Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN)

Phenotyping Study Protocol

Version 8.0

Steering Committee Approval Dates

Version 1.0: April 18, 2014

Version 2.0: July 18, 2014

Version 3.0: September 5, 2014

Version 4.0: September 29, 2014

Version 5.0: January 19, 2015

Version 6.0: March 10, 2015

Version 7.0: June 5, 2015

Version 8.0: June 3, 2016

The Phenotyping Study Protocol, previously called the Prospective Observational Cohort Study Protocol was approved by NIH on August 1, 2014.

1 **Principal Investigator Signature Sheet**

Protocol: LURN Phenotyping Study Protocol V8.0	Approval Date: June 3, 2016
IND: N/A	**LURN DCC Principal Investigator :** Robert Merion, MD
Study Sponsor: The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)	
INSTRUCTIONS: The Principal Investigator must print, sign, and date below. The original signature page should be kept in the site's records. After signature, please scan the signature page and email to the LURN DCC at the address listed below: LURN DCC LURN-Monitors@ArborResearch.org	
I confirm that I have read the above protocol in the latest version. I understand it, and I will work according to the principles of Good Clinical Practice (GCP) as described in the United States Code of Federal Regulations (CFR) – 21 CFR Parts 45, 50, 56, and 312, and the International Conference on Harmonization (ICH) document "Guidance for Industry: E6 Good Clinical Practice: Consolidated Guidance" dated April 1996. Further, I will conduct the study in keeping with local, legal, and regulatory requirements. As the Principal Investigator, I agree to conduct and to carry out the study by the criteria written in the protocol and understand that no changes can be made to this protocol without written permission of the LURN Steering Committee.	
Site Principal Investigator (Print)	
Site Principal Investigator (Signature)	
Date	

Table of Contents

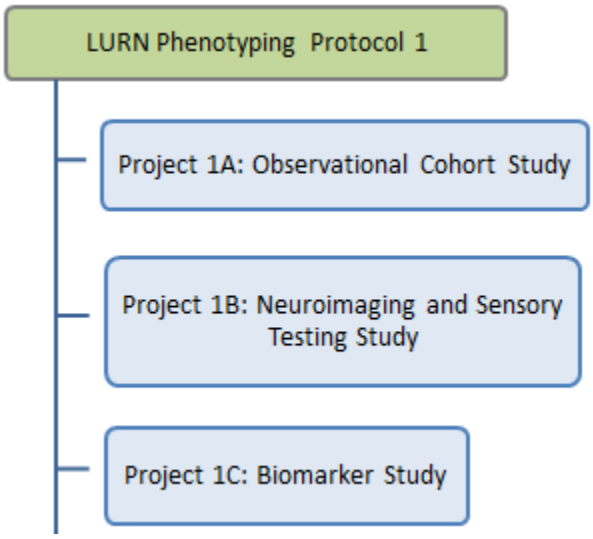
1 Introduction and Overview of LURN Phenotyping Studies

The Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN) is an NIH/NIDDK sponsored research network of six research sites and a Data Coordinating Center (DCC). The long term goals of LURN are to establish an interdisciplinary team of researchers to work collaboratively to increase our understanding of lower urinary tract symptoms (LUTS) by (1) identifying and explaining the important subtypes (phenotypes) of patients with LUTS, (2) improving the measurement of patient experiences of LUTS, and (3) disseminating data, research tools, and biological samples to the research and clinical communities.

Over the course of several years, LURN will conduct clinical studies "to phenotype" LUTS. In the biological sciences, *phenotype* typically refers to the observable characteristics of a person— physical, behavioral, biochemical—as determined by genetic and environmental influences. The "phenotyping" effort in LURN seeks a description of both the observable characteristics of the patient with LUTS, as well as an explanation for why those characteristics are observed in some people and not others.

LURN is pursuing phenotyping research using distinct, but related, projects. Phenotyping Protocol 1 is the overarching effort, and will be divided into three projects (see **Figure 1**). Project 1A will be a large-scale accrual of LUTS patients into a registry. Standardized clinical data, comprised of information typically gathered at the patient clinic encounter, will populate the registry. Using these data, subgroups of patients will be identified for further, more focused and in-depth study. This more focused effort will be conducted as Projects 1B and 1C.

Figure 1: Overview of LURN Phenotyping Protocol Structure



2 LURN and Phenotyping Efforts

Definition of phenotyping in the LURN. In order to improve the care and treatment outcomes of patients with LUTD, it is necessary to better characterize these patients through identification of clinically meaningful subtypes. This phenotyping effort is intended to improve understanding of important differences among patients at several levels including (a) the experience of LUTS, (b) the physical state of the organism, (c) genitourinary (GU) organ system/tissue, and (d) cells/molecules. At any one of these levels, clinically relevant differences might exist among patients with LUTD. Furthermore, the LURN phenotyping effort is intended to explain why differences among patients are observed at one level (e.g., GU organ system/tissue) by linking those differences to differences among patients at another level (e.g., cells/molecules). The explanations that link factors at one level with factors at another level are grounded in mechanistic theories about biological, behavioral, and environmental influences on the person. In the biological sciences, "phenotype" typically refers to the observable characteristics of a person— physical, behavioral, biochemical—as determined by genetic and environmental influences. The phenotyping effort in LURN seeks a description of both the observable characteristics of the person as well as an explanation for why those characteristics are observed in some people and not others.

3 Project 1A: Prospective Observational Cohort Study

3.1 Overview

The Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN) was established by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) to advance our understanding of lower urinary tract dysfunction[1] (LUTD) in women and men. LUTD is a term intended to be comprehensive and to challenge current paradigms about how symptomatic pelvic disorders are defined as 'diseases.' Lower urinary tract symptoms (LUTS)[2] are likely caused and exacerbated by a variety of factors and thus do not represent the manifestation of a single disease. Clinical management of LUTD, including treatment outcomes, remains suboptimal since the biological and psychosocial factors that initiate, exacerbate, and modify this group of symptoms remain largely unknown. As an initial effort to better characterize the biological and psychosocial factors that initiate, exacerbate, and modify LUTS, the LURN investigators will establish a prospective Observational Cohort Study of men and women with LUTS presenting for the first time to LURN physicians. This prospective Observational Cohort Study will be Project 1A of the Phenotyping Protocol.

Information to be obtained from study participants initially (at time of enrollment) includes a standardized clinical examination, medical history, select testing of the lower urinary tract, and participants' self-report of LUTS, pelvic floor symptoms including sexual function and bowel symptoms, depression, anxiety, sleep patterns, stress, metabolic risk factors, and health-related quality of life. We will also collect serum, urine, saliva, and perineal swabs from men and vaginal swabs from women for storage at the NIDDK Sample Repository for future study by the LURN investigators and the broader research community. This information will be used to construct subgroups of patients who have similar symptoms, clinical presentations, comorbidities, pelvic floor dysfunctions, and psychological profiles. These patient characteristics and behaviors likely affect the evaluation, diagnosis, and/or treatment of

[1] Lower urinary tract dysfunction is any disturbance or abnormality of function of the lower urinary tract. The ICS indicates that a dysfunction is accompanied by an observed sign.

[2] Lower urinary tract symptoms are defined by the International Continence Society as subjective reports of an experience that may lead person to seek care from health care professionals. Lower urinary tract symptoms can also indicate pathologies other than lower urinary tract dysfunction. Abrams: Neurourology and Urodynamics 21:167 (2002)

LUTS. Additional information will be collected 3 months and 12 months after enrollment or 3 and 12 months after surgery for patients receiving surgical treatment, and will include an interval clinical history, participants' self-report of LUTS, pelvic floor symptoms including sexual function and bowel symptoms, depression, anxiety, and health-related quality of life. We will also collect biological samples at 3 and 12 months after enrollment.

The long-term goal of the LURN is to better characterize patients with LUTD in order to advance future research on the pathophysiology of these disorders and improve clinical management. The information to be collected during Project 1A will be limited and not sufficient to fully understand the pathophysiology and biology of LUTS. Therefore, the Observational Cohort Study will serve as the basis for an additional LURN study, Project 1B, which is described further in section 4.

3.2 Background, Study Rationale

LUTD affects a large proportion of US men and women, with prevalence increasing with patient age. As examples, the prevalence of non-stress urinary incontinence increases from 4%-5% among women in their 30s to 10%-16% among women in their 60s. Among male Medicare beneficiaries, over 24,000 per 100,000 outpatient office visits listed benign prostatic hyperplasia (BPH) as a relevant diagnosis. These diagnoses significantly affect physical and mental health. Nationally, female urinary incontinence accounted for approximately $2.4 billion in expenditures and BPH accounted for nearly $2 billion in expenditures as recently as 2006. LUTD comprises these conditions as well as additional urinary dysfunctions. Thus, LUTD is an important and impactful public health condition.

Despite its substantial prevalence and resulting effect on public health, there are several challenges in the clinical management of LUTD. LUTD comprises a heterogeneous symptom complex, and patients often have mixed combinations of symptoms. Bladder outlet obstruction, detrusor hypotonicity, and storage LUTS often coexist. Pharmacologic interventions that target these symptoms can have adverse effects that are disproportionately impactful on older patients who are at increased risk for LUTD.[1] In addition, some patients with LUTS will have a cause for their symptoms other than dysfunction of the lower urinary tract, such as nocturnal polyuria.

Population-based epidemiological studies have characterized the prevalence of LUTS and categorized study participants into common symptom profile clusters. From the Boston Area Community Health (BACH) Survey, the European Prospective Investigation into Cancer and Nutrition (EPIC) Study, and the Epidemiology of LUTS (EpiLUTS) study, up to 70% of men and 76% of women have more than minimal LUTS. The symptom clusters derived from these studies categorized patients by predominance and severity of self-reported urinary symptoms. For example, women in EpiLUTS were subdivided into those with one reported symptom, those bothered by stress urinary incontinence, those with urinary urgency, those with terminal dribbling, those with nocturia, and those with mixed urinary symptoms.[1]

Whether these findings are relevant to the patient population that seeks care for their LUTD is not known. Furthermore, the urinary symptom clusters derived from EpiLUTS have not been associated with patient demographic and clinical factors or clinical outcomes that could render these definitions useful in clinical practice. Patients seeking clinical care of their LUTS – in contrast to persons from the community who respond to a survey study – can be heterogeneous and may present with more than four or five profiles of clinically relevant LUTD symptom clusters and likely experience greater bother of their symptoms. As such, patient clusters derived from epidemiological studies may not inform the clinical care of men and women with LUTD. Thus, patients are treated based on anecdote and clinician experience rather than the best available evidence. Furthermore, attempts to classify patients into *obstructive* versus *irritative* or *storage* versus *voiding* categories have shown that few patients fall neatly into either category. For the clinician managing these presenting symptoms, this lack of clarity can be

confusing. Improved disease classifications are needed that better predict future LUTD patient management strategies, outcomes and treatment response.

In this Observational Cohort Study, we will characterize lower urinary tract symptoms using questionnaires. One is the LUTS Tool, which was used in the EpiLUTS study.[1] Another is the Comprehensive Assessment of Self-Reported Urinary Symptoms (CASUS, Appendix C), which was recently developed by LURN investigators to provide a comprehensive assessment of LUTS for phenotyping. In order to refine the CASUS we will evaluate item performance and the validity of scores generated from this instrument. This will result in the elimination of items that have poor variability, are redundant with other items, and/or demonstrate low convergent validity. This process requires administering the CASUS to a large number of men and women with LUTS enrolled in the Observational Cohort Study.

3.3 Study Objectives

The primary objective of this prospective Observational Cohort Study is to characterize a large, geographically diverse group of care-seeking men and women with LUTS in order to identify *clinically relevant* subgroups of patients with similar symptoms, clinical presentations and other factors relevant to LUTD and its treatment. Identification of patient subgroups will also likely guide enrollment and aid in analysis and interpretation of future LURN targeted phenotyping studies.

The aims of the prospective Observational Cohort Study are:

Aim 1: Based on cross-sectional data, characterize urinary symptoms, demographic and clinical characteristics, health-related quality of life, self-reported pelvic floor function (bowel function, sexual function, and pelvic organ prolapse) and psychological factors (stress, anxiety, depression, sleep disturbance) of men and women seeking care for LUTS.

Hypothesis 1a: Health-related quality of life and sexual function will be poorer, and the prevalence of depression, anxiety, bowel disorders, levels of stress and sleep disturbances will be greater in men and women with more severe and more bothersome lower urinary tract symptoms.

Hypothesis 1b: Health-related quality of life, sexual function, bowel function, and the prevalence of depression, anxiety, levels of stress and sleep disturbances will vary by chief urinary care-seeking complaint provided by the study participant.

Hypothesis 1c: Self-reported pelvic floor function (bowel function, sexual function, and pelvic organ prolapse), psychological factors (stress, anxiety, depression) and health-related quality of life will differ among patients with LUTS with and without urinary incontinence.

Hypothesis 1d: Urinary symptoms, clinical assessments by LURN physicians, pelvic floor function (bowel function, sexual function, and pelvic organ prolapse) and psychological factors (stress, anxiety, and depression) will differ in subgroups of individuals seeking care for symptoms of LUTS stratified by:

a) sex (i.e., women will report more severe urinary incontinence and have different associations between urinary symptoms and bowel function, sexual function, and psychological factors than men);
b) age (i.e., older patients with LUTS will report more severe urinary symptoms, bowel symptoms, and sexual dysfunction, have higher prevalence of psychological factors, and have higher post-void residual volumes than younger patients);
c) race/ethnicity (i.e., racial/ethnic subgroups may have variable prevalence of LUTS, pelvic

floor dysfunction and psychological factors);

d) first-degree family history (i.e., patients with first-degree family members diagnosed and/or treated for LUTS will report more severe urinary symptoms);
e) presence or absence of diabetes mellitus (i.e., diabetic patients with LUTS will report more comorbid conditions, more severe and bothersome urinary symptoms, greater bowel and sexual dysfunction, higher prevalence of psychological factors and poorer health-related quality of life than non-diabetics);
f) presence or absence of obesity and other metabolic risk factors (i.e., higher BMI, history of cardiovascular diseases, hypertension, or hyperlipidemia will be associated with more severe and bothersome urinary symptoms, greater bowel and sexual dysfunction, poorer health-related quality of life and higher prevalence of psychological factors).

Hypothesis 1e: Self-reported pelvic floor function, psychological factors and health-related quality of life will differ among patients with LUTS with predominant storage (urgency/frequency) urinary symptoms versus those with predominant voiding (hesitancy/slow flow) urinary symptoms.

Hypothesis 1f: The clinical impression/diagnosis and treatment plan of LURN physicians will be associated with patient responses on the LUTS Tool.

Aim 2: Based on cross-sectional data, identify distinct subgroups (clusters) of study participants based on their urinary symptoms assessed by the LUTS Tool, CASUS, clinical assessments, pelvic floor function and psychological factors utilizing cluster analysis and classification and regression trees (CART).

Hypothesis 2a: Distinct clusters of study participants can be identified that will differ in urinary symptoms, results from clinical assessments, patient-reported pelvic floor function, and psychological factors.

Hypothesis 2b: Urinary symptoms within these LURN clusters will have higher prevalence of mixed and more severe urinary symptoms compared with clusters identified in population-based studies.

Hypothesis 2c: Clusters of study participants derived from the LURN observational cohort will exhibit unique results on targeted phenotyping.

Exploratory Question 2a: Will the number and types of symptom clusters identified from CASUS be similar to the number and types identified by the LUTS Tool?

Aim 3: To prospectively assess the treatments recommended by LURN physicians and associated changes in urinary symptoms and urinary quality of life in men and women seeking care for LUTS.

Hypothesis 3a: Symptom changes, as determined by the LUTS Tool and CASUS, and stratified by the selected treatments, will be associated with specific subgroups of study participants, including those defined at study entry by age, sex, presence of diabetes, and presence of obesity.

Hypothesis 3b: Symptom changes, as determined by the LUTS Tool stratified by the selected treatments, will be associated with pelvic floor function and psychological factors determined at study entry.

Hypothesis 3c: Patient clusters developed in Aim 2 will be associated with initial treatment selection by LURN physicians and response to LUTD-specific treatments as measured by the LUTS Tool.

Hypothesis 3d: Patients that respond to LUTD-specific treatment will move from one LURN symptom cluster to another as a result of lessening of symptom severity.

Aim 4: To evaluate the completion, response variability, and potential overlap of the CASUS items.

Exploratory Question 4a: How do rates of missing items for the CASUS compare with rates for other self-report questions administered in the Observational Cohort Study?

Exploratory Question 4b: Does each CASUS item demonstrate variability across the sample, or are there floor or ceiling effects?

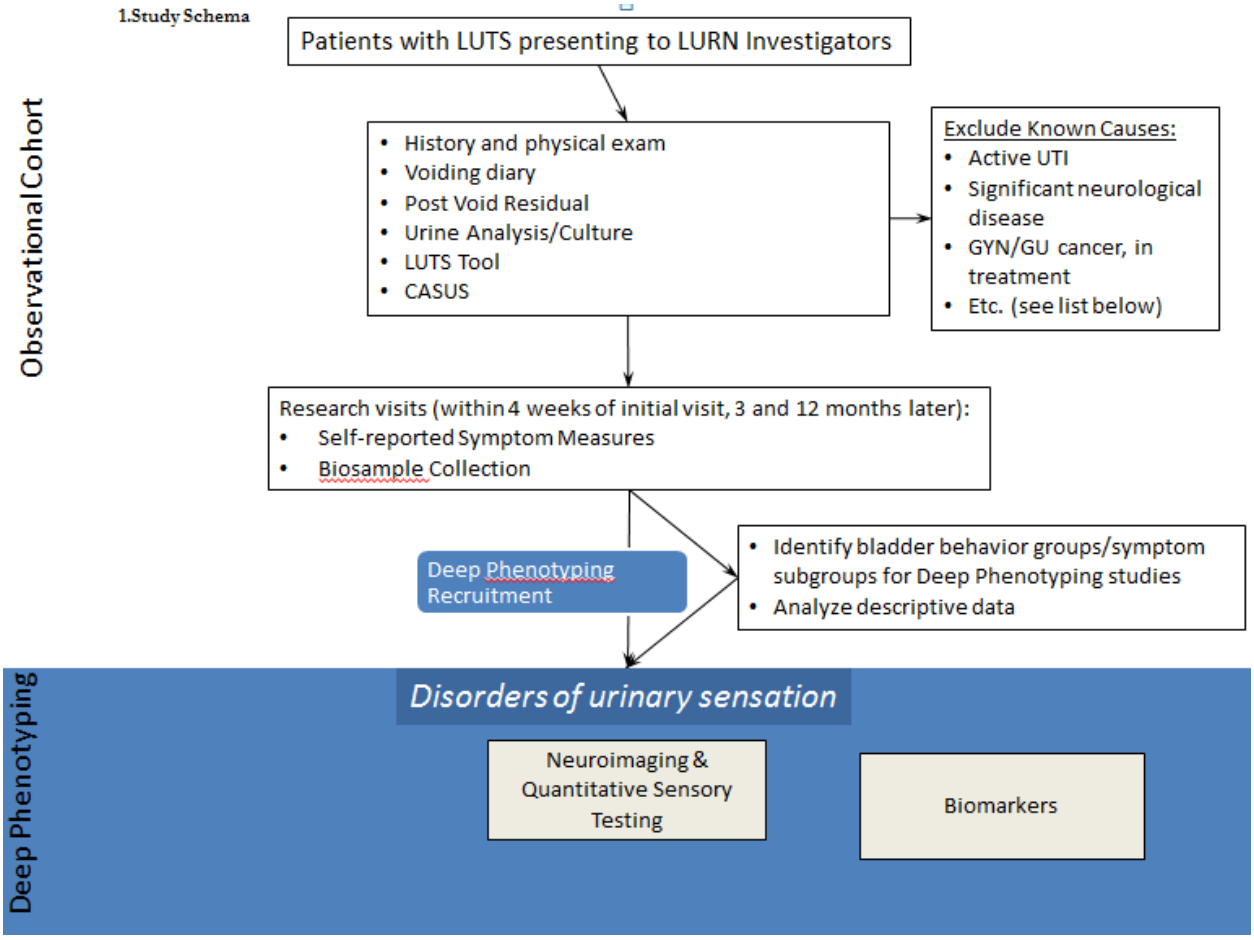
Exploratory Question 4c: Are any CASUS items so highly correlated ($r > .90$) that they are essentially measuring the same thing?

Aim 5: To determine the associations between CASUS items and corresponding items from the LUTS Tool.

Hypothesis 5a: CASUS items will have strong correlations ($r \geq .70$) with corresponding items from the LUTS Tool.

3.4 Methods

3.4.1 Study Schema



3.4.2 Study Methods

This study is a prospective observational study of patients with LUTS presenting for clinical care to one of the LURN physicians. We will collect routine clinical and demographic patient information and validated, self-reported outcome measures, including information on LUTS, pelvic floor symptoms (sexual, bowel, prolapse), health-related quality of life and psychosocial symptoms (anxiety, depression, stress, sleep disturbance) at presentation. Study participants will complete follow-up assessments three and twelve months after their initial assessment to evaluate the trajectory of their symptoms in the context of the treatments they received. Biosample collection will be coordinated with these follow up visits.

3.4.3 Enrollment

Patients with LUTS presenting to LURN clinical sites will be screened for participation based on the inclusion and exclusion criteria (below). We will collect reasons for exclusion of screened patients. Unless listed below, prior or ongoing treatments for LUTD will not preclude patients for participation (i.e., included patients do not necessarily have to be treatment naïve). Eligible patients will be invited to participate in the study. We will collect reasons for non-consent of eligible patients. Consenting participants will complete the self-reported demographic and symptom measures and a 3-day urinary diary before starting any new treatment prescribed by LURN physicians. Standard clinical data (detailed in section 3.4.7) will also be collected. All enrolled patients will be asked for permission to be re-contacted for future participation in other LURN studies.

