Enhanced clinical decisions for management of benign prostatic hyperplasia using patient-reported outcomes: protocol for a prospective observational study

Alexander P. Glaser[1,2], Abigail R. Smith[3], Dacey Maglaque[1], Brian T. Helfand[1,2], Rowida Mohamed[4], Hosanna An[4], Melissa Marquez[4], Pooja Talaty[1], Padraig Carolan[3], Aaron M. Geller[5,6], Francesca R. Farina[4], Sally E. Jensen[7,8] and James W. Griffith[4*]

Abstract

Background Lower urinary tract symptoms (LUTS) due to benign prostatic hyperplasia (BPH) significantly impact quality of life among older men. Despite the prevalent use of the American Urological Association Symptom Index (AUA-SI) for BPH, this measure overlooks key symptoms such as pain and incontinence, underscoring the need for more comprehensive patient-reported outcome (PRO) tools. This study aims to integrate enhanced PROs into routine clinical practice to better capture the spectrum of LUTS, thereby improving clinical outcomes and patient care.

Methods This prospective observational study will recruit men with LUTS secondary to BPH aged ≥ 50 years from urology clinics. Participants will be stratified into medical and surgical management groups, with PRO assessments scheduled at regular intervals to monitor LUTS and other health outcomes. The study will employ the LURN Symptom Index (SI)-29 alongside the traditional AUA-SI and other non-urologic PROs to evaluate a broad range of symptoms. Data on comorbidities, symptom severity, and treatment efficacy will be collected through a combination of electronic health records and PROs. Analyses will focus on the predictive power of these tools in relation to symptom trajectories and treatment responses. Aims are to: (1) integrate routine clinical tests with PRO assessment to enhance screening, diagnosis, and management of patients with BPH; (2) examine psychometric properties of the LURN SIs, including test-retest reliability and establishment of clinically meaningful differences; and (3) create care-coordination recommendations to facilitate management of persistent symptoms and common comorbidities measured by PROs.

Discussion By employing comprehensive PRO measures, this study expects to refine symptom assessment and enhance treatment monitoring, potentially leading to improved personalized care strategies. The integration of these tools into clinical settings could revolutionize the management of LUTS/BPH by providing more nuanced insights into patient experiences and outcomes. The findings could have significant implications for clinical practices, potentially leading to updates in clinical guidelines and better health management strategies for men with LUTS/BPH.

*Correspondence:
James W. Griffith
James.Griffith@bsd.uchicago.edu
Full list of author information is available at the end of the article

Glaser *et al. BMC Urology*   (2024) 24:110

Page 2 of 10

Introduction

Lower urinary tract symptoms (LUTS) attributable to benign prostatic hyperplasia (BPH) are common, affecting approximately 33% of Medicare beneficiaries [1]. As the most common condition contributing to non-neurogenic LUTS in men, BPH is also the focus of frequently updated medical and surgical management guidelines that directly influence provider decision-making and patient care [2, 3]. The burden of LUTS/BPH is expected to increase, both in the US and worldwide, due to shifting demographics and a static urologic workforce [4]. To meet this demand, there is an urgent need to develop and refine efficient strategies to assess, triage, and enhance the care of patients through standardized outcome measures and optimized care coordination, including remote assessment.

Measuring outcomes provides a way to track progress, adjust treatments, monitor quality of care, and help patients achieve optimal outcomes. For decades, the American Urological Association Symptom Index (AUA-SI) has been the patient-reported outcome (PRO) measure used in men with BPH, as recommended in the "Surgical Management" guidelines updated as recently as 2023 [2, 5]. The AUA-SI, however, has several limitations that leave room for improvement. First, the AUA-SI omits urinary incontinence and pain, which in the Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN) cohort, were present in 51% and 17% of treatment-seeking men, respectively, despite exclusion of men with infection or urinary pain as a primary complaint [6, 7]. Urgency urinary incontinence, stress urinary incontinence, and post-void dribbling were all associated with higher LURN-SI scores, AUA-SI scores, and bother, and thus mandate assessment [7, 8]. Furthermore, a proportion of men present with pain, which can be an important differential diagnostic marker (e.g., between BPH and chronic prostatitis/chronic pelvic pain syndrome). Pain may also currently be underassessed in the postoperative setting [9]. Second, the AUA-SI yields a single score for overall LUTS, which is laudable for its simplicity and may facilitate communication with patients. However, a single score may be insufficient when symptoms are differentially impacted by treatment or have different relationships with biomarkers, diagnostic subgroups, and other clinically important variables. Third, the methodology used to develop the AUA-SI was specific to BPH, which is advantageous once the patient has been properly diagnosed, but patients presenting to the urology clinic with LUTS may have other diagnoses such as kidney stones,

infection, chronic prostatitis, pelvic floor dysfunction, and stricture disease. The AUA-SI, although designed for BPH, has been used in a variety of conditions it was not designed for, including prostate cancer and overactive bladder syndrome [10, 11].

