EUROPEAN UROLOGY FOCUS xxx (xxxx) xxx–xxx

available at www.sciencedirect.com
journal homepage: www.europeanurology.com/eufocus

Benign Prostatic Obstruction

Predicting Prostate Surgery Outcomes from Standard Clinical Assessments of Lower Urinary Tract Symptoms To Derive Prognostic Symptom and Flowmetry Criteria

Hiroki Ito[a,b], *Kentaro Sakamaki*[c], *Grace J. Young*[d,e], *Peter S. Blair*[d,e], *Hashim Hashim*[a], *J. Athene Lane*[d,e], *Kazuki Kobayashi*[f], *Madeleine Clout*[d,e], *Paul Abrams*[a], *Christopher Chapple*[g], *Sachin Malde*[h], *Marcus J. Drake*[a,i,*]

[a] *Bristol Urological Institute, North Bristol NHS Trust, Southmead Hospital, Bristol, UK;* [b] *Department of Urology, Yokohama City University, Yokohama, Japan;* [c] *Center for Data Science, Yokohama City University, Yokohama, Japan;* [d] *Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK;* [e] *Bristol Trials Centre, Bristol Medical School, University of Bristol, Bristol, UK;* [f] *Department of Urology, Yokosuka Kyosai Hospital, Yokosuka, Japan;* [g] *Sheffield Teaching Hospitals NHS Trust, Royal Hallamshire Hospital, Sheffield, UK;* [h] *Urology Centre, Guy's and St. Thomas' NHS Foundation Trust, London, UK;* [i] *Department of Surgery and Cancer, Imperial College, London, UK*

Article info

Article history:
Accepted June 22, 2023

Associate Editor: Christian Gratzke

Keywords:
Lower urinary tract symptoms
Prostate surgery
Predictive model
Machine learning
Male
Prognostication

Abstract

Background: Assessment of male lower urinary tract symptoms (LUTS) needs to identify predictors of symptom outcomes when interventional treatment is planned.
Objective: To develop a novel prediction model for prostate surgery outcomes and validate it using a separate patient cohort and derive thresholds for key clinical parameters.
Design, setting, and participants: From the UPSTREAM trial of 820 men seeking treatment for LUTS, analysis of bladder diary (BD), International Prostate Symptom Score (IPSS), IPSS-quality of life, and uroflowmetry data was performed for 176 participants who underwent prostate surgery and provided complete data. For external validation, data from a retrospective database of surgery outcomes in a Japanese urology department (n = 227) were used.
Outcome measurements and statistical analysis: Symptom improvement was defined as a reduction in total IPSS of ≥3 points. Multiple logistic regression, classification tree analysis, and random forest models were generated, including versions with and without BD data.
Results and limitations: Multiple logistic regression without BD data identified age (p = 0.029), total IPSS (p = 0.0016), and maximum flow rate (Q_{max}; p = 0.066) as predictors of outcomes, with area under the receiver operating characteristic curve (AUC) of 77.1%. Classification tree analysis without BD data gave thresholds of IPSS <16 and Q_{max} ≥13 ml/s (AUC 75.0%). The random forest model, which included all clinical parameters except BD data, had an AUC of 94.7%. Internal validation using the bootstrap method showed reasonable AUCs (69.6–85.8%). Analyses using BD data marginally improved the model fits. External validation gave comparable AUCs for logistic

* Corresponding author. Department of Surgery and Cancer, Imperial College, London, UK.
E-mail address: marcus.drake@imperial.ac.uk (M.J. Drake).

https://doi.org/10.1016/j.euf.2023.06.013

regression, classification tree analysis, and random forest models (all without BD; 70.9%, 67.3%, and 68.5%, respectively). Limitations include the significant number of men with incomplete baseline data and limited assessments in the external validation cohort.
Conclusions: Outcomes of prostate surgery can be predicted preoperatively using age, total IPSS, and uroflowmetry data, with prognostic thresholds of 16 for IPSS and 13 ml/s for Q_{max}.
Patient summary: This study identified key preoperative factors that can predict outcomes of prostate surgery for bothersome urinary symptoms, including which patients are at risk of a poor outcome.

1. Introduction

Lower urinary tract symptoms (LUTS) are categorised as storage, voiding, and postmicturition symptoms. More than three-quarters of men older than 40 yr experience at least one LUTS [1], which can compromise quality of life (QoL), employment, and social activities [2]. As the prevalence of LUTS increases with age, projected population ageing underlines the importance of effective management of male LUTS [3]. Current international guidelines recommend offering surgery if conservative treatments have been unsuccessful and LUTS are severe [4–6]. Transurethral prostate surgery for male LUTS associated with bladder outflow obstruction (BOO) is the standard treatment, and assessment of male LUTS for which interventional treatment is being considered needs to identify risk factors that predict potential adverse symptom outcomes.

UPSTREAM (Urodynamics for Prostate Surgery Trial; Randomised Evaluation of Assessment Methods) is a UK National Institute for Health Research–funded multicentre pragmatic randomised controlled trial (ISRCTN56164274) looking at the diagnostic pathway for men being considered for prostate surgery to treat LUTS [7–11]. In a previous paper, we reported on a multivariable model developed using baseline parameters to predict symptom improvement, and identified the variables that modified the effect of surgery on change in the International Prostate Symptom Score (IPSS) as age, number of comorbidities, maximum flow rate (Q_{max}), and symptom scores, particularly voiding symptoms and quality of life (QoL) [12].