3.4.4 Participant Selection

Inclusion criteria:

a. Women presenting for new patient visits for evaluation or treatment of LUTS to one of the LURN physicians.
b. Men presenting for new and returning patient visits for evaluation or treatment of LUTS to one of the LURN physicians.
c. Age ≥ 18 years.
d. The presence of any of the symptoms reported in Table 1, based on responses to the LUTS Tool with a one month recall period (Appendix A1).
e. The ability to give informed consent and complete self-reported questionnaires electronically.

Table 1: LUTS appropriate for study inclusion

Symptom Cluster	Symptom
Storage	Daytime frequency
	Nocturia
	Urgency
	Incontinence/leakage (various types)
	Poor or absent sensation of bladder filling
Voiding	Slow/weak stream
	Splitting or spraying
	Intermittent stream/Double voiding
	Hesitancy
	Straining
	Dribbling at the end of flow
	Paruresis (i.e., shy bladder, shy bladder syndrome)
	Poor or absent sensation of urethra during void
Post-micturition	Feeling of incomplete emptying
	Post-micturition dribble (delayed)
Other or Poorly Characterized	Abnormal bladder or urethral sensations

Deferral criteria:

a. Microscopic hematuria
    - Patient must undergo appropriate evaluation.
b. Positive urine culture.
    - Patient needs to be treated and have a subsequent negative culture before he or she is eligible.
c. Recent (within 6 months) pregnancy.
d. Current sexually transmitted infection.
    - Patient needs to be treated and have a subsequent test before he or she is eligible.

Exclusion criteria:

a. Gross hematuria.
b. Significant neurologic disease or injury, including but not limited to: cerebral vascular accident with residual defect, Alzheimer's dementia, Parkinson's disease, traumatic brain injury, spinal cord injury, complicated spinal surgery, multiple sclerosis.
c. Primary complaint is pelvic pain.
d. Diagnosis of interstitial cystitis, chronic prostatitis, or chronic orchalgia.
e. Pelvic or endoscopic GU surgery within the preceding 6 months (not including diagnostic cystoscopy).
f. Ongoing symptomatic urethral stricture.
g. History of lower urinary tract or pelvic malignancy.
h. Current chemotherapy or other cancer therapy.
i. Pelvic device or implant complication (e.g., sling or mesh complications).
j. Current functioning neurostimulator.
k. Botox injection to the bladder or pelvic structures within the preceding 12 months.
l. In men, prostate biopsy in the previous 3 months.
m. In women, pregnancy.
n. History of cystitis caused by tuberculosis, radiation therapy, or Cytoxan/cyclophosphamide therapy.
o. Augmentation cystoplasty or cystectomy.
p. Presence of urinary tract fistula.
q. Current major psychiatric disorder or other psychiatric or medical issues that would interfere with study participation (e.g., dementia, psychosis, etc.).
r. Inability to relay valid information, actively participate in the study, or provide informed consent (includes uncontrolled psychiatric disease).
s. Difficulty reading or communicating in English.

In addition to the criteria listed above, pregnancy during the study will be a study end point.

Limitations of the LURN Observational Cohort

Patients will be recruited from LURN clinical sites, which may represent, in part, patients that seek care for bothersome LUTD. The LURN clinical sites (section 7.1) include urology and urogynecology practices at university-based clinical settings that care for general LUTD patients visiting a specialist for the first time, complex LUTD patients who have been referred from other specialists, as well as established male patients who present with persistent or worsening LUTD. Thus, findings from the LURN Observational Cohort Study may not be generalizable to LUTD patients presenting to primary care physicians and may over-represent treatment refractory LUTD. Preliminary survey of LURN sites reviewed the diversity of practice types within the network. Investigators estimated that one in four patients seen at LURN practice sites had no prior treatment, most patients had been treated for LUTD for less than five years, and about one in five had longstanding LUTD (treated for more than 5 years). Investigators noted that most of their referrals come from primary care providers. Thus, the LURN Observational Cohort Study will include a substantial portion of treatment naïve patients.

3.4.5 Schedule of Visits

Patients will be screened for eligibility and approached for study participation during their initial visit to LURN physicians. As part of screening, patients will complete the LUTS Tool with a one month recall period (Appendix A1) at the initial visit. For patients screened but not enrolled, we will collect reasons

for exclusion and patient demographic information. Consenting participants will be asked to complete self-reported demographic and symptom measures and a 3-day urinary diary (within 4-weeks of the initial visit and before initiating treatment). The LUTS Tool with a one week recall period (Appendix A2) will be administered to patients along with CASUS and other questionnaires at the baseline visit. If needed, a research coordinator will arrange a separate baseline visit to facilitate completion of the initial survey and biosample collection. At this baseline assessment, the intake questionnaire will be completed using an online module. Participants who are not comfortable using computers will be given the option to complete the questionnaires on paper. The LUTS Tool will be administered twice in one day (first with a one month recall period during screening and then with a one week recall period as part of the baseline questionnaires) if a participant's initial and baseline visits take place on the same day (see Table 2 below).

Table 2: Schedule of Visits

	Initial Visit	Initial and/or Baseline Visit	3 Month Follow-up Visit	6 Month Assessment	9 Month Assessment	12 Month Follow-up Visit
Eligibility Assessment	X					
Demographics	X					
General Clinical Information	X					
Physical Exam Findings	X					
Clinic Testing (Urine Analysis)	X					
LUTS Tool (one month recall period)	X					
LUTS Tool (one week recall period)		X	X			X
CASUS	X					X
3-Day Voiding Diary		X				
Self-report Questionnaires		X	X			X
Biosample Collection (Blood, Urine, Saliva)		X	X			X
Perineal Swab Collection (Men)		X				
Vaginal Swab Collection (Women)		X				
Interval treatments			X	X	X	X

3.4.6 Follow-up Assessments

Patients will be categorized into one of two groups as of their intake assessment: those for whom a surgical treatment is planned (i.e., surgical patients), and those for whom no surgery is planned (i.e., medical patients). We anticipate, based on a survey of LURN investigators, that surgical treatment will be planned for 10% of the study population. For medical patients, follow-up assessments will occur three and twelve months after the baseline assessment. For surgical patients, follow-up assessments will occur three and twelve months after the surgery. Postponement of follow-up assessments based on surgical schedule will minimize the likelihood that a patient is asked to provide data and biosamples

during the perioperative period. If a surgical patient's surgery has not occurred within 3 months of the initial visit, he or she will revert to the schedule of medical patients, with follow-up assessments scheduled based on the date of the initial visit.

Participation in follow-up assessments will consist of repeat assessment with the LUTS Tool (one week recall period) and CASUS, as well as assessment of sexual function, bowel symptoms, depression, anxiety, and health-related quality of life. The research coordinator will also review any interval treatments received, including non-traditional (e.g. herbal remedies), and non-medicinal (e.g. acupuncture) treatments for LUTD. Section 3.4.10 details the analytic methods that will be used to evaluate longitudinal patient data with repeated measures over multiple follow-up visits. To ensure accuracy of patient report of interval treatments between the 3-month and 12-month assessments, the LURN site research coordinators will contact patients at 6 and 9 months to complete short assessments of interval treatments received.

3.4.7 Data Collected

CLINICAL DATA ELEMENTS

Demographics

Demographic information will be collected for all participants including date of birth, sex, race, ethnicity, level of education, employment, and marital status.

History

Patients will be queried regarding past medical and surgical history; diet and use of alcohol, tobacco, and caffeine; history of urinary, vaginal, or sexually transmitted infections; pelvic, prostate, or urologic pain; obstetric history; and menopausal status and use of hormone therapy. We will collect family history with specific attention to identification of first-degree relatives who have been diagnosed and/or treated for LUTS. All current prescription and over-the-counter medications will be recorded. The presence of Metabolic Syndrome will be determined by clinical history and patients' self-reported history or treatment of: elevated blood glucose, hypertension, elevated triglyceride, reduced HDL cholesterol.

Comorbidities

In addition to the health history abstracted above, we will calculate a ***Functional Comorbidity Index*** score for each participant to capture their health status and competing risk of adverse health events. The Functional Comorbidity Index is an 18-item list of diagnoses that discriminates physical function and risk of mortality.

Physical Examination

Patients will undergo standardized physical examination including assessment of height and weight; waist circumference; GU evaluation (penis, scrotum, or vaginal exam with quantification of pelvic organ prolapse using POPQ system); pelvic floor muscles (including pelvic floor muscle contraction strength assessed using the ***Oxford Grading System***) and the rectum.

Tests

Dipstick urine analysis (UA)

Postvoid Residual Urine Volume (PVR) – will be measured within 10 minutes of voiding by ultrasound or straight catheter

Clinical Diagnosis & Treatment Plan

Clinicians will complete a standard form documenting the primary and secondary LUTD diagnoses and their recommended treatments.

SELF-REPORTED SYMPTOM MEASURES

Lower Urinary Tract Symptoms

LUTS Tool is an instrument that assesses the severity and bother of 18 urinary symptoms. There will be two LUTS Tools used, one with a one month recall period (Appendix A1) and one with a one week recall period (Appendix A2).

The Comprehensive Assessment of Self-Reported Urinary Symptoms (CASUS) is a 56-item questionnaire designed for the purposes of capturing a comprehensive set of urinary symptoms, as well as classifying participants into meaningful subcategories (See Appendix C).

American Urological Association Symptom Score Index (AUA-SI) is a validated 9-item measure, which assesses urinary symptoms. (Appendix B)

Urinary Diary – all patients will complete a 3-day urinary diary including fluid intake, voided volumes, leakage episodes, and activity during leakage.

Pelvic Floor Symptoms

Bowel Symptoms

Three ***PROMIS Gastrointestinal Symptom Scales*** are validated instruments to assess constipation (9 items), diarrhea (5 items), and bowel incontinence (4 items). (Appendices D,E and F)

Sexual Function

International Index of Erectile Function (IIEF, men) is a 6-item measure that assesses erectile function in men. (Appendix G)

Pelvic Organ Prolapse/Incontinence Sexual Questionnaire, IUGA-revised (PISQ-IR, women) is a validated measure of sexual function in women with pelvic organ prolapse, incontinence, and/or fecal incontinence. (Appendix H)

Pelvic Floor

Pelvic Floor Distress Inventory – short form (PFDI-20, women) is a 20-item validated measure with three subscales to assess pelvic floor symptoms in women, including urinary, prolapse, and colorectal. (Appendix I)

Pain

Genitourinary Pain Index (GUPI) is a 9-item measure to assess GU pain in men and women. (Appendix J)

Pediatric Disorders

Childhood Traumatic Events Scale is a 6-item measure assessing recollection of events associated with major upheaval such as deaths. (Appendix K)

Psychosocial Symptoms

PROMIS Depression and Anxiety item banks measure mood, affect, negative self-perceptions, negative social perceptions, fear, anxious feelings, hyperarousal, and somatic symptoms related to arousal. (Appendices L and M)

Perceived Stress Scale (PSS) contains 10 items and assesses non-specific subjective stress. (Appendix N)

PROMIS Sleep Short Form is a validated 8-item assessment of sleep patterns. (Appendix O)

General Health-Related Quality of Life

International Physical Activity Questionnaire – Short Form (IPAQ-SF) is a 9-item assessment of four levels of activity. (Appendix P)

PROMIS Physical Function Item Bank, Mobility Subdomain consists of 16 items that measure lower extremity function. (Appendix Q)

3.4.8 Biosample Collection

Blood, urine, and saliva will be collected in all participants at the baseline visit and at the 3 and 12 month visits. Swabs will be collected at the baseline visit. For men, a cotton-tipped swab will be used to culture the perineal area. For women, a cotton-tipped swab will be used to culture the vaginal opening.

All biosamples will be collected according to methodologies outlined in the Manual of Operations. All biosamples will be stored at the NIDDK Sample Repository for future use, including targeted phenotyping studies.

3.4.9 Sample Size and Power Calculations

Estimates of numbers of patients with LUTS who would be available for recruitment into the prospective observational cohort have been provided by investigators at the six LURN clinical sites, and are shown in Table 3.

Table 3: Estimated number of patients available for recruitment from LURN clinical sites

	Men	Women	Total
Monthly Totals	260	183	443
Yearly Totals	3120	2196	5316

We plan to recruit at least 500 men and 500 women and over a 12-month period. We will review quarterly study accrual to confirm our anticipated timeline, update our estimate of eligible patients across the LURN, review recruitment of planned subgroups of patients (i.e., obese patients and diabetics), and update overall recruitment expectations. If our interim assessment indicates that fewer than 10% of participants are obese or have diabetes, we will target additional recruitment to increase representation of these subgroups in the cohort. Additional recruitment may result in a total sample of up to 600 men and up to 600 women, and may extend the recruitment period by 6 months for a total of 18 months. After we have reached our target accrual for the observational cohort, described above, additional participants will continue to be recruited for Project 1B (Neuroimaging and Sensory Testing substudy). These additional participants will undergo only the baseline visit described in Table 2 and none of the follow-up visits.

Rather than presenting power calculations for each hypothesis, we present four series of power calculations for the four basic hypothesis tests we anticipate using: t-tests, logistic regressions, correlations, and chi-square tests (see examples below). All calculations are based on a significance level of 0.05. The power calculations presented below assume that associations are unadjusted for confounding factors. Adjusting analyses using multivariable regression or other techniques will provide at least as much power as an unadjusted analysis and in many cases substantially more power.

In Tables 4 through 7, power is presented for the entire sample size (n = 1000); for group comparisons, subgroup sample sizes are reported in the left-most column. In addition to analyses using the entire sample, certain analyses may be performed separately for men and women. Power calculations for analyses stratified by sex (using only 500 men or 500 women) are presented in parentheses.

Table 4 provides power calculations for two-sample t-tests that will be used to test hypotheses comparing continuous outcomes (e.g., symptom severity, health-related quality of life) between two groups of patients (e.g., men and women) for several potential sample sizes. Differences between groups are expressed in terms of effect sizes. An effect size of 0.25 can be considered small and an effect size of 0.5 can be considered moderate.

Table 4: Statistical power to detect the given effect size using two-sample t-tests using the entire sample (and stratified by sex, in parentheses)

	Effect size			
Total N enrolled	0.2	0.3	0.4	0.5
Subgroups are 50%-50%*				
500 per group	0.885	**0.997**	**>0.999**	**>0.999**
(250 women per group)	(0.607)	(**0.917**)	(**0.994**)	(**>0.999**)
Subgroups are 30%-70%**				
1000: 300 and 700	0.825	**0.991**	**>0.999**	**>0.999**
(150 and 350 women)	(0.534)	(0.866)	(**0.983**)	(**>0.999**)
Subgroups are 10%-90%***				
1000: 100 and 900	0.474	0.811	**0.966**	**0.997**
(50 and 450 women)	(0.268)	(0.519)	(0.764)	(**0.917**)

*An example of a 50-50 split is dividing the sample at the median age and comparing older patients to younger patients.

** An example of a 30-70 split is patients with BMI ≥ 30 compared with patients with BMI < 30.

*** An example of a 10-90 split is patients with diabetes compared with patients without diabetes.

Table 5 provides power calculations for logistic regressions that will be used to test hypotheses comparing continuous predictors (e.g., psychological stress) and dichotomous outcomes (e.g., LUTD with and without urinary incontinence) for several potential sample sizes. Differences between groups are expressed in terms of the odds ratio for a 1 standard deviation increase in the predictor.

Table 5: Statistical power to detect the given odds ratio using logistic regression using the entire sample (and stratified by sex, in parentheses)

	Odds ratio			
Total N enrolled	1.2	1.3	1.4	1.5
Outcome is 50%-50%				
500 per group	0.768	**0.971**	**0.998**	**>0.999**
(250 women per group)	(0.478)	(0.779)	(**0.936**)	(**0.986**)
Outcome is 30%-70%				
1000: 300 and 700	0.696	**0.994**	**0.995**	**>0.999**
(150 and 350 women)	(0.416)	(0.709)	(0.894)	(**0.971**)
Outcome is 10%-90%				
1000: 100 and 900	0.370	0.653	0.858	**0.957**
(50 and 450 women)	(0.210)	(0.384)	(0.573)	(0.738)

Table 6 provides power calculations for correlations that will be used to test hypotheses comparing two continuous variables (e.g., psychological stress and severity of symptoms; Aims 4 and 5) for several potential sample sizes.

Table 6: Statistical power to detect the given correlations using the entire sample (and stratified by sex, in parentheses)

	Correlation		
Total N enrolled	0.10	0.15	0.20
1000	0.887	**0.998**	**>0.999**
(500 women)	(0.610)	(0.921)	(**0.995**)

Table 7 provides power calculations for chi-square tests that will be used to test hypotheses comparing two dichotomous outcomes (e.g., diabetes and no diabetes, LUTD with and without urinary incontinence) for several potential sample sizes. Differences between groups are expressed in terms of proportions of patients in each group within one variable (such as those with diabetes and those without diabetes) that have the other condition (such as urinary incontinence). Proportions of 0.40 and 0.50 would mean 40% of patients with diabetes also have urinary incontinence, while 50% of patients without diabetes have urinary continence.

Table 7: Statistical power to detect the given proportions using chi-square tests using the entire sample (and stratified by sex, in parentheses)

	Proportions (first group vs. second group)		
Total N enrolled	0.40-0.50	0.35-0.50	0.30-0.50
Subgroups are 50%-50%*			
500 per group	0.890	**0.998**	**>0.999**
(250 women per group)	(0.614)	(**0.926**)	(**0.996**)
Subgroups are 30%-70%**			
1000: 300 and 700	0.830	**0.993**	**>0.999**
(150 and 350 women)	(0.538)	(0.877)	(**0.989**)
Subgroups are 10%-90%***			
1000: 100 and 900	0.475	0.823	**0.976**
(50 and 450 women)	(0.265)	(0.522)	(0.783)

*An example of a 50-50 split is dividing the sample at the median age and comparing older patients to younger patients.

** An example of a 30-70 split is patients with BMI ≥ 30 compared with patients with BMI < 30.

*** An example of a 10-90 split is patients with diabetes compared with patients without diabetes.

3.4.10 Statistical Analysis

Aim 1

Aim 1 will characterize men and women with LUTD cross-sectionally, describe the clinical and demographic characteristics of study participants, identify relationships between clinical and demographic characteristics and urinary symptoms and other clinical measures, and identify relationships between the clinical impression and diagnosis. We will do so by testing *a priori* hypotheses.

First, we will report descriptive statistics of the characteristics of the participants. Descriptive statistics will include frequencies and percentages for categorical variables, and means, standard deviations, and ranges for continuous variables. Variables will also be examined separately by subgroups, such as by LURN clinical site sex, race and ethnicity.

We will examine the distribution of responses on health-related quality of life questionnaires and instruments that assess pelvic floor function, and psychological and sleep disorders. If the outcome variables are normally distributed, we will use t-tests to compare the mean values of the parameters among men and women and Pearson correlations to examine associations with symptom severity. If the outcome variables are not normally distributed, we will use Wilcoxon rank sum tests or other non-

parametric tests to compare men and women, and Spearman rank correlations to examine associations with symptom severity. If outcomes are binary or categorical, we will use chi-square tests, logistic regression, or multinomial regression to examine associations.

Using the same techniques, we will also examine whether patients differ in their LUTD symptom severity and bother and self-report of pelvic floor function and psychological factors when stratified by age, sex, race/ethnicity, presence of diabetes mellitus, and presence of obesity, and whether clinicians' impression/diagnosis exhibits variability in urinary symptoms, patient demographic and clinical characteristics, pelvic floor function, and psychological factors.

Additional investigations of subgroup differences will use multivariable analyses to control for demographic characteristics and multiple measures of surgical history, obstetric history, comorbidities, and bother with symptoms at the same time. If the outcome measure is binary or categorical, logistic or multinomial logistic regression will be used. If the outcome measure is continuous, we will use linear regressions. Multivariable models will be created using a best subsets approach, with the final model being the one with the highest likelihood score statistic or explained variance, provided that all covariates are statistically significant at $p < 0.05$. Multivariable models will be adjusted for LURN clinical sites whenever appropriate.

Aim 2

The goal of Aim 2 is to identify subgroups or clusters of study participants. Anticipated subgroups for this aim will not be identified *a priori*, but rather will be identified based on exploratory data mining techniques. Using the basic clinical data obtained above, all patients will then be categorized into groups for future deep phenotyping studies. It is possible that one patient may be assigned into multiple groups.

In separate analyses, we will use cluster analysis to examine subgroups of participants who have a) similar self-reported symptoms (as measured by items on the LUTS Tool and CASUS), and b) voiding diary parameters.

High correlations between clustering variables can be problematic for the identification of valid clusters. Therefore, our first step will be to examine the associations between clustering variables using Pearson correlations, Spearman rank correlations, and exploratory factor analysis as appropriate. If several variables are strongly associated with one another, we will select from among those variables the one with the highest factor loading for inclusion in the cluster analysis.