With an awareness of the need to create an updated outcome tool in urology, LURN created new questionnaires in urology including the 10- and 29-item LURN Symptom Indices (SIs) [12, 13]. The LURN Symptom Index SI-29 was designed as a comprehensive tool, whereas the one-page LURN SI-10 can be extracted from the broader LURN SI-29 and was designed for clinical use analogous to the AUA-SI. Both the LURN SI-29 and SI-10 were developed as PROs for patients with LUTS, using both qualitative and quantitative methods, and have shown initial validity evidence in people with LUTS [8, 12–15]. What is needed now is research on specific conditions, such as rigorous validation of the psychometric properties of these questionnaires in LUTS/BPH. In order for these symptom questionnaires to improve outcomes, they must be connected to actionable phenomena in the clinic (e.g., providing the patients with tailored information, treatment, and/or referrals based on their scores). Toward this end, the present study will administer PROs about both urologic and non-urologic symptoms into the routine care of patients with LUTS/BPH.

Study objectives

Aim 1: Integrate routine clinical tests with PRO assessment to enhance screening, diagnosis, and management of patients with BPH. PROs provide a cost-effective assessment strategy that can improve patient care by allowing for frequent surveillance along with routine objective clinical variables (e.g., post-void residual, prostate volume, intravesical prostatic protrusion), will provide unique predictive power for LUTS severity over time. Understanding a broad set of health factors, such as sleep disturbance and depression, which may exacerbate LUTS, will facilitate counseling and interventions to improve in-clinic and remote evaluation and treatment [16].

Aim 2: To examine psychometric properties of the LURN SIs, including clinically meaningful differences in men with BPH receiving known effective treatment. Longitudinal data will enable calculation of effect sizes of the LURN SI-10 and SI-29 from baseline to follow-up in the context of usual care, including guideline-driven treatment of BPH. LURN SIs will be anchored to patients' ratings of global change (PGI-I), and longitudinal

assessments will provide quantitative estimates of minimally important differences (MIDs) [17, 18]. We hypothesize that MIDs will be approximately 0.5 SDs. Repeated assessment on a subset of patients will allow estimation of test-retest reliability.

Aim 3: Create care-coordination recommendations to facilitate the management of persistent symptoms and common comorbidities measured by PROs. PROs provide a patient-centered way to assess problems that may persist after treatment has been initiated. For example, surgical treatments can lead to transient but distressing symptoms, including pain, urgency, and/or incontinence, which require monitoring after surgery [9]. Both LUTS and comorbid symptoms cause bother for the patient and reduce the quality of life, so it is important to monitor these symptoms and coordinate care so that these symptoms are effectively managed [6, 15]. We will use qualitative interviews with patients and providers to inform the creation of a care-coordination recommendations that can be implemented for patients with BPH. Patients and providers will be interviewed about unmet needs in between clinic visits to develop better ways to follow these symptoms using a combination of in-clinic visits, telehealth, asynchronous contacts, and remote symptom monitoring. We hypothesize that patient and stakeholder input will suggest care-coordination recommendations that have the potential to increase quality of care, and in turn support better outcomes for patients with BPH.

Methods
Ethical approval
This study was approved by the Institutional Review Board of Northwestern University (IRB STU00217126), which served as the single IRB of record. This study is registered in ClinicalTrials.gov (NCT05898932).

Study design, setting, and population
Aims 1 and 2
Men with BPH aged 50 years and older will be recruited from urology clinics at NorthShore University HealthSystem ("NorthShore"). Recruited patients with BPH will undergo standard of care treatment as determined by their urologist. To obtain a sample that is representative of BPH patients, as well as to generalize to patients receiving different kinds of treatment, the sample will be stratified by whether they are to receive medical/non-surgical management ($n=200$) or procedural/surgical management ($n=100$). Procedural/surgical management will include both minimally invasive surgical therapies (e.g., prostatic urethral lift, water vapor thermal therapy) and resective therapies (e.g., transurethral resection of prostate, holmium laser enucleation of prostate, robotic waterjet treatment) as determined by the patient and the

patient's treating urologist. Participants will be allowed to crossover between treatment arms if their clinical treatment type changes as determined by shared decision making with their urologist (e.g., if a participant in the medical management arm decides to undergo surgery during routine clinical care).

Both new and returning patients undergoing medical or surgical treatment for BPH will be eligible. Eligible patients (see Table 1 for inclusion and exclusion criteria) will be identified in two ways: (1) a research coordinator will screen the electronic health record (EHR) for upcoming clinic appointments to identify patients who may be eligible. A coordinator will contact these patients by phone to describe the study and obtain informed consent remotely, which will be documented using the Research Electronic Data Capture System (REDCap). Alternatively, (2) eligible patients who present for a clinic visit (either in-person or telehealth) will be referred to the coordinator to discuss the study, obtain informed consent (again either in-person or remotely), and document the consent and participant details in REDCap. During these interactions, research staff will explain the contents of the consent form, focusing on the purpose of the study, study procedures, study duration, participants' rights and responsibilities, potential risks and benefits, and the use of participant's protected health information for research purposes. Participants will be given the opportunity to ask questions about the study and will be provided with as much time as needed to review information about the study to reach a voluntary decision about participation.

