NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 29 • Issue 1 • January 2025

ISSN 2046-4924

Development of a clinical decision support tool for Primary care Management of lower Urinary tract Symptoms in men: the PriMUS study

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

DOI 10.3310/RGTW5711

NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Development of a clinical decision support tool for Primary care Management of lower Urinary tract Symptoms in men: the PriMUS study

Adrian Edwards[1*] Ridhi Agarwal[2] Janine Bates[3] Alison Bray[4,5] Sarah Milosevic[3] Emma Thomas-Jones[3] Michael Drinnan[4,5] Marcus Drake[6] Peter Michell,[7] Bethan Pell[3] Haroon Ahmed[1] Natalie Joseph-Williams[1] Kerenza Hood[3] Yemisi Takwoingi[2] and Chris Harding[5,8]

[1]Division of Population Medicine, Cardiff University, Cardiff, UK
[2]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[3]Centre for Trials Research, Cardiff University, Cardiff, UK
[4]Northern Medical Physics and Clinical Engineering, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
[5]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK
[6]Faculty of Medicine, Department of Surgery & Cancer, Imperial College London, London, UK
[7]Patient Representative
[8]Department of Urology, Freeman Hospital, Newcastle upon Tyne Hospitals, Newcastle upon Tyne, UK

[*]Corresponding author EdwardsAG@cardiff.ac.uk

Published January 2025
DOI: 10.3310/RGTW5711

This report should be referenced as follows:

Edwards A, Agarwal R, Bates J, Bray A, Milosevic S, Thomas-Jones E, *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**(1). https://doi.org/10.3310/RGTW5711

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 3.5

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 3.5 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal
Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was funded by the HTA programme as award number 15/40/05. The contractual start date was in May 2017. The draft manuscript began editorial review in November 2023 and was accepted for publication in July 2024. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

Background: Lower urinary tract symptoms particularly affect older men and their quality of life. General practitioners currently have no easily available assessment tools to diagnose lower urinary tract symptom causes. Referrals to urology specialists are increasing. General practitioner access to simple, accurate tests and clinical decision tools could facilitate management of lower urinary tract symptoms in primary care.

Objectives: To determine which of several index tests in combination, best predicted three diagnoses (detrusor overactivity, bladder outlet obstruction and/or detrusor underactivity) in men presenting with lower urinary tract symptoms in primary care. To develop and validate three diagnostic prediction models, and a prototype primary care clinical decision support tool.

Design: Prospective diagnostic accuracy study. Two participant cohorts, for *development* and *validation*, underwent simple index tests and a reference standard (invasive urodynamics).

Setting: General practices in England and Wales.

Participants: Men (16 years and over) consulting their general practitioner with lower urinary tract symptoms.

Sample size: Separate calculations for model development and validation cohorts, from literature estimates of detrusor overactivity, bladder outlet obstruction and detrusor underactivity prevalences of 57%, 31% and 16%, respectively.

Predictors and index tests: Twelve potential predictors considered for three diagnostic models.

Main outcome measures: The primary outcome was diagnostic model sensitivity and specificity for detecting bladder outlet obstruction, detrusor underactivity and detrusor overactivity, with 75.0% considered minimum clinically useful performance.

Statistical analysis: Three separate logistic regression models generated with index test variables to predict the presence of bladder outlet obstruction, detrusor overactivity, detrusor underactivity conditions in men with lower urinary tract symptoms.

Results: One model each was developed and validated for bladder outlet obstruction and detrusor underactivity, two for detrusor overactivity (detrusor overactivity main, detrusor overactivity sensitivity analysis 2). Age, voiding symptoms subscore, prostate-specific antigen level, median maximum flow rate, median voided volume were predictors for bladder outlet obstruction. Median maximum flow rate and post-void residual volume were predictors for detrusor underactivity. Age, post-void residual volume and median voided volume were included in detrusor overactivity main model, while age and storage symptoms subscore predicted detrusor overactivity sensitivity analysis 2.

For all four models, sensitivity of 75.0% could be achieved with a specificity of 74.2%, 47.3%, 45.6% and 46.2% for bladder outlet obstruction, detrusor underactivity, detrusor overactivity main and detrusor overactivity sensitivity analysis 2 models, respectively. Similarly, a specificity of 75.0% could be achieved with a sensitivity of 71.3%, 39.8%, 33.3% and 62.7% for bladder outlet obstruction, detrusor underactivity, detrusor overactivity main and detrusor overactivity sensitivity analysis 2 models, respectively.

The prototype tool (not yet intended for use in practice) is available at Primary care Management of lower Urinary tract Symptoms decision aid for lower urinary tract symptoms (shinyapps.io).

General practitioner feedback during tool development and small-scale user-testing in simulated consultation scenarios was favourable. Patients supported such management in primary care.

Strengths/limitations: This was a prospective, multicentre study in an appropriate primary care population. Most of the index tests are possible routinely in primary care or at home by patients. The diagnostic models were validated in a separate cohort from the same population.

Limitations include that target condition prevalences may differ in other populations.

Conclusion: We identified sensitivities and specificities of diagnostic models for detrusor overactivity, bladder outlet obstruction and detrusor underactivity in routine United Kingdom practice and developed a prototype clinical decision support tool.

Future work: Economic modelling, a feasibility trial and powered randomised controlled trial are needed to evaluate the Primary care Management of lower Urinary tract Symptoms tool in practice.

Study registration: Current Controlled Trials ISRCTN10327305.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 15/40/05) and is published in full in *Health Technology Assessment*; Vol. 29, No. 1. See the NIHR Funding and Awards website for further award information.

Contents

List of tables

List of figures

List of boxes

List of abbreviations

BCI	bladder contractility index
BOO	bladder outlet obstruction
BOOI	bladder outlet obstruction index
CPRD	Clinical Practice Research Datalink
CRFs	case report forms
CRN	Clinical Research Network
DO	detrusor overactivity
DRE	digital rectal examination
DU	detrusor underactivity
GP	general practitioner
HCA	healthcare assistant
HCRW	Health and Care Research Wales
HTA	Health Technology Assessment funding programme
ICIQ	International Consultation on Incontinence Questionnaire
IPSS	International Prostate Symptom Score
ISC	intermittent self-catheterisation
LUTS	lower urinary tract symptom(s)
MICE	multiple imputation by chained equations
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research
NuTH	The Newcastle upon Tyne Hospitals NHS Foundation Trust
PPI	patient and public involvement
Prototype tool	PriMUS prototype online decision support tool
PSA	prostate-specific antigen (blood test)
RCT	randomised controlled trial
ROC	receiver operating characteristic
SAE	serious adverse event
SMG	Study Management Group
SSC	Study Steering Committee
SSRI	selective serotonin reuptake inhibitor
TURP	transurethral resection of the prostate
UTI	urinary tract infection
Voice Global	Online PPI platform

Plain language summary

Urinary symptoms such as a weak flow and frequent urination are common in older men and often bothersome. Men visiting their general practitioner with these symptoms are often referred to a specialist because good diagnostic tools are not available in primary care. Three common causes of symptoms are: bladder obstruction due to non-cancerous growth of the prostate, reduced power of the bladder muscle and bladder overactivity.

We aimed to create a tool to help general practitioners manage men with urinary symptoms. This required first to develop mathematical models, which combined results from several simple tests that general practitioners could organise. The web-based tool then constructed would indicate the most likely diagnosis and provide recommendations for treating and managing the condition. The tests included prostate examination, prostate-specific antigen blood test, symptoms questionnaires and home-based urine flow measurements.

To develop the mathematical models, 350 men with urinary symptoms underwent the simple tests and a specialist invasive test called urodynamics, which is currently regarded as providing the best diagnosis. A second group of 251 men also had the simple tests and urodynamics. Their results were used to measure the performance of the models.

The model to diagnose bladder obstruction performed well (close to the invasive urodynamics 'gold standard' test), and those to diagnose reduced power of the bladder muscle and bladder over-activity performed moderately but less well. A prototype version of the web-based tool was developed. We consulted patients and general practitioners to assess the tool's acceptability. General practitioners confirmed their enthusiasm because they find managing bladder symptoms challenging, and patients said they would prefer to be managed in primary care. We received good feedback about the prototype tool and gained ideas for refining it.

Following this project, it would be valuable to estimate the cost, benefits and practicalities of implementing the tool, aided by data from the study, and trial its effectiveness compared with current care.

Scientific summary

This summary contains text reproduced with permission from Pell *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;10:e037634; Milosevic *et al*. Conducting invasive urodynamics in primary care: qualitative interview study examining experiences of patients and healthcare professionals. *Diagn Progn Res* 2021;5:10; and Milosevic *et al*. Managing lower urinary tract symptoms in primary care: qualitative study of GPs' and patients' experiences. *Br J Gen Pract* 2021;71:E685–92. These are Open Access articles distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) licence, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/. The text below includes minor additions and formatting changes to the original text.

Background

Lower urinary tract symptoms (LUTS) particularly affect older men and can lead to poor quality of life, often referred to as the degree of 'bother' experienced. General practitioners (GPs) currently have no easily available assessment tools to effectively diagnose the causes of LUTS and aid discussion of treatment with patients. Men are increasingly referred to urology specialists who often recommend treatments that could have been initiated in primary care. GP access to simple, accurate tests and clinical decision tools could facilitate faster and more effective patient management of LUTS in primary care.

The reference standard test for investigation of LUTS, and thus diagnosis of detrusor overactivity (DO), bladder outlet obstruction (BOO) and detrusor underactivity (DU), is invasive urodynamics, which takes place in secondary care. However, National Institute for Health and Care Excellence (NICE) guidelines suggest that many men referred to specialist care with LUTS are eventually managed conservatively, and so could have remained within primary care. Further, GPs do not have access to validated clinical decision tools giving an indication of the most likely cause of symptoms to guide treatment and management. Making such a tool available should improve treatment efficacy, standardise treatment, reduce unnecessary referrals, expedite referral of those requiring specialist care, and thus improve cost-effectiveness of NHS care.

A primary care-based clinical decision support tool with defined accuracy would, firstly, mean that the men could undergo the necessary simple tests straightaway, organised through the GP surgery and, secondly, would get a quicker result regarding predicted diagnosis and choice of management options that are most likely to be effective. For GPs, a clinical decision support tool could allay uncertainty around both diagnosis and best management.

Objectives

The primary objectives of the Primary care Management of lower Urinary tract Symptoms (PriMUS) study were to:

1. Develop statistical models to predict the likelihood of each urological condition (BOO, DU and DO), based on a series of non-invasive index tests, with urodynamics as the reference standard.
2. Estimate the diagnostic accuracy of the above models in an independent validation cohort.

The study incorporated an internal pilot phase with 'Stop/go' criteria that included quantitative and qualitative assessments. The progression criteria were designed to allow for mitigating strategies to be discussed to allow for some adaptation to recruitment processes in the main study.

The secondary objectives were to:

1. Develop a series of patient management recommendations and thresholds for clinically useful diagnostic prediction by expert consensus and with reference to current clinical guidelines that map to the diagnoses predicted by the statistical model.
2. Combine the statistical model and management recommendations into an online tool that will form the prototype clinical decision support tool.
3. Complete a qualitative study to explore the feasibility of introducing the clinical decision support tool into primary care, including potential acceptability to primary care staff and patients.
4. Collect NHS costs involved in delivering the new pathway and compare with cost of standard pathway calculated from NHS and other sources.

Objectives 1 and 2: development of statistical models and diagnostic accuracy

Methods

Men presenting to their GP with LUTS were recruited prospectively from GP practices in Bristol, Newcastle upon Tyne and Wales. Participants underwent a series of simple index tests and the invasive reference standard (urodynamics). To determine which index tests used in combination best predicted three urodynamic observations (BOO, DU and DO) in men, diagnostic prediction models were developed for each target condition using logistic regression modelling.

Multiple imputation by chain equations was used to handle missing data and fractional polynomial functions were used to fit continuous variables. The discriminative ability of the models was assessed using the c-index, and calibration was assessed using calibration plots and the calibration slope. Internal validation was conducted to assess optimism of the performance statistics using the bootstrapping procedure. External validation was conducted to assess model performance in another sample from a similar population as the development 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 receiver operating characteristic 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. Sensitivity analyses were performed by fitting two alternative models for each target condition. In sensitivity analysis 1, 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) and in sensitivity analysis 2 (SA2), alternative measures or methods of measurement of candidate predictors were considered.

Results

Between March 2018 and June 2022, 350 and 251 men were respectively recruited into the development and validation cohorts. In the development cohort (median age 69), 163 (46.6%), 141 (40.3%) and 253 (72.3%) participants were diagnosed with BOO, DU and DO, respectively. In the validation cohort (median age 67), 112 (44.6%), 87 (34.7%) and 166 (66.1%) participants were diagnosed with BOO, DU and DO, respectively. Two models were developed and validated for DO (DO main and DO SA2), while one model each was developed and validated for BOO (BOO model 3) and DU. Age (participant demographics), voiding symptoms subscore (International Consultation on Incontinence Questionnaire – male LUTS questionnaire), prostate-specific antigen (PSA) test result (blood test), median maximum flow rate (uroflowmetry) and median voided volume (uroflowmetry) were predictors for BOO. Median maximum flow rate and post-void residual volume (bladder ultrasound) were predictors for DU. Age, post-void residual volume and median voided volume were included in DO main model, while age and storage symptoms subscore (International Prostate Symptom Score questionnaire) were predictors in DO SA2 model.

Bladder outlet obstruction model 3 demonstrated good discriminative performance with an optimism-corrected c-index of 0.80. The models for DU, DO main and DO SA2 demonstrated moderate discriminative ability with an optimism-corrected c-index of respectively 0.64, 0.67 and 0.65. Similar estimates of c-index were observed for each model with the validation cohort. The optimism-corrected calibration slope for each model was < 1.00 (BOO model 3: 0.87; DU: 0.77; DO, Main: 0.78; DO SA2: 0.74), suggesting that the models were overfitted. Miscalibration was also observed with the validation cohort for DU {0.82 [95% confidence interval (CI) 0.31 to 1.32]}, DO main [0.72 (95% CI 0.27 to 1.17)] and DO SA2 [1.36 (95% CI 0.78 to 1.94)] models, whereas BOO model 3 demonstrated good calibration performance of 0.99 (95% CI 0.68 to 1.30).

For BOO model 3, a sensitivity of 75.1% could be achieved with a specificity of 74.2% approximately at a threshold of 50.9%. At a threshold of 53.3%, a specificity of 75.5% could be achieved with a sensitivity of 71.3% approximately. For DU model, a sensitivity of 75.3% could be achieved with a specificity of 47.3% approximately at a threshold of 34.2%. At a threshold of 41.4%, a specificity of 75.1% could be achieved with a sensitivity of 39.8% approximately. For DO model from the main analysis, a sensitivity of 75.1% could be achieved with a specificity of 45.6% approximately at a threshold of 63.8%. At a threshold of 75.2%, a specificity of 75.7% could be achieved with a sensitivity of 33.3% approximately. For DO model from SA2, a sensitivity of 75.3% could be achieved with a specificity of 46.2% approximately at a threshold of 63.1%. At a threshold of 71.4%, a specificity of 75.6% could be achieved with a sensitivity of 62.7% approximately.

Conclusions

The models for BOO, DU and DO are a combination of index tests that are simple to perform and less invasive than the reference standard. DO SA2 is the only model to use index tests that are in primary care. The remaining models include predictors that are not currently available or routinely used in primary care. The limited availability of the predictors in primary care could either result in missing data at the time of diagnosing or a delay in receiving the diagnosis for the target condition if the patient needs to be referred to receive an index test that is part of the model. The latter could imply that the models may be more useful for secondary care.

The validation cohort was recruited in a similar prospective manner to the development cohort, following the same inclusion/exclusion criteria, definitions to collect data on the predictors and outcome and recruiting participants from the same study sites. BOO model 3 continuously showed a high discriminative ability, whereas the DU, DO main and DO SA2 models showed moderate discriminative performance. There was large uncertainty around the c-index and calibration slope, likely driven by the smaller sample size of the validation cohort, less than originally targeted, owing to recruitment problems, principally due to the pandemic. Thus, further external validation and recalibration may be required before the models can be used in practice to ensure applicability of the models in settings where case-mix or prevalence could differ.

Objective 3: patient management recommendations

Urology specialists (target n = 15–20) were invited to take part in the process of developing management recommendations to inform the clinical decision support tool. Interviews and questionnaires were used to establish how the urologists would manage a number of different commonly encountered clinical scenarios, focusing on the thresholds at which they would recommend treatment and the strategies they would use when multiple urodynamic abnormalities are diagnosed or suggested. Scenarios were informed by real-life data generated by study participants.

Feedback from the interviews and questionnaire were collated, and Study Management Group members with a background in urology then considered this in conjunction with available evidence/guidelines to inform the development of the draft management recommendations.

Objective 4: a prototype online clinical decision support tool

The prototype tool is available at PriMUS decision aid for LUTS (shinyapps.io). This is a draft and not yet intended to be used in practice (details of evaluation to date under next objective).

The GP is asked to enter data relating to demographics, initial questionnaire and uroflowmetry results and any previous treatments that have been used. A ‘Statistics’ summary of findings is provided, along with further pages with graphical displays (bar charts and crowd figures) of diagnostic probabilities, and management recommendations in accordance with NICE guidance. It is intended that the GP and patient may review these later pages (displays and recommendations) together to move towards a shared decision of management choice.

Objective 5: feasibility of introducing the clinical decision support tool into primary care

To explore GPs' experiences of managing LUTS, together with patients' experiences of and preferences for treatment in primary care, 25 patients and 11 GPs were purposively sampled from 20 GP practices in 3 UK regions. We also conducted initial user-testing of the prototype decision support tool with GPs. Participants were asked to try out the prototype tool prior to the interview so that they were able to give feedback on design and ease of use. We also conducted a simulated consultation workshop involving the study management GPs (AE and HA) and patient and public contributors (n = 3). Field notes were taken during this workshop.

A framework approach was used to analyse interview data. There were four main themes concerning treatment of LUTS in primary care: unresolved symptoms, preference for primary care, satisfaction with involvement in decision-making, and challenges of managing LUTS in primary care. Our findings emphasise the importance of LUTS being managed in primary care where possible, as in addition to cost savings and reduced waiting times, this is a more accessible option for patients, who tend to be more comfortable and confident being treated by familiar clinicians or in more familiar environments.

Feedback from GPs during development and small-scale user-testing with the tool in simulated consultation scenarios was favourable. It is more likely that the tool will be applicable in primary care by *internally referring patients to a GP with Special Interest (in LUTS)* who can undertake the discussion, examination, gather the required data (e.g. voiding symptoms subscore, flow rates and residual volumes; PSA testing) in consultations that are longer than the usual 10 minutes' duration, and review treatments chosen and benefits or side effects ensuing. This 'clinic model' might operate at the single practice or cluster of practices level.

Objective 6: potential National Health Service costs involved in the new pathway

We were unable to meet this objective as originally intended. Up-to-date data on referral rates are required (identified to have risen to 30% of men presenting with LUTS in primary care at the time of commissioning PriMUS study, 2017). The original objective included assessing Clinical Practice Research Datalink data for a reference standard for referrals, and current medical management and to model whether these are likely to diminish by implementing the PriMUS tool. On detailed assessment of feasibility, we concluded that identifying index cases (men presenting with LUTS) may be susceptible to poor coding of presentations in primary care (e.g. when presenting with sometimes vague symptoms). Complex resource intensive work would be required to ascertain cases and outcomes, and the other study pressures (delayed study recruitment and data collection during the pandemic years) meant that this was not feasible.

Given the reasonable diagnostic accuracy data for the PriMUS models, it is now possible to undertake economic modelling research, including the potential resource use effects of adopting the PriMUS tool, its diagnoses and management recommendations. This modelling would provide the basis for constructing decision-analytic models with precision to study the impact of the diagnostic precision models. Key outcomes that determine overall resource use are medical management (treatments and consultation time) and referral rates. Training requirements and set-up of 'clinics' at practice or cluster level are also relevant.

Further research: feasibility trial and randomised controlled trial

If suggested to be potentially cost-effective, then it will be important to evaluate use of the PriMUS tool in practice. This is likely to include a feasibility trial and a powered randomised controlled trial, again in the context of the single/cluster practice model as outlined above. A process evaluation will provide valuable information for wider implementation. As with all models currently used in practice, continued external validation of the models is important, and in populations with greater ethnic diversity.

As above, the primary outcome is likely to be that of referral rates as these are the driver of resource use (secondary care investigations, clinic time etc). Treatment decisions – including medications, review appointments, investigations – are also important outcomes across primary and secondary care, affecting overall resource use. A cost-effectiveness study of implementing the PriMUS tool in routine general practice is required. It will also be important to capture important patient-based outcomes, potentially including patients' confidence in treatment decisions, adherence to treatment decisions (also affecting resource use measures) and patient safety.

In addition, prior to use of the tool in clinical practice, it would be required to undergo the process of regulation and certification as a medical device.

Study registration

Current Controlled Trials ISRCTN10327305.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 15/40/05) and is published in full in *Health Technology Assessment*; Vol. 29, No. 1. See the NIHR Funding and Awards website for further award information.

Chapter 1 Introduction

Background

Lower urinary tract symptoms (LUTS), such as frequent urination, a slow stream and having to wake in the night to urinate, affect a significant proportion of older men and can lead to poor quality of life. Three common causes of LUTS are: instability of the bladder muscle (detrusor overactivity, DO), benign enlargement of the prostate gland causing bladder outlet obstruction (BOO), and weakness of the bladder muscle (detrusor underactivity, DU). These may be present individually or in combination.

The reference standard test for investigation of LUTS, and thus diagnosis of DO, BOO and DU, is invasive urodynamics, which takes place in secondary care. It involves insertion of catheters into the patient's bladder and rectum so that the behaviour of the bladder and outlet can be examined during filling and voiding. Owing to availability, complexity and cost, management decisions for men with LUTS are often based on results from a combination of non-invasive and minimally invasive investigations instead. These include digital rectal examination (DRE) to assess prostate size, symptoms questionnaires, uroflowmetry and measurement of post-void residual.

National Institute for Health and Care Excellence (NICE) guidelines[2] suggest that many men referred to specialist care with LUTS are eventually managed conservatively, and so could have remained within primary care. Male LUTS (MLUTS) account for around four presentations per month in an average-sized general practitioner (GP) practice. This rate of presentation, although high enough to represent a large burden on the NHS, makes it difficult for GPs to gain sufficient expertise to be confident about diagnosis and management. Further, GPs do not have access to validated clinical decision tools giving an indication of the most likely cause of symptoms to guide treatment and management. Making such a tool available should improve treatment efficacy, standardise treatment, reduce unnecessary referrals, expedite referral of those requiring specialist care, and thus improve cost-effectiveness of NHS care.

Rationale for Primary care Management of lower Urinary tract Symptoms

Male LUTS is rarely a threat to life or health, but patients often find the problem very intrusive (often termed 'bothersome') to work and social life. The current concentration of diagnostic assessment and initiation of treatment in secondary care usually means a delay in addressing the problem, and inconvenience and extra embarrassment at having to have hospital assessment at a distance from home.

A primary care-based clinical decision support tool with defined accuracy would, firstly, mean that the men could undergo the necessary simple tests straightaway, organised through the GP surgery and, secondly, would get a quicker result regarding predicted diagnosis and choice of management options that are most likely to be effective. For GPs, a clinical decision support tool could allay uncertainty around both diagnosis and best management.

Specialist urology units would concentrate more on specialist investigation and treating the 10–20% of men who require complex management such as surgery. Those men still requiring early referral or who have not benefitted from simple management options would be referred with a much more informed perspective of their problem, with all the initial diagnostics completed, making planning of further care quicker and seamless.

Reducing rates of referral to secondary care for a number of clinical conditions has been identified as one way of improving delivery of NHS care. Full economic evaluation is not included in this study since the design concentrates on diagnostic accuracy. However, using the rate of consultation and referral documented in the Clinical Practice Research Datalink (CPRD), we estimate that, if referrals to secondary care were halved, the NHS in England would save about £1.3 million per year – in equivalent opportunity costs/savings relating to freed up capacity – from reduced outpatient attendance. Although we estimate that such a change would require one extra 15-minute consultation with the GP, the relatively low prevalence of the condition means this should not noticeably impact overall GP workload. It may also be undertaken outside the 'usual' consultation setting or clinic, such as a GP with Special Interest operating in a cluster of practices, or at a community diagnostic clinic. Assessment and initial treatment costs would be unchanged but will be incurred in the primary rather than secondary care setting. Outcome(s) from the patient perspective should not change since GPs will now be receiving the same treatment recommendation from the clinical decision support tool that they would have previously been sent from the secondary care urology clinic.

Objectives

The primary objectives of the Primary care Management of lower Urinary tract Symptoms (PriMUS) study were to:

1. Develop statistical models to predict the likelihood of each urological condition (BOO, DU and DO), based on a series of non-invasive index tests, with urodynamics as the reference standard.
2. Estimate the diagnostic accuracy of the above models in an independent validation cohort.

The study incorporated an internal pilot phase with 'Stop/go' criteria that included quantitative and qualitative assessments. The progression criteria were designed to allow for mitigating strategies to be discussed to allow for some adaptation to recruitment processes in the main study.

The secondary objectives were to:

1. Develop a series of patient management recommendations and thresholds for clinically useful diagnostic prediction by expert consensus and with reference to current clinical guidelines that map to the diagnoses predicted by the statistical model.
2. Combine the statistical model and management recommendations into an online tool that will form the prototype clinical decision support tool.
3. Complete a qualitative study to explore the feasibility of introducing the clinical decision support tool into primary care, including potential acceptability to primary care staff and patients.
4. Collect NHS costs involved in delivering the new pathway and compare with cost of standard pathway calculated from NHS and other sources.

Chapter 2 Methods

Main study

Study design

This prospective diagnostic accuracy study assessed which of several simple clinical variables (index tests) collected in primary care, and that when used individually or in combination, best predicted a urodynamic diagnosis in men who presented to their GP with LUTS. Two cohorts of participants, one for development of the prototype diagnostic models (one model for each condition – BOO, DO and DU) and the other for validation, underwent a series of simple index tests in the community and an invasive reference standard (urodynamics) at participating GP practices or hospital sites. A qualitative evaluation was also conducted during the main study that aimed to gauge acceptability of tests and study procedures and obtain perspectives of both men and GPs to inform design and potential implementation of the clinical decision support tool in primary care (see *Chapter 5*).

Study procedures

See *Figure 1* for a flow diagram of the patient pathway.

Site set-up

Between February 2018 and December 2021, 73 primary care sites were opened to recruitment across the UK. These were managed through three study hubs, Newcastle, Bristol and Wales. The networks with whom we had agreements to facilitate the study were: Wales Primary Care (originally 'PiCRIS') Research Network; North-East England and North Cumbria Research Network and Western Research Network.

Participants

Men with one or more LUTS were identified in UK primary care settings either opportunistically during a GP consultation or by regular, predefined primary care database searches for symptom presentations. Recruitment commenced in March 2018, with participants being recruited consecutively into the development and validation cohorts.

The first set (350) of participants to complete the uroflowmetry, other index tests and invasive urodynamics procedure comprised the development cohort; subsequently an intended sample size of 325 complete participant data sets was sought to comprise the validation cohort.

Inclusion criteria

- Men aged 16 years and over.
- Men who presented to their GP with a complaint of one or more bothersome LUTS (this includes men on current treatment, but who are still symptomatic).
- Men able and willing to give informed consent for participation in the study.
- Men able and willing to undergo all index tests and reference standard, and who completed study documentation.

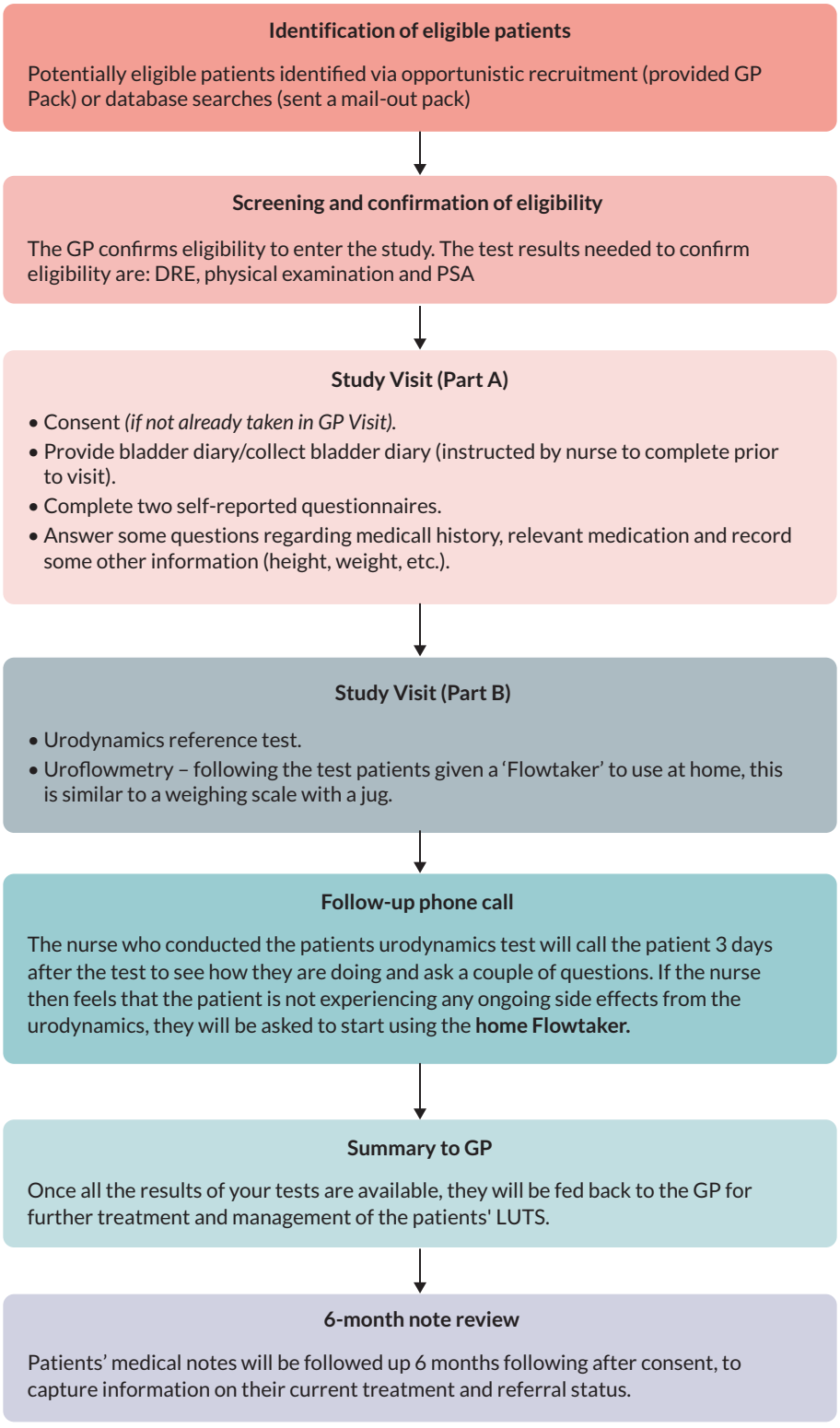


FIGURE 1 Patient pathway.

Exclusion criteria

- Men with neurological disease or injury affecting lower urinary tract function.
- Men with LUTS considered secondary to current or past invasive treatment or radiotherapy for pelvic disease.
- Men with other contraindications to urodynamics, such as heart valve or joint replacement surgery within the last 3 months or immunocompromised/immunosuppressed men.

- Men with indwelling urinary catheters or who carry out intermittent self-catheterisation (ISC).
- Men whose initial assessment included clinical findings that are suggestive of possible:
    - Prostate or bladder cancer according to standard NHS cancer pathways. If later deemed unlikely, they become eligible for study participation. Note that prostate-specific antigen (PSA) level was not in itself an inclusion/exclusion criterion – for example, a man with known raised but stable PSA was eligible if it was considered that (risk of) prostate cancer was not clinically considered as relevant to the LUTS presentation.
    - Recurrent or persistent symptomatic urinary tract infection (UTI). If UTI is successfully treated but LUTS remain, they become eligible for study participation.
    - Urinary retention, for example, palpable bladder after voiding.
- Men unable to consent in English or Welsh where a suitable translator is not available.

Participant identification

Participating clinicians were asked to approach eligible patients opportunistically during routine practice sessions.

Participating practices could also conduct a search of their patient electronic records using pre-defined READ codes to identify all potentially eligible patients who have presented within the previous 6 month(s) of the start of the study at that site and up to five additional times throughout the duration of the study. These patients were 'flagged' in their general practice clinical record using pre-specified READ codes to allow easy identification of patients, when they contacted the surgery, who could be eligible to participate.

Practices had the option to send relevant patients a study information pack (including the Patient Information Sheet, Patient Information Summary Leaflet, Urodynamics Leaflet and Bladder Diary), along with a letter informing them about the study, and invite the patient to attend the practice for a consultation.

General practitioner practices were also asked to run searches, based on relevant 'uroselective' medications that patients were currently being prescribed. Patients who were being prescribed these medications were invited by the practice for a medication review, exploring whether study entry would be suitable (i.e. still experiencing 'bothersome' symptoms and wishing re-evaluation).

Potential participants in Newcastle were also identified where they had already been referred from primary care to secondary care, but not yet been seen by a urologist, as the study team considered that these were relatively unselected patients equivalent to those seen in primary care (i.e. largely unmanaged in primary care before referral). Patients who met the PriMUS eligibility criteria could be entered to the study at that point, where they had not technically left primary care and entered secondary care.

Study posters, leaflets and adverts using electronic visual aids (where appropriate) were used in practice waiting areas to inform patients about the study.

When a patient presented to primary care with a complaint of LUTS the PriMUS study was introduced to that patient and the study information pack provided.

General practitioners had the option to consent a patient to enter the study during the initial consultation visit, if the patient met the eligibility criteria and they felt it appropriate. If the patient was not consented during the initial GP consultation, those who expressed an interest completed the Consent to Contact Form, which was passed to the local research nurse. The patient was then contacted by telephone by a research nurse based at the local research network hub or GP practice and invited to attend a study appointment.

Three screening tests prior to enrolment into the study were required: a physical examination of the abdomen (palpable bladder check), DRE and PSA test. The GP undertook the DRE and PSA test and the physical examination (as outlined

in NICE clinical guideline CG97) and completed the eligibility checklist, confirming the patient was eligible to enter the study. The GP was also required to record the results from these tests on a separate eligibility results form. DRE results were accepted from up to 6 months before the date of screening and PSA results were accepted from up to 12 months prior to the date of screening. This time range was a guideline, and we accepted results ± 1 or 2 weeks outside of this time range.

It was permissible for a nurse practitioner to perform these screening tests, assess eligibility and complete the relevant case report forms (CRFs), provided they had already been appropriately trained to perform the screening tests. During a follow-up telephone call with the research nurse, or delegated other, the study was explained in more detail and those who were happy to enter and were eligible were asked to attend Study Visit Part A and provide their informed consent.

During this appointment, the patient was asked to complete the International Consultation on Incontinence Questionnaire (ICIQ)–MLUTS symptom questionnaire and the International Prostate Symptom Score (IPSS) questionnaire[3] (see *Appendix 5*, *Figures 36* and *37*). The research nurse completed the patient registration and baseline CRFs.

Where possible, the research nurse arranged for the participant to undergo the reference test of urodynamics at the same study visit to minimise patient burden; however, in some circumstances, this was not possible and two separate study visits were carried out. The urodynamics procedures were mainly performed by the urodynamics research nurse at the GP practice but with access to the local urology service if required. In some circumstances, the urodynamics took place in a secondary care setting.

Informed consent

Informed consent was obtained during Study Visit Part A. Once consented, the participants were allocated a unique study number (participant ID).

COVID-19 study adaptations – consent

To mitigate the risk of contracting COVID-19, the informed consent process was made more flexible. Patients could opt to receive, sign and send back the consent form either electronically or via the post. Additionally, verbal consent could be taken over the telephone by the research nurse or healthcare assistant (HCA). Each statement was read out with the patient confirming consent verbally. The research nurse/HCA would then sign off the consent form on the participant's behalf. However, these participants were then required to re-confirm and sign the consent form when they attended for their urodynamics appointment.

Separate informed consent was taken for participation in the qualitative data collection (see *Chapter 5*).

Withdrawal

Participants had the right to withdraw consent for participation in any aspect of the study at any time. The participants were informed that their care would not be affected at any time by declining to participate or withdrawing from the study.

If a participant initially consented but subsequently withdrew from the study, clear distinction was made as to what aspect of the study the participant was withdrawing from. These aspects could be:

1. Withdrawal from index tests/reference test allowing for data already collected and medical records to be used.
2. Withdrawal from NHS medical note review but allowing data already collected to be used.

3. Withdrawal from entire study (index tests/reference test and NHS medical note review) and does not want any data already collected relating to them to be used.

Furthermore, it was important to collect safety data ongoing at the time of withdrawal, especially if the participant withdrew due to a safety event.

A participant could withdraw or be withdrawn from the study for the following reasons:

- Withdrawal of consent for investigation by the participant.
- Any alteration in the participant's condition which justified the discontinuation of the study in the Chief Investigators' opinion.
- Lost to follow-up.
- Other – determined by patient's GP.

COVID-19 study adaptations – Study Visit A and Visit B

Participants were given the option to conduct Study Visit A over the telephone, thereby reducing the number of in-person visits the participant needed.

When arranging the Study Visit B, nurses were required to complete a risk assessment tool with the participant over the telephone, to establish the COVID-19 risk when attending the appointment. For those considered at high risk following assessment, the advice was to delay the study visit until it was safer to do so. All patients considered low risk were able to attend for their urodynamics. Both nurse and participant were required to wear appropriate personal protective equipment. The urodynamics had to be performed within 18 months of the DRE and PSA test results, if outside of the 18 months, then the DRE and PSA tests had to be repeated before urodynamic studies was performed.

Data collection

General practitioners, primary care nurses or an appropriately trained delegate undertook the data collection relating to all the index tests. Specialist trained urodynamic nurses undertook the data collection for the reference urodynamics test.

All data collection was done by electronic data capture using a bespoke database developed by the Cardiff University Centre for Trials Research Clinical Trials Unit (CTR), with paper copies of all CRFs being made available.

Study Visit Part A

Once informed consent was obtained, the data from the index tests were collected. This included a baseline assessment; collecting demographic information, relevant medication, and medical history and two self-reported questionnaires – IPSS and ICIQ-MLUTS. Participants were given a bladder diary to complete for 3 days at home and instructed to bring this to their urodynamic visit (Study Visit Part B).

Study Visit Part B – reference standard

On arrival, the participant was asked to pass urine into a flow meter in private, after which a measurement of post-void residual ultrasound (one of the index tests; see below) was made. A dual lumen catheter (one channel to fill the bladder, and the other to measure intravesical pressure, P_{ves}) was inserted into the bladder via the urethra, and a single lumen catheter inserted into the rectum to measure abdominal pressure (P_{abd}). Detrusor pressure (P_{det}), generated by the bladder muscle itself, was calculated by subtracting P_{abd} from P_{ves}.

Filling phase

The participants were asked to bring their completed bladder diary (one of the index tests) to their urodynamics appointment, providing the urodynamic nurse with an indication of their maximum bladder capacity. The participant's bladder was filled with sterile saline at a maximum rate of 50 ml/minute. They were asked to report the first sensation of bladder filling, followed by the point at which they felt the normal desire to void, and finally the strong desire to void. At this point, bladder filling was stopped and provocation, in the form of running taps and asking the participant to cough, were performed. During the COVID-19 pandemic, participants were asked to perform the Valsalva manoeuvre instead of coughing to mitigate against the risk of spreading COVID-19.

Voiding phase

Following provocation, the patient was given permission to void, marking the start of the voiding phase. Voided volume (V_{void}) and flow rate (Q) were measured as they passed urine into the flow meter.

If either the filled or voided volumes were below 150 ml, the filling and voiding phases were repeated once more using a maximum filling rate of 20 ml/minute.

Diagnostic definitions

Definition of our three target conditions were based upon the following parameters measured during invasive urodynamics and subsequently read from a graphical representation of the test:

1. Maximum detrusor contraction pressure during the filling phase.
2. Maximum flow rate during the voiding phase (Q_{max}).
3. Detrusor pressure at the point of maximum flow rate ($P_{det}Q_{max}$).

If there were no detrusor contractions during filling, DO was deemed as not present. If there were any contractions (contraction pressure > 0), DO was deemed as present.

Diagnosis of BOO was based upon the bladder outlet obstruction index (BOOI), defined as $P_{det}Q_{max} - 2Q_{max}$. BOO was deemed as present if BOOI > 40, and absent if BOOI ≤ 40.

Diagnosis of DU was based upon the bladder contractility index (BCI), defined as $P_{det}Q_{max} + 5Q_{max}$. DU was deemed as present if BCI < 100, and absent if BCI ≥ 100.

