes: To assess whether combined MUS+BPTx is superior to MUS alone for improving change in SUI symptoms at 1 year in women electing surgical treatment for MUI.
-SUI symptoms will be measured using the UDI-stress subscale.

Secondary Alternative Hypothesis: Combined MUS+BPTx is superior to MUS alone for improving change in SUI symptoms in women with MUI at 1 year following MUS surgery.

1.3. Exploratory Aims:

1. Secondary urinary outcomes: To assess whether combined MUS+BPTx is superior to MUS alone for improving the number of urgency and urge incontinence episodes on bladder diary at 1 year following MUS surgery.

2. Time to failure: To compare time to failure between MUS+BPTx versus MUS alone.
-Failure will be defined as initiation of any additional treatment for lower urinary tract symptoms (SUI, UUI/OAB, or voiding dysfunction).

3. Predictors of poor outcomes: To develop models to identify predictors of change of MUI, OAB, and SUI outcomes measured using the UDI between baseline and 1 year post-treatment.

4. Quality of life and global impression: To compare quality of life outcomes and Patient Global Impression-Improvement (PGI-I)[3], Patient Global Impression-Severity (PGI-S)[3] between groups

5. Safety and additional treatments: To describe rates of reoperation (sling revision) for worsening OAB symptoms after MUS and to compare the proportion of women in each group initiating additional treatment for SUI and/or OAB, and the types of additional treatment (BPTx, medications, other)

6. Minimally important difference (MID) and clinical definitions: To determine MIDs and clinically meaningful definitions of MUI that predict clinical outcomes using cut-offs and combinations of standardized measures

7. Pelvic floor muscle strengthening: To compare pelvic floor muscle strength changes between women randomized to combined MUS+BPTx versus MUS alone and to estimate associations between pelvic floor muscle strength improvement and UI symptoms. We will also explore predictors of unsuccessful pelvic floor muscle strengthening.

8. Cost-effectiveness analysis: To determine the cost effectiveness of combined midurethral sling (MUS) and peri-operative behavioral/pelvic floor therapy (BPTx) compared to MUS alone for the treatment of MUI symptoms

2. BACKGROUND INFORMATION AND SCIENTIFIC RATIONALE

2.1. Disease/Condition Background:

Up to 50% of women with incontinence have mixed urinary incontinence (MUI); a complex condition that is significantly challenging for patients, clinicians and researchers.[4-6] For patients, the combination of UUI and SUI is more bothersome compared to either condition alone.[7-9] For clinicians, treatment of MUI is challenging due to higher failure rates, as interventions designed to benefit one symptom often do not benefit the other. For clinicians and researchers, the lack of a clinically useful definition of MUI[10] and the frequent exclusion of MUI patients from randomized trials[11] pose challenges for determining best treatment approaches. The wide variability of patient symptoms and terminology, ranging from “stress-predominant”, “urge-predominant”, “OAB -wet” or “OAB-dry”, further complicates data interpretation and patient management. The current definitions and treatment approaches have failed to provide significant progress in the treatment of this bothersome condition.

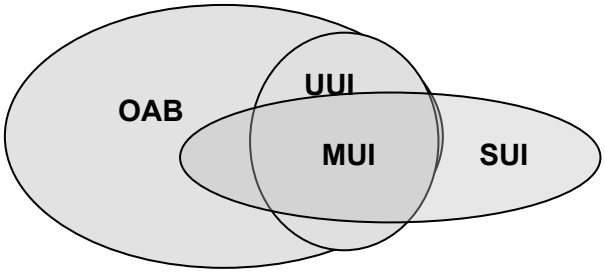


Figure 1. Conceptual model of MUI. Adapted from Katsumi et al[1]

2.2. Challenges with definitions

There is significant variability and controversy regarding the "best" definition of MUI: essentially there is an absence of a universal definition.[10-15] Based on the name alone, it makes sense that "MUI" includes symptoms of both SUI and UUI/OAB. The International Urogynecological Association (IUGA)/International Continence Society (ICS) joint terminology report defines OAB based on symptoms alone as "urgency with or without urgency incontinence, usually with frequency and nocturia": women can be "OAB-wet" or "OAB-dry". MUI is defined by the same group as "the complaint of involuntary leakage associated with urgency and also with exertion, effort, sneezing, or coughing".[16] Clinical challenges with this definition include: 1) it excludes women who may have significant urgency and/or frequency without UUI; 2) it excludes women who have detrusor overactivity in the absence of sensory urgency; and 3) many women do not experience SUI or UUI based on these clear cut definitions. Purely "objective" measures such as urodynamic evaluation (UDE) also do not provide a clear and consistent definition. Further complicating the issue is the lack of consensus and evidence regarding the pathophysiology of MUI. Some experts argue the two conditions should be considered as having completely different pathological processes,[12] whereas others argue that at least in a subset of women, they are directly linked (e.g. proximal urethral funneling causing detrusor overactivity).

Brubaker et al and the Urinary Incontinence Treatment Network (UITN) attempted to develop an empirically derived definition of MUI in 2009.[10] Using data from the Stress Incontinence Surgical Treatment Efficacy Trial (SISTEr trial), a randomized trial comparing fascial sling to Burch colposuspension,[17] the investigators used a series of regressions and attempted to define cut-off values for a variety of standardized measures that could predict clinical outcomes. Standardized measures included the Medical, Epidemiologic and Social Aspects of Aging (MESA),[18] the Urogenital Distress Inventory (UDI),[2] urodynamic studies and a 3-day urinary diary. The investigators created threshold definitions using the MESA (which measures the frequency of SUI and/or UUI), the UDI (which measures the presence and degree of bother for SUI and UUI), and UDE (defined as presence of urodynamic SUI and detrusor overactivity with or without associated leakage). These definitions were evaluated against the trial's clinical outcome, a composite outcome divided into SUI success (negative cough stress test, no SUI re-treatment, and negative MESA SUI) and overall success (stress criteria plus no leakage on diary or pad test). After testing 12 different definitions for MUI, the authors were unable to identify a definition that could accurately reflect clinical outcomes and proposed that both subcomponents of SUI and UUI should be individually described instead of using a one-dimensional descriptor. One limitation is that the SISTEr trial included only women with pure SUI or stress-predominant MUI.

In a second attempt, Brubaker et al used data from the UITN Behavior Enhances Drug Reduction of Incontinence (BE-DRI) to again explore operational definitions of MUI, using various thresholds and combinations of the MESA, UDI and 7-day voiding diary.[19] They were unable to identify strict cut-off values for any of these baseline measures that could predict the study's primary outcome (success defined as a 70% reduction in incontinence episodes). Because of this, the authors again recommended using distinct descriptions of both urgency and stress subcomponents when characterizing subjects with MUI until better definitions are developed. One limitation is that the BE-DRI population included primarily women with urge-predominant MUI.

2.3. Current treatment strategies for MUI: Challenges and old assumptions

Based on expert opinion, the primary treatment strategy for MUI typically begins with segregation of symptoms and focus on the most bothersome symptom (SUI vs UUI). Although many women may clearly have one condition that is more bothersome, many have equally bothersome symptoms, or cannot determine which condition is "most bothersome". Behavioral/pelvic floor therapy (BPTx) has been shown to be effective for all types of incontinence[20], and some experts suggest that BPTx should be the first treatment for MUI, regardless of which symptom is more bothersome because it is minimally invasive. Other authors support the first-line use of anti-muscarinics for MUI, despite that the improvement over placebo

has been shown to be only modest,[21] side effects are common, and discontinuation is high, ranging from 43%-83% within the first 30 days of initial prescription.[22]

Although intuitively it makes sense that non-surgical options should be offered first, these recommendations are based on the following assumptions for MUI:

1. OAB and SUI are separate and unrelated conditions

-There is some evidence to suggest that at least in a subset of women, these 2 conditions may be related (proximal urethral funneling causing detrusor overactivity)

2. Treatment should always be initiated in a stepwise, sequential fashion

-There has been little evidence evaluating the potential benefit of combined treatments, and thus the old paradigm of following stepwise treatment remains unproven

3. Surgical treatment should be reserved for women with SUI-predominant MUI because it will worsen OAB symptoms

-Most studies suggesting worsening of OAB symptoms included traditional bladder neck slings and colposuspension and not MUS

4. All women would prefer to take long-term medications over undergoing a surgical intervention

-Adherence to anticholinergics is poor

Clinically, many women with MUI become dissatisfied with conservative treatment and/or the need to take a medication long-term. In practice, there can be much "cross-over" due to patient dissatisfaction when the outcomes of treatment are focused on only one symptom. Many women with "urge-predominant" MUI who have tried BPTx and/or anti-muscarinic therapy will go on to choose surgical treatment for SUI after becoming dissatisfied with the results. Women with equally bothersome OAB and SUI components commonly choose surgery, with or without a trial of BPTx. This "traditional" treatment paradigm for MUI has not resulted to significant advances and we are now challenged to consider new paradigms for MUI.

2.4. Behavioral/pelvic floor muscle therapy (BPTx)

BPTx includes components of behavioral therapy, designed to change behaviors to encourage continence, and pelvic floor muscle therapy, designed to strengthen the pelvic floor muscles, enhance the physiological closure of the bladder neck, and improve coordination. A recent Cochrane review of pelvic floor muscle exercise found that these treatments were effective for both SUI and MUI compared to placebo or no treatment, but women with pure SUI may have better outcomes.[20] The UITN study BE-DRI by Burgio et al evaluated whether combined anti-muscarinic therapy with behavioral therapy would increase the number of women who could discontinue drug therapy while sustaining a significant reduction in UUI.[23] BE-DRI included women with pure UUI or UUI-predominant MUI. Although the addition of behavioral therapy did not improve drug therapy discontinuation, the study found that the combination of behavioral training and drug therapy yielded improved urinary outcomes compared to drug therapy alone. Specific to the MUI population, there is a paucity of literature evaluating whether combined therapies including BPTx that are designed to simultaneously treat both components (bothersome SUI and bothersome UUI) will improve a patient's outcome and perception of her condition.

2.5. Anti-incontinence surgical treatment outcomes in women with MUI

Although "traditional teaching" is that women with MUI should not undergo anti-incontinence surgery for SUI due to potential risk of worsening OAB, this is not supported by recent literature for MUS outcomes. There continues to be accumulating evidence regarding the efficacy of midurethral sling (MUS) for the treatment of MUI (See Table 1). The MUS has proven to be highly effective for SUI treatment with cure rates up to 80% at 1 year[24] and there is more recent evidence supporting improved OAB outcomes also.

A systematic review by Jain et al in 2011 including six randomized trials and seven prospective studies reported that the overall cure rate of urgency and the UUI component of MUI after MUS was 30-85% at a follow-up of a few months to 5 years.[25] Whether authors consider MUS to be helpful or hurtful for MUI often depends on the point of view of a paper, and may also be highly dependent on the definitions used to define "persistent OAB." Some studies report that more than 50% of women with MUI experience complete resolution or improvement of OAB symptoms after MUS treatment.[26] However, other studies report that MUI is a risk factor for failure of both SUI and OAB outcomes[27] or that MUS may exacerbate OAB symptoms. One study reported a failure rate of 42% compared to 12% for SUI outcomes in women with baseline MUI compared to those with SUI alone.[28] Whether there may be specific patient characteristics that are associated with resolution or exacerbation of OAB symptoms also remains unclear.

The Trial of Mid-Urethral Slings (TOMUS) by Richter et al for the UITN randomized 597 women with pure SUI or SUI-predominant MUI (based on MESA scores) to retropubic versus transobturator MUS.[24] Success was a composite outcome, defined as: 1) negative CST; 2) negative pad test; 3) no retreatment for SUI; 4) no self-reported leakage on 3-day voiding diary; 4) no self-reported SUI symptoms; 5) no self-reported retreatment of SUI. At baseline, 70/589 (12%) had detrusor overactivity on UDE, but overall mean urge scores on MESA were low (5.9-6.6 $\pm$ 4 points). One year postoperatively, 11% had persistent UUI (defined as any MESA urge item response of "sometimes" or "often" or post-operative initiation of anti-muscarinic treatment for UUI). The rate of de novo UUI was 1/597 (0.002%). The UDI-irritative subscale scores improved from a mean of 41.2 (25.4), to 9.2 (15.2), and 8.9 (15.1) at baseline, 6 and 12 months, respectively suggesting improvement in OAB symptoms (unpublished data, personal communication). Higher baseline MESA urge scores increased the risk of overall (objective and subjective) sling failure.[29] In a planned secondary analysis evaluating UDE predictors, detrusor overactivity on preoperative UDE was *not* a risk factor for objective or subjective failure.[30]

Barber et al performed a second trial also comparing retropubic versus transobturator MUS for SUI.[27, 31] Although women with baseline detrusor overactivity were excluded, 71% had baseline UUI based on the UUI item on the PFDI-20 questionnaire[32]. At 1 year postoperative 31% of women reported bothersome UUI and 4-10% had new or worsened UUI. 45% of women were failures, defined as a composite outcome of "abnormal bladder function" defined as: 1) incontinence symptoms of any type; 2) positive CST; 3) retreatment for SUI; 4) postoperative urinary retention. Overall, 79% reported Patient-Global Impression of Improvement[3] (PGI-I) scores as "much better/very much better". The 2 UDI-irritative items in the UDI-6 (UUI and frequency) improved from a median of 3 points at baseline to 0 points at 12 months, also suggesting improvement in OAB symptoms (unpublished data, personal communication). In a secondary analysis, baseline UUI was not a risk factor for recurrent UI 1 year postoperatively, but preoperative use of anti-muscarinic medications was. However, 53% (10/19) of women taking anti-muscarinics at baseline were no longer taking them 1 year postoperatively.

A secondary analysis by Palva et al of another randomized trial comparing retropubic versus transobturator MUS evaluated the prevalence of urinary urgency symptoms after MUS.[33] In the original inclusion criteria, only women with a "detrusor instability score" $\leq$ 7 (suggesting pure SUI) were included. However, the authors found that despite this inclusion criteria, a considerable proportion of women reported at least slightly bothersome urinary frequency and UUI on the UDI-6 (~75% reported urinary frequency and 66% had UUI at baseline that was at least "somewhat bothersome"). At 36 months postoperatively, 51-60% were "cured" of urinary frequency and 73-75% were cured of UUI based on UDI-6 responses. The rate of de novo urgency was 3.1-4.5% at 12 months and 5.6-6.2% at 36 months. The authors go so far as to conclude that MUS "can be recommended in cases of mixed incontinence".

Abdel-fattah performed an RCT comparing two transobturator MUS including 341 women with pure SUI or SUI-predominant MUI. In a secondary analysis evaluating only the subset of women with urodynamic MUI, (n=83/341, 24%), 52% of women were cured of urgency, 23% had persistent urgency, and 25% had worsened urgency.[34] 58% were cured of UUI, 24% had persistent UUI, and 19% had worsened UUI at 12 months postoperative. At 12 months, 75% of women with MUI experienced overall "cure" of incontinence based on the PGI-I $\leq$ 2, although in their original report of their primary trial findings, preoperative UUI was a risk factor for sling failure by PGI-I.[35]

In summary, recent secondary analyses of trials have suggested that over half of women may experience improvement and/or "cure" of OAB symptoms after MUS; however, to date there has not been a study focused on strategies to improve outcomes in women with MUI undergoing surgery.

Table 1. Randomized trials reporting midurethral sling outcomes in women with MUI*†

First Author	No. Pts	Inclusion criteria	Primary objective	% MUI at baseline	Follow-up	% postop OAB and definition	De novo OAB	Other relevant findings
Richter,[24, 29, 30]	597	Pure SUI / SUI-predom by MESA	TVT vs TVT-O or Monarc	-12% DO	1 year	10-12% *persistent* UUI (by MESA or treatment)	0.002% New UUI	-MESA urge score risk factor for failure -Baseline DO not risk factor
Barber[27, 31]	170	Urodynamic SUI and no DO	TVT vs Monarc	-71% UUI (PFDI) -14% preop anticholin	1 year	-31% UUI postop (PFDI) -4-10% new/worse UUI (PFDI) -16% anticholin postop	4-10% New / worse UUI	-79% "Cure" by PGI-I <2
Palva[33]	267	Pure SUI / SUI-predom by "detrusor instability score"	TVT vs TVT-O	-75% frequency (UDI) -66% UUI (UDI)	1 year & 3 year	-1 year: 22% frequency 13% UUI (UDI) -3 years: 36% frequency 21% UUI (UDI)	-1 year: 3-4.5% -3 years: 5.6-6.2%	-Only provides postop prevalence of sxs, -unclear % "persistent" or "cured"
Abdel-fattah[34-36]	341	Pure SUI / SUI-predom (undefined)	TVT-O vs ARIS	-24% DO (N=83) -18% prior antimusc	1 year	By BBUSQ: -23% persistent urgency -25% worsening urgency -24% persistent UUI -19% worsened UUI -~25% worsened OAB taking anticholinergics	4.3% UUI	-52% Cure urgency -58% Cure UUI -75% "cure" by PGI-I < 2.

*Excludes small, under-powered RCTs
†TVT™ (Tension free-vaginal tape, Gynecare, Ethicon Inc); TVT-O™ (Gynecare TVT™ Obturator System, Ethicon Inc); Monarc™ (American Medical Systems, Inc), ARIS® (Transobturator Sling System, Coloplast Pty Ltd)

2.6. Limitations of existing MUS trials for the MUI population

The existing MUS RCT data are limited because they do not focus on women with MUI and the inclusion criteria almost always require one condition to be "predominant" or "more bothersome" (e.g. SUI-predominant for most surgical trials and UUI-predominant for most medication trials). Thus, women with equally bothersome symptoms are typically excluded, or may feel pressured to "choose" a most bothersome condition in order to qualify for a trial. In addition, many MUS trials use a composite outcome to define failure (e.g. any self-reported incontinence or incontinence on diary) and therefore it is difficult to tease out

SUI and OAB outcomes separately, which is highly important when counseling a patient with MUI. Finally, ancillary studies from existing trials are underpowered to determine SUI and OAB outcomes separately.

2.7. MIMOSA Trial: First Network trial attempt focused on MUI population

In 2009 the UITN published on their experience with the "Mixed Incontinence: Medical or Surgical Approach" (MIMOSA) trial.[37] MIMOSA was designed as a pragmatic clinical trial randomizing women to nonsurgical treatment (pharmacological therapy and behavioral therapy) versus surgical treatment (MUS including TVT, TOT, TVT-O, fascial sling and Burch). After 4-5 months of enrollment as a feasibility study, 27 women were randomized out of 1190 women screened and the study was stopped due to low enrollment. The investigators felt recruitment was challenging at least in part due to the divergent treatment approaches, but also because of the practical trial design and strict inclusion criteria.

Based on unpublished data from MIMOSA: of 24 women randomized with complete follow-up data at 6 months, 71% met criteria of optimal outcome (defined as score $\leq$ 2 on PGI-I and a score of $\leq$ 2 on PGI-S), suggesting that surgical treatment may improve MUI symptoms at least in the short term.

To avoid the challenges encountered in MIMOSA, the ESTEEM protocol team carefully designed our treatment to ensure a fair perception of treatment arms in an efficacy trial design, and carefully selected inclusion criteria that would not be overly strict, yet still allow recruitment of a MUI population (See Section 4.2, Inclusion Criteria).

2.8. Summary of known and potential risks and benefits of study treatment

BPTx has been shown to be beneficial when used for the treatment of MUI. Other than time and effort commitment and potential discomfort from a pelvic exam, the risks of BPTx are extremely low. MUS has been shown to be an effective treatment for SUI, and recent evidence suggests possible benefit for MUI populations also. However, there are women with MUI who report persistent or worse UUI/OAB symptoms after MUS and this is one potential risk. The remaining risks of MUS are not expected to be different for the ESTEEM population compared to previous studies including pure-SUI or SUI-predominant MUI subjects.

2.9. Significance of proposed study / Rationale for combined surgical and BPTx approach

In summary, at least three gaps of knowledge contribute to the clinical challenge of treating women with MUI who desire SUI surgery. First, there is a lack of data to guide counseling on expected outcomes, particularly for the OAB component after MUS (what happens to OAB symptoms after MUS?). Second, while persistent / worsened OAB symptoms after surgery are associated with patient dissatisfaction, there have been essentially no trials evaluating how to best treat this component peri-operatively (how do we improve OAB outcomes after MUS? Can combined treatment improve outcomes for MUI?). Third, there is little data on what factors may increase the risk of MUS "failure" in this population (who should or should not get a MUS if they have MUI?). ESTEEM will provide the needed information to address these gaps.

