Urinary Incontinence Treatment Network (UITN)

ValUE:
Value of Urodynamic Evaluation

July 25, 2008

TABLE OF CONTENTS

APPENDICES

FIGURES AND TABLES

A. STUDY AIMS

Primary Aim: To determine if women desiring surgery for diagnosed, uncomplicated predominant stress urinary incontinence (UI) who receive a basic office evaluation only without preoperative urodynamic studies (No UDS arm) have non-inferior treatment outcomes compared to women who receive both office evaluation and preoperative UDS (UDS arm).

The Secondary Aims of this trial are:

- To determine, among women randomized to the UDS arm, how often physicians alter clinical decision-making based on the results of UDS before a planned intervention.
- To compare the amount of improvement in incontinence outcomes in women randomized to the two diagnostic methods.
- To determine the incremental cost and utility of performing UDS compared with not performing UDS (cost-minimization, utility analysis).

B. BACKGROUND AND SIGNIFICANCE

B.1 Urodynamics, Clinical Decision-Making and Clinical Outcomes

Although no reliable and specific figures are available for the total expenditure on UDS, UDS is commonly performed for patients with urinary incontinence (UI) regardless of gender and age. UDS is typically performed prior to incontinence surgery. Urodynamic studies are expensive, time-consuming, and uncomfortable diagnostic investigations. The 3rd ICI reported insufficient evidence with which to answer the following key research questions related to UDS: 1) Do physicians alter clinical decision-making based on results of UDS?, and 2) Do alterations in clinical decisions made in response to UDS results improve the clinical outcomes?

The Cochrane Review on "*Urodynamic investigations for management of urinary incontinence in children and adults*" summarizes the current situation on urodynamics and the need for a randomized urodynamic trial well:

"*A Committee of the International Consultation on Incontinence (ICI) has published an extensive review of all the evidence available for dynamic testing of patients with incontinence (Griffiths 2005). They agree that the evidence for urodynamics should depend on whether the use of the test alters clinical management, and whether (as a consequence) it improves clinical outcomes (Griffiths 2005, page 589). They support research (by randomised controlled trials) that 'may provide objective documentation of the utility of soundly based tests' (Griffiths 2005, page 630). As a preliminary conclusion, however, they state that urodynamics is not needed for patients with uncomplicated stress or urge incontinence, as long as there are no symptoms or signs of voiding difficulties (Griffiths 2005, page 630)….*

There is currently no consensus about whether or not urodynamic investigations need to be performed to guide management. Indeed, no published research supports a need for cystometric testing in routine or basic evaluation of urinary incontinence (AHCPR 1996). For example, in a survey of 442 women treated surgically for stress incontinence, Black and colleagues found that the likelihood of improvement was similar regardless of whether or not urodynamic pressure studies had been conducted before surgery (Black 1997). Although this study did not conclusively show that urodynamics had little or no prognostic value, it suggested that the role of urodynamic testing needed reappraisal….

The need for pre-operative urodynamics is often justified by the consideration that pre-existing detrusor overactivity may be either a contra-indication for surgery or at least carries a worse prognosis. However, in Black's study, women who suffered from urgency before operation

actually reported a reduction in urgency after surgery (Black 1997). Similarly, in a report of the success of tension-free vaginal tape surgery in elderly women (all of whom had preoperative urodynamics), 46% had preoperative urgency cured after surgery, and 21% developed it de novo (Sevestre 2003). This symptom is commonly believed to worsen after surgery for stress incontinence, is often considered to be a contra-indication, and is thought to result in worse outcomes, but the reverse may be the case in practice. It seems likely that this finding would also apply to detrusor overactivity, and therefore its pre-operative diagnosis would seem to be unnecessary as it would not influence management. Similarly, the value of distinguishing between intrinsic sphincter deficiency and bladder neck / urethral mobility has also not been established, and the official position of the ICS is that these are 'simplistic and arbitrary terms, not to be used until further research has been done' (Abrams 2002)….

The value of accurate diagnosis also depends on the availability and effectiveness of appropriate treatments. It is of no clinical value unless it is known, for example, that urodynamics distinguishes between a group for whom surgery is effective and another group for whom it is not effective or contra-indicated or where management needs to be altered in a specific way….

The value of urodynamic investigation in the diagnosis and management of people complaining of incontinence is therefore open to question (Ramsay 1995). Some current consensus statements and practice recommendations do not advocate initial urodynamic tests prior to conservative treatment of stress incontinence (AHCPR 1996; Homma 1999) whereas others recommend them prior to continence surgery (RCOG 2003). Their role before prolapse surgery is controversial (Fowler 2006), and an ICI Committee found no evidence to suggest whether or not women with occult urinary incontinence (revealed by prolapse reduction) in fact manifest the problem after surgery, nor whether urodynamics was helpful in predicting improving this outcome (Griffiths 2005, page 629).

In view of the small numbers and the methodological questions about dropout and numbers followed up, a larger definitive randomised trial should be carried out to determine the place for urodynamics in both routine and specialised clinical care of people with incontinence. This was first suggested by Black et al in their survey of clinical practice (Black 1997) and supported by Black and others in subsequent correspondence (Black 1998; Lose 1998b). Further support for such a trial was provided by the ICI Committee on Dynamic Testing that 'may provide objective documentation of the utility of soundly based tests' (Griffiths 2005, page 630)….

B.2 Urodynamics Prior to Surgical Management of SUI

A randomized trial design for different populations with different diagnoses would unlikely have enough power to answer the question of the value of UDS for specific clinical diagnoses. One widely recognized and yet unresolved issue is whether preoperative UDS improves the outcomes in women with an office-based diagnosis of SUI considering surgery. Several professional organizations including the 3rd International Consultation on Incontinence (**3rd ICI**), the IUGA Guidelines for Research and Practice, and the RCOG recommend performance of UDS prior to surgical management of UI; although no reliable data exist to support these recommendations. In an informal survey at the 2006 San Diego UITN Steering Committee meeting, the vast majority of investigators performed preoperative UDS in nearly all patients. In the SISTEr study of 655 women who received urodynamics before surgery, typical

urodynamic parameters often given as reasons for performing UDS (e.g. VLPP, presence or absence of DO, USI) did not predict stress success or failure, although the absence of USI had a nearly statistically significant lower success rate. We therefore think that the pilot data obtained by SISTEr further justifies the need for a randomized trial of preoperative UDS and provides reassurances to any investigator concerned about the ethics of not providing this diagnostic study.

