Accepted Manuscript

British Journal of General Practice

Diagnostic models for male lower urinary tract symptom management in primary care

Harding, Chris; Agarwal, Ridhi; Bates, Janine; Bray, Alison; Milosevic, Sarah; Thomas-Jones, Emma; Drinnan, Michael; Drake, Marcus; Michell, Peter; Pell, Bethan; Ahmed, Haroon; Joseph-Williams, Natalie; Schatzberger, Tom; Davies, Jane; Hood, Kerenza; Takwoingi, Yemisi; Edwards, Adrian

DOI: https://doi.org/10.3399/BJGP.2025.0513

To access the most recent version of this article, please click the DOI URL in the line above.

Received 23 August 2025
Revised 23 February 2026
Accepted 06 April 2026

Diagnostic models for male lower urinary tract symptom management in primary care

Authors: Harding, Chris; Agarwal, Ridhi; Bates, Janine; Bray, Alison; Milosevic, Sarah; Thomas-Jones, Emma; Drinnan, Michael; Drake, Marcus; Michell, Peter; Pell, Bethan; Ahmed, Haroon; Joseph-Williams, Natalie; Schatzberger, Tom; Davies, Jane; Hood, Kerenza; Takwoingi, Yemisi; Edwards, Adrian

Number of words: 3024
References: 29
Figures: 0
Tables: 3
Corresponding Author:
Prof A Edwards MB BS, PhD.
Division of Population Medicine,
Cardiff University
Email - EdwardsAG@cardiff.ac.uk

Affiliations:

- Chris Harding MB BChir MD FRCS Urol, Professor of Urology, Translational and Clinical Research Institute, Newcastle University. Consultant Urological Surgeon, Freeman Hospital, Newcastle-upon-Tyne.
- Ridhi Agarwal BSc, MSc, PhD. Research fellow, Department of Applied Health Sciences, University of Birmingham. National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, University of Birmingham.
- Janine Bates Mphil, Research Associate, Cardiff University.
- Alison Bray BSc,PhD, Clinical Scientist Certificate of Equivalence, Lead Healthcare Scientist, Northern Medical Physics and Clinical Engineering, The Newcastle upon Tyne Hospitals NHS Foundation Trust and Honorary Researcher, Translational and Clinical Research Institute, Newcastle University.
- Sara Milosevic PhD. Research Associate, Centre for Trials Research, Cardiff University
- Emma Thomas-Jones PhD. Principal Research Fellow, Centre for Trials Research, School of Medicine, Cardiff University
- Michael Drinnan PhD, Professor of Digital Healthcare, Teesside University
- Marcus Drake DM FRCS Urol. Chair in Neurological Urology,Department of Surgery & Cancer - Faculty of Medicine, Imperial College London.
- Peter Michell, BSc. MBCS. PPI Representative
- Bethan Pell, Bsc, MSc. PhD Student, DECIPHer, School of Social Sciences, Cardiff University.
- Haroon Ahmed MB BCh, MRCGP, PhD. Reader in Epidemiology, Division of Population Medicine, Cardiff University.

- Natalie Joseph-Williams BSc PGDip PhD. Reader in Improving Patient Care, Division of Population Medicine, School of Medicine, Cardiff University
- Tom Schatzberger MBBS MRCGP. Clinical Lead CBC Health Federation
- Jane Davies PhD BN RN. Honorary Research Fellow, Cardiff University
- Kerenza Hood BSc (hons), PhD, CStat, Dean of Research & Innovation, College of Biomedical & Life Sciences, Cardiff University
- Yemisi Takwoingi, DVM, MSc, PhD. Professor of Test Evaluation and Evidence Synthesis, Department of Applied Health Sciences, University of Birmingham. National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, University of Birmingham.
- Adrian Edwards MB BS, PhD. Professor of General Practice, Division of Population Medicine, Cardiff University

Conflicts of Interest (CoI):

CH: has received payment or honoraria for lectures, consultancy and/or educational activities from Medtronic, Astellas, GSK, Viatris and Convatec - none are directly related to the topic of this manuscript.

MDrake: has received payment or honoraria for lectures, consultancy and/or educational activities from Astellas not directly related to the topic of this manuscript.

AB and MDrinnan : developed the flowtaker device that was the home uroflowmeter used in the study but receive no individual royalty payment

Diagnostic models for male lower urinary tract symptom management in primary care

Abstract – word count 267

Background - recent research has concluded that primary care assessment and management of lower urinary tract symptoms (LUTS) requires practical resources.

Aim - To develop and independently validate risk-prediction models for three common urological conditions based on a combination of simple, minimally-invasive index tests in men with LUTS presenting to primary care.

Design and Setting - A prospective, multicentre, diagnostic accuracy study with independent development and validation cohorts. Set in 67 general practices in England and Wales.