Patients with MUI who ultimately elect surgery for SUI are often hopeful that their overall urinary condition will improve, but as surgeons we currently cannot assure patients this will be the outcome. Treatments that can optimize both OAB and SUI outcomes in this population are needed. Studies have demonstrated potential clinical benefit of initiating perioperative physical therapy after other procedures including prostatectomy[38, 39] and orthopedic procedures[40]. Perioperative BPTx combined with MUS may have similar effects on improving OAB outcomes in women with MUI. The index surgery may serve as a "teachable moment" that can be used to reinforce principles and adherence of BPTx to optimize outcomes, and/or an opportunity to affect postoperative tissue remodeling and neuromuscular dysfunction.

2.10. Innovation

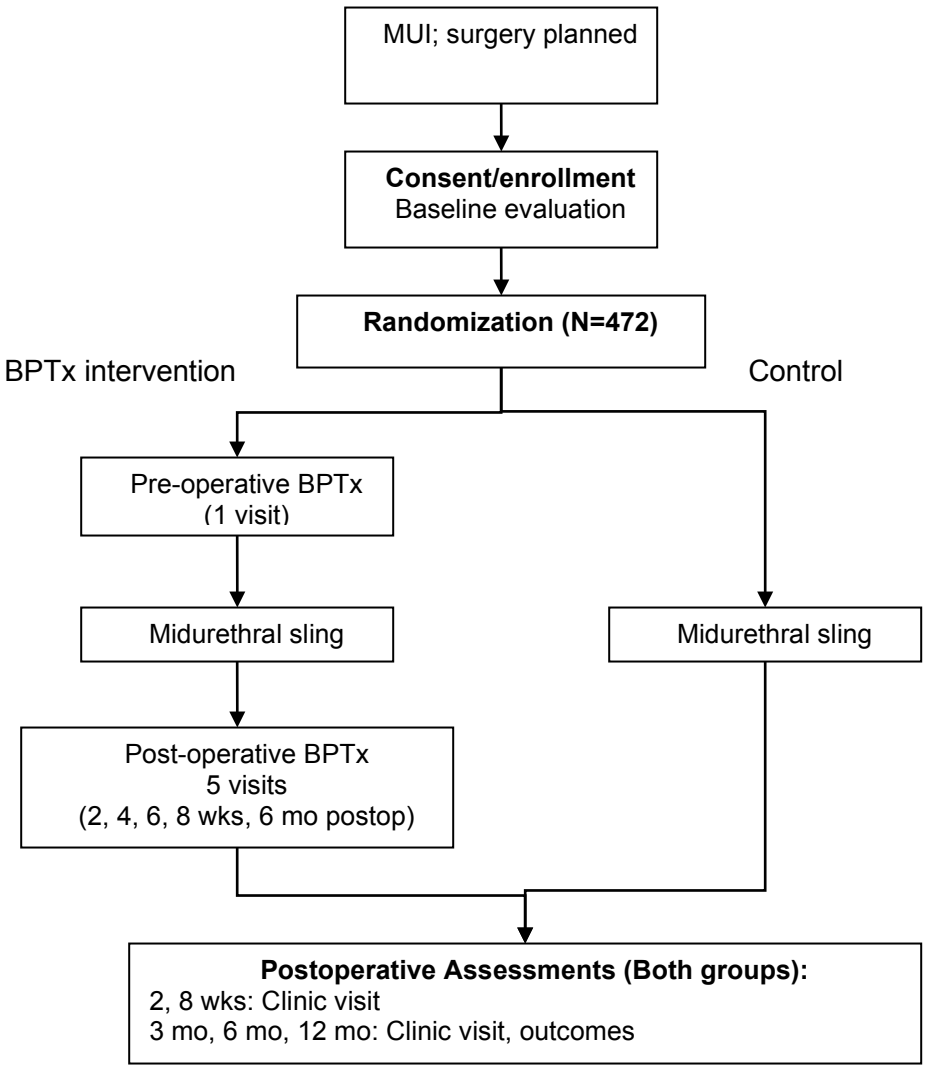
This proposal is innovative for several reasons. First, it studies a population of women who are often excluded from clinical intervention trials but are at high risk for failing segregated SUI and UUI treatments. Second, in contrast to the historical paradigm of initiating treatments separately and stepwise for SUI and UUI, we will evaluate the effect of a combined surgical and non-surgical approach to optimize treatment outcomes. Third, this study will provide critical information regarding OAB outcomes after MUS and will be

powered to allow reporting of OAB and SUI outcomes separately after treatment. Finally, we will gain important predictive information regarding which patients may experience improvement, worsening, or no change in their OAB symptoms. At the completion of this study, we will understand whether a combined behavioral/surgical treatment approach is superior to surgery alone and will have predictive information that will be directly applicable to the clinical care of patients with this challenging condition.

3. STUDY DESIGN

3.1. Description of study design (See Figure 2, Study flow diagram)

Figure 2. Study flow diagram



ESTEEM is a 3-stage, multi-center randomized trial of 472 women with MUI who have elected to undergo surgical treatment for SUI. Participants will be randomized to a peri-operative BPTx program+MUS versus MUS alone. The purpose is to compare combined MUS+BPTx versus MUS alone (control) on improving MUI symptoms at 1 year.

Stage 1: preoperative BPTx versus control
Stage 2: All participants will undergo a MUS
Stage 3: postoperative BPTx versus control (based on initial randomization)

3.2 Masking issues

It is not feasible to mask the patients or interventionists to the BPTx intervention due to the nature of the treatment being studied. The team considered "sham" visits with interventionists; however, based on expert interventionist opinion, sham interventions for UI involving the pelvic floor are extremely difficult to design in a way that is convincing yet maintains the integrity of the intervention itself. We also considered using "general massage" as a potential control group; however, there is some evidence suggesting that psychological stress is associated with OAB and irritative symptoms which could potentially contaminate the control group.[41] Issues of adherence (or over-adherence) in the massage group are also possible, as women could schedule these independently from the study. For these reasons, the team decided it was not feasible to incorporate any sham procedures in the control group.

Study surgeons and outcome assessors will be masked to treatment assignment. All outcome measures will be collected by masked outcome assessors. Study coordinators / clinical staff performing objective measurement of PFM strength will be masked (Aim 7). All patient-reported outcomes (PROs) will be administered prior to other clinical assessments or procedures.

Table 2. Masking in ESTEEM

Study individual	Masking
Study participant	No
Interventionist	No
Outcome assessors (includes clinical staff performing PFM measurement)	Yes
Study surgeon	Yes

Efforts will be made by unmasked research assistant/staff members to remind the patient that the surgeon is masked to her treatment assignment. If she desires additional treatment, it is likely the surgeon would offer BPTx as additional treatment and she will be reminded that she can decline additional BPTx without revealing to her surgeon that she received the BPTx intervention. Such methods have been effective for past PFDN trials (e.g. OPTIMAL trial[42]).

3.3. Randomization and Stratification

Patients will be assigned to one of the two treatment groups with a randomization sequence prepared and maintained centrally by the Data Coordinating Center (DCC). Allocation to the treatment groups will be 1:1. Randomly ordered permuted blocks will be used, with block sizes known only by the DCC. The web-based data management system will provide the treatment assignment for each participant as she is randomized. Thus, the allocation sequence will be concealed from clinical site staff.

Randomization will be stratified by clinical site. It is important that UUI "severity" is comparable in both groups as it is a potential risk factor for treatment failure. Therefore, randomization will also be stratified based on UUI "severity" which will be defined by the number of urgency urinary incontinence episodes (IEs) on diary. This will ensure that women who have more frequent, or more "severe" UUI are equally distributed between the two groups. SUI severity is less of an issue because all subjects will be receiving the same treatment for SUI.

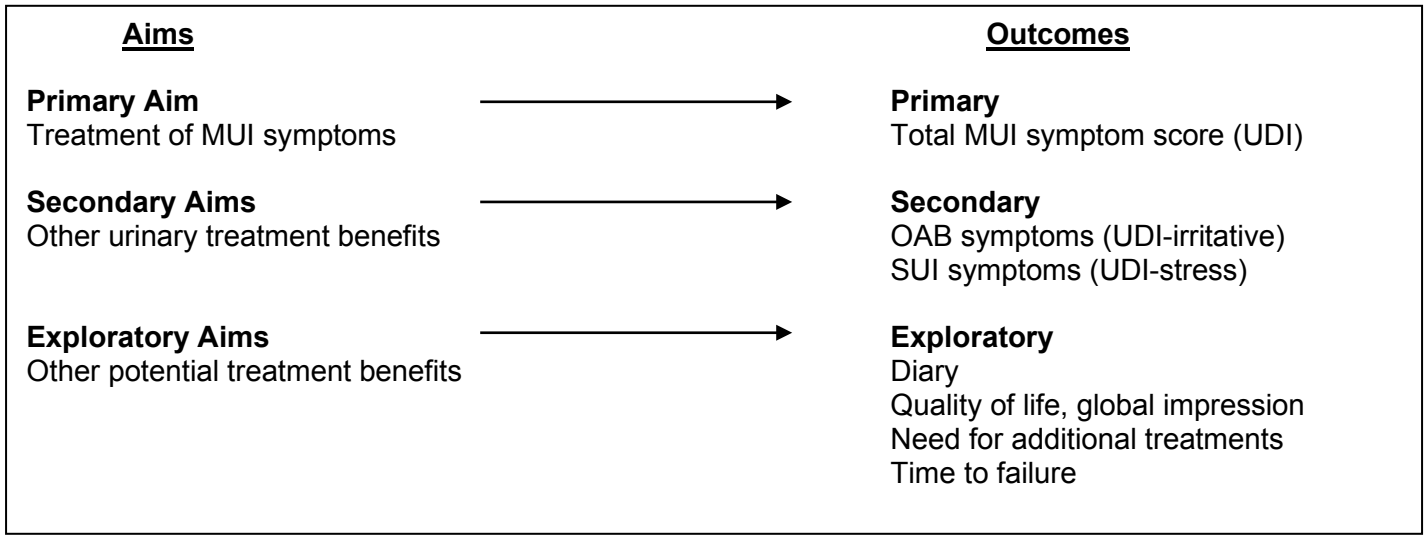
Burgio et al,[43] identified risk factors for unsuccessful behavioral treatment of urge/urge-predominant MUI. Women who had $\geq$10 IE/week on a 7-day diary diary at baseline were much less likely to be completely continent after behavioral treatment. Therefore, for a 3-day diary, this would translate into ~4 IEs on a 3-day diary as a potential risk factor for treatment failure.

There is limited data to support stratification based on presence of preoperative DO on UDE. One study found that up to 38% will have resolution of DO after MUS.[44] Other studies suggest that baseline DO is a risk factor for postoperative UUI.[45] Still other studies even suggest that baseline DO is associated with greater improvement in OAB symptoms postoperatively. Choe et al evaluated 132 women with MUI who underwent MUS and found a higher proportion of women with preoperative DO had complete resolution of OAB symptoms postoperatively compared to those without DO (37% vs. 18%).[46] A secondary analysis of TOMUS data supported that more severe UUI (by MESA score) was a risk factor for non-SUI sling failure after MUS (or failure due to UUI);[29] however, baseline DO was not a risk factor (28% vs 21% objective failure for women with and without DO, respectively).[30]

Based on the existing evidence, the team reached consensus that women should be stratified based on a cutoff of ≥ 4 urge IEs on 3-day diary. The team agreed that there was insufficient data regarding preoperative DO to stratify by this variable; however, this data will be collected for exploratory analyses.

3.4. Outcomes

Figure 3. ESTEEM outcomes



3.4.1. Detailed Description of Primary Study Outcome

The primary outcome for this study is the mean change from baseline in UDI-total score at 1 year postoperative. The UDI is a validated, disease specific patient-reported outcome (PRO) measure. A PRO is a measurement of any aspect of a patient's health status that comes directly from the patient (without interpretation by the physician, researcher, other). In clinical trials, symptom indices and quality of life PRO instruments are being increasingly used as primary outcomes and supported by federal agencies.[37, 47, 48]

The long form of the UDI is a 19 item, validated UI symptom specific questionnaire with 3 subscales: stress, irritative, and obstructive symptoms.[2] Higher scores represent more severe disease or bother from the patient perspective. Construct validity (convergent) was originally established by demonstrating significant correlation between the overall UDI and its subscale scores with the number of IEs on 7-day diary and pad tests. Criterion validity was established by correlating total and subscale scores with physician diagnoses. The UDI can effectively discriminate between known UI clinical groups and diagnoses (specifically genuine SUI, urodynamic detrusor overactivity, or mixed) and is responsive to change. These are some minimum qualities needed for valid interpretation of a PRO in a clinical trial.

Although it is fairly simple to determine the statistical significance of a change in a symptom index, placing the magnitude of these changes in a context that is meaningful for patients is more difficult. *The*

minimum important difference (MID) of a measure is a score change that should reflect a clinically meaningful response to treatment and represents the "between group criterion" that needs to be met or exceeded in order for study results to be considered clinically meaningful. From the patient perspective, MID can be defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial..." [49] It is useful for interpreting questionnaire results for both within-group and between-group differences and represents the magnitude of benefit for which trials should be powered to minimize type 1 and type 2 errors. Although no single approach to determine MID is perfect, a combination of approaches is often used to determine a reasonable range of MID scores. Importantly, there are published MID ranges for the total UDI score and its subscales for urge predominant and pure stress/stress predominant urinary incontinence populations.

Table 3 summarizes the relevant published MID data for the UDI. Dyer et al used the BE-DRI study population to determine MID values for the UDI and UDI-irritative subscale.[50] The BE-DRI population included 94% subjects with urge-predominant MUI based on bladder diary with a baseline mean UDI total score of 120 (49) points and UDI-irritative subscale score of 58 (22) points. Using anchor based methods, the authors recommend an MID of -35 for total UDI and -15 for UDI-irritative scores for this population.

Barber et al used the Ambulatory Treatments for Leakage Associated with Stress Incontinence Trial (ATLAS) study population and determined MID values for the UDI total and UDI–stress subscales.[51] This population was pure/stress-predominant MUI, undergoing conservative treatment for SUI. The baseline mean UDI score was lower at 80 (40) points and UDI-stress was 47 (19) points. Based on their findings, the authors recommend an MID of -11 and -8 for the UDI total and stress subscale scores respectively.

Table 3. Published MIDs for the Urogenital Distress Inventory

Trial/Author	Population	Endpoint/ intervention	UDI component	Anchor-based MID	Distribution-based MID (1/2 SD)	Recommended MID
BE-DRI, Dyer[50]	Pure urge/Urge-predominant MUI	8 month Meds +/- BPTx	UDI-total	-45 to -36	-25	-35
			UDI-irritative	-20 to -18	-11	-15
ATLAS, Barber[51]	Pure SUI/SUI predominant MUI	3 month Pessary vs BPTx vs both	UDI-total	-22.6 to -6.4	-21.9 to -18.8	-11
			UDI-stress	-16.5 to -4.6	-10.6 to -9.1	-8

Published MIDs are important for estimating sample size and interpreting findings, however there are at least 3 different ways we can analyze the UDI scores:

#1. Compare postoperative *mean UDI scores* between groups at 1 year
#2. Compare *mean changes* (delta) in UDI scores from baseline to 1-year between groups *(preferred, see below)*
#3. Dichotomize "success" and "failure" as women who achieved a 35 point improvement versus those who did not (also known as "responder analysis")

We chose not to dichotomize our outcome for many reasons (option #3). Dichotomizing women as "success" or "failure" based on MID could simplify interpretation; however, using purely a responder analysis approach has limitations and some authors recommend avoiding this for primary analyses in trials.[52] One disadvantage of responder analysis is reduced power and efficiency compared to analysis on the original scale, primarily due to the loss of information associated with lumping groups together. Particularly relevant to ESTEEM, some women could have *worse* scores compared to baseline and this information would be lost because they would be grouped with those who may have "slightly" improved, but

just not enough to be classified as a “success”. Also relevant to ESTEEM, if both groups worsened but one group “worsened less”, this information would also be lost using this approach.

Because the trial is randomized and we will be stratifying by UUI severity, we would expect the baseline UDI scores to be similar between groups. If the average baseline score is the same in the two groups, then comparing the mean change in UDI score between groups (option #2) is mathematically equivalent to comparing post-operative UDI scores at 1 year between groups (option #1). However, option #2 has some advantages in that (1) if baseline scores are not the same in the two groups, comparing the mean change in UDI score between groups at 12 months will account for that baseline difference, and (2) UDI scores typically have a distribution that is highly skewed, but differences from baseline should be close enough to being normally distributed that analysis methods that assume a normal distribution can be used.

3.4.1.a Rationale for using UDI as primary study outcome

Due to limitations in how to best define successful treatment of MUI, the investigators had extensive discussion around whether an objective, subjective, or composite outcome would be best for this trial. The team agreed that the primary outcome must remain true to the clinical question and be clinically relevant in capturing both potential benefit and harm of both the control and the intervention for this trial. It is critical that the outcome is meaningful from the patient perspective and will be able to capture OAB improvement, worsening, or no change in symptoms. The team discussed using bladder diary, patient global impression, or OAB PROs. Arguments against each of these were based on the following rationale:

1. Problems with using bladder diary as primary outcome:

a. Diary does not capture a meaningful patient outcome- It is becoming clear that typical clinical trial endpoints such as reduction in IEs, voided volumes, etc do not capture what is meaningful to patients. Counting IEs on diary likely does not capture what is important to a patient (e.g. having 3 large urge leaks a day may be more bothersome than having 20 small stress leaks or having 20 urgency associated voids may be more bothersome than having 1 UUI episode). In addition, diary IEs do not correlate perfectly with patient satisfaction.[53] Finally, bladder diaries have been shown to be less reliable in women with MUI, particularly for the SUI component.[54]

b. Diary cutoffs to define improvement for MUI are unknown-What percent improvement for the SUI component and for the UUI component is clinically important for a woman with MUI? Any cutoffs chosen would be arbitrary.

c. Using IEs on bladder diary as a primary outcome would require a minimum number of IEs (approximately 3-4 IE/3days) at baseline to be able to detect a change. The protocol team felt that setting such strict inclusion criteria would be too limiting to allow recruitment of a good range of MUI severity (see Inclusion Criteria, Section 4.2).

For all of these reasons, the team decided against using bladder diary IEs as the primary outcome and to instead focus on measures that can capture outcomes from the patient perspective.

2. Problems with using global impression measure as primary outcome

A patient’s overall/global impression of improvement would be reflective of her overall urinary condition. Although this outcome would seemingly be ideal for capturing a meaningful outcome, for our trial it could potentially introduce bias. Because it is not feasible to mask subjects in ESTEEM to the intervention (BPTx), a single, subjective global impression item would be subject to bias. For example, if subjects in the control group were more likely to ask for additional treatment and report they were not “improved” because they knew there was another potential treatment available that they did not receive, this would bias our study towards a higher failure rate in the control group (making our intervention seem more effective than it really is). The challenges of masking or designing a sham procedure for the control group for ESTEEM have already been noted above (Section 3.2).

3. Problems with using OAB PRO measure as primary outcome

Finally, we considered using an OAB PRO as the primary outcome, such as the Overactive Bladder Questionnaire (OAB-q)[55] or the UDI-irritative subscale. However, these do not account for SUI symptoms, which are part of the MUI symptom constellation. In addition, it is still unclear whether patients with MUI are at risk of sling failure for SUI and at least 2 studies suggest this may be the case.[27, 28] Finally, there are some women with MUI who may not be able to clearly distinguish all UI episodes as stress- or urge- related and the team felt it would be important to also capture these symptoms. Finally, specifically regarding the OAB-q, there is less validity data in a MUI population compared to the UDI.

For all of these reasons, using the UDI as the primary outcome is ideal and it has all of the characteristics that are important for a MUI population:

1. The overall UDI score includes both a stress and irritative subscale, allowing us to comprehensively capture both SUI and OAB symptom outcomes.
2. It captures a meaningful outcome from the patient perspective, incorporating both the presence and bother of SUI and OAB symptoms.
3. It includes 3 UI items that are not necessarily specific to stress or urge and thus can help capture UI episodes for which patients cannot clearly distinguish as SUI or UUI.
4. It can capture both improvement and worsening of preexisting symptoms, but also the development of new urinary symptoms.[56]

Because MUI includes both SUI and UUI, it is important to be able to report SUI and OAB outcomes separately. There is no clinical rationale for assuming that one component, or that one subscale of the UDI is more important to women than another. Therefore, the UDI-stress subscale and UDI-irritative subscale will be important secondary outcomes for which ESTEEM will be powered to detect differences and each will have a priori analysis plans (see Section 6, Statistical Considerations).