Debrief process and monitoring process

The urodynamic nurses debriefed the participant following the urodynamic procedure, providing them with a post-urodynamics leaflet and safety card. The urodynamic nurse also instructed the participant that they would receive a follow-up phone call approximately 3 days later, to monitor for any related adverse events.

Blinding and review process

Invasive urodynamics is a complex investigation and interpretation can be challenging. Further, because standard practice involves interaction between the reference and some index tests as described earlier, the primary interpretation in this study was not blinded. Therefore, a review process was implemented to ensure the integrity of the reference standard. All urodynamics procedures and assessments were second read by a blinded reviewer to determine the three parameters. If any of the resulting diagnoses differed between the nurse and the reviewer, the case was

sent to a second non-blinded reviewer to make the final decision. The workflow for the final diagnosis of each target condition verified using the reference standard, including how the participants were recorded as indeterminate, is presented in *Appendix 1*, *Figures 24* and *25*.

Follow-up phone call on day 3

The urodynamic nurses contacted the patient 3 days (± flexibility if the third day fell on a weekend day) after their urodynamic procedure to monitor for any adverse events and serious adverse events (SAEs). Any SAEs were subsequently recorded by the urodynamic nurses and processed centrally by CTR.

Uroflowmetry (Flowtaker)

The participant was provided with the Flowtaker at the end of their urodynamic visit. They were provided with an information sheet on how to use the Flowtaker and instructed not to start this until given the green light to do so during their follow-up phone call. The participant was asked to return the device back to the GP practice within 2 weeks. If the patient was unable to do this, they were provided with packaging to post the device back to the research nurses within 2 weeks.

General practitioner summary report

Once the reference and index tests had taken place, results were compiled into a GP summary report which was provided to the GPs, along with a summary of relevant NICE-recommended managements,[2] to help inform management of the participant.

Six-month follow-up

A review of the patient's medical notes took place 6 months after the summary GP report was sent to the GP. This included any medication changes, any representations to the GP with bothersome LUTS and whether there had been any referrals to secondary care.

Safety and pharmacovigilance

Invasive urodynamics has the potential to cause adverse events. A medical doctor was required to be on site while the test was taking place. Due to a 5% risk of a UTI,[4] the urodynamic nurse also provided the participant with a post-urodynamics debrief sheet following the test, informing them on the importance of drinking plenty of water for 24 hours following the test, how to identify signs of an infection, and to seek medical care if they suspected they have one.

Serious adverse events were reported from the time of signature of informed consent, throughout the treatment period up to, and including, 7 days after the participants received their urodynamic assessments.

Adverse events were captured by the urodynamic nurses either during Study Visit Part B or during the 3-day follow-up phone call. For SAEs, an assessment of causality between the event and the study intervention, and the expectedness of the event, was carried out by the (GP site) principal investigator, or delegated urodynamic nurse, and then independently by a clinical reviewer. If the clinical reviewer classed the event as *probably* or *definitely* caused by the intervention, it was classified as a serious adverse reaction.

Primary outcome measures

Sensitivity and specificity of the diagnostic models were evaluated for detecting DU, BOO and DO. The three conditions were coded as binary outcomes (present/absent).

Secondary outcome measures

A patient management algorithm to guide initial treatment for men with LUTS.

A prototype online clinical decision support tool for use in primary care.

Qualitative summary of patients' and clinicians' views on the use of a LUTS clinical decision support tool in the primary care setting.

Costs/savings of implementation of the primary care LUTS clinical decision support tool both from a population and individual patient perspective.

Development cohort

The development cohort was used to develop three separate diagnostic models, which combined results from the index tests to predict a urodynamic diagnosis of BOO, DU and DO.

Validation cohort

The diagnostic accuracy of the models was assessed using data from the validation cohort.

Candidate predictors and index tests

Twelve potential predictors were considered for the three diagnostic models. The candidate predictors and the index tests from which they were derived are listed in *Table 1*. Their selection was informed by subject knowledge using a systematic review included in the NICE guideline CG97[2] updated with a study-specific unpublished selective review by our group in 2015, the judgement of the expert clinical members of the study team, and the stipulations of the funding commissioning brief. Alternative models with fewer or different forms of the predictors were considered in sensitivity analyses (see *Chapters 3* and *4*). All participants were requested to undertake all tests, which were a combination of:

- Tests carried out for eligibility assessment prior to enrolment
- Tests carried out at participant visits to a primary or secondary care location for the purpose of the study following enrolment
- Tests carried out by the participant at home following enrolment.

Digital rectal examination: The clinician inserts a gloved, lubricated finger into the patient's rectum in order to examine the prostate gland. Both the size of the prostate and its hardness and nodularity are assessed. In this study, the presence of hardness or nodularity was used as an eligibility screening test, and the size was used as a candidate index test. NICE guideline CG97 recommends that men with LUTS are offered a DRE at initial GP assessment.

Prostate-specific antigen blood test: PSA is a protein produced by prostate gland cells that is often elevated in the presence of prostate cancer. It is also thought to be correlated with prostate size. Normal ranges are age dependent. In this study, PSA level was used as both an eligibility screening test and as a candidate index test. NICE guideline CG97 recommends that men with LUTS are offered a PSA test if they meet certain criteria, such as an abnormal DRE, or if they are concerned about prostate cancer.

TABLE 1 Index tests and candidate predictors for diagnostic models in the main analysis for BOO, DU and DO

Index test	Candidate predictor *Unit or result*
Participant demographics	Age *Years*
DRE	Size of prostate *Ordinal: Normal = 1/Mild enlargement = 2/Moderate enlargement = 3/Gross enlargement = 4*
Blood test	PSA test result *ng/ml*
ICIQ–MLUTS questionnaire[a]	Incontinence symptoms subscore *Score range: 0–24*
	Voiding symptoms subscore *Score range: 0–20*
Bladder ultrasound[a]	Post-void residual urine *ml*
Uroflowmetry[a]	Median maximum flow rate *ml/second*
	Median voided volume *ml*
	Mean 24-hour frequency *Voids per day*
	Mean nocturia frequency *Voids per night*
Bladder diary[a]	Mean urgency score
	Mean 24-hour fluid intake *ml*

a Not currently available and/or in routine use in primary care.

International Consultation on Incontinence Questionnaire–MLUTS questionnaire: A questionnaire used to assess LUTS and their impact on quality of life. It includes 13 questions scored on a scale from 0 to 4, allowing calculation of a voiding subscore from 0 to 20, an incontinence subscore from 0 to 24, and a storage subscore from 0 to 12. NICE guideline CG97 recommends that men considering any treatment for LUTS are offered assessment of their baseline symptoms using a validated symptom score. The ICIQ–MLUTS and IPSS questionnaires [see the description of sensitivity analysis 2 (SA2) below] are examples of these.

Bladder ultrasound: An ultrasound image of the bladder taken immediately after voiding to measure the post-void residual volume (the amount of urine remaining in the bladder). NICE guideline CG97 recommends that measurement of post-void residual volume is not routinely offered to men with LUTS at initial assessment, but that it is offered at specialist assessment.

Uroflowmetry: A portable home flow meter device (Flowtaker; Laborie, Mississauga, Canada) used by patients at home for several days to record multiple voids. Flowtaker measures urine flow rate, urine volume, and the date and time of each void, and produces summary statistics including median maximum flow rate, median voided volume, mean 24-hour frequency and mean nocturia.

National Institute for Health and Care Excellence guideline CG97 recommends that measurement of urine flow rate is not routinely offered to men with LUTS at initial assessment, but that it is offered at specialist assessment.

Bladder diary: A 3-day diary that a patient completes to record voiding times and volumes, incontinence pad use, fluid intake, sensations of urgency and urine leakage. The PriMUS team devised rules enabling calculation of summary statics from this information including mean urgency score and mean 24-hour fluid intake. A bladder diary is a more detailed version of a frequency-volume chart, which includes voiding times and volumes only. NICE guideline CG97 recommends that men with bothersome LUTS are asked to complete a frequency-volume chart at initial assessment.

Reference standard

The reference standard was invasive urodynamics, a test routinely carried out in a specialist care setting for the investigation of LUTS. Invasive urodynamics was conducted rather than video urodynamics, which is in line with the most contemporary national and international guidelines, and is sufficient to diagnose BOO, DU and DO, to which most non-complicated adult MLUTS can be attributed. Invasive urodynamics was performed using portable equipment (Goby, Laborie, Mississauga, Canada) in either a primary or secondary care location, by specially trained urodynamic nurses, according to International Continence Society standards.[5] A dual-lumen catheter (one channel to fill the bladder and the other to measure intravesical pressure) was inserted into the bladder via the urethra, and a single-lumen catheter was inserted into the rectum to measure abdominal pressure. Detrusor pressure, generated by the bladder muscle itself, was calculated by subtracting abdominal pressure from intravesical pressure.

Quality assurance work on the urodynamics assessments was undertaken. If the urodynamics result did not agree between the first two reviewers, these cases proceeded to third-reviewer arbitration. Results are presented here as they relate to both Development and Validation cohorts. For 110 (31.4%) participants from the development cohort and 78 (31.1%) from the validation cohort, the urodynamics result did not agree between the first two reviewers, and these cases proceeded to third-reviewer arbitration.

Internal pilot study

An internal pilot study was conducted between February 2018 and November 2018.

The aims of the internal pilot were to establish that the protocol for the main trial could be implemented by assessing:

1. Site and patient recruitment rate.
2. The proportion of patients undergoing urodynamic assessments.
3. Those with a complete data set for analysis.
4. The acceptability of the gold standard urodynamic assessment with patients, and potential refinements in processes via a qualitative evaluation.
5. Assess the prevalence of the three urodynamic diagnoses.

The pilot qualitative evaluation is discussed in *Chapter 5*.

All progression criteria were attained, and the Study Steering Committee (SSC) approved the continuation of the study. The full pilot report is included as an appendix along with some changes that were made to the study design and management to improve efficiencies for the main study (see *Appendix 2*).

Statistical analysis methods

Sample size

Sample size calculations were carried out separately for the model development and validation cohorts. For both, estimated prevalence for BOO, DU and DO of 31%, 16% and 57% was used, respectively, based on previous literature[6,7] and clinical expertise.

Development cohort

The sample size for developing the logistic regression models was based on a rule of thumb suggesting that five events per variable were adequate.[8] 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 (DU diagnoses). Over 100 events were expected for the BOO model and 200 for the DO model providing more than an adequate number of events.

Validation cohort

The required sample size 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. Better sensitivity and specificity will give narrower confidence intervals (CIs).

Attrition

To allow an attrition rate of 20–25%, the resulting 675 was increased to give a final sample size of 880.

Missing data

Patterns of missingness were investigated in the development and validation cohorts by tabulating the predictors and reference standard results. This analysis provided information on predictors that could be difficult to obtain in practice, and if alternatives could be considered to ensure the models are clinically relevant, as assessed in the sensitivity analyses described below.

Number of participants that had missing or indeterminate reference standard results were reported but excluded from the analysis. When data were missing for candidate predictors, multiple imputation by chained equations (MICE)[9] was used to impute missing values by replacing missing values with plausible values based on the distribution of the observed data. The methodology compensated for the uncertainty of the imputation procedure and ultimately allowed us to perform the analysis with greater power on most of the participants. Distributions of imputed values were visually checked for comparability with the observed data. The number of imputed data sets that were created was determined by the percentage of participants that had at least one candidate predictor missing. For instance, if 15% of individuals had at least one candidate predictor missing, then 15 imputed data sets were created. If the percentage was < 10%, then 10 imputed data sets were created. Multiple imputation was performed using the mi package in Stata 15 (StataCorp LP, College Station, TX, USA).

Descriptive statistics

Each cohort has been described in terms of demographics and clinical characteristics. Categorical data were summarised using frequencies and percentages. Continuous data were summarised using mean and standard deviation where the data were normally distributed, and median and interquartile range where the data were skewed. The numbers of patients with missing data were also tabulated.

Model development

Model development was performed using results from the first 350 participants to complete the uroflowmetry and urodynamics procedure, and external model validation was performed using the subsequent 325 participants.

The three target conditions were coded as binary outcomes (present/absent) in the logistic regression models. The proportion of men with and without each target condition was reported. The proportion of men classified as indeterminate with each target condition was also reported; these men were excluded from the analyses.

Candidate predictors were selected from those listed in *Table 1*. As predictor distributions should be wide to facilitate reliable predictions, the distribution of each predictor was explored prior to selection. Relationships between predictors were investigated using scatter plots and estimation of correlation coefficients. All selected candidate predictor variables were included in the models without univariable evaluations of association between outcome and predictor for statistical significance. To gain maximum diagnostic information, continuous variables were not categorised. Non-linearity of continuous variables was accounted for by using a multivariable fractional polynomial approach to identify

appropriate transformations. Logistic regression analysis was performed on each imputed data set and the imputation-specific coefficients were combined using Rubin's rules.[10] Each model was developed using backward elimination with a *p*-value of 0.157 to select predictors for inclusion.[11] This *p*-value was chosen because it is known to be a good proxy for the Akaike information criterion approach.

Model performance

To evaluate the predictive performance of each model, the predicted probability and linear predictor were estimated for each participant. The predictive performance of each model was assessed in terms of discrimination, which is the ability to distinguish between those who do or do not have a particular diagnosis, and calibration meaning the agreement between predicted and observed probabilities. Discriminative ability was assessed using the c-index and its 95% CI. For a logistic model, this is equivalent to the area under the receiver operating characteristic (ROC) curve. Calibration was evaluated using calibration plots and the calibration slope. Calibration plots of the average observed probability against predicted probability were used to visually assess calibration. Within each decile of the predicted probability, the average predicted probability was compared with the corresponding observed proportions. The calibration slope of the prognostic index (linear predictor) was estimated using logistic regression with the linear predictor as the covariate. The value for the calibration slope should ideally be one signifying perfect agreement between the predicted probabilities and the observed probabilities. A calibration slope < 1 indicates that a model over-predicts while a calibration slope > 1 indicates under prediction.

Internal validation

Bootstrapping was used for internal validation to assess model overfitting and optimism. For each model, 100 bootstrap samples were obtained from the original data sets, then multiple imputation and the variable selection process were repeated for each sample in order to check for model instability, as recommended in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis guidelines.[12] Optimism refers to the difference between the 'apparent' and 'test' performance statistics, where 'apparent' refers to the apparent performance of the bootstrap model fitted to the bootstrap sample and 'test' refers to the validation performance of the bootstrap model fitted to the original data sets. The average optimism was then determined. The last step involved subtracting the average optimism from the original model's apparent performance. The optimism-corrected c-index and calibration slope were obtained, where the latter was used as the shrinkage factor to correct a model. The apparent coefficients were multiplied by the shrinkage factor to obtain the optimism-corrected coefficients, and the intercept was re-estimated by fitting a logistic model with the shrunken coefficients set as an offset.[13]

Sensitivity analysis

Sensitivity analyses were performed by fitting two alternative models. In sensitivity analysis 1 (SA1), two predictors (mean urgency score and mean 24-hour fluid intake from the bladder diary) were excluded from the list of candidate predictors, because these predictors may be difficult to obtain in practice. This sensitivity analysis assessed the adequacy of using fewer tests.

For SA2, alternative measures or methods of measurement of seven predictors included in the main analysis were considered (*Table 2*). Some predictors may be difficult to obtain in practice from a technological perspective and some men may prefer the use of one method to another (i.e. Flowtaker or bladder diary). Thus, SA2 assessed the robustness of the models to the approach used to obtain the predictors and the use of alternatives. The alternatives were as follows.

International Prostate Symptom Score questionnaire: A shorter and therefore less burdensome questionnaire than the ICIQ–MLUTS, also used to assess LUTS and their impact on quality of life. It includes four questions about voiding symptoms, and three about storage symptoms, all scored from 0 to 5. This allows calculation of a voiding subscore out of 20, and a storage subscore out of 15. In SA2 IPSS storage subscore was used as an alternative to ICIQ–MLUTS incontinence subscore, and IPSS voiding subscore was used as an alternative to ICIQ–MLUTS voiding subscore.

Physical examination of palpable bladder: The clinician examines the patient's bladder by feeling their abdomen to determine whether it is palpable. In SA2 this was used as an alternative to post-void residual urine as it also gives an indication of how well the bladder empties, but without the requirement for an ultrasound measurement.

TABLE 2 Index tests and candidate predictors for diagnostic models in sensitivity analyses for BOO, DU and DO

Main analysis **Candidate predictor** **(Index test)** ***Unit or result***	**SA1** **Candidate predictor** **(Index test)** ***Unit or result***	**SA2** **Candidate predictor** **(Index test)** ***Unit or result***
Age (Participant demographics) *Years*	Age (Participant demographics) *Years*	Age (Participant demographics) *Years*
Size of prostate (DRE) *Ordinal: Normal = 1/* *Mild enlargement = 2/* *Moderate enlargement = 3/* *Gross enlargement = 4*	Size of prostate (DRE) *Ordinal: Normal = 1/* *Mild enlargement = 2/* *Moderate enlargement = 3/* *Gross enlargement = 4*	Size of prostate (DRE) *Ordinal: Ordinal: Normal = 1/* *Mild enlargement = 2/* *Moderate enlargement = 3/* *Gross enlargement = 4*
PSA test result (Blood test) *ng/ml*	PSA test result (Blood test) *ng/ml*	PSA test result (Blood test) *ng/ml*
Incontinence symptoms subscore (ICIQ–MLUTS questionnaire*) *Score range: 0–24*	Incontinence symptoms subscore (ICIQ–MLUTS questionnaire*) *Score range: 0–24*	Storage symptoms subscore (IPSS questionnaire) *Score range: 0–15*
Voiding symptoms subscore (ICIQ–MLUTS questionnaire*) *Score range: 0–20*	Voiding symptoms subscore (ICIQ–MLUTS questionnaire*) *Score range: 0–20*	Voiding symptoms subscore (IPSS questionnaire) *Score range: 0–20*
Post-void residual urine (Bladder ultrasound*) *ml*	Post-void residual urine (Bladder ultrasound*) *ml*	Palpable bladder (Physical examination) *Binary: No = 0/Yes = 1*
Median maximum flow rate (Uroflowmetry*) *ml/second*	Median maximum flow rate (Uroflowmetry*) *ml/second*	Maximum flow rate (One-off flow test*) *ml/second*
Median voided volume (Uroflowmetry*) *ml*	Median voided volume (Uroflowmetry*) *ml*	Voided volume (One-off flow test*) *ml*
Mean 24-hour frequency (Uroflowmetry*) *Voids per day*	Mean 24-hour frequency (Uroflowmetry*) *Voids per day*	Average number of voids (Bladder diary*)
Mean nocturia frequency (Uroflowmetry*) *Voids per night*	Mean nocturia frequency (Uroflowmetry*) *Voids per night*	Nocturnal polyuria score (Bladder diary*)
Mean urgency score (Bladder diary*)	–	Mean urgency score (Bladder diary*)
Mean 24-hour fluid intake (Bladder diary*) *ml*	–	Mean 24-hour fluid intake (Bladder diary*) *ml*

Urine flow rate and volume can also be obtained from a hospital/GP-based urine flow test, which in this study was carried out during the invasive urodynamics session prior to catheterisation. In SA2 maximum flow rate and voided volume from a one-off hospital/GP-based urine flow test were used as alternatives to median maximum flow rate and median voided volume from Flowtaker, respectively.

The bladder diary, described earlier under *Table 1*, can also be used to calculate the average number of voids (the mean number of voids per 24 hours), and nocturnal polyuria score (the mean proportion of nocturnal urine volume). These were used in SA2 as alternatives to mean 24-hour frequency and mean nocturia frequency from Flowtaker, respectively.

External validation

The models were then externally validated, and model performance (c-index, calibration plot and calibration slope) was assessed using the validation cohort. When the models were subjected to overfitting or underfitting, they were re-calibrated to the validation data set. The intercept and coefficients were re-estimated using the calibration slope as a shrinkage factor. From the qualitative research, distributions of the probability (risk) thresholds were investigated for clinical usefulness of the prediction in guiding treatment of each target condition. However, this research found a lack of consensus among the clinicians in defining risk thresholds for BOO, DU and DO. Thus, the sensitivity and specificity were plotted on a ROC plot for each model, which included labels of the risk thresholds in increments of 10% points. Risk thresholds were identified at a sensitivity and specificity of 75%, which was deemed to be the minimum clinically useful performance.

Patient and public involvement

The National Institute for Health and Care Research (NIHR) defines patient and public involvement (PPI) in research in its Public Information Pack[14] as:

> *… research being carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them. It is an active partnership between patients, carers and members of the public with researchers that influences and shapes research.*

Patient and public involvement planning

VoiceNorth Panels

During development of our NIHR Health Technology Assessment (HTA) funding application, an overview of the project was presented in September 2015 to a PPI panel organised through the VoiceNorth PPI platform (now Voice Global, Newcastle University, Newcastle upon Tyne, UK).[15] The panel welcomed the idea of transferring simple tests to a more dignified (i.e. with a bit more privacy) setting, that is, from secondary care to primary care or the home environment. However, they had concerns about possible missed cancers and GP workload, which were acknowledged and addressed in further iterations of the application. The panel also believed that the inclusion of invasive urodynamics may be a recruitment incentive, as it would provide a timely and accurate diagnosis for men included in the study. A VoiceNorth panel were consulted again in March 2016 to review the application *Plain language summary*, and the suggestions gathered from the meeting included in the revised *Plain language summary*.

National Institute for Health and Care Research Health Technology Assessment funding application

The following PPI aspects were included in the NIHR HTA funding application:

- A co-applicant and member of the Study Management Group (SMG) to act as PPI co-ordinator. Early in the application process, this was role was held by the chief executive officer of The Bladder and Bowel Foundation (registered charity number: 1085095). Later, owing to closure of the charity in 2016, this role was allocated to Alison Bray.
- Involvement of patient-representatives in the following activities:
    - Design of the research
    - Management of the research
    - Developing participant information resources
    - Undertaking/analysing the research
    - Contributing to the reporting of the research
    - Dissemination of research findings.

Planned outcomes relating to PPI were as follows:

- Informing recruitment and retention strategies for the patient study.

- Ensuring that all patient-facing literature is appropriately written and emphasises the aspects of most concern to potential participants.
- Obtaining user input into how the decision support tool should look in practice, and what is required to ensure that its component tests can be delivered efficiently in the community.

Funding and payments

In the funding application, £8000 was allocated to PPI activities. In addition, a proportion of the salary cost for the PPI co-ordinator (representing 3% full-time equivalent for 3 years) was included.

Patient-representatives were paid, and their expenses reimbursed, according to NIHR guidelines on Reward and recognition for public contributors: A guide to the payment of fees and expenses[16] as follows:

- £25: Task/activity requiring little or no preparation and equating to approximately 1 hour of activity or less.
- £50: Task/activity likely to require some preparation and equating to approximately 2 hours of activity.
- £75: Task/activity where preparation is required and equating to approximately half a day's activity.
- £150: One-off, all-day meetings.
- £300: All-day meetings that require substantial preparation and/or responsibilities at the meeting.
- Variable: Reimbursement of any reasonable expenses incurred as a result of involvement in the project.

Potential applicants were referred to information on tax considerations relating to these payments, and asked to ensure they understood the implications of these before agreeing to take part. Payments were managed through The Newcastle upon Tyne Hospitals NHS Foundation Trust's (NuTH's) research finance team.

Recruitment of patient-representatives

In March 2017, with the PriMUS project scheduled to begin in May 2017, Bladder Health UK (registered charity number: 1149973) advertised the opportunity on its website and social media but was unable to recruit.

The opportunity was then advertised in May 2017 on the VoiceNorth PPI platform (now VOICE Global[15]), and with posters in NuTH's Urology department. Seven applications resulted from these two methods, and two patient-representatives were chosen to join the PriMUS SMG.

The same platform was used again in April 2021, branded at that time as Voice, for a further patient-representative recruitment exercise. Ten applications were received from which two patient-representatives were chosen, one to join the SMG and one to join the SSC. The other SSC member was recruited through Health and Care Research Wales (HCRW).

Patient-representative members of the Study Management Group/Study Steering Committee

Patient-representative members of the SMG were as follows:

- Malcolm Gristwood from June 2017 to November 2019
- Ray White from June 2017 to December 2021
- Peter Michell from May 2021 to the end of the project.

Patient-representative members of the SSC were as follows:

- Alan David Pryce May 2018 to the end of the project
- Alan Hart from May 2021 to the end of the project.

Patient-representative activities

Study Management Group meetings

Patient-representatives were welcome to participate in any aspect of the SMG meetings, and *Research partner/PPI* was a standing item on the agenda throughout the project. Patient-representatives were included on the e-mail distribution

list for SMG meetings, which included meeting agendas, minutes, study progress reports and any other documents to be considered/discussed by the group. SMG meetings took place via online videoconference, except one face-to-face meeting in London in May 2018, to which patient-representative were invited but were unable to attend.

Patient-representatives attended the SMG meetings regularly from July 2017 to March 2019. During this time, the SMG benefitted from their input into activities focusing on study planning and start-up.

There was a break in their regular attendance from March 2019 to May 2021. This was because recruitment of patients into the study was to target and stable between March 2019 and February 2020, followed by a period of no recruitment due to COVID-19 from March 2020 to September 2020. Recruitment picked up slowly thereafter. This meant that SMG discussions were fairly routine, and the patient-representatives and other SMG members felt that there was little opportunity for PPI contribution. During this time, the PPI co-ordinator acted as a point of contact between the patient-representatives and other SMG members, to allow communication about relevant topics, including boosting patient recruitment when it was able to resume in October 2020 following the pause due to COVID-19.

Regular attendance began again in August 2021 as SMG activities that would benefit from patient-representative input resumed/arose, such as review of the prototype decision support tool, discussed below.

Patient-representatives contributed to a variety of SMG discussions, including on the following topics:

- Patient-facing study materials.
- Project management, such as patient recruitment progress, and action plans to address recruitment issues.
- Study outcome measures.
- Study design and sample size of the qualitative work involving patients.
- Content of study publications.
- Implementation of the decision support tool.
- A serious incident review of mis-classified bladder diary data, which had resulted in some patients having erroneous diagnoses of nocturnal polyuria, providing patient perspective on the solutions, that is information sharing with patients and GPs.

Document reviews

Patient-representatives who reviewed patient-facing material were asked to consider the following aspects:

- Is the language appropriate?
- Is the information clear and well-ordered?
- Is there anything not included that you think should be addressed in these documents?
- Does the study sound appealing?

The content reviewed by patient-representatives was:

- Patient-facing material:
    - Patient information sheet.
    - Urodynamics leaflet.
    - Consent form.
    - Flowtaker instructions.
    - Recruitment poster.
    - Patient information booklet: A condensed version of the patient information sheet, created following feedback from patient-representatives that the patient information sheet was very long.
    - Post-urodynamics leaflet: A leaflet given to patients after their invasive urodynamics test with advice on how to reduce the risk of, and respond to, potential adverse events.
    - Invasive urodynamics patient safety card: A business card-size card given to patients after their invasive urodynamics test with summary advice on how respond to potential adverse events.
    - Patient recruitment animation video display in GP practice waiting rooms.

- This final report PPI section.
- Final report *Plain language summary* chapter.

Publication co-authorship

Our patient-representative Ray White was a co-author of a publication, led by Sarah Milosevic, entitled *Conducting invasive urodynamics in primary care: qualitative interview study examining experiences of patients and healthcare professionals*, which was published in the journal Diagnostic and Prognostic Research in 2021.[17]

The publication describes a qualitative study that aimed to explore the acceptability and feasibility of conducting invasive urodynamics in primary care, based on semistructured interviews with patients and with healthcare professionals. It concluded that patients generally found the invasive urodynamic procedure to be acceptable and valued the GP-based setting due to its convenience and familiarity.

Prototype online decision support tool evaluation

Patient-representatives were involved in a role-play session in December 2021 that aimed to obtain feedback on a prototype of the online decision support tool ('prototype tool'). The online tool itself is not a deliverable of the PriMUS project, but a prototype was created by Michael Drinnan and developed iteratively throughout the project (https://drinnan.shinyapps.io/PriMUS/).[18] The prototype tool incorporated statistical models developed by the PriMUS statistics team, described in the September 2021 report *PriMUS report v3.0 development cohort clean*.

The session was chaired by Alison Bray and involved patient-representatives Ray White, Peter Michell, and Frank Bray (father of Alison, involved as a one-off event), and SMG GP members Adrian Edwards and Harry Ahmed. It was held online using Zoom videoconferencing (Zoom Video Communications, California, USA). The two role-plays took the form of simulated GP consultations between a GP and a man seeking management for LUTS. The session ended with feedback and discussion, including the following:

- GPs referred to the tool during the 'consultation' less than hoped and suggested some treatment options not suggested by the tool. This is likely to improve with training and familiarity with the tool.
- There is not enough time in a standard GP consultation to do the tool justice. Holding a longer consultation would be preferable to simplifying the tool.
- Some of the terminology used in the tool was unclear to the patient-representatives, and to a lesser extent, the GPs.
- The interface was user-friendly and would help guide the patient–GP discussion.
- Data entry was easy and the graphical output was well-designed.
- It was queried whether the tool would provide a meaningful result if one or more of the inputs was unavailable.
- In one of the role-plays, the probabilities of the three diagnoses output by the tool were similar. The patient-representatives commented that ideally one would be high and the others low.

Recommendations for the prototype tool resulting from the session include those listed below. These should be reviewed in any further development of the decision support tool following PriMUS.

1. Review the terminology and language on all screens. Include lay definitions/explanations of specialist medical terms, possibly using pop-ups.
2. Distinguish screens intended to be seen by GPs only versus by both GPs and patients.
3. Include information on normal PSA ranges. Include a warning if PSA is above the normal range for the patient's age group.
4. Include treatment side effects, possibly using pop-ups.
5. Include a 'not available' option for missing tool inputs.

Study Steering Committee

The role of the SSC was to provide overall, independent supervision of the project through regular review meetings. The patient-representative members serving on the SSC ensured that the study was patient-centric, providing insight into study design (frequency of visits, data collection methods and burden on participants), protocol development, patient recruitment and retention and overall patient experience.

Patient and public involvement costs

Throughout the project, £3500 was attributed to PPI costs, broken down (rounded to the nearest £50) as follows:

- £2700: SMG meetings
- £250: Activities/workshops
- £400: Document reviews
- £100: Project introduction/orientation
- £50: Miscellaneous meetings

Conclusion

The PriMUS team has implemented PPI in the project in line with the proposal. There has not yet been patient-representative involvement in dissemination activities, which we plan to address after the project. The PPI activities came in under budget, which is in part due to planned in-person activities and events not taking place due to the COVID-19 pandemic. Patient-representatives have played a key role in shaping the design and conduct of the project and ensuring that the patient voice has been represented throughout. The other members of the project team wish to thank them for their involvement.

Equality, diversity and inclusion

To ensure our findings are generalisable and applicable to various demographic groups, we designed our research methodology to minimise bias and promote inclusivity in participant recruitment.

Participants were recruited from geographical areas that allowed for recruitment from a wide range of socioeconomic backgrounds, ages, ethnicities to ensure diversity within our recruited population.

Sites chosen for PriMUS were purposively spread across the UK. Participants were recruited from 55 GP practices across Newcastle upon Tyne, Wales and Bristol between 19 March 2018 and 4 December 2019 (development cohort) and 28 March 2018 and 8 June 2022 (validation cohort).

Entry criteria for PriMUS were designed to be as inclusive as possible and the baseline demographics of PriMUS participants reflects the primary care population in the UK of men with LUTS.

In the development cohort, the median age was 69 (interquartile range 61–74) years. The majority (97.4%) of the men was of white ethnicity. The validation cohort characteristics were generally similar with the median age of 67 (interquartile range 60–74) years. The majority (94.4%) of the men was of white ethnicity.

Reflections on your research team and wider involvement

The PriMUS research team had substantial PPI in study planning, the funding application, study delivery and prototype clinical decision support tool evaluation. The research team benefited from a diverse set of skills and expertise allowing for a well-defined research question to be formulated.

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.

The following substantial amendments were made during the study:

Substantial Amendment 1 (3 October 2017); Substantial Amendment 2 (10 January 2018); Substantial Amendment 3 (20 April 2018); Substantial Amendment 4 (26 February 2019); Substantial Amendment 5 (6 June 2019); Substantial Amendment 6 (6 September 2019); Substantial Amendment 7 (23 March 2020); Substantial Amendment 8 (25 August 2020).

Chapter 3 Development cohort results

Participants

The results presented in this chapter are based only on participants in the development cohort. The first 350 participants with returned uroflowmetry and urodynamics CRFs were included in the development cohort, with the last participant in this cohort completing the procedure on 4 February 2020.

Recruitment of participants into the development cohort

Participants were recruited from 55 GP practices across Newcastle upon Tyne, Wales and Bristol between 19 March 2018 and 4 December 2019. The number of participants recruited at each GP practice and study hub is shown in *Table 33*. The flow of men (development and validation cohorts) through the study until recruitment end date (8 June 2022) is shown in *Figure 2*. In total, 2470 participants were screened, of which 350 and 251 participants were in the development and validation cohorts, respectively. The number of men excluded, with reasons, is also shown in *Figure 2*.

Withdrawal level 1 = withdrawal from index tests/reference standard allowing data already collected and medical records to be used.

Withdrawal level 2 = withdrawal from medical note review, but allowing data already collected to be used.

Withdrawal level 3 = withdrawal of consent for the entire study and does not want any data already collected to be used.

The results of the reference standard (present, absent or indeterminate) for each target condition are shown in *Table 3*. There were 14 (4.0%), 16 (4.6%), 10 (2.9%) men with indeterminate diagnosis for BOO, DU and DO, respectively. Of the remaining 336, 334 and 340 men, there were 163 (46.6%), 141 (40.3%) and 253 (72.3%) with BOO, DU and DO, respectively. Thus, leaving 173 (49.4%), 193 (55.1%) and 87 (24.9%) without BOO, DU and DO, respectively. The final diagnosis of the 350 participants across the three target conditions is presented in *Table 4*. Eighteen men (5.1%) and 42 (12.0%) men were absent and present with all three target conditions, respectively. There were a high proportion of participants absent with DU and diagnosed with both BOO and DO, 88 (25.1%).

Baseline characteristics of participants

The demographics and clinical characteristics of the 350 men are presented in *Appendix 3*, *Tables 34–36* separately for each target condition. The median age was 69 (interquartile range 61–74) years. 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 inhibitor (SSRI) antidepressants and 25 (7.1%) participants on 5-alpha reductase inhibitors (see *Appendix 3*, *Tables 37–39*).

Adverse events

Three out of 350 participants in the development cohort experienced urodynamics-related adverse events that were severe or moderate, and required hospitalisation. One participant was diagnosed with haematuria (SAE level of severe) and the two other participants were diagnosed with retention of urine *Table 5*.

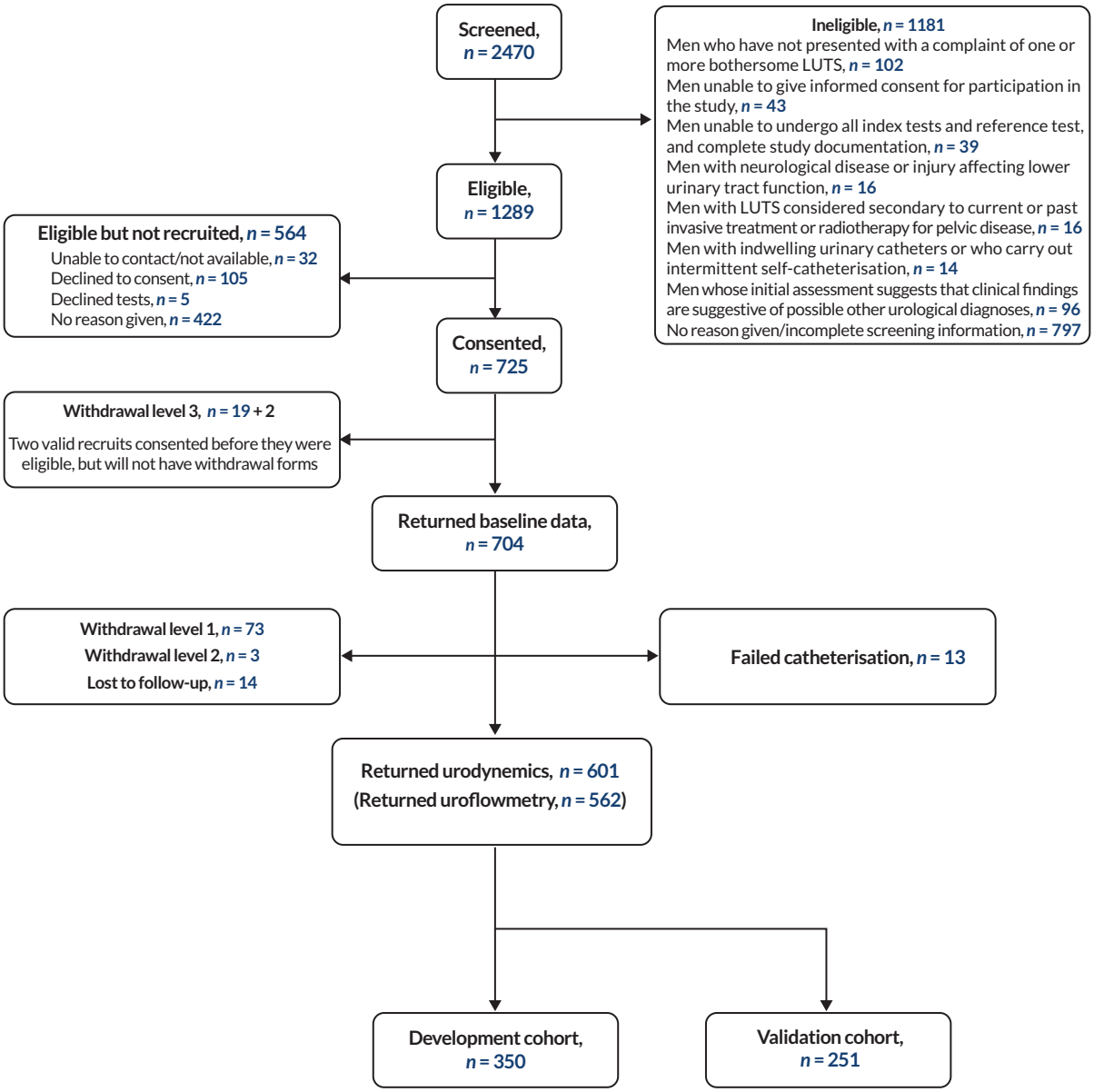


FIGURE 2 Overview of participant recruitment.

Imputation of missing data

The candidate predictors for the development cohort are summarised by the target conditions in *Tables 6–8*. The number of participants with missing data for the candidate predictors is presented in the tables. The multiple imputation procedure produced values for the missing data that were consistent with the non-missing data.

Bladder outlet obstruction

Data from 336 individuals were used to develop the models. For the main analysis, SA1 and SA2, 291 (86.6%), 329 (97.9%) and 241 (71.7%) individuals had complete data for all potential predictors respectively. Data on age, size of prostrate, PSA test result, palpable bladder, voiding symptoms subscore (ICIQ–MLUTS questionnaire), incontinence symptoms subscore (ICIQ–MLUTS questionnaire), storage symptoms subscore (IPSS questionnaire), median maximum flow rate, median voided volume and mean 24-hour frequency were known for all individuals. The percentage of data missing for the candidate predictors in the main analysis and the sensitivity analyses is shown in *Table 6*.

TABLE 3 Reference standard results for each target condition

Target condition		Frequency (%) *n* = 350
BOO	Present	163 (46.6)
	Absent	173 (49.4)
	Indeterminate	14 (4.0)
DU	Present	141 (40.3)
	Absent	193 (55.1)
	Indeterminate	16 (4.6)
DO	Present	253 (72.3)
	Absent	87 (24.9)
	Indeterminate	10 (2.9)

TABLE 4 Cross-classification of participants with, without or indeterminate diagnosis of BOO, DU and DO with urodynamics (reference standard)

	BOO present			BOO absent			BOO indeterminate		
Final diagnosis, *n* (%)	DU present	DU absent	DU indeterminate	DU present	DU absent	DU indeterminate	DU present	DU absent	DU indeterminate
DO present	42 (12.0)	88 (25.1)	2 (0.6)	49 (14.0)	70 (20.0)	–	–	–	2 (0.6)
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
DO absent	13 (3.7)	17 (4.9)	–	36 (10.3)	18 (5.1)	–	–	–	3 (0.9)
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
DO indeterminate	1 (0.3)	–	–	–	–	–	–	–	9 (2.6)
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–

Note
'–' means there was no participants in this particular category.