Method – 601 men presenting to primary care with LUTS: 350 in the development cohort and 251 in the validation cohort underwent index tests - validated symptom assessment, bladder diary, physical examination, serum prostate specific antigen, uroflowmetry and post-void residual volume estimation. Invasive urodynamic studies provided a diagnostic reference standard. Risk prediction models for urodynamic diagnoses were developed and then validated in independent cohorts. Discriminative and calibration performance of the risk prediction models for diagnosing bladder outlet obstruction (BOO), detrusor underactivity (DU) and detrusor overactivity (DO) using the c-index, calibration slope and plot and sensitivity and specificity of the models.

Results - The BOO model from the development cohort demonstrated good discriminative performance (optimism-corrected c-index of 0.80). The models derived from the development cohort for DU and DO demonstrated moderate discriminative ability (optimism-corrected c-indices of 0.64 and 0.67 respectively). Similar estimates of c-index were observed for each model within the independent validation cohort (BOO=0.82, DU= 0.63 and DO =0.62).

Conclusions – Using a combination of simple, non-invasive index tests can accurately predict common urological diagnoses in men with LUTS, potentially facillitating earlier initiation of evidence-based treatments and improved management in primary care.

How this fits in?

Current assessment and management of lower urinary tract symptoms in primary care requires practical resources specifically around diagnostics. The aetiology of LUTS is varied and specifically in older men our results show that it should not be automatically assumed to be secondary to benign prostatic enlargement. The diagnostic decision tool detailed in this manuscript has the potential to increase the level of accuracy surrounding the aetiology of lower urinary tract symptoms and hence guide primary care clinicians towards the most appropriate treatment.

The PRIMUS study has developed unvalidated a diagnostic tool for use in men with lower urinary tract symptoms that potentially improves the accuracy of diagnoses and consequently the appropriateness of first-line treatments.

Trial registration and funding – ISRCTN 10327305. PriMUS was funded by the National Institute for Health Research (NIHR) Health Technology Assessment REF 15/40/05.

Introduction

Lower urinary tract symptoms (LUTS) in men are common, bothersome and increasingly lead to secondary care referral. In the UK, LUTS affect one in four men, with around 300,000 per year consulting their GP. This represents a considerable long-term health burden [1,2,3]. The results from a recent study reported a 7-8% secondary care referral rate for male LUTS. The authors concluded that *"as provision for men with LUTS in primary care is inconsistent, primary care health professionals require practical resources to support the assessment of urinary symptoms"* [4].

Current guidelines, including those from the National Institute of Health and Care Excellence (NICE), [5,6] suggest that many men referred to specialist care with LUTS should be treated conservatively with lifestyle advice or simple medications, which should allow initial management to remain in primary care.

After ruling out urinary tract infection and other serious causes of LUTS, the initial medical management of LUTS differs significantly according to underlying diagnosis. The reference standard diagnostic test for male patients with LUTS is urodynamic testing [6]. This invasive test involves the measurement of bladder and abdominal pressures during both the filling and voiding phases of the micturition cycle. Urodynamics enable diagnoses for the more common causes of urinary symptoms; bladder outlet obstruction (BOO), detrusor underactivity (DU) and detrusor overactivity (DO) to be definitively made. However, urodynamic testing is invasive, requires expensive equipment, highly-trained staff and currently takes place exclusively in secondary care.

With access to simple, accurate tests and clinical decision support tools, GPs could offer accessible, effective, timely management of LUTS in primary care. This would mirror the management of other long-term conditions such as diabetes, asthma and hypertension. In the PRImary care Management of lower Urinary tract Symptoms in men (PriMUS) study [7] our aim was to develop such a tool.

The objectives in this research were:

- to develop models to predict the likelihood of BOO, DU and DO from simple non-invasive index tests that are feasible in primary care;
- to estimate the diagnostic accuracy of these models in an independent validation cohort.

Methods

The study protocol was previously published [7] and more detailed information regarding the trial can also be found in the NIHR Health Technology Assessment report [8]. PPI included design and management of the research, including developing participant information resources, analysing the research, contributing to the reporting and dissemination of research findings. In summary men presenting to their GP with LUTS were recruited prospectively into two cohorts; one for development and one for validation of the diagnostic models. Participants underwent both simple index tests and the reference test (urodynamic studies) to determine which combination of index tests best predicted three common urodynamic diagnoses in men:

BOO, DU and DO. Interpretation of the urodynamic study was conducted by two independent reviewers providing internal validation of reference test results.

The index tests were symptom assessment using International Prostate Symptom Score (IPSS) and International Consultation on Incontinence (ICIQ) [9,10] questionnaires, completion of a bladder diary, physical examination including digital rectal examination, serum prostate specific antigen, home and in-clinic uroflowmetry and post-void residual volume estimation with ultrasound. Patients then underwent invasive urodynamic studies according to a standardised operating procedure to provide a diagnostic reference.