3.4.1.b. Rationale for timing of primary outcome:

There was significant discussion regarding the best timing for the primary outcome. In framing this question, the group considered at which time-point would a difference in outcome lead to recommendation of BPTx as part of clinical practice. Long-term outcomes of 1 year and/or more are “standard” for surgical trials and are important to determine if a surgical treatment is worthwhile. However, outcomes for BPTx trials are often shorter, between 3-6 months and there was concern that longer time points may miss improvements which may not be sustained over time. Clinically, women with MUI who ultimately have persistent OAB symptoms seem to experience a re-occurrence of these symptoms within 3-6 months of the index MUS surgery and therefore, many investigators felt it was important for the outcome to be at least 6 months or greater.

Based on these considerations, the primary outcome will be the change in UDI score from baseline and 12 months postoperative, given the intervention is the combination of BPTx and surgery, with Time 0 = the time of surgery. Note that if a participant is randomized but surgery is not performed, then Time 0 will be the planned surgery date. A secondary outcome will include time to failure; therefore, we will be able to detect any potential early differences that are not sustained at 12 months (See Section 6, Statistical Analysis). Additional assessments will be made at 3 and 6 months postoperatively, which will allow for shorter-term assessments of BPTx effects.

3.4.1.c. Management of subjects who request additional treatment for SUI and/or OAB after MUS:

The overarching goal of ESTEEM is to evaluate the effect of combined treatment on improving both SUI and UUI outcomes in women with MUI. Therefore, any request for additional treatment for any lower urinary tract symptoms (SUI, UUI/OAB, voiding dysfunction) before the 1 year outcome for either of these symptoms will be considered treatment failure. The team agreed it would be difficult to withhold additional treatment from either group for the 1 year study duration; however, any additional treatment should be initiated after the acute postoperative recovery period. Clinically, some women may experience immediate

exacerbation of OAB symptoms after MUS followed by improvement, whereas other women may experience initial improvement but then recurrence of OAB symptoms several months later. Therefore, the team came to a consensus that any additional treatment should be deferred until 3 months postoperatively when OAB symptoms would be expected to have reached a baseline. This will allow enough time for complete physical and tissue recovery from the surgical procedure, will allow for assessment of potential BPTx early benefits, and will provide information on the natural course of OAB symptoms in the early postoperative period which is important for clinical counseling and decision-making. Any subjects receiving additional treatment prior to the 3 month time point will be considered a protocol deviation.

In the event that a randomized participant decides not to undergo surgery but then later changes her mind and has MUS surgery, the surgery will be considered additional treatment. For purposes of calculating follow up windows, the date of the original planned surgery that did not occur will be Time 0.

Subjects who initiate additional treatment will be asked to complete all primary and secondary outcome measures prior to initiation of additional treatment. The type of additional treatment will not be limited and will be left to the physician's clinical judgment. This may include (but is not limited to) behavioral and/or pelvic floor therapy, continence pessary, medical therapy, other procedure-based treatments (posterior tibial nerve stimulation, Botox), and surgical (sling revision, re-placement, sacral neuromodulation). Statistical models designed to specifically account for subjects who initiate additional treatment will be used. Please see Section 6, Statistical Considerations section for details on how the analysis of primary and secondary outcomes measured at 1 year will account for any additional treatment requests for SUI and/or OAB if initiated prior to the 1 year time point.

3.4.2. Secondary outcomes

Consistent with Brubaker et al who emphasized the importance of characterizing the OAB and SUI components separately for MUI populations, we will ensure adequate power of our trial to detect differences in OAB and SUI symptom outcomes separately.

3.4.2.a. Urge urinary incontinence/overactive bladder symptom outcomes

Because the primary clinical problem in this population is the potential for persistent or worsening OAB after MUS, it is highly important to capture and report on the cardinal symptoms of OAB from the patient perspective. The UDI-irritative subscale measures symptom burden, impact, and changes related to OAB which are important aspects that cannot be directly observed or otherwise measured. It is highly responsive to treatment-related change and is able to discriminate among levels of change in all bladder diary variables (urinary urgency, frequency and urge incontinence) and patient ratings of treatment benefit. Particularly for ESTEEM, this comprehensive OAB measure will be important to understanding how MUS may affect all OAB symptoms individually and as a whole.

3.4.2.b. Stress urinary incontinence symptom outcomes

It is also important to be able to report on SUI outcomes separately. The majority of studies have not demonstrated significant differences in efficacy for SUI outcomes for subjects who had MUI preoperatively; however the majority of studies only had small subsets of women with MUI. Two studies have suggested worse SUI outcomes in women with MUI at baseline (see section 2.4 above). One study by Paick et al evaluated 274 women, of which 73 had MUI and reported cure rates for SUI to be 78% for the MUI group and 95% for the pure SUI group.[57] They also reported that maximal urethral pressure at baseline was associated with a greater risk of persistent OAB, suggesting the possibility that profound urethral dysfunction may contribute to persistent symptoms. A study by Gleason et al using data from the University of Alabama including 534 women with MUS found that women with MUI had higher rates of SUI compared to women with SUI only (36% vs 16%, p<.001) with an adjusted OR = 2.7 (95% CI 1.7, 4.2) (unpublished data). In addition, because BPTx can also treat SUI, it is important to know if women randomized to

MUS+BPTx have improved SUI symptom outcomes as well. Therefore, as a secondary outcome, we will also compare SUI outcomes between women randomized to MUS + BPTx versus MUS alone. SUI symptoms will be measured using the UDI-stress subscale.

3.4.3. Other outcomes

3.4.3.a. Other UUI/OAB outcomes

i. Bladder diary – We will assess the change in IE frequency and type, number of urgency episodes, urgency severity with voids, number of diurnal voids, and number of nocturnal voids and compare these variables between groups at 6 months and 1 year.

ii. Overactive Bladder treatment satisfaction (OAB-SAT-q)[58]-The OAB-SAT-q is an 11 item instrument designed to assess patient satisfaction with treatment in a clinical setting. There are three 3-item subscales (Satisfaction, Side Effects, Endorsement) and two single items (Convenience, Preference). Response options are presented on 4-, 5-, and 6-point Likert scales. It has demonstrated good psychometric properties in OAB/UUI patients receiving anticholinergic and anticholinergic + behavioral therapies. We will compare change from baseline in OAB-SAT-q scores at 6 months and 1 year between treatment groups.

iii. Overactive Bladder Questionnaire-Symptom subscale (OAB-q) – The OAB-q is a validated, responsive questionnaire that includes 8 symptom bother items (SS) and 25 health related quality of life (HRQOL) items of 4 subscales (coping, concern, sleep, and social interaction).[55] In a systematic review of UI questionnaires by Avery et al, the OAB-q was rated as “grade A”, highest recommendation specifically for OAB symptoms.[59] Each item is rated on a 6-point Likert scale ranging from “not at all bothered” to “a very great deal bothered” for symptom items and “none of the time” to “all of the time” for HRQOL items. Subscales are summed and transformed into scores ranging from 0-100 with higher bother scores indicating increasing symptom bother and higher HRQOL scores indicating better quality of life.[60,61] We will compare change from baseline in OAB-q scores at 6 months and 1 year between treatment groups.

3.4.3.b. Time to failure

For analyzing time to failure, “failure” will be defined as initiation of any additional treatment for either SUI or UUI/OAB symptoms during the follow-up period. Subjects lost to follow up will be censored at the time of their last visit.

3.4.3.c. Quality of life/global impression

We will compare change from baseline in the scores below at 6 months and 1 year between treatment groups.

a) Incontinence Impact Questionnaire (IIQ)
b) Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ) [62]
c) European Quality of Life-5 Dimensions (EQ-5D)[63]
d) Adaptation Index
e) Patient Global Impression of Improvement (PGI-I)[3] and Patient Global Impression of Severity (PGI-S)[3]:

3.4.3.d. Safety/additional treatment

a) additional re-treatments for SUI or UUI within 12 months of treatment, and type of re-treatment
b) return to OR for sling revision due to worsened OAB symptoms

3.4.3.e. Pelvic floor muscle (PFM) strength

PFM strength has traditionally been measured subjectively by a clinician or interventionist using the Brink score in many previous studies, including Network studies. However, because this is a subjective measure, it may be subject to bias. Although there are many trials showing symptom improvement with pelvic floor therapy[20], there are limited studies evaluating the association between PFM *strength* and improvements in UI symptoms. To contribute to the literature about this issue, in ESTEEM we will objectively assess PFM strength changes using the Peritron Perineometer. Peritron is an advanced pressure biofeedback perineometer specifically designed for pelvic floor assessment. Pelvic floor muscle contraction creates pressure in the sensor that is transferred and displayed on a "Readout Unit" which is small and handheld.

FIGURE 4. Peritron Perineometer

Example: Peritron 9300 Device (www.win-health.com/perineomter.html)

After a thorough search of the literature and discussion with other experts in the field, the protocol investigators concluded that the Peritron device has adequate evidence to support its validity, including test-retest reliability and inter-rater reliability, for both baseline and maximum contraction pressure measurements. In addition, studies support its reliability in "normal", continent controls as well as women with UI.

Studies evaluating the Peritron's reliability properties are in Table 4. A study by Hundley et al supports the reliability of measurements from this device in postmenopausal, parous women (inter-rater reliability for baseline and maximum pressure 0.78 to 0.88).[64] This is supported in normative women as well (correlation r=0.83).[65] The Peritron device provides a potential method of determining an objective measure of PFM strength. Measurement using the Peritron device will be standardized and Principal Investigators at each site will be trained on how to use the device and will be responsible for training their clinical staff and for quality assurance of Peritron use. Clinical staff performing the Peritron measurements will be masked to the intervention the subject received. PFM measures (Maximum squeeze amplitude and duration of squeeze), will be performed at baseline, at the first post-operative visit after surgery (2 weeks), 8 weeks, and at the primary endpoint (12 months) – See Assessment Table 11. Changes in squeeze measures from baseline at 8 weeks and 12 months will be compared between treatment groups.

Table 4. Validity properties of Peritron device

Author (year)	N	Subject characteristics	Study aims	Peritron Findings
Kerschan-Schindal et al (2002)[66]	37	Postmenopausal all with UI (28 SUI; 5 UUI; 4 MUI)	1. To examine the test-retest reliability of several PFM measures. 2. To correlate findings between different measures.	**Peritron Reliability:** -ICC for max contraction = 0.97 -ICC for mean contraction over 5 seconds = 0.95 -Correlation between max force and mean contraction force over 5s = r =0.95. **Correlations with other measures:** - urine stop test r = 0.88 max force - digital exam r = 0.70 max force - pad tests r = -0.33 and -0.28 for max
Hundley et al (2005)[64]	100	Mean age 48 (22 to 85) yrs 46% postmenopausal	1.To compare Brink scores with Peritron measurement 2. Determine intra- and inter-rater reliability for the Peritron.	**Peritron Reliability:** -Interrater reliability max pressure, r = 0.88 **Brink Reliability:** -Interrater Brink for total score = 0.68, pressure = 0.68, vertical displacement = 0.58, and squeeze duration = 0.44 **Correlations with other measures**: - Brink pressure r =0.67
Bo et al, 2005[65]	20	"Normals" PT students Mean age 25.1 (21-38) yrs.	To assess whether max vaginal squeeze pressure differed when measured with 2 different sized probes.	**Peritron Reliability:** Test-retest: $r^2 = 0.83$
Frawley et al (2006)[67]	20	19 female PT (1 unable to contract) Age range 25-65 yrs Some parous subjects reported mild UI and/or prolapse	1. To determine the intra-therapist reliability for digital muscle testing and vaginal manometry on max voluntary contraction and endurance. 2. To establish how reliability varied with different tools and different testing positions.	**Peritron Reliability:** -Test-retest for Max pressure: r=0.91 to 0.96 across positions (supine lowest at 0.91). -pressure endurance r=0.05 to 0.41with hooklying the lowest
Rahmani et al (2009)[68]	15	20-50 yrs	1. Test-retest reliability	**Peritron Reliability:** -Test-retest (same day) Max pressure: ICC=.95 -Test-retest (same day) Endurance: ICC=.94 -Test-retest (between-days) Max pressure: ICC=.88 -Test-retest (between day) Endurance: ICC=.83

3.4.3.f. Cost-effectiveness outcomes

The cost-effectiveness analysis will be conducted from a societal perspective and will be expressed as incremental cost required to produce one additional unit of quality-adjusted life year (QALY). Data on each subject's use of medical and non-medical resources, related to urinary incontinence will be collected during

the follow up period. Direct and indirect costs of the treatment of urinary incontinence with combined midurethral sling (MUS) and peri-operative behavioral/pelvic floor therapy (BPTx) compared to MUS alone and women's preference for health states for improvement in urinary incontinence will be estimated.

We plan to capture incremental direct health care, direct non-medical, and indirect resource use related to study interventions and complications and other urinary incontinence management (such as other UI treatment, UI products and management of side effects). Costs will be estimated using the resource costing method. Direct medical service use collected from each study case report form and direct non-medical and indirect costs collected from patient questionnaires are monetized by multiplying the number of units of each resource use by the average unit cost of this item in dollars. Detailed individual cost data will not be collected. This method allows a consistent capture of resource use when costs are incurred across multiple health systems or payers. Detailed case report forms, that include the interventions performed (e.g. midurethral sling surgery and behavioral/pelvic floor therapy sessions) and clinical events (e.g. complications and additional treatment) will be completed by the study coordinator at study visits. Patient questionnaire on direct non-medical costs (e.g. pads, laundry) and indirect costs (e.g. time, lost productivity) will be completed at study visits 3, 6 and 12 months. Data from medical resource types (physician visits, behavioral/pelvic floor therapy sessions, medications, hospital admissions and emergency room visits) will be collected. Cost for each direct medical service use, direct non-medical items, and indirect items will be assigned based on national Medicare reimbursement rates or other standardized unit costs as indicated in the following Table 5.

Table 5: Resource utilization data collection and price data source, by utilization category

Service	Source Documentation	Price Weight
Surgery: midurethral sling	Case Report Form	Medicare reimbursement
Behavioral/pelvic floor therapy	Case Report Form	Medicare reimbursement
Medication	Case Report Form	Drug Red Book
Physician visit	Case Report Form	Medicare reimbursement
Complication: surgery	Case Report Form	Medicare reimbursement
Complication: hospitalization	Case Report Form	Medicare reimbursement
Complication: ER visit	Case Report Form	Medicare reimbursement
UI products	Questionnaire	Average national cost
UI laundry / dry cleaning	Questionnaire	Average cost
Time	Questionnaire	Average cost
Lost Productivity	Questionnaire	Average cost

Rationale for using the EQ-5D to measure Utility Values

The European Quality of Life-5 Dimensions (EQ-5D) (EuroQol Group, http://www.euroqol.org), preference-based utility index algorithm will be used to calculate each subject's utility index.[69] This instrument will be collected at baseline and follow up study visits (3, 6, and 12 months). The EQ-5D has 5 attributes (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) with 3 levels each for a possible 243 unique health states. The EQ-5D scoring Function is based on the time-tradeoff method with UK Scores ranging from -0.59 to1.00 and US Scores from -0.11 to 1.00. This instrument has been previously validated in women with urinary incontinence (Penn preliminary data, Tables 6 and 7) and used in women with urinary incontinence.[70, 71] These data will be used to compare change in QALYs between the two treatment groups. We are choosing to use a general scale to calculate change in utilities (rather than condition-specific) to allow for comparison of cost-effectiveness results with other interventions and diseases.

A questionnaire to measure direct non-medical costs (e.g. pads, laundry) and indirect costs (e.g. time, lost productivity) will be administered. Based on similar questionnaires used in SISTEr[17] and ValUE[72] studies, this instrument should take approximately 15 minutes for a subject to complete at baseline and 3, 6 and 12 months.

Table 6: Mean utility preference scores for women with urge incontinence, stress incontinence and mixed incontinence.

	UUI (n = 40)	MUI (108)	SUI (n=54)	p-value[a]
HUI-3	0.78 ± 0.23	0.79 ± 0.24	0.86 ± 0.15	0.29
EQ-5D	0.71 ± 0.23	0.73 ± 0.26	0.81 ± 0.16	**0.02**
SF-6D	0.76 ± 0.12	0.74 ± 0.12	0.81 ± 0.11	**0.02**
VAS [b]	0.78 ± 0.15	0.78 ± 0.16	0.80 ± 0.14	0.63

UUI = urge incontinence SUI = stress incontinence, MUI = mixed urge and stress incontinence
[a] Kruskal Wallis [b] VAS scores were divided by 100 to enhance comparability

Table 7: Utility preference score correlations with symptom severity and condition-specific HRQOL measures

	HUI-3	EQ-5D	SF-6D	VAS
	r- value[a]	**r-value[a]**	**r-value[a]**	**r-value[a]**
PFDI-20 score	-0.32	-0.42	-0.37	-0.22
Bladder subscore	-0.16	-0.26	-0.24	-0.23
PFIQ-7 score	-0.45	-0.48	-0.50	-0.32
Bladder subscore	-0.29	-0.31	-0.41	-0.26

Lower scores on the HUI-3, EQ-5D, SF-6D and VAS represent worse utility values while higher scores on the PFDI, ISI and PFIQ represent worse symptom severity and quality of life. [a] Spearman correlation

4. SELECTION OF PARTICIPANTS

Adult women aged 21 or older with bothersome MUI (defined as bothersome SUI and UUI) will be eligible.

4.1. Eligibility Criteria/Rationale for inclusion/exclusion

4.1.1. Defining the ESTEEM MUI population

For ESTEEM, women must demonstrate both subjective bothersome SUI and UUI and objective documentation of both conditions. The team wanted to ensure that our eligibility criteria would identify the appropriate MUI population, but wanted to avoid overly strict criteria that may hinder recruitment such as in MIMOSA.

However, as already discussed, the MUI population is difficult to define. Currently, an instrument that can clearly segregate SUI versus UUI symptoms and assess the magnitude of bother that is predictive of clinical outcomes for MUI *does not exist*. Therefore, defining our inclusion criteria for this MUI population is critical, but we recognize that whatever criteria are selected may not be considered to be strictly "evidence-based".

We reviewed the literature on common definitions of SUI and UUI used in previous clinical trials to help determine our criteria. Trials for SUI often use a subjective report of SUI in combination with a positive cough stress test (CST). CST has a 90-100% test-retest reliability.[73] For OAB and UUI, trials often use bladder diary to document the diagnosis. More invasive UDE has not been shown to predict treatment outcomes for SUI and has a reliability similar to the CST[72, 74, 75]. For OAB, DO is a urodynamic observation but most often is not documented on UDE.[76] There is poor agreement between OAB symptoms and DO and the presence of DO does not predict outcomes of a variety of OAB treatments.[77] Therefore, trials have

moved away from strictly using UDE parameters as criteria and similarly, we will not use strictly UDE parameters as inclusion in ESTEEM.

Because no single measure captures our criteria of providing subjective and objective documentation of both conditions, we will use a combination to define MUI in ESTEEM. This includes subjective documentation of at least moderately bothersome SUI and UUI on UDI, objective documentation of both SUI and UUI on diary, and objective documentation of SUI by CST or UDE.

The team reviewed bladder diary criteria for existing SUI and UUI trials (summarized in Table 8). Ultimately the goal of ESTEEM is to capture those women who have MUI that are most clinically challenging because it is unclear which to treat first and for which a MUS potentially could be efficacious, detrimental, or neutral. It is *not* the patient who has severe UUI who needs sacral neuromodulation that we are interested in recruiting for ESTEEM. In addition, unlike previous UUI trials, because our primary outcome is *not* defined by diary improvement, the diary will be utilized only to document the presence of both SUI and UUI IEs. Therefore, the number of IEs does not have to be set "so high" solely to allow demonstration of outcome improvement.

Therefore the team decided that at least 2 incontinence episodes must be documented on a 3-day diary: a minimum of 1 documented episode of SUI and 1 documented episode of UUI would be appropriate for documenting MUI. In addition, patients must also report at least moderate bother from both SUI and UUI on the UDI to be eligible *and* desire surgical treatment of SUI symptoms. This will allow appropriate documentation of both conditions, but would not be overly strict so as to exclude women on either the mild or severe end of the spectrum.