TABLE 5 Urodynamics-related adverse events (development cohort)

Urodynamics-related adverse events, *n* (%)		Total 3/350
Main diagnosis	Haematuria	1 (0.3)
	Retention of urine	2 (0.6)
Seriousness of the event	Not serious	0 (0.0)
	Resulted in death	0 (0.0)
	Life-threatening	0 (0.0)

continued

TABLE 5 Urodynamics-related adverse events (development cohort) (*continued*)

Urodynamics-related adverse events, *n* (%)		Total 3/350
	Required hospitalisation	3 (0.9)
	Persistent or significant disability/incapacity	0 (0.0)
	Congenital anomaly/birth defect	0 (0.0)
	Other medically important condition	0 (0.0)
	Visible haematuria (blood in urine)/urethral bleeding requiring intervention	0 (0.0)
SAE level	Mild	0 (0.0)
	Moderate	1 (0.3)
	Severe	2 (0.6)

TABLE 6 Summary of candidate predictors for the development of models for BOO

Analysis	Candidate predictor		Diagnosis		
			BOO *n* = 163	No BOO *n* = 173	Total *n* = 350
Main analysis	Age, years (participant demographics)[a,b]	Median (IQR)	69.8 (64.0–74.9)	67.8 (58.9–74.7)	69.0 (61.8–74.9)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Size of prostate (DRE),[a,b] *n* (%)	Normal	42 (25.8)	55 (31.8)	101 (28.9)
		Mild enlargement	44 (27.0)	53 (30.6)	101 (28.9)
		Moderate enlargement	71 (43.6)	61 (35.3)	137 (39.1)
		Gross enlargement	6 (3.9)	4 (2.3)	11 (3.1)
		Missing	0 (0.0)	0 (0.0)	0 (0.0)
	PSA test, ng/ml (blood test)[a,b]	Median (IQR)	2.0 (1.2–3.3)	1.3 (0.6–2.3)	1.6 (0.9–2.8)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Incontinence symptoms subscore (ICIQ–MLUTS questionnaire)[a]	Median (IQR)	4.0 (2.0–7.0)	5.0 (3.0–7.0)	4.0 (3.0–7.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Voiding symptoms subscore (ICIQ–MLUTS questionnaire)[a]	Median (IQR)	8.0 (5.0–10.0)	7.0 (5.0–10.0)	7.5 (5.0–10.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Post-void residual urine, ml (bladder ultrasound)[a]	Median (IQR)	85.0 (36.0–167.0)	52.0 (25.0–117.0)	67.0 (29.0–150.0)
		Missing, *n* (%)	1 (0.6)	2 (1.2)	9 (2.6)

TABLE 6 Summary of candidate predictors for the development of models for BOO (*continued*)

Analysis	Candidate predictor		Diagnosis		
			BOO ***n*** **= 163**	**No BOO** ***n*** **= 173**	**Total** ***n*** **= 350**
	Median maximum flow rate, ml/second (uroflowmetry)[a]	Median (IQR)	9.9 (7.9–12.4)	15.2 (11.5–21.1)	12.1 (9.1–16.6)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Median voided volume, ml (uroflowmetry)[a]	Median (IQR)	170.0 (133.0–219.0)	189.0 (147.0–245.0)	180.0 (139.0–233.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Mean 24-hour frequency (uroflowmetry)[a]	Median (IQR)	7.7 (5.6–9.7)	7.2 (5.6–8.9)	7.4 (5.6–9.2)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Mean nocturia frequency (uroflowmetry)[a]	Median (IQR)	1.6 (1.0–2.4)	1.6 (1.0–2.3)	1.6 (1.0–2.4)
		Missing, *n* (%)	2 (1.2)	2 (1.2)	4 (1.1)
	Mean urgency score (bladder diary)[b]	Median (IQR)	1.5 (1.0–2.1)	1.6 (1.1–2.2)	1.5 (1.1–2.1)
		Missing, *n* (%)	9 (5.5)	11 (6.4)	20 (5.7)
	Mean 24-hour fluid intake, ml (bladder diary)[b]	Median (IQR)	1758.0 (1350.0–2150.0)	1925.0 (1458.5–2334.0)	1830.0 (1412.0–2236.5)
		Missing, *n* (%)	16 (9.8)	17 (9.8)	34 (9.7)
SA2	Storage symptoms subscore (IPSS questionnaire)	Median (IQR)	9.0 (6.0–11.0)	9.0 (6.0–11.0)	9.0 (6.0–11.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Voiding symptoms subscore (IPSS questionnaire)	Median (IQR)	8.0 (4.0–11.0)	7.0 (4.0–11.0)	7.0 (4.0–11.0)
		Missing, *n* (%)	1 (0.6)	0 (0.0)	2 (0.6)
	Palpable bladder (physical examination), *n* (%)	Yes	7 (4.3)	4 (2.3)	11 (3.1)
		No	156 (95.7)	169 (97.7)	339 (96.9)
		Missing	0 (0.0)	0 (0.0)	0 (0.0)
	Maximum flow rate, ml/second (one-off flow test)	Median (IQR)	6.0 (3.0–9.5)	8.3 (5.0–14.2)	7.6 (4.0–11.1)
		Missing, *n* (%)	13 (8.0)	23 (13.3)	36 (10.3)

continued

TABLE 6 Summary of candidate predictors for the development of models for BOO (*continued*)

Analysis	Candidate predictor		Diagnosis		
			BOO *n* = 163	No BOO *n* = 173	Total *n* = 350
	Voided volume, ml (one-off flow test)	Median (IQR)	94.0 (54.0–195.0)	123.5 (62.0–282.0)	114.0 (58.0–238.0)
		Missing, *n* (%)	10 (6.1)	19 (11.0)	29 (8.3)
	Average number of voids (bladder diary)	Median (IQR)	9.3 (7.0–11.3)	9.3 (7.3–11.2)	9.3 (7.3–11.3)
		Missing, *n* (%)	16 (9.8)	17 (9.8)	34 (9.7)
	Nocturnal polyuria score (bladder diary)	Median (IQR)	0.3 (0.3–0.5)	0.4 (0.3–0.4)	0.3 (0.3–0.4)
		Missing, *n* (%)	24 (14.7)	23 (13.3)	48 (13.7)

IQR, interquartile range.
a These variables are candidate predictors in SA1.
b These variables are candidate predictors in SA2.

TABLE 7 Summary of candidate predictors for the development of models for DU

Analysis	Candidate predictor		Diagnosis		
			DU *n* = 141	No DU *n* = 193	Total *n* = 350
Main analysis	Age, years (participant demographics)[a,b]	Median (IQR)	69.2 (63.1–75.1)	68.5 (61.2–74.6)	69.0 (61.8–74.9)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Size of prostate (DRE),[a,b] *n* (%)	Normal	40 (28.4)	56 (29.0)	101 (28.9)
		Mild enlargement	42 (29.8)	55 (28.5)	101 (28.9)
		Moderate enlargement	55 (39.0)	76 (39.4)	137 (39.1)
		Gross enlargement	4 (2.8)	6 (3.1)	11 (3.1)
		Missing	0 (0.0)	0 (0.0)	0 (0.0)
	PSA test, ng/ml (blood test)[a,b]	Median (IQR)	1.6 (1.0–2.6)	1.7 (0.8–3.0)	1.6 (0.9–2.8)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Incontinence symptoms subscore (ICIQ–MLUTS questionnaire)[a]	Median (IQR)	4.0 (3.0–7.0)	5.0 (3.0–7.0)	4.0 (3.0–7.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)

TABLE 7 Summary of candidate predictors for the development of models for DU (*continued*)

Analysis	Candidate predictor		Diagnosis		
			DU *n* = 141	No DU *n* = 193	Total *n* = 350
	Voiding symptoms subscore (ICIQ–MLUTS questionnaire)[a]	Median (IQR)	8.0 (5.0–10.0)	7.0 (5.0–10.0)	7.5 (5.0–10.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Post-void residual urine, ml (bladder ultrasound)[a]	Median (IQR)	89.5 (35.0–206.0)	52.0 (24.0–118.0)	67.0 (29.0–150.0)
		Missing, *n* (%)	1 (0.7)	2 (1.0)	9 (2.6)
	Median maximum flow rate, ml/second (uroflowmetry)[a]	Median (IQR)	10.9 (8.5–13.9)	13.5 (9.8–18.4)	12.1 (9.1–16.6)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Median voided volume, ml (uroflowmetry)[a]	Median (IQR)	168.0 (133.0–203.0)	195.0 (146.0–249.0)	180.0 (139.0–233.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Mean 24-hour frequency (uroflowmetry)[a]	Median (IQR)	7.4 (5.8–9.2)	7.4 (5.5–9.2)	7.4 (5.6–9.2)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Mean nocturia frequency (uroflowmetry)[a]	Median (IQR)	1.5 (1.0–2.5)	1.7 (1.0–2.3)	1.6 (1.0–2.4)
		Missing, *n* (%)	1 (0.7)	3 (1.6)	4 (1.1)
	Mean urgency score (bladder diary)[b]	Median (IQR)	1.5 (1.1–2.1)	1.5 (1.1–2.1)	1.5 (1.1–2.1)
		Missing, *n* (%)	13 (9.2)	7 (3.6)	20 (5.7)
	Mean 24-hour fluid intake, ml (bladder diary)[b]	Median (IQR)	1717.0 (1349.0–2110.0)	1920.0 (1450.0–2341.0)	1830.0 (1412.0–2236.5)
		Missing, *n* (%)	15 (10.6)	18 (9.3)	34 (9.7)
SA2	Storage symptoms subscore (IPSS questionnaire)	Median (IQR)	9.0 (7.0–11.0)	8.0 (6.0–11.0)	9.0 (6.0–11.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Voiding symptoms subscore (IPSS questionnaire)	Median (IQR)	8.0 (4.0–12.0)	7.0 (4.0–10.0)	7.0 (4.0–11.0)
		Missing, *n* (%)	0 (0.0)	1 (0.5)	2 (0.6)
	Palpable bladder (physical examination), *n* (%)	Yes	6 (4.3)	5 (2.6)	11 (3.1)

continued

TABLE 7 Summary of candidate predictors for the development of models for DU (*continued*)

Analysis	Candidate predictor		Diagnosis: DU *n* = 141	Diagnosis: No DU *n* = 193	Total *n* = 350
		No	135 (95.7)	188 (97.4)	339 (96.9)
		Missing	0 (0.0)	0 (0.0)	0 (0.0)
	Maximum flow rate, ml/second (one-off flow test)	Median (IQR)	6.5 (3.7–9.2)	8.0 (4.0–13.5)	7.6 (4.0–11.1)
		Missing, *n* (%)	18 (12.8)	18 (9.3)	36 (10.3)
	Voided volume, ml (one-off flow test)	Median (IQR)	114.0 (51.0–224.0)	113.5 (66.0–243.0)	114.0 (58.0–238.0)
		Missing, *n* (%)	14 (9.9)	15 (7.8)	29 (8.3)
	Average number of voids (bladder diary)	Median (IQR)	9.0 (7.3–11.3)	9.3 (7.3–11.3)	9.3 (7.3–11.3)
		Missing, *n* (%)	14 (9.9)	19 (9.8)	34 (9.7)
	Nocturnal polyuria score (bladder diary)	Median (IQR)	0.4 (0.2–0.4)	0.3 (0.3–0.4)	0.3 (0.3–0.4)
		Missing, *n* (%)	17 (12.1)	29 (15.0)	48 (13.7)

IQR, interquartile range.
a These variables are candidate predictors in SA1.
b These variables are candidate predictors in SA2.

TABLE 8 Summary of candidate predictors for the development of models for DO

Analysis	Candidate predictor		Diagnosis: DO *n* = 253	Diagnosis: No DO *n* = 87	Total *n* = 350
Main analysis	Age, years (participant demographics)[a,b]	Median (IQR)	70.2 (64.0–75.6)	64.5 (56.6–70.7)	69.0 (61.8–74.9)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Size of prostate (DRE),[a,b] *n* (%)	Normal	75 (29.6)	24 (27.6)	101 (28.9)
		Mild enlargement	70 (27.7)	26 (29.9)	101 (28.9)
		Moderate enlargement	101 (39.9)	33 (37.9)	137 (39.1)
		Gross enlargement	7 (2.8)	4 (4.6)	11 (3.1)
		Missing	0 (0.0)	0 (0.0)	0 (0.0)
	PSA test, ng/ml (blood test)[a,b]	Median (IQR)	1.8 (0.9–2.9)	1.3 (0.6–2.3)	1.6 (0.9–2.8)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)

TABLE 8 Summary of candidate predictors for the development of models for DO (*continued*)

Analysis	Candidate predictor		Diagnosis		
			DO *n* = 253	No DO *n* = 87	Total *n* = 350
	Incontinence symptoms subscore (ICIQ–MLUTS questionnaire)[a]	Median (IQR)	5.0 (3.0–7.0)	4.0 (2.0–6.0)	4.0 (3.0–7.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Voiding symptoms subscore (ICIQ–MLUTS questionnaire)[a]	Median (IQR)	7.0 (5.0–10.0)	8.0 (5.0–11.0)	7.5 (5.0–10.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Post-void residual urine, ml (bladder ultrasound)[a]	Median (IQR)	65.0 (28.0–146.0)	77.0 (35.0–174.0)	67.0 (29.0–150.0)
		Missing, *n* (%)	3 (1.2)	4 (4.6)	9 (2.6)
	Median maximum flow rate, ml/second (uroflowmetry)[a]	Median (IQR)	11.7 (8.9–16.2)	13.9 (9.9–17.1)	12.1 (9.1–16.6)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Median voided volume, ml (uroflowmetry)[a]	Median (IQR)	172.0 (136.0–223.0)	203.0 (155.0–255.0)	180.0 (139.0–233.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Mean 24-hour frequency (uroflowmetry)[a]	Median (IQR)	7.5 (5.9–9.4)	6.9 (5.1–8.8)	7.4 (5.6–9.2)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Mean nocturia frequency (uroflowmetry)[a]	Median (IQR)	1.7 (1.1–2.3)	1.3 (1.0–2.4)	1.6 (1.0–2.4)
		Missing, *n* (%)	4 (1.6)	0 (0.0)	4 (1.1)
	Mean urgency score (bladder diary)[b]	Median (IQR)	1.6 (1.1–2.2)	1.3 (1.0–1.9)	1.5 (1.1–2.1)
		Missing, *n* (%)	13 (5.1)	6 (6.9)	20 (5.7)
	Mean 24-hour fluid intake, ml (bladder diary)[b]	Median (IQR)	1783.5 (1408.0–2187.0)	1953.0 (1411.0–2383.0)	1830.0 (1412.0–2236.5)
		Missing, *n* (%)	23 (9.1)	10 (11.5)	34 (9.7)
SA2	Storage symptoms subscore (IPSS questionnaire)	Median (IQR)	9.0 (7.0–11.0)	8.0 (4.0–10.0)	9.0 (6.0–11.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Voiding symptoms subscore (IPSS questionnaire)	Median (IQR)	7.0 (4.0–11.0)	7.0 (5.0–11.0)	7.0 (4.0–11.0)

continued

TABLE 8 Summary of candidate predictors for the development of models for DO (*continued*)

Analysis	Candidate predictor		Diagnosis		
			DO *n* = 253	No DO *n* = 87	Total *n* = 350
		Missing, *n* (%)	1 (0.4)	0 (0.0)	2 (0.6)
	Palpable bladder (physical examination), *n* (%)	Yes	9 (3.6)	2 (2.3)	11 (3.1)
		No	244 (96.4)	85 (97.7)	339 (96.9)
		Missing	0 (0.0)	0 (0.0)	0 (0.0)
	Maximum flow rate, ml/second (one-off flow test)	Median (IQR)	7.0 (3.7–11.0)	8.0 (5.0–12.0)	7.6 (4.0–11.1)
		Missing, *n* (%)	31 (12.3)	5 (5.7)	36 (10.3)
	Voided volume, ml (one-off flow test)	Median (IQR)	99.5 (56.0–209.0)	137.0 (62.0–311.0)	114.0 (58.0–238.0)
		Missing, *n* (%)	27 (10.7)	2 (2.3)	29 (8.3)
	Average number of voids (bladder diary)	Median (IQR)	9.3 (7.3–11.3)	9.0 (7.0–10.7)	9.3 (7.3–11.3)
		Missing, *n* (%)	23 (9.1)	10 (11.5)	34 (9.7)
	Nocturnal polyuria score (bladder diary)	Median (IQR)	0.4 (0.3–0.5)	0.3 (0.2–0.4)	0.3 (0.3–0.4)
		Missing, *n* (%)	34 (13.4)	13 (14.9)	48 (13.7)

IQR, interquartile range.
a These variables are candidate predictors in SA1.
b These variables are candidate predictors in SA2.

In the main analysis, 13.4% (*n* = 45) of individuals had at least one candidate predictor missing. Missing values for post-void residual urine, mean nocturnal frequency, mean urgency score and mean 24-hour fluid intake were imputed. Fourteen imputed data sets were created by replacing missing values with simulated values from a set of imputation models constructed from all potential candidate predictors (age, size of prostrate, PSA test result, voiding symptoms subscore, incontinence symptoms subscore, post-void residual urine, median maximum flow rate, median voided volume, mean 24-hour frequency, mean nocturia frequency, mean urgency score and mean 24-hour fluid intake) and outcome variable (binary outcome for BOO – present or absent).

For SA1, 2.1% (*n* = 7) of individuals had at least one candidate predictor missing. Missing values for post-void residual urine and mean nocturnal frequency were imputed. Ten imputed data sets were created. For SA2, 28.3% (*n* = 95) of individuals had at least one candidate predictor missing and so 29 imputed data sets were created. Missing values for voiding symptoms subscore, voided volume, maximum flow rate, average number of voids, mean urgency score, mean 24-hour fluid intake and nocturnal polyuria score were imputed.

Detrusor underactivity

Data from 334 individuals were used to develop the models. For the main analysis, SA1 and SA2, 289 (86.5%), 327 (97.9%) and 240 (71.9%) individuals had complete data for all potential predictors, respectively. Data on age, size of prostrate, PSA test result, palpable bladder, voiding symptoms subscore (ICIQ–MLUTS questionnaire), incontinence symptoms subscore (ICIQ–MLUTS questionnaire), storage symptoms subscore (IPSS questionnaire), median maximum

flow rate, median voided volume and mean 24-hour frequency were known for all individuals. The percentage of data missing for the candidate predictors in the main analysis and the sensitivity analyses is shown in *Table 7*.

In the main analysis, 13.5% (*n* = 45) individuals had at least one candidate predictor missing and so 14 imputed data sets were created. Missing values for post-void residual urine, mean nocturnal frequency, mean urgency score and mean 24-hour fluid intake were imputed. For SA1, 2.1% (*n* = 7) of individuals had at least one candidate predictor missing and so 10 imputed data sets were created. Missing values for post-void residual urine and mean nocturnal frequency were imputed. For SA2, 28.1% (*n* = 94) of individuals had at least one candidate predictor missing and 29 imputed data sets were created. Missing values for voiding symptoms subscore, voided volume, maximum flow rate, average number of voids, mean urgency score, mean 24-hour fluid intake and nocturnal polyuria score were imputed.

Detrusor overactivity

Data from 340 individuals were used to develop the models. For the main analysis, SA1 and SA2, 292 (85.9%), 329 (96.8%) and 246 (72.4%) individuals had complete data for all potential predictors, respectively. Data on age, size of prostrate, PSA test result, palpable bladder, voiding symptoms subscore (ICIQ–MLUTS questionnaire), incontinence symptoms subscore (ICIQ–MLUTS questionnaire), storage symptoms subscore (IPSS questionnaire), median maximum flow rate, median voided volume and mean 24-hour frequency were known for all individuals. The percentage of data missing for the candidate predictors in the main analysis and the sensitivity analyses is shown in *Table 8*.

In the main analysis, 14.1% (*n* = 48) of individuals had at least one candidate predictor missing and so 15 imputed data sets were created. Missing values for post-void residual urine, mean nocturnal frequency, mean urgency score and mean 24-hour fluid intake were imputed. For SA1, 3.2% (*n* = 11) of individuals had at least one candidate predictor missing and 10 imputed data sets were created. Missing values for post-void residual urine and mean nocturnal frequency were imputed. For SA2, 27.6% (*n* = 94) of individuals had at least one candidate predictor missing and 28 imputed data sets were created. Missing values for voiding symptoms subscore, voided volume, maximum flow rate, average number of voids, mean urgency score, mean 24-hour fluid intake and nocturnal polyuria score were imputed.

Model specification

Separate logistic regression models were built for BOO, DO and DU. The apparent coefficients and odds ratios for each selected predictor in the models for BOO, DU and DO from the main analysis are provided in *Tables 9–13*. The apparent coefficient and odds ratio values for each predictor in the models for BOO, DU and DO from the sensitivity analyses are provided in *Appendix 3*, *Tables 40–45*.

Main analysis

Age, PSA result, voiding symptoms subscore, median maximum flow rate, median voided volume and mean 24-hour frequency were the six predictors in the BOO model 1. Due to the extreme values for the coefficient, odds ratio and CIs for the age fractional polynomial functions (see *Table 9*), this predictor was re-assessed with a single fractional polynomial term in BOO model 2 (see *Table 10*). Age as a linear term was not statistically significant ($p = 0.238$), and there was little difference in the coefficient and odds ratio values of the remaining predictors in this model in comparison with BOO model 1.

The fractional polynomial functions for mean 24-hour frequency and voiding symptoms subscore in BOO model 2 are plotted in *Figure 32*. The odds of BOO appears to decrease as the mean 24-hour frequency increases to 12 ml, after which it starts to increase and increases more rapidly at higher volumes. As the voiding symptoms subscore increases, the odds of BOO appears to increase between the score of 0 and 3, after which it starts to decrease and decreases less rapidly at higher scores. Following clinical input, the fractional polynomial function for voiding symptoms subscore was considered implausible because the odds of BOO was expected to increase for each unit increase in the score. The fractional polynomial terms were unlikely to be modelling genuine effects and were more likely to be influenced by fewer participants having this target condition with increasing voiding symptoms subscore, that is overfitting. Thus,

TABLE 9 Apparent model coefficients and odds ratios and optimism-corrected coefficients for predictors included in the model for BOO (model 1)

Predictors and intercept	Apparent			Optimism-corrected coefficient
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value	
Age_1 (years)	−815.23 (−1411.12 to −219.35)	0.00 (0.00 to 5.48e-96)	0.007	−698.18
Age_2 (years)[a]	644.29 (193.99 to 1094.60)	6.5e + 279 (1.77e+ 84)	0.005	551.78
PSA result (ng/ml)	0.19 (−0.01 to 0.39)	1.20 (0.99 to 1.47)	0.066	0.16
Voiding symptoms subscore_1 (ICIQ–MLUTS questionnaire)	2.73 (1.08 to 4.37)	15.26 (2.96 to 78.76)	0.001	2.33
Voiding symptoms subscore_2 (ICIQ–MLUTS questionnaire)	1.34 (0.48 to 2.20)	3. 82 (1.62 to 9.02)	0.002	1.15
Median maximum flow rate (ml/second)	−0.38 (−0.47 to −0.28)	0.69 (0.62 to 0.76)	0.001	−0.32
Median voided volume (ml)	0.011 (0.005 to 0.016)	1.011 (1.005 to 1.016)	0.001	0.009
Mean 24-hour frequency_1 (voids per day)	−1.01 (−1.85 to −0.17)	0.36 (0.16 to 0.84)	0.018	−0.86
Mean 24-hour frequency_2 (voids per day)	2.09 (0.38 to 3.80)	8.09 (1.46 to 44.66)	0.017	1.79
Intercept	−8.71 (−14.77 to −2.64)	0.0002 (3.83e-07 to 0.07)	0.005	−7.43

$$Age\_1 = \left(\frac{age}{10}\right)^{-2}$$

$$Age\_2 = \left(\frac{age}{10}\right)^{-2} \ln\left(\frac{age}{10}\right)$$

$$Voiding\ symptoms\ \text{subscore\_1} = \left(\frac{Vss + 1}{10}\right)^{-1}$$

$$Voiding\ symptoms\ \text{subscore\_2} = \left(\frac{Vss + 1}{10}\right)^{-1} \ln\left(\frac{Vss + 1}{10}\right)$$

$$Mean\ \text{24-hour}\ frequency\_1 = \left(\frac{MF}{10}\right)^{3}$$

$$Mean\ \text{24-hour}\ frequency\_2 = \left(\frac{MF}{10}\right)^{3} \ln\left(\frac{MF}{10}\right)$$

MF, mean 24-hour frequency; *VSS*, voiding symptoms subscore.
a Upper bound of the CI is extremely large.

Note
Optimism-corrected coefficients multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.86.

voiding symptoms subscore was re-assessed as a single fractional polynomial term in BOO model 3 (see *Table 11*). Voiding symptoms subscore as a linear term ($p = 0.052$) reduced the model with four other predictors: age, PSA result, median maximum flow rate and median voided volume as linear terms. Mean 24-hour frequency became statistically non-significant. There was little difference in the coefficient and odds ratio values of the four additional predictors in BOO model 3 in comparison with BOO model 2.

TABLE 10 Apparent model coefficients and odds ratios and optimism-corrected coefficients for predictors included in the model for BOO (model 2 with age as a linear term)

Predictors and intercept	Apparent			Optimism-corrected coefficient
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value	
Age (years)	−0.02 (−0.05 to 0.01)	0.98 (0.95 to 1.01)	0.238	−0.02
PSA test result (ng/ml)	0.18 (−0.01 to 0.37)	1.20 (0.99 to 1.45)	0.069	0.16
Voiding symptoms subscore_1 (ICIQ–MLUTS questionnaire)	2.62 (1.12 to 4.13)	13.76 (3.05 to 62.05)	0.001	2.28
Voiding symptoms subscore_2 (ICIQ–MLUTS questionnaire)	1.28 (0.51 to 2.05)	3.60 (1.67 to 7.77)	0.001	1.11
Median maximum flow rate (ml/second)	−0.36 (−0.46 to −0.27)	0.70 (0.63 to 0.77)	0.001	−0.32
Median voided volume (ml)	0.010 (0.005 to 0.015)	1.01 (1.00 to 1.02)	0.001	0.009
Mean 24-hour frequency_1 (voids per day)	−0.89 (−1.63 to −0.14)	0.41 (0.20 to 0.87)	0.020	−0.77
Mean 24-hour frequency_2 (voids per day)	1.82 (0.45 to 3.18)	6.17 (1.57 to 24.15)	0.009	1.58
Intercept	1.34 (−1.26 to 3.94)	3.81 (0.28 to 51.17)	0.313	1.18

$$\textit{Voiding symptoms sub-score\_1} = \left(\frac{Vss + 1}{10}\right)^{-1}$$

$$\textit{Voiding symptoms sub-score\_2} = \left(\frac{Vss + 1}{10}\right)^{-1} \ln\left(\frac{Vss + 1}{10}\right)$$

$$\textit{Mean 24-hour frequency\_1} = \left(\frac{MF}{10}\right)^{3}$$

$$\textit{Mean 24-hour frequency\_2} = \left(\frac{MF}{10}\right)^{3} \ln\left(\frac{MF}{10}\right)$$

MF, mean 24-hour frequency; *VSS*, voiding symptoms subscore.

Note
Optimism-corrected coefficients multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.87.

TABLE 11 Apparent model coefficients and odds ratios and optimism-corrected coefficients for predictors included in the model for BOO (model 3 with age and voiding symptoms subscore as a linear term)

Predictors and intercept	Apparent			Optimism-corrected coefficient
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value	
Age (years)	−0.02 (−0.05 to 0.01)	0.98 (0.95 to 1.01)	0.172	−0.02
PSA test result (ng/ml)	0.19 (0.00 to 0.38)	1.21 (1.00 to 1.46)	0.048	0.16
Voiding symptoms subscore (ICIQ–MLUTS questionnaire)	−0.07 (−0.14 to 0.00)	0.93 (0.87 to 1.00)	0.052	−0.06
Median maximum flow rate (ml/second)	−0.34 (−0.43 to −0.25)	0.71 (0.65 to 0.78)	0.001	−0.30
Median voided volume (ml)	0.010 (0.005 to 0.015)	1.010 (1.005 to 1.015)	0.001	0.009
Intercept	3.92 (1.43 to 6.42)	50.6 (4.17 to 614.13)	0.002	3.43

Note
Optimism-corrected coefficients multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.87.

Median maximum flow rate and post-void residual volume were the two predictors in the DU model. The odds ratio of DU is estimated to increase by 0.3% (95% CI 0.1% to 0.5%) for each unit increase in the post-void residual urine and decrease by 8.0% (95% CI 4.0% to 12.0%) for each unit increase in median maximum flow rate (see *Table 12*).

For the DO model, the three predictors were age, post-void residual volume and median voided volume. The odds of DO is estimated to increase by 5.0% (95% CI 3.0% to 8.0%) for each year increase in age and decrease by 0.3% (95% CI 0.1% to 0.5%) for each unit increase in post-void residual urine (see *Table 13*). The fractional polynomial function for median voided volume in the DO model is plotted in *Figure 33*. The odds of DO appears to decrease as the median voided volume increases to 350 ml, after which it starts to increase and increases more rapidly at higher volumes.

Sensitivity analysis 1

For this sensitivity analysis, the mean urgency score and the mean 24-hour fluid intake from the bladder diary were excluded as candidate predictors during model development, because these predictors may be difficult to obtain in practice. Thus, the sensitivity analysis assessed the adequacy of the models developed for BOO, DU and DO with fewer tests.

TABLE 12 Apparent model coefficients and odds ratios and optimism-corrected coefficients for predictors included in the main model for DU

Predictors and intercept	Apparent			Optimism-corrected coefficient
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value	
Post-void residual urine (ml)	0.003 (0.001 to 0.005)	1.003 (1.001 to 1.005)	0.001	0.002
Median maximum flow rate (ml/second)	−0.08 (−0.12 to −0.04)	0.92 (0.88 to 0.96)	0.002	−0.06
Intercept	0.42 (−0.23 to 1.07)	1.52 (0.79 to 2.90)	0.209	0.26

Note
Optimism-corrected coefficients multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.77.

TABLE 13 Apparent model coefficients and odds ratios and optimism-corrected coefficients for predictors included in the main model for DO

Predictors and intercept	Apparent			Optimism-corrected coefficient
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value	
Age (years)	0.05 (0.03 to 0.08)	1.05 (1.03 to 1.08)	0.001	0.04
Post-void residual urine (ml)	−0.003 (−0.005 to −0.001)	0.997 (0.995 to 0.999)	0.004	−0.002
Median voided volume_1 (ml)	−0.19 (−0.32 to −0.07)	0.82 (0.73 to 0.93)	0.002	−0.15
Median voided volume_2 (ml)	0.13 (0.04 to 0.21)	1.13 (1.04 to 1.23)	0.004	0.10
Intercept	−1.32 (−3.07 to 0.44)	0.27 (0.05 to 1.55)	0.142	−0.81

$$\textit{Median voided volume\_1} = \left(\frac{MVV}{100}\right)^3$$

$$\textit{Median voided volume\_2} = \left(\frac{MVV}{100}\right)^3 \ln\left(\frac{MVV}{100}\right)$$

Note
Optimism-corrected coefficients multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.78.

The logistic regression model for BOO, DU and DO from the sensitivity analysis remained the same as BOO model 1, DU model and DO model from the main analysis, respectively (see *Tables 40–42*, *Appendix 3*), which suggests that the models are robust.

Sensitivity analysis 2

This sensitivity analysis assessed the robustness of the models to alternative test methods. The index tests for incontinence symptoms subscore, voiding symptoms subscore, post-void residual urine, median maximum flow rate, median voided volume, mean 24-hour frequency and mean nocturia frequency were replaced with other tests as candidate predictors, as shown in *Table 2*.

Age and PSA results remained as predictors for BOO outcome. Additional variables included maximum flow rate and voided volume, which respectively replaced median maximum flow rate and median voided volume as candidate predictors from the main analysis (see *Table 43*, *Appendix 3*). Age remained as a predictor for DO and included storage symptoms subscore from the IPSS questionnaire (see *Table 45*, *Appendix 3*). Mean 24-hour fluid intake was a candidate predictor in the main analysis but was not included in the DU model. However, in SA2, this variable was a predictor for DU along with maximum flow rate (which replaced the median maximum flow rate from the main analysis) and voided volume (see *Table 44*, *Appendix 3*).

Model performance

Main analysis

The apparent calibration slope and c-index of the models for BOO, DU and DO are reported in *Table 14*. The apparent calibration slope for all the models was 1.00 (see *Table 14*), suggesting perfect calibration as expected, since the models were developed using the same development cohort used to estimate the predicted probabilities.

TABLE 14 Performance statistics of models for BOO, DU and DO

		Discrimination			Calibration			
Outcome	Analysis	Apparent c-index (95% CI)	Optimism c-index (95% CI)	Optimism-corrected c-index	Apparent calibration slope (95% CI)	Optimism calibration slope (95% CI)	Optimism-corrected calibration slope	Calibration plot of the apparent model
BOO, *n* = 342	BOO model 1, main	0.85 (0.81 to 0.89)	0.019 (0.015 to 0.023)	0.83	1.00 (0.76 to 1.24)	0.14 (0.11 to 0.18)	0.86	*Figure 3*
	BOO model 2, main	0.84 (0.80 to 0.88)	0.020 (0.016 to 0.024)	0.82	1.00 (0.76 to 1.24)	0.13 (0.10 to 0.16)	0.87	*Figure 3*
	BOO model 3, main	0.82 (0.77 to 0.86)	0.020 (0.015 to 0.025)	0.80	1.00 (0.75 to 1.25)	0.13 (0.09 to 0.17)	0.87	*Figure 3*
	SA1	0.85 (0.81 to 0.89)	–	–	1.00 (0.76 to 1.24)	–	–	*Figure 34*, *Appendix 4*
	SA2	0.71 (0.65 to 0.77)	0.022 (0.017 to 0.026)	0.69	1.00 (0.63 to 1.36)	0.17 (0.14 to 0.20)	0.83	*Figure 35*, *Appendix 4*
DU, *n* = 340	Main	0.68 (0.62 to 0.73)	0.034 (0.030 to 0.039)	0.64	1.00 (0.63 to 1.37)	0.23 (0.20 to 0.26)	0.77	*Figure 4*
	SA1	0.68 (0.62 to 0.73)	–	–	1.00 (0.63 to 1.37)	–	–	*Figure 34*, *Appendix 4*
	SA2	0.66 (0.60 to 0.73)	0.032 (0.027 to 0.037)	0.63	1.00 (0.58 to 1.42)	0.20 (0.18 to 0.23)	0.80	*Figure 35*, *Appendix 4*

TABLE 14 Performance statistics of models for BOO, DU and DO (*continued*)

Outcome	Analysis	Discrimination			Calibration			
		Apparent c-index (95% CI)	Optimism c-index (95% CI)	Optimism-corrected c-index	Apparent calibration slope (95% CI)	Optimism calibration slope (95% CI)	Optimism-corrected calibration slope	Calibration plot of the apparent model
DO, *n* = 346	Main	0.72 (0.66 to 0.78)	0.041 (0.036 to 0.047)	0.67	1.00 (0.65 to 1.35)	0.22 (0.19 to 0.25)	0.78	*Figure 4*
	SA1	0.72 (0.65 to 0.78)	–	–	1.00 (0.65 to 1.35)	–	–	*Figure 34*, *Appendix 4*
	SA2	0.69 (0.63 to 0.76)	0.042 (0.036 to 0.048)	0.65	1.00 (0.62 to 1.38)	0.26 (0.22 to 0.29)	0.74	*Figure 35*, *Appendix 4*

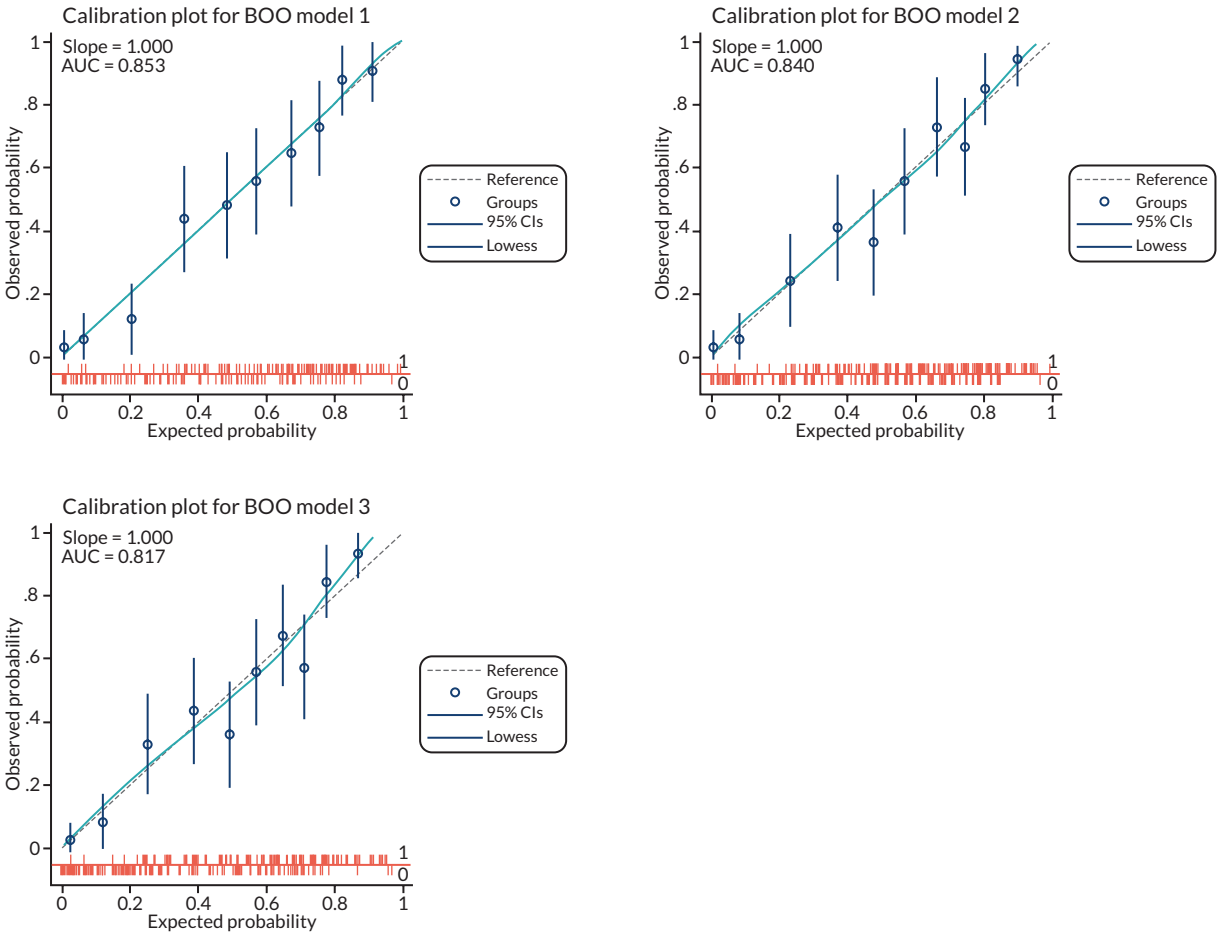


FIGURE 3 Calibration plots comparing observed and predicted probabilities for BOO by decile of risk (main analysis), based on the development cohort. AUC, area under the curve.

The apparent c-index (95% CI) of BOO model 1, BOO model 2 and BOO model 3 were 0.85 (95% CI 0.81 to 0.89), 0.84 (95% CI 0.80 to 0.88) and 0.82 (95% CI 0.77 to 0.86), respectively, suggesting strong discriminative performance of the models. The apparent c-index (95% CI) of the models for DU and DO were 0.68 (95% CI 0.62 to 0.73) and 0.72 (95% CI 0.66 to 0.78), thereby suggesting that these models have moderate discriminative ability. There was large uncertainty around the c-index and calibration slope of the DU and DO models, as seen in the width of the 95% CI. This is likely to be due to the relatively small number of participants observed with these outcomes.

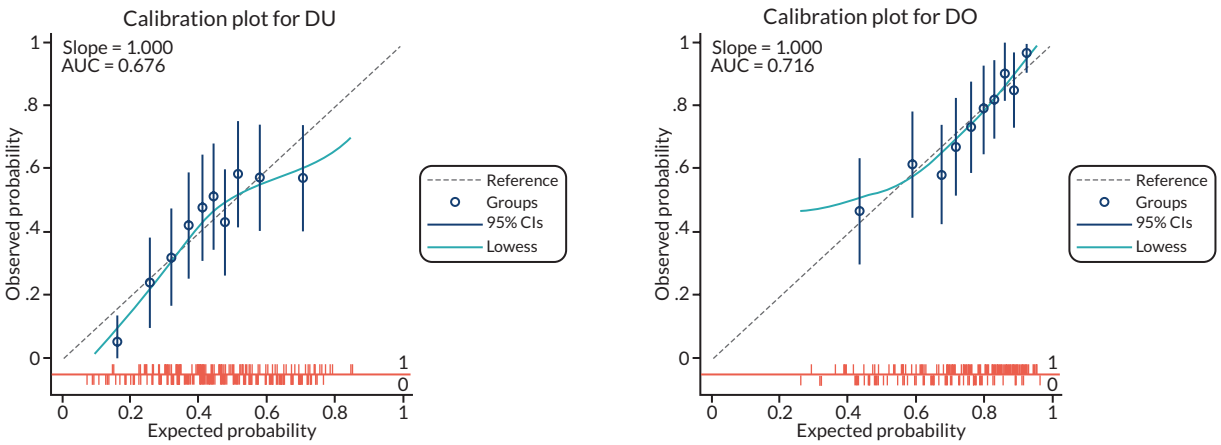


FIGURE 4 Calibration plots comparing observed and predicted probabilities for DU and DO by decile of risk (main analysis), based on the development cohort. AUC, area under the curve.