Sample size calculations were performed separately for the model development and validation cohorts. For both, estimated prevalences for BOO, DU and DO of 31%, 16% and 57% were used, respectively, based on previous literature [11,12] and clinical expertise.

The sample size for developing the risk prediction models was based on a rule of thumb suggesting that five events per variable was adequate [13]. A sample size of 350 was chosen to allow at least 11 variables in each model. This was driven by our lowest estimated prevalence of 16% for DU, giving 56 events (diagnoses). Over 100 events were expected for the BOO model and 200 for the DO model providing a more than adequate number.

The required sample size for the validation cohort was chosen to ensure that estimates of test accuracy could be made with adequate precision. Sensitivity and specificity of 75% were deemed to be the minimum clinically useful performance. A sample size of 325 was chosen, giving estimates of sensitivity of 75% to within 8%, 10% and 14% for DO, BOO and DU respectively, based on 'positive' samples of 185, 101 and 52, and estimates of specificity of 75% to within 8%, 7% and 6%, based on 'negative' samples of 140, 224 and 273. To allow an attrition rate of 20%–25%, the resulting 675 was increased to give a final required sample size of 880.

Sensitivity analyses were performed by fitting two alternative models for each target condition. In sensitivity analysis 1 (SA1), predictors that may be difficult to obtain in practice were excluded from the list of candidate predictors (mean urgency score and mean 24-hour fluid intake from the bladder diary). In sensitivity analysis 2 (SA2), alternative measures or methods of measurement of candidate predictors were considered (for example using measurements from in-clinic flow tests rather than home uroflowmetry).

Risk prediction models were developed for each target condition using logistic regression. Multiple imputation by chained equations was used to handle missing data and fractional polynomial functions were used to fit continuous variables. Discrimination (ability to distinguish between those who do or do not have the target condition) of each model was assessed using the c-index, and calibration (agreement between predicted and observed probabilities) was assessed using calibration plots and calibration slope. Internal validation was conducted to assess optimism of the performance statistics using bootstrapping. External validation was conducted to assess model performance in another sample from a similar population as the development cohort (validation cohort). In both forms of validation, the models were recalibrated using the calibration slope as a shrinkage factor to re-estimate the intercept and model coefficients. Sensitivity and specificity were plotted on a ROC plot for each model. Risk thresholds were identified at a sensitivity and specificity of 75%, which was deemed to be the minimum clinically useful performance.

Results

350 and 251 men were recruited into the development and validation cohorts respectively. The recruitment for the validation cohort was less than planned, largely due to the COVID-19 restrictions imposed during the trial period. Baseline characteristics of men recruited to development and validation cohorts are shown in Table 1. The median age was 69 years (Interquartile range 61-74). Majority (97.4%) of the men were of white ethnicity. Almost half

(n=171, 48.9%) of the men were ex-smokers, 150 (42.9%) had never smoked and 29 (8.3%) were current smokers. Of the 350 men, over two-thirds had comorbidities: 107 (30.6%) had no comorbidities, 188 (53.7%) had a single comorbidity and the remaining 55 (15.7%) had multiple comorbidities. There were 41 (11.7%) men with diabetes mellitus II, 13 (3.7%) with obstructive sleep apnoea and 12 (3.4%) with cerebrovascular accident. There were 102 (29.1%) men on alpha-blockers, 65 (18.6%) on calcium channel blockers, 26 (7.4%) on selective serotonin reuptake inhibitors (SSRIs) antidepressants and 25 (7.1%) participants on 5-alpha reductase inhibitors. Summary statistics of the candidate predictors are shown in Table 2.

Of 601 men recruited, 545 (91%) had at least one of the three common urological conditions on urodynamic testing, with prevalence from 34.7% (DU in validation cohort) to 72.3% (DO in development cohort).

Models were developed and validated for BOO, DU and DO (Table 3). Six index tests were required for the 3 models (age, prostate specific antigen test, voiding symptoms sub-score (ICIQ-MLUTS), median maximum flow rate, median voided volume and post-void residual urine). Optimism-corrected calibration slopes less than 1 suggest that all models were overfitted at development, though the BOO model demonstrated good calibration performance at validation.

In the development cohort, the BOO model demonstrated good discriminative performance with an optimism-corrected c-index of 0.80, whereas the models for DU and DO were moderately discriminative. This was evident in the diagnostic performance of the validation cohort where only the BOO model approached the criterion of >75% sensitivity & specificity.