The bladder diary (BD) is an integral part of structured LUTS assessment [10,12] and can potentially contribute information on factors relevant to initial management [13]. BDs capture frequency, voided volume (VV), and other factors, such as the BD Sensation Scale, for interpretation of storage LUTS and overactive bladder syndrome [14,15]. In the current study we developed novel prediction models for surgical outcomes, including factors prognostic for poor outcomes (either lack of improvement in symptoms, or symptom deterioration). Another aim was to evaluate the predictive role of BD in planning interventional treatment for voiding LUTS and whether BDs are essential given their unreliable completion in practice [16]. Models that can be used to formulate recommendations for patients lacking full BD information are needed. The overall findings were validated in an external "real-life" Japanese population.

2. Patients and methods

BD, International Prostate Symptom Score (IPSS), IPSS-QoL, and uroflowmetry (Q_{max}, VV, postvoid residual volume [PVR]) data were obtained from baseline information for the UPSTREAM trial [11]. In brief, 820 men seeking treatment for bothersome LUTS underwent standard LUTS assessment as detailed in the applicable UK National Institute for Health and Care Excellence (NICE) guideline (summarised in [17]) and were randomised to either undergo or omit additional invasive urodynamics (UDS) assessment. They then chose their treatment, as advised according to clinical recommendations on the basis of the evaluation results. The primary outcome was the IPSS 18 mo after randomisation. The inclusion criterion was men ($\geq$18 yr) seeking treatment for bothersome LUTS. The exclusion criteria were as follows: inability to pass urine without a catheter; relevant neurological disease; active treatment for LUTS, on active surveillance for prostate or bladder cancer; previous prostate surgery; not medically fit; and inability to complete outcome assessments. The trial was approved by the Oxford B Research Ethics Committee (reference 14/SC/0237). A detailed description of the study population, interventions, and outcomes has previously been published [18].

The IPSS is a patient-reported outcome (PRO) with seven separate items scored from 0 to 5 (overall range 0–35), with higher scores indicating more severe symptoms [19]. A validated Japanese translation of the instrument is available [20]. Symptom improvement was defined as a total IPSS reduction of at least 3 points from baseline to 18 mo after randomisation [21]. The International Consultation on Incontinence Questionnaire (ICIQ)-BD [14] is a 3-d BD, specifically referring to three consecutive 24-h periods. Six BD outputs were evaluated, including the average, maximum, and minimum VV per micturition, average 24-h urinary frequency, and the BD Sensation Scale. All parameters were averaged for the 3-d period. The BD Sensation Scale is a 5-point scale described in detail on its front page, with an abbreviated version on the back page of the ICIQ-BD. It includes two scores for normal voids, and three for voids with urgency or urgency incontinence (abbreviated scale: 0 = did not need to go, went just in case; 1 = normal desire to pass urine; 2 = had urgency but it passed away; 3 = had urgency but got to the toilet before leaking; 4 = had urgency and leaked). By calculated a modified Total Urgency and Frequency Score (TUFS) derived from the Patient Perception of Intensity of Urgency Scale [22] by adding the BD Sensation Scale score for each void recorded in a patient's diary and dividing this by the number of days completed.

External validation was undertaken using data from a single-centre routine-practice retrospective outcome database for prostate operations for LUTS at a Japanese urology department. A total of 264 patients underwent prostate surgery between 2016 and 2021, yielding 227 procedures with complete clinical data used for validation. Data for age, total IPSS,

Please cite this article as: H. Ito, K. Sakamaki, G.J. Young 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), https://doi.org/10.1016/j.euf.2023.06.013

IPSS-QoL, and uroflowmetry were extracted from the database. The data set was approved by the institutional review board of Yokosuka Kyosai Hospital (reference YKH20-74).

Three prognostic models (multiple logistic regression, classification tree analysis, and an artificial intelligence [AI]-based random forest) were derived to predict surgical outcomes using the UPSTREAM data set. The models were developed using age, total IPSS, IPSS-QoL, and uroflowmetry data, with and without BD data. To summarise the crude relationship, univariable logistic regressions were initially performed. In multiple logistic regression, variables were selected via the backward elimination method on the basis of *p* values. The area under the receiver operating characteristic curve (AUC) was used to evaluate prediction ability. Internal validation to evaluate overfitting and optimism was performed via the bootstrap method [23] in which 1000 bootstrap samples were repeatedly generated from an original data set. The random forest model is a supervised machine learning algorithm that grows multiple classification trees in which variables are evaluated for more accurate prediction according to the mean decrease in Gini coefficient value (a greater mean decrease indicates stronger predictive potential) [24]. Classification tree analysis is a model that allows selection of predominant factors and determination of significant thresholds for those parameters [25].

Observations in the UPSTREAM trial and the external validation data were excluded from analysis if results were missing for outcomes or any covariates. All analysis was performed in R v4.1.2 using the *rpart*, *randomForest*, and *pROC* packages.

