

Shauna Leighton ORCID iD: 0000-0001-5700-1424

H. Henry Lai ORCID iD: 0000-0003-2691-994X

Claire Yang ORCID iD: 0000-0002-4750-6258

Longitudinal Changes in Symptom Based Female and Male LUTS Clusters

Cindy L. Amundsen MD¹, Margaret E. Helmuth MA², Abigail R. Smith PhD², John O.L. DeLancey MD³, Catherine S. Bradley MD MSCE⁴, Kathryn E. Flynn PhD⁵, Kimberly S. Kenton MD MS⁶, H. Henry Lai MD⁷, David Cella PhD⁶, James W. Griffith PhD⁶, Victor P. Andreev PhD DSc², J. Eric Jelovsek MD MMed¹, Alice B. Liu MPH⁸, Ziya Kirkali MD⁹, Claire C. Yang MD⁸ and the LURN Study Group

¹Duke University Medical Center, Durham, NC

²Arbor Research Collaborative for Health, Ann Arbor, MI

³University of Michigan, Ann Arbor, MI

⁴University of Iowa Carver College of Medicine, Iowa City, IA

⁵Medical College of Wisconsin, Milwaukee, WI

⁶Northwestern University Feinberg School of Medicine, Chicago, IL

⁷Washington University School of Medicine, St. Louis, MO

⁸University of Washington, Seattle, WA

⁹National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD

Short Title: Longitudinal Changes in Symptom Based Female & Male LUTS Clusters

Word Count: 2966

Declarations of Interest: Dr. Bradley, Dr. DeLancey reports grants from NIH/NIDDK, during the conduct of the study; Dr. Jelovsek reports other from UpToDate, from null,

This is the author manuscript accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/nau.24219](https://doi.org/10.1002/nau.24219).

outside the submitted work; Dr. Yang reports grants from National Institutes of Health, during the conduct of the study; Dr. Andreev reports grants from NIDDK, during the conduct of the study; Dr. Lai reports grants from National Institutes of Health, during the conduct of the study; other from Medtronic, other from Allergan, personal fees from Aquinox, personal fees from Teva, outside the submitted work; Ms. Helmuth reports grants from NIDDK, during the conduct of the study; Dr. Smith reports grants from NIDDK, during the conduct of the study; .

Corresponding Author:

Cindy L. Amundsen, MD

Department of Obstetrics and Gynecology,

Duke University Medical Center

5324 McFarland Drive, Suite 310

Durham, NC, USA

Tel: (919) 401-1001

Fax: 919-401-1033

Email: cindy.amundsen@duke.edu

Funding /Support

This study is supported by the National Institute of Diabetes & Digestive & Kidney Diseases through cooperative agreements (grants DK097780, DK097772, DK097779, DK099932, DK100011, DK100017, DK097776, DK099879).

Research reported in this publication was supported at Northwestern University, in part, by the National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgements

This is publication number 18 of the Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN).

The following individuals were instrumental in the planning and conduct of this study at each of the participating institutions:

Duke University, Durham, North Carolina (DK097780): PI: Cindy L. Amundsen, MD, Kevin Weinfurt, PhD; Co-Is: Kathryn Flynn, PhD, Matthew O. Fraser, PhD, Todd Harshbarger, PhD, Eric Jelovsek, MD, Aaron Lentz, MD, Drew Peterson, MD, Nazema Siddiqui, MD, Alison Weidner, MD; Study Coordinators: Carrie Dombeck, MA, Robin Gilliam, MSW, Akira Hayes, Shantae McLean, MPH

University of Iowa, Iowa City, IA (DK097772): PI: Karl Kreder, MD, MBA, Catherine S Bradley, MD, MSCE, Co-Is: Bradley A. Erickson, MD, MS, Susan K. Lutgendorf, PhD, Vince Magnotta, PhD, Michael A. O'Donnell, MD, Vivian Sung, MD; Study Coordinator: Ahmad Alzubaidi

Northwestern University, Chicago, IL (DK097779): PIs: David Cella, Brian Helfand, MD, PhD; Co-Is: James W Griffith, PhD, Kimberly Kenton, MD, MS, Christina Lewicky-Gaupp, MD, Todd Parrish, PhD, Jennie Yufen Chen, PhD, Margaret Mueller, MD; Study Coordinators: Sarah Buono, Maria Corona, Beatriz Menendez, Alexis Siurek, Meera Tavathia, Veronica Venezuela, Azra Muftic, Pooja Talaty, Jasmine Nero. Dr. Helfand, Ms. Talaty, and Ms. Nero are at NorthShore University HealthSystem.