Although studies are needed for the use of UDS for all urinary incontinence disorders, we are proposing a narrower and more feasible first study for both networks that only addresses preoperative UDS for uncomplicated SUI in selected patients with predominant SUI following a basic office evaluation (BOE). We propose to study the clinical utility of UDS in women with stress dominant incontinence because of the high prevalence of, and unresolved usage of UDS for these conditions. Practices vary, but in general, a majority of the clinicians and most societies recommend conducting UDS prior to surgical management of SUI, even primary surgery. There is less controversy about performing UDS preoperatively in women with balanced mixed urinary incontinence, urge –dominant mixed incontinence or those who have failed a previous surgery. There is also less controversy about the use of UDS prior to any conservative therapy; it is generally agreed that it is not necessary. We also recognize that a study based on multiple clinical diagnoses does not answer the question of the value of urodynamics for a specific diagnosis. It is generally accepted in medicine that diagnostic studies should be requested for specific indications to answer specific questions. An underlying assumption of this trial is that urodynamics are a diagnostic tool requested to evaluate a specific clinical diagnosis. In this case, when the clinical diagnosis is primary stress incontinence, do urodynamics, confirmation of urodynamic stress incontinence, and urodynamic measures change diagnoses, treatments, and improve outcomes in a group of women planning surgery. We consider urodynamic investigations before surgery to be the generally accepted standard procedure throughout much of the world, but a procedure for which there is insufficient evidence to support. The question to be addressed in this study is whether women diagnosed with uncomplicated stress urinary incontinence following a basic office evaluation and managed without preoperative UDS have non-inferior outcomes compared to those who receive the more invasive and expensive preoperative UDS.

C. STUDY DESIGN

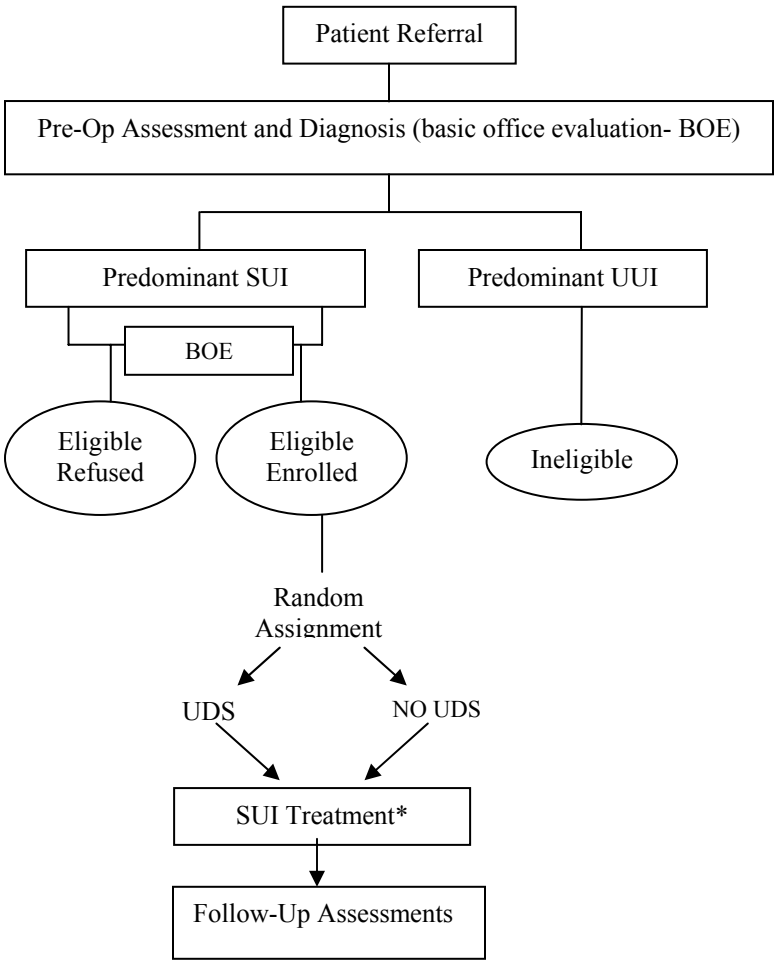
This study is designed as a randomized non-inferiority trial. By randomly assigning women to receive UDS or not, confounding factors due to surgeon and patient preferences as well as differences by clinical center will be virtually eliminated. This would not be the case in a non-randomized observational comparison.

UDS is currently considered the standard of care and the goal is to compare a less invasive alternative, a basic office evaluation without UDS (No UDS), with it to ascertain whether the No UDS arm is non-inferior to the UDS arm. It is anticipated that any difference seen between the UDS and No UDS groups will be subtle. Based on the goal and anticipated result, this fits the framework for a non-inferiority trial. At the conclusion of the study, if the No UDS arm is no more than marginally non-inferior to the UDS arm, it will be possible to deem the No UDS arm non-inferior to the UDS arm and conclude that UDS is not necessary rather than continuing with the standard of care. If this study was designed as a typical superiority trial and the success rates in the No UDS and UDS arms were very similar, the conclusion drawn would

be an inability to detect a difference between the two arms. This is not the same as declaring the No UDS arm as non-inferior to the UDS arm and is less clinically relevant.

The design of the trial is depicted in Figure 1.

Figure 1. Trial Design



*Any approved clinical care method Investigators use in their practice.

D. DEFINITION OF TREATMENT SUCCESS AND FAILURE

Treatment success is defined as:

- 70% reduction in UDI from baseline to 12 months; and
- Score of 1 or 2 ("Very much better" or "Much better") on the PGI-I at 12 months.

Treatment failure is therefore defined by the occurrence of one of the following:

- <70% reduction in UDI from baseline to 12 months; or
- Score ≥ 3 on the PGI-I at 12 months.

E. STUDY POPULATION

The study population will consist of women with predominant stress urinary incontinence (SUI) who are seeking surgical treatment. Recruitment is planned from clinical practices. Advertisement is not anticipated. The basic office evaluation (BOE) is a routine part of the Investigators' clinical assessments prior to surgery. The components of the BOE are described in Section G.1.

In short, following the BOE, (which can be done at the first visit), if the clinical diagnosis is uncomplicated, demonstrable SUI, the provider discusses SUI management options with the patient including conservative and specific surgical options. Any patient who desires non-conservative therapy and meets the inclusion and exclusion criteria will be offered study participation.