Aim 3
Participants will be approached for consent to participate in qualitative interviews following the same procedures used in the longitudinal observational cohort study. We will recruit a representative subset of participants from the parent study into the interviews in terms of age, race, and ethnicity. Interview participants will be recruited until thematic saturation is reached, with a maximum of 20 interviews with health professionals and 40 with patients. Study logistics prevent the interview subsample from being a random subsample (e.g., participant characteristics may be related to which participants agree to participate in the interviews). Thus, the study team will track the demographics of those approached for interviews and will review diversity metrics at a weekly meeting accompanied by a descriptive report prepared by the study statistician. Targets for these qualitative interviews will encompass a diverse group of health care professionals responsible for the care of individuals with LUTS/BPH, including urologists, physician assistants, and nurses, as well as patients being treated for LUTS/BPH with medications, minimally invasive procedures, and traditional BPH surgeries. This method seeks to include

Table 1 Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
1) Male sex	1) Female sex or intersex
2) Age 50 years or older	2) Younger than 50 years of age
3) Diagnosed by physician with BPH	3) Being a prisoner or detainee
4) Able and willing to complete questionnaires	4) Gross hematuria
5) Able and willing to provide informed consent	5) Interstitial cystitis
6) Ability to read, write, and speak in English	6) Pelvic or endoscopic genitourinary surgery within the preceding 6 months (not including diagnostic cystoscopy)
7) No plans to move from study area in next 6 months	7) History of cystitis caused by tuberculosis, radiation therapy, or Cytoxan/cyclophosphamide therapy
	8) Ongoing symptomatic urethral stricture
	9) Current chemotherapy or other cancer therapy
	10) History of lower urinary tract or pelvic malignancy
	11) Severe neurological of psychiatric disorder that would prevent study participation (e.g., bipolar disorder, psychotic disorder, Alzheimer's Disease)
	12) Current moderate or severe substance use disorder

Deferral Criteria

1) Microscopic hematuria without appropriate workup per AUA/SUFU Guidelines

2) Positive urine culture

Table 2.1 Study questionnaires, medical/non-surgical intervention(Schedule of assessments: baseline, inter-visit follow-up, and clinic assessments)

	# Items	Week −2	Baseline (clinic)	Week 4	Week 8	Week 12 (clinic)	Week 16	Week 20	Week 24 (clinic)
LURN SI-29 [12]	28								
PROMIS–29 Profile v2.1 [20]	29								
AUA-SI [5]	8								
PGI-I [21]	1								
National Cancer Institute Quick Food Scan [22]	17								
3-Day Voiding Diary [23, 24]	9								
PROMIS® Squale v1.0 – Gastrointestinal Constipation 9a [25]	9								
PROMIS® Scale v1.0 – Gastrointestinal Diarrhea 6a [25]	6								
SHIM [26]	5								
MSHQ-Ed [27]	4								
SBQ [28]	8								
PROMIS® Item Bank v2.0 – Cognitive Function- Short Form 4a [29]	4								
Patient Feedback Survey (optional)* [30]	6								

Note: Study time points (e.g., Week 4, Week 8) are approximate – Participants will be given a time window of ± 2 weeks. The minimum time between assessments will be approximately 4 weeks. Time windows may be adjusted based on COVID-19 protocols (e.g., if a clinic visit is converted to telehealth)

*Adapted from Atkinson et al.

Abbreviations: LURN SI-29: LURN SYMPTOM INDEX-29; PROMIS: Patient-Reported Outcomes Measurement Information System; AUA-SI: The American Urological Association Symptom Index; PGI-I: Patient Global Impression of Improvement; SHIM: Sexual Health Inventory for Men; MSHQ-Ed: Male Sexual Health Questionnaire-Ejaculatory Dysfunction; SBQ: STOP-BANG questionnaire

a wide array of backgrounds within the participant pool, thus enriching the study's insights by capturing the full spectrum of experiences related to BPH [19].

Data collection

All participants will complete a baseline visit after consent where demographic and clinical information will be collected. Routinely collected objective clinical data from the EHR will be analyzed in combination with PROs. In our protocol, we included health questionnaires specific to the management of LUTS/BPH, listed in Table 2. Participants in the medical arm will complete questionnaires at baseline and every 4 weeks for 24 weeks (Table 2.1); participants in the surgical arm will complete questionnaires at baseline, pre-op, and Weeks 1, 2, 4, 6, 8, 12, and 24 post-surgery (Table 2.2). Patients undergoing surgery will have more frequent queries in the immediate postoperative period (Table 2.2), as many men report

Table 2.2 Study questionnaires, surgical/procedural intervention(Schedule of assessments: baseline, inter-visit follow-up, and clinic assessments)

	# Items	Baseline (clinic)	Preop**	Week 1	Week 2	Week 4 (clinic)	Week 6	Week 8	Week 12 (clinic)	Week 24 (clinic)
LURN SI-29 [12]	28									
PROMIS–29 Profile v2.1 [20]	29									
AUA-SI [5]	8									
PGI-I [21]	1									
National Cancer Institute Quick Food Scan [22]	17									
3-Day Voiding Diary [23, 24]	9									
PROMIS® Squale v1.0 – Gastrointestinal Constipation 9a [25]	9									
PROMIS® Scale v1.0 – Gastrointestinal Diarrhea 6a [25]	6									
SHIM [26]	5									
MSHQ-Ed [27]	4									
SBQ [28]	8									
PROMIS® Item Bank v2.0 – Cognitive Function-Short Form 4a [29]	4									
Patient Feedback Survey (optional)* [30]	6									

Note: Study time points (e.g., Week 4, Week 8) are approximate – Participants will be given a time window of ± 2 weeks. The minimum time between assessments will be approximately 4 weeks. Time windows may be adjusted based on COVID-19 protocols (e.g., if a clinic visit is converted to telehealth)

*Adapted from Atkinson et al.