3. Results

Of 820 participants treated at 26 sites in the UPSTREAM study, 291 received prostate surgery (81% transurethral resection of the prostate or bladder neck incision, 11% laser, 7% Urolift). Of these 291, 176 (60.5%) had complete BD, IPSS, IPSS-QoL, and uroflowmetry data available and were included in the models (baseline values summarised in Supplementary Table 1). Univariable logistic regression analysis showed that the strongest predictive parameter was total IPSS score, followed by age, Q_{max}, and IPSS-QoL (Table 1). The multiple logistic regression model with BD data consisted of age (p = 0.020), total IPSS (p = 0.0007), Q_{max} (p = 0.071), and modified TUFS (p = 0.089), for which the AUC was 79.6% (Table 2 and Fig. 1). The multiple logistic regression model without BD data consisted of age (p = 0.029), total IPSS (p = 0.0016), and Q_{max} (p = 0.066) and yielded an AUC of 77.1%.

The classification tree analysis model with BD data consisted of IPSS and Q_{max}, which yielded an AUC of 75.0% (Fig. 1B). Threshold values for IPSS and Q_{max} were 16 points and 13 ml/s, respectively (Fig. 2). Notably, the same analysis without BD data showed equivalent results. In the model with BD data, the top parameter was total IPSS, followed by Q_{max}, the minimum and average VV per micturition (derived from uroflowmetry), the modified TUFS, and the BD Sensation Scale (derived from the BD; Table 3). In the model without BD data, the top parameter was total IPSS, followed by Q_{max}, age, and the VV derived from uroflowmetry. A model including all the clinical parameters had an AUC of 97.1% with BD data and 94.7% without BD data (Fig. 1C). Internal validation using the bootstrap method for models with and without BD data yielded AUC values of 0.763 and 0.750 for the logistic regression model, 0.697 and 0.696 for the classification tree analysis model, and 0.873 and 0.858, respectively, for the random forest model (Table 4).

A total of 264 operations to relieve BOO were undertaken for male LUTS in Yokosuka Kyosai Hospital between 2016 and 2021. After excluding cases with missing data, clinical data from 227 operations were used for external validation (Supplementary Table 2). External validation showed that AUC values for the logistic regression, classification tree analysis, and random forest models (all without BD data) were similar to or lower than in the UPSTREAM data set (70.9%, 67.3%, and 68.5%, respectively; Fig. 3).

Table 1 – Univariable logistic regression model

Parameter	OR (95% CI)	*p* value	AUC
Age	0.923 (0.860–0.986)	0.020	0.664
Total IPSS	1.143 (1.063–1.238)	0.0005	0.710
IPSS-QoL	1.320 (0.889–1.957)	0.164	0.589
Uroflowmetry			
Q_{max}	0.910 (0.836–0.986)	0.021	0.663
PVR	1.001 (0.997–1.005)	0.704	0.488
Voided volume	0.998 (0.994–1.001)	0.188	0.584
Bladder diary			
Average VV per micturition	0.996 (0.990–1.003)	0.311	0.448
Modified TUFS	0.999 (0.955–1.048)	0.957	0.528
Bladder Diary Sensation Scale	0.903 (0.508–1.617)	0.729	0.518
24-h frequency	1.086 (0.932–1.280)	0.309	0.561
Maximum VV per micturition	0.999 (0.995–1.003)	0.535	0.474
Minimum VV per micturition	0.994 (0.985–1.003)	0.163	0.567

AUC = area under the receiver operating characteristic curve; CI = confidence interval; IPSS = International Prostate Symptom Score; OR = odds ratio; PVR = postvoid residual volume; Q_{max} = maximum flow rate; QoL = quality of life; TUFS = Total Urgency Frequency Score; VV = voided volume.

Table 2 – Multivariable logistic regression model

Parameter	OR (95% CI)	*p* value	AUC
Model with bladder diary			0.796
Age	0.918 (0.851–0.984)	0.020	
Total IPSS score	1.167 (1.072–1.285)	0.0007	
Q_{max}	0.922 (0.839–1.008)	0.071	
Modified TUFS	0.952 (0.899–1.008)	0.089	
Model without bladder diary			0.771
Age	0.924 (0.858–0.989)	0.029	
Total IPSS score	1.133 (1.051–1.230)	0.0016	
Q_{max}	0.920 (0.837–1.005)	0.066	

AUC = area under the receiver operating characteristic curve; CI = confidence interval; IPSS = International Prostate Symptom Score; OR = odds ratio; Q_{max} = maximum flow rate; TUFS = Total Urgency Frequency Score.

4. Discussion

The current study developed three different statistical models to predict symptom outcomes of prostate surgery in a high-quality clinical trial, and explored their performance in a real-life setting using an independent data set.