4.1.2. Targeting a population that is distinct from TOMUS

There were significant improvements in the UDI-irritative subscale scores in the TOMUS trial. Ideally we want to target a population with more severe urge symptoms, since additional effects of BPTx would be difficult to detect in a population too similar to TOMUS. In general the MESA urge score in TOMUS was low at a mean of 5 points. Requiring documentation of UUI on diary and report of at least "moderate bother" from UUI on the UDI will help to ensure a more severe UUI population (with MUI) than TOMUS.

Table 8. Bladder diary inclusion criteria for other relevant trials

Study	Interventions	Inclusion	Final population diary characteristics	Outcome definition
UUI trials utilizing diary for inclusion				
Burgio-BE-DRI[23]	Tolterodine/BPT vs Tolterodine alone	≥7 UIEs on 7-day diary and UUI>SUI on MESA	1% UUI 7-13 IE/wk 1% UUI ≥14 IE/wk 30% MUI 7-13 IE/wk 68% MUI ≥14 IE/wk	70% reduction IEs, no other UUI treatment, withdrawal of antichol at 8 months
Visco-ABC[78]	Anticholinergic vs Botox	≥5 UIEs on 3-day diary and >50% UIE/IE	Mean (SD) IEs/day: 5.6 (3) Urge IEs/day: 5.0 (2.7) Stress IEs/day: 0.8 (1.0) Other IEs/day: 0.1 (.4) Mean voids/day: 7.9 (3) Mean voids/night: 1.6 (1.3)	Change in IEs on 3-day diary monthly, from 1-6 months
Amundsen-ROSETTA	Interstim vs Botox	≥ 6 urge IEs/3-day diary	-	Change in IEs on 3-day diary
Other relevant trials that did not utilize diary for inclusion				

Brubaker-MIMOSA[37]	Initial surgical treatment vs initial non-surgical treatment	No BD -MESA urge>stress or urge score ≥ 7 and moderate or great bother on UDI-6 and moderate or severe UI on PGI-S	-	PGI-I ≥much better and PGI-S normal or mild
Nager-ValUE[72]	Basic office eval vs eval + UDS	No BD MESA stress> urge, +CST		≥70% reduction in UDI and PGI-I ≥ much better
Richter-TOMUS[24]	Retropubic vs transobturator MUS	No BD MESA stress> urge, +CST	Median IE/d = 2.7 10^{th}-90^{th} %=(0.7-6.7)	1) neg CST; 2) neg pad test; 3) no retreatment for SUI; 4) no UI on 3-day diary; 4) no self-reported SUI; 5) no self-reported retreatment of SUI
Barber[31]	Retropubic vs transobturator MUS	No BD criteria SUI on UDE No DO	Range of IE/d = (0-16.3)	Composite: 1) No UI of any type; 2) neg CST; 3) no retreatment for SUI; 4) no postop retention
SISTEr[17]	Burch versus fascial sling	No BD MESA stress> urge, +CST	Mean IE/d = 3.1-3.3	1) no self report UI; 2)pad test; 3) no IE on diary; 4) neg CST; 5) no re-treatment for UI

BD=bladder diary

Based on the above rationale, the ESTEEM inclusion/exclusion criteria are as follows:

4.2. Inclusion Criteria

1) Presence of both SUI and UUI on bladder diary; and ≥ 2 IEs/3 days
   a) ≥ 1 Stress IE/3 day diary
   b) ≥ 1 Urge IE/3 day diary
2) Reporting at least “moderate bother” from UUI item on UDI
   “Do you usually experience urine leakage associated with a feeling of urgency, that is a strong sensation of needing to go to the bathroom?”
3) Reporting at least “moderate bother“ from SUI item on UDI
   “Do you usually experience urine leakage related to coughing, sneezing, or laughing”
4) Diagnosis of SUI defined by a positive cough stress test (CST) or UDE within the past 18 months
5) Desires surgical treatment for SUI symptoms
6) Urinary symptoms ≥3 months
7) Subjects understand that BPTx is a treatment option for MUI outside of ESTEEM study protocol (see Section 5.3 for Rationale)
8) Urodynamics within past 18 months

4.3. Exclusion Criteria

1) Anterior or apical compartment prolapse at or beyond the hymen (≥0 on POPQ), regardless if patient is symptomatic
   a) Women with anterior or apical prolapse above the hymen (<0) who do not report vaginal bulge symptoms will be eligible
2) Planned concomitant surgery for anterior vaginal wall or apical prolapse > 0

a) Women undergoing only rectocele repair or other repair unrelated to anterior or apical compartment (ie: anal sphincter repair) are eligible

3) Women undergoing hysterectomy for any indication will be excluded
4) Active pelvic organ malignancy
5) Age <21 years
6) Pregnant or plans for future pregnancy in next 12 months, or within 12 months post-partum
7) Post-void residual >150 cc on 2 occasions within the past 6 months, or current catheter use
8) Participation in other trial that may influence results of this study
9) Unevaluated hematuria
10) Prior sling, synthetic mesh for prolapse, implanted nerve stimulator for incontinence
11) Spinal cord injury or advanced/severe neurologic conditions including Multiple Sclerosis, Parkinsons
12) Women on overactive bladder medication/therapy will be eligible after 3 week wash-out period
13) Non-ambulatory
14) History of serious adverse reaction to synthetic mesh
15) Not able to complete study assessments per clinician judgment, or not available for 12 month follow-up
16) Women who only report "other IE" on bladder diary, and do not report at minimum 1 stress and 1 urge IE/3 days
17) Diagnosis of and/or history of bladder pain or chronic pelvic pain
18) Women who had intravesical Botox injection within the past 12 months
19) Women who have undergone anterior or apical pelvic organ prolapse repair within the past 6 months

The team discussed the issue of whether bladder capacity should determine eligibility. Historically, some clinicians have used bladder capacity as a criteria for whether a woman with MUI is eligible for an anti-incontinence procedure for SUI, often excluding women with capacities <150-200 cc to avoid exacerbation of OAB symptoms. Upon review of the literature, there is very little evidence to support excluding women with a "small" bladder capacity, or to guide what volume defines a "small" capacity bladder. Gamble et al performed a retrospective study to evaluate predictors of persistent postoperative detrusor overactivity after a variety of slings.[79] They found that the mean maximum cystometric capacity was smaller in women with postoperative persistent DO compared to those with resolved DO. However, the mean capacity in women with persistent DO was 459 cc (SD 185) versus 539 cc (SD 176), which does not support the traditional teaching of avoiding slings in women with capacities less than 150-200 cc. Also, 37% of their study population included traditional bladder neck slings, which may be more obstructive than MUS. Finally, the proportion of women reporting UUI *symptoms* in this study was not different between women who had resolved versus persistent DO, highlighting the limitation of using UDE parameters to predict symptoms. Numerous other studies have failed to demonstrate any specific bladder capacity cutoff that is associated with better or worse outcomes or poses a safety issue for MUS.

Because there is a lack of evidence to support setting a minimum bladder capacity cutoff for this study, women determined to be eligible for a MUS based on their clinician's judgment will be eligible for ESTEEM, regardless of bladder capacity. One advantage of the ESTEEM design is that only women who have been offered a MUS by their clinician will be eligible. Therefore, if the provider determines that the patient is not clinically a candidate for a MUS, she will not be eligible. In addition, we will be excluding women with a history of painful bladder or chronic pelvic pain syndromes who often have "small" capacity bladders. To further contribute to the literature about this issue, we will collect data on both maximum cystometric capacity on UDE and functional bladder capacity based on voiding diaries and evaluate these variables as potential predictors of worsening OAB symptoms in our exploratory analyses.

4.4. Screening for Eligibility

It is anticipated that participants will come from PFDN Clinical Site practices. Women with MUI will be offered the full range of treatment options consistent with routine practice including expectant management, pelvic floor muscle therapy, behavioral therapy, medication and possibly surgery. Those

patients who are offered surgery by their physician and who elect to undergo MUS for SUI will be offered participation in ESTEEM. Subjects will be identified as ESTEEM candidates by their physician. Because in ESTEEM, women must have elected to undergo MUS, it does not compete with the current ongoing PFDN trial, ROSETTA, in which women desiring a MUS are actually excluded from that trial.

Subjects will be approached by study personnel consistent with local IRB requirements. Enrollment will occur after written and verbal consent. If the participant accepts participation in ESTEEM, the UDI will be administered to confirm at least moderate bother from both SUI and UUI and the coordinator will confirm documentation of SUI by either CST or UDE within the past 18 months, and UI symptoms for at least 3 months. The coordinator will also document that the patient understands that behavioral/pelvic floor therapy is a treatment option for MUI outside of ESTEEM (See section 5.3, "What is the best control group"). She will be instructed on how to complete the voiding diary.

To address the issue of overactive bladder medication use, these subjects will be required to have a washout of 3 weeks prior to completing the voiding diary. The anticholinergic with the longest half-life currently on the market is Vesicare with a half life of 45-68 hours. Therefore, by 1 week there should be negligible amounts in the bloodstream and by 2 weeks the drug would be completely out of the system. Therefore, 3 weeks should be adequate time for washout and this time period is consistent with prior PFDN studies (ABC trial[78]). In addition, because we are highly interested in what happens to OAB outcomes after MUS, subjects will need to remain off of overactive bladder medication until 3 months postoperative to allow accurate assessment of these symptoms postoperatively (See statistical analysis plan for details on why 3 months is adequate to allow analyses). Subjects who re-start overactive bladder medication postoperatively will be considered as having "additional treatment". Every effort will be made to schedule the patient's surgery within 3 months from enrollment (see Section 4.6, Appointment scheduling below).

4.5. Baseline Visit

At the baseline visit, the voiding diary will be reviewed to ensure that entries are clear and interpretable. If the first baseline voiding diary is not acceptable, the subject will be allowed one more attempt. If the second baseline voiding diary is not acceptable, the subject will not be eligible for the trial.

Once eligibility is confirmed, pre-treatment information will be obtained including:

- Demographics – age, race/ethnicity, education level
- Medical history – prior urinary incontinence procedures and treatments, prior pelvic surgeries, comorbidities, smoking, medications
- Physical exam – Body mass index, pelvic organ prolapse quantification (POPQ), PFM strength (Peritron and Brink measures)
- Questionnaires – self-administered

4.6. Appointment scheduling and randomization

Once patients are enrolled, surgery should be scheduled within 3 months from enrollment, and randomization should occur 7-35 days prior to the booked surgical date. This will allow enough time for those subjects randomized to the BPTx intervention to have their first preoperative visit scheduled, while minimizing withdrawal from the study due to unforeseen personal circumstances that may require a patient to cancel or change the date of their surgical procedure. Surgery should be performed 7-35 days after randomization and the surgery should be scheduled before randomization occurs. If a participant is randomized but does not undergo surgery, the planned surgery date will serve as Time 0 for calculating windows for follow up visits and phone calls. If surgery is rescheduled but does not occur, then the last planned date of surgery will be Time 0. If the participant decides against surgery but later changes her mind, the planned date of the surgery that did not occur will be Time 0, and the surgery that occurs after she changes her mind will be considered additional treatment.

Postoperatively, all subjects will return for visits at 2 and 8 weeks and 3, 6, and 12 months. Subjects randomized to BPTx will undergo undergo BPTx intervention sessions at 2 weeks preoperatively, and then postoperatively at 2, 4, 6, 8 weeks and 6 months. All subjects (intervention and control) will have visits with a masked assessor for PFM Peritron measurements at baseline, and 2 weeks and 8 weeks, and 12 months

postoperative. (See Assessment Table 11). All subjects will receive calls from research staff to determine AEs and additional treatment 4 and 6 weeks postoperative.

5. DESCRIPTION OF STUDY INTERVENTIONS

5.1. Midurethral sling procedure (both groups)

To address the potential issue that different sling or mesh types may result in different outcomes, MUS types will be standardized. All women (both groups) will receive a MUS which can include the TVT™ (mechanical cut mesh only, Gynecare, ETHICON Women's Health & Urology, Somerville, NJ), TVT-O™ (mechanical cut mesh only, Gynecare), or Monarc™ (American Medical Systems, Minnetonka, MN). In the TOMUS trial and Barber's equivalence trial, these approaches and devices demonstrated equivalence for improving objective success of SUI and were not significantly different for subjective success, persistent UUI or de novo UUI.[24, 31] The Gynecare "laser-cut" slings will not be allowed in this trial due to data from Moalli et al showing that the laser-cut meshes are "stiffer" (less deformation under an applied load), which theoretically may increase risk of mesh complications.[80] Although it is unclear how "laser-cut" meshes may affect clinical outcomes, these types of slings were not included in the TOMUS or Barber's equivalence trials resulting in less published, long-term outcome data. "Mini-sling" or "single-incision" slings will not be allowed. Key aspects of the procedure will be standardized across surgeons and sites.

5.1.a. Surgeon Certification- To address the issue of surgeon certification and to ensure standardized training of all surgeons, all "certified surgeons" will have performed a minimum of 20 midurethral sings of any type, including 5 of the specific MUS allowed in ESTEEM that the surgeon will be using in the study. The site PI must sign off that each participating surgeon has met the criteria.

5.1.b. Standardization of sling procedures:

Detailed standardization of the surgical procedure will be developed and will include the following key points:

1. The participating surgeon must be present and scrubbed for key portions of the procedure. Residents and fellows may participate in procedures as is standard for each Clinical Site
2. All subjects will receive preoperative intravenous antibiotic prophylaxis. The choice of antibiotic will be determined by each surgeon.
3. Deep vein thrombosis prophylaxis is required for all participants. The choice of prophylaxis will be determined by each surgeon.
4. Any concomitant native tissue procedures must be declared prior to randomization. Per exclusion criteria, women clinically requiring anterior vaginal prolapse or apical repairs are ineligible.
5. Tensioning of the sling will be performed in a fashion to ensure that it is a tension-free technique. This can include either by placing a blunt instrument between the sling and the urethra, or by folding a small knuckle of mesh in a Babcock clamp or similar method during tensioning.

5.1.c. Need for postoperative sling revision:

To address the issue of postoperative sling revision, the team developed a plan for several potential scenarios which may require the surgeon to revise the sling, detailed below. Women who undergo a sling revision will all be considered as having "additional treatment" in outcome analyses regardless of indication. Prior to sling revision, subjects will complete all outcome assessments including the primary outcome (UDI).

1. Urinary retention / incomplete bladder emptying (abnormal PVR) – An abnormal post-void residual is defined as PVR > 150 cc in this protocol (consistent with exclusion criteria). This is a known complication after MUS, and there is no evidence to support that this would be higher in women with MUI. Based on Barber's trial which included 70% women with MUI, the sling revision rate was 0-1%, which is also consistent with the TOMUS trial. For retention/incomplete emptying, the postoperative management and need for sling revision will be left up to the surgeon's clinical judgment.

2. Worsening OAB/lower urinary tract symptoms with a **normal** PVR – it is possible that some women may experience worsening OAB symptoms immediately postoperatively. It is unclear from the literature in which women such symptoms may be transient and ultimately resolve once postoperative recovery is complete, or in which women it will persist and/or worsen over time (an aim of ESTEEM). Therefore, for women with a normal PVR complaining of worsening OAB symptoms, sling revision will be deferred until 3 months postoperatively. This will provide important information about the natural course of these symptoms in the immediate postoperative period, and whether BPTx is effective for improving these symptoms early on. If after 3 months the patient desires sling revision due to worsening OAB symptoms, the surgeon can perform the procedure based on his/her clinical judgment. There is no evidence to support any potential harm by delaying sling revision in a woman with OAB symptoms and a normal PVR.

3. Persistent SUI symptoms – For women who have persistent SUI symptoms, sling revision/replacement can be performed after 3 months based on the surgeon's clinical judgment.

5.2. Background for BPTx intervention

To develop the most evidence-based, reproducible, standardized, and logical BPTx intervention protocol, the team reviewed the evidence and determined that bladder training/urge suppression techniques, pelvic floor muscle therapy, and weight loss have high level of evidence for treatment of urinary incontinence. Therefore, weight loss will be discussed with all women, and bladder training/urge suppression and pelvic muscle exercises will be incorporated into the ESTEEM BPTx intervention.

The summary of evidence for 5 key questions relevant to our intervention are summarized below:

5.2.1. What is the evidence for behavioral/lifestyle modification?

There are many components that can be defined as "behavioral" or "lifestyle" modification including caffeine intake, fluid intake, obesity, smoking, constipation and timed voiding. A summary of ICI evidence and recommendations is below:

Table 9. Summary of ICI recommendations

Modification	Level of evidence	Grade of recommendation	Recommendation
1. Caffeine intake	2	B	Caffeine reduction may improve incontinence
2. Fluid intake	3	B	Minor decreases by 25% may be recommended provided baseline consumption is not less than one liter a day
3. Weight loss	1	A	Morbidly and moderately obese women should consider weight loss to reduce UI
4. Smoking	3	None	More research
5. Constipation	3	None	More research
6. Timed voiding	3	C	Two-hour voiding intervals in women with mild UI and infrequent voiding patterns
7. Bladder training/urge suppression	1	A	Recommended for UI reduction

Caffeine: Aside from the volume of fluid ingested with these beverages, caffeine has been shown to have a diuretic effect and may increase OAB symptoms by increasing bladder pressure and bladder muscle excitability.[81-83] In addition, caffeine is a central nervous system stimulant and animal research has suggested that caffeine increases calcium release from smooth muscle leading to excitatory contraction of smooth muscle organs like the bladder.[84] Few well designed studies have addressed the impact of caffeine on bladder symptoms and those that have produced conflicting results, but there are some small studies suggesting decreasing caffeine may improve continence.[85]

Fluid intake: Excessive fluid intake can certainly increase urinary frequency and exacerbate OAB symptoms.[86] Interestingly, excessive restriction of fluid may also exacerbate symptoms due to poor

elimination of irritants from the bladder, decreasing the functional capacity of the bladder and increasing the risk of urinary tract infections.[87] Appropriate fluid intake should be balanced against activity level, climate, and fluid content of ingested foods. For most older adults, fluid intake should be approximately six 8-oz glasses per day.[88]

Weight loss: Obesity, defined as a body mass index greater than or equal to 30 kg/m2, was traditionally considered a risk factor for SUI only but more recently has been appreciated as a risk factor for OAB and UUI as well.[89, 90] Bump et al showed improvement in both SUI and UUI following surgical weight reduction in morbidly obese women.[91] But, even moderate weight loss can improve bladder symptoms in overweight women. A large randomized trial demonstrated that a structured weight loss intervention group resulting in a loss of 8% of body weight was associated with a clinically relevant reduction of 70% or more in the frequency of all IEs (P<.001), SUI (P=.009), and urge IEs (P=.04) compared to a control group which only lost 1.6% of body weight.[90, 92]

Smoking: Smoking, particularly nicotine, has been implicated as a risk factor for OAB and incontinence.[93, 94] Potential etiologies are increased intra-abdominal pressure from chronic cough and increased nicotine induced detrusor overactivity (as shown in cats).[95] Little clinical data is available assessing the impact of smoking cessation on bladder symptoms.

Constipation: Constipation is a common co-morbid complaint among patients with OAB and UI.[96-98] Although several studies in children document that constipation is linked to urinary tract symptoms including infection, enuresis, voiding problems and vesicoureteral reflux, the majority of studies in adults have identified an association but no clear causal link. While patients often report an exacerbation of bladder symptoms during times of constipation, few clinical studies exists to suggest resolving constipation improves OAB symptoms. Promotion of bowel regularity initially through natural methods including increasing dietary fiber, increasing water intake, physical activity and use of stool softeners is often recommended because it is low risk; however the evidence for its effect on improving OAB or UUI symptoms in the general adult population is limited.

Timed Voiding: Timed voiding or prompted voiding is a mechanism to theoretically increase bladder awareness, although firm evidence for its effectiveness for UI does not exist. Timed voiding involves a voiding schedule that starts with interval voiding on a fixed schedule regardless of the desire to go.[99] It involves patient cooperation, adequate mobility, and intact cognitive function. For some patients who delay urination, initially decreasing the voiding interval to every 30-90 minutes may be necessary to decrease incontinence episodes while urgency control strategies are being taught.[100] The maintenance of the timed voiding schedule during nighttime hours is determined by the patient's general sleep pattern (weather he/she awakens naturally to void), their motivation to stay dry (whether he/she sets an alarm to make sure to awaken), and the availability of help if needed.