Sensitivity analysis

The diagnostic performance of the risk prediction models for BOO, DU and DO from SA1 remained the same with the exclusion of the mean urgency score and the mean 24-hour fluid intake, suggesting that the models were robust and the exclusion of these parameters did not affect the model's performance. A total of 6 index tests were required for the 3 models in SA2 (age, prostate specific antigen test, maximum flow rate, voided volume, mean 24-hour fluid intake and storage symptoms sub-score (IPSS). The optimism-corrected c-index reduced significantly for the BOO model (0.69), but was comparable for DU (0.63) and DO (0.65). The optimism-corrected calibration slope was less than one for all models, suggesting overfitting (BOO:0.83, DU:0.80, DO:0.74).

Discussion

Summary

The PriMUS study shows that in a cohort of men presenting to primary care with LUTS, a combination of non-invasive clinical assessments can be used to estimate the likelihood of common urological conditions with a degree of accuracy comparable with existing primary care clinical support tools in other disease areas.

Comparison with existing literature

Various simple tests have been previously assessed for diagnostic accuracy, most often regarding their ability to predict BOO. Symptom assessment alone was associated with poor diagnostic accuracy in two large multi-national studies [14,15]. Maximum urine flow rate, a widely used variable obtained from a non-invasive urine flow study had no correlation with urinary symptoms and was unable to accurately diagnose urological conditions such as BOO. Similarly, post-void residual volume was found to have a weak correlation with BOO [17,18]. However, the diagnostic ability of serum PSA was found to have a significant correlation with

BOO [19]. Despite this the current use of serum PSA measurement is largely confined to its ability to predict underlying prostate cancer.

The existing literature suggests that when used alone, simple tests do not carry sufficient levels of accuracy to be useful in the diagnosis of male LUTS. Other researchers have investigated the use of a combination of measures derived from simple tests to provide a urodynamic diagnosis of BOO [20]. A combination of clinical measurements including prostate volume, maximum flow rate and post-void residual volume was shown to better predict BOO than symptom scores alone. These findings align with our results. The PriMUS study used a larger cohort and more robust methodology than previously published studies and showed that combining results from simple tests can accurately diagnose the cause of male LUTS. More recently, the UPSTREAM study identified that core evaluations used in primary care may predict outcomes of interventional treatment [21,22] further suggesting that simple, non-invasive tests have significant clinical utility.

Another important finding from the PriMUS study is the high prevalence of defined urodynamic abnormalities in men presenting to primary care with LUTS. Our findings are supported by other data which estimate a high prevalence of underlying urodynamic abnormalities but in an unselected population [23].

Implications for practice

Previously published qualitative work from PriMUS indicates that patients prefer initial LUTS management in primary care [24]. These findings mirror the views expressed in a 2024 report which revealed that the public in the UK wants primary and community care to be a higher priority for NHS resources than hospital services [25].

Despite the high prevalence of urological conditions identified in this study, baseline data from participants indicated that most men had not yet tried potentially suitable first-line medical treatments. The PriMUS clinical decision support tool, using these diagnostic models and with its management algorithm based on existing NICE guidelines, could facilitate the initial accurate management of these patients in a primary care setting whilst also avoiding a “trial and error” approach of different medications [8]. For example, most symptomatic men in our study were over 60 years old and an underlying diagnosis of BOO might have been assumed. Less than 50% in both cohorts were found to actually have BOO indicating that a trial of alpha-adrenoreceptor antagonists, which would be common practice in this scenario, would not be successful. A different approach based on simple, rapidly-available diagnostic tests such as illustrated in the PRiMUS study is likely to result in those men getting the most appropriate treatment more quickly and avoiding repeat consultations, thus potentially saving costs.

The levels of diagnostic accuracy we report are comparable to other (non-urological) predictive models that are routinely used in primary care, such as QRisk,[26] HAS-BLED [27] and CHADS-VASC2 [28] albeit with different timespans for predicting risk. These other scales / scores are used routinely and integrated into clinical record systems in primary care. This provides some confidence that with implementation in routine settings, there will be opportunity for the PRIMUS risk prediction models to be made available via the standard primary care record systems (EMIS, Vision, SystemOne).

The variables included in our proposed model and clinical diagnostic tool include simple demographic data, symptom scores and PSA tests, already routinely administered by primary care staff. Other variables required for the model include simple measurements that can be obtained from a home uroflowmetry test. We found 92.5% (556/601) of men successfully used the home uroflowmetry device and provided usable data to input into the model. This suggests that the addition of a simple home uroflowmetry test, used alongside existing investigations carried out in primary care, may enable men presenting with LUTS to be diagnosed and potentially treated within general practice. Another predictive variable (for DO and DU) that does present more challenges is post-void residual volume estimation (from ultrasound). We

envisage implementation via a pathway involving a GP with Special Interest, operating at the level of a cluster or network of practices or community diagnostic clinics. Primary care capacity to undertake such clinical management requires resourcing but could be funded from costs potentially saved by avoiding early secondary care referral.