** If surgery is not immediately scheduled, complete questionnaires every 4 weeks until surgery is scheduled

Abbreviations: LURN SI-29: LURN SYMPTOM INDEX-29; PROMIS: Patient-Reported Outcomes Measurement Information System; AUA-SI: The American Urological Association Symptom Index; PGI-I: Patient Global Impression of Improvement; SHIM: Sexual Health Inventory for Men; MSHQ-Ed: Male Sexual Health Questionnaire-Ejaculatory Dysfunction; SBQ: STOP-BANG questionnaire

bothersome symptoms of frequency, urgency, pain, and/or incontinence in this timeframe; thus, this period, in particular is a target for better characterization and management of symptoms. The use of PROs is already part of the management guidelines for BPH; the overarching hypothesis of this project is that a brief (i.e., clinically feasible) set of questions can aid in the management of BPH and can improve upon the existing method of the AUA-SI. We recognize that the proposed series of measures is longer than one might use in clinical practice, but a comprehensive assembly of questionnaires is necessary for research purposes to clarify which measures would need to be dropped versus retained for optimal use in the real-world clinic. Data analyses will also include the LURN SI-10 score, which is a summary score that can be extracted from the LURN SI-29.

For assessment of test-retest reliability, patients will receive the LURN SI-29 and AUA-SI questionnaires in two instances, approximately 2 weeks apart. We will target 2 weeks between each questionnaire administration but will allow a 1-3-week time interval between administrations. When feasible, patients who provide consent prior to their baseline visit will complete the questionnaires 2 weeks (range 1–3 weeks) prior to baseline and then again at the baseline visit. Patients who provide consent at the baseline visit and have a scheduled diagnostic procedure at least 1 week but no longer than 3 weeks later (e.g., diagnostic cystoscopy and/or prostate transrectal ultrasound), with no treatments or interventions planned in the interim, will complete questionnaires at the baseline clinic visit and again 2 weeks later (range 1–3 weeks). The questionnaires will be delivered via Epic MyChart, if available, or alternately using RED-Cap, on paper, or if needed over the telephone (e.g., in patients with insufficient vision). A research coordinator will contact participants to remind them of these questionnaires in order to minimize missing data. Mode of administration will be recorded.

To assess MIDs, we will administer the Patient Global Impression of Improvement (PGI-I) as a measure of global change at each post-baseline time point. The PGI-I is a single-item questionnaire that asks respondents to rate their condition compared with prior to treatment on a 7-point scale from very much worse (1) to very much better [17]. LURN SI-29 and SI-10 scores at 12 weeks post-baseline will be used to assess changes in scores since baseline. At the end of the study, participants will be invited to complete a patient-response burden questionnaire to provide a better understanding of how the participants perceive the RROs,

Qualitative data Collection

Participants in qualitative interviews will include patients and health care providers. To maximize convenience, interviews will be conducted via telephone or secure virtual interview (e.g., Zoom) by trained interviewers.

Health professionals will be interviewed once; patients will be interviewed at three time points: baseline, post-treatment, and follow-up. The first two interviews with patients as well as the ones with health care professionals will inform Version 1 of the care coordination recommendations. Subsequently, the third patient interview will focus on the refinement of the care coordination recommendations into Version 2 based on participants' feedback. Participants will provide informed consent prior to each interview and will be asked to verbally confirm their consent at the beginning of the interviews. Trained researchers from Northwestern University and the University of Chicago will use semi-structured qualitative interview guides focusing on how PROs can contribute to decision-making and help patients address unmet needs and health care coordination gaps. Interviews will be audio-recorded and transcribed. Each interview is expected to be less than 20 and 60 min for health care professionals and patients, respectively. The interviewing process will continue until no new themes emerge, with no more than 20 interviews with health professionals and 40 interviews with patients. Health care professionals will receive $100 for each interview; patients will receive $40 for each interview, resulting in a total compensation of $120 for patients who complete all three interviews.

Analysis

Patient demographics and characteristics will be described using frequencies and percentages for categorical variables. Continuous data will be summarized using means, standard deviations (SDs), and as appropriate medians and interquartile ranges for variables with non-Gaussian distributions. The distribution of treatments received, and duration of medication treatment will also be assessed using similar statistics. Longitudinal trajectories of questionnaire responses will be explored using spaghetti plots and other graphical strategies.

Aim 1: identify health variables from the PROMIS Profile that predict trajectories of LUTS within the individual patient

The primary hypothesis of Aim 1 is that health variables from routine care combined with PROs on health domains relevant to BPH (e.g., sleep disturbance, depression) will predict changes in LUTS over time in this 6-month study. We will assess these associations with LUTS over time using linear mixed models with random participant intercepts and slopes. Separate models will be run for each urinary outcome of interest (SI-29 subscales: incontinence, pain, voiding, urgency, nocturia, as well as the SI-29 total score and SI-10 score). Covariates of interest will include objective clinical measures as well as PROMIS measures (e.g., sleep disturbance, depression, pain) and the type of treatment received.