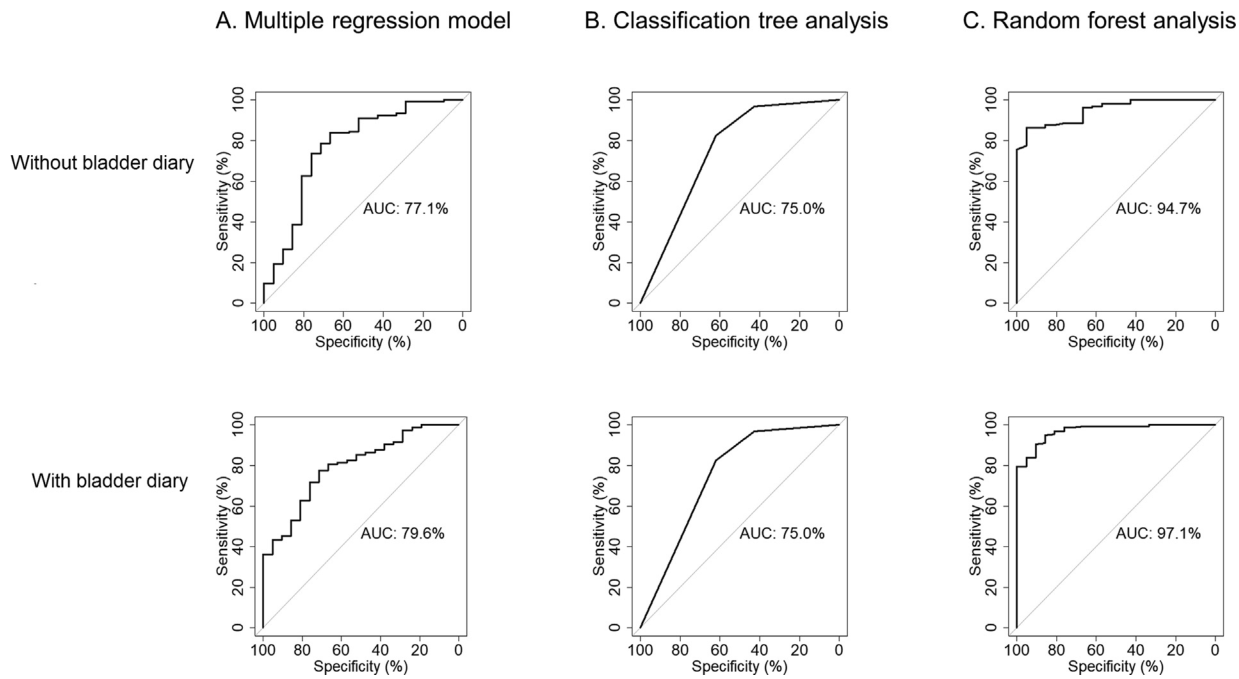

Fig. 1 – Receiver operating characteristic (ROC) curves for the three models generated with age, preoperative IPSS, IPSS-QoL, and uroflowmetry, with or without bladder diary data. AUC = area under the ROC curve; IPSS = International Prostate Symptom Score; QoL = quality of life.

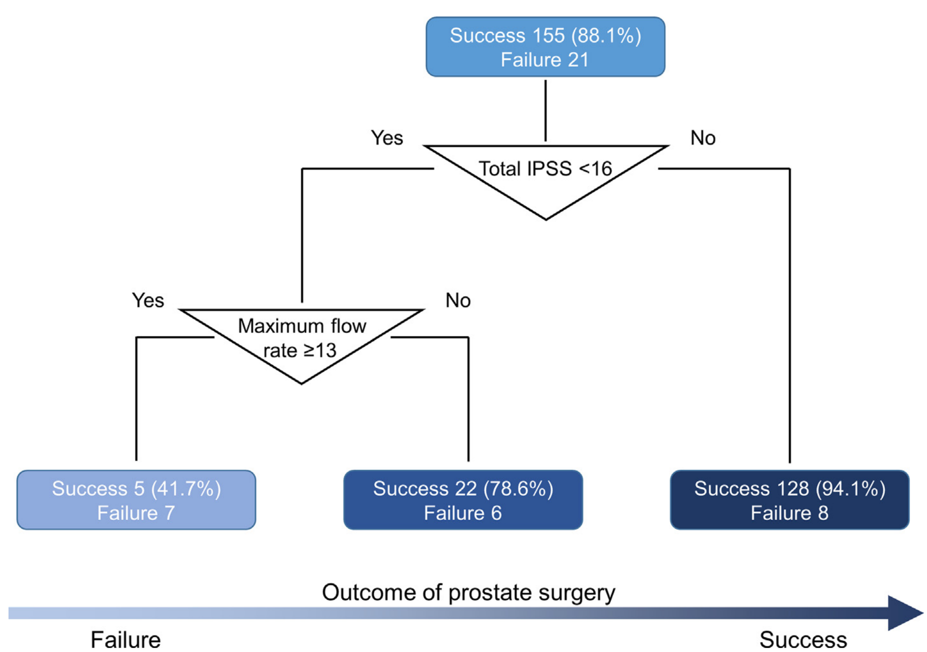

Fig. 2 – Classification tree analysis. IPSS = International Prostate Symptom Score.

One of the key findings is that LUTS outcomes may deteriorate after surgery if baseline Q_{max} is $\geq$13 ml/s. Our previous study established that Q_{max} of $\leq$9.8 ml/s at baseline is a favourable prostate factor for prostate surgery to treat LUTS [12]. That study did not establish a threshold for recommending UDS but concluded that UDS may provide useful additional information when Q_{max} is >10 ml/s, and especially >15 ml/s, with the bladder outflow obstruction index (BOOI) and bladder contractility index (BCI) predicting the change in IPSS most successfully. Hence, the current study is important, as 13 ml/s now provides a clear Q_{max} threshold above which UDS should be recommended to reduce the risk of experiencing symptom deterioration. These conclusions align with observations from the International Continence Society study, which reported that mean Q_{max} was 9.7 ml/s for men with BOO and 12.6 ml/s for men without BOO [26]. That study also revealed that among patients with BOO, 53% had Q_{max} >10 ml/s and 18% had Q_{max}