University of Michigan Health System, Ann Arbor, MI (DK099932): PI: J Quentin Clemens, MD, FACS, MSCI; Co-Is: Mitch Berger, MD, PhD, John DeLancey, MD, Dee Fenner, MD, Rick Harris, MD, Steve Harte, PhD, Anne P. Cameron, MD, John Wei, MD; Study Coordinators: Morgen Barroso, Linda Drnek, Greg Mowatt, Julie Tumbarello

University of Washington, Seattle Washington (DK100011): PI: Claire Yang, MD; Co-I: John L. Gore, MD, MS; Study Coordinators: Alice Liu, MPH, Brenda Vicars, RN

Washington University in St. Louis, St. Louis Missouri (DK100017): PI: Gerald L. Andriole, MD, H. Henry Lai; Co-I: Joshua Shimony, MD, PhD; Study Coordinators: Susan Mueller, RN, BSN, Heather Wilson, LPN, Deborah Ksiazek, BS, Aleksandra Klim, RN, MHS, CCRC

National Institute of Diabetes and Digestive and Kidney Diseases, Division of Kidney, Urology, and Hematology, Bethesda, MD: Project Scientist: Ziya Kirkali MD; Project Officer: Christopher Mullins PhD; NIH Personnel: Tamara Bavendam, MD, Robert Star, MD, Jenna Norton

Arbor Research Collaborative for Health, Data Coordinating Center (DK097776 and DK099879): PI: Robert Merion, MD, FACS; Co-Is: Victor Andreev, PhD, DSc, Brenda Gillespie, PhD, Gang Liu, PhD, Abigail Smith, PhD; Project Manager: Melissa Fava, MPA, PMP; Clinical Study Process Manager: Peg Hill-Callahan, BS, LSW; Clinical Monitor: Timothy Buck, BS, CCRP; Research Analysts: Margaret Helmuth, MA, Jon Wiseman, MS; Project Associate: Julieanne Lock, MLitt

Abstract

Aim: Lower urinary tract symptoms (LUTS) are diverse in type and variable in severity. We examined symptom change within the LURN identified clusters over time and tested associations with treatments received.

Methods: Patient-reported LUTS and treatment data were collected at multiple time points between baseline and 12 months from The Symptoms of Lower Urinary Tract Dysfunction Research Network Observational Cohort study. LUTS Severity Scores were calculated to summarize changes in symptom reporting over time in previously identified LURN clusters. Repeated measures linear regression models tested adjusted associations between cluster membership and severity scores.

Results: 417 men and 396 women were classified into improved, unchanged, and worsened symptoms between baseline and 12 months [men: 44.1%, 40.5%, 15.3%; women: 55.8%, 33.1%, 11.1%, respectively]. Improvement in LUTS Severity Scores varied by cluster [estimated adjusted mean change from baseline range: -0.04 change in standard deviations of severity scores (Δ SD) to -0.67 Δ SD]. Prostate surgery was associated with improved severity scores [-0.63 Δ SD] in men, while stress incontinence surgery was associated with improved severity scores [-0.88 Δ SD] in women.

Conclusion: Symptom improvement varied by cluster indicating response to therapy differs amongst subtypes of patients with LUTS. The differential

improvement of patients in clusters suggests mechanistic differences between clusters and may aid selecting more targeted treatments in the future.

Keywords: Lower urinary tract symptoms; The Symptoms of Lower Urinary Tract Dysfunction Research Network Observational Cohort study; symptom-based clustering; urinary incontinence

Introduction

Lower urinary tract symptoms (LUTS) are diverse in quality and variable in severity, however, current treatment of LUTS is primarily based on the predominant symptom. Treatment decisions are largely a matter of patient preference and alternative treatments are recommended only when improvement is insufficient and symptoms are severe, resulting in a notable failure rate¹. In reality, many individuals report multiple urinary symptoms, and it might be that individuals with certain *groups* of symptoms should be assessed for treatment efficacy, rather than basing treatment efficacy on a single or predominant symptom.

The *Symptoms of Lower Urinary Tract Dysfunction Research Network* (LURN) was assembled in 2012 to increase our understanding of lower urinary tract symptoms (LUTS) by identifying the important subtypes of patients with LUTS²⁻³. The rationale for defining subtypes was the recognition that LUTS patients are a heterogeneous group with regard to presentation, severity, and symptom progression over time. LURN's first step to defining patient subtypes was to use a resampling-based consensus clustering approach. This, using the patient

baseline reported symptom data, allowed for the identification of novel symptom-based clusters⁴. Four male and four female clusters were identified, providing the basis for our current analysis. The first male cluster (M1) was the largest of the four male clusters and reported moderate to severe levels of storage and voiding symptoms, while the second male cluster (M2) reported moderate post-micturition symptoms (post-void dribbling and post-void leaking) along with some weak stream symptoms. The third male cluster (M3) was the least symptomatic, endorsing mostly moderate frequency symptoms. The fourth male cluster (M4) reported severe frequency, urgency and urgency incontinence symptoms.

As in males, the four female clusters were also characterized by multiple symptoms. The first female cluster (F1) reported frequency, post-micturition, and voiding symptoms, at a mild to moderate level. The second female cluster (F2) endorsed more severe frequency and urgency, as well as urgency incontinence. The third female cluster (F3), the largest of the four female clusters, reported moderate to severe levels of incontinence, in addition to frequency, urgency and mild voiding symptoms. The fourth female cluster (F4) reported a moderate to severe levels of all symptoms (Supplemental Table 1). Because complex conditions such as LUTS may be better understood by subtyping patients, LURN's conceptual framework incorporates the investigation of patients with LUTS at multiple levels, including patient's symptoms, characteristics of the genitourinary organs and comorbidities, and cellular/ molecular factors where by meaningful differences among patients may be identified². The male and female clusters were found to differ on self-reported mental health, bowel function, and sexual function measures, indicating that distinct patient subtypes were identified with the symptom-based clusters^{3,4}. In this analysis we examined the differential changes in symptom

severity by cluster when treated with standard therapy. But since symptoms are just one component to the complexity of lower urinary tract dysfunction, further redefining of these clusters by incorporating additional data will be necessary to understand the mechanisms of each subtype.