Strengths and limitations

The PriMUS study was conducted in line with current Good Clinical Practice and topic-specific research guidelines [29] and oversight throughout was provided by an external trial steering committee. We used urodynamic testing as the reference standard for diagnosis and the interpretation of this test was conducted by two independent reviewers providing internal validation of the reference test results.

The sample size of 601 men recruited from different geographical, socioeconomic and ethnic backgrounds makes this one of the largest studies in this field. Our sample was lower than protocol targets mainly because urodynamic studies necessitate regular coughs for quality control purposes and there was reticence regarding clinic attendance during COVID. The prevalence of the three diagnoses was considerably higher than we had assumed for our sample size calculation and therefore this lower sample size did not increase the width of the confidence intervals to a level that would reduce our ability to draw robust conclusions.

A limitation of the study is that in terms of diagnostic accuracy the model only approached the pre-specified sensitivity and specificity threshold for clinically useful test performance for the diagnosis of BOO. We believe that this alone will be useful for initiial LUTS assessment in men presenting to primary care. The majority of LUTS presentations are likely to be from older men and the logical assumption is that the undelying diagnosis would be BOO due to benign prostatic enlargement. However our data would refute that and in our cohort of men with a median age of over 68yrs the most prevalent diagnosis was DO. Hence the abillity to accurately predict the presence or absence of BOO would potentially avoid inappropriate first-line empirical treatment directed at a possible BOO diagnosis which may not be present.

Another limitation of the study is that it does not evaluate the possible impact of pathway change for men presenting with LUTS, specifically with regard to outcomes important to patients and potential cost-savings for the healthcare system. The potential for primary care clinicians to accurately diagnose and initially manage these patients is however demonstrated by our results.

Finally we acknowledge there is large uncertainty around the calibration slope, likely driven by the smaller than planner sample size of the validation cohort during the COVID-19 pandemic. Thus, further external validation and recalibration may be required before the models can be widely used in practice. This may ensure applicability of the models in settings where case-mix or prevalence could differ to ensure the calibration performance of the models. We acknowledge that the “five events per variable” rule is now outdated, since this often leads to insufficient sample size, especially for complex models, which in turn can result in the models overfitting. However, at the time of study development, this rule was accepted.

Implication for research

Further research to evaluate the use of the PriMUS diagnostic tool in practice, particularly in terms of potential effects on patient-important outcomes and resource use (practitioner time, medications, referrals), is necessary before full integration into clinical pathways. Both clinical effectiveness and process evaluations will provide valuable information for wider implementation and examine which patients may most benefit from increased use of simple diagnostic tests in primary care such as those who may have failed empirical first-line treatment.

Conclusion

The PriMUS study has shown the potential for combining the results of simple index tests to improve the diagnosis of the three common causes of LUTS in men in primary care. The high prevalence of urodynamic abnormalities within a cohort of men presenting to primary care with LUTS emphasises the clinical need for better and more widespread access to diagnostics and initial treatment and suggests that this could take place within primary care, aligning with patient preferences.

Table 1: Participant demographics and clinical characteristics of the development and validation cohorts

Participant characteristic		Development cohort, n=350	Validation cohort, n=251
Prevalence, n (%)	BOO	163 (46.6)	112 (44.6)
	DU	141 (40.3)	87 (34.7)
	DO	253 (72.3)	166 (66.1)
Age (years)	Median (IQR)	69.0 (61.8–74.9)	67.6 (60.6–74.0)
	Missing, n (%)	0 (0.0)	0 (0.0)
Height (cm)	Median (IQR)	176.0 (171.0–180.0)	176.0 (172.0–181.0)
	Missing, n (%)	1 (0.3)	5 (2.0)
Weight (kg)	Median (IQR)	83.5 (75.0–96.0)	86.0 (76.7–97.5)
	Missing, n (%)	0 (0.0)	7 (2.8)
Ethnicity, n (%)	White	341 (97.4)	237 (94.4)
	Asian (Pakistani)	1 (0.3)	2 (0.8)
	Asian (Chinese)	1 (0.3)	0 (0.0)
	Asian (Indian)	2 (0.6)	3 (1.2)
	Asian (Other)	2 (0.6)	1 (0.4)
	Black (Caribbean)	2 (0.6)	1 (0.4)
	Black (African)	0 (0.0)	2 (0.8)
	Black (Other)	0 (0.0)	1 (0.4)
	Other	1 (0.3)	2 (0.8)
	Missing	0 (0.0)	2 (0.8)
Smoking status, n (%)	Never	150 (42.9)	132 (52.6)
	Current	29 (8.3)	17 (6.8)
	Ex-smoker	171 (48.9)	98 (39.0)
	Missing	0 (0.0)	4 (1.6)
Medical history, n (%)	None	107 (30.6)	70 (27.9)
	Adrenal insufficiency	0 (0.0)	0 (0.0)
	Cerebrovascular accident	12 (3.4)	17 (6.8)
	Chronic heart failure	2 (0.6)	6 (2.4)
	Chronic venous stasis	0 (0.0)	2 (0.8)
	Dependent oedema	3 (0.9)	2 (0.8)
	Diabetes insipidus	0 (0.0)	0 (0.0)
	Diabetes mellitus I	1 (0.3)	0 (0.0)
	Diabetes mellitus II	41 (11.7)	28 (11.2)
	Hypercalcaemia	3 (0.9)	0 (0.0)
	Liver failure	0 (0.0)	0 (0.0)
	Obstructive sleep apnoea	13 (3.7)	12 (4.8)
	Polyuric renal failure	0 (0.0)	0 (0.0)
	Pyelonephritis	0 (0.0)	0 (0.0)
	Sickle cell anaemia	1 (0.3)	0 (0.0)
	Other	232 (66.3)	167 (66.5)