Time-dependent linear mixed models will allow exploration of longitudinal mediation. For example, we expect obesity and sleep apnea will lead to sleep disturbance, which in turn affects LUTS. In these models, LUTS is a dependent variable regressed on sleep disturbance at the previous time point, which in turn is regressed on the participants obesity and sleep apnea status.

We will also explore other approaches to connect patient phenotype to LUTS trajectory. We will create subgroups based on variables such as treatment received, clinical cut points of PROMIS measures, and objective clinical parameters (e.g., obese vs. non-obese). We will assess changes in SI's from baseline to 12- and 24-weeks and compare these differences across different subgroups. We will further explore latent class trajectory models to create homogenous groups of participants based on their longitudinal urologic symptoms as measured by the SI-10 and SI-29 over follow-up. This statistical method theorizes that a participant's urologic trajectory is driven by their latent class membership. The "latent class" in these models is not directly observed but rather inferred from patterns in the data. The optimal latent class solution will be identified using the Bayesian Information Criterion (BIC) fit statistic (lower BIC indicates a better fit). Once these latent classes are identified, multinomial logistic regression will be used to predict group membership using baseline variables, including objective clinical measures and questionnaire scores. Treatment patterns in each group will also be explored descriptively.

Aim 2: examine psychometric properties of the LURN SIs, including clinically meaningful differences in men with BPH receiving known effective treatment

Test-retest reliability The premise of test-retest reliability is that if a construct (e.g., urinary urgency) does not change, then the questionnaire score should also not substantially change. Thus, changes in scores without intervention can be used as an index of measurement error. LUTS are known to vary over short timeframes; therefore, any changes in questionnaire responses could be due to true changes, not measurement error. We expect, however, the test-retest reliability to be consistent with other LUTS questionnaires.

Test-retest reliability will be assessed for both the LURN SIs and the AUA-SI in two ways. First, we will use an intraclass correlation coefficient (ICC) from a two-way mixed-effects model using absolute agreement assumptions. This model is appropriate for scores obtained multiple times from the same patients, and the absolute agreement measure includes any systematic differences in scores across the two instances (i.e., differences in mean scores) as relevant to the calculation. Given our expected sample size, an ICC estimate above

0.9 would indicate excellent reliability, estimates greater than 0.75 or 0.5 would indicate good or moderate reliability, respectively [31].

Second, we will calculate the coefficient of reliability (CR), an absolute measure of test-retest reliability that quantifies measurement error. The CR is an absolute measure in the sense that it assesses absolute differences (as opposed to consistency, which does not guarantee agreement to be high) and represents the maximum difference between two measurements with 95% probability. It has the advantage of being measured in the same units as the questionnaire, facilitating clinical interpretation. The Pearson correlation will also be calculated as necessary to make comparisons of test-retest reliability with other published questionnaires. The two pre-treatment visits will be the basis for an ICC for both measures. We will also examine graphically the consistency of the LURN SIs over the two-week test-retest period.

Minimally important differences (MIDs)

MIDs are an important psychometric property to establish because statistical differences in PRO measures do not always translate to meaningful differences in patient perception or result in changes in patient management [32]. MIDs should be based on a variety of methods that converge (i.e., triangulation). Three commonly used methods are distribution methods, anchor-based methods, and opinion-based methods; although no single method has been established as preferred, a combination of methods ensures that robust and generalizable MIDs are established. We will use an anchor-based method as primary, with distribution-based and opinion-based methods as secondary/sensitivity analyses to validate our results.

We will administer the PGI-I at each post-baseline time point as the anchor questionnaire. The 7-point PGI-I will be collapsed into three levels (Better, No Change, and Worse), as it is unlikely that we will have sufficient endorsement at all seven ratings for analysis. We will then describe the mean and standard error (SE) of change scores at each of the three levels and compare mean differences across the three ratings of improvement. We hypothesize that mean differences in LURN SI-10 scores and LURN SI-29 subscale scores will be significantly different for patients rating their improvement as Better, No Change, and Worse on the PGI-I based on parametric or non-parametric Analysis of Variance (ANOVA), as appropriate. Inspection of the change difference, i.e., the difference in mean subscale score changes between the better and the average of the No-Change and Worse ratings, will guide the determination of Preliminary MIDs for each subscale score. We will then verify our preliminary MIDs by using them to assess their sensitivity and

specificity for predicting a response of "Better" using Receiver Operating Characteristic (ROC) Curves.

We will then examine the distributions of the LURN SI-10 scores and SI-29 subscales before and after treatment (i.e., baseline vs. 12-week follow-up), as well as the distributions of difference scores. Multiples of the SD of the change scores (e.g., 0.2, 0.5) will be calculated and compared with the preliminary MIDs identified by the anchor approach. We expect that the MIDs identified by the anchor approach will correspond to approximately 0.5 SDs in the change score, representing a meaningful change for many health-related quality of life instruments [33, 34]. MIDs identified by both methods will be shared with stakeholders (i.e., providers who see patients with BPH; patients with BPH) to ensure that these values are clinically sensible. We will also examine changes in the LURN SIs in relationship to the established clinically meaningful change (3 points or more) on the AUA-SI [17]. The final recommended MIDs will be determined by synthesizing these various methods.