Table 3 – Random forest model

Parameter	Mean decrease in Gini coefficient	AUC
Model with bladder diary		0.971
Age	1.270	
Total IPSS score	3.755	
IPSS-QoL	0.281	
Uroflowmetry		
Q_{max}	2.233	
Postvoid residual volume	0.291	
Voided volume	0.904	
Bladder diary		
Average VV per micturition	1.521	
Modified TUFS	1.423	
Bladder Diary Sensation Scale	1.406	
24-h frequency	0.350	
Maximum VV per micturition	0.395	
Minimum VV per micturition	2.112	
Model without bladder diary		0.947
Age	2.376	
Total IPSS score	6.293	
IPSS-QoL	0.555	
Uroflowmetry		
Qmax	3.844	
Postvoid residual volume	0.985	
Voided volume	2.057	

AUC = area under the receiver operating characteristic curve; IPSS = International Prostate Symptom Score; Q_{max} = maximum flow rate; QoL = quality of life; TUFS = Total Urgency Frequency Score; VV = voided volume.

Table 4 – AUC results for internal validation via the bootstrap method for the three models with and without BD data

Model		Actual data	Bootstrap estimate (95% CI)
Multivariable logistic regression	With BD	0.796	0.763 (0.692–0.812)
	Without BD	0.771	0.750 (0.673–0.782)
Classification tree analysis	With BD	0.750	0.697 (0.500–0.852)
	Without BD	0.750	0.696 (0.500–0.827)
Random forest	With BD	0.973	0.873 (0.794–0.941)
	Without BD	0.947	0.858 (0.781–0.927)

AUC = area under the receiver operating characteristic curve; BD = bladder diary; CI = confidence interval.

>15 ml/s. Accordingly, UDS for men with Q_{max} between 10 and 13 ml/s may identify BOO and hence should be discussed with patients to explore their individual preference [27].

BD data did not yield essential information for prediction of symptom outcome from interventional prostate treatment. One of the three models (multivariable logistic regression) suggested that BD data might improve the accuracy. Internal validation for two of the models (multivariable logistic regression and random forest) also showed that BD assessment might increase predictivity. Six parameters were derived from the BD; the modified TUFS and the BD sensation scale appeared to offer some benefit in prediction of prostate surgery outcomes, and it is possible that parameters not evaluated might also contribute. Nonetheless, models were also developed without BD data in order to deal with the known difficulty of full BD completion in clinical practice [16], and these were also effective. Overall, effective decision-making with regard to voiding LUTS can be based on other baseline measurements without a BD being essential. Such a conclusion has no bearing on the key role of BDs in assessing storage LUTS and nocturia, as recommended in guidelines [4,28].

The current study developed independent models in order to add BD information and proceed to external validation. Nonetheless, identification of IPSS <16 and IPSS QoL ≤4 as predictive of poor symptom outcomes in the current study corresponds to the identification of IPSS >16 and IPSS QoL >4 as predictive of good outcomes in the previous study [12].

Prognostic factors in baseline assessments enable proper counselling when considering surgery, since interventional treatment should only be pursued if there is a realistic prospect of improving symptoms. This is important for consent, for avoiding unnecessary surgery, and also reducing the risk of LUTS deteriorating following surgery, which is a clear possibility [10,18]. Until recently, pathways relied principally on expert consensus and provided comparatively non-specific guidance that surgery is usually needed when patients gain insufficient relief of LUTS or PVR after conservative or pharmacological treatments (relative operation indications). Identification of prognostic features in the current study, along with the detailed quantitative and qualitative data from UPSTREAM [10,12,18,29,30], makes it possible to use baseline assessments to advise individual patients on the likely outcomes of surgery for voiding LUTS (Fig. 4). Favourable characteristics (illustrated in green in Fig. 4) include clinical factors identified in the symptom scores and flow rate tests, in particular voiding LUTS with both high symptom severity and bother, and Q_{max} <10 ml/s. These predict good symptom outcomes, provided the intervention achieves effective relief of BOO. UDS is not needed when all these factors are present, which is compatible with the conclusion from the UPSTREAM study that UDS should not be used routinely in male LUTS assessment [10]. For cases that do not have all of the favourable clinical characteristics, UDS identifies men who should experience an improvement in symptoms after surgery on the basis of characteristic low flow/high detrusor pressure even though they did not have all the favourable clinical characteristics, based on the severity of the outflow obstruction (BOOI ≥48) and adequate bladder contractility (BCI ≥123) [12]. If neither clinical nor UDS characteristics are favourable, men need explicit counselling about the possibility of a worse, or at best a neutral outcome (minimal symptom change), with the risk of symptom deterioration. Worsening LUTS is particularly relevant for certain features that can be considered "unfavourable" (illustrated in red in Fig. 4).