IQR = interquartile range

Table 2: Summary of candidate predictors by the development and validation cohorts.

Analyses			Candidate predictor		Development cohort, n=350	Validation cohort, n=251
Main	SA1	SA2				
✓	✓	✓	Age, years (Participant demographics)	Median (IQR)	69.0 (61.8 – 74.9)	67.6 (60.6−74.0)
				Missing, n (%)	0 (0.0)	0 (0.0)
✓	✓	✓	Size of prostate (Digital rectal examination), n (%)	Normal	101 (28.9)	68 (27.1)
				Mild Enlargement	101 (28.9)	79 (31.5)
				Moderate Enlargement	137 (39.1)	95 (37.9)
				Gross Enlargement	11 (3.1)	7 (2.8)
				Missing	0 (0.0)	2 (0.8)
✓	✓	✓	Prostate specific antigen test, ng/ml (Blood test)	Median (IQR)	1.6 (0.9 – 2.8)	1.3 (0.7 – 2.5)
				Missing, n (%)	0 (0.0)	2 (0.8)
✓	✓		Incontinence symptoms sub-score (ICIQ MLUTS questionnaire)	Median (IQR)	4.0 (3.0 – 7.0)	4.0 (3.0 – 7.0)
				Missing, n (%)	0 (0.0)	0 (0.0)
✓	✓		Voiding symptoms sub-score (ICIQ MLUTS questionnaire)	Median (IQR)	7.5 (5.0 –10.0)	7.0 (4.0 –10.0)
				Missing, n (%)	0 (0.0)	0 (0.0)
✓	✓		Post-void residual urine, ml (Bladder ultrasound)	Median (IQR)	67.0 (29.0 – 150.0)	48.0 (18 –114)
				Missing, n (%)	9 (2.6)	8 (3.2)
✓	✓		Median maximum flow rate, ml/s (Home Uroflowmetry)	Median (IQR)	12.1 (9.1 – 16.6)	12.7 (9.2 – 17.2)
				Missing, n (%)	0 (0.0)	39 (15.5)
✓	✓		Median voided volume, ml (Home Uroflowmetry)	Median (IQR)	180.0 (139.0 – 233.0)	183.0 (137–254)
				Missing, n (%)	0 (0.0)	40 (15.9)
✓	✓		Mean 24-hour frequency (Home Uroflowmetry)	Median (IQR)	7.4 (5.6 – 9.2)	7.8 (5.9 – 10.1)
				Missing, n (%)	0 (0.0)	40 (15.9)
✓	✓		Mean nocturia frequency (Home Uroflowmetry)	Median (IQR)	1.6 (1.0 – 2.4)	1.8 (1.2 – 2.8)
				Missing, n (%)	4 (1.1)	41 (16.3)
✓		✓	Mean urgency score (Bladder diary)	Median (IQR)	1.5 (1.1 – 2.1)	1.6 (1.0 – 2.3)
				Missing, n (%)	20 (5.7)	44 (17.5)
✓		✓	Mean 24-hour fluid intake, ml (Bladder diary)	Median (IQR)	1830.0 (1412.0 –2236.5)	1833 (1433 – 2340)
				Missing, n (%)	34 (9.7)	60 (23.9)
		✓	Incontinence symptoms sub-score (IPSS questionnaire)	Median (IQR)	9.0 (6.0 – 11.0)	9.0 (6.0 – 11.0)
				Missing, n (%)	0 (0.0)	0 (0.0)
		✓	Voiding symptoms sub-score (IPSS questionnaire)	Median (IQR)	7.0 (4.0 – 11.0)	7.0 (3.0 – 11.0)