Qualitative data analysis

Qualitative interviews with patients and providers will be transcribed and imported into the qualitative analysis software *Dedoose* [35]. Two coders will review a sample of five transcripts to create an initial codebook. They will then meet to discuss and revise the coding scheme. The two coders will independently recode the same transcripts using the updated coding scheme, followed by a second meeting to further refine existing codes and add new ones. Subsequent transcripts will be coded separately by the two coders, who will have regular meetings to maintain consistency. Regular meetings with the research team at large will inform the coding process and resolve any coding discrepancies. The codebook will represent themes expressed by the participants across the interviews. Themes, concerns, and symptoms from the qualitative interviews will be used to determine prevalent topics for inclusion in care coordination. We will then organize the themes around clinical problems that could be resolved by concrete action (e.g., referral to sleep medicine for insomnia, referral to psychology/psychiatry for anxiety).

Sample size and power considerations

Based on prior publications, we expect – under usual treatment – approximately 70%, 20%, and 10% of men to report Better, No change, and Worse, respectively, on the PGI-I at the 12-week assessment [36]. Assuming a sample size of 300 from Aim 1, this represents 210, 60, and 30 participants, respectively. Using a conservative Type I error rate of alpha=0.01 to account for pairwise testing across the three groups, we estimate that we will have at least 80% power to detect effect sizes as small as d=0.5

(a medium-sized effect) [37] comparing the better and no-change groups. We would also have 80% power for effects as small as $d=0.7$ comparing the better and worse groups, and as small as $d=0.8$ comparing the no-change and worse groups. Although some of these effects are large in size (i.e., $d=0.8$), prior work using the PGI-I to assess changes in the AUA-SI in four placebo-controlled randomized controlled trials found effect sizes of 0.75–1.75 between the three groups [17]. Therefore, our study will be sufficiently powered with 300 participants assuming similar magnitude of change in the LURN SIs. For test-re-test reliability we calculated that 60 patients will be needed to obtain an ICC estimate of 0.8 with a 95% confidence interval half-width of no more than 0.1; therefore, participants will be approached until 60 patients have completed two scorable questionnaires [38].

Data monitoring and quality control

The Safety Monitor of the study will provide oversight for data and safety monitoring for this study. The Safety Monitor and the study teams will meet bi-annually to review the data and any adverse events.

Discussion

The goal of this project is to use newly developed LURN SIs to improve the clinical care of patients with LUTS/BPH. The integration of new patient-centered tools will improve evaluation and clinical decision-making by including symptoms not commonly measured in men using standard questionnaires, such as urinary incontinence. Brief PROs will allow for frequent monitoring of LUTS through remote assessment. Using care-coordination recommendations, the health care team can be more responsive to post-treatment symptom changes, resulting in reduced bother from LUTS and higher quality of life in patients with BPH.

Strengths of this study include its longitudinal design, questionnaires that capture symptoms relevant to BPH, integration of questionnaires with data from the medical record, and mixed methods comprising quantitative and qualitative methods. We have taken care to accommodate various preferences by allowing patients to provide PROs through multiple means, including EPIC MyChart, traditional paper-and-pen surveys, or phone interviews. This approach ensures that individuals with varying levels of technological proficiency can participate, minimizing the risk of missing data and bolstering the overall representativeness and reliability of our study's findings. Several limitations are important to note. Patient recruitment for our study is primarily conducted within a single health care organization (NorthShore), which could limit the generalizability of our results. However, analysis and qualitative work is spread across three major academic institutions in Chicago (NorthShore, Northwestern University, and the University of Chicago). Variations in practice patterns, health care staff expertise, and patient perspectives across different settings, institutions, or regions may impact the applicability of our findings to other contexts. A potential limitation of our study is self-selection bias, wherein our sample is composed of participants who choose to respond to our recruitment efforts. This could introduce systematic differences between participants and non-participants, potentially affecting the generalizability of our findings. Another limitation is the observational rather than experimental nature of the study. This was an intentional choice as our goal is to build toward a randomized controlled trial. Lastly, we must address potential biases related to self-reported data, including social desirability bias and memory recall issues. Respondents may provide answers that they perceive as socially acceptable or desirable, which may not reflect their true thoughts or behaviors. Moreover, memory recall issues could introduce errors or inaccuracies into our data. We will be vigilant in considering these potential biases throughout our analysis and interpretation of the results.

In the future, we hope to incorporate the findings of this study, including the use of the LURN SIs, into the clinical practice guidelines for patients with BPH. Future plans may also include randomized-controlled trials on potential drivers of LUTS identified in this study. For example, a clinical test of Aim 3 would be a trial of whether implementation of the care coordination recommendations results in better outcomes for men with BPH and/or decreased health care utilization in these men. Results of this study may also inform the development of patient self-management tools. Finally, although this project focuses on the population of patients with LUTS secondary to BPH, the LURN SIs can be implemented in a variety of disease states, including stricture disease, pelvic floor dysfunction, and prostate cancer, which creates additional avenues for research.

The primary goal of PROs is to give the patient a voice in their health care by having them report on perceived symptoms, bother, and impact. PROs have already had a long history of success with the AUA-SI in clinical care and research. However, after decades of the AUA-SI, it is time for some innovative changes and improvements that more broadly capture the symptoms and needs of men with LUTS/BPH, and to directly connect them to comorbidities and QOL. Implementation of this study will ensure that the LURN SIs do more than merely collect data, but rather use them to guide treatment, thereby improving the care of men with BPH.