E.1 Inclusion Criteria

1. Female
2. Predominant SUI as evidenced by all of the following:
    a. Self-reported stress-type UI symptoms, of duration ≥3 months*
    b. MESA stress symptom score (percent of total possible stress score) greater than MESA urge symptom score (percent of total possible urge score)
3. Observation of leakage by provocative stress test at any volume
4. Eligible for randomization to either treatment group
5. Eligible for SUI surgery
6. Desires non-conservative therapy for SUI
7. PVR <150ml by any method. (May repeat once if initial measure is abnormal)
8. Negative urine dipstick (negative result = trace or less for leukocytes & nitrites) or negative UA or negative culture
9. Available to initiate SUI treatment within 6 weeks of randomization
10. Available for 12-months of follow-up and able to complete study assessments, per clinician judgment.
11. Signed consent form.

*Patient can be rescreened after respective time interval has been met.

E.2 Exclusion Criteria

1. Age <21 years*
2. Currently undergoing or has had recommended treatment of apical or anterior prolapse
3. No anterior or apical prolapse ≥ +1 on standing straining prolapse exam
4. Pregnant or has not completed child bearing.
5. <12 months post-partum*†

6. Active malignancy of cervix, uterus, fallopian tube(s) or ovary > Stage I, or bladder of any Stage
7. History of pelvic radiation therapy
8. Previous incontinence surgery
9. Current catheter use
10. Neurological disease known to affect bladder storage (e.g. MS, Parkinsonism, CVA)
11. Previous (i.e. repaired) or current urethral diverticulum
12. Prior augmentation cystoplasty or artificial sphincter
13. Implanted nerve stimulators for urinary symptoms or previous botox bladder injections.
14. Any pelvic surgery within the last 3 months*
15. Previous placement of synthetic mesh on a vaginal approach in the anterior compartment
16. Participation in another treatment intervention trial that might influence results of this trial.
17. A urodynamic result reviewed by the investigator in the preceding 12 months or any recollection by the investigator of urodynamic results on that subject.

*Patient can be rescreened after respective time interval has been met.

†"Partum" is defined as a delivery or other termination that occurs after 20 weeks gestation.

If more than 12 months transpire between completion of any baseline measure and randomization, any such "expired" measure(s) must be repeated prior to randomization to ensure current eligibility for the trial as well as to obtain current baseline values for critical measures that are subject to change. The one exception is the urine dipstick; this measure expires if a patient is not randomized within 3 months of screening. (See Section H.6 and Appendix A.)

F. RANDOMIZATION

Patients will be randomized to either UDS or No UDS. The process of obtaining the randomized treatment assignment will be accomplished by an automated randomization system at the Biostatistical Coordinating Center (New England Research Institutes). A back-up system of randomization using sealed envelopes will be made available in the case of technical problems. Randomization will be stratified by surgeon, using permuted blocks. Patients and surgeons will not be blinded to their treatment assignment.

G. DIAGNOSTIC PROCEDURES

G.1 Office Evaluation

The basic office evaluation will include:

- Self-reported stress-type UI symptoms, by MESA and duration of symptoms
- Provocative stress test at a non-standardized volume (may be repeated with a retrograde fill if initial test was at low bladder volume)
- PVR by any method
- Dipstick urinalysis
- Assessment of urethral mobility by one of the following methods: Q-tip test, visualization, palpation, point Aa on POP-Q exam, or lateral cystogram
- A standing, straining prolapse exam.

G.2 Urodynamic Studies

Patients randomized to the UDS arm will undergo the following core urodynamic investigations:

1. Noninstrumented uroflowmetry (NIF) with comfortably full bladder
2. PVR obtained with catheter (after NIF)
3. Filling CMG with VLPP determination attempts.
4. Pressure flow study (PFS)

Other optional UDS investigations (e.g. Urethral Pressure Profiles, Videourodynamics) may be performed if those investigations are the investigators customary practice for this patient population. Because of the poor sensitivity in diagnosing DO in the supine position (Abrams N&U, 2008), supine filling cystometry is discouraged unless the patient is physically disabled. Specific testing and procedural details of the urodynamic core procedures should conform to ICS Good Urodynamic Practice guidelines and the results and interpretation of results should conform to ICS nomenclature. Consistent with a practical clinical trial design, a specific UDS protocol for this study will not be developed; the local urodynamic testing should be performed according to standard local high quality urodynamic practice. The results of the urodynamic tests will be recorded on a standard form using ICS recommended nomenclature and diagnoses.

H. MEASUREMENT

H.1 Primary Outcome Measures

The primary outcome will be measured at 12 months and includes self-reported urinary incontinence, irritative and obstructive symptoms using the Urogenital Distress Inventory and the Patient Global Index – Improvement.

a. Urogenital Distress Inventory (UDI)

The Urogenital Distress Inventory (UDI)[1] is a patient reported questionnaire that consists of approximately 20 questions that cover stress incontinence, urge incontinence, urgency, frequency, and voiding dysfunction. It has been shown to be valid, reliable, and responsive. Significant reductions in values are typically observed after SUI surgery. Based on SISTEr data, at 24-months post-surgery, the "success" subjects had an 88% mean reduction in UDI, and the "failure" subjects had a 65% reduction; overall, there was a 78% mean reduction at 24-months, indicating sensitivity to change.

b. Global Impression Index – Improvement

The Patient Global Index (PGI-I) is a global, patient-oriented outcome measure that has been shown to have good construct validity in incontinence trials.[2,3] This patient-oriented measure will be used as accumulating data show that patient satisfaction after medical treatment is primarily determined by patient expectations (Ford RC 1997, Lochman JE 1983). The PGI-I correlates significantly with incontinence episode frequency (r=0.36), pad test weights (r=0.20) and the IQOL (r=-0.50) in a population of women with predominant SUI.[4] Global impression scales are advantageous because they are simple, direct, and intuitively understandable to both patients and investigators. This particular measure assesses both stress and urge incontinence components; construct validity has been demonstrated in a population of women with predominant SUI.[4] The PGI-I asks subjects to check the one number that best describes how one's urinary tract condition (bladder) is now, compared to how it was before they received treatment for urinary leakage. Response choices are: 1) "very much better", 2) "much better", 3) "a little better", 4) "no change", 5) "a little worse", 6) "much

worse", and 7) "very much worse". See Appendix B. Subjects who report scores of 1 or 2 (Very much better or Much better) will meet one component of the treatment success definition.