We will use several methods of cluster analysis, with different properties, and compare results. One method will be k-means cluster analysis, which is a widely used nonhierarchical method, but which tends to find clusters with similar numbers of observations and can be influenced by the seeds.[2] Another method will be nucleated agglomerative clustering, which is based on k-means clusters but which tends to perform better.[2] A third method will be rotated principle component clustering, which tends to be very accurate but does not perform well with very small sample sizes.[3]

Prior to performing the cluster analysis, voiding diary parameters will be standardized using z scores. If they are not normally distributed, log transformation will be used if appropriate. If some variables, such as the count of nighttime voids, cannot be transformed to achieve normality, we will use latent class cluster analysis (LCCA) instead of z scores. LCCA easily handles a mix of count, binary, nominal, ordinal, and continuous variables.[4]

We will examine which clustering variables contribute to differences between the clusters using discriminant function analysis. The results of this, along with an examination of the means of each contributory variable in each cluster, will help us identify the clinical profile of patients belonging in each

cluster. The number of clusters retained in the final solution will be based on clinical interpretability, and, as appropriate given the type of cluster analysis chosen, any of the following: a) aggregation error, b) the gap statistic, c) Akaike information criteria (AIC) and Bayesian information criteria (BIC), and d) model entropy.[5]

For Exploratory Question 2a, we will create separate cluster analyses for the LUTS Tool and CASUS. We hypothesize that these analyses will yield a similar number of clusters based on both questionnaires. Moreover, we anticipate that both questionnaires will assign the same participants to the same clusters. We will, however, examine any differences between clustering across the two measures to determine whether the types of clustered identified are different across the two questionnaires (e.g., if the bother items included in the LUTS Tool, but not CASUS, result in a different cluster of participants).

We will also use classification and regression trees (CART) to examine subgroups of participants who are more or less likely to have a binary characteristic (classification tree) or have higher or lower means of a continuous variable (regression tree). Separate CARTs will be examined for dependent variables, including each LUTS Tool item, CASUS items, and voiding diary parameters. (For the description below, we will assume that the dependent variable is continuous and that the analysis is a regression tree. The same process holds for classification trees.)

The first stage in each CART will be to grow the tree. Variables that are believed to have the potential to distinguish between groups of patients, which may include demographic characteristics, surgical and obstetric history, comorbidities, bother with symptoms, and symptoms or voiding diary parameters that are not the dependent variable of interest, will be tested for predictive ability. The variable that best predicts the dependent variable will be identified, and will split the tree into two nodes or subgroups that have low mean and high mean values of the dependent variable. Each node is split again by the most predictive variable for just the patients in that node, until the remaining group is homogenous or there are no additional splits possible. The minimum group size will be set at 20 patients per node.

After growing each CART, we will prune the tree to avoid over-specification. This will be done by cross-validating the tree on a sample of the data and will be used to minimize expected misprediction error.

Aim 3

Aim 3 involves longitudinal hypotheses about LUTS, health-related quality of life, pelvic floor function, and psychological factors. Associations between patient clusters and treatment selection will be examined as described in as part of Aim 2. We will examine changes in symptoms over time (Hypotheses 3a and 3b) using repeated measures ANOVA or mixed models with random within-person effects. Repeated measures ANOVA and mixed models with random within-person effects will allow for analysis of a cohort of patients with multiple follow-up visits within the 12-month study time frame of the LURN prospective observational cohort. To address Hypothesis 3c, we will categorize patients into types of treatment: medication, surgical intervention, watchful waiting, and other. We will examine the associations between symptom cluster and these categories of treatments using chi-square tests and multinomial regressions. Patients will also be categorized into a treatment-responsive group and a treatment-unresponsive group. We will examine associations between treatment response and cluster, as well as other patient characteristics, using logistic regression. To address Hypothesis 3d, we will use multinomial logistic regressions to predict symptom cluster at 12 months using baseline cluster and treatment responsiveness, controlling for patient characteristics.

Aim 4

For the baseline and 12 month assessments of the CASUS, we will compute rates of missingness separately for each item and compare these to similarly computed rates of missingness for items from

the other self-report instruments (Exploratory Question 4a). We will examine distributions of item responses to identify any items that have little or no variability (Exploratory Question 4b). Finally, we will compute the Pearson (or Spearman) correlations between all pairs of CASUS items separately at baseline and the 12 month assessments, along with their 95% confidence intervals (Exploratory Question 4c). Any pair of items with a correlation $\geq$ .90 will be flagged for discussion by the LURN investigators. Very high correlations suggest that the items essentially measure the same thing, such that only one item from the pair is necessary. We will also examine the mean change score for each CASUS item, along with LUTS Tool items, stratified by treatment status.

Aim 5

Separately for the baseline and at 12 month assessments, we will examine the relationship between CASUS items and corresponding items on the LUTS Tool using scatterplots with superimposed LOESS curves, and Pearson (or Spearman) correlations as appropriate (Hypothesis 5a). Table 8 lists the corresponding CASUS and LUTS Tool items that we will examine. Correlations $\geq$ .70 will be considered evidence of the convergent validity of the specific CASUS items.

3.4.10.1 Missing Data

Every effort will be made to obtain complete data for all variables. In any publication of results from this study, the percent missing for each variable will be reported, and any sample size reduction due to missing data will be acknowledged. Preliminary analyses, performed prior to the end of data collection and cleaning, will be performed using complete cases (that is, we will drop a participant from the analysis if one or more of the participant's data points of interest are missing). Once all data have been collected, we will examine patterns of missing data and will also consider whether the data can be assumed to be missing at random (MAR). A test for MAR will be carried out using logistic regression to predict missingness (Y/N) separately for each variable with missing data, using all relevant measured variables as potential predictors. Any variables found to be predictive of missing outcomes will be included in any analysis of that outcome. Missingness related to unmeasured variables cannot be tested. To address missing covariate data in regression models, we will perform multiple imputations using IVEware software to give 5-10 estimates for each missing value, followed by analyses to combine the results from each of the 5-10 imputation datasets. The final results incorporate both between- and within-imputation variance, and assuming MAR the results will yield unbiased estimates of both the parameters and standard errors.

724 **Table 8: Items from the Comprehensive Assessment of Self-Reported Urinary Symptoms and their Analogues from the LUTS Tool**

CASUS Item Number	CASUS Item	Corresponding LUTS Tool Item
A1	In the past 7 days, during waking hours, how many times did you typically urinate?	Had frequent daytime urination?
A2	In the past 7 days, during a typical day, how much time typically passed between urinations?	Had frequent daytime urination?
A3	In the past 7 days, during a typical day, how often did you urinate twice or more within a few minutes?	Had frequent daytime urination?
B1	In the past 7 days, during a typical night, how many times did you wake up and urinate?	During a typical night, how many times do you wake up because you need to urinate?
B2	In the past 7 days, how often did you wakeup at least once during the night because you had to urinate?	Had frequent nighttime urination?
B5	In the past 7 days, how often did you leak urine during the night, including wetting a pad or the bed?	Leaked urine when you were sleeping?
C5	In the past 7 days, how often did you have pain or discomfort in your bladder while it was filling?	Had pain or discomfort in your pubic or bladder area?
C7	In the past 7 days, how often did you have pain or discomfort in your bladder when it was full?	Had pain or discomfort in your pubic or bladder area?
C9	In the past 7 days, how often did you have pain or discomfort while urinating?	Had a burning feeling when you urinate?
D1	In the past 7 days, how often did you feel a sudden need to urinate?	Had a sudden need to rush to urinate?
D3	In the past 7 days, how often did you have a sudden need to rush to urinate for fear of leaking urine?	Had a sudden need to rush to urinate for fear of leaking urine?
E1	In the past 7 days, how often did you have to push when urinating?	Had to push or strain while urinating?
E2	In the past 7 days, how often did you have a delay before you urinated?	Had a delay before you start to urinate?
E5	In the past 7 days, how hard did you have to push during urination?	Had to push or strain while urinating?
F1-M	In the past 7 days, how often did you have splitting or spraying of your urine stream?	Had splitting or spraying of your urine stream?
F1-F	In the past 7 days, how often did you have splitting or spraying of your urine stream?	Had splitting or spraying of your urine stream?
F2	In the past 7 days, once you started urinating, how often did your urine flow stop and start again?	How often did your urine flow start and stop while you were urinating?
F3	In the past 7 days, how often was your urine flow slow or weak?	Had a weak urine stream?
F4	In the past 7 days, how often did you have a trickle or dribble at the end of your urine flow?	Had a trickle or dribble at the end of your urine flow?

G3	In the past 7 days, how often did you leak urine or wet a pad after feeling a sudden need to urinate?	Leaked urine in connection with a sudden need to rush to urinate?
G4	In the past 7 days, how often did you leak urine or wet a pad while laughing, sneezing, or coughing?	Leaked urine in connection with sneezing, coughing, or other physical activities?
G5	In the past 7 days, how often did youlleak urine or wet a pad when doing physical activities, such as exercising or lifting a heavy object?	Leaked urine in connection with sneezing, coughing, or other physical activities?
G9	In the past 7 days, how often did you leak urine or wet a pad without any reason you could identify?	Leaked urine for no reason?
H2	In the past 7 days, how often did you feel that your bladder was not completely empty after urination?	Had the feeling your bladder was not empty after urinating?
H3	In the past 7 days, how often did you dribble urine just after zipping your pants or pulling up your underwear?	Leaked urine after you have finished urinating?

725

726 **3.5 Project 1A Timeline**

Key Tasks	**Target Completion Date**
Approval of the revised Protocol	*July 18, 2014*
EEP review and the EEP Response.	*July, 2014*
NIDDK approval of the Protocol	*August 1, 2014*
IRB submission of the Protocol	*January, 2015*
Finalize and distribute the Biosample Collection and Observational Cohort MOO to the LURN	*February, 2015*
IRB approval of the Protocol	*March, 2015*
Study Coordinator Training	*March 5, 2015*
Study site orientation/ activation	*March, 2015*
Begin subject enrollment	*April, 2015*
End enrollment	*April, 2016*

727

4 Project 1B: Neuroimaging and Sensory Testing Study

4.1 Background, Study Rationale

The LURN Neuroimaging and Sensory Testing Study will investigate abnormal sensation of the lower urinary tract at the level of the organism (**Figure 2**). Examples of abnormal sensation include urinary urgency, frequency, nocturia, sensation of incomplete bladder, etc. These sensations require processing of the afferent signals by the brain and the somatosensory nervous system. In this study, we will investigate whether **abnormal brain connectivity or sensory processing contribute to abnormal sensation of the lower urinary tract or LUTS**. We are asking the question, "What are the sensory processing factors contributing to disorders of urinary sensation?" The prototypical LUTS or abnormal sensation that we will focus on is urinary urgency. Urinary urgency is defined as the complaint of a sudden compelling desire to pass urine which is difficult to defer, in accordance to the 2002 International Continence Society (ICS) terminology.

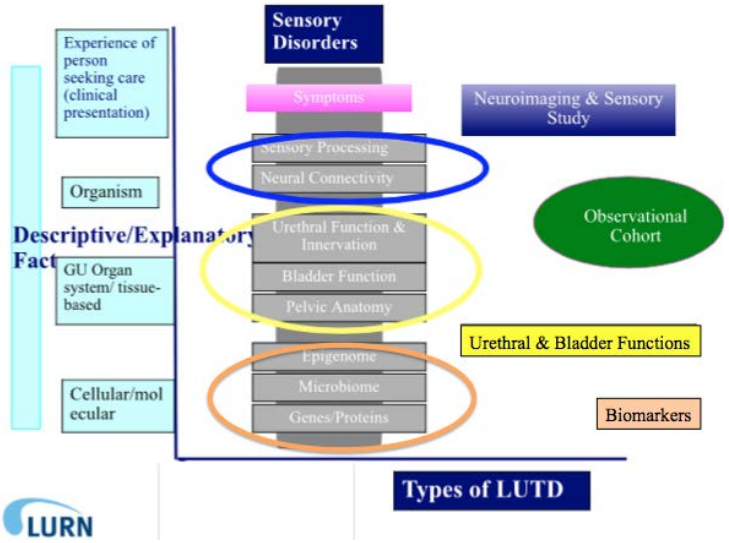


Figure 2: Overview of LURN Phenotyping Effort

Sensations of the body, including sensations associated with the urinary tract (e.g., urinary pain and urgency), will necessarily engage the nervous system in processing, interpreting and modulating the sensation. Simply put, without involvement of the brain and the somatosensory system, there is no perceived sensation to speak of. With this in mind, we propose to phenotype patients with urinary urgency using magnetic resonance neuroimaging and multimodal quantitative sensory testing (QST) methods. **The overarching hypothesis is that patients with urinary urgency (with or without urgency incontinence) will demonstrate abnormal sensory processing of the nervous system, which will be manifested as: (1) abnormal functional connectivity of brain regions involved in urinary sensation, and (2) abnormal sensory hypersensitivity, involving multiple sensory modalities including pressure and auditory sensitivity.**

With neuroimaging (functional MRI), we will examine functional and structural connectivity between brain regions implicated in sensory processing and motor control of the urinary tract. With quantitative sensory testing (QST), we will examine generalized or "global" sensory sensitivity by characterizing patient responses to somatic (i.e., pressure) and non-somatic (i.e., auditory) stimulation. These methods are complimentary in that neuroimaging explores the neural substrates that underlie the subjective sensory percepts evoked and measured in QST.

The mechanistic phenotyping strategy proposed here permits the identification of patient subgroups based on objective neuroimaging connectivity patterns and behavioral responses to multimodal sensory stimulation. Currently there is a limited understanding of how to conceptualize empirically LUTD, and as a result comprehensive assessment and treatment of LUTS is limited. Understanding the pathophysiologic mechanisms involved in LUTS is therefore critical to developing effective and

individualized therapies.

4.1.1 Phenotyping by Neuroimaging

Rationale of fMRI Studies for LUTS. Bladder continence and voiding control depend on proper functioning of the brain neural network that provides the ability to voluntarily postpone voiding during bladder filling. Contemporary functional MRI studies show that abnormal activation and/or deactivation of specific brain areas may contribute to the symptoms of overactive bladder (OAB), a term used to describe a urinary symptom complex defined by the presence of urinary urgency, with or without urgency incontinence, usually with frequency and nocturia, in the absence of urinary tract infection or other identifiable causes. Griffiths and Tadic evaluated elderly female patients with urgency incontinence, and found altered neural activity in the limbic region, including the anterior cingulate gyrus (ACG), insula, and prefrontal cortex, compared to patients without urinary symptoms. Other areas of the brain, e.g., the parieto-temporal lobes, thalamus, periaqueductal gray (PAG), and pontine micturition center (PMC), are also involved in bladder control. Thus, many discrete areas of the brain are recognized to be involved in bladder function and control.

Although specific cortical areas have been described, much less is known about how alteration of communication, also known as connectivity, between these cortical areas may contribute to the pathophysiology of urinary urgency. This is a logical step in research, as we leap from imaging individual brain centers (activation/deactivation) to understanding how these brain centers communicate with each other. A secondary analysis of the functional MRI (fMRI) data in a small numbers of patients revealed a shift of brain connectivity to the parieto-temporal complex, and a change of overall cortical connectivity, compared to controls. Although these data are promising, the sample size was too small (n=11) to draw definitive conclusions. Additional brain connectivity studies are needed to understand the central nervous system (CNS) contribution to urgency. Of note, interstitial cystitis/bladder pain syndrome (IC/BPS), a pelvic pain symptom complex that shares overlapping symptoms with OAB (e.g. urgency), has recently been shown to have alterations in resting state activities and connectivity within the sensory and motor networks in the brain.

Recent studies also showed that brain white matter hyperintensities (WMH), a measure of structural defects in the brain's white matter, are associated with increased prevalence of urgency, increased severity of urgency incontinence,[1] and the presence of detrusor overactivity during urodynamic testing. Brain WMH burden is also correlated with alteration of brain activities in neural circuits involved in bladder control. Collectively, the data suggest that **damage to brain white matter may affect functional connectivity between cortical regions involved in bladder control**.

There are several major limitations of the studies to date: most functional MRI studies had enrolled only geriatric women with urgency incontinence. Many of these studies did not have a matched control group. Thus, it remains unclear if the abnormalities are also present in younger patients, in male patients, or in urgency patients without urgency incontinence. Here we propose to study male and female patients with urgency, with and without urgency incontinence, across the age spectrum (see **Table 9**). In addition, all published functional MRI studies so far have utilized a block design that involves repetitive, alternating, rapid cycles of bladder infusion and fluid removal via a catheter. Unfortunately, this bladder stimulation paradigm is non-physiologic, invasive, and may sensitize the bladder. Clinically, the feeling of a strong urge to void during rapid, repetitive, artificial bladder infusion/withdrawal may not recreate the everyday experience. In this study, we plan on using a more natural dieresis protocol without using a catheter.

Table 9: Comparison of proposed LURN neuroimaging studies to previous imaging studies.

Previous functional MRI studies	Proposed LURN neuroimaging studies
Functional MRI studies (fMRI) that focused primarily on the activation/deactivation of cortical areas	(a) Connectivity studies using resting state functional MRI (fMRI) to examine alterations in brain networks and abnormal communication between cortical areas involved in bladder control; (b) diffusion tensor imaging (DTI) to examine structural alterations in brain white matter tracts.
Mostly women	Men and women (1:1 ratio)
Predominantly elderly	Across all age groups
Mostly with urgency incontinence	Patients across a spectrum of urgency, with or without urgency incontinence
Small no. of participants, typically single site studies	Large no. of participants across six LURN sites (n=256 for this proposal)
Repeated bladder filling and withdrawal via a catheter (unnatural)	Use a diuresis protocol to fill bladder without a catheter (more natural filling)
	Integration with detailed phenotyping data available through the umbrella LURN Observational Cohort Study
	Integration with quantitative sensory testing (QST) as part of this protocol

In this neuroimaging study we shall use two innovative neuroimaging technologies – **resting state functional MRI (RSfMRI) and diffusion tensor imaging (DTI)** – to investigate functional connectivity and structural white matter tracts in subjects with urinary urgency, and compare the results to matched controls. RSfMRI provides a picture of functionally related regions of the brain to examine the sensory, motor, and default mode networks at rest. RSfMRI offers an improvement over stimulus-based fMRI in not requiring repeated bladder filling via a catheter. DTI provides a map to study specific brain white matter tracts. **RSfMRI and DTI studies represent a critical next step, as we leap from imaging activities in individual brain centers to understanding how these centers communicate and function as an integrated network**, and how this network may be compromised in patients with urinary urgency. This study will address the deficits of the literature, expand our understanding of the brain-bladder network, and represent a new paradigm of LUTS research.

Besides providing insights into the CNS contribution to urgency, RSfMRI data may also help to decipher the mechanistic difference between patients with or without incontinence at the level of the brain. It is generally believed that patients with urgency incontinence have involuntary bladder contractions

and/or abnormalities in the pelvic floor, to permit leakage to occur when the bladder contracts. Thus "motor defects" of the bladder and/or pelvic floor might be involved. On the other hand, the mechanisms that underlie urgency without urinary incontinence are poorly understood. These patients may have "sensory defects" involving the bladder, afferent (sensory) nervous system, and/or the central nervous system (e.g. the brain). It is currently unknown whether urgency without or without incontinence might represent a true continuum reflecting different degrees of continence control, or the two entities might have different underlying pathophysiology. RSfMRI studies will allow us to investigate the sensory and motor networks in the brain, to determine if: (1) patients' severity of incontinence is positively correlated to alterations in the motor network of the brain that controls the pelvic floor, and (2) patients with abnormal urgency have differential alterations in brain connectivity in the sensory network that governs visceral sensation compared to healthy controls.

4.1.2 Phenotyping by Quantitative Sensory Testing (QST)

Rationale of QST Studies for LUTS. A significant percentage of OAB patients do not have involuntary bladder contractions and/or urgency incontinence. This raises the question whether some patients with urgency might have abnormal processing of their sensation by the nervous system. **We hypothesize that a subset of patients with urinary urgency will demonstrate sensory hypersensitivity compared to healthy controls. This sensory hypersensitivity may be generalized and involve multiple sensory modalities including somatic mechanical (i.e., pressure) and auditory sensitivity.** QST has been used extensively to phenotype clinical conditions that are characterized by sensory hypersensitivity, such as fibromyalgia or interstitial cystitis/bladder pain syndrome (IC/BPS). Here we shall use QST to investigate sensory hypersensitivity in patients with urinary urgency.

Definition of QST. Quantitative sensory testing (QST) refers broadly to procedures that assess perceptual responses to quantifiable physical stimuli in an effort to measure gain or loss in sensory function.[20-22] In pain research, for example, QST can detect increased pain sensitivity (hyperalgesia), decreased pain sensitivity (hypoalgesia), pain in response to normally non-painful stimulation (allodynia), and altered endogenous pain modulation. During QST, sensations are evoked by stimuli (e.g., mechanical or thermal) applied in a systematic manner to one or more body regions. Subject responses to these stimuli, such as ratings of perceived intensity, are correlated to stimulus intensity or duration to provide a quantifiable index of experimental sensory sensitivity. QST has been used extensively to characterize sensory function in individuals, and investigate pharmacological efficacy and mechanistic differences between groups. In addition, pre-treatment/baseline QST has been shown to predict treatment outcomes for both behavioral and pharmacological pain interventions. Taken together, these studies support our view that mechanistic phenotypes determined by QST may be useful in the development of patient subgroups and personalized treatment algorithms in LUTS. Overall, QST studies will help us understand whether LUTS patients with abnormal sensation in the urinary tract might also have global abnormalities in sensory processing.

QST in Chronic Pain and LUTD. Sensory hypersensitivity, whereby a particular sensation is perceived at a lower than expected threshold during QST, has been found in a wide variety of chronic pain conditions, such as fibromyalgia, chronic back pain, and vulvodynia. This hypersensitivity can be present in both painful/symptomatic and pain-free/non-symptomatic body sites. Neuroimaging studies, including those by the University of Michigan group, have found that sensory sensitivity correlates with increased brain activity in the insula, anterior cingulate gyrus, prefrontal cortex, and thalamus. Notably, these areas are nearly identical to the brain areas that are activated during urine storage.