Abbreviations

AUA-SI   American Urological Association Symptom Index
BPH   Benign prostatic hyperplasia
LUTS   Lower urinary tract symptoms

LURN	Symptoms of Lower Urinary Tract Dysfunction Research Network
LURN SI-29	LURN Symptom Index 29
LURN SI-10	LURN Symptom Index 10
MSHQ-Ed	Male Sexual Health Questionnaire-Ejaculatory Dysfunction
PGI-I	Patient Global Impression of Improvement
PRO	Patient-reported outcome
PROMIS	Patient-Reported Outcomes Measurement Information System
SBQ	Stop-BANG questionnaire
SHIM	Sexual Health Inventory for Men
UUI	Urgency urinary incontinence
SD	Standard deviation

Data availability
Data sharing is not applicable to this article as no datasets were generated or analyzed in this protocol paper. However, the datasets used and/or analyzed that will be generated during the course of the currently described study will be available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
This study was approved by the Institutional Review Boards of Northwestern University (IRB STU00217126), which served as the single IRB of record. This study is registered in ClinicalTrials.gov (NCT05898932). A coordinator will contact patients by phone or in person to describe the study and obtain informed consent, which will be documented using the Research Electronic Data Capture System (REDCap).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Surgery, Division of Urology, Endeavor Health (Formerly NorthShore University HealthSystem), Evanston, IL, USA
[2]Department of Surgery, Division of Urology, University of Chicago Pritzker School of Medicine, Chicago, IL, USA
[3]Feinberg School of Medicine, Department of Preventive Medicine, Division of Biostatistics, Northwestern University, Chicago, IL, USA
[4]Pritzker School of Medicine, Department of Obstetrics and Gynecology, University of Chicago, Chicago, IL, USA
[5]Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Department of Physics & Astronomy, Northwestern University, Chicago, IL, USA
[6]IT Research Computing and Data Services, Northwestern University, Chicago, IL, USA
[7]Feinberg School of Medicine, Department of Medical Social Sciences, Northwestern University, Chicago, IL, USA
[8]Feinberg School of Medicine, Department of Surgery, Northwestern University, Chicago, IL, USA

References
1. Urologic Diseases in America P: UDA Data Report: Epidemiology of non-malignant urologic disease in the United States: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD. 2003; 2023.
2. Sandhu JS, Bixler BR, Dahm P, Goueli R, Kirkby E, Stoffel JT, Wilt TJ. Management of lower urinary tract symptoms attributed to Benign Prostatic Hyperplasia (BPH): AUA Guideline Amendment 2023. J Urol. 2024;211(1):11–9.
3. Gravas S, Gacci M, Gratzke C, Herrmann TRW, Karavitakis M, Kyriazis I, Malde S, Mamoulakis C, Rieken M, Sakalis VI, et al. Summary Paper on the 2023 European Association of Urology Guidelines on the management of non-neurogenic male lower urinary tract symptoms. Eur Urol. 2023;84(2):207–22.
4. Launer BM, McVary KT, Ricke WA, Lloyd GL. The rising worldwide impact of benign prostatic hyperplasia. BJU Int. 2021;127(6):722–8.
5. Barry MJ, Fowler FJ Jr., O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, Cockett AT. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. J Urol. 1992;148(5):1549–57. discussion 1564.
6. Glaser AP, Mansfield S, Smith AR, Helfand BT, Lai HH, Sarma A, Yang CC, Taddeo M, Clemens JQ, Cameron AP, et al. Impact of Sleep Disturbance, physical function, depression and anxiety on male lower urinary tract symptoms: results from the symptoms of Lower Urinary Tract Dysfunction Research Network (LURN). J Urol. 2022;208(1):155–63.
7. Helfand BT, Smith AR, Lai HH, Yang CC, Gore JL, Erickson BA, Kreder KJ, Cameron AP, Weinfurt KP, Griffith JW, et al. Prevalence and characteristics of urinary incontinence in a treatment seeking male prospective cohort: results from the LURN Study. J Urol. 2018;200(2):397–404.
8. Glaser AP, Kristof TW, Judge C, Javitch G, Hirsch S, Talaty P, Helfand BT, Smith AR, Griffith JW. The 10-item LURN Symptom Index (LURN SI-10) detects additional symptoms and shows convergent validity with the IPSS in men presenting with lower urinary tract symptoms. Urology. 2023;171:184–9.
9. Hines L, Mali K, Campbell T, Jain R, Quarrier SO. Advancing Holmium Laser Enucleation of the prostate outcomes reporting: an Assessment of patient-reported outcomes in the literature. Urology; 2024.
10. Yaxley JW, Coughlin GD, Chambers SK, Occhipinti S, Samaratunga H, Zajdlewicz L, Dunglison N, Carter R, Williams S, Payton DJ, et al. Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study. Lancet. 2016;388(10049):1057–66.
11. Barnard J, Liaw A, Gelman J. Long-term follow-up suggests high satisfaction rates for bulbomembranous radiation-induced urethral stenoses treated with anastomotic urethroplasty. World J Urol. 2023;41(7):1905–12.
12. Cella D, Smith AR, Griffith JW, Flynn KE, Bradley CS, Gillespie BW, Kirkali Z, Talaty P, Jelovsek JE, Helfand BT, et al. A new outcome measure for LUTS: symptoms of lower urinary tract Dysfunction Research Network Symptom Index-29 (LURN SI-29) questionnaire. Neurourol Urodyn. 2019;38(6):1751–9.
13. Cella D, Smith AR, Griffith JW, Kirkali Z, Flynn KE, Bradley CS, Jelovsek JE, Gillespie BW, Helfand BT, Talaty P, et al. A new brief Clinical Assessment of

Lower urinary tract symptoms for women and men: LURN SI-10. J Urol. 2020;203(1):164–70.