H.2 Secondary Outcome Measures

a. MESA

Self-reported stress-type UI symptoms will be collected using selected items regarding stress and urge-type UI symptoms taken from the questionnaire for the Medical, Epidemiologic, and Social Aspects of Aging Project (MESA) conducted at the University of Michigan.[5] The items query the patient's description of how urine loss occurred. Questions referring to loss of urine at times of exertion such as laughing, sneezing, lifting, or bending over define SUI. Questions referring to urine loss preceded by an urge to void, or uncontrollable voiding with little or no warning define urge incontinence. Symptoms associated with urine loss of both urge and stress-types define "mixed" incontinence. For the purposes of this trial, predominant SUI is defined as the percent of stress-type symptoms > the percent of urge-type symptoms. Agreement between the MESA questions and a clinician's assessment has been reported as 87% for women.[5] The authors further noted that self-reported stress-type symptoms had an accuracy of 69% in predicting a urodynamic diagnosis of SUI.

b. Quality of Life

Health-related QOL is a multidimensional concept which encompasses well-being that is related to health and is distinguished from measures of health and functional status. Measures of general health-related QOL and UI-specific QOL are included.

General health-related QOL will be measured with the Short Form-12 Health Survey. The SF-12 is a comprehensive, psychometrically sound, and efficient way to measure of health status and health-related QOL. Designed for self-administration and completion in less than 3 minutes without assistance, the SF-12 includes 8 concepts commonly represented in health surveys: physical functioning, role functioning physical, bodily pain, general health, vitality, social functioning, role functioning emotional, and mental health. Results are expressed in terms of two meta-scores: the Physical Component Summary (PCS) and the Mental Component Summary (MCS). The SF-12 is a reliable, valid and responsive measure and has been reported to be comparable to that of the SF-36.[6]

A condition-specific QOL measure, i.e. the Incontinence Impact Questionnaire (IIQ-7)[1] will be used in this trial so as to be sensitive enough to detect change. This 30 item instrument assesses the impact of urinary incontinence on activities, social roles and emotional states in women, covering 4 domains, physical activity, social relationships, travel and emotional health. The 30-item measure was condensed into the 7-item short form. Regression analysis confirmed adequate validity and consistency.[7]

c. Severity

The Hunskaar/Sandvik Incontinence Severity Index (ISI) [8] will be used to assess severity. The ISI correlates well with the pad test and will be used as a pad test surrogate in this trial.

d. Patient Global Impression Index – Severity (PGI-S)

The Patient Global Impression Index of Severity [22] [4] question correlated with incontinence episode frequency, stress pad test, and Incontinence Quality of Life Questionnaire results, respectively (all $p < .0001$). This established construct validity of this global assessment questions for baseline severity and treatment response. The PGI-S scale asks patients to that best describe how one's urinary tract condition is now. Response choices are: 1) "normal", 2) "mild", 3) "moderate", 4) "severe". See Appendix B.

e. Patient Satisfaction with Treatment Outcome

Satisfaction with treatment outcome will be measured with a 5-point Likert scale (very dissatisfied to very satisfied) that asks the patient to rate her satisfaction with how the treatment has affected 1) urine leakage;2) urine leakage related to feeling of urgency; 3) urine leakage related to physical activity, coughing or sneezing; and 4) frequency of urination.

f. Voiding Function

Voiding function will be evaluated by a PVR at initial screen. Only subjects with normal PVR's will be candidates for this trial. Postoperatively, any use of a catheter and/or medical therapy to facilitate bladder emptying at or beyond 6 weeks post-surgery, or any surgical therapy to facilitate bladder emptying at anytime post-surgery will be considered voiding dysfunction.

g. Stress Test

At the 12-month visit, a provocative stress test standardized to volume (i.e. 300ml) will be performed for direct observation of urine leakage.[9] Observed urine loss from the urethra coincidental with the Valsalva maneuver or cough is a positive test. To minimize potential bias, the stress test observation should be performed by any qualified examiner who is blinded to the randomization group. This examiner may be an Investigator, physician colleague, fellow, resident, nurse, or clinic personnel who regularly perform this test. The stress test should not be performed by the study surgeon

h. Additional Treatment and Evaluation

Additional treatment or urodynamic evaluation is defined as patient report of any of the following (any such reports require Investigator confirmation):

- Any subsequent surgical procedure for treatment of urinary incontinence; and/or
- Any injection of urethral bulking agents; and/or
- Pelvic floor muscle therapy (PFMT) program ; and/or
- Formal bladder training; and/or
- Medication(s); and/or
- Pessary, if used for treating urinary incontinence.
- Any subsequent urodynamic studies performed – not part of the initial randomization

i. Economic and Cost-utility Measures

Direct Medical Costs. Direct costs of medical care both within the study (i.e. costs associated with surgery) and medical care utilization outside of the study will be assessed. For study-related medical resource use, each component of the intervention will be recorded, including type of surgery, inpatient days, complications/morbidity (any adverse deviation from the normal intra-operative or postoperative course), medications and additional therapies). Intervention costs will be calculated as the product of resource utilization and unit costs. Medication cost will be estimated using the minimum average wholesale price of commonly prescribed medications [34]. Marginal use of resources (provider visits) will be estimated between groups. Direct costs will be calculated using a proxy for societal cost, Medicare resource-based relative value scale charges for physician services [35].

Direct Non-Medical Costs: Incontinence Management Expenses. Participants will record numbers of pads, protection, laundry, dry cleaning, personal hygiene products, household protection, and household cleaning products consumed specifically for their incontinence. They will also record indirect resources related to incontinence (time spent on incontinence-related

healthcare, limits on employment or volunteer work due to incontinence). The Incontinence Resource Use Questionnaire was derived from other published expense surveys and is included in the PRIDE, DAISy and RRISK studies [36-38]. The average national cost of each product will be determined by a survey of several retail and wholesale stores.