Although QST studies in OAB patients are lacking, several studies have utilized QST to evaluate sensory sensitivity in patients with IC/PBS, a type LUTD. In one of the earliest such studies, Clauw et al.,

demonstrated that female IC/PBS patients have significantly decreased pressure pain thresholds, meaning increased pain sensitivity, throughout the body at traditional fibromyalgia tender points compared to healthy controls. Results from the remaining QST studies conducted on this patient population seem to depend largely upon the pain modality assessed. One group found hyperalgesia to bladder filling but no difference in cutaneous electrical thresholds between subjects with painful bladder syndrome and controls. Ness et al. showed that pressure pain and ischemic pain thresholds were significantly decreased in IC patients when measured at the forearm. More recently, Lai et al. demonstrated increased pressure sensitivity in the suprapubic region of IC patients compared to controls. Thermal pain sensitivity has also been assessed; one group identified a significant decrease in sensitivity in 1 of 4 tested dermatomes among patients with IC/PBS compared to controls, while other studies failed to find any significant abnormalities in thermal pain sensitivity in IC/PBS patients. These studies indicate that, at least in IC/PBS, pressure is the most consistent QST modality for detecting sensitivity differences between patients and controls, regardless of testing site.

Quantitative sensory testing has also been used to demonstrate that some chronic pain patients exhibit increased sensitivity to non-somatic stimulation, including auditory and visual stimuli. There is also evidence that somatic pain and auditory sensitivities are often interrelated suggesting a global state of CNS sensory amplification might play a role in the pathogenesis of many chronic pain disorders and these measures may highlight an important individual patient phenotype. The biological plausibility of this proposition is supported by neuroimaging studies showing the insula, a brain region that plays a polysensory integration function, is hyperactive in most individuals with chronic pain. Interestingly, although is it currently unknown whether OAB patients also exhibit non-somatic hypersensitivity, the insula is hyperactive in this population as well.

Urinary Urgency – Urinary Pain Continuum. Even though OAB patients do not report chronic pain, emerging evidence suggested that OAB and IC/BPS have overlapping symptoms (e.g. urgency), and the two syndromes may share similar pathophysiology processes (e.g. abnormal sensory processing). In fact, some investigators have considered OAB and IC/BPS as part of a continuum of bladder hypersensitivity disorder, and lumped both conditions under the category of "sensory/afferent abnormalities." The distinction between urinary urgency and urinary pain is not always clear, and both could be on a continuum of sensory hypersensitivity. Indeed, a recent report indicated substantial overlap in self-reported urinary pain, urgency, frequency, and incontinence symptoms in IC/PBS and OAB. It is also possible that mechanisms which are responsible for the development or maintenance of bladder pain may also contribute to urinary symptoms via a global sensory hypersensitivity phenomenon. These hypotheses are not mutually exclusive.

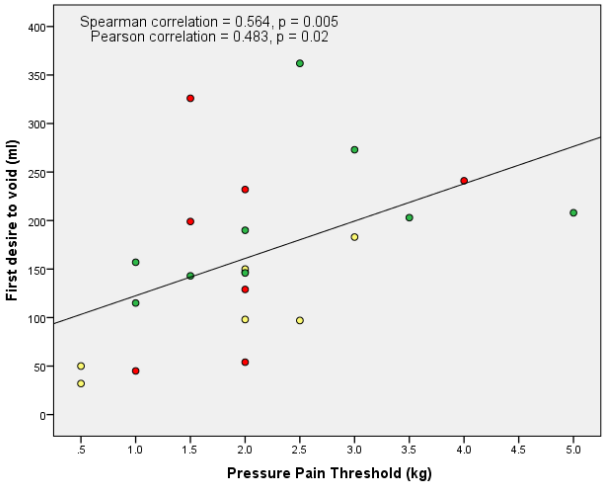


Figure 3: Relationship between thumbnail pressure pain sensitivity and bladder sensitivity in females with IC (green), OAB (red), and IC+OAB (yellow).

This second hypothesis was evaluated in a small pilot study conducted at the University of Michigan (Clemens & Harte, unpublished data). Female patients with IC (n=9), OAB (n=8), and IC+OAB (n = 6) underwent pressure pain QST at the thumbnail and urodynamic testing. During urodynamic testing, water was infused into the empty bladder via a catheter at a rate of 50 ml/min with the patient standing. The amount of

water delivered into the bladder was recorded when patients indicated: 1) first sensation of fluid in bladder, 2) first desire to void, and 3) strong desire to void. Maximum cystometric capacity was also measured. Pain and bladder sensitivity, and bladder capacity, were not significantly different between groups. However, as shown in **Figure 3**, there appears to be a moderate correlation between bladder sensations and thumbnail pain, such that increased pain sensitivity (lower pain threshold) is associated increased bladder sensitivity (less water required to evoke first desire to void). The type of symptoms (IC vs. OAB) does not appear to impact this association. A similar but not significant correlation was also observed for pressure pain tolerance and strong desire to void. These data, while preliminary, suggest a relationship between pain and bladder sensitivity with the possibility of a shared mechanism of global sensory hypersensitivity. This construct will be explored in LURN by examining the association of non-urological somatic pressure pain and auditory sensitivity measured by QST with neuroimaging, urodynamic testing, and self-reported urinary symptoms.

Thumbnail Pressure Pain Sensitivity. We propose using QST to measure experimental pressure pain in urinary urgency patients. Threshold and suprathreshold indices of pain sensitivity will be assessed by pressure applied to the thumbnail. The use of thumbnail pressure as an evoked pain stimulus and its validity in the measurement of CNS pain and sensory processing has been discussed extensively. Experimentally, the easily accessible thumbnail provides an ideal stimulation site because of its dense innervation of mechanical receptors and large representation in the primary somatosensory cortex. Thus, pain sensations can be readily evoked by low intensity, non-tissue damaging pressures. More importantly, the thumb is a "neutral site" that is not associated with LUTD or other chronic conditions, and is remote from the site of primary symptom complaint (i.e., the bladder). Thus, findings of increased sensitivity at the thumbnail, as opposed to the pelvic region or bladder alone, suggest a CNS mediated mechanism of generalized or global sensory hypersensitivity/hyperalgesia. Accordingly, it was previously demonstrated that experimental pain evoked by thumbnail pressure is associated with overall body tenderness, measures of clinical pain, functional neuroimaging, and brain levels of glutamate, and is lowered following analgesic treatment.

Recently, pressure pain sensitivity was evaluated in 346 participants at six discovery sites of the NIH-sponsored Multidisciplinary Approach to Chronic Pelvic Pain (MAPP) research network (Harte et al., unpublished MAPP data). A series of pressures were delivered to the thumbnail using the University of Michigan-designed MAST QST system (see below). Results indicated that female and male patients with chronic urological pain (primarily IC/PBS) and positive control patients (primarily fibromyalgia) exhibited significantly increased pressure pain sensitivity compared to healthy controls (all $p<0.02$). Pain sensitivity at baseline was also a significant predictor of urological symptom change over a 6-month period. Importantly, pain sensitivity variables of interest were not associated to potentially confounding psychological factors, including anxiety, depression, affect and coping. These findings, in conjunction with the findings discussed above, support the value and feasibility of this QST method in LURN. The application of chronic pain investigative techniques to quantify sensitivity in urinary urgency is novel and presents of fruitful area of discovery. Furthermore, by adopting a similar QST method as used in the MAPP network, results obtained in LURN subjects can be compared with those from MAPP subjects to determine potential mechanistic differences (or similarities) between these types of LUTD patients.

Auditory sensitivity. As mentioned above, in addition to hypersensitivity to somatic stimuli, many chronic pain patients also experience hypersensitivity to non-somatic stimuli. As part of the MAPP study, auditory sensitivity was assessed in 38 subjects with irritable bowel syndrome (IBS) (15 M, 23 F), 34 subjects with chronic urological pain (18 M, 16 F), and 52 healthy controls (28 M, 24 F). The experimental paradigm was based on an earlier study of auditory sensitivity from Hollins in which two tones of different frequencies were combined to produce mildly unpleasant sounds. These sounds were

delivered at varying loudness levels in random order. After each trial, the combined tones were rated for intensity and unpleasantness on a scale of 0-100. For both unpleasantness and intensity, there was a significant group x sex interaction (p <0.05); post-hoc analyses revealed that for unpleasantness, these differences were driven by women, and for intensity, the differences were driven by men. The results reveal differences in non-somatic perception between groups and suggest that this simple, non-invasive measure may be useful as a phenotyping variable for subgrouping subjects on widespread sensitivity to environmental events and perhaps central augmentation of unpleasant sensory stimuli. We propose adopting a similar auditory paradigm in the LURN network. Whereas thumbnail pressure pain sensitivity necessarily involves both peripheral and central nervous system mechanisms, auditory sensitivity is considered a more "pure" CNS mediated test modality. Therefore, the inclusion of auditory testing to the LURN QST phenotyping battery improves our ability to detect central mechanisms of sensory amplification.

4.1.3 Summary

The overall objective of the neuroimaging and sensory testing protocol is to provide LURN a comprehensive yet feasible set of neuroimaging and QST methods. Examining for the presence of specific neuroimaging and sensory sensitivity abnormalities in urinary urgency (with or without urgency incontinence) may lead to more evidence-based categorization and treatment paradigms for LUTD, rather than relying upon nonspecific, symptom-based categorization. Evidence of global, centrally-mediated sensory abnormalities in some of these patients may suggest different etiologic factors as the cause the symptoms, and may provide a rationale for individualized therapy targeted at sensory abnormalities.

4.2 Study Objectives

Aim 1: To use resting state functional MRI (RSfMRI) and diffusion tensor imaging (DTI) to phenotype patients with urinary urgency, with or without urgency incontinence.

Hypothesis 1a: Patients with urinary urgency will demonstrate different brain functional connectivity (RSfMRI), including changes in inter- and intra-network connectivity of the control and salience networks, compared to controls.

Hypothesis 1b: Patients with urinary urgency will demonstrate altered brain white matter tract integrity (DTI), including reduced anisotropy within the prefrontal cortex and in the limbic region, compared to controls. The alteration in brain white matter tract integrity (DTI) will further correlate with changes in brain functional connectivity (RSfMRI) in patients with urinary urgency.

Hypothesis 1c: The degree of MRI abnormalities (RSfMRI, DTI) will have a positive correlation to the severity of urgency incontinence in patients.

Aim 2: To use quantitative measures of global sensitivity to phenotype patients with urinary urgency, with or without urgency incontinence.

Hypothesis 2a: Patients with urinary urgency will demonstrate increased sensitivity to non-pelvic somatic pressure stimuli and auditory stimuli compared to controls.

Hypothesis 2b: Global sensory abnormalities will be less common in patients with urinary urgency than in pelvic pain patients recruited through the NIDDK MAPP Research Network.

Hypothesis 2c: The degree of sensory sensitivity will have a positive correlation to the severity of urgency incontinence in patients.

Aim 3: To assess the interaction between the neuroimaging and multimodal sensory testing aims.

Hypothesis 3: Patients with abnormal functional connectivity of the brain in the RSfMRI study will demonstrate abnormalities in multimodal sensory testing.

4.3 Methods

4.3.1 Study Methods

There are two parts to this protocol:

(1) Neuroimaging studies including RSfMRI and DTI, and

(2) Sensory testing using multimodal QST

Neuroimaging Protocol:

DTI and RSfMRI will be used to collect data for the Neuroimaging and Sensory Testing Study. DTI and RSfMRI are currently being performed by many LURN sites for the MAPP study; this will facilitate standardization across sites. Participating LURN sites for the neuroimaging study will be: (1) Washington University in St Louis, (2) University of Michigan at Ann Arbor, (3) Northwestern University, (4) Duke University, (5) University of Washington at Seattle, and (6) University of Iowa. Data from neuroimaging studies will be transferred to the central imaging repository. Central readings of the images will be performed at Washington University. Washington University investigators and their technical team will be blinded to the identity of the subjects (patients versus healthy control).

Overview of the neuroimaging sequence, including approximate amounts of time for each step:

(1) Participant completes self-reported questionnaires, about 20-30 minutes,
(2) Participant will first void prior to entering the MRI scanner. The voided volume will be measured in mL using a measuring cup. Sites requiring pregnancy testing the day of neuroimaging will use this sample to perform the pregnancy test. This sample will be used for a urine dipstick test for controls.
(3) Participant will drink 350 mL of water, about 5 minutes,
(4) Participant will be asked to rate the severity of urgency verbally (**Query 1**), about 1 minute,
(5) Participant receives MRI scan instructions,
(6) After about 20 minutes from Query 1, participant will be asked to rate the severity of urgency verbally (**Query 2**), about 1 minute,
(7) Participant will be asked to wear pull-ups (or diapers) before going into the scanner room, about 4 minutes,
(8) MRI set-up and localizer scans, about 6 minutes,
(9) Participant will be asked to rate the severity of urgency verbally (**Query 3***), about 1 minute,
(10) **Urgency RSfMRI scan (RS1):** resting state functional connectivity MRI data acquisition (3.0 mm^3, TR = 2.2s, 10 minutes), during which the participants will be asked to stay still and awake while looking at a cross hair.
(11) Participant will be asked to rate the severity of urgency verbally (**Query 4**), about 1 minute,
(12) Participant will exit the scanner, and void, which will be measured using a measuring cup, about 5 minutes,
(13) Second MRI set-up and localizer scans since the subject has moved, about 6 minutes,
(14) Participant will be asked to rate the severity of urgency verbally (**Query 5****), about 1 minute,
(15) **Empty bladder RSfMRI scan (RS2):** resting state functional connectivity MRI data acquisition (3.0 mm^3, TR = 2.2s, 10 minutes), during which the participants will be asked to stay still and awake while looking at a cross hair,

(16)Participant will be asked to rate the severity of urgency verbally (**Query 6*****), about 1 minute,
(17)3-D Magnetization prepared rapid acquisition gradient echo (MPRAGE), high resolution T-1 (1.0 mm^3), about 6 minutes,
(18)Participant will be asked to rate the severity of urgency verbally (**Query 7******), about 1 minute,
(19)**DTI scan**: (Spin-Echo EPI, 60 dir sequence, several *b*=0, 2X25 or 60 *b*=1000 mm2/s, 2.0 mm^3), 11 minutes,
(20)Participant will leave the MR scanner,
(21)Participant will be asked to rate the severity of urgency verbally (**Query 8**), about 1 minute.

If a patient is unable to hold their bladder until the next step is finished, a contingency plan has been developed so patients may void and then complete the rest of the scan (see manual of operation for details).

It is anticipated that scanner time will be approximately 60 minutes (up to 70 minutes).

The time between finishing water ingestion to the end of the Urgency RSfMRI scan is about 52 minutes.

<u>Standardization of MRI scans across all participating LURN sites:</u>

Washington University is responsible for all aspects of MRI (RSfMRI and DTI) images including determining specific MRI pulse sequences, site qualification, quality assurance/quality control (QA/QC) of all MRI data, tracking all MRI data acquisition and processing, and performance of all MRI data processing. To assure that acquisition sequences are standardized, Washington University will provide each participating LURN site with these protocols or confirm that the site's routine scanning protocols will be adequate.

Each participating site will identify a neuroimaging lead (a LURN investigator who will be responsible for overall performance of the site, including subject recruitment and data quality), a protocol lead (a personnel such as study coordinator to ensure the neuroimaging sequences described above are followed), and a technical lead (a personnel usually from the imaging center that executed the specific MRI parameters).

Initial calibration for the LURN Imaging protocol will be conducted using both a physical phantom and a human subject. The phantom will be the fBIRN agar ball phantom. The phantom will be imaged with the 3-plane localizer, T1-weighted MP-RAGE, resting state fMRI, and diffusion imaging from the LURN imaging protocol (see **Figure 4**). The human test subject will be scanned using the entire LURN imaging protocol except that the prior water ingestion will not be performed (starting with the second S+L localization, then RS2, T1 and DTI). The agar phantom and the human subject exam should be completed within one week of each other. All imaging data for these two calibration scans will be uploaded to the central imaging repository. The scans will be reviewed for image quality and scan parameters by the central imaging data repository. The sites will notified if they passed the site qualification or if changes need to be made to the protocol and qualification scans reacquired.

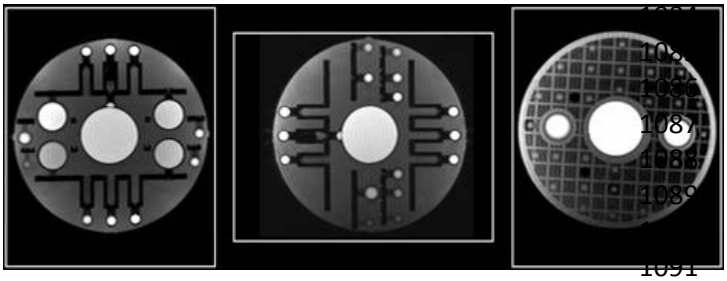


Figure 4: MP-RAGE phantom acquisition.

In addition, one human volunteer at each site (a different person at each site) will be scanned once as a human phantom to further enhance multi-site QA/QC and standardization. This individual will be used to obtain scans that can be used for quality assurance. The phantom exam for the human volunteer should be the same as the agar gel phantom exam.

After the initial site qualification scan, the fBIRN phantom will be scanned on a yearly basis to assess scanner stability. It will be the responsibility of the DCC to notify sites when the calibration scan is required. The DCC will also monitor the timeliness of data uploads to the central imaging data repository. Scanner stability will be assessed and issues concerning scanner stability will be sent to the sites in a timely manner.

QA/QC and standardization will be certified by the central site before site activation and subject enrollment. QC/QA will also be assessed on a regular basis by the central site based on: (1) the anatomic scan data quality, (2) functional connectivity data quality, and (3) diffusion tensor data quality. Scans that failed QC/QA (as determined by the central site) will be addressed and resolved as an ongoing basis between the technical personnel of the central site and the participating LURN sites.

Data transfer and data analysis:

A LURN neuroimaging internet site will be set up at the central imaging repository. Through this site, all investigators will be easily able to upload subject images to the central image site. All investigators will have access to images via the DCC upon request, both in raw and processed form.

The investigators and their technical team at the central imaging repository will be blinded to subject identity (e.g. HPI information and membership to the patient versus control groups) during neuroimaging data processing. Post-processing data will be shared with the DCC to further integrate with the broader demographics and deep phenotyping information.

Quantitative Sensory Testing (QST) Protocol:

Procedures for QST will adhere to standardized experimental protocols. Pressure pain testing will be conducted using the University of Michigan MAST system. Auditory sensitivity will be tested using a pure-tone audiometer. The MAST device and audiometer will undergo yearly calibration to maintain reliability and consistency across testing sites. Instructions will be scripted and participants will undergo extensive training before testing. All procedures have been evaluated for reliability and safety, and are well tolerated by urology patients, causing no more than temporary mild discomfort. However, subjects can stop testing at any time if the procedures become unbearable.

Overview of the neuroimaging sequence:

(1) MAST familiarization (left thumb), 5-7 minutes,
(2) MAST ascending series (right thumb), 5-7 minutes,
(3) Short break, 5 minutes,
(4) Hearing screening (left and right ear separately), 5 minutes,
(5) Auditory ascending series (left and right ears together), 5 minutes,
(6) Auditory randomized series (left and right ears together), 10 minutes,
(7) Short break, 5 minutes,
(8) MAST randomized series (right thumb), 5-7 minutes.

Figure 5: MAST handset

Pressure Pain Sensitivity:

Pressure pain sensitivity will be assessed using the MAST system. The MAST system is a non-significant-risk investigational device that applies a computer-controlled pressure stimulus to the thumbnail at a precisely controlled intensity for a specified duration. The MAST system consists of two tablet computers, one of which is an experimenter-controlled server that manages the test procedure, and the other a touch-screen patient interface that can display instructions and that the

participant uses to enter responses. The system also includes a hand-held force actuator, or handset, that applies pressure stimuli to the thumbnail bed. The handset is a pistol-grip-style unit manufactured in cast urethane for easy cleaning and ergonomically designed to be held comfortably in either hand by 95% of all U.S. adults with a slot into which the participant inserts his or her thumb (**Figure 5**). Pressure is applied to the participant's thumbnail by a conformal rubber probe with an area of 1 cm^2. The probe is attached to a cylindrical transducer driven by a miniature servo-motor. A dynamic, closed-looped control system uses digital load-cells to measure the exact pressures applied to the thumb, and self-adjusts motor output to the resistance of the thumb and any movement to ensure accurate and repeatable force delivery. The MAST System incorporates a series of redundant mechanical, electrical, and software safety features to prevent patient injury in the event of user error or device failure. MAST systems are currently being used in several clinical trials at the University of Michigan, and elsewhere, including the NIDDK MAPP Network. Of note, all LURN sites except one (Duke University) are also part of the MAPP network, and have the MAST equipment and trained personnel required for this test, thus streamlining its implementation.

Participants will undergo a familiarization procedure prior to the actual test. The purpose of familiarization is: 1) to teach participants how to perform the test correctly, 2) to reduce test anxiety, and 3) to acclimate participants to the sensations (pressures, sounds, etc.) experienced during the task. Pressures will be applied to the non-dominant (left) thumb during the familiarization procedure. One or two light "test" pressures (0.2 kg/cm^2, 2 seconds) will be applied to the participant's thumbnail in advance of the familiarization procedure to ensure proper thumb positioning. The familiarization test consists of a series of ascending pressures beginning at 0.5 kg/cm^2 and increasing in 0.5 kg/cm^2 increments. Each pressure will be delivered for 5 seconds. After the pressure is released, pain intensity will be rated on a 0-100 numerical rating scale (NRS) displayed on the patient interface (0 indicating "no pain", and "100" indicating "most intense pain imaginable"). The familiarization procedure will be terminated when the subjects reaches asks to stop the test, provides a pain intensity rating of $\geq$ 50/100, or a maximum pressure of 10 kg/cm^2 has been applied. This process will take approximately 5-7 minutes.