The objective of this analysis is to assess how the symptom profiles of patients presenting with similar combinations of LUTS at baseline change over time. We sought to identify symptom change within the LURN identified clusters in the 12 months after presentation and test associations with treatments received.

Materials and Methods

Study Design and Population

The LURN Observational Cohort Study is a one-year, multi-center study. Men and women over 18 years presenting to a LURN clinic with at least one LUTS as reported on the LUTS Tool⁵ were enrolled. Patients were not required to be treatment naive at enrollment. Patient's medical history, self-reported LUTS, physical exam findings and patient reported outcome (PRO) questionnaires reporting on bowel functioning, psychological health, urologic pain, pelvic floor, and sexual and erectile function were collected at the baseline visit. Follow-up in clinic visits at 3 and 12 months collected data on treatment history for LUTS and other PROs. Additional phone visits collected treatment history for LUTS at 6 and 9 months⁶.

Measures

Self-reported LUTS were recorded using the LUTS Tool and the American Urological Association Symptom Index (AUA-SI). The LUTS Tool is a 44-item questionnaire assessing severity and bother of 22 urinary symptoms. The questionnaires collected include the PROMIS gastrointestinal constipation, diarrhea, and bowel incontinence subsets⁷, PROMIS Depression and Anxiety Short Forms⁸, PROMIS Sleep Disturbance Short Form⁹, Perceived Stress Scale¹⁰, Genitourinary Pain Index [GUPI]¹¹, Pelvic Floor Distress Inventory [PFDI]¹², International Index of Erectile Function [IIEF]¹³ and Pelvic Organ Prolapse/Incontinence Sexual Questionnaire, IUGA-revised¹⁴.

LUTS treatments information prior to enrollment were collected at baseline and patients subsequently reported specific LUTS treatments (surgeries, medications, behavioral or physical therapies) at 3, 6, 9 and 12 months. Specific treatments were collapsed into categories relating to either the type of treatment (e.g. nutraceutical/herbal remedies) or target symptom (e.g. medication for overactive bladder [OAB]) (Supplemental Table 2). Participants were treated by their physicians according to standard care practices.

An overall LUTS Tool Severity Score, used to show the total LUTS severity, was calculated as the Euclidean length of all LUTS Tool severity questions and was weighted by the correlations between baseline LUTS Tool responses, to account for redundancy between questions¹⁵. The overall LUTS Tool Severity Score was scaled from 0 to 100, with 0 being the least severe (no

symptoms), and 100 being the most severe (most severe rating for all 22 symptoms). Additionally, a cluster-specific severity score was calculated using the same method separately for each cluster only including those symptoms highly endorsed by that cluster at baseline. “High endorsement” was defined as a median cluster response of “sometimes” or greater (Supplemental Table 3). This score was used to show the combined severity of symptoms that were highly endorsed within each cluster, which were likely to be the focus of treatment. Using both scales we are able to observe changes in the most severe symptoms for each patient (cluster specific scale) in addition to capturing changes in new symptoms that may develop over time (overall severity scale).

The cluster-specific severity score was used to classify participants based on changes in the most severe symptoms within each cluster. Participants were classified into improved, unchanged, or worsened symptom severity at 12-months by comparing baseline and 12-month cluster-specific severity scores. A decrease or increase in cluster-specific severity score of more than $\frac{1}{2}$ of a standard deviation of the baseline score was defined as improved or worsened, respectively, as this is generally accepted as a “clinically meaningful difference”¹⁶. Participants with cluster-specific severity scores at 12-months within $\frac{1}{2}$ standard deviation of their baseline score were classified as having unchanged symptom severity at baseline and 12-months.

Statistical Analysis

For participants completing at least two thirds of the LUTS Tool at each visit, multiple imputation was performed on self-reported data due to data missingness between 1% and 6% for LUTS Tool questions. Ten imputed data sets were generated using a sequential regression technique in IVEware version 2.0^{17,18}. Participants who met the threshold to have imputation performed at baseline, 3 months and 12 months were included in all analyses. Participant characteristics were reported as means and standard deviations or medians and interquartile ranges, as appropriate, for continuous variables, and percentages for categorical variables, stratified by improvement status and sex. Comparisons between improvement status within each sex were made using chi-square tests and Wilcoxon two-sample tests or Kruskal-Wallis tests for categorical and continuous variables, respectively. Comparisons between baseline and 12-month LUTS symptoms, stratified by cluster, were made using Wilcoxon two-sample tests, with an adjustment for multiple comparisons using the false discovery rate (FDR) correction¹⁹.

Repeated measures linear regression models were used to assess associations between severity scores measured at baseline, 3 months, and 12 months, and baseline patient characteristics including age, race, ethnicity, BMI, comorbidities (as assessed by the Functional Comorbidity Index), alcohol consumption, smoking status, education, and marital status, PRO's including bowel functioning, psychological symptoms and sleep disturbance, and cluster membership.