5.2.2. What is the evidence for bladder training/urge suppression?

Bladder training through urgency control and suppression techniques has been an effective means of decreasing the intensity of urgency and incontinence in well motivated patients. Bladder training, sometimes referred to as bladder retraining, bladder reeducation or bladder drills, may be effective as the result of rewiring of complex circuitry between the bladder and the brain.[101] The training consists of three important components, (1) education about bladder function, dysfunction and urgency control strategies; (2) a timed voiding regimen that evolves to gradually increase the interval between voids; and (3) positive feedback and reinforcement by caregivers.[102, 103] Utilization of relaxation techniques including slow deep breathing and distraction techniques (mental concentration on other tasks) are most popular during urgency suppression.[100] Additional strategies including rapid contractions of the pelvic floor, or quick flicks (described below) and the use of self-motivating statements ("I can do it," "I am in control.") are also

popular.[104] Furthermore, patients are instructed to avoid running or walking fast to the bathroom as this may increase intra-abdominal pressure and promote leakage. Bladder training is used to slowly increase the interval between voids in attempts of reestablishing normal voiding intervals, break previously formed voiding habits, and diminishing urgency. In general, the voiding interval is increased on a weekly basis by approximately 15 to 30 minutes until a voiding interval of every 3-4 hours is reached.[104] A randomized controlled trial of 123 women with mixed urinary incontinence showed a 57% reduction in incontinence episodes and a 54% reduction in quantity of urine loss after implementation of a bladder training program.[105] **The ICI rated the level of evidence a 1 (based on scant evidence) and the grade of recommendation an A for the impact of bladder training on reduction in urinary incontinence.**

5.2.3. What is the evidence for Pelvic Floor Muscle Training (PFMT)?

A recent Cochrane review titled "Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in women" reported on 12 PFMT trials.[20] Of the 12 PFMT trials meeting their inclusion criteria, 3 provided no details of the PFMT method used. Per the review, most existing trials were at moderate to high risk of bias. There was considerable heterogeneity in interventions used, study populations and outcome measures. Women who did PFMT were more likely to report subjective improvement, cure and improvement in quality of life compared to those who did not. Women who did PFMT also reported fewer incontinence episodes per day, and less leakage on short office based pad test compared to those that did not. The authors concluded that PFMT should be considered first-line conservative treatment for SUI, UUI, or MUI. The effect seemed greatest in women with pure SUI and for programs that were at least 3 month in duration; however the authors recommend additional research to support these conclusions.

5.2.4. What is the best approach to PFMT for treatment of urinary incontinence?

The same Cochrane review[20] above also attempted to separate trials by those that increase: 1) Strength 2) Endurance, and/or 3) Coordination (for urgency suppression). Based on the descriptions of training, two trials had PFMT programs that clearly or predominantly targeted coordination[106] or strength training[107]. Miller and colleagues described a short (one week) program to improve coordination between a voluntary pelvic floor muscle contraction (VPFMC) and a rise in intra-abdominal pressure.[106] Bø et al recommended a program that comprised 8 to 12 high intensity (close to maximal) VPFMC, with six to eight second hold and three to four fast contractions added at the end of each hold, six second rest between contractions three times per day. Exercises were done in different body positions included lying, kneeling, sitting, standing; all with legs apart[107].

It was difficult to characterize the other PFMT programs, because they were either a mixed program (for example strength and endurance) or had not described a key training parameter (for example amount of voluntary effort per contraction). This Cochrane review highlighted some gaps and opportunities for future research in this field. Recommendations from the authors included research in which one arm would comprise a supervised PFMT program derived from sound exercise science, confirmation of a correct voluntary pelvic floor muscle contraction, and incorporate appropriate supervision and adherence measures to promote maintenance of knowledge acquisition. The choice of program would have to be set against the resource implications of intensively supervised individual programs and the opportunity cost this represents. The reporting of formal economic analysis would have to be added to the study. Careful clinical judgment would be needed about what sort of program could actually be applied in everyday practice and in different countries with their different health care delivery systems while still delivering an effective intervention.

A second relevant Cochrane review[108] titled "Comparisons of approaches to pelvic floor muscle training for urinary incontinence in women" also attempted to compare different approaches and/or components. These included: 1) differences in training supervision (amount, individual versus group), 2) approach (one versus another, the effect of an additional component) and 3) exercise training (type of contraction, frequency of training). Overall, the review concluded that there was insufficient evidence

regarding the best approach to PFMT; however, more frequent visits resulted in improved subjective outcomes (women receiving “regular” supervision were more likely to report improvement compared to little or no supervision).

5.3. What is the best “control” group for this study?

The team discussed whether women randomized to the control arm should receive baseline educational materials about behavioral and/or pelvic floor therapy. Educational materials that are routinely provided to women with MUI considering treatment options (before deciding on surgery) from each site were collected. The majority of sites (7/8) currently provide routine written material to patients on Kegel’s/pelvic floor muscle exercises. The majority also routinely provide information on: 1) urge suppression/kegel (7/8 sites); caffeine (7/8 sites); other bladder irritants (5/8 sites), and excessive fluid intake (6/8 sites). All sites were in agreement that these are routinely offered to women prior to moving forward with surgical intervention, although not all women choose to use these behavioral strategies.

The team considered the possibility of providing educational pamphlets to the control group; however, the ESTEEM population includes women who have already elected to proceed with surgery. In clinical practice, women who have decided on surgery have already been offered other conservative options and it is not routine practice to provide pamphlets again about other options at a preoperative visit. Therefore, this would not mirror what happens in the “real world”.

Because of these reasons, the team agreed the control group in ESTEEM should be MUS only. However to balance this, as part of our inclusion criteria, women will be reminded that BPTx is a treatment option for MUI (even outside of the study) to ensure they have been offered behavioral therapy and/or physical therapy outside of ESTEEM. (See Inclusion Criteria, Section 4.2). Along these lines, women who previously tried other behavioral or pelvic therapy will not be excluded. If the patient meets eligibility for ESTEEM, she would still have bothersome MUI by inclusion criteria. If the patient was not aware that behavioral/physical therapy was an option, she would be offered a referral at that point for which she can either accept (and cancel her surgery), or decline (and still be eligible for ESTEEM). The research coordinator will ask this screening question using similar wording that has been used in previous PFDN protocols.

Although routine educational pamphlets may be provided to subjects prior to their enrollment into ESTEEM per usual care at each site, once enrolled, no additional educational pamphlets may be provided to either control or intervention subjects outside of the protocol. The control group will complete bladder diaries and undergo PFM assessments at the same time intervals as the intervention group to control for any potential independent effects that bladder diary completion may have.

Rationale for including women who have previously tried behavioral and/or physical therapy:

There are many reasons to include women who have previously tried behavioral and/or pelvic floor physical therapy. First, women eligible for ESTEEM must have at least moderately bothersome MUI and desire surgery; therefore, even though these women have had treatments in the past, they did not improve enough to forego additional treatment. In addition, ESTEEM is evaluating the effect of ***combined surgical and BPTx treatment and not just BPTx alone.*** Therefore, women who have previously failed BPTx alone in the past may still significantly improve with combined surgical/BPTx treatment or surgery alone and there is no evidence to support their exclusion from this trial. *This is the most important reason why these women should be included.* Second, many women who have previously tried behavioral and/or physical therapy may have had a wide range of non-standardized interventions to varying degrees, durations, and with various components. Therefore, it is difficult to conclude that they may be at “higher risk” for failure, or that they will not benefit from the ESTEEM intervention. In ESTEEM, the BPTx protocol is based on existing evidence for specific BPTx components and the expertise of interventionists focused solely on improving MUI symptoms. This standardized protocol can potentially enhance the surgical effects for women with MUI. The protocol does provide the opportunity to identify risk factors for failure of a standardized BPTx intervention which will help build additional evidence for future trials.

5.4. Intervention - *See Appendix A for the full BPTx Intervention Protocol*
As stated above, for the intervention the team focused on evidence-based BPTx strategies. When evidence was lacking, the team made decisions based on the most logical and pragmatic rationale with a focus on developing a reproducible and standardized protocol.

For the purposes of this proposal "Behavioral training" (BPTx) will include:
1. Pelvic floor muscle training
2. Urge strategies defined in the field (included in intervention handout)
3. Stress strategies defined in the field (included in intervention handout)
4. Delayed voiding techniques (included in intervention handout)

The intervention will include 1 preoperative BPTx intervention visit and 5 post-operative intervention visits at 2, 4, 6, and 8 weeks and 6 months postoperative. Data from the ATLAS trial demonstrated that adherence with BPTx strategies decreased after 6 months, corresponding to a potential decrease in benefit.[110] Therefore, a 6 month BPTx intervention session is part of the intervention in ESTEEM. Participants randomized to intervention will receive BPTx implemented by an experienced registered nurse, nurse practitioner or physical therapist. Patients will be monitored using an adherence questionnaire.

The intervention will be **standardized** through the following mechanisms:
a. Certification of all interventionists through passing of e-learning modules and attendance and demonstration of hands-on skills at a 2-day, in-person interventionist training session
b. There will be an interventionist checklist to ensure the same components have been performed across subjects
c. There is a detailed protocol for the PFM exercise progression
d. There is a detailed protocol for "special circumstances" for when the standard PFM exercise progression protocol cannot be followed (ie: weak muscle) that the interventionist will be required to follow
e. Subject handouts will be developed for the 4 components (PFME, Urge strategies, stress strategies, and delayed voiding techniques) and the interventionists will be required to refer only to the handouts during the education component
f. All intervention sessions will be audiotaped and a subset will be audited by behavioral therapy experts to ensure adherence to protocol. Any protocol deviations will be addressed as necessary.
g. Phone calls between interventionists and behavioral experts will take place as needed to ensure adherence to protocol and address any issues and deviations.

Preliminary data from the OPTIMAL trial suggest that perioperative BPTx was not effective for improving urinary, prolapse, or colorectal symptoms at 6 months (unpublished data); *however, the study population in OPTIMAL is significantly different from ESTEEM*. Regarding baseline urinary symptoms, subjects in OPTIMAL were required to have an affirmative response to one SUI item only on the UDI whereas subjects in ESTEEM will be required to have an affirmative response to *both* the SUI and UUI items on the UDI and these symptoms must be *at least moderately bothersome*. Only 40% of women in the OPTIMAL trial reported mixed UI. In addition, all women in OPTIMAL had at least stage 2 symptomatic pelvic organ prolapse and all underwent apical prolapse suspension procedures as part of the intervention. Existing data support that urgency and urge incontinence symptoms may be associated with severe prolapse and surgical correction of prolapse may improve OAB symptoms.[111] In addition, there is solid evidence supporting that MUS is an effective treatment for SUI and therefore it is plausible that BPTx may not provide any additional effect in the OPTIMAL study population. However, there is minimal high-quality data regarding outcomes in MUI and there is evidence supporting that MUI is a risk factor for MUS failure. Finally, the BPTx component in OPTIMAL was developed as a prophylactic intervention, whereas the combined effect of MUS and BPTx is designed as a treatment intervention in ESTEEM. For all of these

reasons, we believe that the early findings from OPTIMAL do not directly address the aims proposed in ESTEEM and are not applicable to a MUI population.

The intervention in ESTEEM has been designed to focus on SUI and UUI symptoms and includes only components that address these 2 symptom constellations. Differences between the ATLAS, OPTIMAL and ESTEEM interventions are presented in Table 10.

Table 10. ATLAS and OPTIMAL behavioral therapy interventions and control compared to ESTEEM

Study	**ATLAS**	**OPTIMAL**	**ESTEEM**
Study design	Pessary vs BPTx vs both	Periop BPTx vs control + vaginal suspension	Combined periop BPTx+MUS vs control
Study population	-SUI or SUI predominant desiring non-surgical treatment	-Stage 2-4 prolapse with presence of SUI -All women underwent vaginal vault suspension -SUI defined as affirmative response to SUI item on UDI and objective confirmation	-No significant prolapse -No vaginal vault repair allowed -*Bothersome* mixed UI desiring midurethral sling (defined as at least *moderate bother* for both SUI and UUI items on UDI and confirmation on bladder diary)
Primary outcome definition	PGI-I and PFDI ≤ somewhat bother for SUI items	UDI (urinary outcome) -Urinary outcome powered to detect 11 point diff in UDI	UDI(total)-long form -Powered to detect 35 point diff in UDI(total), 15 point in UDI(irrit), and 8 points UDI(stress) scales.
Primary outcome time point	3 months	-Urinary-short term 6 months for urinary sxs -Prolapse-long term 2 years	12 months
# visits	4	5	6
Duration of active treatment	6 weeks	2 weeks preop to 3 months postop	2 weeks preop to 6 months postop
Interval between visits	Q2-3 weeks	Postop: (Q2-4 wks) 2, 4, 8 wks 3 months	Postop: 2, 4, 6, 8 weeks 6 months
Intervention components			
1. Bladder diary review	Yes	No	Yes
2. PFMT, technique eval	Yes	Yes	Yes
3. Standardized protocol for PFMT exercise progression	No	No	Yes
3. PFMT standardized "special circumstances"	No	No	Yes
3. SUI strategies	Yes	Yes	Yes
4. UUI strategies	Yes	Yes	Yes
5. Dysfx void strategies	No	Yes	Yes
6. Colorectal Sx strategies	No	Yes	No
7. Verbal/written home PFME Px	Yes	Yes	Yes
8. PFMT Adherence	Yes	Yes	Yes
9. Addressing other PFD Sx	No	Yes	No
10. Other written educational materials	SUI, UUI, PFME, Diary	SUI, UUI, PFME, Postop instructions, lifting, healthy bladder, healthy bowel	SUI, UUI, PFME, Diary

Control group			
	-Completed diaries same as intervention	-"Usual care" – routine periop teaching and standardized postop handouts -No diaries	-Will complete diaries at same time intervals as intervention group -Will have PFM measures same as intervention
Methods to standardize intervention			1. Interventionist checklist 2. Protocol for exercise progression 3. Interventionist protocol for "Special Circumstances" 4. Subject handouts that interventionist will review during education
Findings	-BPTx superior for SUI symptoms: 33% vs 49% for pessary vs BPTx (P=.006) -No difference in PGI-I -Higher satisfaction in BPTx: 63% vs 75% pessary vs BPTx (P=.02) -Combination better than pessary alone, but not BPTx	Preliminary: 6 months no diff in UDI score between groups	N/A

5.5. Patient management and follow-up

5.5.1. Baseline Procedures

In addition to information collected to determine eligibility and standardized questionnaires, the following information will be obtained for all randomized patients by chart review or patient report:

a. Demographic information: age, race, ethnicity, insurance status, education

b. Medical history: vaginal parity, comorbidities, height, weight, prior pelvic surgeries, medications, estrogen status, previous treatments for pelvic floor disorders

c. Social history: tobacco use

d. Pelvic, rectal exam, neurological examination, POP-Q, PFM strength (collectively will include Brink and Peritron measurement), post-void residual, urinary stress test

e. Standardized urodynamic evaluation (UDE) will be performed preoperatively – There continues to be controversy regarding the usefulness of UDE for preoperative evaluation of SUI. However, it is often recommended in women who have a "mixed" UI picture and there are no definitive studies to determine if UDE parameters may be helpful in predicting outcomes after surgery in women with MUI. Therefore, the protocol team agreed that patients in ESTEEM should undergo UDE testing, primarily to allow evaluation of variables that may predict clinical outcome. Because eligibility includes women electing surgery, and because this is a complex population, many patients may already have UDE results prior to enrollment. For those women who have not, they will undergo testing preoperatively, although there are no specific UDE parameters that determine eligibility for this trial. Urodynamic tests performed within the past 18 months will be allowed.

f. Patient-reported outcomes and questionnaires – includes UDI, IIQ, EQ5D, Adaptation questionnaire, PGI-I, PGI-S, OAB-q, OAB-sat-q, PISQ,

5.5.2. Postoperative visits and procedures

Patients will undergo clinical and PRO assessments at 3 months, 6 months, and 12 months postoperatively. (See Table 11 above). The primary outcome will be at 12 months. Additional treatment for patients with persistent OAB symptoms should not be offered in the first 3 months, given this time period may still represent recovery from acute events related to surgery. Patients requesting additional treatment in the first 12 months will be considered treatment failures, and will complete PRO assessments at the time of initiation of additional treatment. Any additional long term follow up beyond 12 months, consideration would need to be given to the natural history of progression and remission of OAB.[112, 113]

Table 11. Timeline of visits, events, and data collection

	Baseline	Random-ization visit (T1-5 wks preop)	Preop BTPx visit (range 1-5 wks preop)	Surg MUS (T0)	Call (2-4d post-op)	2 wk post-Clinic	4 & 6 wks post	8 wk post-	3 mo post-Clinic and QoL	6 mo post-Clinic and QoL	12 mo post-Clinic and QoL
Estimated duration of clinic and/or BPTx visit for each group		Both: 1.5-2hr	Control: N/A Interv: 1.5hr		Contr: N/A Interv: 15 min	Control: 1.5hr Interv: 2.5hr	Control: N/A Interv: 1hr	Control: 1hr Interv: 2hr	Both: 1.5hr	Control: 1.5hr Interv: 2.5 hr	Both: 1.5-2hr
All subjects											
Consent	X										
Coordinator visit	X	X				X		X	X	X	X
Masked clinical staff visit (for PFM measures)		X				X		X			X
Hx/PE (update)						X		X	X	X	X
Medication audit	X					X		X	X	X	X
UDE	X										
UDI (inclusion and primary outcome)	X								X	X	X
Other PRO questionnaires		X							X	X	X
Voiding diary	X*					X*	X	X*		X*	X*
PFM measures		X				X		X			X
Additional treatment**						X	X (both groups by phone)	X	X	X	X
Adverse events				X		X	X (both groups by phone)	X	X	X	X
Voiding function (PVR)	X					X					
Subjects randomized to intervention only											
BPTx visit			X			X	X	X		X	
BPTx self-efficacy questionnaire		X								X	X
BPTx Adherence / Barrier questionnaire						X	X	X		X	X

* Data will be keyed into iMedidata

**For subjects who request/initiate additional treatment, all outcome measures will be completed prior to initiation of additional treatment.

6. Statistical considerations

6.1. Sample size estimates

6.1.1. Primary aim and secondary aims:

This study is designed to compare the efficacy of MUS+BPTx versus MUS alone on improving MUI symptom outcomes. Because OAB and SUI symptoms are highly important secondary outcomes as stated previously, we felt strongly that our sample size should provide adequate power to detect differences for the separate UDI-irritative and UDI-stress subscales in addition to the UDI total score. Our initial sample size estimates were based on published MIDs for the UDI total score and subscales; however, we recognize that the populations on which those MIDs were based might differ from the target population for ESTEEM. A secondary aim of ESTEEM is to estimate the MIDs for UDI scores in this study population, and it is possible that the MIDs in this population could be smaller than values previously published, particularly for the UDI total score. Thus, our goal was to power the study to detect a statistically significant difference between groups in change from baseline in UDI total score at 1 year that was smaller than the published MID but still in a range of what we think may be a clinically important difference in our population.

Sample size estimates are based on simulations using analysis methods accounting for both the rate of additional treatment in the two groups as well as UDI total score or subscore values over the 12 month follow up period (refer to the statistical analysis plan for details). We assumed that 30% of women in the MUS only group and 20% of women in the MUS+BPTx group would request additional treatment. In TOMUS, 10-12% of women who had baseline MUI had persistent UUI postoperatively based on MESA responses and/or initiation of anticholinergic treatment.[24] In Barber's TVT vs TOT equivalence trial, 70% reported baseline MUI and postoperatively, 30% of all women reported bothersome UUI with 16% of subjects on anticholinergic treatment postoperatively.[31] In Abdel-Fattah's transobturator MUS trial, 25% reported worsening OAB and almost all of these women were on anticholinergic treatment postoperatively.[34] In Palva's TVT vs TVT-O trial, 174 women reported preoperative UUI and of these, 7 women (4%) had tried anticholinergics postoperatively after 3 years. Therefore, based on existing MUS trials, the rate of additional treatment for OAB ranges from 4-25%, supporting our conservative assumption that 30% of women will request additional treatment in the MUS only group.

i. Primary outcome: MUI symptoms = UDI-total score

The MID for the UDI-total score published by Dyer et al is estimated to be 35 points.[50] Assuming a two-sided alpha of .05, SD of 50.4, and true difference in mean change from baseline in UDI-total scores at 1 year between treatment groups of 35, 75 women per group would provide 90% power to detect a statistically significant difference between groups.

ii. Secondary outcome: OAB symptoms= UDI-irritative subscale: For the UDI-irritative subscale, the published MID estimate is 15 points.[50] Assuming a two-sided alpha of 0.05, SD of 25.6, and true difference in mean change from baseline in UDI-irritative scores at 1 year between treatment groups of 15, 92 women per group would provide 90% power.

iii. Secondary outcome: SUI symptoms = UDI-stress subscale: For the UDI-stress subscale, the published MID is 8 points.[51] Assuming a two-sided alpha of 0.05, SD of 21.5, and true difference in mean change from baseline in UDI-stress scores at 1 year between treatment groups of 8, 200 women per group would provide 90% power to detect a statistically significant difference between groups.