To assess pressure pain sensitivity, the MAST System will first deliver an ascending series of pressures (5-s duration; 4 $kg/cm^2/s$) to the dominant thumbnail at 20-s intervals, beginning at 0.5 kg/cm^2 and increasing in 0.5 kg/cm^2 steps. As during the familiarization procedures, pain intensity will be rated after each stimulus on a 0-100 numerical rating scale (NRS) displayed on the patient interface. The ascending series will be terminated when the subjects reaches his or her personal tolerance (i.e., wanting to stop), a pain intensity rating of $\geq$ 80/100, or a maximum pressure of 10 kg/cm^2. Patient responses obtained in the ascending series will be used to compute a set of 5 stimuli within that subject's range of tolerable pressures. The ascending series will take approximately 5-7 minutes.

After auditory sensitivity testing (see below), the patient's range of tolerable pressures will be delivered and rated 2X each (5-s duration, 20-s inter-stimulus interval) in random sequence. This procedure will require approximately 5-7 minutes to complete.

Auditory Sensitivity:

In the auditory sensitivity portion of the protocol, we will examine whether individuals with urinary urgency exhibit increased sensitivity to sounds, also termed hyperacusis. Each subject will complete a hyperacusis questionnaire through which they will indicate their experiences of real-life auditory sensitivity. The questionnaire will be administered with other questionnaires shortly before the neuroimaging procedures, or shortly after exiting the scanner if the participant needs more time to complete the questionnaire.

After the familiarization and ascending series of the MAST testing (see above), a hearing screening will

be performed according to American Speech-Language-Hearing Association guidelines for screening hearing impairments in adults. This involves a brief case history and a 25 dB HL pure-tone screen at 1000, 2000, and 4000 Hz in both ears separately using a calibrated audiometer (MAICO MA 33, MAICO Diagnostics, Eden Prairie, MN). Testing will be conducted in a quiet environment using earphones. Participants who fail to respond to either (left or right) of the 2000 Hz screening tones will be excluded from further testing. Participants who fail to respond to 1000 Hz or 4000 Hz tones can proceed with the auditory testing, provided they have passed both the left and right 2000 Hz screening tones.

For the auditory sensitivity test, participants will listen to a series of audiometer-generated pure tone acoustic stimuli. A total of 6 tones, 3 seconds in duration, will be presented binaurally at ascending intensity levels (40-90 dB, 2000 Hz). After each tone, the subject will rate separately the intensity and unpleasantness of the tone on standard numerical rating scales. Ratings are from 0-100 with the endpoints "none" to "most intense imaginable" or "most unpleasant imaginable." If a participant cannot tolerate or does not wish to hear a sound above a certain level (e.g., 80 dB), the ascending series will be stopped and the participant will not be presented with louder tones.

After the ascending series, the randomized series will begin without a break. In the randomized series, participants will be presented with up to 6 tones, three times each in random order. If a participant previously indicated that he or she could not tolerate or did not want to hear a sound at or above a certain level (e.g. 80 dB), the sound and all louder sounds will be skipped during the randomized series.

The entire auditory screening and testing procedure will require 20 minutes to complete. Subjects' response to the auditory testing will be compared to their response for the pressure pain sensitivity paradigm, fMRI, and symptom data.

<u>Timeline of the Neuroimaging and Sensory Testing Studies:</u>

May 2015 to April 2017: protocol standardization, recruitment, scanning, and sensory testing

May 2017 onwards: data analysis and preparation of publications

4.3.2 Enrollment

The target population will be clinic patients with complaints of urinary urgency, with or without urgency incontinence, usually with frequency and nocturia, consistent with the symptom complex commonly known as overactive bladder (OAB). Although the exact diagnostic workup of urinary urgency is at the discretion of the treating physician, it is recommended that the workup outlined in the 2012 AUA/SUFU OAB Guideline be followed. In addition, only patients who are eligible for, consented for, and are being extensively phenotyped in the LURN Observational Cohort Study will be eligible. This is to ensure that phenotyping data and biospecimens are available for participants who underwent neuroimaging and sensory testing. Additionally, age-matched healthy controls without urinary urgency and other LUTS will be recruited.

The LURN consortium plans to recruit participants with early (minimal) and late (severe) symptomatic disease to reflect the spectrum of patients to be seen in clinics. Participants in the neuroimaging and Sensory Testing Study will reflect this distribution. In the Neuroimaging and Sensory Testing Study, half of the participants with urgency will have significant urgency incontinence, which can be considered a more severe form of the syndrome, and the other half will have no significant urgency incontinence, see **Table 10.** As we also want to investigate potential differences in pathophysiology between patients who are able to maintain continence at the time of urgency to urinate, versus those who cannot maintain continence (e.g. differences in connectivity to motor area of the brain controlling pelvic floor function), both patient groups (with or without incontinence) will be recruited. Although we are recruiting both patient groups to ensure variability in the sample, the severity of incontinence will be treated as a

continuous spectrum in analyses.

Participants will be recruited equally across the following groups (see sample size calculation section also).

Table 10: Recruitment Table

	Urgency with significant urgency incontinence	Urgency without significant urgency incontinence	Controls (without urgency or other LUTS)
Male	42	42	42
Female	42	42	42

Half of the participants will be males, and the other half will be females. In general, neuroimaging data and sensory testing data cannot be compared across sex. Within each sex, the three groups will be age-matched also, as age can be a confounding factor for the neuroimaging studies (e.g. white matter hyperintensities caused by chronic vascular conditions in older subjects may complicate DTI interpretation) and for sensory pain testing. To control for the effects of age, recruitment will be age-stratified (i.e., less than 60 years old, 60 years old and older) to prevent a skewed age distribution. We anticipate that men with incontinence or women without incontinence will take longer to recruit compared to the other groups in each sex. In the unlikely event of a subject withdrawing from the study before all three tests are completed, only participants with the full complement of neuroimaging and both sensory tests with usable data will be counted toward the total sample size.

It is anticipated that each of the six participating LURN sites will recruit about 42 subjects over the course of 2 years (or 7 subjects per site for each of the cells in **Table 10** above), for a total of 252 participants across the entire LURN Research Network. Participants will be compensated for their effort in this study.

4.3.3 Participant Selection

Only a subset of patients who enrolled in the LURN Observational Cohort Study (Project 1A) with urinary urgency (with or without urgency incontinence) will be eligible for the neuroimaging and sensory testing study. The inclusion and exclusion criteria are as followed:

Urgency subjects:

Inclusion criteria:

ALL of the following criteria have to be fulfilled to be eligible:

a. Enrollment in the LURN Observational Cohort Study, including collection of samples for biomarker analysis.
b. Symptoms of urinary urgency, with or without urgency incontinence, usually with frequency and nocturia, consistent with the 2002 ICS definition of overactive bladder (OAB).
c. Answered “sometimes”, “often”, or “always” on question 6 of **LUTS Tool – 1 month version** (“During the past month, how often have you had a sudden need to rush to urinate?”). Subjects

who answered "never" or "rarely" are not eligible since they are not deemed to have significant urgency symptom. Subjects will be assigned into two subgroups using the following:

For assignment into the sub-group with significant urgency incontinence:

d. Answered "sometimes", "often", or "always" on question 16b of the **LUTS Tool – 1 month version** ("How often in the past month have you… Leaked urine in connection with a sudden need to rush to urinate?")

For assignment into the sub-group without significant urgency incontinence:

e. Answered "never" or "rarely" on question 16b of the **LUTS Tool – 1 month version** ("How often in the past month have you… Leaked urine in connection with a sudden need to rush to urinate?").

Deferral criteria: (See the deferral criteria for the Observational Cohort Study in section 3.4.4.)

a. Microscopic hematuria

- Patient must undergo appropriate evaluation.

b. Positive urine culture.

- Patient needs to be treated and have a subsequent negative culture before he or she is eligible.

c. Current sexually transmitted infection.

- Patient needs to be treated and have a subsequent test before he or she is eligible.

d. Recent (within 6 months) pregnancy.

Exclusion criteria:

a. See exclusion criteria for the Observational Cohort Study in section 3.4.4. In addition,
b. Answered "never" or "rarely" on question 6 of **LUTS Tool – 1 month version** ("During the past month, how often have you had a sudden need to rush to urinate?"). *[Subjects who answered "never" or "rarely" to the urgency question are not eligible since they are not deemed to have significant urgency symptom.]*
c. Any contraindication to MRI scanning, including:*
    1) Left-handed individuals *[Handiness will influence the laterality analysis of imaging results.]*
    2) Participant has CNS diseases, including structural brain abnormalities (e.g., neoplasms, subarachnoid cysts), cerebrovascular disease, ongoing infectious disease (e.g., abscess), history of other neurological disease, including stroke or seizure disorders.
    3) Participant has claustrophobia: Potential participants will be questioned about possible discomfort with being in an enclosed space (e.g., MRI scanner). Those who report such problems will be excluded.
    4) Participant has vision or hearing impairments that would impede completion of study procedures.
    5) Participant has any metal implants, devices, or jewelry that would be unsafe in the MRI, or meets any other exclusionary criteria as specified by the MRI Screening form. Presence of InterStim bladder neurostimulator (whether or not it is functioning, or whether it is turned on or off) is a contraindication to MRI. Patients with a non-functioning InterStim may enter the LURN Observational Cohort Study but not into the Neuroimaging and Sensory Testing Protocol.
d. Any contraindication to QST sensory testing, including:*
    6) Current, habitual, or previous use (within the last 12 months) of artificial nails, nail enhancements, or nail extensions that cover any portion of the thumbnail. Exceptions,

including brief and/or occasional use, may be permissible at the discretion of the study team. *[Nail products interfere with pain testing at thumbnails.]*

7) Menière's disease or the use of a hearing aid in either ear. *[These will interfere with auditory testing.]*

e. Use of opioids, including tramadol, and sedatives including benzodiazepines, in the absence of a 1-week washout periods for those subjects undergoing neuroimaging and QST. If the participant is non-compliant with the 1-week washout, he/she will be EXCLUDED from the study and will not proceed to testing.

f. Participants are permitted to use the following medications on an as-needed basis: over-the-counter or prescribed analgesics (NSAIDs, acetaminophen), muscle relaxants, nasal decongestants (pseudoephedrine, phenylephrine); however, participants will be asked to refrain from taking these medications for a minimum of 24 hours prior to their QST and neuroimaging study visit. *[Pain medications will interfere with testing].* In addition, participants will be asked to refrain from the following prior to QST and neuroimaging visits: alcohol (24 hours), nicotine (2 hours), and caffeine (6 hours). A compliance check will be conducted at the start of the visit to determine if participants followed these instructions and to record instances of non-compliance. If the participant is non-compliant to these instructions, he/she can still PROCEED with the testing; however, this protocol deviation(s) will be recorded on the CRF.

(*Note: Contraindications to either MRI or QST will preclude recruitment since it is anticipated that participants will undergo both neuroimaging and sensory testing.)

Control subjects:

Controls will be individuals without urinary urgency or other LUTS. Controls should have no urinary frequency (<8 voids/day), nocturia (0-1 void/night), urgency, or any urinary incontinence, including urgency incontinence, as assessed by the LUTS Tool – 1 month version; in addition they must with minimal to mild LUTS as assessed by the **AUA Symptom Index** (AUASI <8).

Inclusion criteria:

ALL of the following criteria have to be fulfilled to be eligible as a control:

a. 18 years of age or older, and
b. Answered "1-3 times a day" or "4 to 7 times a day" on question 2 of the **LUTS Tool – 1 month version** ("During a typical day in the past month, how many times did you urinate during waking hours?"), and
c. Answered "none" or "1 time a night" on question 3 of the **LUTS Tool – 1 month version** ("During a typical night in the past month, how many times did you wake up because you needed to urinate?"), and
d. Answered "never" or "rarely" on question 6 of the **LUTS Tool – 1 month version** ("During the past month, how often have you had a sudden need to rush to urinate?"), and
e. Answered "never" or "rarely" on question 15 of the **LUTS Tool – 1 month version** ("During the past month, how often did you leak urine?"), and
f. Answered "never" or "rarely" on question 16b of the **LUTS Tool – 1 month version** ("How often in the past month have you… Leaked urine in connection with a sudden need to rush to urinate?"), and
g. **AUA Symptom Index** (7-item) scores of 0 to 7.

Deferral criteria: (See the deferral criteria for the Observational Cohort Study in section 3.4.4.)

a. Positive urine culture.

- A urine dipstick will be performed. If positive for nitrite on the urine dipstick, recruitment is deferred and the subject is recommended to undergo a urine culture with their physician. Subject may be recruited if the urine culture result is negative. Subject with positive urine culture needs to be treated and have a subsequent negative culture before he or she is eligible.

b. Recent (within 6 months) pregnancy.

Exclusion criteria:

a. See exclusion criteria for the Observational Cohort Study in section 3.4.4.
b. A clinical diagnosis of overactive bladder (OAB).
c. Currently using medications specifically for LUTS (e.g., anti-cholinergics, beta-agonists, alpha-agonists, 5-alpha-reductases, PDE5-inhibitors for urinary problems).
d. A post-void residual of 150 CC or more
e. Contraindications to MRI scanning, as described in the *MRI exclusion criteria* above.
f. Contraindications to QST sensory testing, as described in the *QST exclusion criteria* above.

4.3.4 Schedule of Visits

For patients (urgency subjects with or without urgency incontinence), preferably, neuroimaging and sensory testing will be performed on the same day of the LURN Observational Cohort Study when the biological specimens and detailed questionnaire data are collected. This same-day-visit allows the strongest possible correlation between the different dataset (urologic and non-urologic symptoms, psychosocial measures, biomarkers, imaging, sensory testing) without large temporal gaps between them. If scheduling conflict does not permit a same day visit when the biological specimens and detailed questionnaire data are collected, neuroimaging and sensory testing should be performed within four weeks of that visit (prior to or after). Prior to scanning, the following questionnaires will be administrated: (1) **ICIQ-UI** (urinary incontinence), (2) **ICIQ-OAB** (overactive bladder), (3) **UDI-6** (urinary distress inventory), (4) **IIQ-7** (incontinence impact questionnaire), (5) **OAB-q** short form, (6) a symptom burden questionnaire **PSPS-Q**, (7) **BPI** (brief pain inventory), (8) a **hyperacusis** questionnaire, (9) the **MAPP-2 Body Map**, (10) an **Urgency Catastrophizing Scale**, modified from a pain catastrophizing scale, and (11) the **Complex Medical Symptom Inventory** (CSMI). Participants should drink what they normally would and should not be dehydrated prior to the imaging study. Neuroimaging should be performed **prior to** sensory testing on the same day, since residual effects of QST may interfere with RSfMRI results. Usually the QST testing equipment will be physically located in a different building from the functional MRI scanner.

Table 11: Schedule of Visits for Neuroimaging & Sensory Testing Case Subjects

	Observational Protocol Baseline/Initial Visit	**Neuroimaging & Sensory Testing Visit**
All Components of Baseline/Initial Visit for Observational Protocol (Project 1A) Listed in Table 2, including biosample and DNA collection	**X**	
fMRI		**X**
Additional Surveys Listed Above in Section 4.3.4		**X**
MAST Testing		**X**
Auditory Stimulation		**X**

For control subjects, they will provide the detailed questionnaire data and the biological specimens in the same manner as patients who enrolled in the Observational Cohort Study. Blood, urine, saliva, and genital swabs will be collected for storage at the NIDDK Sample Repository for future study by the LURN investigators and the broader research community. Preferably, neuroimaging and sensory testing will be performed on the same day when the biological specimens and detailed questionnaire data are collected. If scheduling conflict or logistic issues does not permit a same day visit when the biological specimens and detailed questionnaire data are collected, neuroimaging and sensory testing should be performed within a four weeks of that visit (prior to or after).

Table 12: Schedule of Visits: Neuroimaging and Sensory Testing Control Subjects

	Screening Assessment (remote)	Neuroimaging & Sensory Testing Visit
Eligibility Assessment	**X**	
Screening Demographics	**X**	
LUTS Tool – One Month		**X**
AUA Symptom Index		**X**
On-Line Self Report Questionnaires		**X**
Urine Analysis (dipstick)		**X**
Pregnancy Test		**X**
Biosample Collection (Blood, Urine, Saliva)		**X**
Genital Swab Collection		**X**
DNA Collection		**X**
fMRI		**X**
Additional Surveys Listed Above in Section 4.3.4		**X**
MAST Testing		**X**
Auditory Stimulation		**X**

4.3.5 Data Collected

Resting state functional MRI (RSfMRI):

Overview of RSfMRI Data: Two sets of data (**Urgency RS1 & Empty bladder RS2**) will be obtained from each subject. By comparing the "empty bladder" and "urgency" scans within the same subject, we will examine the status of the various resting state networks when the bladder is empty versus when the subjects reported urgency.

RSfMRI Regions of Interest (ROIs) Selection: The first RSfMRI analysis will use a standard set of ROISs selected within different networks — Default, Dorsal Attention, Ventral Attention, Auditory, Vision, Somatosensory, and Cognitive/Control will be used. Additionally we will include specific regions of interest (ROIs) that have previously been proposed to be involved in bladder control (e.g.

periaqueductal gray, pontine region, insular cortex, anterior cingulate gyrus, frontal cortex, cerebellum, pontine micturition center, and pre-optic hypothalamus). While traditional RSfMRI and DTI studies tend to focus on cortical areas, here we will expand our analyses to include the brainstem. For example, we will examine brainstem areas such as the (periaqueductal gray and rostral ventrolateral medulla which may be involved in descending control of bladder function and sensation.

Functional Connectivity: Correlation coefficients based on the time-course of BOLD signals will be estimated amongst brain regions creating a connectivity matrix (giving correlations between pairs of brain regions). Functional correlation maps will be produced by extracting the BOLD time course from a seed region (a ROI within a network of interest), then computing the correlation coefficient between that time course and the time course from all other brain voxels. Correlation values will be converted to a normal distribution using Fischer's r-to-z transformation and a random effects analysis corrected for multiple comparisons will be performed. A composite RSfMRI map for each of the distinct networks will be calculated for each subject by averaging the z scores from each of the ROIs of the respective network. Group averages will be overlaid on structural brain images and compared for changes in inter- and intra-network average connectivity using the methods in Brier et al. Much of this analysis will be done using an open-source, software package for structural and functional analyses of the cerebral and cerebellar cortex developed at Washington University as part of the Human Connectome Project. Activation levels by brain region as well as connectivity will be examined.

The canonical resting state networks in each group average with will be classified using the methods presented in one of our papers. Briefly, a fine grain connectivity matrix between all gray matter voxels will be created for each subject, similar to that described for ROIs above. The connectivity matrices within each group will be averaged and classified using the fuzzy c-means algorithm producing average maps of the canonical resting state networks in the two groups. A comparison of the different networks between groups will be performed using the technique of Support Vector Machine.

Diffusion Tensor Imaging (DTI):

The raw diffusion data will be converted to DTI data using the standard log linear least squares method. DTI parameters will include:

(a) Mean Diffusivity — The overall average value of water diffusion, not sensitive to the direction of diffusion.

(b) Fractional Anisotropy — The extent to which water diffusion has directional asymmetry. Typically normal white matter tracts have high anisotropy, and injured tracts have lower anisotropy.

(c) Axial Diffusivity — The diffusion value of water in the fastest direction, along the predominant direction of the axons.

(d) Radial Diffusivity — The diffusion in directions perpendicular to the axons fibers.

DTI Region Of Interest (ROIs) Selection: Measurement of DTI parameters will be performed in selected ROIs and will be compared between groups. ROIs will include:

(a) Regions To Assess Global Measures Of White Matter Structural Integrity — Centrum semi-ovale, frontal, parietal, and occipital white matter.

(b) Somatosensory Regions — Thalamus and the subcortical precentral gyrus.

(c) Attention Regions — Intraparietal sulcus, temporoparietal junction, and ventral frontal cortex.

(d) Specific White Matter Tracts that might be involved with urgency — ATR (anterior thalamic radiation), UNC (uncinate fasciculus), IFO (inferior fronto-occipital fasciculus), SFO (superior longitudinal

fasciculus), IFO (inferior longitudinal fasciculus). Studies suggested that ATR and SLF might be involved in urgency.

Pressure Pain Sensitivity:

Pain ratings from the ascending and random tests obtained from the MAST system will be used to compute psychophysical functions of each subject's pressure pain sensitivity, with pressure intensity (kg/cm^2) and response magnitude (0-100 NRS: intensity or unpleasantness) represented on the x- and y-axes, respectively. These curves will be used to compare single subject and group differences in pain sensitivity by analysis of slope and area of the curve (AUC). In addition, a modified three-parameter logistic model will be used to fit stimulus-response data from the ascending series. The midpoint between the minimum and maximum stimulus intensity will be estimated within-person using the SAS NLIN procedure to derive a measure of suprathreshold pressure pain sensitivity, referred to as Pain50. Pressure pain threshold (PPT) and pressure pain tolerance (Tol) will also be determined for each subject from the ascending series. PPT is defined as the first pressure in a string of at least two consecutive pressures that elicited a NRS pain rating > 0. Tol is the last pressure recorded in the stimulus response profile.

Auditory sensitivity:

Participant responses to the hyperacusis questionnaire and auditory screening will be collected. Patient intensity and unpleasant ratings (0 to 100 NRS) of auditory tones will be processed in manner consistent to the pressure ratings. For each subject, stimulus-response functions will be created with sound intensity (dB) and mean response magnitude (0-100: intensity or unpleasantness) represented on the x- and y-axes, respectively. Overall ratings loudness intensity and unpleasantness for the entire procedure will also be collected for each participant. Stimulus response curves will be used to compare single subject and group differences in auditory sensitivity.