Utilities: 1) HUI3. We will assess health-related quality of life (utilities) for health outcomes experienced by the participants using a well-tested patient preference measure, the Health Utilities Index Mark 3 (HUI3) [40].[10] The HUI3 instrument has fifteen multiple choice questions or attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain) with five to six severity levels per attribute, ranging from highly impaired to normal. While none of the domains is directly associated with incontinence, the HUI3 captures areas affected by illness and is independently sensitive to incontinence. These categorical data on health status can be translated into a utility score that reflects global HRQL and will be used in cost-utility analyses. Since the HUI3 is a multi-attribute model validated in a broad cross-section of healthy people, the HUI3 score represents the societal perspective even when administered to individuals with a specific condition (e.g. urinary incontinence).

k. Physician Diagnosis and Management

Prior to randomization, the surgeon will specify the diagnosis and planned treatment based on results of the basic office evaluation and clinical judgment. If randomized to UDS, the surgeon will report any change in diagnosis, change in treatment plan, or modifications of treatment as a result of additional information provided by the UDS. At 3 and 12-months, any further change(s) in treatment management as well as additional urodynamic testing will be reported.

l. Patient Preparedness Assessment

It is possible that subjects undergoing UDS will feel more prepared than those who do not receive UDS. We will measure this preparedness with a preparedness instrument modified from one developed at Loyola. (Appendix C)

H.3 Independent Variables

There are three groups of independent variables:

- Sociodemographic characteristics race; marital status; education; occupation;
- Risk factors for UI: age [11, 12]; parity [13-18]; menopause status and use of HRT [15, 19]; BMI [18-21]; previous pelvic surgery [11, 15]; current bladder and urethral medications [22], comorbidities;
- Type and severity of UI measured by the following: urethral mobility, MESA, Hunskaar/Sandvik Severity Score.

H.4 Intervening Variables

Intervening variables that might affect the outcomes of interest consist of the following categories:

Intraoperative considerations: surgical and medical complications;

Postoperative considerations: time to return to adequate voiding; urinary retention; de novo or persistent urinary urgency or urge incontinence; voiding dysfunction; medical complications;

UI Treatments: including surgery or conservative treatment.

H.5 Physical Exam and History

The following data will be collected from the pre-operative history and physical exam.

a. History & Sociodemographic Information:

- age
- race
- pregnancies (parity, number of vaginal deliveries)
- past pelvic surgeries
- history of estrogen replacement therapy
- duration of incontinence
- smoking history
- past non-surgical incontinence treatments
- medication audit of bladder or urethral drugs only
- frequent UTIs, defined as >3 in previous 12 months

b. Examination:

- provocative stress test at any volume to observe leakage
- height and weight (for BMI); these data will also be collected at 12 months follow-up
- PVR by any method
- Dipstick urinalysis
- Assessment of urethral mobility by one of the following methods: Q-tip test, visualization, palpation, point Aa on POP-Q exam, or lateral cystogram
- A prolapse exam that includes a standing straining exam

H.6 Schedule of Measurements

The schedule of measurements is included in Appendix A. Data will be collected during clinic visits at baseline and 12-months, and by mail/telephone or clinic visit at 3-months. Randomization and Operative data will also be collected.

If more than 12 months transpire between completion of a baseline measure and randomization, any such "expired" measures must be repeated prior to randomization to ensure current eligibility for the trial as well as to obtain current baseline values for critical measures that are subject to change. The one exception is the urine dipstick; this measure expires if a patient is not randomized within 3 months of screening. (See Appendix A.)

I. Patient Safety

Adverse events (AEs) will be reported in accordance with the Department of Health and Human Services (DHHS) code of federal regulations (Title 45, Part 46). As required, a Data and Safety Monitoring Board (DSMB) has been organized. The DSMB is scheduled to meet very 6 months, once by teleconference and once in-person in a 12-month period.

At the bi-annual meetings of the DSMB, summary reports of all reportable AEs will be reviewed. The BCC is responsible for preparing these summary reports. Interim monitoring by the DSMB is also discussed in Section J2. After review, the Executive Secretary of the UITN DSMB is responsible for generating a brief summary report for distribution to all clinical sites, the BCC, and NIDDK. The report will document that a review of data and outcomes across all centers took place on a given date. It will also inform Investigators of the DSMB's conclusion with respect to progress or need for modification of the protocol. Investigators are required to submit a copy of this report to their local IRB. If there is any correspondence with the DSMB

outside of the planned meetings that has study-wide implications, it is the responsibility of the Executive Secretary of the UITN DSMB to draft an appropriate memo for distribution to all clinical sites, the BCC, and NIDDK. Again, Investigators are responsible for submitting a copy of this correspondence to their IRB.

J. STATISTICAL CONSIDERATIONS

J.1 Sample Size Determination for Primary Endpoints

This study is designed as a non-inferiority trial. At the conclusion of the study, it will be determined whether it can be ruled out "beyond a reasonable doubt" that baseline office assessment alone (No UDS arm) is inferior to the baseline office visit plus UDS (UDS arm) by a clinically significant amount. Four possible conclusions will be drawn: 1) No UDS is better than UDS (superiority), 2) No UDS is non-inferior to UDS (non-inferiority), 3) No UDS is inferior to UDS (inferiority), or 4) unable to show that No UDS is either better than or non-inferior to UDS (inconclusive).

It will be assumed that 70% of women in each group will "succeed" where success is defined as a reduction of 70% in UDI scores from baseline to 12 months and a score of 1 or 2 on the PGI (as defined earlier). P_1 is the proportion of women in the UDS group who succeed and P_2 is the proportion of women in the No UDS group who succeed. Delta (δ) or the equivalence margin is the largest clinically relevant difference that would be allowed for the two arms to differ by and still say that the No UDS arm is non-inferior to the UDS arm. This will be defined as 11%. Alpha (α) is the probability of declaring that the No UDS group is not inferior to the UDS group given that the No UDS group is indeed inferior by at least δ. A (1- α)% (95%) two-sided confidence interval will be used for decision-making. Power is the probability of declaring that the No UDS group is not inferior to the UDS group given that the No UDS group is not inferior. The study will be designed to have 80% power.

The sample size calculations were performed under the assumption that the success rates for the two groups are identical. It is also assumed that a test of the difference in proportions using the normal approximation to the binomial will be used to determine non-inferiority. Using a two-sided 95% confidence interval, an equivalence margin of 11% (δ) and that the true proportion of successes in each group is 70%, 270 women are needed in each arm to have 80% power for determining whether No UDS alone is non-inferior to UDS (Hintze, J. (2004). NCSS and PASS. Number Cruncher Statistical Systems. Kaysville, UT). To account for a dropout rate of 10%, a sample size of 300 women per arm (600 women total) will be required.