Treatments received over the course of the study period were also included as

time-dependent covariates. Repeated measures linear regression models allow for the inclusion of treatments for the same patient at multiple time-points, which accounts for patients who received multiple, sequential treatments over the course of the study, resulting in a better estimate of the association between severity score and specific treatments. The model also controls for the non-independence of multiple time points for each patient.

Separate models were fitted for men and women for the overall severity score. Stepwise selection with entry level at $p=0.15$ and stay level at $p=0.20$ was used for variable selection. Interactions between visit and highly reported treatment ($>10\%$ use in each cluster) and visit and cluster membership were tested in all models to assess differences in severity score trajectories over time by treatment use and cluster membership. Parameter estimates for differences in severity scores are reported on the per-unit (raw score) and per-standard-deviation of the severity score (i.e., effect size) scale. All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NY).

Results

Among 813 participants with sufficiently complete questionnaires at all three time points that were eligible for analysis, 184 (44.1%), 169 (40.5%), and 64 (15.3%) men and 221 (55.8%), 131 (33.1%), and 44 (11.1%) women were classified as improved, unchanged, and worsened, respectively, based on change in cluster specific severity score from baseline to 12 months. The majority were white, married or living with a partner, overweight or obese, and

middle age (Table 1, Supplemental Table 4). There were no baseline demographic or clinical differences between improved, unchanged, or worsened groups with the exception of ethnicity in women, with Hispanic women more likely to be improved at 12 months; and marital status in men, with men who were married or living with a partner were more likely to have unchanged symptom severity between baseline and 12 months.

Radar plots, presented in Figure 1, show baseline and 12-month symptom signatures for each of the clusters. At baseline, clusters M1, M4, F3 and F4 presented with many symptoms (range 10-18) with median severity reported as “sometimes” or greater, while clusters M2, M3, F1 and F2 presented with fewer symptoms (range 2-6) with median severity reported as “sometimes” or greater. At 12 months, there were statistically significant decreases in post-micturition (-1.3 unit change [Δ] in average LUTS tool response, $p=0.004$), voiding (-1.7 Δ , $p=0.02$), pain (-0.8 Δ , $p=0.02$), urgency (-1.3 Δ , $p<0.001$) and frequency symptoms (-1.4 Δ , $p=0.01$) in cluster M1, in urgency (-1.3 Δ , $p<0.001$) and incontinence symptoms (-1.2 Δ , $p=0.02$) in cluster M4, in post-micturition symptoms (-0.9 Δ , $p=0.05$) in cluster M2 and in frequency symptoms (-0.9 Δ , $p=0.01$) in cluster M3. Cluster M3 also saw a statistically significant increase in incontinence with urgency (0.6 Δ , $p=0.02$), but this cluster reported the lowest level of overall symptoms at both time-points. Cluster F1 reported significant decreases in post-micturition (-1.0 Δ , $p=0.01$), voiding (-0.6 Δ , $p=0.04$), pain (-0.8 Δ , $p=0.01$) and frequency symptoms (-1.1 Δ , $p=0.01$). Cluster F2 and F3 reported significant decreases in frequency (-1.5 Δ , $p=0.003$, -1.3 Δ , $p<0.001$ respectively)

and urgency symptoms (-1.4 Δ , $p=0.01$, -1.5 Δ , $p<0.001$ respectively), with cluster F3 also reporting decreases in post-micturition (-0.7 Δ , $p=0.03$), voiding (-0.6 Δ , $p=0.01$) and incontinence (-2.4 Δ , $p=0.001$) symptoms. Cluster F4 saw a significant decrease in frequency (-1.3 Δ , $p=0.01$), post-micturition (-1.7 Δ , $p=0.002$) and incontinence symptoms (-1.6 Δ , $p=0.04$). On average, symptoms mostly improved or stayed the same between baseline and 12 months, but there was variability within each symptom category and across clusters.

Comparisons between the overall severity score and the cluster specific severity scores allowed us to examine the difference between changes in highly reported symptoms within each cluster and change in all LUTS. Changes in cluster-specific severity scores were reflected in the overall severity score as demonstrated by the strong positive relationship between the two scores (Supplemental Figure 1). This indicates that on average, participants reported stability or slight improvement in their cluster-specific symptoms and minimal changes in other symptoms, and the symptom signatures of the clusters stayed largely the same.

Treatments received differed between men and women. Overall, more men reported medication use (55% vs 30%, $p\text{-value} <0.001$), while more women reported behavioral or physical therapy (74% vs 30%, $p\text{-value} <0.001$). More men reported no treatment (28% vs 10% respectively, $p\text{-value} <0.001$) (Figure 2). Surgery was less frequently reported in men (8%-17% per cluster) and women (8%-31% per cluster), with the most common procedures being surgery for stress incontinence in F3 (23%) and prostate surgery in M1 (14%) (Figures 2 & 3).

Women reported a higher proportion of not receiving treatment prior to enrollment than men (51% vs 40% respectively: p -value=0.002).

Results from multivariable repeated measures linear regression in men demonstrated improvement at 3 and 12 months compared to baseline for all four clusters, but similar scores between 3 and 12 months, indicating that scores tended to level off after initial improvement. The degree of decrease varied by cluster (range 0.07-0.63 SD decrease), with the largest decreases observed in cluster M1 at 3 months (0.55 SD) and 12 months (0.63 SD). Prostate surgery was associated with decreased severity scores (0.63 SD) in all clusters. Baseline patient variables associated with worse overall severity scores in men included higher anxiety levels and more sleep disturbance (Table 2).