				Missing, n (%)	2 (0.6)	0 (0.0)
		✓	Palpable bladder (Physical examination), n (%)	Yes	11 (3.1)	8 (3.2)
				No	339 (96.9)	241 (96.0)
				Missing	0 (0.0)	2 (0.8)
		✓	Maximum flow rate, ml/s (In-clinic flow test)	Median (IQR)	7.6 (4.0 – 11.1)	7.7 (5.0 – 13.0)
				Missing, n (%)	36 (10.3)	23 (9.2)
		✓	Voided volume, ml (In-clinic flow test)	Median (IQR)	114.0 (58.0 – 238.0)	123.5 (70 –244)
				Missing, n (%)	29 (8.3)	21 (8.4)
		✓	Average number of voids (Bladder diary)	Median (IQR)	9.3 (7.3 – 11.3)	9.0 (7.3 – 11.0)
				Missing, n (%)	34 (9.7)	59 (23.5)
		✓	Nocturnal polyuria – proportion of 24-hr urine output produced during sleeping hours (Bladder diary)	Median (IQR)	0.3 (0.3 – 0.4)	0.3 (0.2 – 0.4)
				Missing, n (%)	48 (13.7)	74 (29.5)

✓ = included in specified analysis

IQR = interquartile range

SA1 – sensitivity analysis 1

SA2 – sensitivity analysis 2

Table 3: Performance measures of the risk prediction models in the main analysis.

	Development cohort		Validation cohort			
	Optimism-corrected Calibration slope	Optimism-corrected C-index	Calibration Slope (95% CI)	C-index (95% CI)	Sensitivity at 75% specificity (threshold)	Specificity at 75% sensitivity (threshold)
BOO	0.87	0.80	0.99 (0.68, 1.30)	0.82 (0.76, 0.88)	71.3% (0.533)	74.2% (0.509)
DU	0.77	0.64	0.82 (0.31, 1.32)	0.63 (0.56, 0.71)	39.8% (0.414)	47.3% (0.342)
DO	0.78	0.67	0.72 (0.27, 1.17)	0.62 (0.55, 0.70)	33.3% (0.752)	45.6% (0.638)

Funding: PriMUS was funded by the National Institute for Health Research (NIHR) Health Technology Assessment REF 15/40/05.

Ethical approval: The Wales REC 6 approved the study (17/WA/0155) on the 20 June 2017 and subsequent research and development approval obtained for Wales on 21 August 2017 and HRA approval on 23 August 2017.

Conflicts of Interest (CoI):

CH: has received payment or honoraria for lectures, consultancy and/or educational activities from Medtronic, Astellas, GSK, Viatris and Convatec - none are directly related to the topic of this manuscript.

MDrake: has received payment or honoraria for lectures, consultancy and/or educational activities from Astellas not directly related to the topic of this manuscript.

AB and MDrinnan : developed the flowtaker device that was the home uroflowmeter used in the study but receive no individual royalty payment

REFERENCES

1. Boyle P, Robertson C, Mazzetta C et al: UrEpik Study Group. The prevalence of lower urinary tract symptoms in men and women in four centres. The UrEpik study. Journal of the British Association of Urological Surgeons.2003; 92: 409 –414.
2. Clifford GM, Logie J, Farmer RD. How do symptoms indicative of BPH progress in real life practice? The UK experience. Eur Urol. 2000;38 Suppl 1:48-53.
3. Treagust J, Morkane T, Speakman M. Estimating a population's needs for the treatment of lower urinary tract symptoms in men: what is the extent of unmet need? J Public Health Med. 2001 Jun;23(2):141-7. doi: 10.1093/pubmed/23.2.141. PMID: 11450931.
4. Drake MJ, Worthington J, Frost J et al. Treatment of lower urinary tract symptoms in men in primary care using a conservative intervention: cluster randomised controlled trial. BMJ2023;383:e075219 http://dx.doi.org/10.1136/bmj-2023-075219
5. NICE CG97 Lower urinary tract symptoms: full guidance [Internet]. London: National Institute for Health and Clinical Excellence; 2015 [accessed Feb 2024]. https://www.nice.org.uk/guidance/cg97/resources/lower-urinary-tract-symptoms-in-men-management-pdf-975754394053
6. Cornu JN, Gacci M, Hashim H et al. EAU guidelines on management of non-neurogenic male lower urinary tract symptoms (LUTS), incl. Benign Prostatic Obstruction (BPO), Arnhem, the Netherlands: EAU Guidelines Office; 2023.
7. Pell B, Thomas-Jones E, Bray A et al. PRImary care Management of lower Urinary tract Symptoms in men: protocol for development and validation of a diagnostic and clinical decision support tool (the PriMUS study). BMJ Open. 2020 Jun 30;10(6):e037634. doi: 10.1136/bmjopen-2020-037634. PMID: 32606065; PMCID: PMC7328815.
8. Edwards A, Agarwal R, Bates J et al. Development of a clinical decision support tool for Primary care Management of lower Urinary tract Symptoms in men: the PriMUS study. Health Technol Assess 2025;29(XX). https://doi.org/10.3310/RGTW5711
9. Barry MJ, Fowler FJ, O'Leary MP, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. J Urol. 1992;148(5):1549-1557
10. Abrams P, Avery K, Gardener N, Donovan J, ICIQ Advisory Board . The International Consultation on Incontinence Modular Questionnaire: http://www.iciq.net. JUrol. 2006;175:1063–1066. - PubMed
11. Kuo HC. Videourodynamic Analysis of Pathophysiology of Men with Both Storage and Voiding Lower Urinary Tract Symptoms. Urology. 2007 August;70(2):272-6. PubMed PMID: 47362251.
12. Van Doorn B, Blanker MH, Kok ET, Westers P, Bosch JLHR. Prevalence, incidence, and resolution of nocturnal polyuria in a longitudinal community-based study in older men: The Krimpen study. European Urology. 2013 March;63(3):542-7. PubMed PMID: 52261027.
13. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and cox regression. American Journal of Epidemiology. 2007 March;165(6):710-8. PubMed PMID: 46348374.
14. Neal DE. The National Prostatectomy Audit. British Journal of Urology, Supplement. 1997;79(2):69-75. PubMed PMID: 28321216
15. Abrams P, Donovan JL, De La Rosette JJMCM, Schafer W. International Continence Society 'Benign Prostatic Hyperplasia' Study: Background, aims, and methodology. Neurourology and Urodynamics. 1997;16(2):79-91. PubMed PMID: 27066750.