J.2 Interim Monitoring

In superiority trials, early stopping rules are used to monitor for safety and traditionally would stop a trial if a treatment difference is so large that it is unethical to continue treating the other group of patients with the "inferior" treatment. Interim monitoring of the outcomes for making a decision to end the trial early is not appropriate in this trial for several reasons. In non-inferiority trials, such as this one, the argument of stopping the trial early when there is definitive evidence of non-inferiority does not apply. Say, for example, that early data suggests that the No UDS arm is non-inferior to the UDS arm. Therefore, it is ethical to continue current practice (use of UDS), and one would want extremely strong evidence to abandon a study in favor of changing standard practice.

This trial is comparing two commonly used diagnostic assessment methods, not two treatments. Treatment of patients is a matter of each clinician's clinical judgment. Thus, there would be no need to stop assessing patients by one method or the other due to a difference in treatment outcomes. Also, the more invasive of the two arms is the UDS procedure; however, this is current medical practice and thus any adverse events have been well-documented. The "less invasive" arm is simply an office visit and does not pose any significant threat to the patients.

Finally, the primary endpoint is not collected until 12 months after randomization and it is expected that accrual will take only slightly more than 12 months. Thus, accrual into the trial will be completed before the 12 month visit data are collected and analyzed. Hence, if early stopping were designed into the trial, it would not show any benefit to the patients since they all were already randomized. For these reasons, the trial will not be designed with a planned early look at the data.

However, the trial will be carefully monitored by the DSMB for any unforeseen issues including adverse events, complications, slow recruitment and excess crossovers between the two arms.

J.3 Analytic Approach

To determine whether No UDS is non-inferior to UDS, a two-sided 95% confidence interval for the difference in the proportion of successes according to the primary outcome definition in the UDS group (P_1) and the No UDS group (P_2) ($P_1 - P_2$) will be constructed. The normal approximation to the binomial distribution will be used in its construction. This confidence interval will be two-sided, as is convention in a conservative non-inferiority trial (Pocock, 1983). Decisions regarding non-inferiority will be made according to Figure 1 based on the confidence interval. Basically, if the upper limit of the two-sided confidence interval for the difference is less than the delta limit, only then can non-inferiority for the No UDS group compared with the UDS group be declared. According to Figure 1, this occurs only when the upper end of the confidence limit does not cross the dotted line (the δ value of 11%). In the two inconclusive scenarios in Figure 1 (scenarios 2 and 5), the confidence limit crosses the δ value, so it is not certain that the No UDS arm is truly non-inferior to the UDS arm; in this case, non-inferiority cannot be declared. In scenario 1, both the lower and upper limit of the confidence interval exceed δ; in this case, the No UDS group can be declared inferior to the UDS group.

To prepare for the possibility that the No UDS group is in fact superior to the UDS group, two sequential tests will be done while maintaining the planned overall α level of 0.05 (Dunnett and Gent, Statistics in Medicine, 1996). As described above, the test for non-inferiority will be done and determined according to the two-sided level 1-α confidence interval. If the null hypothesis is not rejected, then we cannot conclude that the No UDS arm is non-inferior to the UDS arm, and no further testing will be done. However, if the null hypothesis is rejected, then the same confidence interval will be used to assess whether the No UDS arm is better than the UDS arm. If the interval does not overlap zero, then the No UDS arm is deemed better than the UDS arm (scenario 3 in Figure 1). If this null hypothesis cannot be rejected, then the No UDS arm is deemed non-inferior to the UDS arm. However, if the null hypothesis can be rejected, then the No UDS arm is deemed better than the UDS arm.

To minimize bias towards non-inferiority, only women treated "per protocol" will be considered in the primary endpoint (Snapinn, Current Controlled Trials, 2000). "Per protocol" is defined as considering only women in the UDS arm who have UDS performed and only women

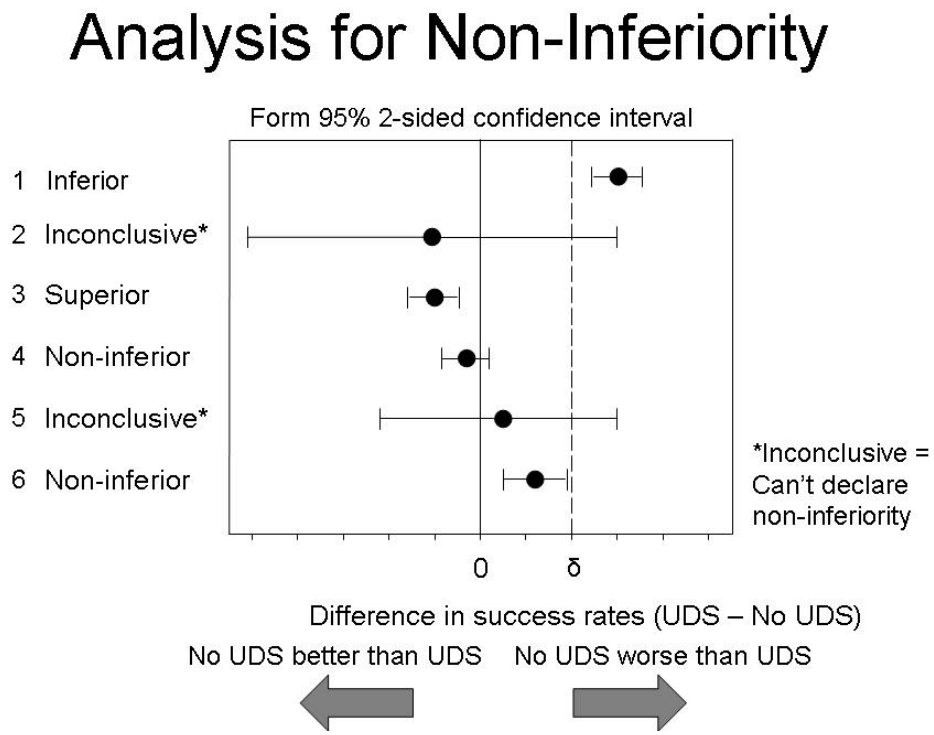
in the No UDS arm who only have an office assessment (and no UDS); e.g. include only those who receive the assessment that they are randomized to in the analysis. Intention-to-treat (ITT) analysis will also be performed, but will be considered secondary.