The calibration plots of the BOO models from the main analysis are shown in *Figure 3*. The apparent calibration was good for BOO model 1 and BOO model 2 and adequate for BOO model 3. All risk groups overlapped the reference line in BOO models 1, 2 and 3. The deciles were spread across the reference line in all three models indicating risk of the outcome varied among participants. The lowess marginally drifted away from the reference line at the top right of the plot for BOO model 1, suggesting there was miscalibration at individual level in participants with higher risk, but this was very slight. The lowess drifted slightly further away from the reference line in BOO model 2 due to the group of participants with high risk not being as close to the reference line, but again, this was very slight. Unlike BOO model 1 and BOO model 2, the lowess for BOO model 3 did not overlay the reference line in the middle and bottom left of the plot, but remained close, suggesting slight miscalibration at individual level in participants with low and moderate risk.

In contrast, the apparent calibration was poor in the DU model (*Figure 4*). Nine groups of participants with moderate risk overlapped the reference line in the middle. The deciles were clustered in the bottom left and middle of the plot indicating participants had low to moderate risk of the outcome. The lowess drifted away from the reference line at the bottom left and top right of the plot, suggesting there was miscalibration at individual level in participants with lower and higher risk. There were less data at the higher risk probabilities as indicated by the spike plot at the bottom and there were fewer participants with the outcome in the low-risk groups when compared with the groups of participants with moderate risk.

The apparent calibration was moderate in the DO model (see *Figure 4*). All groups of participants overlapped the reference line. The deciles were clustered in the middle and top right of the plot indicating participants had moderate to high risk of the outcome. The lowess drifted considerably away from the reference line in the middle of the plot, suggesting there was high miscalibration at the individual level in participants with low and moderate risk. However, there was less data at the low and moderate risk probabilities as indicated by the spike plot at the bottom.

Sensitivity analyses

The models for BOO, DU and DO from SA1 were the same as BOO model 1, DU model and DO model from the main analysis. Thus, there was little to no difference in the apparent c-index, apparent calibration slope and calibration plot (see *Appendix 4*, *Figure 34*). The BOO model from SA2 had reduced discriminative performance in comparison with the BOO models from the main analysis with an apparent c-index of 0.71 (95% CI 0.65 to 0.77), presented in *Table 14*. The calibration plot *Figure 35* (see *Appendix 4*) shows the lowess strays considerably from the reference line at the top right.

In contrast, the apparent c-index for the DU and DO models from SA2 were only slightly reduced, when compared to their respective models from the main analysis. The apparent c-index for the DU and DO models in SA2 were 0.66 (95% CI 0.60 to 0.73) and 0.69 (95% CI 0.63 to 0.76), respectively. The improvement in model performance was seen in the

calibration plots, shown in *Figure 35* (see *Appendix 4*). The lowess in the plots for SA2 appear to diverge slightly away from the reference line rather than considerably in comparison with the models from the main analysis (*Figure 4*).

Internal validation

The optimism-corrected c-index and calibration slope for each model in the main analysis and SA2 are presented in *Table 14*. Internal validation was not performed on models built from SA1, since the candidate predictors and fractional polynomial terms selected in these models were the same as the models from the main analysis.

For the remaining models, there was small optimism adjustment to the c-index, indicating good internal performance in terms of discrimination. However, the calibration slope for each model suggested a large amount of shrinkage was required to adjust the models for overfitting, that is the predictions were too low for low probabilities and the predictions were too high for high probabilities. The overfitting was likely introduced through the variable selection procedure in the bootstrap method. The models for BOO required less optimism adjustment to the c-index and calibration slope than the DU and DO models, suggesting a marginally better internal performance.

The optimism-corrected calibration slope from each model was used as the shrinkage factor in order to adjust for the overfitting. The new optimism-corrected coefficients for each predictor in the BOO, DU and DO models are provided in *Tables 9–13* for the main analysis and in *Appendix 3* for the SA2. *Box 1* details how to obtain the predicted probability of having BOO, DU and DO from the prediction models. Predictive probabilities were obtained in this manner for each participant in the validation cohort to evaluate the performance of the diagnostic models.

Discussion

Diagnostic prediction models were developed for the detection of BOO, DU and DO using data collated from 350 participants across 55 GP practices in Bristol, Newcastle upon Tyne and Wales. The models were internally validated using the bootstrap method, which replicated the model development process, including the variable selection

BOX 1 Using the diagnostic prediction models

The logistic regression models can be expressed as:

$$\log\left(\frac{P_{LUTS}}{1-P_{LUTS}}\right) = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k$$

where P_{LUTS} denotes the probability of a man having BOO, DU or DO, α denotes the model intercept and β_i denotes the coefficient for predictor x_i. Alternatively, the model can be represented in the following form for calculating the predictive probability of a man having BOO, DU or DO:

$$P_{LUTS} = \frac{e^{\alpha+\beta_1 x_1+\ldots+\beta_k x_k}}{1+e^{\alpha+\beta_1 x_1+\ldots+\beta_k x_k}}$$

For example, the diagnostic model for DU from the main analysis after correcting for optimism can be expressed as:

$$\log\left(\frac{P_{DU}}{1-P_{DU}}\right) = 0.26 + 0.002PVR - 0.06MMF$$

Then to calculate the predictive probability of a man having DU, the model can be presented as:

$$P_{DU} = \frac{e^{0.26+0.002PVR-0.06MMF}}{1+e^{0.26+0.002PVR-0.06MMF}}$$

For example, a man with post-void residual urine of 85 ml and median maximum flow rate of 15.2 ml/second would have a predicted probability of 0.36 for having DU.

procedure. The models for BOO showed strong discriminative performance, whereas the models for DU and DO provided a moderate degree of discrimination. All models from the main analysis and sensitivity analyses appeared to be overfitted and were recalibrated using shrinkage factors from the internal validation.

The candidate predictors considered in the logistic regression models were assessed in an efficient way by avoiding categorisation of continuous variables and employing multiple fractional polynomial functions in order to keep as much predictive information as possible. Furthermore, MICE was used to impute missing values to avoid bias and make best use of available data. Complete data were available for most predictors. Data were largely missing for the candidate predictors collected from the one-off flow test (maximum flow rate and voided volume) and bladder diary (mean urgency score, mean 24-hour fluid intake, average number of voids and nocturnal polyuria score). The BOO and DU models from SA2 had maximum flow rate, voided volume or mean 24-hour fluid intake as predictors. The limited availability of these predictors raises a concern about the utility of the model in medical practice. For this reason, BOO model 3 and DU model from the main analysis, which does not include these predictors, may be more suitable and were chosen as the final models.

There are some candidate predictors in the main analysis and sensitivity analyses that are not currently available or routinely used in primary care. Some of these predictors have continued to be included in the final models for the target conditions. The limited availability of the predictors in primary care raises a concern about the utility of the models in practice. It could either result in missing data at the time of diagnosing or a delay in receiving the diagnosis for the target condition if the patient needs to be referred to receive an index test that is part of the model. The latter would imply that the models may be more useful for secondary care. The model for DO in SA2 is the only model to use index tests that are in primary care. For this reason, this model and the DO model from the main analysis were externally validated using the validation cohort.

The fractional polynomial functions for age had extreme values for the coefficient, odds ratio and CIs in BOO model 1. Thus, this variable was simplified to a linear term in BOO model 2. Age as a linear term did not change the coefficient and odds ratio values of the remaining predictors in this model when compared with BOO model 1. Furthermore, the fractional polynomial terms for voiding symptom subscore in BOO model 1 and BOO model 2 were reduced to a simple linear term in BOO model 3, because the odds of BOO from the complex terms were not clinically plausible following clinical input. Voiding symptom subscore as a linear term did not change the coefficient and odds ratio values of the remaining predictors in this model when compared with BOO model 2 but removed mean 24-hour frequency as a statistically significant predictor. The calibration plot, optimism-corrected c-index and calibration slope of all models were generally similar. Thus, BOO model 3, the less complex and more parsimonious model was chosen as the final BOO model. This model also had a stronger discriminative performance in comparison with the BOO model from SA2.

The optimism-corrected calibration slope for the models in the main analysis and sensitivity analyses was < 1.0, suggesting that the models were overfitted. The models could be overfitted due to unimportant variables included by chance during the backwards elimination procedure, inflated importance of predictors in the models as well as the variable selection procedure in the bootstrap method for internal validation. The models developed in each bootstrap sample may have been different to the models developed using the original sample if the original predictors or functional forms were not commonly chosen during the variable selection procedure. Thus, before the models can be considered applicable in practice, the models require additional validation using the validation cohort and potentially further recalibration.

Chapter 4 Validation cohort results

Participants

The results presented in this section are based only on participants in the validation cohort. The validation cohort consisted of 251 participants with the last participant completing uroflowmetry and the urodynamics procedure on 11 August 2022.

Recruitment of participants into the validation cohort

Participants were recruited from 55 GP practices across Newcastle upon Tyne, Wales and Bristol between 28 March 2018 and 8 June 2022. The number of participants recruited at each GP practice and study hub is shown in *Appendix 4*, *Table 46*.

The results of the reference standard (present, absent or indeterminate) for each target condition are shown in *Table 15*. There were 14 (5.6%), 12 (4.8%) and 7 (2.8%) men with indeterminate diagnosis for BOO, DU and DO, respectively. Of the remaining 237, 239 and 244 men, there were 112 (44.6%), 87 (34.7%) and 166 (66.1%) with BOO, DU and DO, respectively. Thus, leaving 125 (49.8%), 152 (60.6%) and 78 (31.1%) without BOO, DU and DO, respectively. The final diagnosis of the 251 participants across the three target conditions is presented in *Table 16*. Sixteen men (6.4%) and 15 (6.0%) men were absent and present with all three target conditions, respectively. There were a high proportion of participants 60 (23.9%) absent with DU and diagnosed with both BOO and DO. The prevalence of BOO was similar between the development and validation cohorts, whereas the prevalence of DU and DO was slightly lower in the validation cohort.

Baseline characteristics of participants

The demographics and clinical characteristics of the 251 men are presented in *Appendix 4*, *Tables 47–49* separately for each target condition. Generally, the baseline characteristics were similar between the participants in the development and validation cohorts. The median age was 67 (interquartile range 60–74) years. Majority (94.4%) of the men were of white ethnicity. Just over half (*n* = 132, 52.6%) of the men were never smokers, 98 (39.0%) were ex-smokers and 17 (6.8%) were current smokers. The validation cohort consisted of a greater number of never smokers (52.6%) and fewer ex-smokers (39.0%) than the development cohort (42.9% never smokers and 48.9% ex-smokers).

Of the 251 men in the validation cohort, over 70% had comorbidities: 70 (27.9%) had no comorbidities, 136 (54.2%) had a single comorbidity and the remaining 43 (17.1%) had multiple comorbidities. There were 28 (11.2%) men with

TABLE 15 Reference standard results for each target condition

Target condition		Frequency (%) *n* = 251
BOO	Present	112 (44.6)
	Absent	125 (49.8)
	Indeterminate	14 (5.6)
DU	Present	87 (34.7)
	Absent	152 (60.6)
	Indeterminate	12 (4.8)
DO	Present	166 (66.1)
	Absent	78 (31.1)
	Indeterminate	7 (2.8)

TABLE 16 Cross-classification of participants with, without or indeterminate diagnosis of BOO, DU and DO with urodynamics (reference standard)

	BOO present			BOO absent			BOO indeterminate		
Final diagnosis, *n* (%)	**DU present**	**DU absent**	**DU indeterminate**	**DU present**	**DU absent**	**DU indeterminate**	**DU present**	**DU absent**	**DU indeterminate**
DO present	15 (6.0)	60 (23.9)	–	30 (12.0)	57 (22.7)	–	1 (0.4)	1 (0.4)	2 (0.8)
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
DO absent	18 (7.2)	18 (7.2)	–	22 (8.8)	16 (6.4)	–	–	–	4 (1.6)
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
DO indeterminate	1 (0.4)	–	–	–	–	–	–	–	6 (2.4)
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–

Note
'–' means there were no participants in this particular category.

Diabetes mellitus II, 17 (6.8%) with cerebrovascular accident and 12 (4.8%) with obstructive sleep apnoea. There were 83 (33.1%) men on alpha-blockers, 50 (19.9%) on calcium channel blockers, 20 (8.0%) on SSRI antidepressants and 30 (12.0%) participants on 5-alpha reductase inhibitors (see *Appendix 4*, *Tables 50–52*).

Adverse events

Two out of 251 participants in the validation cohort experienced urodynamics-related adverse events that required hospitalisation. One participant was diagnosed with UTI (SAE level of moderate) and another participant was diagnosed with acute urinary retention (*Table 17*).

Imputation of missing data

The candidate predictors for the validation cohort are summarised by the target conditions in *Tables 18–20*. The number of participants with missing data for the candidate predictors in the main analysis and SA2 (for the validation of DO model) is presented in the tables. The multiple imputation procedure produced values for the missing data that were consistent with the non-missing data. Generally, there was a higher percentage of candidate predictors missing data in the validation cohort in comparison with the development cohort.

Bladder outlet obstruction

Data from 237 individuals were used to validate the models from the main analysis, of which 149 (62.9%) individuals had complete data for all candidate predictors. Data on age, voiding symptoms subscore (ICIQ–MLUTS questionnaire) and incontinence symptoms subscore (ICIQ–MLUTS questionnaire) were known for all individuals. The percentage of data missing for the candidate predictors is shown in *Table 18*. In the main analysis, 37.1% (n = 88) individuals had at least one candidate predictor missing and so 38 imputed data sets were created. Missing values for size of prostrate, PSA test result, post-void residual urine, mean nocturnal frequency, median maximum flow rate, median voided volume, mean 24-hour frequency, mean urgency score and mean 24-hour fluid intake were imputed.

TABLE 17 Urodynamics-related adverse events (validation cohort)

Urodynamics-related adverse events, *n* (%)		Total 2/251
Main diagnosis	UTI	1 (0.4)
	Acute urinary retention	1 (0.4)
Seriousness of the event	Not serious	0 (0.0)
	Resulted in death	0 (0.0)
	Life-threatening	0 (0.0)
	Required hospitalisation	2 (0.8)
	Persistent or significant disability/incapacity	0 (0.0)
	Congenital anomaly/birth defect	0 (0.0)
	Other medically important condition	0 (0.0)
	Visible haematuria (blood in urine)/urethral bleeding requiring intervention	0 (0.0)
SAE level	Mild	0 (0.0)
	Moderate	1 (0.4)
	Severe	1 (0.4)

TABLE 18 Summary of candidate predictors for the validation of models for BOO

			Diagnosis		
Analysis	Candidate predictor		BOO *n* = 112	No BOO *n* = 125	Total *n* = 251
Main analysis	Age, years (participant demographics)	Median (IQR)	69.0 (63.5–74.6)	67.5 (57.0–73.0)	67.6 (60.6–74.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Size of prostate (DRE), *n* (%)	Normal	19 (17.0)	43 (34.4)	68 (27.1)
		Mild enlargement	34 (30.4)	40 (32.0)	79 (31.5)
		Moderate enlargement	53 (47.3)	39 (31.2)	95 (37.9)
		Gross enlargement	4 (3.6)	3 (2.4)	7 (2.8)
		Missing	2 (1.8)	0 (0.0)	2 (0.8)
	PSA test, ng/ml (blood test)	Median (IQR)	1.7 (0.9–2.7)	1.1 (0.7–2.0)	1.3 (0.7–2.5)
		Missing, *n* (%)	2 (1.8)	0 (0.0)	2 (0.8)
	Incontinence symptoms subscore (ICIQ–MLUTS questionnaire)	Median (IQR)	4.5 (3.0–7.0)	4.0 (3.0–7.0)	4.0 (3.0–7.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)

TABLE 18 Summary of candidate predictors for the validation of models for BOO (*continued*)

Analysis	Candidate predictor		Diagnosis		
			BOO *n* = 112	No BOO *n* = 125	Total *n* = 251
	Voiding symptoms subscore (ICIQ–MLUTS questionnaire)	Median (IQR)	7.0 (5.0–11.0)	6.0 (3.0–9.0)	7.0 (4.0–10.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Post-void residual urine, ml (bladder ultrasound)	Median (IQR)	77.0 (26.5–151.0)	34.0 (15.0–79.0)	48.0 (18.0–114.0)
		Missing, *n* (%)	0 (0.0)	4 (3.2)	8 (3.2)
	Median maximum flow rate, ml/second (uroflowmetry)	Median (IQR)	10.0 (7.3–12.7)	16.6 (12.7–23.2)	12.7 (9.2–17.2)
		Missing, *n* (%)	12 (10.7)	25 (20.0)	39 (15.5)
	Median voided volume, ml (uroflowmetry)	Median (IQR)	169.0 (133.0–224.0)	215.0 (141.0–277.0)	183.0 (137.0–254.0)
		Missing, *n* (%)	13 (11.6)	25 (20.0)	40 (15.9)
	Mean 24-hour frequency (uroflowmetry)	Median (IQR)	8.9 (6.5–11.0)	7.3 (5.6–9.1)	7.8 (5.9–10.1)
		Missing, *n* (%)	13 (11.6)	25 (20.0)	40 (15.9)
	Mean nocturia frequency (uroflowmetry)	Median (IQR)	2.0 (1.3–3.0)	1.5 (1.1–2.1)	1.8 (1.2–2.8)
		Missing, *n* (%)	13 (11.6)	25 (20.0)	41 (16.3)
	Mean urgency score (bladder diary)	Median (IQR)	1.5 (1.0–2.4)	1.6 (1.0–2.2)	1.6 (1.0–2.3)
		Missing, *n* (%)	26 (23.2)	18 (14.4)	44 (17.5)
	Mean 24-hour fluid intake, ml (bladder diary)	Median (IQR)	1697.5 (1415.0–2283.0)	1954.0 (1453.0–2387.0)	1833.0 (1433.0–2340.0)
		Missing, *n* (%)	30 (26.8)	28 (22.4)	60 (23.9)

IQR, interquartile range.

Detrusor underactivity

Data from 239 individuals were used to validate the models from the main analysis, of which 150 (62.8%) individuals had complete data for all candidate predictors. Data on age, voiding symptoms subscore (ICIQ–MLUTS questionnaire) and incontinence symptoms subscore (ICIQ–MLUTS questionnaire) were known for all individuals. The percentage of data missing for the candidate predictors is shown in *Table 19*. In the main analysis, 37.2% (*n* = 89) individuals had at least one candidate predictor missing and so 38 imputed data sets were created. Missing values for size of prostrate, PSA test result, post-void residual urine, mean nocturnal frequency, median maximum flow rate, median voided volume, mean 24-hour frequency, mean urgency score and mean 24-hour fluid intake were imputed.

Detrusor overactivity

Data from 244 individuals were used to validate the models. For the main analysis and SA2, 152 (62.3%) and 142 (58.2%) individuals had complete data for all candidate predictors, respectively. Data on age, voiding symptoms subscore (both ICIQ–MLUTS and IPSS questionnaires) and incontinence symptoms subscore (ICIQ–MLUTS questionnaire) and storage symptoms subscore (IPSS questionnaire) were known for all individuals. The percentage of data missing for the candidate predictors in the main analysis and SA2 is shown in *Table 20*. In the main analysis, 37.7% (*n* = 92) of individuals had at least one candidate predictor missing and so 38 imputed data sets were created. Missing values for size of prostrate, PSA test result, post-void residual urine, mean nocturnal frequency, median maximum flow rate, median voided volume, mean 24-hour frequency, mean urgency score and mean 24-hour fluid intake were imputed. For SA2, 41.8% (*n* = 102) of individuals had at least one candidate predictor missing and 42 imputed data sets

TABLE 19 Summary of candidate predictors for the validation of models for DU

			Diagnosis		
Analysis	**Candidate predictor**		**DU** *n* = 87	**No DU** *n* = 152	**Total** *n* = 251
Main analysis	Age, years (participant demographics)	Median (IQR)	69.2 (59.7–76.5)	67.9 (62.2–73.3)	67.6 (60.6–74.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Size of prostate (DRE) *n* (%)	Normal	20 (23.0)	42 (27.6)	68 (27.1)
		Mild enlargement	28 (32.2)	47 (30.9)	79 (31.5)
		Moderate enlargement	36 (41.4)	57 (37.5)	95 (37.9)
		Gross enlargement	2 (2.3)	5 (3.3)	7 (2.8)
		Missing	1 (1.2)	1 (0.7)	2 (0.8)
	PSA test, ng/ml (blood test)	Median (IQR)	1.4 (0.6–2.2)	1.4 (0.8–2.7)	1.3 (0.7–2.5)
		Missing, *n* (%)	1 (1.2)	1 (0.7)	2 (0.8)
	Incontinence symptoms subscore (ICIQ–MLUTS questionnaire)	Median (IQR)	4.0 (2.0–6.0)	5.0 (3.0–7.5)	4.0 (3.0–7.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Voiding symptoms subscore (ICIQ–MLUTS questionnaire)	Median (IQR)	7.0 (4.0–10.0)	7.0 (4.0–10.0)	7.0 (4.0–10.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Post-void residual urine, ml (bladder ultrasound)	Median (IQR)	61.0 (25.0–150.0)	37.0 (18.0–101.0)	48.0 (18.0–114.0)
		Missing, *n* (%)	1 (1.2)	3 (2.0)	8 (3.2)
	Median maximum flow rate, ml/second (uroflowmetry)	Median (IQR)	11.3 (8.6–15.3)	13.7 (9.7–20.0)	12.7 (9.2–17.2)

TABLE 19 Summary of candidate predictors for the validation of models for DU (*continued*)

Analysis	Candidate predictor		Diagnosis		
			DU *n* = 87	No DU *n* = 152	Total *n* = 251
		Missing, *n* (%)	15 (17.2)	22 (14.5)	39 (15.5)
	Median voided volume, ml (uroflowmetry)	Median (IQR)	163.5 (127.0–226.5)	197.0 (149.0–276.0)	183.0 (137.0–254.0)
		Missing, *n* (%)	15 (17.2)	23 (15.1)	40 (15.9)
	Mean 24-hour frequency (uroflowmetry)	Median (IQR)	7.9 (6.4–9.9)	7.8 (5.8–10.5)	7.8 (5.9–10.1)
		Missing, *n* (%)	15 (17.2)	23 (15.1)	40 (15.9)
	Mean nocturia frequency (uroflowmetry)	Median (IQR)	1.9 (1.2–2.4)	1.7 (1.2–2.8)	1.8 (1.2–2.8)
		Missing, *n* (%)	15 (17.2)	24 (15.8)	41 (16.3)
	Mean urgency score (bladder diary)	Median (IQR)	1.3 (1.0–2.0)	1.7 (1.1–2.4)	1.6 (1.0–2.3)
		Missing, *n* (%)	10 (11.5)	34 (22.4)	44 (17.5)
	Mean 24-hour fluid intake, ml (bladder diary)	Median (IQR)	1690.5 (1389.0–2165.0)	1968.5 (1483.5–2393.5)	1833.0 (1433.0–2340.0)
		Missing, *n* (%)	19 (21.8)	40 (26.3)	60 (23.9)

IQR, interquartile range.

were created. Missing values for size of prostrate, PSA test result, palpable bladder, voided volume, maximum flow rate, average number of voids, mean urgency score, mean 24-hour fluid intake and nocturnal polyuria score were imputed.

Model performance

Bladder outlet obstruction model 3, DU and DO models from the main analysis and DO model from SA2, developed in *Chapter 3*, were externally validated using the validation cohort. The models were fitted as proposed in this data set to assess the discrimination and calibration performance in another sample from a similar population as the development cohort. The calibration slope and c-index of the models are reported in *Table 21*. The c-index for BOO model 3 was 0.82 (95% CI 0.76 to 0.88), suggesting a strong discriminative performance. The c-index (95% CI) of the models for DU and DO from the main analysis were 0.63 (95% CI 0.56 to 0.71) and 0.62 (95% CI 0.55 to 0.70), respectively, suggesting that these models have adequate discriminative ability. The DO model from SA2 had an improved discriminative performance in comparison with the DO model from the main analysis with a c-index of 0.70 (95% CI 0.64 to 0.77).

The calibration slope for BOO model 3 was 0.99 (95% CI 0.68 to 1.30), suggesting near-perfect calibration. On the other hand, the calibration slope was 0.82 (95% CI 0.31 to 1.32) and 0.72 (95% CI 0.27 to 1.17) respectively for the DU and DO models from the main analysis, suggesting that the models could be overfitted as predictions were too extreme for low and high probabilities. The calibration slope for DO model from SA2 was > 1, 1.36 (95% CI 0.78 to 1.94), suggesting that the model could be underfitted, that is the predictions were not low enough for low probabilities and the predictions were not high enough for high probabilities. There was large uncertainty around the c-index and calibration slope of the DU and DO models, as seen in the width of the 95% CI. This is likely to be due to the few participants observed with these outcomes and the validation cohort having a small sample size.

TABLE 20 Summary of candidate predictors for the validation of models for DO

Analysis	Candidate predictor		Diagnosis		
			DO *n* = 166	No DO *n* = 78	Total *n* = 251
Main analysis	Age, years (participant demographics)[a]	Median (IQR)	68.7 (62.4–74.7)	65.5 (56.4–71.5)	67.6 (60.6–74.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Size of prostate (DRE),[a] *n* (%)	Normal	46 (27.7)	19 (24.4)	68 (27.1)
		Mild enlargement	51 (30.7)	26 (33.3)	79 (31.5)
		Moderate enlargement	61 (36.8)	32 (41.0)	95 (37.9)
		Gross enlargement	7 (4.2)	0 (0.0)	7 (2.8)
		Missing	1 (0.6)	1 (1.3)	2 (0.8)
	PSA test, ng/ml (blood test)[a]	Median (IQR)	1.5 (0.8–2.6)	1.2 (0.6–2.3)	1.3 (0.7–2.5)
		Missing, *n* (%)	1 (0.6)	1 (1.3)	2 (0.8)
	Incontinence symptoms subscore (ICIQ–MLUTS questionnaire)	Median (IQR)	5.0 (3.0–8.0)	4.0 (2.0–5.0)	4.0 (3.0–7.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Voiding symptoms subscore (ICIQ–MLUTS questionnaire)	Median (IQR)	7.0 (3.0–10.0)	7.0 (5.0–11.0)	7.0 (4.0–10.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Post-void residual urine, ml (bladder ultrasound)	Median (IQR)	39.0 (17.0–102.0)	66.0 (29.0–136.0)	48.0 (18.0–114.0)
		Missing, *n* (%)	5 (3.0)	1 (1.3)	8 (3.2)
	Median maximum flow rate, ml/second (uroflowmetry)	Median (IQR)	12.7 (8.4–18.2)	12.4 (9.9–17.1)	12.7 (9.2–17.2)
		Missing, *n* (%)	27 (16.3)	11 (14.1)	39 (15.5)
	Median voided volume, ml (uroflowmetry)	Median (IQR)	173.0 (133.0–239.0)	204.0 (151.0–280.0)	183.0 (137.0–254.0)
		Missing, *n* (%)	27 (16.3)	12 (15.4)	40 (15.9)
	Mean 24-hour frequency (uroflowmetry)	Median (IQR)	8.1 (6.3–10.8)	7.2 (5.7–8.9)	7.8 (5.9–10.1)
		Missing, *n* (%)	27 (16.3)	12 (15.4)	40 (15.9)

TABLE 20 Summary of candidate predictors for the validation of models for DO (*continued*)

Analysis	Candidate predictor		Diagnosis		
			DO *n* = 166	No DO *n* = 78	Total *n* = 251
	Mean nocturia frequency (uroflowmetry)	Median (IQR)	2.0 (1.3–3.0)	1.6 (0.9–2.2)	1.8 (1.2–2.8)
		Missing, *n* (%)	28 (16.9)	12 (15.4)	41 (16.3)
	Mean urgency score (bladder diary)[a]	Median (IQR)	1.7 (1.2–2.4)	1.1 (1.0–1.7)	1.6 (1.0–2.3)
		Missing, *n* (%)	32 (19.3)	12 (15.4)	44 (17.5)
	Mean 24-hour fluid intake, ml (bladder diary)[a]	Median (IQR)	1813.0 (1400.0–2332.0)	1902.5 (1515.0–2350.0)	1833.0 (1433.0–2340.0)
		Missing, *n* (%)	40 (24.1)	20 (25.6)	60 (23.9)
SA2	Storage symptoms subscore (IPSS questionnaire)	Median (IQR)	9.0 (7.0–12.0)	7.0 (5.0–10.0)	9.0 (6.0–11.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Voiding symptoms subscore (IPSS questionnaire)	Median (IQR)	6.0 (3.0–11.0)	7.0 (4.0–12.0)	7.0 (3.0–11.0)
		Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
	Palpable bladder (physical examination), *n* (%)	Yes	7 (4.2)	0 (0.0)	8 (3.2)
		No	158 (95.2)	77 (98.7)	241 (96.0)
		Missing	1 (0.6)	1 (1.3)	2 (0.8)
	Maximum flow rate, ml/second (one-off flow test)	Median (IQR)	8.0 (5.2–15.0)	6.5 (4.0–9.9)	7.7 (5.0–13.0)
		Missing, *n* (%)	19 (11.4)	3 (3.8)	23 (9.2)
	Voided volume, ml (one-off flow test)	Median (IQR)	136.5 (75.5–225.5)	106.5 (57.0–264.5)	123.5 (70.0–244.0)
		Missing, *n* (%)	18 (10.8)	2 (2.6)	21 (8.4)
	Average number of voids (bladder diary)	Median (IQR)	9.0 (7.3–11.3)	8.7 (7.0–10.3)	9.0 (7.3–11.0)
		Missing, *n* (%)	39 (23.5)	20 (25.6)	59 (23.5)
	Nocturnal polyuria score (bladder diary)	Median (IQR)	0.3 (0.2–0.4)	0.3 (0.2–0.4)	0.3 (0.2–0.4)
		Missing, *n* (%)	51 (30.7)	22 (28.2)	74 (29.5)

IQR, interquartile range.
a These variables are candidate predictors in SA2.

TABLE 21 Performance statistics of BOO (model 3, main analysis), DU (main analysis) and DO (main analysis and SA2)

Outcome	Analysis	c-index (95% CI)	Calibration slope (95% CI)	Recalibrated calibration slope (95% CI)	Calibration plot
BOO, *n* = 237	BOO model 3, main	0.82 (0.76 to 0.88)	0.99 (0.68 to 1.30)	1.00 (0.69 to 1.31)	*Figure 5*
DU, *n* = 239	Main	0.63 (0.56 to 0.71)	0.82 (0.31 to 1.32)	1.00 (0.38 to 1.62)	
DO, *n* = 244	Main	0.62 (0.55 to 0.70)	0.72 (0.27 to 1.17)	1.00 (0.38 to 1.62)	
	SA2	0.70 (0.64 to 0.77)	1.36 (0.78 to 1.94)	1.00 (0.58 to 1.42)	

Potential reasons for why the models for DU and DO (both main analysis and SA2) are mis-calibrated compared to BOO model 3 could include difference in prevalence between the development and validation cohorts. The prevalence of BOO was similar between the development and validation cohorts, whereas the prevalence of DU and DO was slightly lower in the validation cohort. Another reason could be variation in case-mix between the development and validation cohort in the observed predictors or other unobserved characteristics, due to temporal factors such as the COVID-19 pandemic during the validation cohort recruitment.

The calibration plot for the four models is shown in *Figure 5*. The calibration plot was good for BOO model 3. Nine out of the 10 risk groups overlapped the reference line. Similarly to the development cohort, the deciles were spread across the reference line indicating risk of the outcome varied among participants. The observed and expected probabilities were close to one another, as seen by points lying close to the reference line. The lowess marginally drifted away from

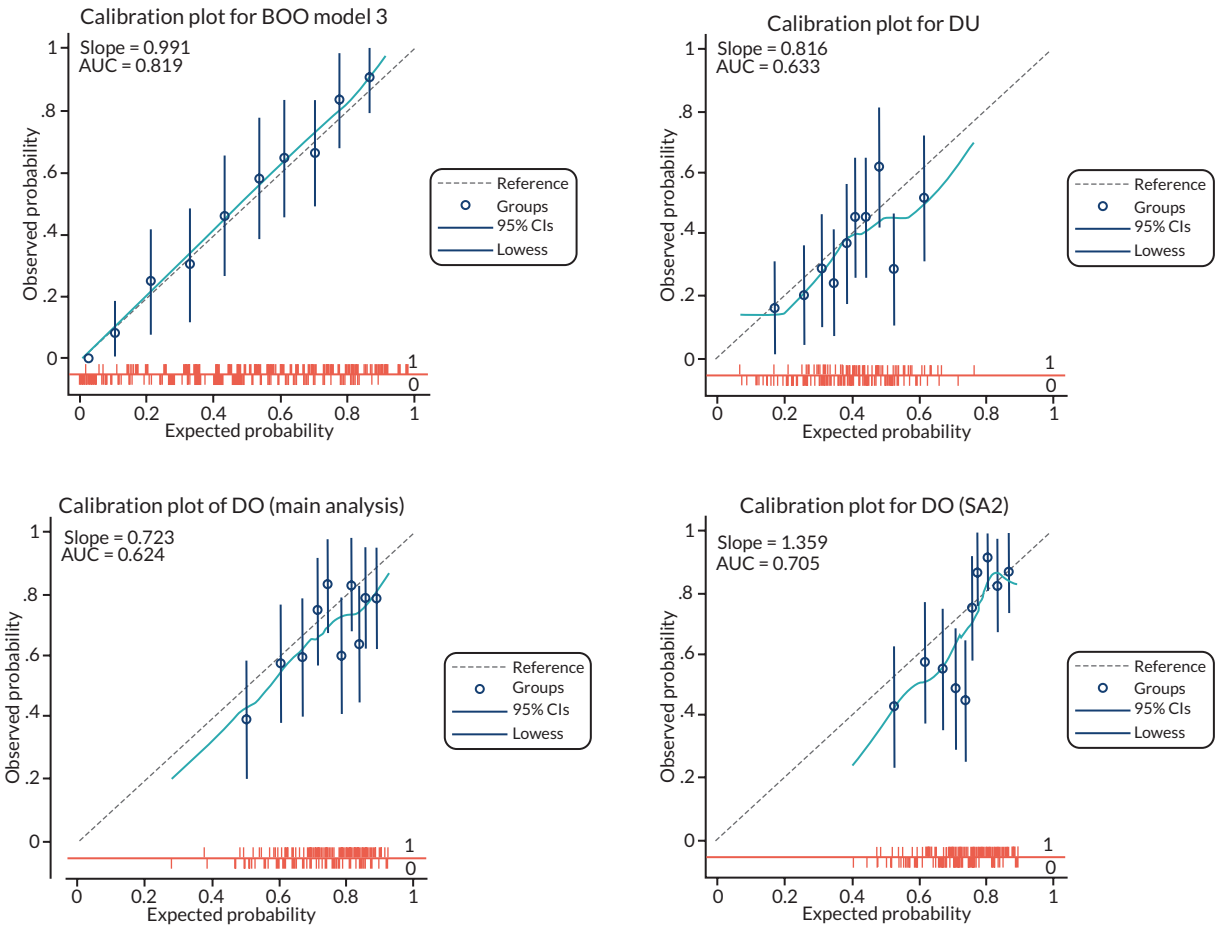


FIGURE 5 Calibration plots comparing observed and predicted probabilities for BOO (model 3, main analysis), DU (main analysis) and DO (main analysis and SA2), based on the validation cohort. AUC, area under the curve.

the reference line as the risk of the target condition increased, suggesting slight miscalibration at individual level in participants with higher risk. For the higher risk groups, the expected probability was less than the observed probability, suggesting the model is underfitting in these groups.

The calibration plot was poor for the DU model. The deciles were clustered in the middle of the plot indicating majority of the participants had low to moderate risk of the outcome (probability between 0.2 and 0.6). There were fewer data at the higher risk probabilities, as indicated by the spike plot at the bottom, and there were fewer participants with the outcome in the low-risk groups when compared with the groups of participants with moderate risk. Nine out of 10 groups of participants overlapped the reference line. The lowess generally stayed below the reference line, such that the expected probability was greater than the observed probability, suggesting that the model was overfitting at individual level in participants.

Similarly to the DU model, the calibration plots for the DO models from the main analysis and SA2 were poor. Nine groups of participants overlapped the reference line with the DO model from the main analysis, whereas eight groups of participants overlapped the reference line with the model from SA2. The deciles were clustered in the middle and top right of the plot indicating participants had moderate to high risk of the outcome (probability between 0.5 and 0.9). Similarly to the DU model, the lowess generally stayed below the reference line, such that the expected probability was greater than the observed probability. This suggests that the model was overfitting at an individual level in participants. However, in the DO model from SA2, the expected probability was less than the observed probability in the high-risk groups (as seen by the lowess being above the reference line), suggesting underfitting. This could be due to a greater number of participants observed with the outcome in the high-risk groups than without, as indicated by the spike plot at the bottom.

Recalibration

To correct for the miscalibration in the calibration slope, the models were re-calibrated to the validation data set. The model intercept was re-estimated to account for the difference in prevalence in the validation cohort and the model coefficients were re-estimated using the calibration slope as a shrinkage factor to correct the model for overfitting or underfitting. A large amount of shrinkage was required for the DU and DO models from the main analysis and the DO model from SA2 to adjust for miscalibration in comparison with BOO model 3, suggesting the latter model had a better predictive performance. Significant improvement in the calibration plots were therefore seen for DU model and DO models presented in *Figure 6* where the lowess intersected the reference line on multiple occasions. Following recalibration, the calibration slopes for all models were 1.00 (*Table 21*).

The new model coefficients for each predictor in BOO model 3, DU and DO models are presented in *Tables 22–25*. Full model equations can also be found in *Box 2*. Note, the recalibration re-estimated model parameters based on the validation cohort; thus, the calibration performance has improved in this sample. However, performance in new cohorts is not guaranteed; therefore, further external validation may be required to reassess the calibration performance.

Estimates of diagnostic accuracy

The ROC plots of the models following recalibration are presented in *Figure 7*. The plots include labels of the predicted probabilities to inform the sensitivity and specificity at different risk thresholds. Sensitivity and specificity of 75% were deemed to be the minimum clinically useful performance. Thus, for:

- BOO model 3, at a threshold of 50.9%, a sensitivity of 75.1% could be achieved with a specificity of 74.2% approximately. At a threshold of 53.3%, a specificity of 75.5% could be achieved with a sensitivity of 71.3% approximately.
- For the DU model, at a threshold of 34.2%, a sensitivity of 75.3% could be achieved with a specificity of 47.3% approximately. At a threshold of 41.4%, a specificity of 75.1% could be achieved with a sensitivity of 39.8% approximately.

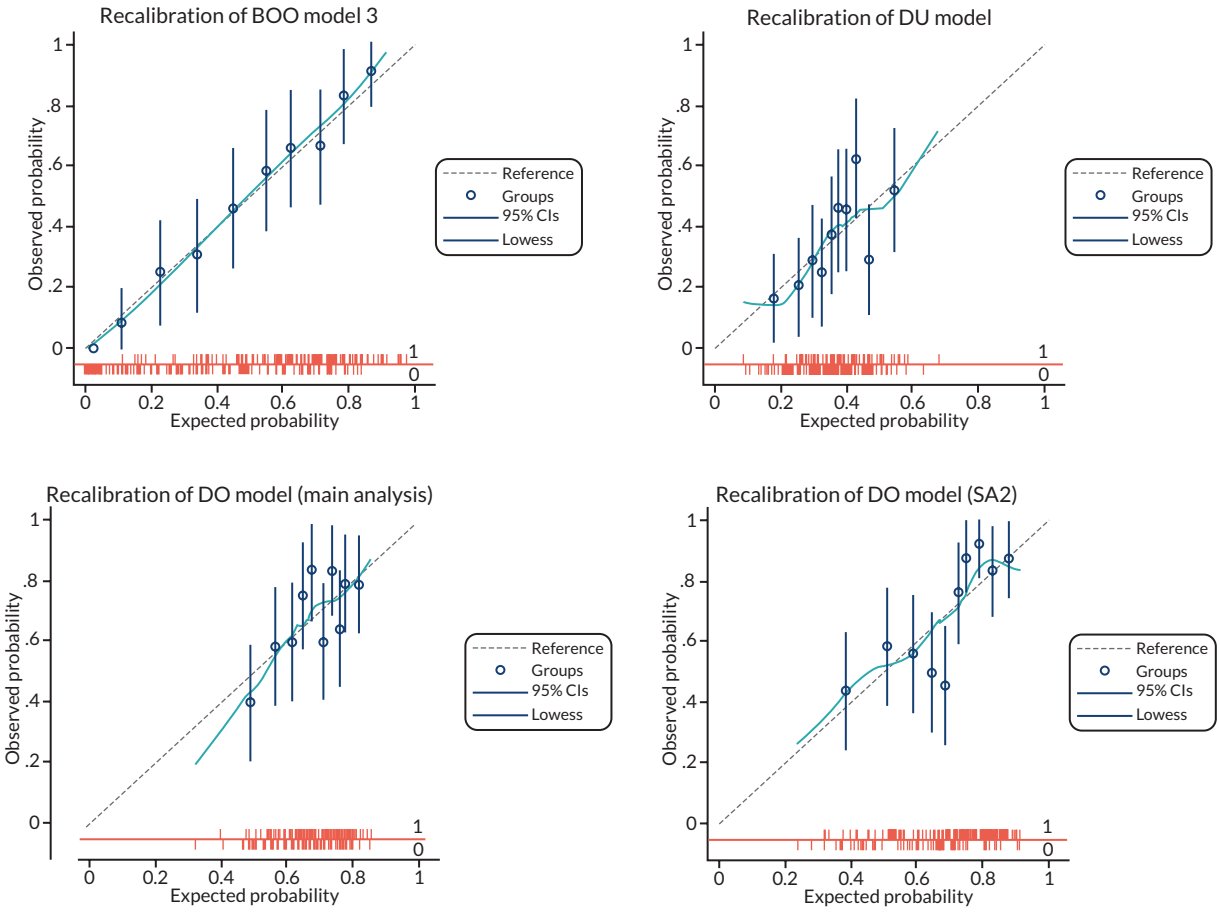


FIGURE 6 Calibration plots comparing observed and predicted probabilities for BOO (model 3, main analysis), DU (main analysis) and DO (main analysis and SA2), based on the validation cohort following recalibration.