4.3.6 Sample Size and Power Calculations

A total of 252 participants who are able to complete all three tests (MRI, pressure and auditory tests) with usable data will be recruited across all participating LURN sites. See **Table 10 (recruitment table)** above. One-third (n=84) will be urgency patients with urgency incontinence, one-third (n=84) will be urgency patients without urgency incontinence, and one-third (n=84) will be age-matched normal volunteers (controls) without urgency or other LUTS. Half of the participants will be males (n=126), and half will be females (n=126). Age will be evenly distributed across two age bins (<60, ≥60 year old). Among the 252 participants, all of them will undergo both neuroimaging and sensory testing.

T-tests will be used to compare functional connectivity, diffusivity and anisotrophy, sensitivity thresholds and tolerances between urgency patients and controls. **Table 13** shows the statistical power for various effect sizes. The power calculations presented below assume that associations are unadjusted for confounding factors. Adjusting analyses for age, sex, or other characteristics of the samples will provide at least as much power as an unadjusted analysis and in many cases substantially more power.

Table 13: Statistical power to detect the given effect size using two-sample t-tests using the entire sample, n=252 participants total.

	Effect size		
Number enrolled	0.4	0.5	0.6
168 with urgency, 84 controls	0.847	0.961	0.994

Note: An effect size of 0.4 would be achieved assuming a whole-sample average correlation standard deviation of 0.2, a control group mean of 0.5, and a group mean of 0.58. For an effect size of 0.6, the group mean would be 0.62.

Since RSfMRI and DTI have never been published in any urinary urgency studies, it is difficult to calculate the sample size precisely. Based on other RSfMRI and DTI studies, we estimate that n=84 for each of the three groups (urgency with urgency incontinence, urgency without urgency incontinence, and matched healthy controls) will allow us to detect an effect size of 0.5 with >90% statistical power, using two-tailed analysis with alpha=0.05.

Statistical power to assess associations of incontinence severity and functional connectivity, diffusivity and anisotrophy, sensitivity thresholds and tolerances will be based on correlation coefficients as a proxy for the linear regression used at the analysis stage.

Table 14: Statistical power to detect the given correlation among urgency patients, n=168 participants total

	Correlation coefficient (r)		
Number enrolled	0.20	0.25	0.30
84 with urgency incontinence versus 84 without urgency incontinence	0.743	0.908	0.979

We estimate that n=84 for each of the sub-groups (with urgency incontinence, versus without urgency incontinence) will allow us to detect a correlation coefficient of 0.25 with >90% statistical power, assuming alpha=0.05 (see **Table 14**).

4.3.7 Statistical Analysis

First, we will report descriptive statistics of the characteristics of the participants. Descriptive statistics will include frequencies and percentages for categorical variables, and means, standard deviations, and ranges for continuous variables. Variables will also be examined separately by subgroups, such as by LURN clinical site sex, race and ethnicity.

Aim 1

Aim 1 focuses on examination of RSfMRI and DTI results. We will examine the distribution of incontinence severity, diffusivity and anisotrophy. Functional connectivity will already be normalized as described in section 4.3.4. For ROI pairs of interest, we will use t-tests to compare the mean values of the subject-specific correlation estimates among urgency patients and controls. We will use linear regression to examine associations of these connectivity measures with incontinence severity, adjusted for the subject's age as needed. Additional investigations will use linear regression to control for other

demographic characteristics and potentially confounding variables. Variable inclusion will be guided by a best subsets approach, with the final model being the one with the highest likelihood score statistic or explained variance in which all covariates are statistically significant at $p < 0.05$. Multivariable models will be adjusted for LURN clinical sites whenever appropriate. We will also consider the distribution of activation levels at each ROI and compare the distribution among subject groups both visually and using t-tests to test whether mean levels differ between subjects with vs. without a full bladder, between those with vs. without urinary urgency, and between those with urgency with vs. without incontinence.

Between-groups comparison will also be performed in males and females separately: urgency as a whole [n=84 each sex] versus control [n=42 each sex].

Aim 2

In Aim 2, we will be focused on sensory testing results. We will examine the distribution of sensitivity thresholds and tolerances. If the thresholds and tolerances are normally distributed or can be transformed to achieve normality, we will use t-tests to compare the mean values among urgency patients and controls and Pearson correlations to examine associations with incontinence severity. If the thresholds and tolerances are not normally distributed, we will use Wilcoxon rank sum tests or other non-parametric tests to compare urgency patients and controls and Spearman rank correlations to examine associations with incontinence severity. As in Aim 1, between-groups comparison will also be performed in males and females separately and multivariable models will be examined.

Aim 3

Aim 3 involves an examination of both the RSfMRI and sensory testing results. We will use Pearson correlations or Spearman rank correlations, as appropriate, to examine associations between functional connectivity and sensory thresholds and tolerances. As in Aim 1, between-groups comparison will also be performed in males and females separately and multivariable models will be examined.

Aim 4

For Aim 4, we will examine whether patients' RSfMRI and sensory testing results are related to variables collected in other LURN phenotyping studies. We will examine the distribution of self-reported storage symptoms (collected in the Observational Cohort study), biomarker load (assessed in the Biomarker study), and urodynamic/urethral sensitivity (assessed in the Organ Based study). We will use Pearson correlations or Spearman rank correlations, as appropriate, to examine associations between these variables and functional connectivity and sensory thresholds and tolerances. As in Aim 1, between-groups comparison will also be performed in males and females separately and multivariable models will be examined.

4.4 Project 1B Timeline

Key Tasks	**Target Completion Date**
Study Coordinator Training	*March 6, 2015*
Steering Committee Approval of the revised Protocol	*April, 2015*
IRB submission of the Protocol	*May, 2015*
IRB approval of the Protocol	*August, 2015*
Study site orientation/ activation	*August, 2015*
Begin subject enrollment	*September, 2015*
End enrollment	*September, 2017*

5 Project 1C – Biomarker Pilot Protocol

5.1 Introduction and Overview

According to request for applications (RFA) that initiated the Lower Urinary Tract Dysfunction Research Network (LURN) study, one of the main goals of the whole LURN project is to find biomarkers of symptom initiation, flare, and progression. We consider the pilot project described below as an important step towards this goal.

Project 1C describes a pilot exploratory study for a potential, larger-scale project aimed at determining whether the biologic signatures measured by the SomaLogic assay SomaScan can distinguish unique subtypes of lower urinary tract symptoms (LUTS). The present protocol provides the rationale for the exploratory biomarker platform, a justification of the "bottom-up" approach for establishing biomarker signatures, the Specific Aims and Methodologies of the pilot project, and a description of how the results of the pilot will provide the foundation for a larger-scale biomarker study.

5.2 Goal of Biomarker Working Group (BWG): A Larger-Scale Biomarker Study

The long-term aim of the BWG is to determine biomarkers that can identify unique subgroups of men and women with LUTS.

LUTS has a negative impact on health-related quality-of-life (HRQoL) and has far-reaching effects on personal functionality and productivity. Based on a prevalence of 19%, the social costs of symptoms of lower urinary tract dysfunction (LUTD) have been estimated at nearly $25 billion per year [6]. LUTD is associated with many systemic diseases, e.g. type 2 diabetes mellitus [7], obesity [8], and atherosclerosis [9]. As such, its pathophysiology and response to common medical and surgical therapies is not uniform.

LUTS are prevalent and commonly experienced by both men and women [10]. Previous research has identified a multitude of risk factors outside of the genitourinary (GU) system (herein referred to as non-urologic factors), including depression, anxiety, psychological distress, diabetes, obesity, aging, and genetic predisposition, which directly contribute to the development and severity of LUTS [7,8, 9, 11, 12, 13]. As such, LUTS have been extraordinarily difficult for researchers to fully characterize. Therefore, the overarching purpose of the LURN is to carry out deeper phenotyping studies that can ultimately improve upon the characterization and treatment of men and women with LUTS.

LUTS are difficult for patients to adequately describe and for clinicians to characterize and treat. This is largely due to the fact that the patient experience of urinary symptoms can be variable and that the presence and severity of urinary symptoms may be the result of a multitude of pathological processes. Therefore, it is difficult to define subtypes of LUTS based only on the predominant symptoms reported by patients. The identification of subtypes of LUTS based on factors other than self-reported symptoms is critical to advance our understanding of LUTS pathology and to effective clinical management and treatment of LUTS. Novel tools that can accurately quantitate the presence, types, and severity of LUTS are needed, and biological markers are one such type of tool.

There is a need to identify biomarkers that can ultimately be used in clinical practice as a tool to provide a quantitative measure of the presence and severity of a patient's LUTS. Biomarkers can provide unique data that are complementary to clinical variables in distinguishing subsets of patients with specific urinary disorders, or can be predictive of differences in response to treatment. Furthermore,

identification of biomarkers can provide additional insights into the pathophysiologic mechanisms underlying LUTS in men and women. Evidence-based biomarkers could provide a tool for clinicians to "personalize" treatment strategies for their patients in order to initiate more effective treatments and monitor clinical response.

The premise for our long-term study is based upon the fact that the pathophysiology underlying LUTS is heterogeneous in origin, or that a common symptom might be caused by various underlying mechanisms. Many clinicians believe that this may explain why patients who report similar urinary symptoms respond differently to the same therapy. Therefore, the ultimate goal of a future study is to perform classification based on the levels of biomarkers, without regard to symptoms. It is anticipated that this unbiased approach will enable better understanding of molecular mechanisms of subtypes of LUTD and potentially personalized targeted interventions. In order to reach these long-term goals, this protocol is focused on a novel "bottom-up" approach, as described below. Most researchers have used a "top-down" approach, in which patients with different types of urinary symptoms were first identified and then biomarkers were measured and compared among the groups or with controls. The "bottom-up" approach is different, in that we propose to first identify clusters of patients based on "biomarker signatures", i.e. groups of up- and down-regulated biomarkers and then compare the clinical characteristics based upon these clusters. To achieve this long-term "bottom-up" goal, we will measure the concentrations of a large panel of biomarkers contained within the biospecimens obtained from randomly-selected LUTS patients recruited in the LURN Observational Cohort Protocol. We will also measure the same biomarker panel in a group of control subjects without LUTS matched by age, race, sex, and comorbidities. Unsupervised classification will be performed as a way to identify distinct biomarker groups/signatures. These biomarker groups/signatures will be compared with clinical characteristics and self-reported symptoms of LUTS.

5.2.1 SomaLogic Platform

The SomaLogic assay is a commercially-available test that measures a large panel of biomarkers representative of many different pathways. The test demonstrates exceptional dynamic range, quantifying proteins that span over 8 logs in abundance (from femtomolar to micromolar) and excellent reproducibility (4.6% median %CV). Specifically, the test measures the levels of 1310 proteins (including 330 inflammatory, 80 neurological, 180 stress response, 110 metabolic/endocrine, 70 aging-related, 70 renal and fibrosis markers, and 180 immune response biomarkers, among others). The SomaLogic assay measures proteins that have been implicated in a wide range of physiologic and pathologic processes. This is, therefore, an ideal platform to apply towards a project aimed at phenotyping LUTS since their etiology is multifactorial.

The SomaLogic platform has been previously used to characterize protein profiles in many disease phenotypes, including cancer, bowel disease, and aging. Multiple scientific groups have demonstrated the platform's high sensitivity (38 femtomole [fMol] limits of detection [LOD]) and reproducibility at a 4.6% coefficient of variation level. (See link to SomaLogic publications list (n=48): http://www.somalogic.com/Resources/Publications.aspx.) SomaLogic's assay is used in other National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) studies (e.g., Childhood Liver Disease Research Network [ChiLDReN], where serum of children with biliary atresia is assayed) and was recently licensed by the National Institutes of Health (NIH). Of particular relevance to the LURN study is a recent SomaLogic paper describing the identification of 11 proteins as a molecular signature of aging. The initial SomaLogic study on 202 subjects was confirmed in an independent study of 667 subjects, and validated in 384 subjects by using RNA-Seq technology [14].

5.2.2 Unsupervised Classification Methodology

Unsupervised classification is a common methodology in many LURN study protocols. This is a well-established pattern classification technique [15] that incorporates statistical methods, including k-means clustering, fuzzy k-means clustering, hierarchical clustering, principle component analysis, nonlinear component analysis, independent component analysis, multidimensional scaling, and self-organizing maps. Recently, this group of methods was complemented by an even more sensitive classification technique called topological data analysis [16], which proved to be useful in a broad range of multidimensional data analysis applications, from detecting subtypes of breast cancer [17] to exploring the states of folding pathways of biopolymers [18], and classification of the voting patterns of the Members of the U.S. House of Representatives [19]. Unsupervised classification (including clustering) is a well-established field with numerous applications in both research and clinical medicine [15, 17, 18].

These experimental techniques have biological relevance for characterizing disease phenotypes, such as LUTD. For example, a topological data analysis approach of gene expression microarray data was used to identify a subclass of Estrogen Receptor-positive (ER+) breast cancers that express high levels of c-MYB and low levels of innate inflammatory genes. When looking back at clinical data, this subclass of patients with this particular molecular signature exhibited 100% survival and no metastasis. The group has a clear and statistically distinct molecular signature, which highlights coherent biology but would not have been identified if classical techniques had been utilized [17]. Another recent example of these methodologies includes a study that was designed to determine whether biomarkers could classify a group of patients with inflammatory bowel disease that experienced different clinical outcomes or phenotypes [20]. This study involved only 35 patients with inflammatory bowel disease. The authors determined the gene expression profiles of patients and blindly binned them into different subgroups based upon their expression levels using unsupervised clustering techniques. After the biomarker groups were determined, the biomarker phenotypes were compared with clinical outcomes. Interestingly, the clusters were able to predict clinical outcomes. Taken together, these studies demonstrate how a "biomarker-driven approach" can be used to define clinical phenotypes.

5.3 Rationale

Before embarking upon a large-scale project, several questions and concerns need to be answered. For example, it is currently unknown whether SomaScan can be used to measure proteins that are relevant to LUTS. As such, it is important to determine if there are differences in the concentrations of proteins in patients with LUTS compared with those without LUTS symptoms.

While previous studies have identified proteins contained within plasma and urine that are associated with the presence and severity of LUTS, the ideal biologic specimen for this purpose remains unknown. Therefore, it would be prudent to determine the best biospecimen medium (plasma vs. urine) that will provide the most robust results.

Finally, while biospecimens have been collected as part of the LURN Observational Cohort, before a large-scale study is performed it will be important to verify if the samples were appropriately collected by Research Sites for SomaScan assay.

Although SomaScan was used in multiple biomarker studies of complex common diseases, it has not been applied towards patients with LUTS. Therefore, prior to using this methodology in a large-scale study, we will perform a pilot project, to inform the design of a future study. In this pilot study, we plan to assess the feasibility of conducting a larger-scale study by:

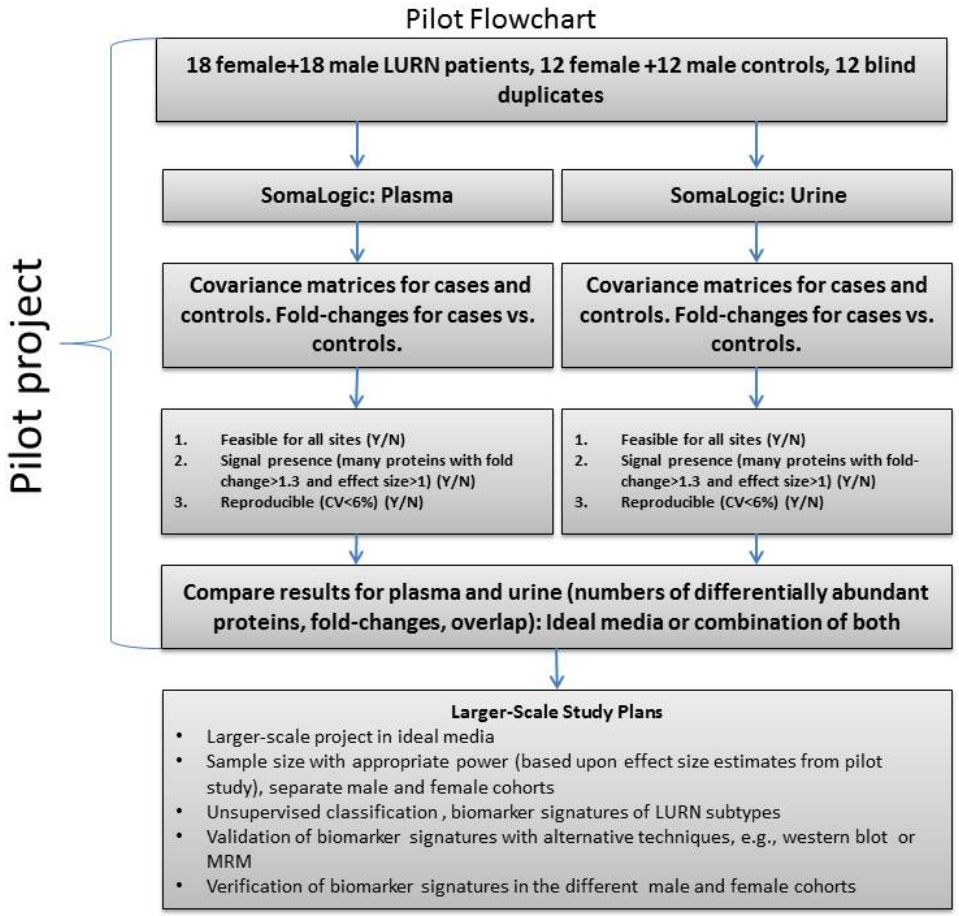
- determining the ideal media for measuring proteins related to LUTS (plasma, urine, or both)

- estimating the number and effect size of biomarkers in a LUTS signature group
- estimating covariance matrices of SomaScan targets in LUTS patients and in controls
- evaluating the reproducibility of SomaScan assay by using blind duplicates of the samples
- evaluating the quality of sample collection/storage at each of the six LURN Research Sites

5.4 Materials and Methods

A schematic overview for the pilot project is presented in Figure 6.

Figure 1: Pilot Biomarker Study Flowchart



5.5 Participant Selection

For this study, females with LUTS (n=18) and males with LUTS (n=18) that are enrolled in the LURN Observational Cohort Study will be blindly and randomly selected (3 males and 3 females from each participating LURN site) from individuals with severe LUTS symptoms (at least one symptom with severity level 4 or higher, as justified in Section 4.3.3). 24 evaluable controls, (12 females and 12 males), without LUTS will also be recruited for this study (see Inclusion and Exclusion criteria below). Controls will be frequency matched with cases by age, sex, race, body mass index (BMI), diabetes status, and LURN institutional site. Controls should not have significant LUTS, but are not required to be completely

healthy. Exclusion criteria for controls are the same as exclusion criteria for LURN patients, plus the presence of LUTS. The detailed inclusion and exclusion criteria for controls are presented in Section 4.1.2. Plasma and urine samples for selected cases will be requested from the NIDDK Biorepository. In order to evaluate SomaScan reproducibility, one female and one male case and one female and one male control subjects will be randomly selected. For each of these controls, 3 additional aliquots will be prepared and labeled in a way that does not allow identifying them as duplicates. These blind duplicate samples will be added to the whole collection of samples (18 female LUTS patients + 18 male LUTS patients + 12 female controls + 12 male controls + 12 blind duplicates = 72). The total of 72 plasma and 72 urine samples will be analyzed with SomaScan at SomaLogic, Inc. The proposed numbers of cases and controls in the pilot study are based on literature recommendations [21, 22, 23, 24] for the sample size in pilots, which vary from 10-15 per group to 24-36 per group, but are generally not more than 10% of the planned larger-scale study.

We will record the quality of ongoing biospecimen collections for LURN by obtaining the details surrounding the specimens being used in this pilot study (e.g., timing of collection, time required to place at -80 degrees, etc.). Only samples that have been processed within 2 hours from collection will be used for this study. We will send the biospecimens to SomaLogic for analysis on the SomaScan. The SomaScan will be used to analyze a protein panel of 1310 proteins contained within the plasma and urine. Results will be sent directly to the LURN data coordinating center (DCC). Data derived from plasma and urine will be handled independently.

5.6 Methodology

5.6.1 Evaluation of reproducibility of SomaScan

We will evaluate the reproducibility (measurement error) of a SomaScan assay by performing it on blind replicates of plasma and urine samples from two different control subjects and two different LUTS subjects. We will randomly select male and female cases and controls and will send for blinded analysis the original sample, plus three additional aliquots for each of these subjects. If possible, we will send these in different batches to include both intra- and inter-assay variability in the assessment of reproducibility. The presence of four replicates for each of the two cases and two controls will enable us to estimate the standard deviation, mean, and coefficient of variation (CV) for both controls for each of the 1310 biomarkers. These CV estimates will be compared with the distribution of CV levels across the biomarkers (median 4.6%, interquartile range 3.9% to 7.3%) advertised by SomaLogic and provided in their SOMAscan™ Proteomic Assay Technical White Paper available online (http://www.somalogic.com/Technology/SOMAscan-basic-info.aspx).

5.6.2 Evaluation of the quality of sample collection/storage

Strict adherence to the sample collection and storage procedure (for plasma not more than 2 hours from hand to -80 refrigerator) is crucial for the success of SomaScan assay. Output of the SomaScan assay is a 1310-dimensional vector with each dimension characterizing abundance of one of 1310 target proteins in a given patient sample. Length of this vector:

$$L_i = \sqrt{\sum_{j=1}^{1310} A_{ij}^2}$$

(i-subject index, j-protein index, A_{ij}-abundance of protein "j "in the sample of patient "i") represents overall protein abundance for each patient sample. If L_i is substantially smaller for the patient sample from one of the sites relative to other sites, it could mean that samples at this site were not

collected/stored properly. Sample collection and storage at this site will be reexamined and recommendations will be provided to study coordinators.