To assess secondary aim 1, descriptive statistics will be calculated to measure how often physicians alter their clinical decision-making after UDS is performed. Specifically, the rates of altered diagnoses and number of modifications to the initial treatment will be compared using chi-square tests or t tests, as appropriate. For secondary aim 2, we will compare women in the two groups with regard to the two components of the primary outcome: the percent improvement in UDS and PGI-I score. We will compute the mean percent improvement in UDI in the two groups and the 95% confidence interval of the difference in mean improvement. For PGI-I score we will use ordinal logistic regression analysis to estimate the relative odds of a lower score (more improvement) in the PGI-I of the No UDS vs. UDS diagnostic groups and its 95% confidence interval.

An underlying, but untested, premise of this trial is that availability of UDS results can influence how a physician performs the incontinence surgery and therefore the outcome of surgery. To address this important clinical question, we have specified one sub-group analysis to compare the outcomes of women who received surgical treatment in the two randomization groups. For this analysis, a two-sided 95% confidence interval for the difference in the proportion of successes between the two arms will be conducted only among women who received surgery. The definition of success for this comparison is the primary outcome measure plus a negative standard volume stress test.

Additional analyses will be performed comparing the two arms in terms of quality of life, urinary incontinence severity, voiding dysfunction, and other measures as necessary. Multivariable linear regression will be used for continuous measures and logistic regression for dichotomous measures controlling for relevant baseline characteristics.

Figure 2. Examples of Scenarios for Non-Inferiority Trials with Trial Conclusions



J.4 Cost Analysis

For a non-inferiority trial, the primary outcome of the trial is not expected to differ. In this case, with the effectiveness outcome of the cost-effectiveness analysis expected to be the same, the cost-effectiveness analysis can reduced to a cost-minimization analysis where the study of interest is the total cost difference between women who receive pre surgical UDS and those who do not. The less costly assessment will be a preferred choice from the cost standpoint. The study focus will be the incremental cost. That is to say, we are interested in the cost components that are different between comparison arms. Those cost components that are likely to be the same between comparison arms will have zero cost increments; therefore, their information will not be critical to collect if the study budget is the concern.

For direct cost, we will follow the patients for 12 months and collect the number of utilizations for hospitalizations, emergency-room visits, outpatient care, physician visits, laboratory tests, and prescribed medications that are related to their incontinence problem or the assessment they received. Standard cost (DRG, CPT, and AWP) will be applied to each utilization.

Cost-utility and cost-effectiveness analyses will be performed to compare the two groups. For cost-utility analyses, we will estimate net cost and net QALY in each arm of the model and calculate the marginal cost per quality adjusted life year (QALY) between the groups. For cost-effectiveness analyses, we will determine the marginal cost between the groups per treatment outcome.

K. INFORMED CONSENT PROCEDURES

Patients will undergo clinical assessment for type of urinary incontinence and candidacy for surgery. If a patient appears to be eligible for the study and desires surgical treatment of SUI, she will be approached about the study. After a verbal explanation of the study, the patient will be presented with the informed consent form by study personnel, either the surgeon or research nurse. Written informed consent will be obtained prior to proceeding with any further eligibility determination, testing, or data collection activities.

Templates for the informed consent forms will be used by all of the centers, modifying the content or format as necessary to meet the requirements of their respective institutional human subjects committees.

L. DATA MANAGEMENT

A customized data management system (DMS) has been developed for this study using the Biostatistical Coordinating Center's Web-based Advanced Data Entry and Protocol Tracking (ADEPT) system. ADEPT integrates all aspects of study data collection including: participant screening and enrollment data; real time accrual reporting; tracking of participant study appointments for follow-up; all study follow-up data collection; censoring/loss to follow-up data; and monitoring timeliness and quality of data collection. Data are stored in an Oracle relational database at the BCC. CTCs enter data via a customized ADEPT secure web application that provides real-time field level validations and context sensitive help. HTML data entry forms are enhanced with client side JavaScript code to ensure rapid entry of study forms, proper validations of all data fields and proper skip patterns within study data forms. Interim background data submittals prevent loss of data due to loss of internet connections.

L.1 Data Entry

ADEPT includes a number of standard features designed to ensure consistently high quality data. Integrated into the data entry system are real time validations, including both inter- and intra-instrument data checks. Inconsistent or questionable values are flagged during entry, and an edit report is automatically generated. These edit reports are designed to be human readable, listing the participant ID, the instrument name, and a detailed description of why each specific data item was flagged. The edit report can be printed out and reviewed by a supervisor or returned to the data collector for resolution. The ADEPT system also tracks missing data rates by instrument and data collector. Data entry quality is monitored through a sample-based double data entry system. A self-adjusting algorithm is used to enforce a higher double data entry rate on keyers who have higher error rates.

L.2 Data Security and Integrity

The web based components of the data management system utilize several levels of security to ensure privacy and integrity of the study data:

- Web Access requires use of assigned user names and passwords.

- Passwords will be changed every 90 days.
- Web based entry uses secure socket layer data encryption.
- Firewall protects against unauthorized access to study data.

All study data will be stored on the BCC's NT based Oracle server. Access to data on this server (from both inside and outside the data center) is controlled by Oracle's extensive security features. Oracle archiving and backup system ensures minimal data loss even in the most catastrophic system failure. Backup data tapes are stored off site in a nearby bank safety deposit box.

M. QUALITY CONTROL ACTIVITIES

The BCC has primary responsibility for QC/QA activities. The BCC also requires that the sites complete certain QC activities, most of which are monitored by the BCC.

The key QC/QA activities are:

- Development of a study Manual of Operations;
- Clearly formatted and carefully constructed Data Forms, clearly linked to the research questions and research models, with clear, up-to-date manuals of instruction;
- Sign-Off Procedures for all study forms;
- Central training and certification of all CTC data collection staff with the use of standardized checklists;
- Central training and certification of CTC data managers;
- Verification of patient eligibility;
- On-going monitoring of all protocols/data collection activities;
- Completion of reliability and/or pilot studies for *key measurements* as appropriate;
- Inclusion of repeat measurements, as feasible, in the course of the study; and
- Site visits to CTCs with pre-specified goals.