In women, overall severity scores at 3 and 12 months were lower than baseline for all clusters with decrease in scores ranging from 0.20 SDs to 0.67 SDs. Clusters F3 and F4 had on average adjusted overall severity scores >0.5 SD lower than baseline at 12 months. Severity score trajectories were similar to those seen in men, with no differences between 3 and 12 month scores detected. Having a surgery for stress incontinence was associated with better overall severity scores (0.88 SD decrease) in all clusters. In addition, more comorbidities, more sleep disturbance, and worse pelvic floor distress inventory scores (all assessed at baseline) were associated with higher levels of overall severity at all-time points.

Discussion

Fifty percent of participants in this study showed LUTS improvement during the one-year follow-up. This low rate of improvement substantiates the need to find a better way to diagnose and treat LUTS and to continue to add additional data to redefine the clusters. Most symptoms showed improvement across all clusters; however, the degree of improvement varied by cluster. Both the overall LUTS severity score and cluster specific severity score changed consistently over time and we did not see substantial increase in severity or bother of a specific LUTS that was not highly endorsed at baseline, which indicates that our cluster signatures remained stable over time. This suggests that the clusters may be mechanistically distinct rather than different stages of LUTS. The incorporation of more diagnostic data to redefine the clusters may help prove this finding.

Despite recruiting from academic urologic and urogynecologic centers, 46% of the cohorts reported no prior LUTS treatment at baseline; therefore, men and women utilized a wide variety of behavioral treatments in each cluster during the 12-month follow-up. Furthermore, in 3 of the 4 male clusters, over 30% did not report using LUTS therapy during the follow up period. Interestingly, within the M2 and M4 clusters, the proportions reporting unchanged or worsened symptoms at 12 months were similar for those who did and did not report treatment of any kind. This may be due to low baseline symptoms in those clusters, making a salutary change in symptoms more difficult to detect. Although those with high severity scores at baseline improved the most, there were still a significant number of unchanged or worsened in each cluster, which supports the

need to better understand the pathophysiology of these individuals. By continuing to redefine the current LURN clusters by incorporating more data that include diagnostic information and possibly cellular/molecular factors, more meaningful differences among patients may be identified.

Associations between treatments (or combinations of treatments) and symptom changes did not vary by cluster. We saw an improvement in symptoms for women who had stress incontinence surgery during the study period, regardless of cluster. However, treatment decisions were made per standard of care without knowledge of cluster membership, therefore further assessment of treatment targeted to cluster membership is needed to determine potential differences in disease etiology and treatment response.

We found that certain factors were predictive of worse symptoms (such as multiple comorbidities, high anxiety, high scores on disease specific quality of life PROs and sleep disturbance); nonetheless, we do not understand the underlying pathophysiologies for why these patients might be more difficult to treat and reemphasizes a need for pathophysiologic studies to better understand LUTS. The largest decreases in overall severity scores were in the clusters with more symptoms. For men, prostate surgery was associated with symptom improvement in all clusters; and only for the female cluster (cluster group F3) with a high proportion reporting stress urinary incontinence, did we find midurethral slings correlated with reduced LUTS. Therefore, since various pathophysiologies can produce similar LUTS, we need to understand the

underlying causes for an individual to better guide treatment decisions and to find better therapies.

A primary limitation of our study is that our symptom-based clusters need to be refined using additional data (e.g. bladder diary and urodynamics data). Work is currently underway to integrate other data domains. In addition, the applicability of these clusters to other populations (e.g. non care-seeking) with higher levels of racial and LUTS severity diversity is yet to be determined. Clinical validation of these clusters should be the subject of future studies. Finally, from this observational study design, causal relationships between treatment and symptom change cannot be determined nor can a comparison be made between the different therapies used for similar symptoms. There was variability in the frequency of individual treatments used (e.g. midurethral slings as compared with bulking agents) and these results were combined. This does not allow for the assessment of associations between the different individual treatments and symptom severity, which may differ from the group estimates presented. Further prospective studies will be required to evaluate how specific individuals change within each cluster.

Conclusions

Symptom improvement varied by cluster indicating response to therapy differs amongst subtypes of patients with LUTS. While specific treatments could not be associated with cluster-specific symptom improvement in this observational study, the differential improvement by cluster can be used to generate further

hypotheses regarding mechanistic differences between clusters that could be used to apply more targeted treatments in the future.

References:

1. Gormley EA, Lightner DJ, Burgio KL, Chai TC, Clemens JQ, Culkin DJ, Das AK, Foster HE Jr, Scarpero HM, Tessier CD, Vasavada SP; American Urological Association; Society of Urodynamics, Female Pelvic Medicine & Urogenital Reconstruction. Diagnosis and treatment of overactive bladder (non-neurogenic) in adults: AUA/SUFU guideline. *J Urol.* 2012;188(6 Suppl):2455-2463.
2. Yang CC, Weinfurt KP, Merion RM, Kirkali Z, LURN Study Group. Symptoms of Lower Urinary Tract Dysfunction Research Network. *J Urol.* 2016;196:146-152.
3. Andreev VP, Liu G, Yang CC, Smith AR, Helmuth ME, Wiseman JB, Merion RM, Weinfurt KP, Cameron AP, Lai HH, Cella D, Gillespie BW, Helfand BT, Griffith JW, DeLancey JOL, Fraser MO, Clemens JQ, Kirkali Z; LURN Study Group. Symptom Based Clustering of Women in the LURN Observational Cohort Study. *J Urol.* 2018;200:1323-1331.
4. Liu G, Andreev VP, Helmuth ME, Yang CC, Lai HH, Smith AR, Wiseman JB, Merion RM, Erickson BA, Cella D, Griffith JW, Gore JL, DeLancey JOL, Kirkali Z, and the LURN Study Group. Symptom-based Clustering of Men in the Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN) Observational Cohort Study. *J Urol.* 2019 [In Press].