The ICIQ MLUTS voiding subscore was the best predictor in our previous study [12] and was thus included in Figure 4. The ICIQ MLUTS was not included in the models developed in the current study as it was not administered to the external validation population, which is a limitation of the study. In clinical practice, direct questioning is needed if IPSS is

Please cite this article as: H. Ito, K. Sakamaki, G.J. Young 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), https://doi.org/10.1016/j.euf.2023.06.013

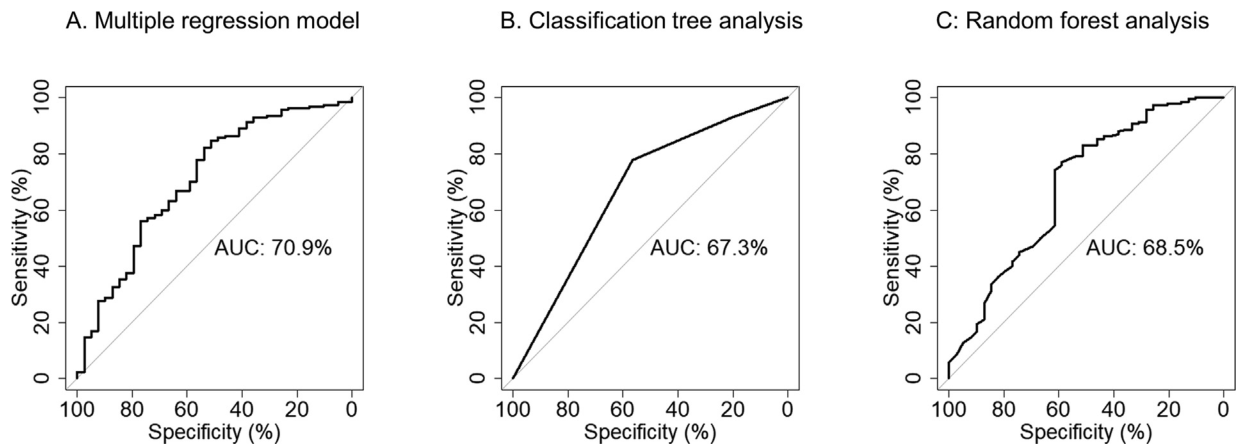

Fig. 3 – Receiver operating characteristic (ROC) curves from external validation of the three models generated with age, preoperative IPSS, IPSS-QoL, and uroflowmetry without bladder diary data. AUC = area under the ROC curve; IPSS = International Prostate Symptom Score; QoL = quality of life.

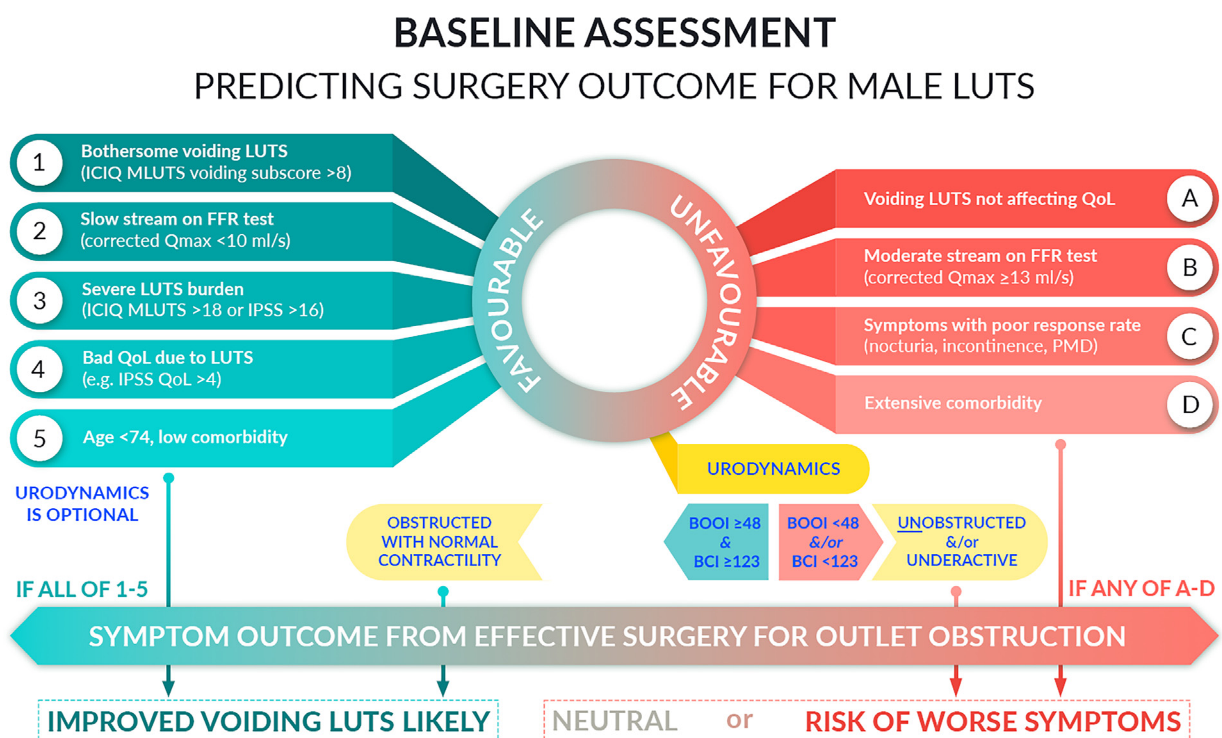

Fig. 4 – Key baseline clinical parameters for prediction of outcomes from surgery to treat BOO in male LUTS, incorporating overall findings from the UPSTREAM study. Two groups of men have a good chance of voiding symptom improvement from surgery, provided that BOO is properly relieved ("effective surgery"): (1) those with all of the favourable predictive factors (shown in green on the left); and (2) those for whom a urodynamics test finds BOOI ≥48 and BCI ≥123. Without these features, the outcome may be neutral (symptom score change below the minimally important difference) or even symptom deterioration, particularly in cases with any of the unfavourable factors shown in red. BCI = bladder contractility index; BOO = bladder outflow obstruction; BOOI = BOO Index; FFR = free flow rate; ICIQ MLUTS = International Consultation on Incontinence Questionnaire on male LUTS; IPSS = International Prostate Symptom Score; LUTS = lower urinary tract symptoms; PMD = postmicturition dribble; Q_{max} = maximum flow rate; QoL = quality of life.