REFERENCES

1. 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.
2. Wagner TH, Patrick DL, Bavendam TG, Martin ML, Buesching DP. Quality of life of persons with urinary incontinence: development of a new measure. Urology 1996; 47:67-71; discussion 71-2.
3. Patrick DL, Martin ML, Bushnell DM, Yalcin I, Wagner TH, Buesching DP. Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL). Urology 1999; 53:71-6.
4. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. Am J Obstet Gynecol 2003; 189:98-101.
5. 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.
6. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996; 34:220-33.
7. Uebersax JS, Wyman JF, Shumaker SA, McClish DK, Fantl JA. Short forms to assess life quality and symptom distress for urinary incontinence in women: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program for Women Research Group. Neurourol Urodyn 1995; 14:131-9.
8. Hunskaar S, Vinsnes A. The quality of life in women with urinary incontinence as measured by the sickness impact profile. J Am Geriatr Soc 1991; 39:378-82.
9. Shull BL, Halaska, M., Hurt, G., Kinn, A., Laycock, J., Palmtag, H., Reilly, N., Yong Yang & Zubieta, R. Physical Examination. In: Abrams P, Saad, K. & Wein, A., ed. Incontinence: Proceedings of the 1st International Consultation on Incontinence - June 28-July 1, 1998 - Monaco. St. Helier, England: Health Publications Ltd., 1999.
10. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. Health Utilities Index. Pharmacoeconomics 1995; 7:490-502.
11. Milsom I, Ekelund P, Molander U, Arvidsson L, Areskoug B. The influence of age, parity, oral contraception, hysterectomy and menopause on the prevalence of urinary incontinence in women. J Urol 1993; 149:1459-62.
12. Yarnell JW, Voyle GJ, Richards CJ, Stephenson TP. The prevalence and severity of urinary incontinence in women. J Epidemiol Community Health 1981; 35:71-4.
13. Burgio KL, Locher JL, Zyczynski H, Hardin JM, Singh K. Urinary incontinence during pregnancy in a racially mixed sample: characteristics and predisposing factors. Int Urogynecol J Pelvic Floor Dysfunct 1996; 7:69-73.
14. Foldspang A, Mommsen S, Lam GW, Elving L. Parity as a correlate of adult female urinary incontinence prevalence. J Epidemiol Community Health 1992; 46:595-600.
15. Jolleys JV. Reported prevalence of urinary incontinence in women in a general practice. Br Med J (Clin Res Ed) 1988; 296:1300-2.
16. Sommer P, Bauer T, Nielsen KK, et al. Voiding patterns and prevalence of incontinence in women. A questionnaire survey. Br J Urol 1990; 66:12-5.

17. Thomas TM, Plymat KR, Blannin J, Meade TW. Prevalence of urinary incontinence. Br Med J 1980; 281:1243-5.
18. Yarnell JW, Voyle GJ, Sweetnam PM, Milbank J, Richards CJ, Stephenson TP. Factors associated with urinary incontinence in women. J Epidemiol Community Health 1982; 36:58-63.
19. Burgio KL, Matthews KA, Engel BT. Prevalence, incidence and correlates of urinary incontinence in healthy, middle-aged women. J Urol 1991; 146:1255-9.
20. Brown JS, Seeley DG, Fong J, Black DM, Ensrud KE, Grady D. Urinary incontinence in older women: who is at risk? Study of Osteoporotic Fractures Research Group. Obstet Gynecol 1996; 87:715-21.
21. Mommsen S, Foldspang A. Body mass index and adult female urinary incontinence. World J Urol 1994; 12:319-22.
22. Montella AJ, Wordell, C.J. The effects of drugs on the lower urinary tract. In: Ostergard D, Bent, A., ed. Urogynecology and urodynamics. Baltimore: Williams and Wilkins, 1996.

Appendix A. Schedule of Measurements

	Screen	Baseline	Operative	3 months	12 months
UDI	✓**			✓	✓
PGI-I				✓	✓
PGI-S		✓**		✓	✓
Stress Test*	✓*,**				✓*
UDS		✓†,**			
Additional Treatment‡				✓	✓
SF-12		✓**		✓	✓
IIQ-7		✓**		✓	✓
ISI		✓**		✓	✓
MESA	✓**			✓	✓
Voiding Function				✓	✓
Patient Satisfaction				✓	✓
Patient Preparedness		✓**			
Adverse Events		✓	✓	✓	✓
Urine Dipstick	✓***				
Costs & Utilities		✓**			✓
Sociodemographic		✓**			
Comorbidities		✓**		✓	✓
H&P	✓**	✓**			✓
PVR	✓**				✓
Medication Audit		✓**		✓	✓
Prolapse Assessment	✓**				
Urethral Mobility	✓**				
Operative Measures			✓		
MD Dx & Tx Info.		✓**	✓	✓	✓

* Screening Stress Test is <u>not</u> standardized to volume; Follow-Up test is standardized (300ml).

† UDS performed per randomization assignment.

** Measure must be repeated if patient is not randomized within 12 months of completion.

*** Measure must be repeated if patient is not randomized within 3 months of completion.

‡ Patient report of additional treatment(s) requires Investigator confirmation.

Appendix B: PGI-I and PGI-S

Patient Global Impression of Improvement (PGI-I) Scale

Check the one number that best describes how your urinary tract condition is now, compared to how it was before you received treatment for your urinary leakage

1. Very much better
2. Much better
3. A little better
4. No change
5. A little worse
6. Much worse
7. Very much worse

Patient Global Impression of Severity (PGI-S) scale

Check the one number that best describes how your urinary symptoms are now

1. Normal
2. Mild
3. Moderate
4. Severe

Appendix C. Patient Preparedness Instrument

Instructions:

We are interested in learning more about how prepared women feel for pelvic floor surgery and how this relates to their surgical experience. Your doctors will not see the answers to these questions until after your surgery is completed, so please ***ask*** your doctor any questions that you still have. The results of this questionnaire will be kept confidential. Please place an "X" under the most appropriate box.

		Strongly Agree	Agree	Somewhat Agree	Somewhat Disagree	Disagree	Strongly Disagree
1	I know about the *alternatives* to the planned surgery.						
2	I understand the *purpose* of the planned surgery (what this surgery can accomplish).						
3	I understand the *benefits* of the planned surgery (How this surgery should help me).						
4	I understand the risks of the planned surgery (what are the chances of something not going the way my doctor and I want it to go).						
5	I understand the *complications* of the planned surgery (what problems can come from this surgery)						
6.	I feel prepared about what to expect after surgery while I am in the hospital						
7.	I feel prepared about what to expect after surgery when I am at home.						
8.	I feel prepared to cope with a catheter after the surgery while I am in the hospital						
9.	I feel prepared to cope with a catheter after the sugery when I am at home						
10.	My doctors and nurses have spent enough time preparing for my upcoming surgery						
11.	Overall, I feel prepared for my upcoming surgery						