TABLE 22 Recalibration of the model coefficients for predictors included in the model for BOO (model 3)

Predictors and intercept	Development optimism-corrected coefficient	Re-calibrated coefficient
Age (years)	−0.02	−0.02
PSA test result (ng/ml)	0.16	0.16
Voiding symptoms subscore (ICIQ–MLUTS questionnaire)	−0.06	−0.06
Median maximum flow rate (ml/second)	−0.30	−0.29
Median voided volume (ml)	0.009	0.009
Intercept	3.43	3.46

Note
Re-calibrated coefficients multiplied with a shrinkage factor (calibration slope) of 0.99.

TABLE 23 Recalibration of the model coefficients for predictors included in the model for DU (main analysis)

Predictors and intercept	Development optimism-corrected coefficient	Re-calibrated coefficient
Post-void residual urine (ml)	0.002	0.002
Median maximum flow rate (ml/second)	−0.06	−0.05
Intercept	0.26	−0.02

Note
Re-calibrated coefficients multiplied with a shrinkage factor (calibration slope) of 0.82.

TABLE 24 Recalibration of the model coefficients for predictors included in the model for DO (main analysis)

Predictors and intercept	Development optimism-corrected coefficient	Re-calibrated coefficient
Age (years)	0.04	0.03
Post-void residual urine (ml)	−0.002	−0.002
Median voided volume_1 (ml)	−0.15	−0.11
Median voided volume_2 (ml)	0.10	0.07
Intercept	−0.81	−0.62

$$Median\ voided\ volume\_1 = \left(\frac{MVV}{100}\right)^3$$

$$Median\ voided\ volume\_2 = \left(\frac{MVV}{100}\right)^3 \ln\left(\frac{MVV}{100}\right)$$

Note
Re-calibrated coefficients multiplied with a shrinkage factor (calibration slope) of 0.72.

TABLE 25 Recalibration of the model coefficients for predictors included in the model for DO (SA2)

Predictors and intercept	Development optimism-corrected coefficient	Re-calibrated coefficient
Age (years)	0.03	0.05
Storage symptoms subscore (IPSS questionnaire)	0.10	0.14
Intercept	−2.05	−3.43

Note
Re-calibrated coefficients multiplied with a shrinkage factor (calibration slope) of 1.36.

- For DO model from the main analysis, at a threshold of 63.8%, a sensitivity of 75.1% could be achieved with a specificity of 45.6% approximately. At a threshold of 75.2%, a specificity of 75.7% could be achieved with a sensitivity of 33.3% approximately.
- For DO model from SA2, at a threshold of 63.1%, a sensitivity of 75.3% could be achieved with a specificity of 46.2% approximately. At a threshold of 71.4%, a specificity of 75.6% could be achieved with a sensitivity of 62.7% approximately.

Discussion

In *Chapter 3*, BOO model 3, DU and DO models from the main analysis and DO model from SA2 were chosen for additional validation using the validation cohort. These models were externally validated using data collated from further 251 participants across 55 GP practices in Bristol, Newcastle upon Tyne and Wales. The BOO model appeared to have strong discriminative performance, whereas the models for DU and DO from the main analysis had moderate discriminative ability. The DO model from SA2 had an improved discriminative performance when compared to the DO model from the main analysis. In comparison with BOO model 3, the remaining models appeared to be substantially mis-calibrated. All models were recalibrated using the calibration slope as a shrinkage factor to the validation data sets.

Complete data were available for fewer predictors in the validation cohort in comparison with the development cohort: age, voiding symptoms subscore and incontinence symptoms subscore from ICIQ–MLUTS questionnaire and storage symptoms subscore from IPSS questionnaire. The percentage of participants with a least one missing predictor in the validation cohort was nearly three times the amount that was missing in the development cohort for the main analysis.

BOX 2 Full model equations of BOO model 3, DU and DO models following analysis using the validation cohort

Model equation for BOO model (main analysis), corrected for optimism

Predictor	Fractional polynomial and coefficient ($\beta_i x_i$)
Age, years	−0.0174566 *age*
PSA test result, ng/ml	0.1621547 *PSA*
Voiding symptoms subscore (*VSS*) (ICIQ MLUTS questionnaire)	−0.0612032 *VSS*
Median maximum flow rate (*MMF*), ml/second	−0.2933156 *MMF*
Median voided volume (*MVV*), ml	+0.0087792 *MVV*
Intercept (α)	+3.449188

Model equation for DU (main analysis), corrected for optimism

Predictor	Fractional polynomial and coefficient ($\beta_i x_i$)
Post-void residual urine (*PVR*), ml	0.0020115 *PVR*
Median maximum flow rate (*MMF*), ml/second	−0.0518618 *MMF*
Intercept (α)	−0.0150367

Model equation for DO (main analysis), corrected for optimism

Predictor	Fractional polynomial and coefficient ($\beta_i x_i$)
Age, years	0.0299694 *age*
Post-void residual urine (*PVR*), ml	−0.0017761 *PVR*
Median voided volume (*MVV*), ml	$-0.1085548\left(\frac{MVV}{100}\right)^3 + 0.0707095\left(\frac{MVV}{100}\right)^3 \times \ln\left(\frac{MVV}{100}\right)$
Intercept (α)	−0.6244504

Model equation for DO (SA2), corrected for optimism

Predictor	Fractional polynomial and coefficient ($\beta_i x_i$)
Age, years	0.0462669 *age*
Storage symptoms subscore (*SSS*), (IPSS questionnaire)	+0.1401884 *SSS*
Intercept (α)	−3.433648

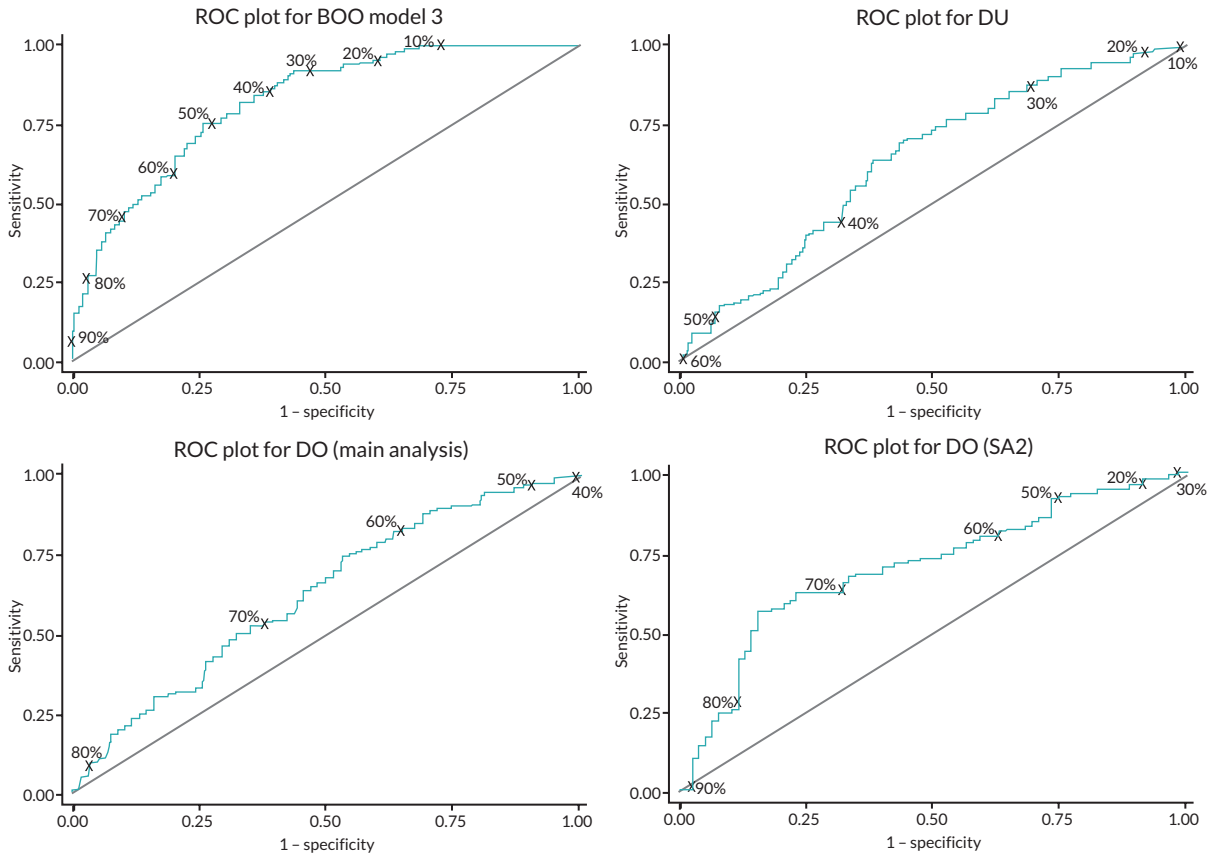


FIGURE 7 Receiver operating characteristic plots for BOO model 3, DU (main analysis) and DO (main analysis and SA2), based on the validation cohort following recalibration.

Data were largely missing for the predictors collected from the uroflowmetry (median maximum flow rate, median voided volume, mean 24-hour frequency and mean nocturia frequency), one-off flow test (maximum flow rate and voided volume) and the bladder diary (mean urgency score, mean 24-hour fluid intake, average number of voids and nocturnal polyuria score). BOO model 3, DU and DO models from the main analysis included median maximum flow rate or median voided volume in their model. Their limited availability may be partly due to the pandemic affecting the practicality of data collection at times, but in the validation data set, this raises a concern about the utility of these models in medical practice. For this reason, the DO model from SA2 may be a more suitable model than DO model from the main analysis. The former model consisted of predictors with no missing data, index tests that are currently available in primary care and this model had a better discriminative performance than DO model from the main analysis.

The validation cohort was recruited in a similar prospective manner to the development cohort, following the same inclusion/exclusion criteria, definitions to collect data on the predictors and outcome and recruiting participants from the same study hubs. The main difference that existed between the development and validation cohorts was the time periods in which both samples of participants were recruited. This type of external validation is likely to have resulted in reduced case-mix variation between the development and validation cohorts and is called temporal split-sample. The limitation of this form of external validation is that the models may have behaved similarly in the validation cohort to the development cohort, resulting in an inflated calibration performance, as potentially observed for BOO model 3. Another limitation is that the models have not been validated in a different population, such as outside of Bristol, Newcastle upon Tyne and Wales, where patient demographics or prevalence could differ. The generalisability of the results in other settings is likely to remain unknown and, therefore, further validation and recalibration may be required before the models can be used in practice.

The validation cohort sample size of 325 participants was not achieved, with only 251 participants recruited due to the decelerated recruitment during the COVID-19 pandemic. Furthermore, the study's use of split-sampling is not recommended and is often regarded as inefficient. The potential implications of both small sample size and split-sampling were reduced power, an increased risk of the models mis-calibrating, and imprecise estimates of model performance measures with wide CIs, as observed with the DU and DO models (*Tables 21* and *23*). Rather, it is recommended that all data are used for model development to reduce the chance of overfitting and the variability around the model performance measures; and an independent external validation is conducted by an independent analysis team to prevent bias in the interpretation of results.

Chapter 5 Qualitative evaluation

Background

Lower urinary tract symptoms are highly prevalent among men aged over 40, with over 70% reporting at least one symptom[20] and are associated with reduced emotional well-being, productivity and quality of life.[21,22] Despite this, only a minority of men with LUTS consult their GP about their symptoms,[23,24] and even fewer receive treatment.[25,26] The frequent comorbidity of MLUTS with prostatic disease, together with a commonly multifactorial aetiology,[27] means diagnosis and management can be complex.

Urodynamics is a specialist test for investigating the causes of LUTS.[28] It is typically conducted in secondary care and involves inserting catheters into the bladder and rectum to enable measurements pertaining to bladder function. A growing number of patients presenting with LUTS are eventually treated conservatively and thus could potentially be effectively managed in primary care.[29,30] Managing LUTS in primary care settings could result in benefits for the patient and cost benefits for the NHS.

Invasive urodynamics is a complex procedure with risk of adverse effects, yet very few qualitative studies have explored patient acceptability of this test. One interview study[31] found patients experienced anxiety and embarrassment about the procedure, which was alleviated by healthcare professionals through effective interpersonal and communication skills. Previously, men participating in a randomised controlled trial (RCT) involving urodynamics[32] found the procedure acceptable and valued the comprehensive insight into their symptoms. As healthcare professionals have been found to play an important role in the patient experience of urodynamics,[31,32] it is important to further explore their perspectives. Furthermore, no study has explored in-depth the attitudes of patients who have declined urodynamics, who may view the procedure differently.[32]

Research relating to outcomes for patients with LUTS has tended to focus on comparison of specific interventions rather than patient satisfaction with their care as a whole, an area that has been little explored.[33] Limited evidence indicates that treatment is frequently ineffective from a patient perspective. For example, in one observational study of males being treated for LUTS in primary care,[34] around half of men reported unsatisfactory outcomes, such as persisting or worsening symptoms. A qualitative interview study also found that most participants experienced no or only partial relief of their LUTS after consulting a clinician, although they generally reported satisfaction with the care they received.[33]

Primary care Management of lower Urinary tract Symptoms was the first large-scale study to implement invasive urodynamics in a primary care setting; no study has yet investigated the feasibility and acceptability of conducting this procedure in a non-specialist setting. As part of the qualitative evaluation within PriMUS we sought (Objective 5 of main study) to explore the feasibility and acceptability of providing invasive urodynamics in primary care, including experiences of recruiting to a urodynamic study, encompassing the perspectives of patients (including those who declined the procedure) and healthcare professionals. Including these participant groups provided further insight into attitudes towards and experiences of urodynamics, and the feasibility of providing urodynamics in primary care, and sought to inform processes or interventions that could improve acceptability of this invasive procedure. This section of the qualitative study is published in the BMC journal *Diagnostic and Prognostic Research*.[17]

Despite the challenges of managing LUTS in primary care, little research has explored in depth the perspectives of GPs themselves or investigated the experiences of patients. Through qualitative interviews PriMUS also explored

GPs' experiences of diagnosing and managing LUTS, together with patients' experiences of and preferences for treatment of LUTS in primary care. Findings of this part of the qualitative study are published in the *British Journal of General Practice*.[19]

Another part of the qualitative evaluation within PriMUS was to develop a series of patient management recommendations and thresholds for clinically useful diagnostic prediction by expert consensus and with reference to current clinical guidelines (Objective 3 of main study) that map to the diagnoses predicted by the statistical model (as described in *Chapter 5*, Workstream 2).

By combining the statistical model and the patient management recommendations we developed a prototype online clinical decision support tool for use in primary care assessed patients' and clinicians' views on the use of this clinical decision support tool in the primary care setting.

Objectives

The PriMUS qualitative work comprised three workstreams:

1. Study pilot phase evaluation/qualitative exploration of GP and patient experiences of LUTS.
2. Development of management recommendations to inform the clinical decision support tool.
3. User-testing of the prototype clinical decision support tool.

Workstream 1: Study pilot phase evaluation/qualitative exploration of general practitioner and patient experiences of lower urinary tract symptoms

The objectives of the PriMUS pilot phase qualitative evaluation were to assess acceptability of the reference urodynamic assessment and other study procedures, and identify barriers and facilitators to study recruitment, from the perspective of health professionals and patients. The qualitative study also included exploration of patients' experiences of LUTS and preferences for treatment, and GPs' experiences of managing LUTS. Feedback on the proposed clinical decision support tool was obtained from GPs.

Methods

Study design

Semistructured telephone interviews were conducted with men invited to take part in the main PriMUS study. Telephone and face-to-face interviews were also carried out with healthcare professionals involved in study delivery, including GPs and practice nurses involved in study recruitment, and research nurses who performed the urodynamic procedure.

Sampling and recruitment

Purposive sampling was used to support maximum variation in terms of study site, decision to participate in the main PriMUS study (for patients) and role in study delivery (for healthcare professionals). Patients invited to participate in the main study could indicate whether they consented to be contacted for a follow-up interview; this included both patients who consented to the main study and those who declined to participate. Patients consenting to be contacted were invited to take part in an interview once they had undergone all study procedures or had made the decision not to participate in the main PriMUS study. Interview participants were offered a £10 voucher for their contribution. Healthcare professionals were approached to take part once they had experience of study processes (e.g. recruiting patients or performing urodynamics). Recruitment continued until data saturation was reached. Informed consent (written or verbal) was obtained for all face-to-face and telephone interviews after participants had sufficient time to read the interview study information sheet.

Data collection

Semistructured interview topic guides were developed in consultation with clinicians and patient representatives on the study team. Topics explored in patient interviews included thoughts about the main study information, the decision of whether to take part, concerns about the study, and how others could be encouraged to participate. Topic guides also aimed to explore patients' experiences of LUTS, their decision to visit the GP about their symptoms, satisfaction with treatment, preferences for treatment in primary versus secondary care, and preferences for involvement in treatment decisions. Patients participating in the main study were additionally asked about their experience of study processes and specifically the urodynamic test. Topics covered in healthcare professional interviews included (as applicable) experiences of study recruitment, perceived patient acceptability of study processes, staff experiences of performing the urodynamic test, and the feasibility of doing so in primary care. GP interviews additionally aimed to explore experiences of identifying causes and treatment options and managing LUTS in general practice, and to gain feedback on the utility and content of the proposed clinical decision support tool. Topic guides were developed iteratively throughout the data collection period to allow for the exploration of previously unanticipated themes arising from the interviews. For example, after issues relating to urodynamic test results were highlighted by GPs, this was specifically explored in subsequent interviews.

Interviews were conducted between May 2018 and February 2019. All health professional interviews and most patient interviews were conducted by an experienced qualitative health researcher. A sample of patient interviews was conducted by medical students, supervised by the main interviewer. To encourage respondents to give honest, unbiased feedback about their experience of the main study, the interviewers were not previously known to participants and had no involvement in main study procedures. All interviews were audio-recorded with the permission of participants and transcribed verbatim for analysis.

Analysis

A framework approach was used to analyse interview data. Interview transcripts were read in full by the study qualitative researcher, and a coding framework was developed based on emerging themes and topics covered in the interview guides. NVivo 11 (QSR International, Warrington, UK) was used to organise the data into the themes identified in the framework. Ten per cent of interview transcripts was independently coded by a second qualitative researcher to ensure consistency in the way the codes were applied. Separate tables were compiled for health professionals and patients, to summarise the experience of each interview respondent in relation to each of the identified themes. The study's qualitative researchers met to discuss these data tables and identify and agree on key themes in relation to the research objective.

Analysis was carried out separately for findings related to: (1) the study pilot phase evaluation (which assessed acceptability of study procedures and identified barriers and facilitators to study recruitment), (2) the exploration of patients' and GPs' experiences of LUTs and (3) health professional feedback on the proposed clinical decision support tool; therefore, key themes and findings are presented separately.

Findings

Interviews were conducted with 25 patients and 18 healthcare professionals, from 22 GP practices across Newcastle upon Tyne, South Wales and Bristol (*Table 26*). Interviews with patients lasted between 8 and 44 minutes (mean 23.0); interviews with healthcare professionals lasted between 9 and 30 minutes (mean 18.8).

Findings are presented separately for each of the three components of this qualitative workstream: (1) the study pilot phase evaluation, (2) the exploration of patients' and GPs' experiences of LUTs and (3) health professional feedback on the proposed clinical decision support tool.

Study pilot phase evaluation

Eight subthemes were identified under the three main framework categories: Acceptability of invasive urodynamics in primary care (three subthemes), Feasibility of invasive urodynamics in primary care (three subthemes) and Recruiting to a urodynamic study (two subthemes) (see *Table 27*). Participant quotes in this section are labelled with a unique participant identification number, with the prefix main study participant for patients who participated in the main study, interview participant for patients who participated in the interview element only, GP for GPs and UN for urodynamic nurses.

TABLE 26 Participant characteristics

Patients (*n* = 25)		Healthcare professionals (*n* = 18)	
	N		*N*
Participant in main study		***Role***	
Yes	22	GP	11
No	3	Practice nurse	3
		Urodynamic nurse	4
Geographical region		***Geographical region***	
Newcastle upon Tyne	10	Newcastle upon Tyne	6
South Wales	9	South Wales	7
Bristol	6	Bristol	5
Age group[a]			
46–55	3		
56–65	6		
66–75	8		
76–85	5		

a Participant age was not recorded for patients who did not participate in the main PriMUS study.

TABLE 27 Key framework categories and emergent subthemes

Key framework categories	Emergent subthemes
Acceptability of invasive urodynamics in primary care	Apprehension and embarrassment Communication Preference for primary care
Feasibility of invasive urodynamics in primary care	Training and support Logistical issues Difficulties receiving and using results
Recruiting to a urodynamic study	Importance of proactive recruitment Reasons for participation and non-participation

Acceptability of invasive urodynamics in primary care

Apprehension and embarrassment

All patients who underwent urodynamics reported finding the procedure acceptable, with most finding it as they had expected, or better than expected. Around half of them reported discomfort, although this was universally described as brief or mild. Some patients described feeling apprehensive about the test; those with experience of similar medical procedures (e.g. cystoscopy) reported lower anxiety. The invasive nature of the procedure was mentioned by several patients; some commented that this was not an issue for them, while others found it embarrassing, particularly where the procedure was conducted by female healthcare professionals. While one suggested this was due to his older age, another explained how he accepted that getting older meant invasive procedures were more likely. Patients who had opted not to take part in the main study primarily reported that this was due to not wishing to undertake the urodynamic procedure.

It was a little bit … uncomfortable at first, but I mean not greatly so … It was only two seconds … I was a little bit … nervous because I didn't know how painful it was going to be. But it was actually … nothing like what I thought it was going to be you know. It was … a lot nicer or better …

MSP 2015, age 65

She explained it was going to feel a bit uncomfortable when they did it, and it did, it tingled a bit but that was about it, it was very much as I expected.

MSP 1023, age 73

No concerns at all really … I'd had that camera or catheter, or whatever you call it, in my bladder twice before. So, I, it didn't put me off at all.

MSP 3103, age 74

My biggest problem … I did feel embarrassed with some of the procedures, you know … [I'm] old fashioned, that's the trouble.

MSP 1117, age 84

If you're male and you've got three women … you know it's a bit embarrassing … I just convinced myself like you're an old man now, you know, they see these things all the time. So I managed to rationalise that … and I realise as I get older you might have to put up with more of that stuff.

MSP 1025, age 57

Communication

A key factor in patient acceptability was the extent to which nurses explained the test, supported patients with information provision in advance and through the procedure, and made them feel at ease. Patients reported they had been given the right level of information, so they understood the purpose of the test, and that the nurses made the procedure as comfortable as possible, which reduced their anxiety. The 'respectful' and 'professional' manner of nurses in discussing and conducting urodynamics also helped to reduce patient embarrassment.

They explained everything well … I thought it was quite an uncomfortable test and quite deep. But … they were very professional, the way they talked about it and dealt with it … they were friendly, they spoke to you during it, you know, you weren't just like lay there … they tried to make you as comfortable as possible.

MSP 2002, age 55

I found it quite easy, really, with the people I saw, you know, all the way through … That was nice not to have to worry about the people … you knew that they were explaining it to you and helping you, and guiding you, really.

MSP 3103, age 74

I was a bit worried at the start … but the girls that did it were, made you feel at ease and were brilliant, so I didn't really feel under any pressure.

MSP 3110, age 56

Preference for primary care

Undergoing the urodynamic test in primary rather than secondary care – and particularly at their own GP practice – was viewed positively by patients, mainly due to convenience (e.g. reduced waiting and travel time and ease of parking) and familiarity with surgery staff. Several patients explained they had 'full confidence' in their GP practice and felt relaxed about visiting the surgery, while they would be more apprehensive if they had to attend hospital for the same procedure. Some suggested that they would not have agreed to undergo urodynamics if the test had been conducted in hospital.

I [can't] see any benefit [to having the test done in hospital], I think it was better at my doctors, it was a lot more relaxed, because obviously I knew the nurse and I knew the doctor and you know, I had chatted with them about the test.

MSP 2002, age 55

It saved me going to hospital and … getting a taxi and one thing and another. I think that's more why I did it sort of thing I think at the doctors rather than traipsing round the country.

MSP 3110, age 56

It's more personal with the doctor. And he knows your history and that.

MSP 2006, age 77

Feasibility of invasive urodynamics in primary care

Training and support

Urodynamic nurses reported that facilitating urodynamics in primary care was generally a positive experience. Most had not independently performed invasive urodynamics before receiving training for the PriMUS study and were apprehensive about conducting the procedure in a community setting, but all felt their confidence grew as they gained experience. They valued the initial training and ongoing support provided by the study team, particularly regular teleconferences with peers and access to ad hoc telephone advice.

It was difficult to begin with and I like being challenged so that was interesting … Being able to go to the … urodynamics course was fantastic, that was really, well essential but really, really interesting and useful as well and the peer support that we get [from] the fortnightly nurse teleconferences [was] particularly useful at the beginning.

UN 506

We had issues on our very first patient and we were lucky that we're able to phone up … and say what shall we do about this … Having someone on the end of the line is reassuring.

UN 607

Logistical issues

Several logistical issues related to performing urodynamics in primary care were identified. Facilities at GP practices varied, and the procedure was sometimes carried out in unsuitable rooms, for example with a carpeted floor that was difficult to clean, insufficient space or difficulty accessing a sluice. Remote working meant nurses had to transport heavy equipment and sometimes experienced internet connectivity issues, which had led to clinic cancellations.

[The equipment is] heavy. It's really bulky. And it's sensitive equipment as well … It's been hard graft. You've got to get it all in the car, get it all out … up hills and over steps and in to little rooms. So it's not good for the equipment. It's not really good for my back [laughs] … It's quite stressful … making sure I've got every single thing to run the test.

UN 606

We don't always have the most appropriate room available to us in the GP surgeries … Sometimes the rooms are tiny, and you literally are falling over each other … sometimes the electricity sockets are not in the right place, so we have to use our extension lead … there's lots of sort of improvisation … I think we manage it quite well, but it is a challenge.

UN 607

Difficulties receiving and using results

General practitioners highlighted problems with receiving and utilising urodynamic results in the initial stages of the study. Study-specific quality assurance procedures meant GP summaries took longer to be returned than expected, and where concurrent diagnoses were identified, GPs were unsure which to treat first. It was suggested this limited the potential benefit of the study to some patients. As GP interviews were conducted in the pilot phase of the PriMUS study, their feedback enabled changes to be made to site training and the process of obtaining results.

You tell [patients that urodynamics] is the gold standard of investigation … And that it could help further … define what the problem is, and … target treatment a bit better … However … I've never had to look at … urodynamics or reports before. And therefore I feel what they're actually getting out of it is … some half-hearted interpretation of … what might

be the best management plan. So although they may be getting gold standard investigation, they're not necessarily getting gold standard advice.

GP 604

Recruiting to a urodynamic study

Importance of proactive recruitment

Main study recruitment was opportunistic or via primary care database searches, with most GP practices using a combination of these approaches. Healthcare professionals emphasised the importance of having a recruitment lead at each practice, to ensure database searches were carried out regularly and to remind others to recruit opportunistically. Maintaining telephone contact with patients invited to take part was highlighted as particularly effective in improving study recruitment and retention.

If you phone [the patient] a day or two beforehand, they're more likely to come in for [their study appointment] … Because, yeah, a couple of people have said, you know if you didn't phone I don't think I'd be here.

UN 606

Really you need somebody … leading the recruitment … making sure that that database search is done, the list is checked, and the appropriate letter sent out … Which … might slip off people's radar a little bit if it's not something that's done regularly.

GP 401

Reasons for participation and non-participation

When explaining the study to patients, clinicians emphasised they would have quicker access to a comprehensive diagnostic test not normally available in primary care, that could help in the diagnosis and treatment of their LUTS. Accordingly, most patients identified this as a key factor in their decision to take part. Another common reason for participation was the altruistic opportunity to contribute to research with the potential to improve medical practice and benefit others in the future. Some participants specifically desired to raise the profile of LUTS in men. Others wanted to participate to help their own GP surgery, particularly where they had built up a good relationship with their GP. Conversely, one patient who declined to take part in the main study explained he was not familiar with any of the GPs at his surgery. For patients who opted not to take part in the study, this was mainly because they did not want to have an invasive test. Those who participated appreciated that the research could be carried out at a local GP surgery, while those who declined generally mistakenly believed they would have to attend hospital.

To improve study recruitment, patients suggested clinicians should emphasise that participants would gain prompt access to a thorough assessment of their symptoms without needing to go to hospital. Additional reassurance about the urodynamic test was recommended, for example, explaining that most patients who experience the test find it acceptable, or giving a detailed timeline of the procedure to show that discomfort associated with catheter insertion is short-lived. Other suggestions included advertising the research more widely, and particularly having the study recommended by a familiar GP. Patients believed those who were uncomfortable with the thought of the urodynamic test would not take part in any instance.

I should benefit from it and so should other people. So to me, even though it was a little bit intrusive the test, I still think it was for the right reasons … just the knowledge that someone else may benefit from it, makes you feel better.

MSP 2002

[The doctor] said … the surgery [were] participating in the PriMUS study and would I be prepared to take part and I said yes because he's a fabulous doctor.

MSP 1117

> *I said yes, because it was at the surgery, had it … been at the hospital, I would have said oh no … you're talking an hour, an hour and a half one way … but because it was round the corner to my house, it was totally different.*
>
> *MSP 2002*

> *I don't like people prodding and poking around my private parts or anything [laughs] … I thought no way am I letting them mess around with me … because it's not a very sort of, what shall we say, palatable thing is it really?*
>
> *IP 301*

> *I didn't want [to do anything that] involved having to go to hospital … I would prefer not to go to hospital … to do any testing at all … If the GP wanted to do it, yes I don't mind that.*
>
> *IP 201*

Exploration of patients' and general practitioners' experiences of lower urinary tract symptoms

Interview data relating to patient and GP experiences of LUTs were organised into four main themes: unresolved symptoms, preference for primary care, satisfaction with involvement in decision-making and challenges of managing LUTS in primary care. Quotes in this section are labelled with each participant's unique identification number, prefaced by 'P' for patients and 'GP' for GPs.

Unresolved symptoms

None of the patients interviewed reported that their symptoms had been fully resolved following their visit/s to the GP. Most had received no treatment for their LUTS (treatment here refers to prescribed medication as opposed to lifestyle advice or watchful waiting). While in some cases this was because tests were still in progress, for other patients tests had been completed and no course of treatment had been prescribed. Patients believed this was because nothing could be done or because symptoms were normal for their age. Despite ongoing symptoms, some expressed satisfaction with GP consultations, as they had been reassured that there was no serious underlying cause such as prostate cancer.

> *[When my symptoms] first started happening obviously you read about things … the cancer thing … I was worried about that, and I went to see the doctor. And he took the blood tests … and everything, which came back okay, you know. So … that's a relief, takes a lot off my mind … I just accept the fact now it's part of growing older I suppose.*
>
> *P 2015*

> *It seems there was nothing [the doctor] could have done, I don't think. Because I'm not on any treatment, he didn't give me anything for it.*
>
> *P 3103*

Of those who had received treatment for their LUTS, some said this had no effect, while others reported that it had made some difference but not completely resolved bothersome symptoms.

> *I took [tablets] for 12 months, and I saw no difference whatsoever, so I stopped taking them.*
>
> *P 1002*

> *I had to get up about five or six times a night … it was really getting me down… They changed my medication to… Tamsulosin… and it, it just sort of keeps me down to about three times a night but it's still at least three times a night.*
>
> *P 1023*

Intolerable side effects of medication were reported, which had led in most cases where they occurred to treatment being discontinued or substituted. Several patients had tried multiple treatment options.

> *[The doctor] gave me some tablets, because of the frequency of getting up in the night. But I … wasn't very good with those, they swelled my ankles and I felt a bit dizzy with them. So I said to him, 'I'm not taking them because I'm not very happy with it' … So I've just put up with it, more or less. I kind of control it by not drinking. Which is not the best thing to do, of course, because you get dehydrated.*
>
> *P 3133*

> *I have had quite a difficult time with the tablets I was originally given … it knocked me for six, I never felt so awful taking a tablet that was supposed to help … I have been given some others, but I had to come off them, because I was so muzzy headed, weak …*
>
> *P 202*

Some patients were dissatisfied with the way their LUTS had been managed, due to the feeling that their symptoms had not been adequately explained or treated.

> *[The doctor] gave me some tablets … and I tried them … they give me a lot of nightmares and one thing and another. So I went back to [the doctors] and … he gave me the name of some herbal remedy to try and he said it's sort of due to my age … Which I was a bit disappointed about … I'm only fifty-six … so it's not exactly old.*
>
> *P 3110*

Patients did not all feel that their LUTS had been thoroughly examined and appreciated the opportunity to have a thorough diagnostic test as part of the PriMUS study.

Preference for primary care

Patients expressed a preference for having their LUTS treated in primary rather than secondary care. This tended to be because visiting the GP was more convenient, either due to the locality of the GP practice or due to shorter waiting times.

> *You go to the doctors, you can see the doctor the same day or the next day. You go to the hospital … you are going to wait for hours on end, because they are so busy. I wouldn't go to the hospital unless I felt I had a serious condition, that my GP would have referred me to the hospital [for] anyway.*
>
> *P 1002*

Patients also commented that they felt more comfortable at their GP surgery as staff were familiar and knew their history. This meant they were more relaxed about having potentially invasive tests. Some emphasised that they felt completely confident in their GP to provide their care.

> *For me, going to my GP makes it so much easier because he knows my history. Whereas in a hospital you start again … and then you get passed on to someone else and it just goes on and on and on … [my doctor] sort of remembered me, so that was so much easier than being with another, you know, strange doctor.*
>
> *P 2002*

> *You feel comfortable [at the GP practice], it's a more comfortable surrounding and all. You just go in there just in the room, one, one person you know … I would be totally confident in [my GP to run the tests that were done in hospital]. Everything that was done there, if it could be done at the GPs it would be brilliant.*
>
> *P 2015*

Patients generally felt there would be no benefit in attending hospital for the diagnosis and treatment of their LUTS. They believed secondary care would only be useful in certain circumstances, for example, if specialist advice or equipment were needed, for invasive and complex procedures, or in an emergency.

> *I suppose in a hospital, you're more likely to get specialist advice … [But going to hospital is] something to be endured, it's not really somewhere you want to go is it?*
>
> *P 2004*

Satisfaction with involvement in decision-making

Patients stated they would be more likely to adhere to recommended treatment for their LUTS if they felt involved in decision-making. Levels of involvement varied, with some patients following doctors' recommendations, others being informed of the reasons for their treatment, and others fully involved in decision-making. Despite this variation, all were satisfied with their level of involvement.

> *I think the more you understand … why the treatment is there, what it's aiming to do, I think it's easier to stick to it … In fairness … all the treatment I've had over the years, I've always felt my doctors have said, look this is why we're doing this … so I've always felt that I've been kept informed.*
>
> *P 1058*

Some patients reported that they would prefer treatment decisions to be made solely by their doctor, although it was acknowledged that the decision as to whether to proceed with medication was their own.

> *I was quite happy for [my GP] to just organise it all and get on with it.*
>
> *P 2001*

> *The way I see it, the doctors, they're the professionals. So basically, you go along with their advice … I'd always listen to what they say, but at the end of the day the decision's with yourself isn't it, you know. But I always take, try to take advice.*
>
> *P 2015*

Challenges of managing lower urinary tract symptoms in primary care

Challenges identified by GPs could be separated into those relating to diagnosis and those relating to treatment (*Figure 8*).

In terms of diagnosis, a key challenge was that the cause of LUTS can be multifactorial, and therefore difficult to establish. Compounding this, GPs reported that patients often present with mixed symptoms (i.e. both voiding and storage symptoms), making treatment decisions more complex. It was acknowledged that GPs may be less likely to identify uncommon causes of LUTS, such as urethral stricture. Diagnoses tended to be largely based on patient history, due to a lack of diagnostic tools available in primary care. However, reliance on patient reporting could be problematic, for example with difficulties obtaining accurate reports of patients' fluid intake.

> *The cause of lower tract infections can be multifactorial … that is one of the challenges, so [are symptoms] due to increased fluid intake, is it due to caffeine? You do have to take quite a detailed history and sometimes patients don't think about [what contains caffeine] or, you know, 'I don't really drink', [but they have] two glasses of whisky at night time.*
>
> *GP 401*

> *Sometimes you meet men where it's very clear what the cause of their LUTS is, but in the majority it's, it comes across as a mixed picture of urgency symptoms, voiding symptoms and you know, that can be difficult, and what you end up often doing is picking what you think is most likely and trying a treatment and then reviewing the patient.*
>
> *GP 502*

General practitioners highlighted the difficulty of differentiating between prostate and bladder symptoms to eliminate the possibility of prostate cancer. Greater awareness among patients due to public health campaigns meant this was a particular concern and often the reason patients in this study decided to visit their GP. As LUTS are common among older men, it was considered that increased awareness of prostate cancer symptoms may result in undue concern.

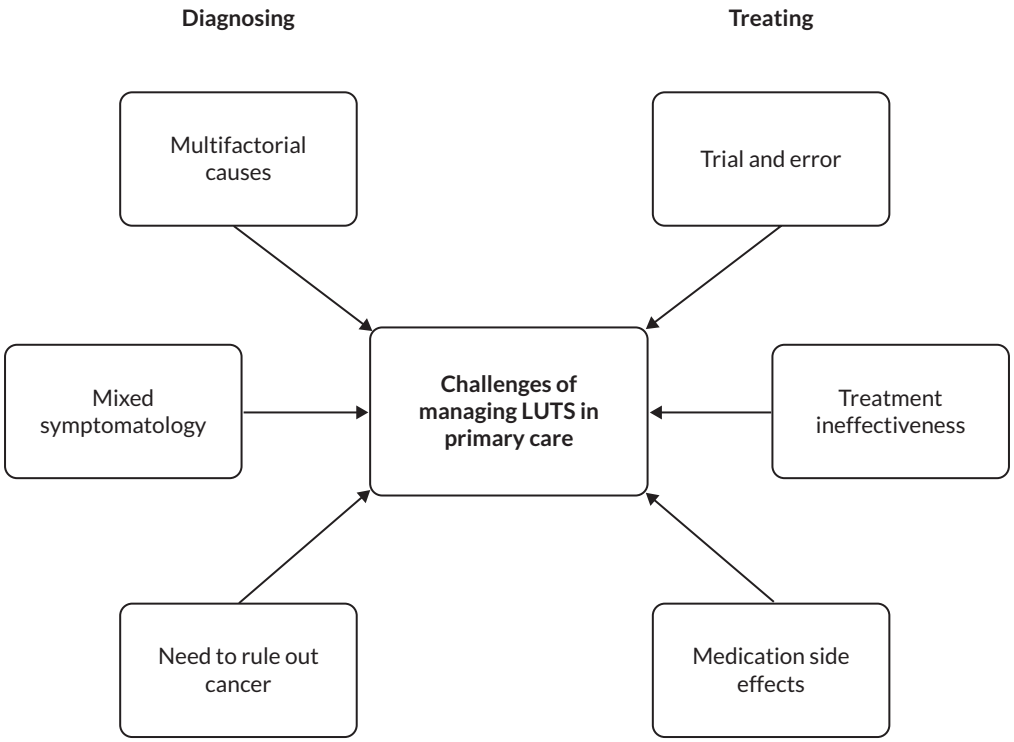


FIGURE 8 Challenges of managing LUTS in primary care.