5. Coyne KS, Barsdorf AI, Thompson C, Ireland A, Milsom I, Chapple C, Kopp ZS, Bavendam T. Moving towards a comprehensive assessment of lower urinary tract symptoms (LUTS). *Neurourol Urodyn*. 2012;31:448-454.
6. Cameron AP, Lewicky-Gaupp C, Smith AR, Helfand BT, Gore JL, Clemens JQ, Yang CC, Siddiqui NY, Lai HH, Griffith JW, Andreev VP, Liu G, Weinfurt K, Amundsen CL, Bradley CS, Kusek JW, Kirkali Z; LURN Study Group. Baseline Lower Urinary Tract Symptoms in Patients Enrolled in LURN: A Prospective, Observational Cohort Study. *J Urol*. 2018;199:1023-1031.
7. Spiegel BM, Hays RD, Bolus R, Melmed GY, Chang L, Whitman C, Khanna PP, Paz SH, Hays T, Reise S, Khanna D. Development of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) gastrointestinal symptom scales. *Am J Gastroenterol*. 2014;109:1804-1814.
8. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D; PROMIS Cooperative Group. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. *Assessment*. 2011;18:263-283.
9. Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, Johnston KL, Pilkonis PA.. Development of short forms from the PROMIS sleep

- disturbance and Sleep-Related Impairment item banks. *Behav Sleep Med.* 2011;10:6-24.
10. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav.* 1983;24(4):385-396.
 11. Clemens JQ, Calhoun EA, Litwin MS, McNaughton-Collins M, Kusek JW, Crowley EM, Landis JR; Urologic Pelvic Pain Collaborative Research Network. Validation of a modified National Institutes of Health chronic prostatitis symptom index to assess genitourinary pain in both men and women. *Urology.* 2009;74:983-987, quiz 987 e981-983.
 12. Barber MD, Chen Z, Lukacz E, Markland A, Wai C, Brubaker L, Nygaard I, Weidner A, Janz NK, Spino C. Further validation of the short form versions of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ). *Neurourol Urodyn.* 2011;30:541-546.
 13. Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology.* 1997;49:822-830.
 14. Rogers RG, Rockwood TH, Constantine ML, Thakar R, Kammerer-Doak DN, Pauls RN, Parekh M, Ridgeway B, Jha S, Pitkin J, Reid F, Sutherland SE, Lukacz ES, Domoney C, Sand P, Davila GW, Espuna Pons ME. A new measure of sexual function in women with pelvic floor disorders

- (PFD): the Pelvic Organ Prolapse/Incontinence Sexual Questionnaire, IUGA-Revised (PISQ-IR). *Int Urogynecol J*. 2013;24:1091-1103.
15. Helmuth ME, Smith AR, Andreev VP, Liu G; Lai HH, Cameron AP, Siddiqui NY, LURN Study Group. Use of Euclidean length to measure urinary incontinence severity based on the lower urinary tract symptoms tool. *Am J Obstet Gynecol*. 2018;218:357-359.
 16. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582-592.
 17. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology*. 2001;27(1):85-96.
 18. Raghunathan TE, Solenberger PW, Van Hoewyk J. IVEware: Imputation and variance estimation software. Ann Arbor, MI: Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan. 2002.
 19. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995;57(1):289-300.

Figures

Figure 1: Radar Plots of Mean Cluster Symptoms Reported at Baseline and 12-months

The exterior colored point on each spoke of the radar plots represents symptoms reported at baseline, while the interior outlined white point on the spoke represents symptoms reported at 12 months. The colored band between these two points represents symptoms reported at 12 months. The colored band between these two points represents the change in symptoms between baseline and 12 months. Larger regions of color indicate larger decreases in the mean symptom severity and bother within that cluster. Any symptom where the outlined white section expands beyond the colored section represents symptoms that were reported as on average worse at 12 months compared to baseline. Spokes on the radar plots (symptom items) that were statistically significantly different between baseline and the 12-month visit are denoted by asterisks.