5.6.3 Estimation of the number and effect size of potential biomarkers; Pathway analysis; Covariance matrices for cases and controls

The relative abundances of the majority of the proteins in biological systems are not 'fine-tuned' and can vary both in time for a given control and across controls. However, it has been suggested that the relative abundances of potential biomarkers of LUTS are different in cases compared with controls to a larger extent than the natural biological variability. It is typical in proteomics studies to consider proteins differentially abundant when abundance differs by more than 30% from the mean normal value for the given protein [25]. Therefore, we will first estimate the number of differentially abundant proteins by comparing protein concentrations in cases versus matched controls. Counting all differentially abundant proteins in the above case-control pairs will provide us with the upper estimate of the number of potential biomarkers. We will then calculate mean abundances of each protein in cases and controls separately. Comparison of the mean abundances of the proteins in all cases versus all controls will provide us with the lower estimate of the number of differentially abundant proteins (potential biomarkers). Comparison of the difference of the mean protein abundances in cases and controls with the standard deviation of the abundances of this protein in the controls (measure of natural variability) will provide the rough estimate of the effect size of the potential candidate biomarker.

Lists of potential candidate biomarkers generated as described above will be submitted into the pathway analysis software, MetaCore (Thomson Reuters) and geneXplain (geneXplain), for enrichment analysis to determine the most affected pathways in LUTS cases versus controls. Pathway analysis will provide us with the information on whether the potential candidate biomarkers are independent or likely regulated by several common master regulators. We will also calculate covariance matrices for protein abundances in cases and in controls to evaluate if the observed differentially abundant proteins are correlated.

5.6.4 Comparison of the results for plasma and urine. Determination of the ideal biological media

The analysis described above will first be performed separately for plasma and urine samples. Then we will compare the lists of potential biomarkers generated from plasma and urine samples and determine if and to what extent they overlap. We will also determine if the differentially abundant proteins observed in plasma and urine belong to the same pathways and if there is strong correlation of protein abundances observed in plasma and urine. Based on the above comparison, we will decide if the combination of plasma and urine data provides important additional information, or if one of the media is sufficient for the study.

5.6.5 Determination of the feasibility of a larger-scale biomarker study

It is necessary to exercise caution when using effect size estimated from the pilot project since the 95% confidence interval can be quite large due to the limited sample in the pilot study [22]. Nevertheless, the pilot study provides information about the most probable values of the effect size and therefore decreases uncertainty in the design of the larger-scale study. In our case of the multiple outcomes, levels of abundance of the potential candidate biomarkers, we can use the pilot project to estimate the likelihood of a certain number of potential biomarkers to be up- or down-regulated relative to controls, with the effect size above certain threshold value. With this information and the information on the covariance matrices candidate biomarkers for LUTS subjects and for controls, we can calculate expected misclassification error by using the 'in-house' developed simulator of unsupervised learning [26]. Our preliminary simulations showed that misclassification error below 5% (across 5 biomarker-based

clusters) is expected in cases of 40 differentially abundant proteins out of 1310 having effect size ≥1.2, when the sample size of the case cohort is ≥150. This number of differentially abundant proteins is not unusual for proteomics studies; for example: (1) 44 proteins were found significantly differentially abundant in the SomaScan study of serum of 51 patients with Duchenne muscular dystrophy versus 17 age-matched controls [27]; (2) 248 differentially abundant proteins were observed in the study of cerebrospinal fluid of patients with age-related neurodegeneration versus controls [14]; (3) 239 proteins were shown significantly differentially abundant in the SomaScan study of serum of 39 patients after 8 weeks of pulmonary tuberculosis treatment relative to the baseline [28].

Results of the pilot study will provide us with the estimates of the number and effect size of the differentially abundant proteins and therefore will enable more accurate estimation of the sample size for the larger-scale biomarker study. Importantly, as described above, it will also provide information on the feasibility of the larger-scale study by evaluating the reproducibility of SomaScan assay and quality of sample collection/storage at each site.

Caution needs to be exercised when combining data from the pilot study with data from the larger-scale main study, especially if important changes in the protocol are implemented based on the results of the pilot [22]. Since the pilot study is of substantial size and cost, we plan to take all measures (e.g., unbiased random selection of the subjects for the study) to retain the possibility of combining data from the pilot study with the larger-scale study. We do not anticipate changes to the protocol, other than possible elimination of either urine or plasma from the larger-scale study, as described in Section 5.7.

5.6.6 Potential limitations/pitfalls

The main goal of this pilot project is to determine the feasibility of the large-scale biomarker study based on the unsupervised clustering approach to discovery of biomarker signatures of subtypes of LUTS by using the SomaScan assay. SomaScan technology is well-established, targets multiple biological pathways and processes relevant to LUTS, and has been used in more than 30 studies. However, it is possible, although not very likely, that it will fail in detecting a substantial number of differentially abundant proteins (potential candidate biomarkers) in LUTS cases versus controls. If this happens, we will examine other assays, e.g., targeted multiple reaction monitoring (MRM) proteomics and metabolomics, to search for potential candidate biomarkers.

5.7 Future Directions: Large-Scale Biomarker Study

As stated above, the results of the pilot study will provide answers to many questions, including the ideal media to perform future analyses and the feasibility of the study. We anticipate that both urine and plasma media will demonstrate differences between cases and controls. As mentioned above, we will ultimately endorse the medium that contains the most proteins with large effect size differences between cases and controls. Based upon the effect sizes noted, we will be able to define whether a larger study is feasible and determine the required sample size.

With these results, we will plan for a larger study, likely using one medium (plasma or urine). This larger study will include larger cohorts of women and men (both cases and controls) enrolled in the LURN Observational Cohort Study (sample size to be determined based on the results from this pilot study). It is expected that these studies will yield meaningful clusters of biomarkers associated with specific subtypes of patients with LUTD.

If the potentially large study yields positive results, i.e., determine the biomarker signatures of LUTS subtypes, we plan to verify those findings by looking at these signatures with the alternative analytical techniques, e.g., western blot or targeted MRM proteomics, and then test in validation cohorts.

5.8 Statistical Analysis in Large-Scale Study

In the large-scale study, we will perform unsupervised clustering by using and comparing the results of several classification algorithms, including k-means clustering, fuzzy k-means clustering, hierarchical clustering, nonlinear component analysis, independent component analysis, multidimensional scaling, and self-organizing maps. These methods generate complementary information, e.g., hierarchical clustering is useful for revealing the substructure of the groups, while nonlinear component analysis helps when interactions of the candidate biomarkers are of importance. We will perform the above unsupervised classification analysis with functions available using MATLAB software, with the Bioinformatics and Statistical Toolboxes, and will evaluate and compare the quality of clustering with the MATLAB function "*evalclusters.m*", which calculates four commonly used criteria for comparison of within-cluster and between-cluster distances.

In a separate step, we will combine our data on the differentially abundant candidate biomarkers with existing biological knowledge of metabolic and signaling pathways and networks by using the MetaCore (GeneGo, Thomson Reuters) mapping and enrichment analysis software tools. We will repeat all the above unsupervised classification procedures at the level of pathways. The advantage of the pathway level analysis is that it: (1) helps to reveal the biological meaning and the mechanism of the discovered effect; (2) decreases the role of biological variability; and (3) typically improves the significance level. In the above analysis, we will correct the significance levels for multiple testing by using the Benjamini-Hochberg false discovery rate control procedure [29], which allows keeping type I error as desired, with much lower type II error (and therefore higher power) than a Bonferroni correction. Therefore, adding candidate biomarkers cannot hurt, but can increase likelihood of biomarker discovery.
Finally, we will reveal the LUTS for the analyzed cases and compare the symptom-blinded and the symptom-based classifications. We will examine whether some of the symptom-based clusters are represented by two or more distinct biomarker-based clusters. The last step will be to combine biomarker and symptom information and perform clustering based on the combined information. We will evaluate how the combination of clinical and biomarker data improves the quality of characterization and suggest the combined "clinical symptoms plus biomarkers" diagnostic/predictive tool for further validation.

5.9 Potential Limitations of the Large-Scale Study/Pitfalls for "Biomarker-Driven" Approach

The proposed study is centered around the "bottom-up" or "biomarker-driven" approach, which is considered to be a novel methodology in the study of LUTS. We believe that this sort of novel approach is greatly needed for a deep phenotyping and understanding of LUTS and LUTD. However, there are some potential limitations of this methodology that have to be considered. For example, we may ultimately need a larger sample size to develop meaningful biomarker clusters that can distinguish LUTS phenotypes. While previous studies of other disease phenotypes have utilized much smaller sample sizes, it is possible that the biomarkers associated with LUTS are much more complex and involve even greater sample sizes. If the results of clustering are contradictory across the classification methods, we will need to increase the sample size for the large scale biomarker project.

Another limitation of this methodology involves the LUTS phenotypes. Previous studies using this technique have involved disease processes with relatively discrete pathologic findings and clinical outcomes. The present study proposes to study phenotypes that are not necessarily associated with

concrete pathologic findings. While we view this as an advantage to our analysis, it is possible that biomarkers will not be able to cluster without well-defined pathologic pathways/processes. In addition, it is possible that we may uncover clusters of biomarkers that are associated with the presence of clinical characteristics other than LUTS (e.g., diabetes and BMI). To avoid this possibility, we will utilize control subjects (matched for race, age, comorbidities) to correct for the presence of these variables.

5.10 General Methodology for Subject Enrollment and Biospecimen Collection

Participants of this pilot study will include patients with LUTS (18 male and 18 female) randomly selected from the Observational Cohort of the LURN Phenotyping Study Protocol and controls without LUTS (12 male and 12 female) recruited separately for this pilot study. Study participants with LUTS in the LURN Phenotyping Study Protocol will have met eligibility requirements, signed informed consent, and provided biospecimens for use by LURN and other investigators. Since enrollment of the patients with LUTS is described in Section 3.4.3, we will not repeat it here and will concentrate on the procedure for selection of control subjects for the pilot study.

Similarly, the biospecimen collection procedure for the LUTS patients is already described in Section 3.4.8 and will not be repeated here. Biospecimens, including whole blood, serum, plasma, saliva, genital swabs and urine collected from all participants enrolled in the LURN Phenotyping Study Protocol and from the controls recruited for this sub-protocol will be stored at the NIDDK Biorepository for use. A formal requisition request will be made prior to disbursement of any biological specimens related to LURN.

5.11 Overview of Study Participant Enrollment

5.11.1 Selection of Study Participants with LUTS for the Biomarker Pilot Study

Patients with LUTS will be selected randomly and blindly to symptoms and demographics from the participants of the LURN Phenotyping Study in order to get a representative sample of possible subtypes of LUTS. We will use a threshold for selection based on the severity level of LUTS to avoid the situation where some of the subtypes will be presented by the patients with low levels of severity (potentially possible due to small sample size of the pilot). Table 1 in Appendix AD provides information on the severity levels of LUTS in patients recruited to the LURN Phenotyping Protocol as of December 2, 2015. Using an inclusion criterion “at least one symptom with severity level >=4” allows selecting 60% of males and 74% of females uniformly distributed across the Research Sites, and therefore provides a representative pool for random selection of LUTS patients for the pilot study. After the random selection of the Biomarker Pilot Study patients, their demographics (i.e. age, race) and BMI and presence or absence of diabetes will be revealed, recorded, and used for selection of controls with the frequency matched demographics, obesity, diabetes, and LURN site.

5.11.2 Recruitment of Controls for the Biomarker Pilot Study

Controls for the pilot study will be recruited after completion of selection of cases as described above. Therefore, information on the desired frequency match in terms of demographics, obesity, diabetes, and LURN site will be available to the recruiting study coordinators at the Research Sites. Information on the controls already recruited for the Biomarker Pilot Study will be made available to study coordinators in a timely manner so that they will know what type of controls (demographics, obesity, diabetes) are still missing. That will permit coordinators to recruit the remaining controls with the appropriate characteristics. We anticipate that the number of recruited controls and biospecimens collected could be twice higher (e.g., 24 males and 24 females, with each site recruiting 4 male and 4 female controls) than required for this pilot study. Twelve male and 12 female controls’ samples will be selected for

SomaScan analysis based on the best matching cases (above) with LUTS. The rest of the samples will be stored for the future study. The aim of moderate over-recruitment of controls is to ensure the proper frequency matching with the blindly selected LUTS cases.

Importantly, to be considered for inclusion into the controls, volunteers should meet the criteria below. Upon providing informed consent and meeting entry criteria, controls will come to the Research Site for one visit to donate biospecimens.

Inclusion criteria for controls:

a. Answered "1-3 times a day" or "4 to 7 times a day" on question 2 of the **LUTS Tool** – 1-month version ("During a typical day in the past month, how many times did you urinate during waking hours?"); and
b. Answered "none" or "1 time a night" on question 3 of the **LUTS Tool** – 1-month version ("During a typical night in the past month, how many times did you wake up because you needed to urinate?"); and
c. Participants respond "never" or "rarely" on every other item of the LUTS Tool; and
d. Age ≥ 18 years old; and
e. The ability to give informed consent; and
f. American Urological Association Symptom Index (7-item) scores of 0 to 7 (This would exclude patients with significant obstructive symptoms.); and
g. Normal urinalysis.

Exclusion criteria for controls:

a. Currently undergoing or have previously received treatment for LUTD;
b. Have reported or been treated for a urinary tract infection in the past 90 days;
c. Gross hematuria;
d. Significant neurologic disease or injury, including but not limited to: cerebral vascular accident with residual defect, Alzheimer's disease, dementia, Parkinson's disease, traumatic brain injury, spinal cord injury, complicated spinal surgery, multiple sclerosis;
e. Primary complaint is pelvic pain;
f. Diagnosis of interstitial cystitis, chronic prostatitis, or chronic orchialgia;
g. Pelvic or endoscopic GU surgery within the preceding 6 months (not including diagnostic cystoscopy);
h. Current sexually transmitted infection;
i. Ongoing symptomatic urethral stricture;
j. History of lower urinary tract or pelvic malignancy;
k. Current chemotherapy or other cancer therapy;
l. Pelvic device or implant complication (e.g., sling or mesh complication);
m. Current functioning neurostimulator;
n. Botox injection to the bladder or pelvic structures within the preceding 12 months;
o. In men, prostate biopsy within the previous 3 months;
p. In women, pregnancy;
q. History of cystitis caused by tuberculosis, radiation therapy, or Cytoxan/cyclophosphamide therapy;
r. Augmentation cystoplasty or cystectomy;
s. Presence of urinary tract fistula,
t. Current major psychiatric disorder or other psychiatric or medical issues that would interfere with study participation (e.g., dementia, psychosis, etc.);

u. Inability to relay valid information, actively participate in the study, or provide informed consent (includes uncontrolled psychiatric disease);
v. Have received pelvic radiation;
w. Have an elevated post-void residual (PVR) urine volume >150 ml;
x. Medical expulsion therapy for symptomatic kidney or ureteral stone within 90 days;
y. Microscopic hematuria;
z. Individual must undergo appropriate evaluation;
aa. Positive urinalysis or urine culture;
bb. Individual needs to be treated and have a subsequent negative culture, and wait at least 90 days before he or she is eligible;
cc. Recent (within 6 months) pregnancy;
dd. Breastfeeding.

5.11.3 Schedule of Visits for Controls

We will recruit controls without more than minor symptoms (as defined in inclusion criteria above) LUTS from the community. Recruitment will be aided by online advertisements (e.g., Craigslist). Controls will have a single baseline visit. During this visit, biospecimens will be collected, including whole blood, serum, plasma, saliva, genital swabs and urine; controls will also answer LUTS tool and AUA questionnaires, as well as the battery of self-reported measures outlined in Section 3.4.7. Plasma and urine will be used in this pilot study; the rest of the samples will be kept for future studies.

Table 15: Schedule of Visits for Biosample Pilot Protocol Control Subjects

	Initial Visit
Eligibility Assessment	X
Demographics	X
General Clinical Information	X
Clinic Testing (Urine Analysis)	X
LUTS Tool (one month recall period)	X
Self-report Questionnaires, including CASUS LUTS Tool – 1 week recall period AUA Symptom Index	X
Biosample Collection (Whole Blood, Serum, Plasma, Saliva, Urine)	X
Genital Swab Collection	X

5.11.4 Urine Collection for Controls

Urine has great utility as a testing matrix. It is easily accessible, can be collected noninvasively, and provides information on numerous physiological processes. Urine is a source of numerous potential biomarkers, including metabolites, cells, proteins, and nucleic acids. To be used successfully for biomarker discovery and validation, various urine specimen parameters must be harmonized, including collection method, volume collected, and timing of collection, processing, and storage. As such, we have previously outlined standard methodology for urine collection (refer to Biomarker Collection section of the LURN Phenotyping Study Protocol Manual of Operations [MOO]). This is briefly summarized below:

1) Either a catheterized specimen or a spontaneously voided mid-stream specimen will be obtained from female patients, depending on provider practice. Male patients will provide spontaneously voided mid-stream specimens.

2) A total volume of 50-100ml will be collected, and aliquots of uncentrifuged urine will be obtained.
3) Urine will be processed as described in the MOO.

5.11.5 Blood Collection for Controls

Blood and its components are commonly used as a testing matrix associated with minimal risks to study participants. Controls enrolled in this pilot study will be asked to donate blood for analysis. It is recommended that controls will be fast for 8 hours prior to blood draw unless it is medically contraindicated. The time of the last meal and the time of blood draw will be recorded. Standard protocols for blood collection have been developed and are described in detail in the Observational Cohort Study MOO. Briefly, we will collect blood for serum and plasma studies. Plasma samples will be collected according to the manufacturer's (SomaLogic, Inc.) protocol and processed as recommended. All plasma will be processed and stored at -80 degrees Celsius within 2 hours from collection.

6 Human Subjects

6.1 Protection of Human Subjects

6.1.1 Institutional Review Board

This study and analysis will be performed under Institutional Review Board (IRB) oversight. Prior to the initiation of the study, an IRB approval for study of human subjects will be obtained separately from the IRB of each of the participating LURN clinical study centers and the DCC. Revisions to the study protocol and changes in the study design will also be submitted to the individual IRBs for approval prior to implementation.

Subjects will be enrolled in the LURN Phenotyping protocol with full and written informed consent, which will include the gathering of privileged health information (PHI) and permission to be contacted about possible participation in subsequent LURN studies.

Each participating center will be responsible for obtaining such human subjects research authorization and will create an informed consent document detailing the procedures described above in the language required by their respective organizations. All key personnel at the participating centers will have successfully completed IRB-required training and certification for human subjects research. Additionally, participants will satisfy HIPAA researchers' privacy requirements.

6.1.2 Patient Confidentiality

Special procedures for ensuring patient confidentiality will be implemented. Data transmission and the distributed data systems will have multiple layers of security as discussed in Section 8, Study Management. Each study subject will be assigned an identification number. Only this number will be used to identify subjects in any individual tabulation. The PHI that is collected will represent the minimum necessary to successfully execute the study.

PHI entered into the database at the site level will only be visible to study personnel accessed through a triple password regimen. The PHI is encrypted at the site level. Site personnel will have the decryption key, and it will not be available to the DCC. It is expected that only group data will be published. If individual subject data are to be published, no identifying information will be included. The study files will be maintained in a secure location. Access to computerized data will be restricted to study personnel. Password authorization will be enforced. Previous use of this security system and a secured server indicates that this technique is very successful in assuring the protection of confidential

information.

All neuroimaging images will be kept in the central imaging data repository in a de-identified manner using their study ID. The list linking subject back to the study ID will be kept by the study coordinators at the clinical sites. All folders with identifiable information will be password protected. The list of identifiers will be kept for 7 years after the completion of research. At which time the list will be destroyed and information will not be able to be linked back to individual participants. Washington University will not have access to the PHI information of participants from other participating LURN sites.

Authorized representatives of the Sponsor, the National Institute of Diabetes, Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH), participating LURN clinical study centers, DCC monitoring staff, as well as the IRBs at each site, will have access to medical records and records from participants in this study. Such access is necessary to ensure the accuracy of the findings.

The DCC has obtained and will maintain a Certificate of Confidentiality from the NIH. The Certificate prevents researchers from being forced to disclose participants' identifying information, even by a court subpoena, in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings.

6.1.3 Risks to the Patient and Adequacy of Protection Against Risk

Patients enrolled in the LURN Phenotyping Study will experience more than the normal amount of testing that is customary for patients with LUTD. Individuals may experience psychological discomfort in answering repeated, longitudinal assessment questions related to LUTS, demographic and clinical characteristics, health-related quality of life, self-reported pelvic floor function (bowel function, sexual function, and pelvic organ prolapse) and psychological factors (stress, anxiety, depression, sleep disturbance Venipuncture carries risks of pain and bruising at the puncture site. With respect to potential discomfort developing during clinical assessment, we note that study personnel will be trained by the investigators to be sensitive to participant discomfort and concerns.

There is a potential risk of breach of confidentiality that is inherent in all research protocols, and steps to minimize this risk are described above. Steps to minimize risk and address any psychological discomfort are addressed below. Recruitment and Informed Consent: At each LURN site, individuals eligible for Project 1A, the Observational Cohort Study (based on criteria described in Section 3.4.4) , Project 1B, the Neuroimaging and Sensory Testing Study (based on criteria described in Section 4.3.3), and 1C, the Biomarker Pilot Protocol (based on criteria described in Section 5.5, will be approached by a LURN investigator for release of their protected health information and contact information so that study staff may approach them to describe the study and obtain informed consent. All consent forms will be HIPAA-compliant. A copy of the signed consent forms will be kept by the study participant, and one will be kept in the research records at the site where the participant was enrolled. Participation in the Project 1A will require completion of standard clinical assessments, a survey comprising the self-reported measures in Section 5.5, and survey assessment with the LUTS Tool at intake and 3-months and 12-months after the intake assessment and CASUS at intake and 12-months after the intake assessment, or after a planned surgical intervention. We anticipate that these assessments will require 45-60 minutes to complete the survey. Participation in Projects 1B and 1C will require full participation in Project 1A.