> *A good thing is greater awareness [of] prostate cancer of course … but it can come with problems as well, because … the majority of men are going to have lower urinary tract symptoms at some point in their life, especially as they get older, and then I suppose a lot of men in that situation, their main concern is prostate cancer. So … if there was a really good prostate screening tool … that would be really helpful.*
>
> *GP 405*

Due to difficulties faced in diagnosing LUTS, treatment was described as a process of trial and error. A lack of guidance on treatment options was reported, resulting in uncertainty particularly where symptoms were mixed.

> *It is a bit of a stab in the dark very often … you kind of take [the patient's] history and you try some drugs, and you try some other drugs.*
>
> *GP 604*

> *I think sometimes it is just trying a few different medications and seeing which helps … It's not exact science … we try trial and error.*
>
> *GP 401*

General practitioners identified that available treatments for LUTS were often ineffective, and expressed concerns about the possibility of side effects, particularly for older patients. As reported by patients, this resulted in medication being discontinued in some cases and LUTS remaining untreated. GPs suggested the less harmful alternative of non-pharmacological approaches was not always considered.

> *The treatments I must say aren't the best, certainly we have quite a lot of unsuccessful … medication trials. Now whether that's due to patient compliance, side-effects or, you know, the expectations. So we are often changing medications over to an alternative one, and sometimes we just give up in the end.*
>
> *GP 403*

> *I guess the problem … is particularly in elderly men, a lot of these treatments have side-effects … there's lots of other non-pharmacological treatments … to give the patient that might be safer and more effective for them.*
>
> *GP 502*

Health professional feedback on the proposed clinical decision support tool

Twelve of the interviews conducted as part of the qualitative study were carried out with health professionals involved in consulting with patients with LUTS: 11 GPs and 1 Practice Nurse. These health professionals were asked to comment on the potential utility of the proposed clinical decision support tool, and on their preferences for the format and structure of the tool.

Potential utility of the tool

Health professionals agreed that a clinical decision support tool would be useful in their practice, helping in their decision-making process and in terms of guiding treatment choices. They reported finding other evidence-based tools useful and suggested that the tool could help standardise practice, which could be particularly useful for GPs dealing less frequently with LUTS in men. It could also act as a useful prompt to look for particular symptoms. It was identified that using a decision support tool could help facilitate shared decision-making with patients, allowing GPs to show and discuss with them the diagnosis and treatments suggested by the tool. Therefore, it was highlighted that the tool would need to use clear, understandable language so that it is accessible to patients.

Interviewees agreed that the tool could be a feasible alternative to urologist referral in some cases, which would benefit patients by allowing them to receive treatment in primary care and avoid long waiting times. Conversely, one GP suggested that the tool could be helpful in highlighting symptoms that may need management in secondary care.

Suitability of the tool for different patient groups

Health professionals identified a need for guidance as to which patients the tool would be appropriate for, as it may be unsuitable for those with comorbidities or suggest treatment options not appropriate for patients on particular medication. One GP emphasised the need to ensure that the tool is not adhered to rigidly but used to support decisions, with GPs considering their own knowledge of the individual patient in prescribing potential treatments. One interviewee reported that they may only use the decision support tool where they were unsure of a patient's diagnosis. It was highlighted that if patients were involved in using the tool and discussing treatment options, this could exclude some patient populations who may not be able to understand the tool.

Anticipated patient reaction

Health professionals identified that the main benefit of the tool for patients would be the ability to have their problem treated in primary rather than secondary care where possible. It was suggested that patients preferred being treated in primary care for convenience in terms of reduced waiting and travelling times, and familiarity.

Interviewees identified that, as previously stated, the tool could facilitate shared decision-making, helping patients to feel more involved in their treatment. It was suggested that this would make patients feel that they were being listened to and taken seriously. If patients had a copy of the tool, this could also help in telephone consultations. Additionally, if it could be partially completed by patients (e.g. if they could enter their symptoms into an online tool), this may reduce embarrassment for those uncomfortable discussing or raising their problems with a GP. However, it was highlighted that self-completion could discourage some patients, for example those with low literacy levels who may struggle to use the tool. Therefore, it was suggested that any self-completion element should be used selectively.

One interviewee emphasised the importance of ensuring GPs explained the tool fully to patients so that it did not appear as though the computer was making the decision about their treatment.

Concerns about the tool

Most health professionals had no concerns about the proposed clinical decision support tool and said that they would be confident in using it provided that it was evidence-based and validated. It was identified that if the tool recommended treatments that were controversial, such as PSA testing, this could be problematic. If it always recommended referral to secondary care this would limit the utility of the tool. Time taken to complete the tool was a key concern, and interviewees emphasised the importance of ensuring ease and speed of use.

Format and structure

Most interviewees stated that they would prefer an electronic tool, integrated into their existing computer system. This would enable the tool to be automatically pre-populated with patient information (such as age, test results, etc.), and thus make it quicker and easier to use. One GP suggested that the tool would be unlikely to be used unless it was integrated into the system and easily available. However, some interviewees stated that a paper-based flow chart would be useful, so that they could have a copy on their noticeboard or in their bag. In addition, if patients were involved in the use of the tool, it was suggested that older patients would tend to prefer a paper copy. One health professional proposed that although most patients would be comfortable with a web-based tool, it may be useful if it was available as a printable PDF so that a copy could be given to patients or stored in paper records.

Summary of feedback

- The clinical decision support tool should be evidence-based and validated, so that GPs are confident in using it.
- There is a need for accompanying guidance so that it is clear which patients the tool is appropriate for.
- Outcomes and treatment options to be suggested by the tool should be considered carefully, to maximise the usefulness of the tool.
- The tool should have the ability to be integrated into GPs existing systems, thus enabling pre-population of key patient details and ensuring speed and ease of use.
- Health professionals should have the ability to easily print out parts of the tool for their own or patient use.
- If the clinical decision support tool is intended to be used by or alongside patients, it should.
    - Use clear, understandable language.
    - Be available in both paper and online formats.

Workstream 2: Development of management recommendations to inform the clinical decision support tool (consensus study)

Objective

The objective of the consensus study was to inform the development of management recommendations to ultimately inform the clinical decision support tool. Specifically, this workstream aimed to establish how urologists would manage a number of different commonly encountered clinical scenarios, focusing on the thresholds at which they would recommend treatment and the strategies they would use when multiple urodynamic abnormalities are diagnosed or suggested.

Methods

Urologists were invited via e-mail to take part in the consensus study. They were e-mailed a copy of the study information sheet and given the option of participating in a telephone interview or completing an online questionnaire.

Questionnaire participants completed a short tick-box form at the beginning of the questionnaire to give consent. Consent for telephone interviews was given verbally at the beginning of the interview and was audio-recorded. Urologists participating in a telephone interview were e-mailed a copy of the scenarios prior to the interview. Eight clinical scenarios were presented in the online questionnaire, while four were presented in the telephone interview to minimise the burden on respondents. Scenarios were informed by real-life data generated by study participants.

Findings

There were 14 participants – 13 completed the online survey and 1 participated in a telephone interview. Participants were urology specialists with between 4 and 30 years of experience and were representative of 10 NHS Trusts in England and Wales.

Scenario 1: David

David, who is 70 years old, reports urinary symptoms for the last 8 years. He has a non-palpable bladder and his PSA is 2.0 ng/ml. His prostate on DRE is non-enlarged and non-nodular. He does not take any uroselective medication at present. He describes moderate storage symptoms. His bladder diary indicates that he voids 8 times per day, averaging 1520 ml output in 24 hours, and usually experiences urgency but manages to get to the toilet. His home flow study shows a maximum flow rate of 22 ml/second and voided volume of 180 ml on average. His post-void residual is 50 ml.

What would be your initial management of this patient?

A conservative management approach was proposed. All respondents said they would suggest lifestyle changes focused on fluid management, for example, reducing caffeine consumption. Eight of the 14 suggested bladder retraining. Pharmacological management was also proposed by eight respondents, mainly as a second-line treatment if lifestyle changes did not improve symptoms. Medication suggested was an anticholinergic/antimuscarinic, with one participant suggesting an alpha-blocker followed by an anticholinergic.

You use a diagnostic decision aid to help establish the cause of the patient's symptoms and this suggests a likely diagnosis of DO. Would this change your management of the patient?

All 14 participants responded 'no'.

The decision aid will provide you with a percentage likelihood of DO. What would the minimum percentage likelihood of DO have to be for you to decide to treat David for this problem?

Thirteen responses were received, ranging from 0% to 80% (*Figure 9*). Throughout this section, where respondents have given a percentage range (e.g. 60–70%), the middle value (e.g. 65%) has been used for reporting.

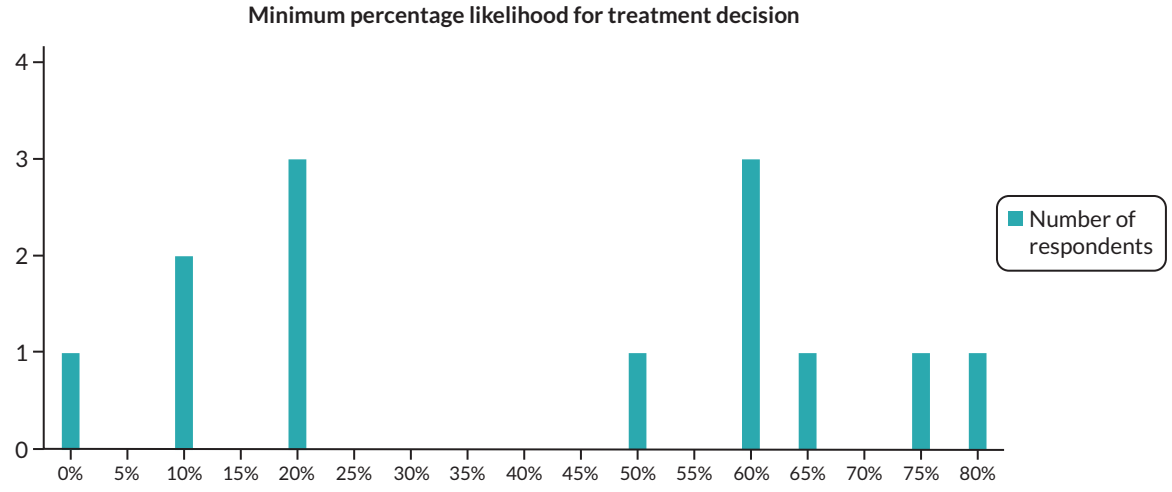


FIGURE 9 Scenario 1 – minimum percentage likelihood for treatment decision.

Of those who gave a reason for their response, two respondents explained that they would want a greater than average (50%) chance of the treatment working. Several said that they would not rely on the results of a diagnostic test to inform management of patients.

Scenario 2: Chris

Chris, a 54-year-old man, reports urinary symptoms for the last 2 months. He has a non-palpable bladder and his PSA is 1.8 ng/ml. His prostate on DRE is very enlarged and non-nodular. He does not take any uroselective medication at present. He describes moderate mixed voiding and storage symptoms. His bladder diary indicates that he voids 9 times per day, averaging 2990 ml output in 24 hours, and usually experiences urgency that passes away before he has to visit the toilet. His home flow study shows a maximum flow rate of 15 ml/second and voided volume of 370 ml on average. His post-void residual is 280 ml.

What would be your initial management of this patient?

The majority of respondents (10 of 14) suggested a combination of lifestyle advice (in terms of reducing fluid/caffeine intake and/or bladder retraining exercises or double voiding) and medical management. Some suggested further investigations, such as urodynamics (3) or a urine dipstick test (2). Of those proposing medical management, the majority (10) suggested an alpha-blocker (such as tamsulosin), with three suggesting combination therapy of an alpha-blocker and a 5-alpha reductase inhibitor (finasteride). One respondent suggested prescription of an overactive bladder medication (e.g. tolterodine) if the patient's urgency persists.

You use a diagnostic decision aid to help establish the cause of the patient's symptoms and this suggests a likely diagnosis of BOO. Would this change your management of the patient?

Thirteen participants responded 'no'; one responded 'yes'.

What would the minimum percentage likelihood of BOO have to be for you to decide to treat Chris for this problem?

Fourteen responses were received, ranging from 10% to 80% (*Figure 10*).

Of those who gave a reason for their response, those giving higher percentages suggested they would want there to be an above-average chance of the diagnosis before starting treatment, with this increasing for those considering surgery. Those who responded with lower percentages reported that given the patient's symptoms, they would start with medical management even without the decision aid.

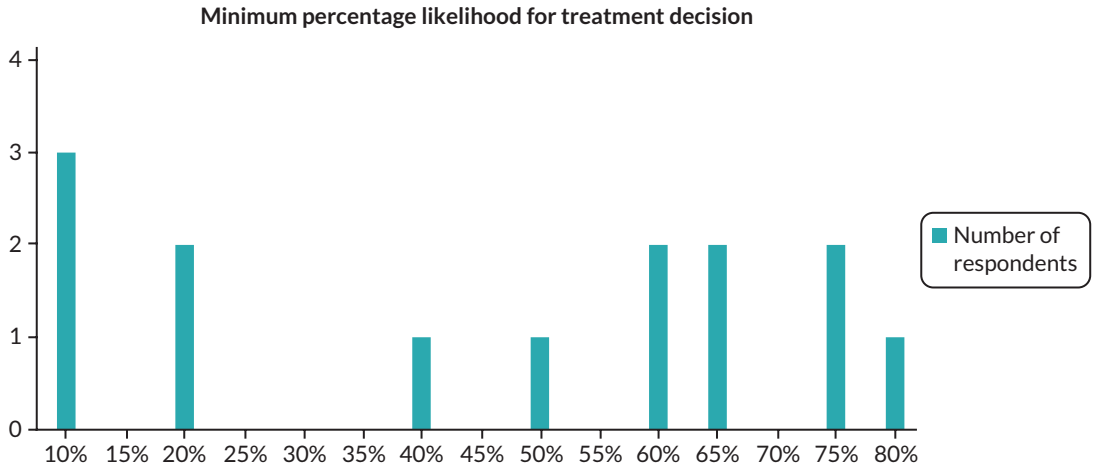


FIGURE 10 Scenario 2 – minimum percentage likelihood for treatment decision.

Scenario 3: Gordon

Gordon, a 69-year-old man, reports urinary symptoms for the last 7 months. He has a non-palpable bladder and his PSA is 0.4 ng/ml. His prostate on DRE is mildly enlarged and non-nodular. He currently takes tamsulosin 400 mcg and finasteride 5 mg/day. He describes moderate predominant voiding symptoms. His bladder diary indicates that he voids 9 times per day, averaging 2000 ml output in 24 hours, and usually has a normal desire to pass urine and no urgency. His home flow study shows a maximum flow rate of 11 ml/second and voided volume of 163 ml on average. His post-void residual is 320 ml.

What would be your initial management of this patient?

Most respondents (9 of 13) would consider surgical management, such as transurethral resection of the prostate (TURP) or Urolift (prostatic urethral lift procedure). Three suggested regular ISC and two suggested lifestyle modification. Eight would carry out further investigations, such as checking how long Gordon has been on finasteride and tamsulosin, finding out how bothersome his symptoms are, a repeat flow test, ultrasound scan, urodynamics/video-urodynamics, flexible cystoscopy or cystometry.

You use a diagnostic decision aid to help establish the cause of the patient's symptoms and this suggests a likely diagnosis of DU. Would this change your management of the patient?

Eight participants responded 'yes'; five responded 'no'.

Of those who said they would change their management of Gordon, four said they would consider ISC, three said they would arrange urodynamics, two would avoid surgery and one would discontinue the medications Gordon is taking. One respondent said that they did not understand how DU could be suggested by a diagnostic aid.

What would the minimum percentage likelihood of DU have to be for you to decide to treat Gordon for this problem?

Eleven responses were received, ranging from 20% to 80% (*Figure 11*).

Of those who gave a reason for their response, all had responded between 60% and 80%. Two felt that a decision support tool could not be used to diagnose DU, and that a test such as urodynamics or cystometry would be needed to confidently confirm the diagnosis. A further respondent also suggested that they would prefer to use urodynamics to obtain a more accurate assessment of Gordon. Two said that they had selected a higher percentage as they needed to be certain of the diagnosis, while one said they would require a high likelihood of diagnosis as Gordon does not have risk factors for DU (such as neurological conditions, diabetes or previous pelvic surgery).

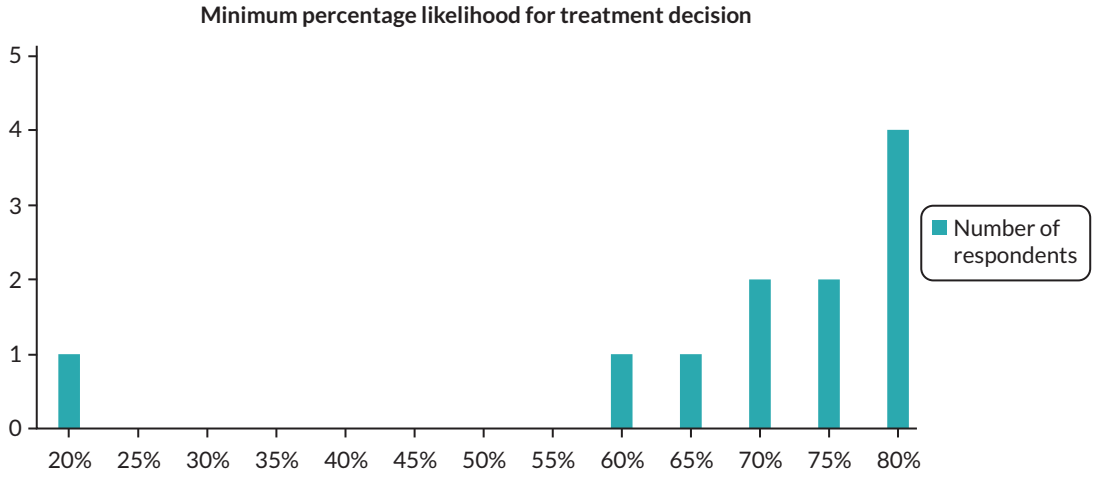


FIGURE 11 Scenario 3 – minimum percentage likelihood for treatment decision.

Scenario 4: Paul

Paul is a 69-year-old man, who reports urinary symptoms for the last 5 years. He has a non-palpable bladder and his PSA is 0.4 ng/ml. His prostate on DRE is non-enlarged and non-nodular. He currently takes tamsulosin 400 mcg and finasteride 5 mg/day. He describes moderate mixed voiding and storage symptoms. His bladder diary indicates that he voids 13 times per day, averaging 1430 ml output in 24 hours, and usually experiences urgency that passes away before he has to visit the toilet. His home flow study shows a maximum flow rate of 5 ml/second and voided volume of 90 ml on average. His post-void residual is 300 ml.

What would be your initial management of this patient?

Six of the 13 respondents said they would consider surgical options, such as TURP or bladder neck incision. Three suggested lifestyle measures (such as fluid advice, bladder retraining and/or double voiding), two suggested an anticholinergic and two suggested ISC. Seven respondents reported that they would arrange further investigations: five suggested urodynamics, while two proposed assessment of how bothersome Paul's symptoms are.

You use a diagnostic decision aid to help establish the cause of the patient's symptoms and this suggests a likely diagnosis of both DO and BOO. Would this change your management of the patient?

Six participants responded 'yes'; seven responded 'no'. Of those who said they would change their management of Paul, three suggested the decision aid outcome would result in them considering surgery: 'It would allow me to consider him for surgery and help with consenting'. One respondent said that urodynamic testing would be needed before a decision could be made about surgery or medical management. One suggested that the diagnostic decision aid could 'replace the need for urodynamics', depending on its specificity.

When treating Paul, which diagnosis would you prioritise for treatment?

Nine respondents said they would prioritise BOO, while four said they would prioritise DO.

Of those who would prioritise BOO for treatment, it was suggested that this was likely to be the underlying problem, and tackling this may help resolve Paul's DO. Two suggested that high post-void residual was their main concern as it increases the risk of infections, kidney damage and retention.

Of those who would prioritise DO, one respondent suggested that as 'detrusor overactivity is usually the more symptomatic component of LUTS', it would make sense to treat Paul's predominant symptoms. However, one respondent who had said they would prioritise BOO suggested that outflow obstruction appeared to be Paul's predominant symptom.

What would the minimum percentage likelihood of DO have to be for you to decide to treat Paul for this problem?

Thirteen responses were received, ranging from 0% to 90% (*Figure 12*).

Three participants gave a reason for their response. One who responded 10% said this was because treatment for DO is reversible. One who responded 70% said that BOO was a more common diagnosis than DO in men of Paul's age, especially with significant residuals. One who responded 90% suggested that even where DO was predominant with high residuals and slow flow, they would still treat Paul's bladder obstruction first.

What would the minimum percentage likelihood of BOO have to be for you to decide to treat Paul for this problem?

Thirteen responses were received, ranging from 10% to 80% (*Figure 13*).

Two participants gave a reason for their response. One who responded 50% said that BOO was the 'obvious' diagnosis, so they would tend to treat this unless there was good evidence of another diagnosis. One who responded 80% said this was because 'a high degree of certainty is required for surgical options'.

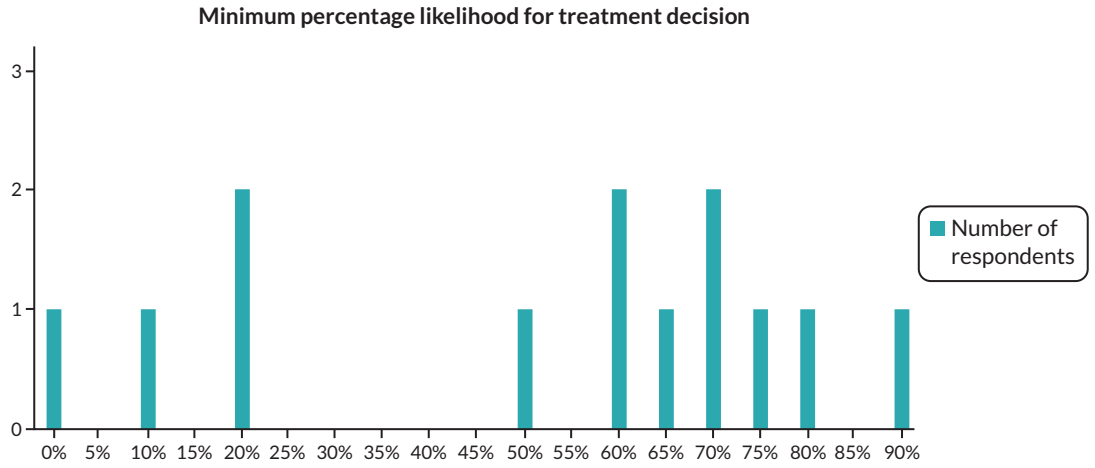


FIGURE 12 Scenario 4A – minimum percentage likelihood for treatment decision.

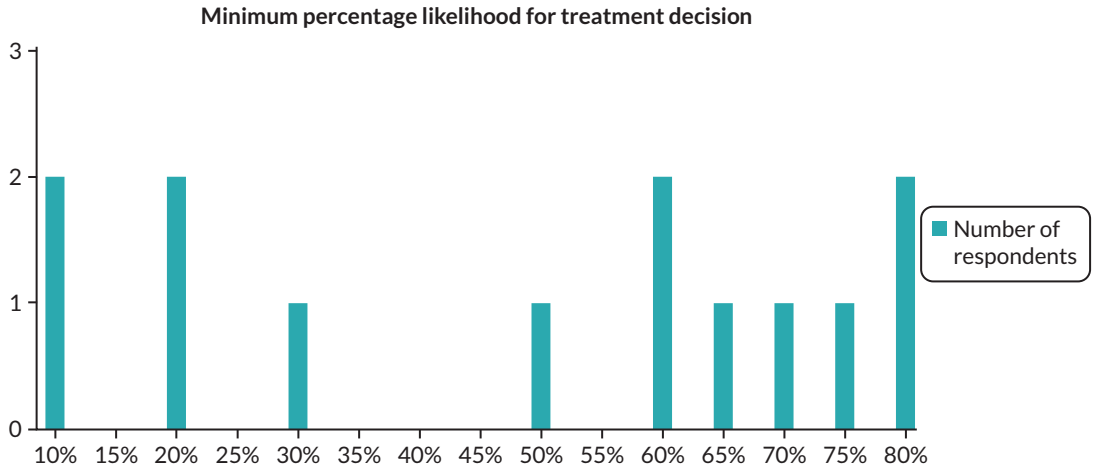


FIGURE 13 Scenario 4B – minimum percentage likelihood for treatment decision.

Scenario 5: Michael

Michael, who is 82, reports urinary symptoms for the last 2 years. He has a non-palpable bladder and his PSA is 3.1 ng/ml. His prostate on DRE is non-enlarged and non-nodular. He currently takes tolteradine 2 mg/day. He describes severe mixed voiding and storage symptoms. His bladder diary indicates that he voids 10 times per day, averaging 1110 ml output in 24 hours, and usually has a normal desire to pass urine and no urgency. His home flow study shows a maximum flow rate of 15 ml/second and voided volume of 140 ml on average. His post-void residual is 220 ml.

What would be your initial management of this patient?

Eleven of the 13 respondents recommended pharmacological management, although suggested medication varied. Three suggested Michael should stop taking tolterodine. Three proposed he should start taking mirabegron, and six suggested he should start taking an alpha-blocker (most commonly tamsulosin). Two suggested Michael's anticholinergic dose should be increased, while one suggested trialling a different anticholinergic. One suggested that due to Michael's age, if lifestyle measures do not work, a B3 agonist should be tried. One suggested combination treatment for outflow obstruction.

Lifestyle measures (such as fluid management) alone were proposed by two respondents, while two suggested a combination of lifestyle measures and pharmacological management. ISC was suggested by two respondents. One respondent did not consider Michael's symptoms to be severe, so questioned whether any treatment was necessary. One respondent said they would arrange urodynamics to investigate further.

You use a diagnostic decision aid to help establish the cause of the patient's symptoms and this suggests a likely diagnosis of both DO and DU. Would this change your management of the patient?

Seven participants responded 'yes'; six responded 'no'. Of those who said this would change their management of Michael, five of the seven said that following the results they would be more likely to consider ISC. Three said they would discontinue the alpha-blocker (one would replace with solifenacin).

When treating Michael, which diagnosis would you prioritise for treatment?

Seven respondents said that they would prioritise DO, while five would prioritise DU.

Of those who would prioritise DO, they reported that this was because this diagnosis was associated with the most bothersome symptoms for Michael.

Those who would prioritise DU suggested that it was important to tackle Michael's high post-void residual. It was proposed that this may also help his overactivity symptoms. One respondent suggested that most treatments for DO could exacerbate Michael's emptying symptoms.

One respondent queried how it was possible for patients to have both DO and DU.

What would the minimum percentage likelihood of DO have to be for you to decide to treat Michael for this problem?

Twelve responses were received, ranging from 10% to 80% (*Figure 14*).

Two participants gave a reason for their response. One who responded 60% said this was because they suspected DO from Michael's history. One who responded 50% said that as treatment would be given in response to Michael's clinical symptoms rather than the presence or absence of DO or DU, making a definitive diagnosis is not so important.

What would the minimum percentage likelihood of DU have to be for you to decide to treat Michael for this problem?

Eleven responses were received, ranging from 20% to 80% (*Figure 15*).

Three participants gave a reason for their response, all of whom gave 80% as the minimum percentage likelihood. Two suggested that, as DU is more difficult and invasive to treat, they would want a high degree of diagnostic certainty. One said that as BOO would be treated differently, they needed to be sure Michael did not have this diagnosis in order to not pursue treatment for this.

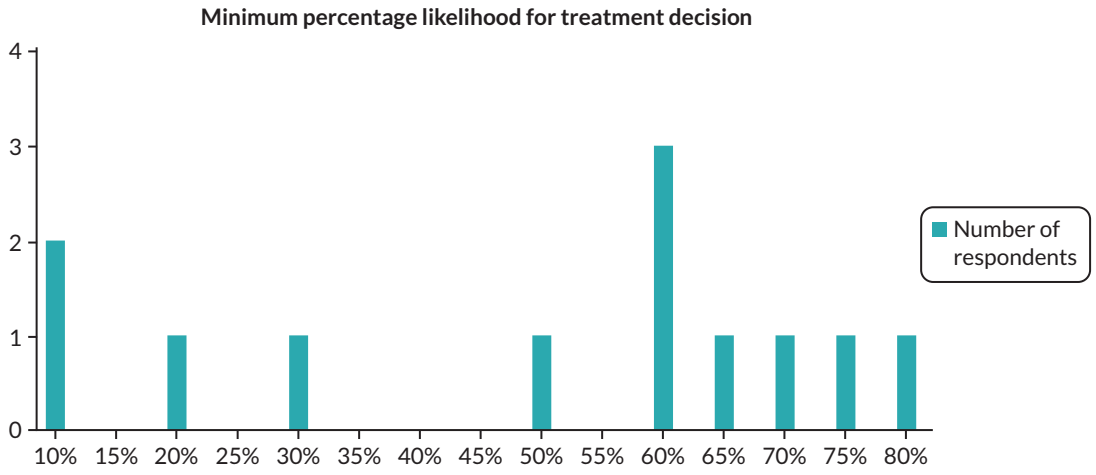


FIGURE 14 Scenario 5A – minimum percentage likelihood for treatment decision.

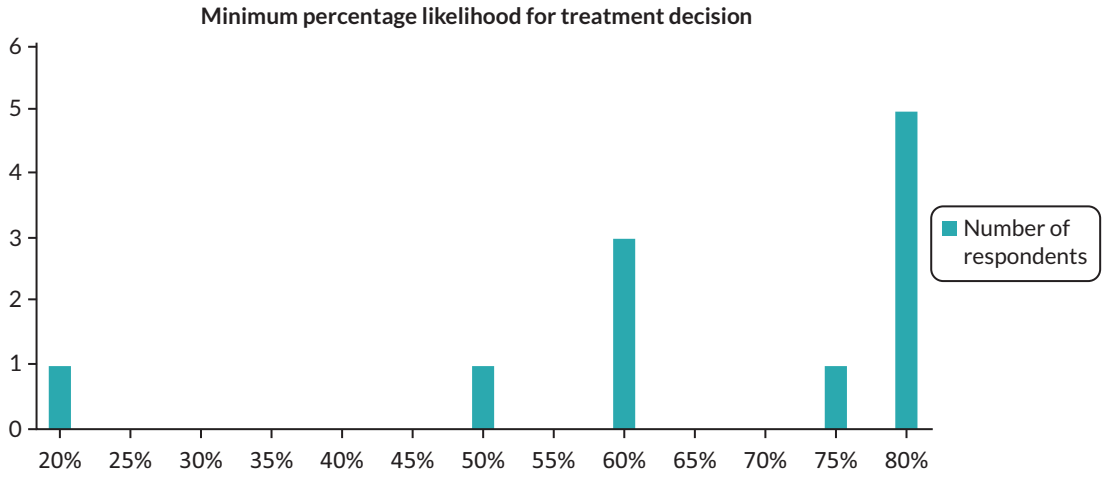


FIGURE 15 Scenario 5B – minimum percentage likelihood for treatment decision.

Scenario 6: Arjun

Arjun, a 66-year-old man, reports urinary symptoms for the last 12 months. He has a non-palpable bladder and his PSA is 3.9 ng/ml. His prostate on DRE is mildly enlarged and non-nodular. He does not take any uroselective medication at present. He describes moderate predominant voiding symptoms. His bladder diary indicates that he voids 12 times per day, averaging 1810 ml output in 24 hours, and usually has a normal desire to pass urine and no urgency. His home flow study shows a maximum flow rate of 6 ml/second and voided volume of 130 ml on average. His post-void residual is 300 ml.

What would be your initial management of this patient?

Twelve of the 14 respondents suggested pharmacological management. Six said they would prescribe an alpha-blocker alone (most commonly tamsulosin). Five suggested combination therapy (e.g. an alpha-blocker and a 5-alpha reductase inhibitor such as finasteride). Five proposed lifestyle modification, in the form of fluid advice, bladder retraining and/or double voiding. Four suggested discussion of bladder outflow tract surgery.

Three respondents reported they would arrange further investigations, including a repeat PSA, rechecking flow rate and post-void residual, checking renal function, a urine dipstick test and a LUTS questionnaire.

You use a diagnostic decision aid to help establish the cause of the patient's symptoms and this suggests a likely diagnosis of both BOO and DU. Would this change your management of the patient?

Three participants responded 'yes'; 10 responded 'no'. Of those who said they would change their management of Arjun, one said they would arrange urodynamics, one said they would recommend ISC, and one said they would check his post-void residual and monitor his renal function.

When treating Arjun, which diagnosis would you prioritise for treatment?

Twelve respondents said they would prioritise BOO, while two said they would prioritise DU. Of those who would prioritise BOO, it was suggested this was the most treatable (4) and more bothersome (2) diagnosis. It was also suggested by three respondents that treatment for BOO might decrease the post-void residual and make DU clinically less relevant. Those who said they would prioritise DU did not provide a reason why.

What would the minimum percentage likelihood of BOO have to be for you to decide to treat Arjun for this problem?

Fourteen responses were received, ranging from 20% to 80% (*Figure 16*).

Two participants gave a reason for their response. One who responded 50% suggested that if there is a chance Arjun has BOO, he should be offered TURP 'as there is little to lose by giving this a try'. One who responded 60–70% said that at Arjun's age, BOO is a more common problem than DU.

What would the minimum percentage likelihood of DU have to be for you to decide to treat Arjun for this problem?

Thirteen responses were received, ranging from 0% to 90% (*Figure 17*).

Two participants gave a reason for their response. One who responded 80% said they would need to be certain Arjun has DU as otherwise he could potentially benefit from TURP. One who responded 30–40% said that at Arjun's age, BOO is a more common problem than DU.

Scenario 7: Donald

Donald, an 81-year-old man, reports urinary symptoms for the last 10 years. He has a non-palpable bladder and his PSA is 1.2 ng/ml. His prostate on DRE is moderately enlarged and non-nodular. He currently takes tamsulosin 400 mcg/day. He describes moderate mixed voiding and storage symptoms. His bladder diary indicates that he voids 9 times per day, averaging 1280 ml output in 24 hours, and usually has a normal desire to pass urine and no urgency. His home flow study shows a maximum flow rate of 6 ml/second and voided volume of 180 ml on average. His post-void residual is 280 ml.

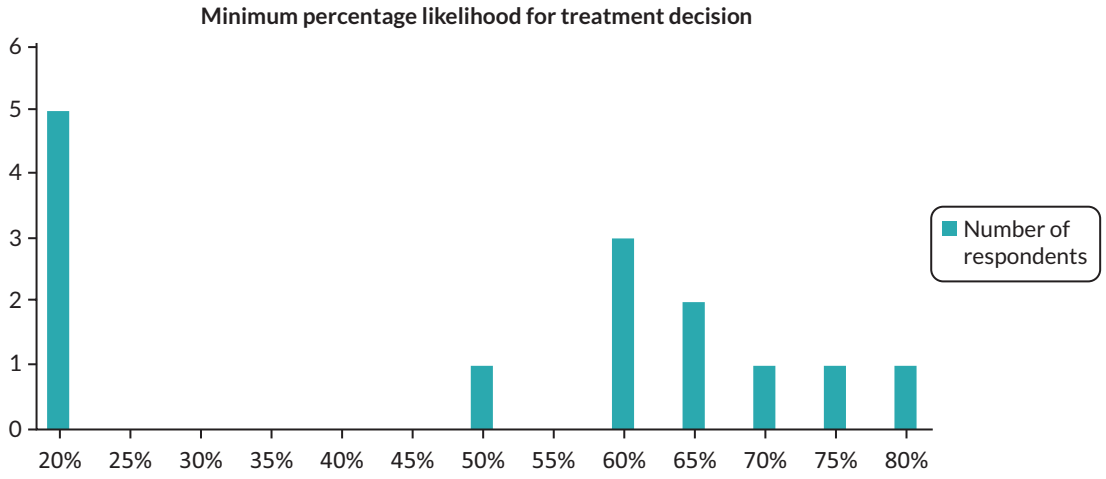


FIGURE 16 Scenario 6A – minimum percentage likelihood for treatment decision.

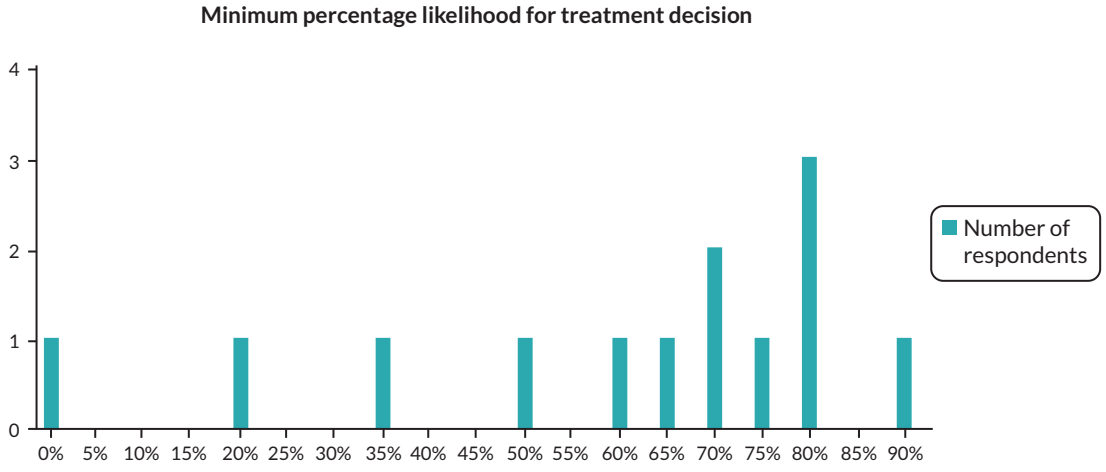


FIGURE 17 Scenario 6B – minimum percentage likelihood for treatment decision.

What would be your initial management of this patient?

Ten of the 14 respondents said they would prescribe finasteride, with one suggesting the addition of an anticholinergic. Lifestyle measures (fluid management and/or bladder retraining) were proposed by three respondents. Two suggested the possibility of no treatment, two said they would consider surgery (e.g. TURP), and two suggested ISC. Several proposed further investigations, including assessing how bothersome Donald's symptoms are (3), checking his renal function (1), videourodynamics (1) and repeating the flow test in 3–6 months (1).

You use a diagnostic decision aid to help establish the cause of the patient's symptoms and this suggests a likely diagnosis of DO, BOO and DU. Would this change your management of the patient?

Four participants responded 'yes'; 10 responded 'no'. Of those who would change their management of Donald and gave a reason for their answer, one said they would be more likely to consider treating Donald for DO and would monitor his post-void residual. Two said they would consider surgery for BOO (such as TURP), and one said they might consider ISC.

When treating Donald, which diagnosis would you prioritise for treatment?

Eleven respondents said they would prioritise BOO, two said they would prioritise DO and none said they would prioritise DU. Of those who would prioritise BOO, four suggested that treating this diagnosis might have a positive impact on Donald's other conditions. Three proposed that treating BOO would give a chance of symptomatic improvement. One suggested Donald's symptoms were more aligned with BOO, as he is not suffering from urgency but has a slow flow, and one said that treating DO could exacerbate Donald's retention and poor flow. Those who would prioritise DO for treatment said that this could produce the most bothersome symptoms.

What would the minimum percentage likelihood of DO have to be for you to decide to treat Donald for this problem?

Twelve responses were received, ranging from 20% to 80% (*Figure 18*).

Two participants gave a reason for their response. One who responded 'not applicable' said that the presence of DO did not matter unless Donald had significant storage symptoms, and that they would target Donald's voiding issues and retention initially in any case. One who responded 50% said that due to Donald's long history, his symptoms could be a result of just BOO or DU.

What would the minimum percentage likelihood of BOO have to be for you to decide to treat Donald for this problem?

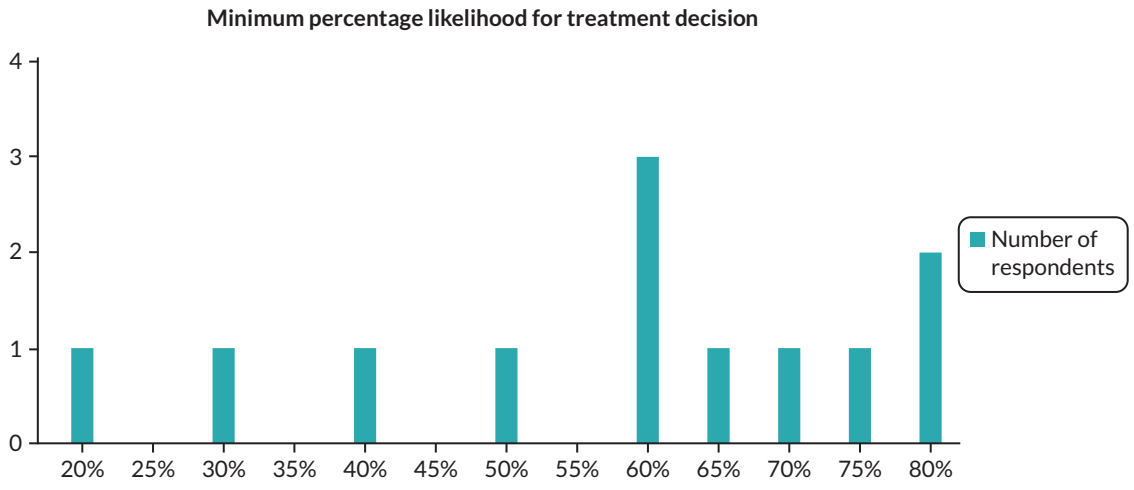


FIGURE 18 Scenario 7A – minimum percentage likelihood for treatment decision.