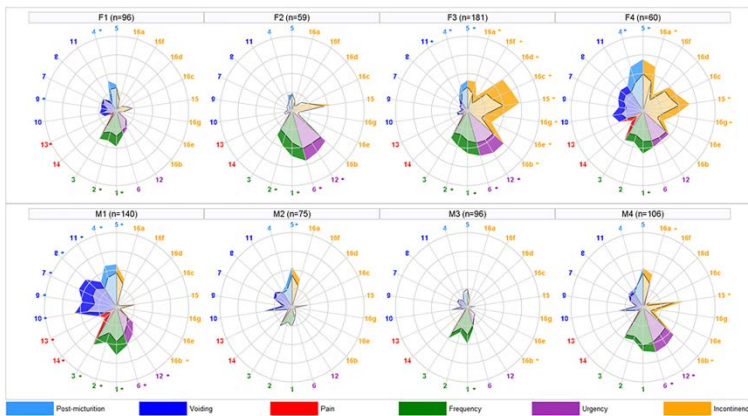


Figure 2: Distribution of Treatments Across Clusters and Proportion of Patients with Changes in Symptoms within those Treatments: Men Footnote: Overall 28% reported using more than one treatment (17% reported 2 treatments, 7% reported 3 treatments)

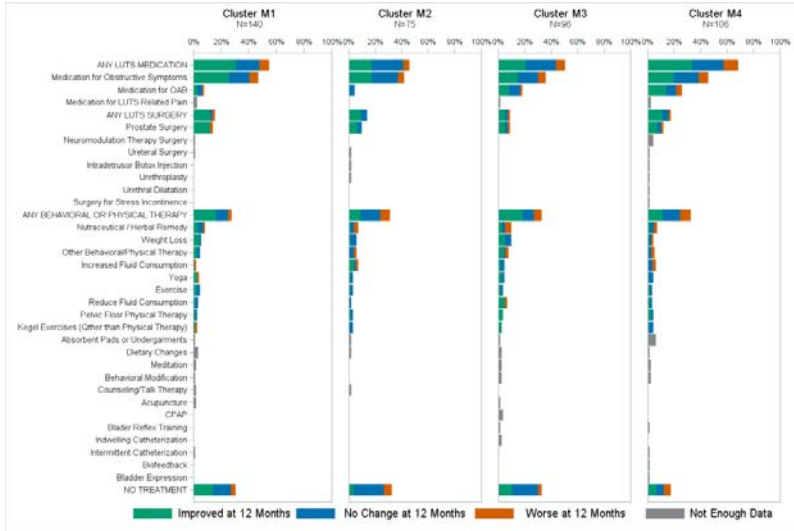
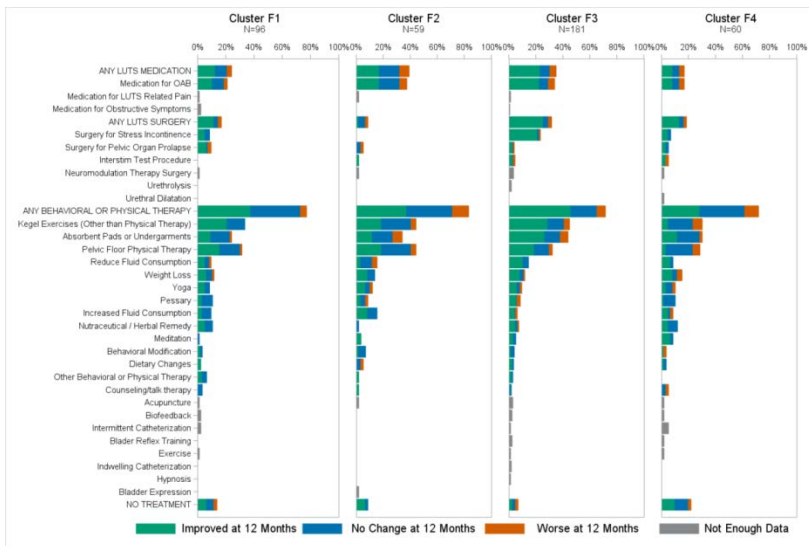


Figure 3: Distribution of Treatments Across Clusters and Proportion of Patients with Changes in Symptoms within those Treatments: Women Footnote: Overall 62% reported using more than one treatment (23% used 2 treatments, 15% used 3 treatments)



Tables

Table 1: Patient Characteristics and Patient-reported Outcomes in Men and Women by 12-month Improvement status

	Women Worse at 12 months N=44	Women Same at 12 months N=131	Women Improved at 12 months N=221	P- value	Men Worse at 12 months N=64	Men Same at 12 months N=169	Men Improved at 12 months N=184	P- value
Age	58.3 (49.8- 65.8)	60.0 (52.2- 69.4)	56.5 (47.5- 67.9)	0.153	63.2 (56.0- 70.9)	65.1 (56.4- 70.2)	63.5 (56.6- 70.7)	0.958
Race				0.858				0.575
African-American	3 (7%)	17 (13%)	28 (13%)		6 (9%)	16 (9%)	18 (10%)	
White	38 (86%)	106 (81%)	179 (81%)		52 (81%)	131 (78%)	152 (83%)	
Other Race	3 (7%)	8 (6%)	14 (6%)		6 (9%)	22 (13%)	14 (8%)	
BMI	31.6 (25.0- 40.4)	28.9 (24.4- 35.8)	29.9 (25.4- 33.8)	0.364	27.4 (24.5- 31.5)	28.4 (25.8- 32.1)	28.8 (25.6- 32.7)	0.179
Functional Comorbidity Index	2.0 (1.0- 3.0)	2.0 (1.0- 4.0)	2.0 (1.0- 4.0)	0.685	2.0 (0.0- 3.0)	2.0 (1.0- 4.0)	2.0 (1.0- 3.0)	0.104
PROMIS Physical Functioning*	47.3 (41.3- 53.4)	46.3 (37.7- 54.2)	48.5 (39.8- 60.3)	0.462	51.2 (45.7- 60.3)	50.8 (43.7- 60.3)	49.8 (42.9- 60.3)	0.347
PROMIS Depression*	50.1 (44.5- 54.6)	49.2 (44.5- 55.3)	47.8 (38.2- 54.5)	0.422	47.8 (44.5- 51.1)	46.5 (38.2- 52.2)	47.8 (38.2- 52.2)	0.684