used as the PRO, to ascertain how severe and how bothersome voiding LUTS are for an individual patient. Likewise, specific questioning is required for poorly responding symptoms (incontinence, postmicturition dribble) as they are not captured in the IPSS. Overall, the predictive importance of the voiding subscore and comprehensive assessment of LUTS [9] suggest key benefits for clinical practice and support adoption of the ICIQ MLUTS as a standard of care.

AI, machine learning, and deep learning techniques show outstanding potential performance [31]. Our random forest model based on machine learning techniques showed remarkably better accuracy than the other models, suggesting that the model is effective. External validation, however, indicated a moderate fit to real-world data in a different setting, which may reflect the complexity of QoL, the subjective nature of outcomes, and the influence of nonclinical factors such as cultural differences. Further limitations of the study include a reliance on comprehensive baseline assessments (ie, availability of fully completed BDs, PROs, and uroflowmetry for analysis), which excluded a significant proportion of UPSTREAM participants, resulting in a comparatively small overall sample and subgroup sizes. In the external validation population, the timing of outcome evaluation (3 mo after surgery) differed from UPSTREAM, in which outcome assessment was 18 mo after randomisation (noting that the time between randomisation, assessment, and surgery varied between centres, as previously detailed [18]). However, on the basis of the results reported here, we believe that external validation using such a different cohort points to the consistency and reliability of the modelling. In line with NICE guidelines [17], prostate volume was not routinely captured in the study databases.

5. Conclusions

Three models using preoperative parameters including age, total IPSS, IPSS-QoL, and uroflowmetry predicted outcomes of prostate surgery. Versions including BD data only slightly improved the accuracy in terms of the AUC. $Q_{max} \geq 13$ ml/s, IPSS <16, and IPSS-QoL ≤ 4 were each associated with high risk of a poor surgical outcome. The AI-based model showed remarkably high accuracy in the research setting, but only moderate accuracy in a routine practice context.

Author contributions: Marcus J. Drake had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Ito, Sakamaki, Young, Kobayashi, Drake.
Acquisition of data: Ito, Sakamaki, Hashim, Drake.
Analysis and interpretation of data: Ito, Sakamaki, Young, Blair, Lane, Kobayashi, Malde, Drake.
Drafting of the manuscript: Ito, Lane, Drake.
Critical revision of the manuscript for important intellectual content: Hashim, Malde, Abrams, Chapple.
Statistical analysis: Ito, Sakamaki.
Obtaining funding: Blair, Hashim, Lane, Abrams, Chapple, Drake.
Administrative, technical, or material support: Clout.
Supervision: None.
Other: None.

Financial disclosures: Marcus J. Drake certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Peter S. Blair, J. Athene Lane, Paul Abrams, Christopher Chapple and Marcus J. Drake were co-applicants for the UPSTREAM trial and thus received grant funding. J. Athene Lane was a member of the Clinical Trials Unit funded by the National Institute for Health Research during the conduct of this trial. Grace J. Young was funded by the UPSTREAM trial. Hashim Hashim has been a speaker for Medtronic. Christopher Chapple has been a consultant for Muvon, Pro-verum, Symimetic, and Urovant, and an author for Astellas Pharma and Ferring. Marcus J. Drake has received personal fees from Astellas Pharma. The remaining authors have nothing to disclose.

Funding/Support and role of the sponsor: The UPSTREAM project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme (project number 12/140/01). The views and opinions expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This study was designed and delivered in collaboration with the Bristol Trials Centre, a UKCRC-registered clinical trials unit (CTU) in receipt of NIHR CTU support funding. The UPSTREAM trial was sponsored by North Bristol NHS Trust (Bristol, UK). Study data were collected and managed using REDCap [32], hosted at the University of Bristol. The sponsors played a role in the design and conduct of the study and in data collection and management.

Acknowledgements: We thank all the men who participated in the UPSTREAM trial. We also thank the participating NHS hospitals, in particular the principal investigators, research nurses, clinical trial assistants, administrators, and all other support staff, for their dedication and hard work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.euf.2023.06.013.