Thirteen responses were received, ranging from 10% to 80% (*Figure 19*).

Two participants gave a reason for their response. One who responded 50% said that if Donald has BOO and this is treated, he has a reasonable chance of improvement of his symptoms, so that as long as risks of surgery are not high, it would be reasonable to try treating him for this diagnosis. One who responded 30–40% said that it would be very common for a patient with a long history like Donald to have BOO.

What would the minimum percentage likelihood of DU have to be for you to decide to treat Donald for this problem?

Twelve responses were received, ranging from 0% to 100% (*Figure 20*).

Two participants gave a reason for their response. One who responded 'not applicable' said that DU was more a diagnosis of exclusion, and that if there was no evidence of BOO (or if Donald did not improve after treatment for BOO), he would have to be treated conservatively or try ISC. One who responded 60–70% explained: 'With his age, if he doesn't have symptoms I'm not just going to treat his post-void residual. The likelihood for him to have symptoms is very high'.

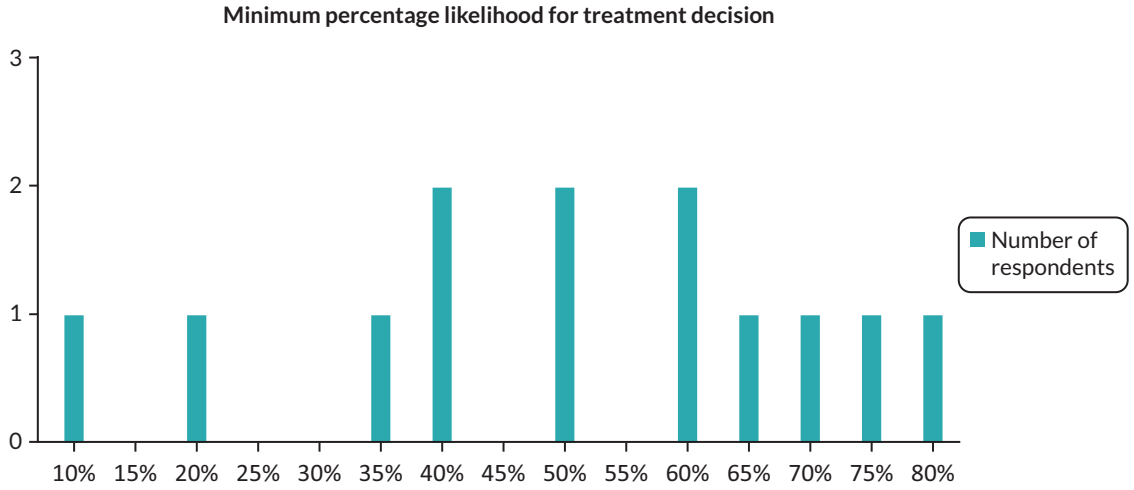


FIGURE 19 Scenario 7B – minimum percentage likelihood for treatment decision.

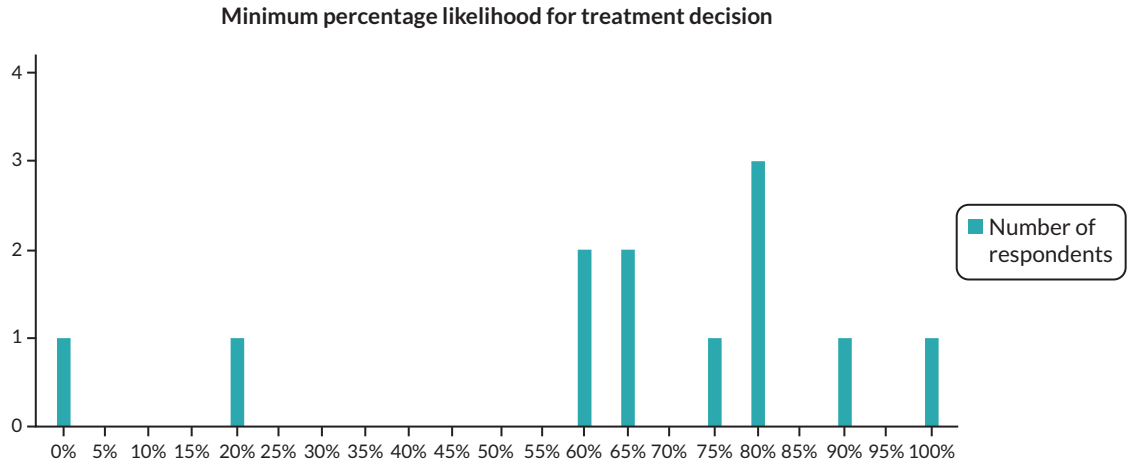


FIGURE 20 Scenario 7C – minimum percentage likelihood for treatment decision.

Scenario 8: Jim

Jim, who is 78 years old, reports urinary symptoms for the last 10 months. He has a non-palpable bladder and his PSA is 0.4 ng/ml. His prostate on DRE is mildly enlarged and non-nodular. He does not take any uroselective medication at present. He describes very mild mixed voiding and storage symptoms. His bladder diary indicates that he voids 6 times per day, averaging 1940 ml output in 24 hours, and usually has a normal desire to pass urine and no urgency. His home flow study shows a maximum flow rate of 24 ml/second and voided volume of 300 ml on average. His post-void residual is 20 ml.

What would be your initial management of this patient?

A conservative management approach was proposed. Nine of the 12 respondents said they would give Jim lifestyle advice, for example, regarding his fluid or caffeine intake. Two also suggested bladder retraining. It was suggested that Jim did not need any treatment; however, one respondent proposed a trial of anticholinergics.

You use a diagnostic decision aid and the result is normal – i.e. it shows no DO, BOO or DU. Would this change your management of the patient?

One participant responded 'yes'; 11 responded 'no'. The respondent who said the decision aid result would change their management of Jim said it would reassure him that no treatment was necessary and lead to him giving lifestyle advice.

Application of findings

Primary care Management of lower Urinary tract Symptoms Study Management Group members with a background in urology considered the above findings in conjunction with available evidence/guidelines to inform the development of the draft management recommendations.

Workstream 3: User-testing of the prototype clinical decision support tool

An early prototype version of the clinical decision support tool was developed using hypothetical data sets (without the mathematical modelling) and underwent user-testing with GPs. The objective of the user-testing was to build on the interviews conducted as part of the pilot phase evaluation to gather participants' views of the proposed tool, in terms of content, design, and perceived acceptability and feasibility of using the tool in primary care settings. Feedback was used to improve and refine the tool.

Methods

An e-mail was sent to all GPs involved in the PriMUS study (with the exception of GP members of the SMG) asking if they would be willing to test out a prototype clinical decision support tool and provide feedback via a brief telephone interview. Those who wished to take part were e-mailed a link to the tool and asked to try entering fictional patient data and generating a diagnostic report. They were provided with no instructions of how to use the tool other than those embedded within the tool itself. They were then asked to arrange a convenient time to take part in an interview.

Telephone interviews were conducted in January and February 2020 with 10 GPs; 6 from GP practices in Wales (across 3 health boards), 2 from practices in Bristol and 2 from practices in Newcastle. Interviews took between 11 and 20 minutes (mean 15.1 minutes). Interviews were audio-recorded, transcribed verbatim and coded for key themes using NVivo version 11.

Findings

Three key aspects of the tool were explored: (1) its design and ease of use; (2) its content and (3) feasibility and acceptability of using the tool in a specialist GP clinic.

Design and ease of use

Ease of use and layout

All GPs reported that the tool was very user-friendly. They found it logical, easy to navigate and straightforward to use. They commented positively about the sliders used for data entry (*Figure 21*) and found them preferable to text boxes. It was easy and quick to generate a results report. Most felt no additional instructions or guidance would be needed to accompany the tool. However, one GP commented that in the IPSS input section, where symptoms are given a score of 0–5, it would be useful to have a brief explanation within the tool of what each number constitutes. GPs liked the layout of the tool and found the text size suitable.

Web-based versus integrated tool

All GPs interviewed would prefer the tool to be integrated within the clinical system rather than being web-based, although some said they would still be happy to use a standalone web-based tool. The advantage of an integrated tool was that some of the input measures could be completed automatically from patient records, and the results report would be integrated into the record for easy reference. Using a web-based tool would be more time consuming as GPs would have to manually transfer information between patient records and the tool; some felt this sort of tool would be used less.

One GP suggested that the tool would be an opportunity to set up user-friendly integration between patients' smartphones and the clinical system, so that patients could input information about their symptoms into an app, which could communicate with the clinical system and the decision support tool.

Format of results

It was suggested that a downloadable report of results would be useful (possibly in PDF format) which could be added to patient records and printed if necessary, for example so that patients could take a copy home. Being able to save results would give a good record of management decisions and enable GPs and patients to see changes in symptoms over time.

Content

Input measures

There are four tabs to be completed in order to generate a results report:

1. Demographics/self-report
2. IPSS
3. Uroflowmetry
4. Management history

General practitioners generally reported that they would have no problem completing the Demographics/self-report, IPSS and Management history tabs. However, they would not be able to complete the Uroflowmetry information as

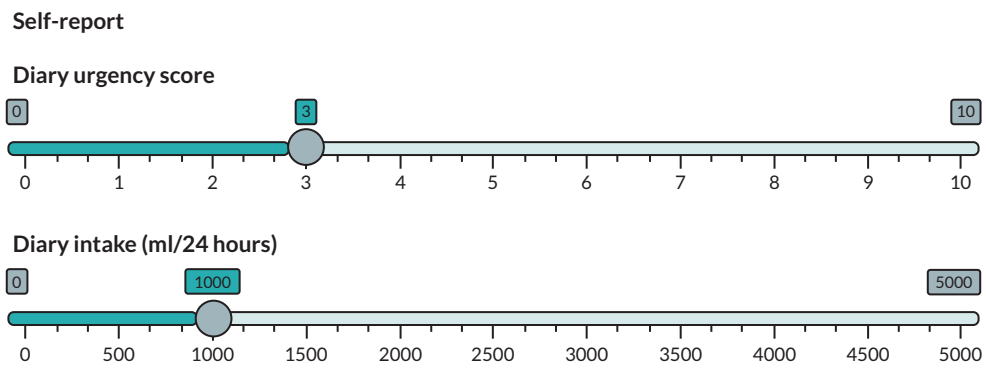


FIGURE 21 Example of sliders used for data entry.

they did not have flow meters available in primary care to give to patients. Even if flow meters were available, it was pointed out that post-void residual would be difficult to ascertain accurately without a bladder scan.

Some GPs were not familiar with the diary urgency score and intake measures – although these were measures in the PriMUS study, it did not appear they were used routinely in practice. However, once it was explained that this would be a bladder diary completed by the patient, GPs felt this would be relatively straightforward to obtain. One GP suggested that fluid intake may be difficult for patients to measure unless they only drank from one bottle. It was also pointed out that while other measures (such as IPSS score and management history) could be completed during the consultation, asking patients to complete a diary would necessitate more than one appointment. It was suggested it would be useful if there was information within the tool explaining how self-report measures could be obtained. Some GPs felt that the process of completing input measures would be too time consuming for a standard 10-minute GP consultation, but more practical if this took place in a specialist clinic.

A common query was whether the tool could still give valid results if not all input measures were completed – it was felt that this should be made clear within the tool. There should also be the option to indicate that measures had not been completed, rather than having to leave them blank or set at zero.

Results

General practitioners felt that the results report (*Figure 22*) was clearly laid out, easy to interpret and included all necessary information. They found the visual representation of disease probabilities useful.

Some felt that the statistics report was slightly long and duplicated information elsewhere; the initial rows were populated from inputs, so they did not give any new data, and the disease probability rows were duplicated in the bar chart of predicted disease probabilities. However, some did think having a clear record of symptoms could be helpful for treatment reviews, to enable GPs and patients to see changes over time.

There was an option within the tool to display detailed results with several more complex charts; however, GPs said they would probably not use these.

Management recommendations

General practitioners felt that the management recommendations generated by the tool were straightforward and easy to follow. Several suggestions for useful additions to this section were made:

- Links to printable patient advice leaflets or a brief explanation of terms such as fluid intake, bladder training, double voiding etc.
- Brief information of what to include in discussions with patients (or information printouts to give to patients) about particular medications, for example, side effects.
- Flagging up of serious symptoms (e.g. abnormal PSA) that require immediate referral.
- Specific examples of medications to try (linking in with local guidelines) rather than generic suggestions. This should also take account of the patient's age and comorbidities, so more information may be required.
- A disclaimer statement emphasising that the tool cannot be completely accurate in all cases, and to refer to a urologist if there is suspicion about other disease (e.g. if there is blood in the urine).

It was pointed out that if there was a similar or identical probability of more than one diagnosis, GPs would not know which to treat. If probabilities were very similar, GPs would need to trial treatment for one diagnosis and then review the patient, which is what they would do without the tool. It would therefore be useful for guidance on treatment priorities or combinations to be built into the tool, as diagnoses are not always clear-cut. Similarly, as there is a separate flow chart for each diagnosis, if the recommendation were to treat for DO, the flow chart would direct GPs to try one type of medication and then to review this at 4–6 weeks and 6–12 months. If unsuccessful, the next step would direct GPs to refer to a urologist, without suggesting combination or alternative treatment, or taking into consideration the probabilities of other diagnoses. In many cases, symptoms are mixed, so accounting for combinations of treatment and more than one diagnosis was considered important.

Statistics

	Description	Value	Units
Age	Age in years on date of study	50	years
PSA	Prostate-specific antigen	10	ng/ml
Storage	IPSS storage symptoms	9	/15
Voiding	IPSS voiding symptoms	12	/20
Intake	Diary daily fluid intake	1000	ml
Urgency	Diary urgency symptom score	3	
Qmax	Maximum urine flow rate	10	ml/s
Vvoid	Mean voided volume	250	ml
Frequency	Mean daily frequency	10	voids/day
Nocturia	Mean nocturia	1	voids/night
Residual	Residual bladder volume	0	ml
P(DO)	Probability of detrusor overactivity	67	%
P(DU)	Probability of detrusor underactivity	42	%
P(BOO)	Probability of bladder outlet obstruction	49	%

Predicted disease probabilities for one patient

This graph summarises the probability for your patient for each of the three diseases. Probabilities are calculated using the derived logistic regression model.

Predicted disease probabilities

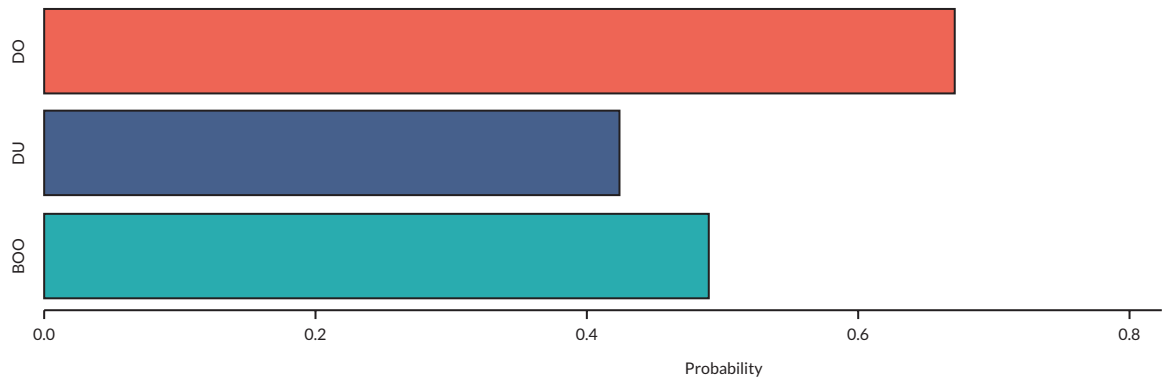


FIGURE 22 Results report.

Acceptability and feasibility

Acceptability of using the tool in practice

General practitioners found the tool straightforward to use and felt that it would be helpful in practice, by reminding them of factors to consider, supporting their diagnosis and guiding treatment. It was suggested that the tool could standardise investigation and management of LUTS. GPs generally felt that the tool would change practice, particularly in terms of the selection of treatment options and the order in which these were trialled. They proposed that the tool had the potential to reduce urology referrals, as it could help symptoms to be better managed in primary care, provided

that uroflowmetry data were accessible. As long as the tool was validated and the issues with uroflowmetry data could be resolved, GPs said they would have no concerns about using the tool in their practice.

It was felt that the tool would be helpful for all patients presenting with LUTS (with the exception of those with suspected prostate problems), provided that it was validated with different groups. Some GPs said they would only use the tool where there was not an obvious diagnosis or management option. They emphasised the importance of the tool providing clear management recommendations where patients present with mixed symptoms.

Feasibility of using the tool in practice

It was explained to GPs that the decision support tool was designed to be used as part of an extended consultation in a specialist primary care clinic, rather than in the usual 10-minute GP consultation. GPs agreed that this would be feasible if a specialist clinic service were set up, although this would need to be funded and patients consulted. Some felt that an extended consultation would be necessary as it would not be practical to use the tool in a shorter time. However, a specialist clinic model could result in more steps to take for patients and GPs: for example, GPs would need to refer patients to the specialist clinic, who would carry out the tests and pass the treatment recommendations back to the GP who could then prescribe medication. Some GPs felt that the tool would be useful in their own surgery, and that it could be used over two 10-minute appointments. One suggested that the tool would be much more useful if it could be used by GPs in usual consultations. However, it was identified that specialist clinic referral would be needed in order to obtain uroflowmetry data.

General practitioners felt that much of the patient data required by the tool could be entered by practice support staff. However, some suggested they would want to input data themselves so that they were aware of the patient's symptoms and PSA score, etc. which could help inform diagnosis and management decisions. If completing previous medical management information, support staff may need additional guidance on how to identify trialled medications, for example, alpha-blockers or anticholinergics.

General practitioners felt that the tool would be helpful in involving patients more in their treatment, as it would allow them to see their symptoms being input and how this led to results and management recommendations. This would enable them to see how their data suggested the likely diagnosis, rather than just being told by the GP. GPs suggested this link would also make it easier to show patients how different treatments could affect different symptoms. Most felt that results were presented in a way that would be clear to patients and that the bar chart and flow charts were useful visual tools which would aid understanding and could also help patients to see how their symptoms changed over time. Having results displayed in several ways was identified as positive as this could cater to different preferences and understanding. GPs felt that the disease probabilities would be easy for patients to understand, as they could explain that, for example, '60 out of 100 patients with symptoms like yours have bladder outlet obstruction'. Having probabilities displayed as percentages throughout the results report (rather than as a probability between 0 and 1 on the bar chart) would make results more straightforward to explain to patients.

It was suggested that if GPs showed patients the management recommendations, having three flow charts on the screen could be confusing, so it would be helpful if there was clear highlighting or separation of the management flow chart recommended by the tool. If there was a very low probability of a particular diagnosis, possibly the flow chart for that diagnosis should not be displayed at all. Some felt that the language used was too clinical and not patient-friendly (e.g. 'predicted disease probabilities'; 'logistic regression'). It was suggested this could either be simplified or that there could be a separate tab with results in a more patient-friendly format.

Summary and recommendations

General practitioners reported that the tool had a user-friendly design and layout, and that it was easy to generate a report of results. Management recommendations were considered useful and straightforward to follow. GPs felt that the tool would be helpful in practice and had the potential to reduce urology referrals. They suggested that it could facilitate greater involvement of patients in their treatment, as it provided useful visual displays of likely diagnoses and management recommendations.

A number of recommendations were made relating to the design, content and feasibility of the tool:

Design

1. The tool should be integrated into the clinical system, so that input measures can be populated automatically, and the results report easily added to the patient record.
2. Disease probabilities should be presented as percentages in the bar chart rather than probabilities of 0–1, to aid patient understanding.
3. The recommended management flow chart should be highlighted to make the suggested course of action clearer.

Content

4. There should be a brief explanation within the tool of input measures that GPs may not be familiar with, such as diary urgency score and intake measures.
5. The option for more detailed results could be removed as it is unlikely to be referred to.
6. Links to printable patient advice leaflets within the management recommendations would be useful, for example to explain terms, such as double voiding, bladder training, etc.
7. Specific examples of medications to try (linking in with local guidelines) rather than generic suggestions would be helpful.
8. More patient-friendly language could be used in the results report; alternatively, there could be separate tabs for GPs and patients.
9. Serious symptoms (such as abnormal PSA) that require immediate referral should be flagged up and a disclaimer statement added, emphasising to refer to a urologist if there is suspicion of other disease.

Feasibility

10. The tool should be validated with a range of patient groups, to ensure GP confidence.
11. Clarity is needed as to whether management recommendations are still valid if there are missing data, particularly uroflowmetry data which GPs may not be able to obtain.
12. Consider making the tool available for use in usual consultations, subject to the availability of/necessity for uroflowmetry data.
13. The tool should be able to guide GPs regarding combinations of treatment approaches (either concurrent or successive) where patients have mixed symptoms, rather than suggesting urology referral after one treatment has been unsuccessful.

Chapter 6 Discussion

Principal findings

The PriMUS study, a diagnostic accuracy study, set in routine UK NHS practice, has shown the ability of a clinical assessment tool, derived from a mathematical model, to estimate the likelihood of a range of urodynamic conditions using simple index tests which could be carried out in primary care.

The variables that can be used in our proposed mathematical model and clinical diagnostic tool include simple demographic data, symptom scores, data derived from a patient completed bladder diary and a blood test (PSA) that is routinely taken by primary care staff. Other variables required for the model include simple measurements that can be obtained from a home uroflowmetry test. This study has demonstrated that a home uroflowmetry test can generate valuable data which are helpful in making a urological diagnosis. Results from the trial show that 92.5% (556/601) of men (with useable data) were able to successfully use the home uroflowmetry device and provide usable data to input into the mathematical model. This work would suggest that the addition of a simple home uroflowmetry test, used alongside existing investigations already carried out in primary care, may enable men presenting with LUTS to be diagnosed and potentially treated within general practice, thus potentially avoiding referral to specialist urology clinics. The embedded qualitative work has highlighted the preference of patients to undergo initial investigation and management outside of the hospital setting and the results from this study may encourage pathway development which could change the initial clinical assessment of men presenting with LUTS.

A striking finding from the study was the significant prevalence of urological conditions in a cohort of men presenting to primary care with LUTS. Of the 601 men recruited to the study, 545 (91%) were found to have at least one urological condition on urodynamic testing. Baseline demographic data from participants indicate that most men had not yet tried but would be suitable for first-line medical treatment of these conditions. This suggests that they would not necessarily need to be referred to secondary care and could have their initial management supervised by primary care. For example, in both the development and validation cohorts in this study, around 90% of men with DO had not yet tried an anticholinergic medication or a beta adrenoreceptor agonist and around 70% of men with BOO had not yet trialled an alpha adrenoreceptor antagonist. The use of our management algorithm, based on existing NICE guidelines, could further facilitate the initial management of these patients within a primary care setting.

Impact of patient and public involvement

Patient and public involvement was embedded in the PriMUS study from conception to completion. The active involvement of our PPI representatives in reviewing the design, execution and documentation made a valuable contribution to the study, as did regular PPI participation in the project review meetings. The impact of this was as follows:

- The study design was acceptable, and the study documentation appropriate, to enable us to recruit a sufficient number of participants to achieve the study aims.
- We believe that study participants were provided with full and clear information about their involvement and had a positive experience of taking part in the research.
- We responded to a serious incident in a way that respected the best interests of the participants.
- We have created a healthcare intervention that is feasible for and acceptable to patients.
- We have patient-centric feedback on improvements to the design and function of the prototype online decision support tool that can be implemented in further development.
- We have created a *Plain language summary* that is comprehensible to patients and the public, and our discussion and conclusions have been reviewed from a patient-representative perspective.
- We have developed a PPI model and resources that can be used in future research.

Strengths and weaknesses of the Primary care Management of lower Urinary tract Symptoms study

This trial was conducted in line with current Good Clinical Practice[35] research guidelines and oversight throughout was provided by an external Trial Steering Committee. The study has demonstrated the feasibility of clinical studies conducted largely in the primary care setting within the topic area of urology. The sample size of over 600 men recruited from a range of different geographical, socioeconomic and ethnic backgrounds makes this one of the largest studies in this field, with strong internal and external validity. The results provide a significant contribution to the existing literature. This serves to increase the transferability of the study findings. The study methodology included the use of 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 main limitation of the study is that it does not detail the potential impact of pathway change for men presenting with LUTS. The potential ability of primary care to accurately diagnose and initially manage these patients is demonstrated by both our quantitative and qualitative results. This is the first reported study to examine patient preferences for LUTS to be managed in primary care.[19] The health economic impact of such a pathway change will need to be explored before policy change or practice recommendations can be made in this topic area.

Critical appraisal[36] with Joanna Briggs Institute checklist for observational studies is shown below (*Figure 23*),[36] indicating which criteria were met or not. A convenience sample of those recruited in primary care was derived, limiting generalisability, although the size of the sample (gained during a time of extreme clinical research pressures during the

	Yes	No	Unclear	Not applicable
1. Was a consecutive or random sample of patients enrolled?	☐	☒	☐	☐
2. Was a case control design avoided?	☑	☐	☐	☐
3. Did the study avoid inappropriate exclusions?	☑	☐	☐	☐
4. Were the index test results interpreted without knowledge of the results of the reference standard?	☑	☐	☐	☐
5. If a threshold was used, was it pre-specified?	☐	☐	☐	☒
6. Is the reference standard likely to correctly classify the target condition?	☑	☐	☐	☐
7. Were the reference standard results interpreted without knowledge of the results of the index test?	☑	☐	☐	☐
8. Was there an appropriate interval between index test and reference standard?	☑	☐	☐	☐
9. Did all patients receive the same reference standard?	☑	☐	☐	☐
10. Were all patients included in the analysis?	☐	☒	☐	☐

FIGURE 23 Joanna Briggs checklist.

pandemic) may support transferability in UK primary care. Clinical diagnoses were made from the invasive urodynamics investigation. The extent of data collection, including some data to be submitted from the patient's home (bladder diary, uroflowmetry), made missingness of some data inevitable, and some attrition from the recruited sample.

Context of other literature

The invasive nature of urodynamic testing with its requirement for catheterisation of the bladder and rectum has led research to explore the diagnostic accuracy of simple tests for men presenting with LUTS. Most of the existing literature in this topic area examines the ability of simple measurements to diagnose BOO. Very little work exists examining the ability of simple tests to diagnose other urological conditions, such as DU or DO.

The role of symptom assessment was explored within the United Kingdom National Prostatectomy Audit and the authors concluded that making an accurate diagnosis was precluded if 'one relies on symptom scores and clinical examination alone'.[37] Another large study conducted by the International Continence Society suggested that symptoms alone can be misleading and are not diagnostic of BOO.[38] The existing literature acknowledges the association of LUTS and the urodynamic diagnosis of BOO but highlights that they are not necessarily causally related. The limited clinical utility of symptom assessment necessitates further investigation for most patients presenting with LUTS.

Measurements made from less invasive tests of urinary tract function such as urine flow studies have also been investigated. Maximum urine flow rate is a widely used measurement obtained from a urine flow study and has been extensively investigated as a possible objective measure of LUTS or diagnostic indicator of BOO. Another large study from the International Continence Society found a lack of correlation between any of the urinary symptoms and maximum urine flow rate.[39] Results from this trial showed that using a maximum flow rate threshold of 10 ml/second or below was associated with a high level of specificity in terms of making a diagnosis of BOO; however, the sensitivity associated with this threshold was poor.

Similarly, researchers have previously investigated the diagnostic utility of using post-void residual volume (which is the volume remaining in the bladder following completion of micturition[40]). The authors concluded that the presence of obstruction only weakly predicted a significant post-void residual volume. Furthermore, the association of LUTS and post-void residual volume has also been studied in a trial involving around 500 community dwelling men.[41] Using regression analysis, this study found no association between urinary symptoms and post-void residual volume.

Finally, the diagnostic ability of a blood test measuring serum PSA has been investigated in relation to providing an accurate diagnosis in men presenting with LUTS. One study examined over 300 men with LUTS and reported a significant correlation between PSA level and the urodynamic diagnosis of BOO.[42] The authors suggest that PSA level may be related to prostate size and hence this explains its correlation with BOO. Despite these encouraging results, the current use of serum PSA measurement is largely confined to its ability to predict underlying prostate cancer.

The existing literature seems to confirm that when used alone simple tests do not carry sufficient levels of accuracy to be useful in the diagnosis of men with urinary symptoms. The results from the PriMUS study have shown that increased accuracy can be obtained by using a combination of these simple tests within our mathematical model. Other researchers have investigated the use of a combination of measures derived from simple tests to provide a urodynamic diagnosis of BOO.[43] In a study involving over 150 men with urinary symptoms, authors concluded that the use of symptom scores alone did not provide sufficient diagnostic accuracy but suggested that a combination of factors including prostate volume, maximum flow rate and post-void residual volume was a better predictor of underlying BOO. These findings align with the results from the PriMUS study as we too have identified maximum flow rate and voided volume among our predictors of BOO.

Another important finding from our work was the demonstration of a significant prevalence of urodynamic abnormalities in men presenting to primary care with LUTS. We report that over 65% of our cohort demonstrated the presence of DO, over 45% demonstrated BOO and over 35% demonstrated DU. These conditions were not mutually

exclusive. Our findings are supported by data published from another study which estimated a high prevalence of underlying urodynamic abnormalities in an unselected population, the majority of whom did not have LUTS.[44] In this population-based study, the estimated prevalence of BOO was just over 20% and the estimated prevalence of overactive bladder was just over 10%. These figures are lower than those reported in the study which is likely to be due to the different populations sampled. Our reported prevalences are not from the general population but from a cohort of men with pre-existing urinary symptoms.

In summary, the PriMUS study has underlined the importance of combining the results of simple index tests when diagnosing the cause of LUTS, as this can mitigate the inherent inaccuracy of using the tests in isolation. We have demonstrated the high prevalence of urodynamic abnormalities within a cohort of men presenting with urinary symptoms and hence the clinical need for better and more widespread access to diagnostics and initial treatment.

Methodological: bladder diaries

One of the candidate index tests for potential inclusion in the diagnostic models was a hand-written bladder diary. This involves the patient recording information including voided volumes, fluid intake and sensations of urgency, for 3 days. Definition of the logic to allow calculation of bladder diary summary statistics that would work for every diary was very complex. In addition, the calculations required each diary to be transcribed and quality checked. This is time consuming, is likely to give rise to transcription errors (we noted 1 error per 6 diaries transcribed, but 50 diaries were double coded to assess and ensure consistency) and is highly unlikely to be feasible in primary care. Specific descriptions of algorithms for analysis are hard; details such as how to handle outlier cases are very difficult to get right. In contrast though, the home uroflowmetry device used in this study provides information on voided episodes and volumes, without the transcription requirements.

Implications for policy and practice

Qualitative studies

Our findings[19] emphasise the importance of LUTS being managed in primary care where possible, as in addition to opportunity cost savings and reduced waiting times, this is a more accessible option for patients, who tend to be more comfortable and confident being treated by familiar clinicians or in more familiar environments. The extent to which these views are generalisable to other GPs, patients and settings requires further research. There was evidence in our qualitative evaluation that urodynamics is also to an extent feasible in the primary care setting.[17] However, this was in the research setting, and informs future *research* rather than practice. Urodynamics remains an invasive investigation, requiring substantial investment in training, equipment and quality assurance, making it largely unpractical for primary care. The aim of PriMUS was to develop diagnostic models that do not depend on invasive diagnostic investigation, and which are more practical and preferred by patients for the primary care setting.

To address the identified challenges of managing LUTS in primary care, prostate cancer risk management or LUTS diagnostic tools would be helpful. It appears that bothersome LUTS are in some cases dismissed as a normal part of the ageing process.[45] Ensuring that such symptoms are managed well is essential to enhance patients' quality of life.

Potential use of models

The BOO model showed good discrimination and calibration statistics; it also had adequate sensitivity and specificity to levels that are likely to be clinically useful.

The DO and DU models involve more trade-off of sensitivity and specificity. To be useful in managing individual patients through guideline recommended pathways, GPs are likely to require *specificity* for a given diagnosis, to be confident in offering treatments/options for a given patient with whom they are consulting. Poorer sensitivity – failing to identify all patients with the condition – is less relevant *at that time* when managing the individual patient presenting with symptoms.

The models need to be integrated into a clinical decision support tool. Ideally, this should integrate with and extract relevant data from the clinical record (e.g. age, PSA level), but can, at least temporarily, be envisaged as stand-alone and requiring input of all required variables. The decision support tool needs to present diagnostic probabilities in ways that are accessible for both clinicians and patients and a basis for (shared) decision-making, including easy access to management advice and guidelines. It needs to be easily accessible within the consultation, and to offer printable summaries for patients to consider, review and revisit with clinicians.

Prototype Primary care Management of lower Urinary tract Symptoms clinical decision support tool

The prototype PriMUS clinical decision support tool can be viewed at: https://drinnan.shinyapps.io/primus (accessed July 2024). This is not yet intended to be used in practice.

Feedback from GPs during development (see *Chapter 2*, *Prototype online decision support tool evaluation*) and small-scale user-testing with the tool in simulated consultation scenarios (see *Chapter 2*, *Prototype online decision support tool evaluation*) was favourable. It is more likely that the tool will be applicable in primary care by *internally referring patients to a GP with Special Interest (in LUTS)* who can undertake the discussion, examination, gather the required data (e.g. voiding symptoms subscore, flow rates and residual volumes; PSA testing) in consultations that are longer than the usual 10 minutes' duration, and review treatments chosen and benefits or side effects ensuing. This 'clinic model' might operate at the single practice, cluster of practices or the community diagnostic clinic level. We note that such posts are not particularly common currently (though members of the Primary Care Urology Society were management Group advisors in the PriMUS study). With both the Society's interest and profile and recent rapid developments in primary care, such as the community diagnostic clinic models, the capacity to grow in this area is quite significant.

As above, patients have indicated strong preferences for management in primary care, and the PriMUS tool is intended to support that primary care management. Key outcomes will include GP and patient confidence in establishing the diagnosis(es) and choosing appropriate management options, revisiting and revising these at intervals where required. Referrals will still be preferred (by patients or GPs) on occasions, but the aim is to reduce referrals overall. Training for GPs will need to orientate and familiarise with the tool, including its strengths and weaknesses, the implications of missing data, and the extent to which apparent diagnoses alter (or sometimes very little) depending on data entry.

The demonstrated level of accuracy is comparable with other such models used regularly by GPs. The reported diagnostic accuracy of the four PriMUS models (BOO, DO, DO-SA and DU) have been presented in relation to other commonly used scores, such as QRisk,[46] HAS-BLED,[47] CHADS-VASC2[48] and FeverPAIN,[49] albeit that other tools have different time spans for predicting risk (see *Chapter 3*, *Model performance* and *Chapter 4*, *Model performance*). Overall, the accuracy figures are comparable to these other scales/scores that are used routinely and integrated into clinical (record) systems in primary care, although only the BOO model reached the required thresholds of both 75% sensitivity and specificity which we deemed at the outset to be clinically useful performance. This provides some confidence that with implementation in routine settings, there will be opportunity for the PriMUS diagnostic models to be made available via the standard primary care clinical systems (EMIS, Vision, SystemOne) as a decision support tool.

Further research

Economic modelling

Given the reasonable diagnostic accuracy data for the PriMUS models, it is now possible to undertake economic modelling research, including the potential resource use effects of adopting the PriMUS tool (including one-off investment, set-up and training costs, and ongoing costs such as additional time for organising/analysing test results, purchasing Flowtakers), its diagnoses and implications of the management recommendations.

The cost data required for evaluating the implementation of the management recommendations should encompass detailed costing of the relevant lower urinary tract conditions and associated pathways. Key outcomes that determine overall resource use are medical management (treatments and primary care consultation time) and referral rates (with costs including secondary care consultation time, investigations, etc.) Training requirements and set-up of 'clinics' at practice or cluster level are also relevant. Up-to-date data on referral rates will be required (identified to have risen to

30% or men presenting with LUTS in primary care at the time of commissioning PriMUS study, 2017), to model whether these are likely to diminish by implementing the PriMUS tool.

As above, the BOO model had adequate sensitivity and specificity to levels that are likely to be clinically useful (75% for both was deemed to be the minimum clinically useful performance). In contrast, the DO and DU models may require a trade-off of sensitivity and specificity. The higher the value of the risk threshold, the lower the sensitivity and the higher the specificity. It would increase the chance of the number of false negatives (missed cases) but decrease the number of false positives (incorrect positive diagnoses). On the other hand, the lower the value of the risk threshold, the higher the sensitivity and the lower the specificity. This would increase the false-positive rate but decrease the false-negative rate. These (particularly false-positive rates) will have impacts on potential costs of implementation in LUTS pathways and require sensitivity analyses within that economic modelling (see *Further research*).

Feasibility trial and randomised controlled trial

If suggested to be potentially cost-effective, then it will be important to evaluate the use of the PriMUS tool in practice. In line with the Medical Research Council complex intervention development framework, this is likely to include a feasibility trial and a powered RCT, again in the context of the single/cluster practice model as outlined above. A process evaluation will provide valuable information for wider implementation. As with all models currently used in practice (see *Prototype Primary care Management of lower Urinary tract Symptoms clinical decision support tool* section), continued external validation of the models is important, and in populations with greater ethnic diversity.

As above, the primary outcome is likely to be that of referral rates, as these are the drivers of resource use (secondary care investigations, clinic time etc.)

Treatment decisions – including medications, review appointments and investigations – are also important outcomes across primary and secondary care, affecting overall resource use. A cost-effectiveness study of implementing the PriMUS tool in routine general practice is required, both for its models and the associated index tests which may not currently be in routine use in primary care versus the referral of participants to receive the invasive urodynamics procedure.

Nevertheless, it will be important also to capture important patient-based outcomes, potentially including patients' confidence in treatment decisions, adherence to treatment decisions (also affecting resource use measures), quality-of-life and patient safety.

One potential adverse event assessed during this PriMUS study concerned a patient who had been assessed in 2019, with benign disease, and minimal medical management ensued. After the pandemic 'lockdown(s)' period, the patient presented again and was found to have a substantially raised PSA of 38, thus highly likely to have prostate cancer. The Study Management team considered whether the patient may have derived reassurance from the PriMUS participation, potentially leading to delayed presentation later (though clearly the pandemic influences were substantial also). Patient safety from adopting PriMUS is an essential outcome domain for further evaluation.

As above also, a formal process evaluation of the adoption and adaptation of using PriMUS in practice will be important to inform wider implementation (training requirements, clinic set-up experiences, variations in patient pathways by locality etc.).

The influences of age, smoking, multi-morbidity and polypharmacy on lower urinary tract symptoms, presentations and treatment preferences

Further research could explore how patients' beliefs that symptoms are age-related impact on treatment expectations among patients, and the role of comorbidity and polypharmacy in influencing readiness to seek medical help for these symptoms. Considering the reported lack of effectiveness and intolerable side effects of some LUTS medications, greater exploration of non-pharmacological treatment would be beneficial.[19] A study of urology outpatients[50] has shown promising results for the effectiveness of sessions promoting the self-management of LUTS; however, the authors suggest a large RCT is needed to confirm findings. Patients in this PriMUS study indicated a strong preference for their symptoms to be managed in primary care. Given that preference for primary versus secondary care appears to

vary between different patient groups, it would be valuable to further explore the effect of patient characteristics on treatment preferences.

We observed that current smoking and polypharmacy are strongly associated with the severity of LUTS experienced by men presenting to primary care.[45] Proposed mechanisms for the effect of smoking on LUTS, such as atherosclerosis and increased sympathetic tone activity,[51] suggest smoking may have a causative effect, meaning some symptomatic benefit may be incurred from smoking cessation. The exact mechanism however remains unclear and previous associations between smoking and LUTS have been inconsistent, so further research to clarify this would be valuable.[52]

Further research into the contribution and mechanism of polypharmacy influences would also be valuable. In this study,[45] the medication categories included uroselective medications, examining for extent and thus potential for interactions. The association with major polypharmacy indicates this finding is not simply a function of being on uroselective medications themselves. Mechanisms relating to polypharmacy are multiple and may overlap with accompanying comorbidities, adding to the complexity of managing these symptoms. Further research could examine this.

Importantly, smoking and polypharmacy are modifiable factors which are already regularly managed with evidence-based strategies within primary care.[53,54] Consideration given to targeting these factors could support decision-making when treating male patients with LUTS in primary care, and the interaction of this with the use of a tool to support diagnosis and decision-making, as shown feasible by the PriMUS study, will be an important element of future research on the PriMUS tool.