PROMIS Anxiety*	49.0 (43.1-56.2)	50.3 (44.7-54.5)	50.0 (43.2-56.2)	0.922	48.4 (43.1-51.5)	48.1 (37.1-52.5)	49.4 (43.1-53.1)	0.326
PROMIS Sleep Disturbance*	52.8 (48.8-55.8)	52.0 (46.7-57.9)	53.3 (46.7-58.5)	0.769	50.1 (45.0-56.7)	51.6 (44.8-56.7)	52.1 (48.5-57.8)	0.244
Pelvic Floor Distress Inventory [^]	84.4 (49.0-113)	62.5 (37.5-100)	69.8 (41.7-103)	0.114	-	-	-	
Perceived Stress Scale [#]	12.0 (9.0-16.0)	12.0 (7.0-18.0)	12.0 (6.0-17.0)	0.590	10.0 (6.0-14.0)	10.0 (5.0-16.0)	10.0 (6.0-16.0)	0.963
International Index of Erectile Function [%]	-	-	-		13.0 (4.0-29.0)	17.0 (3.0-27.0)	15.0 (3.0-27.0)	0.942
LUTS Treatment Prior to Baseline	22 (50%)	61 (47%)	107 (49%)	0.916	36 (56%)	96 (57%)	116 (63%)	0.380

*Higher scores indicate more physical functioning, depression, anxiety or sleep disturbance (general population: mean=50 standard deviation=10)

[^]Higher scores indicate more pelvic floor distress (PFDI ranges from 0-300)

[#]Higher scores indicate more perceived stress (PSS ranges from 0-40)

[%]Higher scores indicate more erectile functioning (IIEF ranges from 5-25)

Table 2: Multivariable Repeated Measures Linear Regression Results

	Mean Estimate	Mean Confidence Limits	Standardized Estimate*	P-value
Overall Severity Score: Men				
Intercept	17.636	9.727 25.545	--	<.001
PROMIS Anxiety (per 5 unit increase) (centered at a mean score of 50)	1.469	0.754 2.183	0.108	<.001

PROMIS Sleep Disturbance (per 5 unit increase) (centered at a mean score of 50)	1.375	0.838	1.912		<.001
				0.102	
3 month vs. baseline: Cluster M1	-7.461	-9.533	-5.390	-0.551	<.001
3 month vs. baseline: Cluster M2	-2.390	-4.544	-0.237	-0.176	0.030
3 month vs. baseline: Cluster M3	-0.889	-2.631	0.853	-0.066	0.317
3 month vs. baseline: Cluster M4	-3.470	-5.599	-1.341	-0.256	0.001
12 month vs. baseline: Cluster M1	-8.547	-	-6.415	-0.631	<.001
			10.679		
12 month vs. baseline: Cluster M2	-1.335	-3.458	0.789	-0.099	0.218
12 month vs. baseline: Cluster M3	-0.485	-2.523	1.554	-0.036	0.641
12 month vs. baseline: Cluster M4	-4.287	-6.633	-1.942	-0.317	<.001
Prostate Surgery	-8.536	-	-4.341	-0.630	<.001
			12.730		
<hr/>					
Overall Severity Score: Women					
<hr/>					
Intercept	16.980	10.968	22.992	--	<.001
Functional Comorbidity Index (per 1 unit increase) (centered at a mean FCI of 2.4)	0.817	0.324	1.309		0.001
				0.059	
PROMIS Sleep Disturbance (Per 5 unit increase) (centered at a mean score of 50)	0.947	0.369	1.524		0.001
				0.068	
PFDI (Per 5 unit increase) (centered at a mean PFDI score of 77.9)	0.583	0.477	0.689		<.001
				0.042	
3 month vs. baseline: Cluster F1	-3.203	-5.659	-0.747	-0.229	0.011
3 month vs. baseline: Cluster F2	-3.707	-6.277	-1.136	-0.266	0.005
3 month vs. baseline: Cluster F3	-6.359	-8.249	-4.469	-0.456	<.001
3 month vs. baseline: Cluster F4	-8.535	-	-3.900	-0.611	<.001
			13.169		
12 month vs. baseline: Cluster F1	-3.687	-6.185	-1.189	-0.264	0.004
12 month vs. baseline: Cluster F2	-2.739	-5.571	0.093	-0.196	0.058
12 month vs. baseline: Cluster F3	-7.582	-9.665	-5.499	-0.543	<.001

12 month vs. baseline: Cluster F4	-9.327	- -4.058	<.001
	14.596		-0.668
SUI Surgery	-12.289	- -8.483	<.001
	16.096		-0.880

*Estimates are standardized by the standard deviation of the outcome and represent the mean estimate's proportion of one standard deviation.

Bolded covariates highlight clinically meaningful differences in estimates, based on a standardized estimate of greater than 0.5 or less than -0.5.