References

[1] Irwin DE, Milsom I, Kopp Z, Abrams P, Artibani W, Herschorn S. Prevalence, severity, and symptom bother of lower urinary tract symptoms among men in the EPIC study: impact of overactive bladder. Eur Urol 2009;56:14–20.
[2] Kim TH, Han DH, Ryu DS, Lee KS. The impact of lower urinary tract symptoms on quality of life, work productivity, depressive symptoms, and sexuality in Korean men aged 40 years and older: a population-based survey. Int Neurourol J 2015;19:120–9.
[3] Milsom I, Irwin DE. A cross-sectional, population-based, multinational study of the prevalence of overactive bladder and lower urinary tract symptoms: results from the EPIC study. Eur Urol Suppl 2007;6:4–9.
[4] Gratzke C, Bachmann A, Descazeaud A, et al. EAU guidelines on the assessment of non-neurogenic male lower urinary tract symptoms including benign prostatic obstruction. Eur Urol 2015;67:1099–109.
[5] Lerner LB, McVary KT, Barry MJ, et al. Management of lower urinary tract symptoms attributed to benign prostatic hyperplasia: AUA guideline part I—initial work-up and medical management. J Urol 2021;206:806–17.
[6] Lerner LB, McVary KT, Barry MJ, et al. Management of lower urinary tract symptoms attributed to benign prostatic hyperplasia: AUA guideline part II—surgical evaluation and treatment. J Urol 2021;206:818–26.
[7] Bailey K, Abrams P, Blair PS, et al. Urodynamics for Prostate Surgery Trial; Randomised Evaluation of Assessment Methods (UPSTREAM) for diagnosis and management of bladder outlet obstruction in men: study protocol for a randomised controlled trial. Trials 2015;16:567.
[8] Young GJ, Lewis AL, Lane JA, Winton HL, Drake MJ, Blair PS. Statistical analysis plan for the Urodynamics for Prostate Surgery Trial; Randomised Evaluation of Assessment Methods (UPSTREAM). Trials 2017;18:455.
[9] Ito H, Young GJ, Lewis AL, et al. Grading severity and bother using the International Prostate Symptom Score and International Consultation on Incontinence Questionnaire male lower urinary tract symptoms score in men seeking lower urinary tract symptoms therapy. J Urol 2020;204:1003–11.
[10] 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;78:701–10.
[11] Lewis AL, Young GJ, Abrams P, et al. Clinical and patient-reported outcome measures in men referred for consideration of surgery to treat lower urinary tract symptoms: baseline results and diagnostic

findings of the Urodynamics for Prostate Surgery Trial; Randomised Evaluation of Assessment Methods (UPSTREAM). Eur Urol Focus 2019;5:340–50.

[12] Young GJ, Metcalfe C, Lane JA, et al. Prostate surgery for men with lower urinary tract symptoms: do we need urodynamics to find the right candidates? Exploratory findings from the UPSTREAM trial. Eur Urol Focus 2022;8:1331–9.

[13] Bright E, Drake MJ, Abrams P. Urinary diaries: evidence for the development and validation of diary content, format, and duration. Neurourol Urodyn 2011;30:348–52.

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

[15] Cameron AP, Wiseman JB, Smith AR, et al. Are three-day voiding diaries feasible and reliable? Results from the Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN) cohort. Neurourol Urodyn 2019;38:2185–93.

[16] Ito H, Abrams P, Lewis AL, et al. Use of the International Consultation on Incontinence Questionnaires bladder diary in men seeking therapy for lower urinary tract symptoms. Eur Urol Focus 2022;8:66–74.

[17] Jones C, Hill J, Chapple C. Management of lower urinary tract symptoms in men: summary of NICE guidance. BMJ 2010;340:c2354.

[18] Lewis AL, Young GJ, Selman LE, et al. Urodynamics tests for the diagnosis and management of bladder outlet obstruction in men: the UPSTREAM non-inferiority RCT. Health Technol Assess 2020;24:1–122.

[19] Barry MJ, Fowler Jr FJ, O'Leary MP, et al. The American Urological Association symptom index for benign prostatic hyperplasia. J Urol 1992;148:1549–57.

[20] Homma Y, Tsukamoto T, Yasuda K, Ozono S, Yoshida M. Shinji MLinguistic validation of Japanese version of International Prostate Symptom Score and BPH impact index. Jpn J Urol 2002;93:669–80.

[21] Barry MJ, Cantor A, Roehrborn CG. Relationships among participant international prostate symptom score, benign prostatic hyperplasia impact index changes and global ratings of change in a trial of phytotherapy in men with lower urinary tract symptoms. J Urol 2013;189:987–92.

[22] Chapple CR, Drake MJ, Van Kerrebroeck P, et al. Total urgency and frequency score as a measure of urgency and frequency in overactive bladder and storage lower urinary tract symptoms. BJU Int 2014;113:696–703.

[23] Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W.

[24] Han H, Guo X, Yu H. Variable selection using mean decrease accuracy and mean decrease Gini based on random forest. In: Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science, 2016. https://doi.org/10.1109/ICSESS.2016.7883053.

[25] Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. Boca Raton, FL: Chapman & Hall/CRC; 1984.

[26] Reynard JM, Shearer RJ. Failure to void after transurethral resection of the prostate and mode of presentation. Urology 1999;53:336–9.

[27] 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;38:320–9.

[28] Smith M, Dawson S, Andrews RC, et al. Evaluation and treatment in urology for nocturia caused by nonurological mechanisms: guidance from the PLANET study. Eur Urol Focus 2022;8:89–97.

[29] Aiello M, Jelski J, Lewis A, et al. Quality control of uroflowmetry and urodynamic data from two large multicenter studies of male lower urinary tract symptoms. Neurourol Urodyn 2020;39:1170–7.

[30] Selman LE, Clement C, Ochieng CA, et al. Treatment decision-making among men with lower urinary tract symptoms: a qualitative study of men's experiences with recommendations for patient-centred practice. Neurourol Urodyn 2021;40:201–10.

[31] Huang JC, Tsai YC, Wu PY, et al. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. Comput Methods Programs Biomed 2020;195:105536.

[32] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–81.