Using 200 per group as our base estimate and adjusting for 15% dropout post-operatively results in a total sample size of 472 randomized to treatment.

Additionally, this sample size will provide approximately 90% power to detect a difference as small as 19 between treatment groups for the UDI-total score, and a difference as small as 16.5 points with 80% power.

6.1.2. Potential limitations of the UDI and primary outcome:

One potential limitation of using change from baseline score as the primary outcome is that point estimates of the difference in means between 2 groups may mask important changes for individual patients that are meaningful. However, this would also be the case if we dichotomized the outcome into "success" versus "failure". In addition, the published MID used for our primary outcome is derived from the BE-DRI population, an urge-predominant MUI population and MID estimates can vary depending on the study population. The published estimate for UDI-total MID for the BE-DRI urge-predominant population is 35 points based on Dyer et al[50] whereas Barber et al found the MID for pure stress/stress-predominant population to be 11 points in the ATLAS population.[51] One advantage of the BE-DRI population is that 96% had MUI, which is more similar to the anticipated ESTEEM population. It is possible that women with UUI require larger improvements compared to pure/SUI predominant women to be meaningful. This is consistent with many previous studies showing that women with UUI experience worse impact and bother than SUI patients and that the UUI component drives patient perception of severity and satisfaction after treatment.

Although we do not definitively know whether 35 is an accurate MID for determining success or failure in this study population, we consider this MID estimate from BE-DRI to be the published MID that is most applicable to our target population. In addition, because our total sample size is 400 subjects (before adjustment for drop out), our study will have 90% power to detect a statistically significant difference in UDI-total scores if the true difference is as small as 19 points between groups and 80% power to detect a difference if the true difference is as small as 16.5 points. This difference is smaller than the conservative, distribution-based MID estimate of -24.8 based on the BE-DRI population. Thus, the planned sample size will allow for analyses to assess whether the true MID in this population is smaller than 35.

Finally, the UDI total score includes 3 subscales: stress, irritative and obstructive. Therefore, our primary outcome will include a total score combining all 3 of these subscales. We believe the inclusion of the obstructive subscale is appropriate for the following reasons:

1. Although obstructive symptoms related to prolapse are not a focus of ESTEEM, some items in this subscale may still be relevant to the MUI population (ie: "general urine leakage not related to urge or activity"; symptoms of "difficulty emptying"; and "incomplete emptying").
2. Because women with symptomatic prolapse will be excluded in both groups, it is unlikely that the inclusion of this subscale in the primary outcome will lead to bias.
3. The published MID for the UDI in the BE-DRI population also includes all 3 subscales for an urge-predominant MUI population.[50]

6.1.3. Management of women who drop out prior to receiving MUS

It is possible that some women in both groups may cancel their surgical MUS procedure due to personal reasons, or other. It is also possible that women randomized to BPTx may cancel their surgical procedure if they receive preoperative BPTx treatment and experience improvement. These women will still be included from an ITT perspective.

6.2. Statistical analysis plan

6.2.1. Primary aim

The mean change from baseline in UDI scores will be compared between groups at 1 year. As explained previously, participants will be permitted to seek additional treatment for SUI and/or OAB after 3 months following MUS. Because such treatment is expected to impact the participant's UDI score at 1 year, we will use an analysis method that accounts for the impact of additional treatment. Specifically, a general linear mixed model will be constructed to model change from baseline in UDI scores using scores recorded

at time points up to 1 year following MUS. For participants who request additional treatment, only UDI measurements up to the time of additional treatment will be included in the model, and measurements taken between additional treatment and 1 year will be considered missing for the purpose of the primary analysis. The model will include fixed effects for treatment group, time, request for additional treatment, and interactions between those variables. It will also be adjusted for the design effects of stratification by center and by baseline urge IE group. Thus, the models will allow for different trajectories of change for women who are or are not randomized to BPTx and for those who do or do not request additional treatment. A statistical test based on the model will be conducted to assess whether mean changes from baseline in UDI scores at 1 year are significantly different between the two treatment groups, accounting for the percent of women in each group who request additional treatment. Sensitivity analysis will be conducted to test the robustness of test results to model specifications.

We will report whether change in total UDI score between baseline and one year is significantly different in the two groups. If the difference is statistically significant, the potential clinical significance of the difference will be discussed. We recognize that our sample size would allow us to find a difference between groups that is statistically significant yet smaller than published MIDs for total UDI score for women with MUI. However, published MIDs were calculated based on populations that may be somewhat different from the one targeted for enrollment in ESTEEM, and a secondary aim of ESTEEM is to explore whether the true MID in this population differs from previously published values.

6.2.2. Secondary aims

The mean change from baseline in UDI-irritative and UDI-stress scores at 1 year will be compared between groups using the same analysis methods described for the primary outcome. If the difference is statistically significant, the potential clinical significance of the difference will be discussed. Additional analyses will be conducted to determine whether the MIDs in this MUI population differ from previously published MIDs.

6.2.3. Exploratory aims

a. Other UUI/OAB outcomes

Bladder diary

We will compare change in number of urge IEs and urgency-episodes and nocturia episodes between groups from baseline to 6 and 12 months. Of note, not all four symptoms of OAB (frequency, urgency, nocturia, and UUI) are required to be present at baseline for eligibility into this trial (only UUI required). Changes from baseline in bladder diary outcomes will be calculated and analyzed using the methods described for the analysis of the primary outcome.

For urinary frequency, women reporting on average >8 voids/24 hours at baseline will be considered symptomatic, and normalization of voiding frequency will be defined as $\leq$ 8 voids/24 hours at 1 year. A 50% improvement will be defined as a reduction by half in the number of voids that patients had at baseline. The number of women who had normalization of voiding frequency and 50% improvement will be compared between groups separately and collectively. We will also assess the proportion of women who had worsening of urinary frequency (includes women who developed de novo frequency and those who worsened). These dichotomous outcomes will be analyzed using logistic regression, controlling for the design effects of stratification by center and by baseline urge IE group. To assess the impact of additional treatment prior to 1 year, sensitivity analyses will be conducted in which women who request additional treatment will be assigned the less-favorable outcome.

OAB-SAT-q and OAB-q

For these scales and associated subscales, differences from baseline will be calculated for the OAB-q, and methods described for analysis of the primary outcome will be used to test for differences between treatment groups at 12 months. For the OAB-SAT-q, differences in post-treatment scores will be compared between groups.

b. Differences in time to failure between groups

Although our primary outcome is at 12 months, the team was interested in whether perioperative BPTx may be associated with a delayed *time to failure* compared to Control. In other words, is BPTx associated with a significant effect, but the effect is not sustained at the 12 month time point? For example, if BPTx could delay the need for anti-muscarinics for up to 9 months, this would be relevant information for counseling women and perhaps clinically recommending perioperative BPTx. As described previously, failure will be defined as initiation of any additional treatment for either SUI or UUI/OAB symptoms.

A class of survival model which can account for interval censoring (outcomes measured at pre-planned time points as opposed to continuously over time) will be used to determine if combined MUS+BPTx is associated with a decrease time to failure compared to MUS alone between 3-12 months. Depending on the distribution of the observed data, an accelerated failure time frailty model or a Bayesian survival model may be used. The model will be adjusted for the design effects of stratification by center and by baseline urge IE group.

c. Predictors of treatment success and failure

Regression models will be created to identify predictors of change from baseline to 1 year for UDI total score and stress and irritative subscale scores. Participants who request additional treatment prior to 1 year will not be included in the predictive models. Potential predictors will include age, diary parameters such as number of UUI episodes/3 days, functional bladder capacity, bother severity at baseline. The relationship between potential predictors and outcomes will be explored in models that include one predictor plus stratification factors (center and baseline urge IE group). Predictive models will be constructed using backward selection of predictors. The impact of collinearity between predictors will be assessed and the final model modified as necessary.

d. Quality of life/global impression

For these scales and associated subscales, differences from baseline will be calculated and methods described for analysis of the primary outcome will be used to test for differences between treatment groups from baseline and 6 and 12 months.

e. To describe safety and initiation of additional treatment for worsening OAB and/or persistent SUI

We will describe rates of sling revision due to worsening OAB symptoms and rates of additional treatment.

f. To determine MIDs and clinically meaningful MUI definitions that predict clinical outcomes.

We will explore potential MIDs for UDI total score and stress and irritative subscores for this MUI population. MIDs will be calculated using anchor- and distribution-based approaches. Potential anchors include global impression of change, incontinence episodes from the bladder diary, and request for additional treatment.

We will attempt to create threshold definitions, based on baseline measures of the UDI, IIQ, OAB-q, UDE, and baseline bladder diary parameters in isolation and in combination, that are predictive of clinical success at 1 year. Definitions of success will be based on a change from baseline in total UDI score, UDI-irritative score or UDI-stress score at least as large as the MID for this MUI population.

g. To compare pelvic floor muscle strength changes between women randomized to combined MUS+BPTx versus MUS alone, to estimate associations between pelvic floor muscle strength improvement and UI symptoms, and to identify predictors of unsuccessful pelvic floor muscle strengthening and urge suppression and their effects on urinary outcomes in women randomized to BPTx

As mentioned above, all women will undergo PFM strength measurements using the Peritron device by masked coordinators at baseline, postoperative at 2 weeks, 8 weeks (end of intervention), and 12 months (primary endpoint). The difference in the maximum pelvic floor muscle contraction pressure (maximum amplitude) will be compared between the BPTx and the control groups. A table of comparative

studies using the Peritron device to measure PFM strength changes with PFM therapy is provided in Table 12 below.

Based on the existing comparative studies using the Peritron, continent women have a maximum amplitude PFM contraction between 36-45 cm H2O. Incontinent women have significantly lower maximum contractions, ranging from 15.5 to 26.5 cm H2O, with most studies showing a maximum contraction of 25 cm H2O. In these studies, incontinent women can improve their maximum contraction pressure up to 34-41 cm H2O with PFM training, which is comparable to continent women. In addition, these studies report women experience significant improvement in UI symptoms, although there is limited information on the direct specific relationship between PFM strength changes and UI symptom changes.

Assuming that women in ESTEEM will have a mean baseline PFM maximum contraction amplitude of 25 cm H2O, and that women randomized to control will not demonstrate significant improvement postoperatively (no change from mean maximum amplitude of 25 cm H2O (SD 13), and that women randomized to BPTx will demonstrate improvement to 35 (SD 13) to 40 (SD 16) cm H2O at 6-12 months, the power to detect a difference between the groups with the current ESTEEM sample size of 400 women would be greater than 0.99. Also, the difference from 25 (SD 13) that we could detect with 80% power is 3.66 cm H2O between groups and with 90% power we could detect a difference as small as 4.23 cm H2O.

For analyses, we will compare the mean change from baseline in PFM maximum contraction strength between the BPTx and control groups at 8 weeks and at 12 months. General linear mixed modeling will be used, controlling for stratification factors and time (8 weeks and 12 months). We will test whether there is significant interaction between treatment group and time. Because additional treatment is not expected to impact this outcome, it will be ignored for the purpose of this analysis. We will estimate the correlation between PFM strength and UI symptoms at baseline and at 12 months. Using regression models, we will also explore potential predictors of unsuccessful pelvic floor muscle strengthening and urge suppression and their effects on urinary outcomes. We will assess the effect of self-efficacy[114, 115], adherence, and barriers to performing pelvic muscle contractions and behavioral therapy.

Table 12. Comparative studies using Peritron measurement of pelvic floor muscle strength

Author	**Pop**	**Study details**	**Baseline PFM strength (SD)***	**Post-treatment PFM strength (SD)**	**P-value**	**Notes**
Rett 2007[116]	SUI	N=26 Single cohort PFME with sEMG biofeedback	Max amp= 24.5 (16)	After 12 sessions: 40.0 (17)	<.0001	No info on "subjective improvement" and PFM strength Overall cohort: Obj cure = 61.5% Subj cure = 23% Subj "almost cure"=65.4%
Gameiro 2010[117]	Any UI	N=103 RCT G1 =vag cones G2 =APFMT	Max amp: G1=24.4 (12.5) G2=20.0 (12.9)	6 mos: G1=40.8 (15.73) G2=35.16 (11.05) 12 mos: G1=34.98 (13.2) G2=34.12 (9.84)	P<.05 for both	*No specific correlation btwn subjective "cure" and PFM strength; however: a. Reduction of pads better for G1 b. # micturitions, nocturia, UI episodes, urgency, pad test ND btwn grps
Amaro 2005[118]	SUI	N=101 Comparative cohort G1 = SUI	Max amp: G1=26.1 (1.15) G2=38.4 (1.33)		P<.001 for all 3 baseline	

		G2=controls	Mean amp: G1=15.4 (.62) G2=28.1 (1.22) Duration (s) G1=8.9 (.17) G2=11.8 (.96)		comparisons	
Gilling 2009[119]	SUI	N=70 RCT G1=Estim G2=Sham	Max amp: G1=17.3 (1.8) G2=15.5 (1.9)	8 wks: G1=19.2 (2) G2=15.1 (1.9)	ND	Subgroup findings: "Patients with poor initial PFM ctx by perinometer randomized to Estim had better UI outcomes than sham" but cannot tease out their PFM scores
Hung 2011[120]	Any UI	N=23 PMT Prospective cohort, pre- post- PFM program 65% SUI 35% MUI 39% UUI	Max amp: 27 (15.0)	4 mos: 41 (24.9)	<.001	
Gamerio[121]	SUI vs UUI	N=51 Cross-sectional G1=SUI (N=22) G2=UUI (N=29)	Max amp: G1=26.5 (3) G2=21.7 (.79) Mean peak G1=16.56 (1.19) G2=13.72 (0.56) Duration: G1=9.54 (0.18) G2=8.43 (.42)		P<.001 for all 3 baseline comparisons	Unclear clinical meaning

h. To determine the cost effectiveness of combined midurethral sling (MUS) and peri-operative behavioral/pelvic floor therapy (BPTx) compared to MUS alone on successful treatment of MUI symptoms

Differential mean costs and differential mean QALYs between the two treatment groups will be estimated using multiple regression analysis. Specifically, a generalized linear model with appropriate link function (e.g., log-link) and response probability distribution (e.g., gamma distribution) will be used to analyze costs due to the potential skewness and heteroscedasticity of medical expenditure data, while an ordinary least squares regression will be used for analyzing QALY data. The models will account for treatment group, study site and stratification factors, as well as other characteristics of the subjects that are found to differ significantly between the groups. When estimating QALYs, we will also adjust for subjects' baseline utility scores to account for potential imbalance in baseline utility between the two treatment groups.[122]

We will calculate the incremental cost-effectiveness ratio (ICER), which is the differential mean costs divided by the differential mean QALYs between the two groups, to assess the additional costs associated with each additional QALY gained. Our base case analysis will be conducted based on subjects with complete data. Sensitivity analysis will be conducted to include subjects with incomplete data using the multiple imputation method. Non-parametric bootstrapping resampling technique will be used to derive the 95% confidence interval for the ICER.[118, 123] In addition, cost-effectiveness acceptability curve (CEAC) will be generated to illustrate the likelihood that one treatment is more cost-effective than the other with various ceiling cost-effectiveness ratios.

In the case that a statistically significant difference in changes in utilities (as measured by EQ-5D) between the treatment groups is not detected, we plan to conduct supplemental analyses using alternative

outcome measures, such as incremental cost per treatment success, incremental cost per HRQOL, or incremental cost per satisfaction.

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

6.3. Interim data monitoring

Safety outcomes will be assessed at each DSMB meeting. This will include the need for sling revision due to worsening OAB symptoms. Rates of sling revision and other safety outcomes will be compared between treatment groups using Fisher's exact tests and provided to the DSMB. There is no established guidance regarding what sling revision rate is "appropriate" for worsening OAB symptoms in this population: this is one of the exploratory aims of this study.

Since we expect to enroll ESTEEM within 2 years, and since the primary outcome is attained at 12 months following surgery, we propose that no interim analyses of outcomes will be performed. Thus, reports to the DSMB will not include outcome data until primary outcomes have been attained for all participants. At each meeting, the DSMB will be presented with information about enrollment and outcome data attainment (for example, the percent of expected clinic visits that have been completed) to allow them to determine that the study is making reasonable progress.

7. Ethical Concerns/Safety

7.1. Ethical Concerns

As discussed in the background section, current clinical practice varies with respect to treatment of MUI and likely reflects training and experiential bias. Although treatment with behavioral modifications and Kegel exercises have been described as effective first line treatments for mild stress, urge, and mixed urinary incontinence, many patients go on to request further therapy for their condition. For moderate symptoms of SUI or UUI additional therapeutic options are generally offered based on treatment paradigms geared toward each of these conditions. When patients have MUI, clinicians must decide which component (the SUI or the UUI) should be addressed first. There is very little evidence to support a defined treatment strategy in this patient population and most recommendations are based on expert opinion. The only way to test the superiority of one approach over another is in the setting of a randomized clinical trial. We have carefully designed this trial to balance the risks and benefits to subjects. All patients in this study have elected to undergo surgery for SUI. Therefore, they will have already been offered more conservative therapies. We will be assured that women will have either previously tried behavioral or pelvic floor therapy or at least have been offered this treatment because it is an inclusion criteria. In addition all patients will be treated with a midurethral sling and half the patients will be randomized to perioperative supervised BPTx. The potential benefits of the BPTx intervention are improvement of MUI symptoms while the risks are very small. The benefits of BPTx in SUI and UUI alone and MUI have been documented as has the benefit of MUS for patients with SUI. Several studies have also documented an improvement in OAB and MUI symptoms following sling. The added benefit of a combined approach of sling plus BPTx in patients with MUI has not been defined and is the subject of this RCT. Any subject can request additional treatment after 3 months postoperative.

7.2. Informed Consent

Subjects will be clinically examined as part of screening and to ensure eligibility for the study. Those subjects who are candidates for and agree to undergo sling surgery and behavioral treatment for MUI will

be approached for enrollment into the trial. Clinical and research staff will describe the study in detail and answer any questions the subject may have. Written informed consent for trial participation will be obtained at that time. A common template for the research informed consent form will be used by all of the clinical sites, modifying the content or format as necessary to meet the requirements of their respective institutional human subjects committees. This protocol must be approved by the IRBs at the clinical sites and DCC before study implementation.

7.3. Data Safety Monitoring Board

The National Institutes of Health has set up a Data Safety Monitoring Board (DSMB) to oversee all PFDN studies, including this study. Members of the DSMB are independent of the study investigators and represent Urology, Urogynecology and Biostatistics, as well as having a lay member. The DSMB meets every 3 months, or more frequently if requested by the Chair, either in person or by teleconference. This protocol has been approved by the DSMB prior to implementation. Safety outcomes will be assessed in a descriptive manner at each DSMB meeting without formal statistical tests. This will include the need for sling revision due to worsening OAB symptoms. There is no established stopping rule to guide what sling revision rate is “appropriate” for worsening OAB symptoms in this population.

7.4. Reporting of serious adverse events

Each clinical investigator is responsible for reporting serious adverse events (SAEs) to the IRB per their IRB guidelines at their institution, and to the DCC. The DCC Safety Specialist reviews and summarizes the SAE per DCC SAE reporting procedures for the PFDN.

7.5. Adverse events

Adverse events are defined as untoward medical events that are temporally-related to participation in a clinical study, regardless of whether they are causally-related to the study. Adverse events will be collected during the course of this study and reported to the DSMB as described above.

Sling surgery is a commonly performed operation for the treatment of SUI and MUI. Like all surgical interventions it has the risk of bleeding, infection, and injury to surrounding structures. In addition, the sling procedure utilizes polypropylene mesh which can introduce additional risk of mesh complication. These include vaginal mesh extrusion, mesh infection, and bladder or urethral mesh erosion. Complications specific to sling placement include bladder perforation, retropubic hematoma, obturator nerve or vessel injury, groin pain, worsening incontinence, and worsening OAB. The FDA has recently issued guidelines on the use of surgical mesh and has recommended it only be used by trained surgeons. All surgeons participating in this study will be specifically trained to use surgical mesh.