14. Weinfurt KP, Griffith JW, Flynn KE, Cella D, Bavendam T, Wiseman JB, Andreev VP, Lai HH, Liu AB, Kirkali Z, et al. The Comprehensive Assessment of Self-reported urinary symptoms: a New Tool for Research on subtypes of patients with lower urinary tract symptoms. J Urol. 2019;201(6):1177–83.

15. Amundsen CL, Helmuth ME, Smith AR, DeLancey JOL, Bradley CS, Flynn KE, Kenton KS, Henry Lai H, Cella D, Griffith JW, et al. Longitudinal changes in symptom-based female and male LUTS clusters. Neurourol Urodyn. 2020;39(1):393–402.

16. Coyne KS, Kaplan SA, Chapple CR, Sexton CC, Kopp ZS, Bush EN, Aiyer LP. Risk factors and comorbid conditions associated with lower urinary tract symptoms: EpiLUTS. BJU Int. Suppl 2009;3:24–32.

17. Viktrup L, Hayes RP, Wang P, Shen W. Construct validation of patient global impression of severity (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower urinary tract symptoms secondary to benign prostatic hyperplasia. BMC Urol. 2012;12:30.

18. Administration FD. Patient-focused drug development guidance public workshop: methods to identify what is important to patients & select, develop or modify fit-for-purpose clinical outcomes assessments. In: *2018*: Workshop date: October (pp. 15–16). 2018.

19. Suri H. Purposeful sampling in qualitative research synthesis. Qualitative Res J. 2011;11:63–75.

20. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, et al. The patient-reported outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010;63(11):1179–94.

21. Guy W. ECDEU assessment manual for psychopharmacology. Rockville, Md: National Institute of Mental Health; 1976.

22. Eating at America's Table Study. Quick Food Scan. [Accessed April 19, 2024]. [https://epi.grants.cancer.gov/diet/shortreg/instruments/percent-energy-from-fat-screener.pdf]

23. Bright E, Cotterill N, Drake M, Abrams P. Developing and validating the International Consultation on Incontinence Questionnaire bladder diary. Eur Urol. 2014;66(2):294–300.

24. Bright E, Cotterill N, Drake M, Abrams P. Developing a validated urinary diary: phase 1. Neurourol Urodyn. 2012;31(5):625–33.

25. 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(11):1804–14.

26. Rosen RC, Cappelleri JC, Smith MD, Lipsky J, Peña BM. Development and evaluation of an abridged, 5-item version of the International Index of

Erectile function (IIEF-5) as a diagnostic tool for erectile dysfunction. Int J Impot Res. 1999;11(6):319–26.

27. Rosen RC, Catania JA, Althof SE, Pollack LM, O'Leary M, Seftel AD, Coon DW. Development and validation of four-item version of male sexual health questionnaire to assess ejaculatory dysfunction. Urology. 2007;69(5):805–9.

28. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, Khajehdehi A, Shapiro CM. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. Anesthesiology. 2008;108(5):812–21.

29. Lai JS, Wagner LI, Jacobsen PB, Cella D. Self-reported cognitive concerns and abilities: two sides of one coin? Psychooncology. 2014;23(10):1133–1141.

30. Atkinson TM, Schwartz CE, Goldstein L, Garcia I, Storfer DF, Li Y, Zhang J, Bochner BH, Rapkin BD. Perceptions of Response Burden Associated with Completion of patient-reported outcome assessments in Oncology. Value Health. 2019;22(2):225–30.

31. Koo TK, Li MY. A Guideline of selecting and reporting Intraclass correlation coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63.

32. Barry MJ, Williford WO, Chang Y, Machi M, Jones KM, Walker-Corkery E, Lepor H. Benign prostatic hyperplasia specific health status measures in clinical research: how much change in the American Urological Association symptom index and the benign prostatic hyperplasia impact index is perceptible to patients? J Urol. 1995;154(5):1770–4.

33. Norman GR, Sloan JA, Wyrwich KW. The truly remarkable universality of half a standard deviation: confirmation through another look. Expert Rev Pharmacoecon Outcomes Res. 2004;4(5):581–5.

34. 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):582–92.

35. Salmona M, Lieber E, Kaczynski D. Qualitative and mixed methods data analysis using dedoose: a Practical Approach for Research across the Social Sciences. Sage; 2019.

36. Viktrup L, Hayes RP, Wang P, Shen W. Construct validation of patient global impression of severity (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower urinary tract symptoms secondary to benign prostatic hyperplasia. BMC Urol. 2012;12(1):30.

37. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd edition). Hillsdale, NJ: L. Erlbaum Associates; 1988.

38. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med. 2002;21(9):1331–5.

Publisher's Note