Psychological discomfort during study procedures: (i.e., during study surveys): With regard to participants' psychological discomfort and overall well-being, we noted above that the study personnel will be specifically trained to be sensitive to subjects' discomfort and concerns. If a participant finds the research procedures to be upsetting, he/she will have the option to withdraw from the study. Subjects who express current/recent thoughts or an intention to harm him/herself or others or answer positively

to the current/recent answers to Questions 56-59 on the PSPS-Q will be referred immediately for psychological care. In this situation, confidentiality would have to be broken in order to protect the participant. The participant will be made aware of this contingency in the informed consent form.

Risks of MRI scanning for Project 1B: Functional MRI scans do not involve injections or any radioactive tracers. Although the long-term risk of exposure to magnetic fields and radiofrequencies associated with MRI is not known, the possibility of any long-term risk is extremely low in view of the information accumulated over the past twenty years. Some people experience dizziness or a metallic taste in their mouth if they move their head rapidly in the magnet. However, this is only a temporary effect, and is not experienced if the head is kept still. The scanner produces loud sounds at times and insulated earphones will be provided to reduce the audible noise. There may be slight discomfort associated with having bladder urgency or urinary incontinence inside the scanner. If unrestrained iron or steel objects are accidentally brought near the MRI magnet, they can be pulled very quickly toward the magnet and can strike people in or near the magnet. Such an event is very unlikely, because precautions are taken to prevent such objects from being brought near the magnet. Subjects are screened for iron or steel implants or clips from surgery, or metallic objects, such as shrapnel or metal slivers in their bodies, and are excluded from study if present. Dental fillings do not present a hazard.

There is a remote possibility that the fMRI will show an abnormal incidental finding either at the time it is performed or during a later review. If the incidental finding is noted at the time of the fMRI, the site's research staff will refer the subject for clinical follow-up. If the incidental finding is noted during a later review, then the central imaging data repository of Project 1B, the Neuroimaging Study, will contact the clinical site and inform them of the results and the subject's study ID. Then the research staff will refer the subject for follow-up clinical care. All interactions regarding incidental findings will be documented up through the referral step. It will be the subject's responsibility to access further clinical care once incidental results and clinical referral information are provided.

Risks of sensory testing for Project 1B: Pressure sensitivity testing may cause some temporary physical discomfort on the thumbnail. The MAST system includes multiple software, electrical, and mechanical safeguards to ensure that the amount of pressure applied does exceed safe limits, including a safety release pin that the subject can turn to immediately release the pressure actuator from their his or her thumb. The test is terminated at or before 10 kg/cm^2 of pressure which is a commonly used maximum pressure level in human sensory testing and does not result in physical injury. Participants will always have personal control over the stimulus and can stop it at any time or express instructions to stop the stimuli. They can also withdraw their thumb from the device. Auditory sensitivity testing may also cause some temporary unpleasantness. Maximum intensity level and duration of an auditory stimulus is 90 dB SPL presented for 5 s, with a minimum interval of 10 s before the next stimulus. These parameters are within the permissible range of safe noise exposure (OSHA 29 CFR 1910.95, Table G-16). Participants can stop testing at any time however if the auditory stimuli become unbearable.

6.1.4 Unauthorized Data Release

The data sets will be stored on a secure server with restricted access (requires a unique username and password) at the DCC and every precaution will be taken to keep the information private. However, there is always the possibility of unauthorized release of data about subjects. Such disclosure would be extremely unlikely to involve a threat to life, health, or safety. It is conceivable that such disclosure could have psychological, social, or legal effects on the patient. Using the standard security procedures (described above under patient confidentiality) can effectively minimize the risk of unauthorized disclosure of data. All study personnel who have access to patient data will be educated regarding the need to protect confidentiality and the procedures to be followed to ensure such protection. All staff

will also be required to sign a standard medical record confidentiality agreement. The computer system on which data are maintained uses standard password protection procedures to limit access to authorized users. After the study is completed, the database will be stored on the NIDDK Data Repository. The database in the Repository will be de-identified to obviate further privacy and security considerations.

6.1.5 Adverse Event Monitoring and Reporting

6.1.5.1 Definition of an Adverse Event

An adverse event (AE) is any untoward medical occurrence or unfavorable and unintended sign in a research subject that occurs during or as a result of a research procedure. For this study, each center will review the list of study procedures and identify the specific procedures that are not standard-of-care at their institution and these will be considered research procedures. Complications that are a result of research procedures will be reported and tracked as adverse events.

Since Project 1A is primarily an observational study, and research procedures (phlebotomy, survey response) present minimal risk, we anticipate few adverse events. The research procedures associated with Project 1B (fMRI, survey response, MAST and audiometer testing) are rarely associated with severe adverse events and are often considered to be no more than minimal risk to the subject. All adverse events must be recorded. The onset and end dates, severity and relationship to study procedure(s) will be recorded for each adverse event. All adverse events will be reported by LURN investigators to the LURN DCC. Any action or outcome (e.g., hospitalization, additional therapy, etc.) will also be recorded for each adverse event. Subjects will be questioned and/or examined by the investigator or his/her designee for evidence of adverse events.

6.1.5.2 Assessment of event severity and relationship to treatment

The modified World Health Organization (WHO) grading system will be used for grading severity of AEs (Appendix AE). For AEs not covered by the modified WHO grading system, the following definitions will be used:

Mild:	awareness of sign, symptom, or event, but easily tolerated
Moderate:	discomfort enough to cause interference with usual activity and may warrant intervention
Severe:	incapacitating with inability to do usual activities or significantly affects clinical status, and warrants intervention
Life-threatening:	immediate risk of death

The investigator must also assess the relationship of any adverse event to the research procedure, based on available information, using the following guidelines:

Unlikely related:	no temporal association, or the cause of the event has been identified; or the procedure cannot be implicated
Possibly related:	temporal association, but other etiologies are likely to be

	the cause; however, involvement of the procedure cannot be excluded
Probably related:	temporal association; other etiologies are possible, but unlikely

6.1.5.3 Definition of serious adverse events

A serious adverse event (SAE) is any adverse experience that results in any of the following outcomes:

- Death;
- Life-threatening AE (i.e., one that places the subject, in the view of the investigator, at immediate risk of death from the AE as it occurs);
- Persistent or significant disability/incapacity;
- Required in-patient hospitalization, or prolonged hospitalization;
- Congenital anomaly or birth defect.

Additionally, important medical events that may not result in death, be life-threatening, or require hospitalization may be considered a serious adverse event when, if based upon appropriate medical judgment, they may jeopardize the subject and may require medical or surgical intervention to prevent one of the outcomes listed in this definition.

6.1.5.4 Reporting Responsibility

All adverse events must be recorded. The onset and end dates, severity and relationship to study procedure(s) will be recorded for each adverse event. Any action or outcome (e.g., hospitalization, additional therapy, etc.) will also be recorded for each adverse event.

All AEs and SAEs must be reported by the investigator to the LURN DCC. The DCC will review reports of all related SAEs and other relevant information immediately, and may request additional information from sites for analysis of these events. Sites will report SAEs according to the time frames outlined below.

All events that are serious and related (possibly or probably) must be reported to the DCC within 24 hours of the investigator being informed of the event. Follow-up information about a previously reported serious and related adverse event may be reported to the DCC within 7 working days of the investigator receiving the information; however, important follow-up information must be submitted within 24 hours. All deaths connected to a study procedure must be reported to the DCC within 24 hours of the investigator being informed of the event.

6.2 Benefits to the Patient

There are no direct benefits to the patients for participation in the study.

6.3 Inclusion of Women

Approximately 50% of the study participants will be women. Recruitment will be monitored to ensure adequate representation of women.

6.4 Inclusion of Minorities

Racial and ethnic minorities will be recruited into the study. We anticipate that the representation of

racial and ethnic minorities will correspond to the fraction of minorities in the population presenting to the participating clinics as patients. Recruitment will be monitored to ensure that the representation of minority groups parallels the racial/ethnic composition of patients seen for visits at LURN Clinical Sites.

6.5 Inclusion of Children

Children under the age of 18 will not be enrolled into this study as the LURN physicians do not have children in the practices.

6.6 Data Safety and Monitoring Plan

Accepted principles of data and safety monitoring will be observed throughout the conduct of the LURN study. The NIH has appointed an independent External Expert Panel (EEP) that will provide study oversight. The EEP will review the study protocol prior to enrollment and will also review all subsequent protocol revisions. The EEP will also evaluate the occurrence of adverse events related to study participation as well as study accrual updates that include the demographics, clinical characteristics, and symptom profiles of enrolled patients to ensure maintenance of recruitment targets and clinical relevance of the study population.

LURN principal investigators will be responsible for monitoring the enrollment of subjects, submission of data to the DCC, and monitoring and reporting of adverse events related to study participation. The DCC will be responsible for monitoring for effective conduct of the protocol and accurate and timely data submission.

IRBs will be provided feedback on a regular basis.

Training of study coordinators and study monitoring activities will be conducted by the DCC to ensure patient confidentiality and privacy and to maximize the reliability, accuracy, and timeliness of study data.

The LURN Clinical Sites, the DCC, and relevant research center staff will conduct regular meetings to review recruitment/enrollment progress, data collection activities, and participant retention. The DCC will produce regular reports regarding enrollment, data quality, and timeliness and share the reports with NIDDK, the Steering Committee, and the participating clinical center. Data will be routinely exported from the system, examined for accuracy and completeness, and backed up to secure storage devices. Upon completion of data collection, final processing and cleaning of data will be conducted. A technical report detailing specific project methodology, response rates, and other details will be produced.

7 Study Organization

7.1 Clinical Centers

The participating LURN clinical study centers will have primary responsibility for developing the study protocol, maintaining high rates of follow-up and data collection, obtaining data of high quality, and interpreting, presenting, and publishing findings from the study.

Northwestern University
Chicago, IL
Principal Investigators: David Cella, PhD and Brian T. Helfand, MD, PhD

University of Iowa
Iowa City, IA
Principal Investigators: Karl J. Kreder, MD, MBA and Catherine S. Bradley, MD, MSCE

Duke University
Durham, NC
Principal Investigators:
Kevin P. Weinfurt, PhD (Steering Committee Co-chair) and Cindy L. Amundsen, MD

University of Washington
Seattle, WA
Principal Investigator: Claire C. Yang, MD (Steering Committee Co-chair)

University of Michigan
Ann Arbor, MI
Principal Investigator: J. Quentin Clemens, MD, FACS, MSCI

Washington University in St. Louis
St. Louis, MO
Principal Investigators: Gerald L. Andriole, Jr., MD and H. Henry Lai, MD

7.2 Data Coordinating Center

The DCC contributes biostatistical expertise and shares in scientific leadership of the research group. The DCC has developed a communication infrastructure that includes meetings, teleconferences, email and bulletins, interactive Web-based encounters, and written correspondence. The DCC assists in protocol development and preparation of scientific publications. The DCC has the major responsibility of creating a database and data collection systems for the participating LURN clinical study centers, ongoing evaluation of data quality, performance monitoring of the LURN clinical study centers, and statistical analyses of the data. The DCC has also created a comprehensive Manual of Operations (MOO) that will govern the conduct of the study. The manual details the protocols, protocol clarifications and amendments, summary of the regulatory requirements for the study, instructions for enrollment, data collection, data management, visit schedules, and detailed instructions on the use of the electronic data submission. The DCC is responsible for clinical monitoring of the study.

Arbor Research Collaborative for Health
Ann Arbor, MI
Principal Investigator: Robert M. Merion, MD, FACS

7.3 Steering Committee

The primary governing body of the study is the Steering Committee, consisting of each of the Principal Investigators of the LURN clinical study centers, the Principal Investigator of the DCC, and the NIDDK Project Scientist. The Steering Committee develops policies for the study pertaining to access to patient data, performance standards, and publications and presentations. It develops the study protocol and meets to discuss the progress of the study and to consider problems arising during its conduct. The Steering Committee may establish subcommittees to further develop specific components of the study protocol. Small working groups may be established to prepare manuscripts and presentations.

8 Study Management

8.1 Data Collection, Data Collection Forms, Data Entry

The DCC will utilize the Web-based ArborLink as the data management nucleus for the LURN phenotyping studies. ArborLink is a database platform developed by Arbor Research Collaborative for Health (Arbor Research). The DCC will utilize ArborLink to create electronic case report forms to capture

all relevant study data for the core study and all investigational/research protocols that are developed and implemented during the course of LURN. The ArborLink system allows real-time monitoring of study data for protocol adherence, quality assurance, adverse event reporting, discrepancy reporting, and other trends.

8.2 Data Management

Study data for Project 1a and 1c will be entered into the electronic data entry system by study coordinators at each study site. These data will be encrypted and transferred to the DCC and stored on a secure server at Arbor Research. Access to the server and data entry system is limited and requires a unique username and password combination. The servers are backed up daily and physically stored in a locked facility.

Study data for Project 1b will be collected in three ways. DTI and RSfMRI will be transferred from each site to the central imaging data repository for quality assurance and central reading (see Section 4.3.1). Data will then be sent to the DCC for incorporation into the final study-wide data sets and for analysis. Pressure pain sensitivity data will be saved on local MAST servers at each site and transferred to the DCC using a secured SSH file transfer server. Auditory sensitivity data and additional study data (such as protocol deviations and timing of study procedures) will be entered into ArborLink by study coordinators at each study site.

All analysis of the data sets will utilize de-identified (coded) data sets.

8.3 Quality Control and Database Management

The first steps in ensuring protocol compliance are good protocol design and careful orientation of study personnel. Following final agreement on protocols, and prior to study initiation at any of the LURN clinical study centers, the DCC will organize a Training and Certification session for LURN study coordinators/data entry personnel.

The electronic data entry system will have built-in data checks as part of study quality assurance. Protocol compliance will be assessed by monitoring the submission of data at required intervals. Data inconsistencies and discrepancy reports will be reviewed by the Clinical Monitors so that necessary queries can be generated and sent to the LURN clinical study centers for verification and resolution.

Periodic requests may be generated for the submission of random source documents to assess the quality of data acquisition and data entry at each site. In addition, the Clinical Monitor or Project Manager will visit each site at least once a year to review source documents, monitor regulatory compliance, and assess protocol adherence.

In addition to source document verification, the Clinical Monitor and Project Manager will produce reports from the database to look for inconsistencies in submitted data, particularly for repeated measures data elements, even if data do not fall outside of built-in validation routines.

Studies of intra-subject and inter-subject data variability by LURN clinical study center as well as intra-center and inter-center data variability will be used to further ascertain random or systematic data quality issues.

8.4 Data Security/ Data Transfer

For the Observational Cohort Study, personnel at each study center will collect and enter data into the Web-based data entry system. The following data security contingencies are in place:

- Compliance with Industry Standards Regarding Data Security (HIPAA and 21 CFR Part 11)

- Audit trails are maintained for all activity and all changes to any data element
- All servers, Web servers, firewalls, etc. are configured and maintained according to industry best practice guidelines for backup, security, continuity of operations, and protection of PHI
- All data are available only to authorized users from each site after secure login with encryption, with all site activity audited at the user level
- All transmissions between the Internet and the database are encrypted using a 128-bit encryption algorithm
- There is a comprehensive security plan in place

Detailed instructions on the use of the database platform, data element definitions, and a code list will be provided in a MOO. Each study site will be provided a copy of the MOO and the entire manual will be available on the study website, and in the Help area of the database user interface.

8.5 Resource Sharing Plan

During the study, data and biosamples will be shared with internal and external investigators according to the guidelines agreed upon by the Steering Committee.

Upon study completion, study data and materials will be transferred to the NIDDK Data Repository. Minutes of the meetings of the Steering Committee, Project Executive Committee, subcommittees, and the External Expert Panel will be kept on file at the DCC.

Whole blood for creation of cryopreserved lymphocytes and biosamples collected during the study will reside at the NIDDK Genetics and Biosample Repositories.

9 References

1. Andersson, K. and A. Wein, *Chapter 68: Pharmacologic Management of Lower Urinary Tract Storage and Emptying Failure*, in *Campbell-Walsh Urology, 10th edition*, K. LR, et al., Editors. 2011.
2. Coyne, K. S., L. S.. Matza, Z. S. Kopp et al, *Examining lower urinary tract symptom constellations using cluster analysis*. BJU Int, 2008. **101**(10): 1267-73.
3. Gong, X. and M. Richman, *On the application of cluster analysis to growing season precipitation data in North America east of the Rockies.* J of Climate, 1995. **8**(4): p. 897-931.
4. Ganesalingam, J., et al., *Latent cluster analysis of ALS phenotypes identifies prognostically differing groups.* PLoS One, 2009. **4**(9): p. e7107.
5. Tibshirani, R., G. Walther, and T. Hastie, *Estimating the number of clusters in a data set via the gap statistic.* J R Statist Soc B, 2001. **63**(2): p. 411-423.
6. Sexton CC, Coyne KS, Vats V, Kopp ZS, Irwin DE, Wagner TH. Impact of overactive bladder on work productivity in the United States: results from EpiLUTS. Am J Manag Care 2009;15:S98-S107.
7. Danforth KN, Townsend MK, Curhan GC, Resnick NM, Grodstein F. Type 2 diabetes mellitus and risk of stress, urge and mixed urinary incontinence. J Urol 2009;181:193-197.
8. Lee RK, Chung D, Chughtai B, Te AE, Kaplan SA. Central obesity as measured by waist circumference is predictive of severity of lower urinary tract symptoms. BJU Int 2012;110:540-545.
9. Russo GI, Castelli T, Privitera S, Fragala E, Favilla V, Reale G, Urzi D, La Vignera S, Condorelli R, Calogero AE, Cimino S, Morgia G. Increase of Framingham risk score is associated with severity of lower urinary tract symptoms. BJU Int 2015;116(5):791-796.
10. Kupelian V, Wei JT, O'Leary MP, Kusek JW, Litman HJ, Link CL, McKinlay JB, Investigators BS. Prevalence of lower urinary tract symptoms and effect on quality of life in a racially and ethnically diverse random sample: the Boston Area Community Health (BACH) Survey. Arch Intern Med 2006;166:2381-2387.
11. Parsons JK. Lifestyle factors, benign prostatic hyperplasia, and lower urinary tract symptoms. Curr Opin Urol 2011;21:1-4.
12. Abdollah F, Briganti A, Suardi N, Castiglione F, Gallina A, Capitanio U, Montorsi F. Metabolic syndrome and benign prostatic hyperplasia: evidence of a potential relationship, hypothesized etiology, and prevention. Korean J Urol 2011;52:507-516.
13. Aktas BK, Gokkaya CS, Bulut S, Dinek M, Ozden C, Memis A. Impact of metabolic syndrome on erectile dysfunction and lower urinary tract symptoms in benign prostatic hyperplasia patients. Aging Male 2011;14:48-52.
14. Baird G, Nelson S, Keeney T, Stewart A, Williams S, Peskind E, Montine T. Age-dependent changes in the cerebrospinal fluid proteome by SOMAmer array. Am J Pathology 2012;180(2):446-456.
15. Duda RO, Hart PE, Stork DG. Pattern classification, 2nd ed. Wiley, New York; 2001.

16. Carlsson G. Topology and data. Bulletin of the American Mathematical Society 2009;46:255-308.

17. Nicolau M, Levine AJ, Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proc Natl Acad Sci USA 2011;108:7265-7270.

18. Yao Y, Sun J, Huang X, Bowman GR, Singh G, Lesnick M, Guibas LJ, Pande VS, Carlsson G. Topological methods for exploring low-density states in biomolecular folding pathways. J Chem Phys 2009;130(14):144115.

19. Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, Carlsson G. Extracting insights from the shape of complex data using topology. Sci Rep 2013;3:1236.

20. Montero-Melendez T, Llor X, Garcia-Planella E, Perretti M, Suarez A. Identification of novel predictor classifiers for inflammatory bowel disease by gene expression profiling. PLoS One 2013;8:e76235.

21. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, Robson R, Thabane M, Giangregorio L, Goldsmith CH. A tutorial on pilot studies: the what, why and how. BMC Med Res Methodol 2010;10:1.

22. Hertzog MA. Considerations in determining sample size for pilot studies. Research in Nursing & Health 2008;31:180-191.

23. Johanson GA, Brooks GP. Initial scale development: Sample size for pilot studies. Educational and Psychological Measurement 2010;70:394.

24. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011;45(5):626-629.

25. Robotti E, Manfredi M, Marengo E. Biomarkers discovery through multivariate statistical methods: A review of recently developed methods and applications in proteomics. J Proteomics Bioinform 2014;S3: 003.

26. Andreev VP, Gillespie BW, Helfand BT, Merion RM. Simulation of unsupervised classification of lower urinary tract dysfunction patients based on the protein microarray data. 2nd International Caparica Conference on Urine Omics and Translational Nephrology. Caparica, Portugal, September 2015.

27. Hathout Y, Brody E, Clemens PR, Cripe L, DeLisle RK, Furlong P, Gordish-Dressman H, Hache L, Henricson E, Hoffman EP, Kobayashi YM, Lorts A, Mah JK, McDonald C, Mehler B, Nelson S, Nikrad M, Singer B, Steele F, Sterling D, Sweeney HL, Williams S, Gold L. Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. Proceedings of the National Academy of Sciences. USA 2015;112(23):7153-7158.

28. De Groote MA, Nahid P, Jarlsberg L, Johnson JL, Weiner M, Muzanyi G, Janjic N, Sterling DG, Ochsner UA. Elucidating novel serum biomarkers associated with pulmonary tuberculosis treatment. PLOS ONE 2013;8(4):e61002.

29. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289-300.

10 Appendices