8. Feasibility

The proposed study population has already chosen to undergo surgical treatment and the BPTx intervention is low risk. We have taken care to have comparable arms in a clinical efficacy trial design with inclusion criteria that are not overly-strict; therefore, we do not anticipate particular difficulty in recruitment of MUI patients as encountered in MIMOSA.[37] If needed in the postoperative period, medical therapy will not be withheld after 3 months postoperative. Women reporting bothersome OAB symptoms for which they desire additional treatment will be presented their options (additional BPTx and/or FDA approved OAB pharmacologic therapy, or other procedures or surgeries), and additional treatment will be offered. Request for additional treatment for either OAB or SUI postoperatively will be driven by patient preference and clinician judgment in both groups.

9. References

1. Katsumi HK, Rutman MP. Can we predict if overactive bladder symptoms will resolve after sling surgery in women with mixed urinary incontinence? Curr Urol Rep;11:328-37.
2. Shumaker SA, Wyman JF, Uebersax JS, McClish D, Fantl JA. Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program in Women (CPW) Research Group. Qual Life Res 1994;3:291-306.
3. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. Am J Obstet Gynecol 2003;189:98-101.
4. Melville JL, Katon W, Delaney K, Newton K. Urinary incontinence in US women: a population-based study. Arch Intern Med 2005;165:537-42.
5. Karram MM, Bhatia NN. Management of coexistent stress and urge urinary incontinence. Obstet Gynecol 1989;73:4-7.
6. Stewart WF, Van Rooyen JB, Cundiff GW, et al. Prevalence and burden of overactive bladder in the United States. World J Urol 2003;20:327-36.
7. Monz B, Chartier-Kastler E, Hampel C, et al. Patient characteristics associated with quality of life in European women seeking treatment for urinary incontinence: results from PURE. Eur Urol 2007;51:1073-81; discussion 81-2.
8. Dooley Y, Lowenstein L, Kenton K, FitzGerald M, Brubaker L. Mixed incontinence is more bothersome than pure incontinence subtypes. Int Urogynecol J Pelvic Floor Dysfunct 2008;19:1359-62.
9. Subak LL, Brubaker L, Chai TC, et al. High costs of urinary incontinence among women electing surgery to treat stress incontinence. Obstet Gynecol 2008;111:899-907.
10. Brubaker L, Stoddard A, Richter H, et al. Mixed incontinence: comparing definitions in women having stress incontinence surgery. Neurourol Urodyn 2009;28:268-73.
11. Dmochowski R, Staskin D. Mixed incontinence: definitions, outcomes, and interventions. Curr Opin Urol 2005;15:374-9.
12. Petros PE. Mixed urinary incontinence--time to uncouple urgency from stress? Int Urogynecol J;22:919-21.
13. Khullar V, Cardozo L, Dmochowski R. Mixed incontinence: current evidence and future perspectives. Neurourol Urodyn;29:618-22.
14. Tyagi R, Staskin DR. Mixed incontinence: the misclassification of patients and limitations of clinical trials. Curr Urol Rep 2005;6:424-8.
15. Murray S, Lemack GE. Overactive bladder and mixed incontinence. Curr Urol Rep;11:385-92.
16. Haylen BT, de Ridder D, Freeman RM, et al. An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction. Neurourol Urodyn;29:4-20.
17. Albo ME, Richter HE, Brubaker L, et al. Burch colposuspension versus fascial sling to reduce urinary stress incontinence. N Engl J Med 2007;356:2143-55.
18. Herzog AR, Diokno AC, Brown MB, Normolle DP, Brock BM. Two-year incidence, remission, and change patterns of urinary incontinence in noninstitutionalized older adults. J Gerontol 1990;45:M67-74.
19. Brubaker L, Lukacz ES, Burgio K, et al. Mixed incontinence: comparing definitions in non-surgical patients. Neurourol Urodyn;30:47-51.
20. Dumoulin C, Hay-Smith J. Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in women. Cochrane Database Syst Rev:CD005654.
21. Khullar V, Hill S, Laval KU, Schiotz HA, Jonas U, Versi E. Treatment of urge-predominant mixed urinary incontinence with tolterodine extended release: a randomized, placebo-controlled trial. Urology 2004;64:269-74; discussion 74-5.

22. Sexton CC, Notte SM, Maroulis C, et al. Persistence and adherence in the treatment of overactive bladder syndrome with anticholinergic therapy: a systematic review of the literature. Int J Clin Pract;65:567-85.
23. Burgio KL, Kraus SR, Menefee S, et al. Behavioral therapy to enable women with urge incontinence to discontinue drug treatment: a randomized trial. Ann Intern Med 2008;149:161-9.
24. Richter HE, Albo ME, Zyczynski HM, et al. Retropubic versus transobturator midurethral slings for stress incontinence. N Engl J Med;362:2066-76.
25. Jain P, Jirschele K, Botros SM, Latthe PM. Effectiveness of midurethral slings in mixed urinary incontinence: a systematic review and meta-analysis. Int Urogynecol J;22:923-32.
26. Tahseen S, Reid P. Effect of transobturator tape on overactive bladder symptoms and urge urinary incontinence in women with mixed urinary incontinence. Obstet Gynecol 2009;113:617-23.
27. Barber MD, Kleeman S, Karram MM, et al. Risk factors associated with failure 1 year after retropubic or transobturator midurethral slings. Am J Obstet Gynecol 2008;199:666 e1-7.
28. Houwert RM, Venema PL, Aquarius AE, Bruinse HW, Roovers JP, Vervest HA. Risk factors for failure of retropubic and transobturator midurethral slings. Am J Obstet Gynecol 2009;201:202 e1-8.
29. Richter HE, Litman HJ, Lukacz ES, et al. Demographic and clinical predictors of treatment failure one year after midurethral sling surgery. Obstet Gynecol;117:913-21.
30. Nager CW, Sirls L, Litman HJ, et al. Baseline urodynamic predictors of treatment failure 1 year after mid urethral sling surgery. J Urol;186:597-603.
31. Barber MD, Kleeman S, Karram MM, et al. Transobturator tape compared with tension-free vaginal tape for the treatment of stress urinary incontinence: a randomized controlled trial. Obstet Gynecol 2008;111:611-21.
32. Barber MD, Walters MD, Bump RC. Short forms of two condition-specific quality-of-life questionnaires for women with pelvic floor disorders (PFDI-20 and PFIQ-7). Am J Obstet Gynecol 2005;193:103-13.
33. Palva K, Nilsson CG. Prevalence of urinary urgency symptoms decreases by mid-urethral sling procedures for treatment of stress incontinence. Int Urogynecol J;22:1241-7.
34. Abdel-fattah M, Mostafa A, Young D, Ramsay I. Evaluation of transobturator tension-free vaginal tapes in the management of women with mixed urinary incontinence: one-year outcomes. Am J Obstet Gynecol;205:150 e1-6.
35. Abdel-Fattah M, Ramsay I, Pringle S, et al. Randomised prospective single-blinded study comparing 'inside-out' versus 'outside-in' transobturator tapes in the management of urodynamic stress incontinence: 1-year outcomes from the E-TOT study. BJOG;117:870-8.
36. Abdel-fattah M, Ramsay I, Pringle S, Hardwick C, Ali H. Evaluation of transobturator tapes (E-TOT) study: randomised prospective single-blinded study comparing inside-out vs. outside-in transobturator tapes in management of urodynamic stress incontinence: short term outcomes. Eur J Obstet Gynecol Reprod Biol;149:106-11.
37. Brubaker L, Moalli P, Richter HE, et al. Challenges in designing a pragmatic clinical trial: the mixed incontinence -- medical or surgical approach (MIMOSA) trial experience. Clin Trials 2009;6:355-64.
38. Goode PS, Burgio KL, Johnson TM, 2nd, et al. Behavioral therapy with or without biofeedback and pelvic floor electrical stimulation for persistent postprostatectomy incontinence: a randomized controlled trial. JAMA;305:151-9.
39. MacDonald R, Fink HA, Huckabay C, Monga M, Wilt TJ. Pelvic floor muscle training to improve urinary incontinence after radical prostatectomy: a systematic review of effectiveness. BJU Int 2007;100:76-81.
40. Husby VS, Helgerud J, Bjorgen S, Husby OS, Benum P, Hoff J. Early postoperative maximal strength training improves work efficiency 6-12 months after osteoarthritis-induced total hip arthroplasty in patients younger than 60 years. Am J Phys Med Rehabil;89:304-14.
41. Bradley CS, Nygaard IE, Mengeling MA, et al. Urinary incontinence, depression and posttraumatic stress disorder in women veterans. Am J Obstet Gynecol;206:502 e1-8.
42. Barber MD, Brubaker L, Menefee S, et al. Operations and pelvic muscle training in the management of apical support loss (OPTIMAL) trial: design and methods. Contemp Clin Trials 2009;30:178-89.

43. Burgio KL, Goode PS, Locher JL, et al. Predictors of outcome in the behavioral treatment of urinary incontinence in women. Obstet Gynecol 2003;102:940-7.
44. Botros SM, Abramov Y, Goldberg RP, et al. Detrusor overactivity and urge urinary incontinence [corrected] following midurethral versus bladder sling procedures. Am J Obstet Gynecol 2005;193:2144-8.
45. Paick JS, Oh SJ, Kim SW, Ku JH. Tension-free vaginal tape, suprapubic arc sling, and transobturator tape in the treatment of mixed urinary incontinence in women. Int Urogynecol J Pelvic Floor Dysfunct 2008;19:123-9.
46. Choe JH, Choo MS, Lee KS. The impact of tension-free vaginal tape on overactive bladder symptoms in women with stress urinary incontinence: significance of detrusor overactivity. J Urol 2008;179:214-9.
47. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed June.
48. Nager CW, Brubaker L, Litman HJ, et al. A randomized trial of urodynamic testing before stress-incontinence surgery. N Engl J Med;366:1987-97.
49. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
50. Dyer KY, Xu Y, Brubaker L, et al. Minimum important difference for validated instruments in women with urge incontinence. Neurourol Urodyn;30:1319-24.
51. Barber MD, Spino C, Janz NK, et al. The minimum important differences for the urinary scales of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire. Am J Obstet Gynecol 2009;200:580 e1-7.
52. Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. Trials 2007;8:31.
53. Goode PS, Burgio KL, Kraus SR, Kenton K, Litman HJ, Richter HE. Correlates and predictors of patient satisfaction with drug therapy and combined drug therapy and behavioral training for urgency urinary incontinence in women. Int Urogynecol J;22:327-34.
54. Lowenstein L, Kenton K, FitzGerald MP, Brubaker L. Clinically useful measures in women with mixed urinary incontinence. Am J Obstet Gynecol 2008;198:664 e1-3; discussion e3-4.
55. Coyne KS, Matza LS, Thompson CL. The responsiveness of the Overactive Bladder Questionnaire (OAB-q). Qual Life Res 2005;14:849-55.
56. Nygaard I, Chai TC, Cundiff GW, et al. Summary of Research Recommendations From the Inaugural American Urogynecologic Society Research Summit. Female Pelvic Med Reconstr Surg;17:4-7.
57. Paick JS, Ku JH, Kim SW, Oh SJ, Son H, Shin JW. Tension-free vaginal tape procedure for the treatment of mixed urinary incontinence: significance of maximal urethral closure pressure. J Urol 2004;172:1001-5.
58. Margolis MK, Fox KM, Cerulli A, Ariely R, Kahler KH, Coyne KS. Psychometric validation of the overactive bladder satisfaction with treatment questionnaire (OAB-SAT-q). Neurourol Urodyn 2009;28:416-22.
59. Avery KN, Bosch JL, Gotoh M, et al. Questionnaires to assess urinary and anal incontinence: review and recommendations. J Urol 2007;177:39-49.
60. Coyne KS, Matza LS, Thompson CL, Kopp ZS, Khullar V. Determining the importance of change in the overactive bladder questionnaire. J Urol 2006;176:627-32; discussion 32.
61. Coyne KS, Matza LS, Thompson C, Jumadilova Z, Bavendam T. The responsiveness of the OAB-q among OAB patient subgroups. Neurourol Urodyn 2007;26:196-203.
62. Rogers RG, Coates KW, Kammerer-Doak D, Khalsa S, Qualls C. A short form of the Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-12). Int Urogynecol J Pelvic Floor Dysfunct 2003;14:164-8; discussion 8.
63. Brooks R. EuroQol: the current state of play. Health Policy 1996;37:53-72.
64. Hundley AF, Wu JM, Visco AG. A comparison of perineometer to brink score for assessment of pelvic floor muscle strength. Am J Obstet Gynecol 2005;192:1583-91.

65. Bo K, Raastad R, Finckenhagen HB. Does the size of the vaginal probe affect measurement of pelvic floor muscle strength? Acta Obstet Gynecol Scand 2005;84:129-33.
66. Kerschan-Schindl K, Uher E, Wiesinger G, et al. Reliability of pelvic floor muscle strength measurement in elderly incontinent women. Neurourol Urodyn 2002;21:42-7.
67. Frawley HC, Galea MP, Phillips BA, Sherburn M, Bo K. Reliability of pelvic floor muscle strength assessment using different test positions and tools. Neurourol Urodyn 2006;25:236-42.
68. Rahmani N, Mohseni-Bandpei MA. Application of perineometer in the assessment of pelvic floor muscle strength and endurance: a reliability study. J Bodyw Mov Ther;15:209-14.
69. The EuroQol Group. EuroQol: A new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.
70. Dumville JC, Manca A, Kitchener HC, Smith AR, Nelson L, Torgerson DJ. Cost-effectiveness analysis of open colposuspension versus laparoscopic colposuspension in the treatment of urodynamic stress incontinence. BJOG 2006;113:1014-22.
71. Manca A, Sculpher MJ, Ward K, Hilton P. A cost-utility analysis of tension-free vaginal tape versus colposuspension for primary urodynamic stress incontinence. BJOG 2003;110:255-62.
72. Nager CW, Brubaker L, Daneshgari F, et al. Design of the Value of Urodynamic Evaluation (ValUE) trial: A non-inferiority randomized trial of preoperative urodynamic investigations. Contemp Clin Trials 2009;30:531-9.
73. Swift SE, Yoon EA. Test-retest reliability of the cough stress test in the evaluation of urinary incontinence. Obstet Gynecol 1999;94:99-102.
74. Bosch JL, Cardozo L, Hashim H, Hilton P, Oelke M, Robinson D. Constructing trials to show whether urodynamic studies are necessary in lower urinary tract dysfunction. Neurourol Urodyn;30:735-40.
75. Nager CW, Kraus SR, Kenton K, et al. Urodynamics, the supine empty bladder stress test, and incontinence severity. Neurourol Urodyn;29:1306-11.
76. Hashim H, Abrams P. Is the bladder a reliable witness for predicting detrusor overactivity? J Urol 2006;175:191-4; discussion 4-5.
77. Rovner ES, Goudelocke CM. Urodynamics in the evaluation of overactive bladder. Curr Urol Rep;11:343-7.
78. Visco AG, Brubaker L, Richter HE, et al. Anticholinergic versus botulinum toxin A comparison trial for the treatment of bothersome urge urinary incontinence: ABC trial. Contemp Clin Trials;33:184-96.
79. Gamble TL, Botros SM, Beaumont JL, et al. Predictors of persistent detrusor overactivity after transvaginal sling procedures. Am J Obstet Gynecol 2008;199:696 e1-7.
80. Moalli PA, Papas N, Menefee S, Albo M, Meyn L, Abramowitch SD. Tensile properties of five commonly used mid-urethral slings relative to the TVT. Int Urogynecol J Pelvic Floor Dysfunct 2008;19:655-63.
81. Riesenhuber A BM, Posch M, Aufricht C. Diuretic potential of energy drinks. Amino Acids 2006;31:81-3.
82. Creighton SM, Stanton SL. Caffeine: does it affect your bladder? . Br J Urol 1990;6:613-14.
83. Lee JG, Wein AJ, Levin RM. The effect of caffeine on the contractile response of the rabbit urinary bladder to field stimulation. . Gen Pharmacol 1993;24:1007-11.
84. Lee JG, Wein AJ, Levin RM. The effect of caffeine on the contractile response of the rabbit urinary bladder to field stimulation. General Pharmacology 1993;24:1007-11.
85. http://www.icsoffice.org/Publications/ICI_4/files-book/recommendation.pdf. Accessed March, 2012.
86. Fitzgerald MP, Stablein U, Brubaker L. Urinary habits among asymptomatic women. Am J Obstet Gynec 2002;187.
87. Dowd TT, Bampbell JM, Jones JA. Fluid intake and urinary incontinence in older community-dwelling women. J Community Health Nursing 1996;13:179-86.
88. Panel on Dietary Reference Intakes for Electrolytes and Water. Dietary Reference Intakes for Water, Potassium, Sodium, Chloride, and Sulfate. In: Standing committee on the scientific evaluation of dietary reference intakes, ed. Food and Nutrition Board of the Institute of Medicine of the National Academies,. Washington, DC: The National Academies Press; 2004:73-185.

89. Dallosso HM, McGrother CW, Matthews RJ, Donaldson MM. . The association of diet and other lifestyle factors with overactive bladder and stress incontinence: a longitudinal study in women. BJU Int 2003;92:69-77.
90. Subak LL, Wing R, West DS, et al. Weight loss to treat urinary incontinence in overweight and obese women. N Engl J Med 2009;360:481-90.
91. Bump RC, Sugerman HJ, Fantl JA, McClish DK. . Obesity and lower urinary tract function in women: effect of surgically induced weight loss. Am J Obstet Gynecol 1992;167:392-7.
92. Subak LL, Whitcomb E, Shen H, Saxton J, Vittinghoff E, Brown JS. Weight loss: A novel and effective treatment for urianry incontinence. J Urol 2005;174:190-5.
93. Nuotio M, Jylha M, Koivisto AM, Tammela TL. Association of smoking with urgency in older people. Eur Urol 2001;40:206-12.
94. Bump RC, McClish DK. Cigarette smoking and urinary incontinence in women. Am J Obstet Gynecol 1992;167:1213-8.
95. Koley B, , Koley J, Saha JK. The effects of nicotine on spontaneous contractions of cat urinary bladder in situ. Br J Pharmacol 1984;83:347-55.
96. Coyne KS, Sexton CC, Irwin DE et al. The impact of overactive bladder, incontinence and other lower urinary tract symptoms on quality of life, work productivity, sexuality and emotional wellbeing in men and women: results from the EPIC study. BJU Int 2008;101:1388-95.
97. Jelovsek JE, Barber MD, Paraiso MF, Walters MD. Functional bowel and anorectal disorders in patients with pelvic organ prolapse and incontinence. Am J Obstet Gynecol 2005;193:2105-11.
98. Raza-Khan F, Cunkelman J, Lowenstein L, Shott S, Kenton K. Prevalence of bowel symptoms in women with pelvic floor disorders. Int Urogynecol J;21:933-8.
99. Godec CJ. 'Timed voiding' – a useful tool in the treatment of urinary incontinence. Urology 1984;23:97-100.
100. Wyman JF, Fantl JA. Bladder training in ambulatory care management of urinary incontinence. Urol Nurs 1991;11:11-7.
101. Wilson PD, Berghamns B, Hagen S, Hay-Smith J, Moore K, Nygaard I, et al. Adult conservative management. In: Abrams P CL, Khoury S, Wein AJ, ed. Incontinence: Proceedings from the third international consultation on incontinence. Plymouth, UK: Health Publications, Ltd; 2005:855-964.
102. Fantl J, Newman DK, Colling J, DeLancey JO, Keeys C, Loughery R, et al. for the Urinary Incontinence in Adults Guideline Update Panel. Urinary incontinence in adults: Acute and chronic managment. Clincial practice guideline No. 2: Update (AHCPR Publication No 96-0692). Agency for Health Care and Policy Research 1996.
103. Newman DK. Behavioral treatments: Implementing toileting, bladder training, and pelvic floor muscle rehabilitation programs. In: Newman DK, Wein AJ, ed. Managing and treating urinary incontinence. Baltimore: Health Professions Press; 2009:233-43.
104. Wyman JF. Behavioral interventions for the patient with overactive bladder. Journal of Wound, Ostomy, and Continence Nursing 2005;32:S11-5.
105. Fantl JA, Wyman JF, McClish DK, Harkins SW, Elawick RK, Taylor JR, et al Efficacy of bladder training in old women with urinar incontinence. JAMA 1991;265:609-13.
106. Miller JM, Ashton-Miller JA, DeLancey JO. A pelvic muscle precontraction can reduce cough-related urine loss in selected women with mild SUI. J Am Geriatr Soc 1998;46:870-4.
107. Bo K, Talseth T, Holme I. Single blind, randomised controlled trial of pelvic floor exercises, electrical stimulation, vaginal cones, and no treatment in management of genuine stress incontinence in women. BMJ 1999;318:487-93.
108. Hay-Smith EJ, Herderschee R, Dumoulin C, Herbison GP. Comparisons of approaches to pelvic floor muscle training for urinary incontinence in women. Cochrane Database Syst Rev;12:CD009508.
109. Richter HE, Burgio KL, Goode PS, et al. Non-surgical management of stress urinary incontinence: ambulatory treatments for leakage associated with stress (ATLAS) trial. Clin Trials 2007;4:92-101.
110. Borello-France D, Burgio KL, Goode PS, et al. Adherence to behavioral interventions for urge incontinence when combined with drug therapy: adherence rates, barriers, and predictors. Phys Ther;90:1493-505.