16. Reynard JM, Yang Q, Donovan JL et al. The ICS-'BPH' Study: Uroflowmetry, lower urinary tract symptoms and bladder outlet obstruction. British Journal of Urology. 1998;82(5):619-23. PubMed PMID: 28497742.
17. Kranse R, Van Mastrigt R. Weak correlation between bladder outlet obstruction and probability to void to completion. Urology. 2003 01 Oct;62(4):667-71. PubMed PMID: 37205002.
18. Kolman C, Girman CJ, Jacobsen SJ, Lieber MM. Distribution of post-void residual urine volume in randomly selected men. J Urol. 1999 January;161(1):122-7. PubMed PMID: 29422305.
19. Laniado ME, Ockrim JL, Marronaro A, Tubaro A, Carter SS. Serum prostate-specific antigen to predict the presence of bladder outlet obstruction in men with urinary symptoms. BJU International. 2004 December;94(9):1283-6. PubMed PMID: 40052094.
20. Vesely S, Knutson T, Fall M, Damber JE, Dahlstrand C. Clinical diagnosis of bladder outlet obstruction in men with lower urinary tract symptoms: Reliability of commonly measured parameters and the role of idiopathic detrusor overactivity. Neurourology and Urodynamics. 2003;22(4):301-5. PubMed PMID: 36760257.
21. Drake MJ, Lewis AL, Young GJ et al. Diagnostic Assessment of Lower Urinary Tract Symptoms in Men Considering Prostate Surgery: A Noninferiority Randomised Controlled Trial of Urodynamics in 26 Hospitals. Eur Urol. 2020;78(5):701-10.
22. Ito H, Sakamaki K, Young GJ et al. Predicting Prostate Surgery Outcomes from Standard Clinical Assessments of Lower Urinary Tract Symptoms To Derive Prognostic Symptom and Flowmetry Criteria. Eur Urol Focus. 2023.
23. Irwin DE, Kopp ZS, Agatep B, Milsom I, Abrams P. Worldwide prevalence estimates of lower urinary tract symptoms, overactive bladder, urinary incontinence and bladder outlet obstruction. BJU International. 2011 October;108(7):1132-8. PubMed PMID: 362557900.
24. Milosevic S, Joseph-Williams N, Pell B et al. Managing lower urinary tract symptoms in primary care: qualitative study of GPs' and patients' experiences. The British journal of general practice : the journal of the Royal College of General Practitioners. 2021;71(710):e685-e92.
25. Cameron G, Buzelli L, Duxbury K et al. The public's views on the future of the NHS in England: findings from deliberative research and polling with the public and what it means for the next government. Health Foundation; 2024 (https://doi.org/10.37829/HF-2024-P01).
26. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. Bmj. 2017 May 23;357:j2099. PubMed PMID: 28536104.
27. Pisters R, Lane DA, Nieuwlaat R et al. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest. 2010 Nov;138(5):1093-100. PubMed PMID: 20299623.
28. Lip GYH, Nieuwlaat R, Pisters R et al. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: The Euro Heart Survey on atrial fibrillation. Chest. 2010 01 Feb;137(2):263-72. PubMed PMID: 358305527.
29. Selman LE, Ochieng CA, Lewis AL, Drake MJ, Horwood J. Recommendations for conducting invasive urodynamics for men with lower urinary tract symptoms: Qualitative interview findings from a large randomized controlled trial (UPSTREAM). Neurourol Urodyn. 2019 01;38(1):320-9. PubMed PMID: 30311686.