111. Miranne JM, Lopes V, Carberry CL, Sung VW. The effect of pelvic organ prolapse severity on improvement in overactive bladder symptoms after pelvic reconstructive surgery. Int Urogynecol J.
112. Wennberg AL, Molander U, Fall M, Edlund C, Peeker R, Milsom I. A longitudinal population-based survey of urinary incontinence, overactive bladder, and other lower urinary tract symptoms in women. Eur Urol 2009;55:783-91.
113. Heidler S, Mert C, Temml C, Madersbacher S. The natural history of the overactive bladder syndrome in females: A long-term analysis of a health screening project. Neurourol Urodyn;30:1437-41.
114. Broome BA. Psychometric analysis of the Broome Pelvic Muscle Self-Efficacy Scale in African-American women with incontinence. Urol Nurs 2001;21:289-97.
115. Broome BA. Development and testing of a scale to measure self-efficacy for pelvic muscle exercises in women with urinary incontinence. Urol Nurs 1999;19:258-68.
116. Rett MT, Simoes JA, Herrmann V, Pinto CL, Marques AA, Morais SS. Management of stress urinary incontinence with surface electromyography-assisted biofeedback in women of reproductive age. Phys Ther 2007;87:136-42.
117. Gameiro MO, Moreira EH, Gameiro FO, Moreno JC, Padovani CR, Amaro JL. Vaginal weight cone versus assisted pelvic floor muscle training in the treatment of female urinary incontinence. A prospective, single-blind, randomized trial. Int Urogynecol J;21:395-9.
118. Amaro JL, Moreira EC, De Oliveira Orsi Gameiro M, Padovani CR. Pelvic floor muscle evaluation in incontinent patients. Int Urogynecol J Pelvic Floor Dysfunct 2005;16:352-4.
119. Gilling PJ, Wilson LC, Westenberg AM, et al. A double-blind randomized controlled trial of electromagnetic stimulation of the pelvic floor vs sham therapy in the treatment of women with stress urinary incontinence. BJU Int 2009;103:1386-90.
120. Hung HC, Hsiao SM, Chih SY, Lin HH, Tsauo JY. Effect of pelvic-floor muscle strengthening on bladder neck mobility: a clinical trial. Phys Ther;91:1030-8.
121. Gameiro MO, Moreira EC, Ferrari RS, Kawano PR, Padovani CR, Amaro JL. A comparative analisys of pelvic floor muscle strength in women with stress and urge urinary incontinence. Int Braz J Urol;38:661-6.
122. McCulloch CE SS. Generalized, linear, and mixed models. New York: John Wiley; 2001.
123. Efron B TR. An introduction to the bootstrap. New York: Chapman and Hall, 1993.

10. ESTEEM Ancillary Study: Goa**l**s am**o**ng **w**omen with mixed urinary incontinence undergoing midurethral sling surgery randomized to behavioral therapy or no behavioral therapy (**GIoW**)

Patient reported outcome (PRO) measures are of critical importance in the evaluation of functional disorders because anatomical and physiologic tests do not precisely correlate with patient experience. Symptom severity and quality of life questionnaires partly fill this gap. The Urinary Distress Inventory (UDI), a measure of pelvic floor symptoms, the Pelvic Floor Impact Questionnaire (PFIQ) and the Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ), measures of health related quality of life, are commonly used symptom and quality of life questionnaires. Within the PFDN, these questionnaires are used in conjunction with physical exam and physiologic testing to measure disease burden and to assess cure. While these questionnaires characterize the severity of symptoms and their impact on quality of life, they do not rank symptom importance nor do they provide an individualized blueprint of what women hope to achieve with treatment. More recently, goal attainment scaling (GAS) has emerged as an established methodology of determining individual women's goals and whether or not they meet personalized goals following treatment.

In goal attainment scaling, patients are asked to list goals and rank their importance; following treatment, women rate whether or not the goal was achieved. Patient-identified goals have been described as the "fourth dimension" of pelvic floor disorder assessment, after physical findings, symptoms, and quality of life. (Lowenstein, 2008) Individualized goals are not adequately captured by traditional symptom severity or quality of life measures. For example, among a group of 200 women seeking care for pelvic floor dysfunction, continence goals were ranked more highly than resolution of bulge symptoms, despite the presence of advanced (Stage 3) prolapse on exam and bother reported on the PFDI.(Elkardy, 2013) In a UITN randomized trial with standardized video consent (SIStr), women undergoing SUI surgery had high expectations for treatment of not only SUI symptoms, but also for treatment of their urgency and frequency, despite being told in that study that the midurethral sling (MUS) was not designed to resolve their urgency symptoms, documentation of stress incontinence on urodynamics and bother and quality of life changes consistent with SUI reported on the PFDI and PFIQ.(Mallett, 2008) Among women with a variety of pelvic floor disorders, patient goals and expectations vary and are linked to treatment satisfaction. (Elkardy, 2003; Hullfish 2004; Komesu 2008) Conversely, unmet goals are closely associated with patient dissatisfaction after treatment. (Elkardy, 2003; Hullfish 2004; Komesu 2008) Despite the importance of individualized goal setting, prior goal attainment scaling studies in urogynecology are limited by inclusion of small numbers of women with an array of pelvic floor dysfunction, lack of assessment of the difference between short and long term goals, and have not consistently followed women after treatment to determine whether their goals are achieved. ***A key gap in our understanding of mixed urinary incontinence and women's expectations following treatment is accurate goal characterization and determination of whether or not goals are attained in the short and long term following treatment.*** ESTEEM provides an ideal study setting in which to answer this question.

ESTEEM will compare the effect of peri-operative behavioral/pelvic floor therapy (BPTx) plus MUS to MUS alone on MUI treatment in 472 women. This trial provides an ideal setting in which to describe individualized goals for MUI treatment as well as the importance of goal attainment on women's impression of cure. This, in turn, will enable providers to ultimately negotiate expectations so that providers and patients have better communication regarding the benefits and limitations of various treatments for mixed urinary incontinence. The *long-term goal* of this supplementary study to ESTEEM is to better understand patient expectations following treatment for MUI in order to provide patients and providers an informed platform for discussion of treatment options and realistic outcome expectations. The *objectives* of this proposal are to describe patient centered goals among a group of women with MUI undergoing midurethral sling surgery with and without BPTx as well as determine whether or not these goals were met following treatment using the validated Self-Assessment Goal Achievement (SAGA) questionnaire. Our *expectation* is that a better understanding of individualized patient goals will improve patient-provider communication,

and provide a unique aspect of patient reported outcomes not currently measured with standard symptom severity and quality of life measures.

Aim 1: To describe patient reported goals and goal ranking among women consenting to ESTEEM. *We hypothesize that women's goals vary and are not currently captured by standard symptom severity and QOL measures.*

Aim 2: To determine whether or not women achieve self-reported goals following treatment for MUI and to compare those who achieve their goals to those who do not in both the intermediate (6 months) and longer term (12 months). *We hypothesize that women who report and rank continence related goals are more likely to achieve those goals than goals related to general health and specific activities and that goal achievement is related to patient's PGI-I scores.*

Significance: PROs are critical to the assessment of pelvic floor dysfunction, yet standardized measures of symptom severity and quality of life may not capture an individual women's motivation and expectations for seeking treatment. Goal attainment scaling is an established methodology of describing and ranking individual goals and has been used in a variety of fields including treatment of pelvic floor dysfunction.(Khuller, 2013) Goal attainment scaling offers unique insight into individual concerns regarding common disorders, such as MUI. While it is known that the impact of pelvic floor dysfunction varies between individuals with similar physiologic measures of disease, the underpinnings of what explains the differences in bother and impact on quality of life are less well characterized. In addition, patient expectations are likely to drive care seeking as well as adherence to treatment regimens and are, in turn, correlated with satisfaction with those treatment outcomes. MUI is a common disorder with lack of consensus regarding treatment; ESTEEM will test whether or not BPTx is beneficial prior to and following sling surgery. A key aspect of understanding women's satisfaction with these treatment options is determining the importance of various lower urinary tract symptoms to individuals and what individualized goals women have for treatment.

Innovation: Mixed urinary incontinence is bothersome to women and often presents a treatment conundrum to providers. The symptom of urinary leakage is what concerns the patient most, yet the etiology of the UUI and SUI are thought to be different and the treatments for one may lead to exacerbation of the other. While ESTEEM will measure symptom severity and quality of life for both SUI and UUI symptoms, currently the protocol does not contain a measure of the importance of alleviating specific symptoms to individual women. In addition, women participating in the trial likely have unique goals and concerns not currently captured with standard symptom severity and quality of life measures presently included in this study. Inclusion of the SAGA questionnaire at baseline, six months and one year after MUS with or without BPTx will offer the PFDN the opportunity to characterize treatment goals in a large number of women with MUI undergoing MUS surgery and assess whether or not those goals are achieved. While goal attainment scaling is an established method of assessing individual goals, until recently, a standardized and valid measure of assessing goals was not available. The SAGA questionnaire has been validated among women with lower urinary tract symptoms and fills that void.(Brubaker, 2013) SAGA consists of nine standardized goals regarding urinary symptoms, and asks women to rate the importance of these standardized goals on a scale from 0 (not applicable) to 5 (very important goal). In addition to these common goals, women are asked to record up to five of their individualized goals and rank them in a similar fashion. At follow-up following treatment, women are asked to rate whether or not they achieved their goals on a scale from 1 (did not achieve goal) to 5 (greatly exceeded goal). Importantly, the common goal list of 9 items was generated from patient and expert interviews, and has undergone validation both within the US and abroad. Adequate face, concurrent, known-groups, and convergent validity and item distribution validity have been determined in a pilot study of 104 subjects and re-evaluated on an international basis in an additional 29 subjects. Reliability and internal consistency testing was not performed because goals were assumed to vary between individuals. *This proposal is innovative, in our opinion, because it will assess goal setting using a newly validated questionnaire in a large group of women with MUI, a common condition which is difficult for patients to understand and for providers to explain, and will determine whether goal attainment is linked to patient global impression of improvement both in the short and long term.* Finally, this innovative proposal offers the PFDN the opportunity to add an translational aim to ESTEEM by linking the clinical

science of a comparative effectiveness trial to individual patients seeking care. This ultimately may inform community dwelling women's decisions to pursue or not to pursue care.

Approach:
Aim 1: To describe patient reported goals and ranking of goals among women consenting for ESTEEM. *We hypothesize that women's goals vary and are not currently captured by standard symptom severity and QOL measures.*

Introduction: Assessment of individualized patient goals offers a unique perspective of expectations and goals with treatment that current PROs do not capture. The *objective* of this aim is to administer the SAGA questionnaire at women's baseline visit in ESTEEM and to describe their ranking of the nine standardized questions in SAGA. In addition, women will be asked to list up to five individualized goals for treatment and also rank their importance. Our *working hypothesis* is that the importance of the 9 common goals will vary between individuals. In addition we hypothesize that women will list a variety of individual goals which are not presently represented by symptom severity and quality of life measures. We will *achieve this aim* by administering the SAGA questionnaire at baseline in women recruited to ESTEEM. Our *expectation* is that the description of baseline goals of women recruited to ESTEEM will offer insight into what women are seeking with treatment for their MUI.

Methods: Women will be administered the baseline set of nine pre-specified goals as well as be asked to list up to 5 individual goals for their therapy. Goals will be ranked on a 6 point scale from 0-Not applicable to 5 – very important goal. In the original validation study of the SAGA, women randomly completed either the pre-specified or self-specified goals first. No order effects were noted in the numbers of goals listed or ranking of nine pre-specified goals. For this study will complete the nine pre-specified goals followed by listing their individual goals. Individualized goals will be transcribed and entered into the patient database; these goals will then be presented to the patient in follow-up assessment of goal achievement in Aim #2. Table 1 is the SAGA questionnaire.

Table 1: Self-Assessment Goal Achievement

Item	Not applicable (0)	Not very important goal (1)	Very important goal (5)
Reduce the number of times I go to the bathroom throughout the day			
Reduce the number of times I get up at night to go to the bathroom			
Reduce the sensation of pressure in my lower abdomen			
Reduce the difficulties I have completely emptying my bladder			
Reduce the difficulty starting or maintaining a urinary stream			
Reduce the urine loss when I cough, laugh or sneeze			
Reduce my urine leakage			
Reduce the sudden need to rush to the bathroom			
List up to 5 goals of your own following surgery and rank them. Goal #1: Goal #2: Goal #3:			

Aim #1 is descriptive in nature therefore the analyses are qualitative versus quantitative. For self-selected goals, goals will be classified by the study working group into categories. The working group will review the goals in order to generate categories; goals will be categorized and then compared across individual categorization. Development of categories and categorization will be by consensus. For the analysis of ranking, each subject's goal 'selection' will be ranked with #1 for their 1st choice, #2 for their 2nd choice and #3 for their 3rd choice. We will rank goals on a preferential ballot which will be ultimately based on the number of goal categories identified by qualitative expert review for individualized goals and for nine categories in the pre-specified goals.

A preferential ballot allows for comparison of goal rankings between individuals and assigns a value to each goal subdomain listed per individual, and a standard value to any goal subdomain identified in the entire population but not listed for a particular individual. A preferential ballot is used in political elections, but can also be used to prioritize preferences across individuals and is referred to as a "Borda count". Originally designed for political elections when there were multiple candidates on a ballot, Borda counts determine the

"winner" of a ballot by giving each candidate on the ballot points corresponding to the position in which the candidate is ranked by each voter. Once all votes have been counted, the candidate with the most points is the winner. In our analysis, a modified Borda preferential ballot consists of candidates (here, the list of goal subdomains) and a ballot for each participant, in which a rank of 1 is assigned for the participant's highest ranked goal, a rank of 2 for their second highest ranked goal, and so on. If a participant doesn't rank all of the candidates (goal subdomains), then the mid-rank of the un-used ranks are assigned – this assures each ballot receives equal weight in the ballot count. The derivation of mid-rank was calculated by summing the remainder of the ranks divided by number of left over ranks. The first winner is based in lowest average rank across all ballots, the second winner is based in second lowest average rank, and so on. The analysis of ballots can be done simply by computing the average rank for each goal category across all ballots.

Since this aim is qualitative in nature a formal power analysis was not computed; given the number of women who will be recruited to ESTEEM, we should have more than enough subjects to reach saturation on goal categories and to evaluated even small difference between goal rankings.

Potential Problems and Solutions: It may be that women have difficulty in generating individualized goals, although prior research has documented that women, on average, do not have difficulty generating up to 4 goals in prior studies. If women have difficulty generating goals, they will be prompted by the coordinators to list what they wish to achieve with their treatment; this will be done without prompting for specific goals to avoid bias.

Aim 2: To determine whether or not women achieve self-reported goals following treatment for MUI. *We hypothesize that women who report and rank specific continence related goals are more likely to achieve those goals than goals related to general health and specific activities and that achievement of these goals will be related to patient's PGI-I scores. In addition, we hypothesize that women whose goals are achieved will report better global improvement in continence and quality of life than women who do not achieve their goals.*

Introduction: Achievement of patient goals offers a unique perspective of patient's assessment of outcome of treatment. The *objective* of this aim is to administer the SAGA follow-up goal achievement questionnaire at women's follow-up visits in ESTEEM at 6 and 2 months. We will describe women's goal achievement and compare which goal categories are more likely to be achieved. In addition, we will correlate goal achievement with PGI-I scores to further evaluate which goals are best correlated with patient's global impression of improvement. Finally, we will observe whether goal achievement is stable between 6 and 12 months by comparing goals achievement at the two timepoints. Our *working hypothesis* is that goal achievement varies between individuals and that women who rank continence goals will be more likely to achieve those goals than goals not related to continence. In addition we hypothesize that goal achievement changes over time and that more women will achieve goals at 6 months than do at one year. Finally, we hypothesize that goal achievement will be significantly correlated with PGI-I scores. We will *achieve this aim* by administering the SAGA follow-up questionnaire at 6 months and one year in women recruited to ESTEEM. Our *expectation* is that goal achievement varies over time and between individuals based on baseline goal setting and treatment efficacy.

Methods: Women will be administered the follow-up SAGA questionnaire at 6 months and one year follow-up in the ESTEEM trial. The follow-up questionnaire is similar to the baseline questionnaire in that it still contains the 9 pre-specified goals and a list of the patient's self-determined goals established at baseline. The response categories for follow-up are 1 (did not achieve goal) to 5 (greatly exceeded goal). The number of goals for each patient will vary as women may report less that 5 self-determined goals, and they may have ranked some of the nine pre-specified goals as "not-applicable" at baseline.

Again a Borda count system will be used to rank in this cohort goals and goal categories that were most likely to be achieved and we will describe the goals that were more likely to be achieved in this cohort at 6

months and one year. In the original validation study, a cut off T score of > 50 was determined to indicate women who achieved their goals, versus women who scored </= 50 who did not achieve goals according to the formula provided by Kiresuk and Sherman (T-scores with mean = 50 and SD = 10). (Kiresuk, 1968) Weights will be applied to women's individualized goal achievement ratings. We will compare women who achieve goals to those who do not at 6 and 12 months to determine both if there are baseline differences between those women who achieve goals and do not, and if there is a different pattern of goal attainment at short term (6 months) and longer term (12 month) follow up.

Potential problems and Strategies to overcome them: It is possible that women will rank goal attainment highly for all listed goals and that there will not be differences noted between continence goals and self stated goals. In this instance the data generated are still valuable because the negative findings are informative to the counseling of patients.

References:

Adams S, Dramitinos P, Shapiro A et al. Do patient goals vary with stage of prolapse? Am J Obstet Gynecol 2013; 205:502.e1-6

Brubaker L, Piault EC, Tully SE et al. Validation study of the self-assessment goal achievement (SAGA) questionnaire for lower urinary tract symptoms. Int J Clin Pract 2013; 67(4): 342–350.

Brubaker L, Khullar V, Piault E et al. Goal attainment scaling in patients with lower urinary tract symptoms: development and pilot testing of the Self-Assessment Goal Achievement questionnaire. Int Urogynecol J (2011) 22:937–946

Elkadry EA, Kenton KS, FitzGerald MP, Shott S, Brubaker L. Patient-selected goals: a new perspective on surgical outcome. Am J Obstet Gynecol. 2003 Dec;189(6):1551-7; discussion 1557-8.

Hullfish KL, Bovbjerg VE, Steers WD. Patient-centered goals for pelvic floor dysfunction surgery: long-term follow-up. Am J Obstet Gynecol. 2004 Jul;191(1):201-5.

Khuller V, Espuna-Pons M, Marschall-Kehrel D et al. Clinical value of a patient-related goal attainment measure: The global development of Self-Assessment Global Achievement (SAGA) Questionnaire for patients with lower urinary tract symptoms. Neurourol Urodynam 2013; epub ahead of print.

Khullar V, Marschall-Kehrel D, Espuna-Pons M et al. European content validation of the Self-Assessment Goal Achievement (SAGA) questionnaire in patients with overactive bladder. Int Urogynecol J 2013; DOI 10.1007/s00192-012-2039-x

Kiresuk T, Sherman R. Goal attainment scaling: a general method of evaluating comprehensive community mental health programs. Community Ment Health J 1968; 4: 443–53.

Komesu YM, Rogers RG, Rode MA, Craig EC, Schrader RM, Gallegos KA, Villareal B. Patient-selected goal attainment for pessary wearers: what is the clinical relevance? Am J Obstet Gynecol. 2008 May;198(5):577.e1-5.

Lowenstein L, Fitzgerald MP, Kenton K, et al. Patient-selected goals: the fourth dimension in assessment of pelvic floor disorders. Int Urogynecol J Pelvic Floor Dysfunct 2008;19:81-4.

Mallet V, Brubaker L, Stoddard AM et al. The expectations of patients who undergo surgery for stress incontinence. Am J Obstet Gynecol 2008;198:308.e1-308.e6.