Conclusions

This large primary care-based study has developed diagnostic models for the BOO, DO and DU conditions among men presenting with LUTS. The BOO model meets thresholds of sensitivity and specificity deemed clinically useful, but the other models entail more trade-off of sensitivity and specificity. A prototype clinical decision support tool has been developed and was felt to be feasible and acceptable to GPs and patients. With lower than required recruitment for the validation sample, further external validation and recalibration may be required before the models can be used in practice to ensure applicability of the models in settings where case-mix or prevalence could differ. Economic modelling and feasibility and randomised trial stages are required, also including a process evaluation to examine implementation issues. With the high and rising referral rate for men presenting with LUTS, there is the potential for such evaluations to demonstrate cost savings and improved patient experience, with limited impact on primary care resource use.

Equality, diversity and inclusion

Seeking for our findings to be generalisable and applicable to various demographic groups, we designed our research methodology in a pragmatic way to minimise bias and promote inclusivity in participant recruitment. Recruitment sites were purposively spread across the UK including sites in Southwest England, Northern England and Wales that allowed for broad geographical representation across socioeconomic backgrounds, ages, ethnicities to ensure diversity within our recruited population. Nevertheless, as described earlier, the sample of participants actually recruited was of very limited diversity, which limits generalisability, and warrants further work in more diverse populations to validate the PriMUS models.

We plan to develop dissemination materials and videos for promoting the findings and will provide translation in diverse languages. We will also ensure that patient-facing materials for future studies are as widely accessible as possible and encompass a full range of diversity considerations.

Additional information

CRediT contribution statement

Adrian Edwards (https://orcid.org/0000-0002-6228-4446): Conceptualisation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – editing and reviewing.

Ridhi Agarwal (https://orcid.org/0000-0002-6930-5617): Formal analysis, Methodology, Writing – original draft, Writing – editing and reviewing.

Janine Bates (https://orcid.org/0000-0003-3610-2415): Data curation, Project administration, Supervision, Writing – original draft, Writing – editing and reviewing.

Alison Bray (https://orcid.org/0000-0003-1402-804X): Conceptualisation, Data curation, Funding acquisition, Investigation, Methodology, Resources, Software, Writing – original draft, Writing – editing and reviewing, PPI contributions.

Sarah Milosevic (https://orcid.org/0000-0003-1973-8286): Formal analysis, Investigation, Writing – original draft, Writing – editing and reviewing.

Emma Thomas-Jones (https://orcid.org/0000-0001-7716-2786): Conceptualisation, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – editing and reviewing.

Michael Drinnan (https://orcid.org/0000-0002-2181-8202): Conceptualisation, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Supervision, Validation, Visualisation.

Marcus Drake (https://orcid.org/0000-0002-6230-2552): Conceptualisation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – editing and reviewing.

Peter Michell: Writing – editing and reviewing, PPI contributions.

Bethan Pell (https://orcid.org/0000-0002-0786-6339): Data curation, Methodology, Project administration.

Haroon Ahmed (https://orcid.org/0000-0002-0634-8548): Funding acquisition, Investigation, Methodology, Writing – editing and reviewing.

Natalie Joseph-Williams (https://orcid.org/0000-0002-8944-2969): Conceptualisation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – editing and reviewing.

Kerenza Hood (https://orcid.org/0000-0002-5268-8631): Conceptualisation, Funding Acquisition, Methodology, Supervision, Writing – editing and reviewing.

Yemisi Takwoingi (https://orcid.org/0000-0002-5828-9746): Formal analysis, Funding acquisition, Project administration, Supervision, Writing – editing and reviewing.

Chris Harding (https://orcid.org/0000-0002-9407-382X): Conceptualisation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – editing and reviewing.

Acknowledgements

The authors gratefully acknowledge numerous people who have participated in PriMUS since inception. This particularly includes grateful thanks to the late Professor Rob Pickard who drove the idea formulation and secured initial funding. We also thank others directly involved in study delivery:

Alex Hall, Alison Edwards, Charlotte Scoble, Danielle Dimuzio, Debra Barnett, Dianne Wake, Emily Perry, Gareth Kennard-Holden, Jane Davies, Jane Davies, Jill Deane, Jill Ducker, Jo Euden, Joanne Sullivan, Joanne Thompson, Jon Rees, Jonathan Deeks, Joy Allen, Laura Bevan, Lisa Mellish, Lucy Marsh, Lucy Shipp, Marta Cobos-Arrivabene, Nicky Todd, Peter Murphy, Rebecca Cousins, Rob Forder, Samantha Clarke, Sarah Nash, Stuart Daniel, Tom Schatzberger, Tracy Scott, Victoria Garner, Wendy Robson.

We also thank key public contributors to study management and direction: Ray White (PPI on SMG), Malcolm Gristwood (PPI on SMG), Frank Bray (PPI for one-off event), Alan Bowen (PPI on SMG), Colin Thomson (PPI on SSC), Alan David Pryce (PPI on SSC), Alan Hart, (PPI on SSC). And other SSC members: Professors Tom Fahey (Chair), Rafael Perea, Gail Hayward, Ian Pearce.

Finally, we thank the 55 GP practices that participated, their staff, and the 725 men who agreed to participate in the PriMUS study.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Available data can be obtained from the corresponding author. Access to anonymised data may be granted following review.

Ethics statement

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.

Information governance statement

Cardiff University is committed to handling all personal information in line with the UK Data Protection Act (2018) and the General Data Protection Regulation (EU GDPR) 2016/679. Under the Data Protection legislation, Cardiff University is the Data Controller, and you can find out more about how we handle personal data, including how to exercise your individual rights and the contact details for our Data Protection Officer here (www.cardiff.ac.uk/public-information/policies-and-procedures/data-protection/research-participants-data-protection-notice).

Disclosure of interests

***Full disclosure of interests*:** Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/RGTW5711.

***Primary conflicts of interest*:** Alison Bray: Royalties or licenses in respect of the Flowtaker device, paid to The Newcastle upon Tyne Hospitals NHS Foundation Trust (NuTH) (not to the individual). Payment or honoraria to NuTH for delivery of a training course.

Michael Drinnan: Royalties or licenses in respect of the Flowtaker device, paid to The Newcastle upon Tyne Hospitals NHS Foundation Trust (NuTH) (not to the individual).

Marcus Drake: Grants/contracts from: NIHR (HTA, i4i, RfPB programmes), MRC, EPSRC, Rosetrees Trust, payment/honoraria and payment for expert testimony from Astellas, Trustee of the International Continence Society.

Peter Michell: PPI Honoraria payments from grant on current project.

Kerenza Hood: NIHR grant for current project. Leadership or fiduciary role in other board etc: NIHR HTA General Board, NIHR Research Professors panel, was a member of NIHR CTU Standing Advisory Committee, HTA General Committee and HTA Funding Committee Policy Group.

Chris Harding: NIHR – 2× HTA grants for unrelated work, 1× grant for this work, The Urology foundation – grant for unrelated work. Consulting fees: Viatris, Teleflex medical, Medtronic (all unrelated to this work). Payment or honoraria: Allergan, Viatris, Convatec (all unrelated to this work). Participation on a Data Safety Monitoring Board or Advisory Board: NIHR SIMS TRIAL, NIHR CADET TRIAL (all unrelated to this work). Leadership or fiduciary role in other board etc: EAU female LUTS guideline, ICS Urodynamics Committee, ICS Mesh Committee.

Nothing to declare: Adrian Edwards, Ridhi Agarwal, Janine Bates, Sarah Milosevic, Emma Thomas-Jones, Bethan Pell, Haroon Ahmed, Natalie Joseph-Williams.

Publications

Pell B, Thomas-Jones E, Bray A, Agarwal R, Ahmed H, Allen AJ, *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;**10**:e037634. https://dx-doi-org.abc.cardiff.ac.uk/10.1136/bmjopen-2020-037634

Milosevic S, Joseph-Williams N, Pell B, Cain E, Hackett R, Murdoch F, *et al.* Conducting invasive urodynamics in primary care: qualitative interview study examining experiences of patients and healthcare professionals. *Diagn Progn Res* 2021;**5**:10. https://dx.doi.org/10.1186/s41512-021-00100-y

Milosevic S, Joseph-Williams N, Pell B, Cain E, Hackett R, Murdoch F, *et al.* Managing lower urinary tract symptoms in primary care: qualitative study of GPs' and patients' experiences. *Br J Gen Pract* 2021;**71**:E685–92. https://dx-doi-org.abc.cardiff.ac.uk/10.3399/BJGP.2020.1043

Purchase T, Trilloe G, Ahmed H, Agarwal R, Bray A, Hood K, *et al.* Polypharmacy and smoking as potentially modifiable risk factors associated with symptom severity in men with lower urinary tract symptoms in primary care: findings from the PriMUS study. *Eur Urol Focus* 2022;**8**:8–10. https://dx-doi-org.abc.cardiff.ac.uk/10.1016/j.euf.2022.01.001

Conferences and presentations

30 November 2017	Cwm Taf UHB R&D Conference	Poster
13 March 2018–14 March 2018	South-West SAPC 2018	Oral
14 June 2018	Cardiff and Vale R&D Conference	Poster
30 November 2017	Bristol Research Service Innovation Conference Poster	
30 November 2018	Cwm Taf UHB R&D Conference	Poster
13 March 2019–14 March 2019	South-West SAPC 2019	Poster
13 March 2019–14 March 2019	South-West SAPC 2019	Poster
11 June 2019	Aneurin Bevan R&D Conference	Poster
October 2020	HCRW Conference	Oral
30 June 2021–1 July 2021	SAPC	Poster

References

1. Pell B, Thomas-Jones E, Bray A, Agarwal R, Ahmed H, Allen AJ, *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;**10**:e037634. https://dx-doi-org.abc.cardiff.ac.uk/10.1136/bmjopen-2020-037634
2. National Institute for Health and Care Excellence. *Lower Urinary Tract Symptoms in Men: Management*. 2010. URL: www.nice.org.uk/guidance/cg97 (accessed 14 November 2022).
3. 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**:1549–57.
4. Foon R, Toozs-Hobson P, Latthe P. Prophylactic antibiotics to reduce the risk of urinary tract infections after urodynamic studies. *Cochrane Database Syst Rev* 2012;**10**:CD008224.
5. Drake MJ, Doumouchtsis SK, Hashim H, Gammie A. Fundamentals of urodynamic practice, based on International Continence Society good urodynamic practices recommendations. *Neurourol Urodyn* 2018;**37**:S50–60.
6. Kuo HC. Videourodynamic analysis of pathophysiology of men with both storage and voiding lower urinary tract symptoms. *Urology* 2007;**70**:272–6.
7. van Doorn B, Blanker MH, Kok ET, Westers P, Bosch JL. Prevalence, incidence, and resolution of nocturnal polyuria in a longitudinal community-based study in older men: the Krimpen study. *Eur Urol* 2013;**63**:542–7.
8. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and cox regression. *Am J Epidemiol* 2007;**165**:710–8.
9. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;**18**:681–94.
10. Rubin D. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons; 1987.
11. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *J R Stat Soc Ser C Appl Stat* 1999;**48**:313–29.
12. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Diabet Med* 2015;**32**:146–54.
13. Steyerberg E. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer; 2009.
14. National Institute for Health and Care Research. *NIHR Public Information Pack*. URL: www.nihr.ac.uk/documents/public-information-pack-pip-how-to-get-involved-in-nhs-public-health-and-social-care-research/27388 (accessed 21 August 2023).
15. Global V. *Voice Global PPI Platform*. URL: https://voice-global.org (accessed 21 August 2023).
16. National Institute for Health and Care Research. *NIHR Guidelines on Reward and Recognition for Public Contributors: A Guide to the Payment of Fees and Expenses*. URL: Payment guidance for researchers and professionals | NIHR (accessed 21 August 2023).
17. Milosevic S, Joseph-Williams N, Pell B, Cain E, Hackett R, Murdoch F, *et al*. Conducting invasive urodynamics in primary care: qualitative interview study examining experiences of patients and healthcare professionals. *Diagn Progn Res* 2021;**5**:10.
18. Drinnan M. *PriMUS Prototype Online Decision Support Tool*. URL: https://drinnan.shinyapps.io/primus/ (accessed 21 August 2023).

19. Milosevic S, Joseph-Williams N, Pell B, Cain E, Hackett R, Murdoch F, *et al*. Managing lower urinary tract symptoms in primary care: qualitative study of GPs' and patients' experiences. *Br J Gen Pract* 2021;**71**:e685–92.
20. Coyne KS, Sexton CC, Thompson CL, Milsom I, Irwin D, Kopp ZS, *et al*. The prevalence of lower urinary tract symptoms (LUTS) in the USA, the UK and Sweden: results from the Epidemiology of LUTS (EpiLUTS) study. *BJU Int* 2009;**104**:352–60.
21. Irwin DE, Milsom I, Kopp Z, Abrams P, Cardozo L. Impact of overactive bladder symptoms on employment, social interactions and emotional well-being in six European countries. *BJU Int* 2006;**97**:96–100.
22. Robertson C, Link CL, Onel E, Mazzetta C, Keech M, Hobbs R, *et al*. The impact of lower urinary tract symptoms and comorbidities on quality of life: the BACH and UREPIK studies. *BJU Int* 2007;**99**:347–54.
23. Isa NMM, Aziz AFA. Lower urinary tract symptoms: prevalence and factors associated with help-seeking in male primary care attendees. *Korean J Fam Med* 2020;**41**:256–62.
24. Solvang M, Elnegaard S, Jarbol DE. Urological symptoms among 23,240 men in the general Danish population – concerns about symptoms, their persistence and influence on primary care contacts. *Scand J Prim Health Care* 2018;**36**:227–36.
25. Rosen R, Altwein J, Boyle P, Kirby RS, Lukacs B, Meuleman E, *et al*. Lower urinary tract symptoms and male sexual dysfunction: the multinational survey of the aging male (MSAM-7). *Eur Urol* 2003;**44**:637–49.
26. Soler R, Averbeck MA, Koyama MAH, Gomes CM. Impact of LUTS on treatment-related behaviors and quality of life: a population-based study in Brazil. *Neurourol Urodyn* 2019;**38**:1579–87.
27. Chapple CR, Roehrborn CG. A shifted paradigm for the further understanding, evaluation, and treatment of lower urinary tract symptoms in men: focus on the bladder. *Eur Urol* 2006;**49**:651–8.
28. Nitti VW. Pressure flow urodynamic studies: the gold standard for diagnosing bladder outlet obstruction. *Rev Urol* 2005;**7**:S14–21.
29. Elkoushy MA, Elshal AM, Elhilali MM. Changing patients' profile presenting for surgical management of benign prostatic hyperplasia over the past 16 years: a single-centre perspective. *Can Urol Assoc J* 2015;**9**:372–8.
30. Young MJ, Elmussareh M, Morrison T, Wilson JR. The changing practice of transurethral resection of the prostate. *Ann R Coll Surg Engl* 2018;**100**:326–9.
31. Shaw C, Williams K, Assassa PR, Jackson C. Patient satisfaction with urodynamics: a qualitative study. *J Adv Nurs* 2000;**32**:1356–63.
32. 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.
33. Welch LC, Botelho EM, Joseph JJ, Tennstedt SL. A qualitative inquiry of patient-reported outcomes: the case of lower urinary tract symptoms. *Nurs Res* 2012;**61**:283–90.
34. Fourcade RO, Lacoin F, Roupret M, Slama A, Le Fur C, Michel E, *et al*. Outcomes and general health-related quality of life among patients medically treated in general daily practice for lower urinary tract symptoms due to benign prostatic hyperplasia. *World J Urol* 2012;**30**:419–26.
35. ICH E6 (R2). *Good Clinical Practice – Scientific Guideline*. The European Medicines Agency. URL: europa.eu
36. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al*.; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36.
37. Neal DE. The national prostatectomy audit. *Br J Urol* 1997;**79**:69–75.
38. Abrams P, Donovan JL, De La Rosette JJ, Schäfer W. International Continence Society 'Benign Prostatic Hyperplasia' Study: background, aims, and methodology. *Neurourol Urodyn* 1997;**16**:79–91.

39. Reynard JM, Yang Q, Donovan JL, Peters TJ, Schafer W, De La Rosette JJ, *et al*. The ICS-'BPH' Study: uroflowmetry, lower urinary tract symptoms and bladder outlet obstruction. *Br J Urol* 1998;**82**:619–23.

40. Kranse R, van Mastrigt R. Weak correlation between bladder outlet obstruction and probability to void to completion. *Urology* 2003;**62**:667–71.

41. Kolman C, Girman CJ, Jacobsen SJ, Lieber MM. Distribution of post-void residual urine volume in randomly selected men. *J Urol* 1999;**161**:122–7.

42. 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 Int* 2004;**94**:1283–6.

43. 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. *Neurourol Urodyn* 2003;**22**:301–5.

44. 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 Int* 2011;**108**:1132–8.

45. Purchase T, Trilloe G, Ahmed H, Agarwal R, Bray A, Hood K, *et al*. Polypharmacy and smoking as potentially modifiable risk factors associated with symptom severity in men with lower urinary tract symptoms in primary care: findings from the PriMUS study. *Eur Urol Focus* 2022;**8**:8–10.

46. 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;**357**:j2099.

47. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. 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;**138**:1093–100.

48. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. 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;**137**:263–72.

49. Little P, Hobbs FD, Moore M, Mant D, Williamson I, McNulty C, *et al*. PRImary care Streptococcal Management (PRISM) study: in vitro study, diagnostic cohorts and a pragmatic adaptive randomised controlled trial with nested qualitative study and cost-effectiveness study. *Health Technol Assess* 2014;**18**:vii1–xxv, 1–101.

50. Brown CT, Yap T, Cromwell DA, Rixon L, Steed L, Mulligan K, *et al*. Self management for men with lower urinary tract symptoms: randomised controlled trial. *BMJ* 2007;**334**:25–8.

51. Ponholzer A, Temml C, Wehrberger C, Marszalek M, Madersbacher S. The association between vascular risk factors and lower urinary tract symptoms in both sexes. *Eur Urol* 2006;**50**:581–6.

52. Bradley CS, Erickson BA, Messersmith EE, Pelletier-Cameron A, Lai HH, Kreder KJ, *et al*.; Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN). Evidence of the impact of diet, fluid intake, caffeine, alcohol and tobacco on lower urinary tract symptoms: a systematic review. *J Urol* 2017;**198**:1010–20.

53. National Institute of Health and Care Excellence. *Multimorbidity and Polypharmacy [KTT18]*. 2019. URL: www.nice.org.uk/advice/ktt18/chapter/Key-points (accessed 1 November 2021).

54. Verbiest M, Brakema E, van der Kleij R, Sheals K, Allistone G, Williams S, *et al*. National guidelines for smoking cessation in primary care: a literature review and evidence analysis. *NPJ Prim Care Respir Med* 2017;**27**:2.

Appendix 1

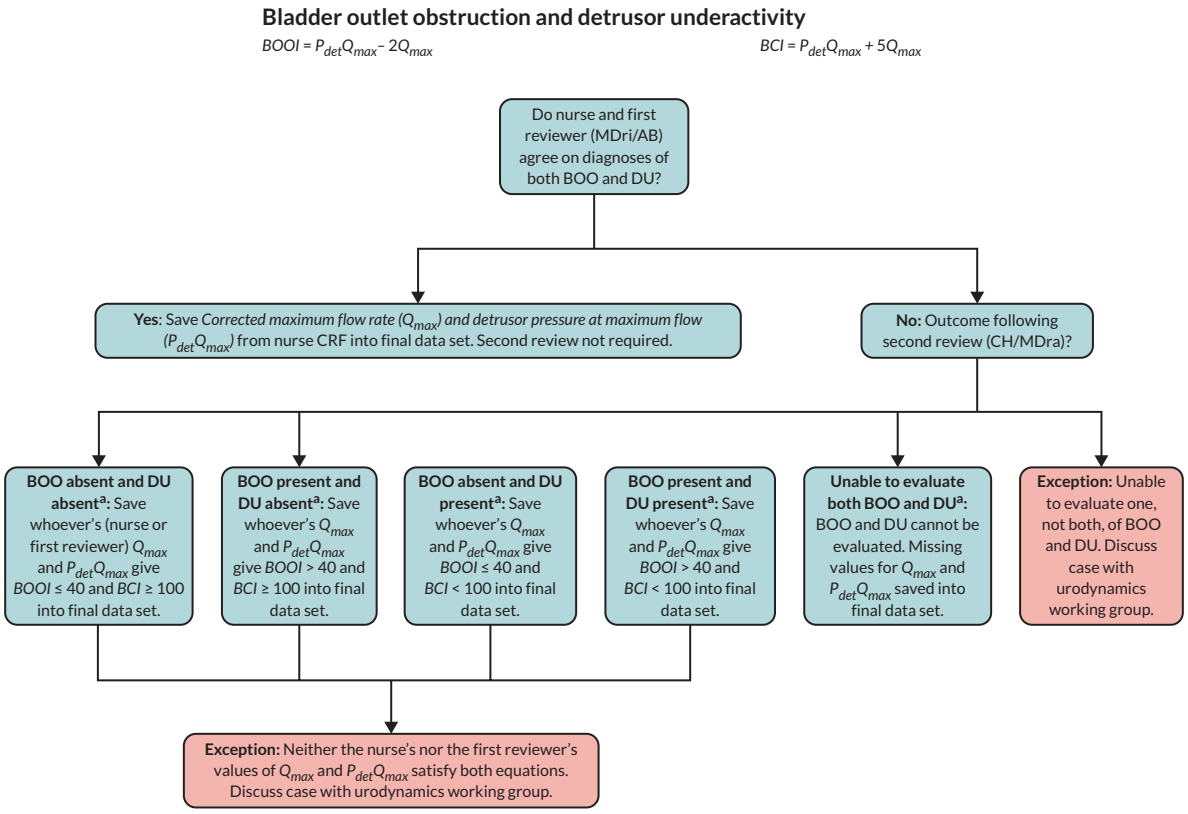


FIGURE 24 Workflow for the final diagnosis of the target conditions by the reference standard.a, These are the two diagnostic decisions resulting from the nurse, first reviewer and second reviewer, i.e. either the nurse and first review agree on one, and the other decision comes from the second reviewer, or the nurse and first reviewer disagree on both, and both decisions come from the second reviewer.

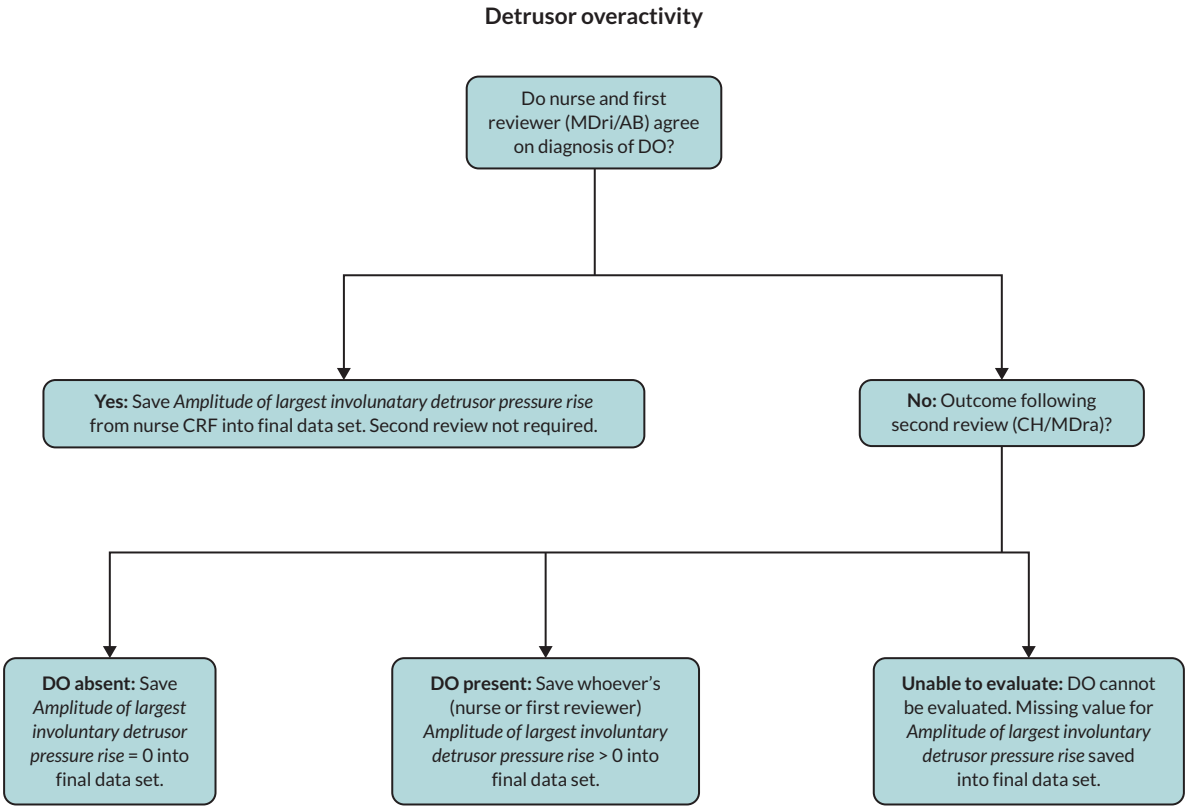


FIGURE 25 Detrusor Overactivity (DO).

Appendix 2 The Primary care Management of lower Urinary tract Symptoms study – internal pilot report

Study summary

The PriMUS Study is a diagnostic accuracy study which aims to develop a clinical decision support tool to help GPs more accurately diagnose and therefore treat and manage MLUTS. See *Figure 26* for Study schema and *Figure 27* for the participant flow diagram. Eligible participants will all receive the same tests; a series of simple (index) tests, following NICE Clinical Guidelines and a urodynamic (reference) test. We will compare the results of the simple (index) tests with the results of the urodynamic (reference) test, to identify which combination of simple (index) tests gives the best predictor of the urodynamic result. This combination of simple (index) tests will be carried forward into the development of a prototype clinical decision support tool, which will be developed and validated with two separate participant cohorts, so that the simple (index) tests can be used in primary care to obtain an accurate diagnosis. Follow-up is to be conducted at approximately 6 months, collating the treatment and management decisions made as a result of these procedures. Our recruitment target is 880 male participants from 3 research hubs in Newcastle, Bristol and Wales.

A qualitative evaluation has also been included in the internal pilot. The aims of this qualitative evaluation were to assess patients' acceptability of the reference urodynamic assessment, identifying barriers and facilitators to the uptake of, or satisfaction with, the test. Further aims were to understand patients' experiences of LUTS, their management of the symptoms, their decision to seek medical advice, and the factors that matter most to patients. We also wanted to understand healthcare professionals' experiences of recruiting patients to the study and the study processes (including feedback from the nurses conducting the urodynamic testing) and identify other barriers/facilitators to participation in the study (from perspective of both healthcare professionals and patients). We ultimately wanted to use the results to inform strategies that will maximise recruitment and retention.

This report describes the findings of the internal pilot study, conducted between February 2018 and November 2018 (recruitment months 1–10).

Primary care Management of lower Urinary tract Symptoms internal pilot study

The aim of the internal pilot was to establish that the protocol for the main trial could be implemented by assessing:

1. Site and patient recruitment rate,
2. The proportion of patients undergoing urodynamic assessments
3. Those with a complete data set for analysis.
4. The acceptability of the gold standard urodynamic assessment with patients, and potential refinements in processes via a qualitative evaluation.

We wanted to use the knowledge gained from experience within the first 6–9 months of recruitment to allow the study team to further develop and refine study processes, should such refinements and developments prove necessary, to continue into the main recruitment phase.

The results were discussed with our SSC, before reporting to the NIHR HTA Programme at Month 19 (originally 17, but changed to 19 with the delay to recruitment opening date), for permission to proceed.

In accordance with the HTA guidance on internal pilot studies, we excluded the first 2 months of recruitment from our calculation of the recruitment rate as we anticipated a 'lag phase' during which practices were still being registered and participating clinicians were developing the confidence and competence in recruiting patients.

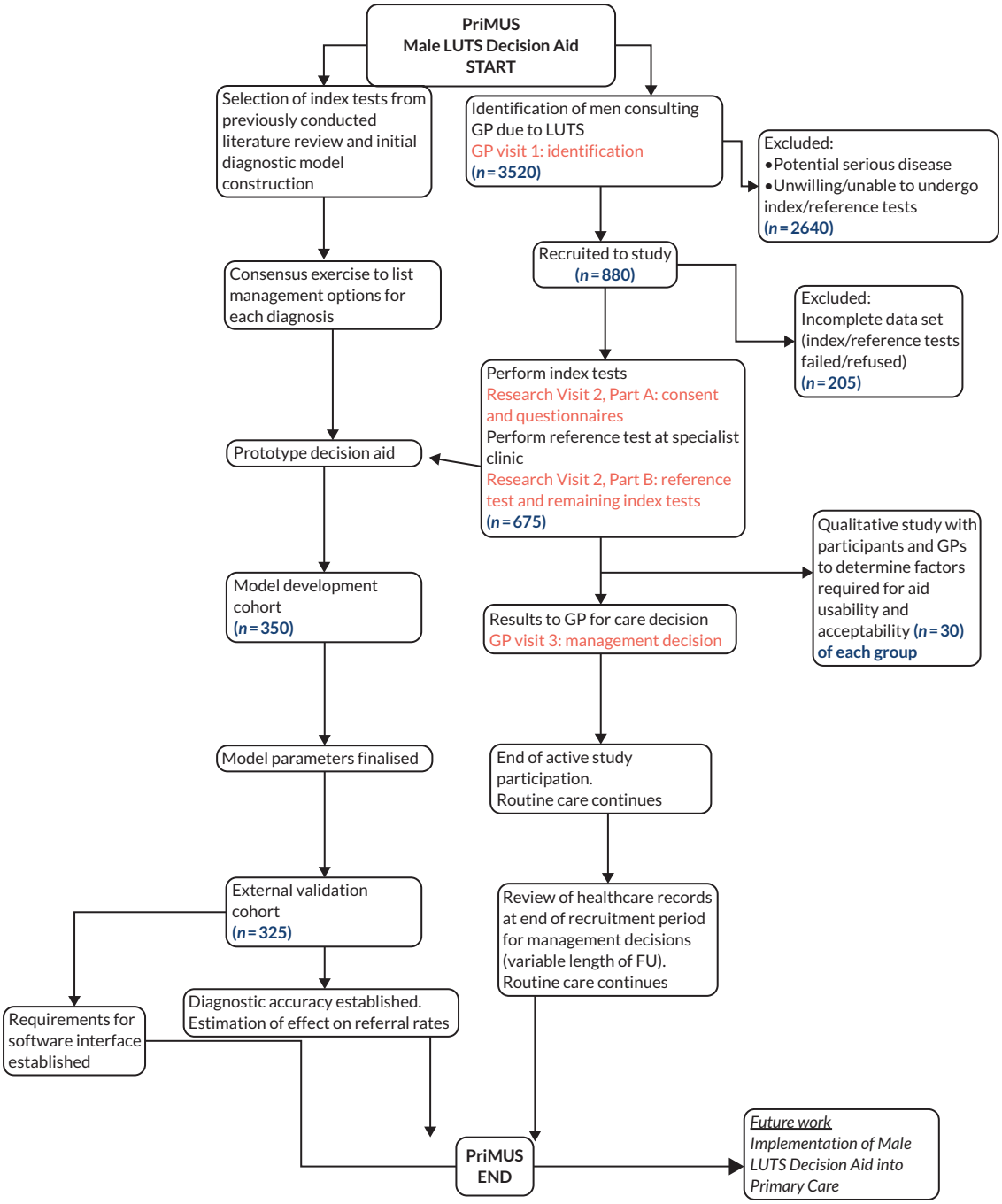


FIGURE 26 Study schema.

Internal pilot progression criteria

The internal pilot progression criteria were pre-specified in the grant application that needed to be met, in the internal pilot study in order for PriMUS to progress to the main study. These are outlined below (*Table 29*).

It was accepted that adjustments to recruitment and study processes (e.g. numbers of GP Practices recruited) could be made based on the internal pilot experience prior to progression to the main recruitment phase.

TABLE 28 Table of assessments

Test	Details	Result	Visit
Relevant demographics	Following assessment for eligibility and agreement to participate	Age in years	Study Visit, Part A
Physical examination of abdomen	Carried out by competent primary care health professional. To palpate for a distended bladder	Bladder palpable/not palpable	First GP consultation
DRE	Carried out by competent primary care health professional. To determine prostate enlargement and likelihood of locally advanced prostate cancer.	Prostate mild/moderate/severe enlargement Further assessment for prostate cancer required/not required	First GP consultation
PSA	Single blood test. To compare value against NHS-defined thresholds	PSA value – established thresholds for further assessment for prostate cancer (typically > 3 ng/mL) or benign enlargement (typically ≥ 1.5 ng/mL) For decision aid: continuous variable in ng/mL	First GP consultation
ICIQ-MLUTS symptom questionnaire	Patient-completed validated questionnaire which defines presence, type and severity of LUTS	Total score (0–52); voiding symptom score (0–20), storage symptom score (0–24), bother scores (0–10).	Study Visit, Part A
IPSS questionnaire	Patient-completed validated questionnaire which defines presence, type and severity of LUTS	Total score (0–35)	Study Visit, Part A
Bladder Diary	A patient-completed/automated diary (at least 3 days) of the volumes and timing of urine passed	Waking (day) time frequency, sleeping (night) time frequency, 24 hour voided volume, daytime voided volume, nocturnal voided volume, average volume voided each void, total urgency scores	Given in Patient Information Pack and Patient completes at home/ provided in Study Visit Part A
Uroflowmetry (Home Flowtaker)	A measurement of urine flow rate and voided volume as a function of time, either in the clinic or at home. Patients in the PriMUS study will perform uroflowmetry at home using the Flowtaker device	Maximum flow rate, voided volume against normal age-adjusted range. Single value in ml/second	Given in Study Visit, Part B – Patient completes at home
Post-void residual	Simple abdominal ultrasound scan to determine the volume of urine remaining in the bladder after urination	Residual volume against normal age-adjusted range. Single value in mL	At Study Visit, Part B

Internal pilot phase findings

General practitioner practice recruitment

Fifty-seven GP practices were successfully recruited to the pilot phase in Newcastle, Wales and Bristol Research Hubs (*Figure 28*). The first two GP practices opened to recruitment on 20 February 2018 in Newcastle. The first GP practice to open in Wales opened on 27 March 2018 and the first GP practice to open to recruitment in Bristol opened on 12 July 2018.

Overall, 97 GP practices have expressed an interest to participate in the PriMUS study. Two GP practices have now declined due to issues with staffing, so inability to commit to research. As described in *Modifications to design and/ or study processes*, 57 GP practices were recruited as part of the internal pilot and will continue to be involved in the main study. Forty additional GP practices have expressed an interest in taking part in the main study (9 in Wales, 8 in Newcastle, 23 in Bristol). Bristol and Newcastle have adopted a 'Hub and Spoke' model, whereby all study activity (apart from screening and identification) takes place in the Hub GP Practice. All Hub Practices sign a GP agreement

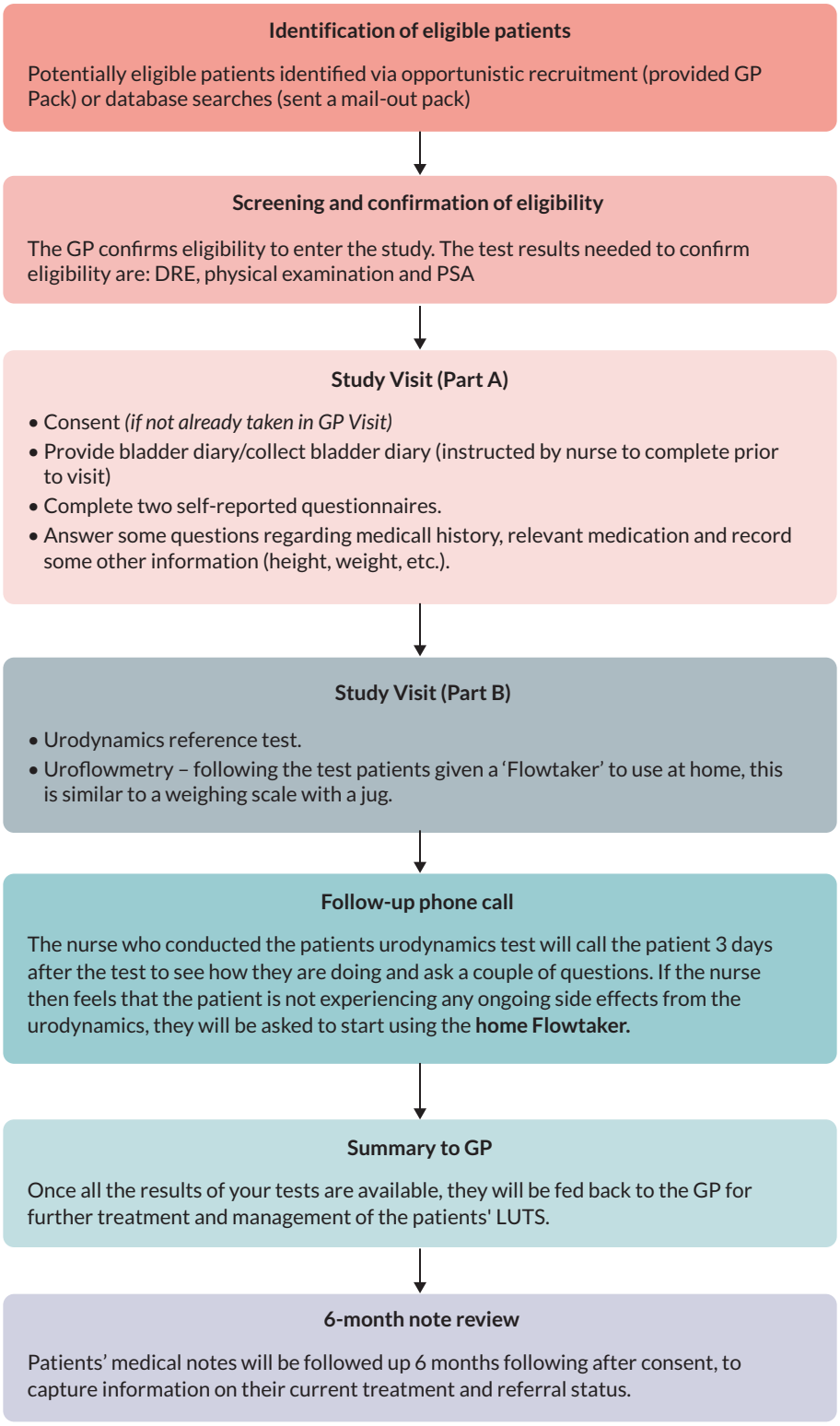


FIGURE 27 Participant flow diagram.

between the practice and CTR. The Spoke practices that already form part of a Clinical Commissioning Group do not need to sign an agreement between the Hub and Spoke, as they already have an existing data-sharing agreement in place. Those Spoke Practices that site outside of a Clinical Commissioning Group now have a GP agreement template, which needs to be signed to ensure an agreement between the Hub and Spoke Practices. This is all part of the site opening procedure.

TABLE 29 Progression criteria

Criteria	Level	Action	Result
Site recruitment	30 sites open	Monitor site recruitment and time to first recruit. At least 0.3 patients per practice per month. Discuss potential mitigating strategies	July: 37 August: 42 September: 48 October: 51 November: 57
Recruitment rate[a]	70% 40–70% < 40%	GO Discuss potential mitigating strategies STOP?	July: (45/54)*100 = 83% Aug: (62/79)*100 = 78% Sep: (87/113)*100 = 76% Oct: (129/155)*100 = 83% Nov: (169/202)*100 = 83%
Gold standard undertaking (and evaluable?)	70% 50–70% < 50%	GO Discuss potential mitigating strategies STOP?	105/116 evaluable: 90% 11/116 not evaluable: 10%
Acceptability of gold standard urodynamic assessment	Qualitative interviews	Detailed in *Table 28* and *Chapter 6*	Detailed in *Table 32* and *Chapter 6*

a Allowing for 2-month lag phase.

We have also received interest from research hubs not currently involved in the PriMUS Study. A Urology Consultant in Birmingham has expressed interest in becoming involved and we are currently in discussions with him to establish feasibility of running PriMUS in this locality.

Patient recruitment

See *Figure 29* for CONSORT table. Six hundred and thirteen patients have been screened for eligibility, of whom 242 have been eligible (39.5%) and 371 have been ineligible (60.5%). Seventy-three patients screened were eligible but not recruited (30.2%).

Reasons for patient not being eligible:

- Men with LUTS considered secondary to current or past invasive treatment or radiotherapy for pelvic disease
- Men with indwelling urinary catheters or who carry out ISC
- Men who have not presented with a complaint of one or more bothersome LUTS
- Not able and willing to undergo all index tests and reference test, and complete study documentation

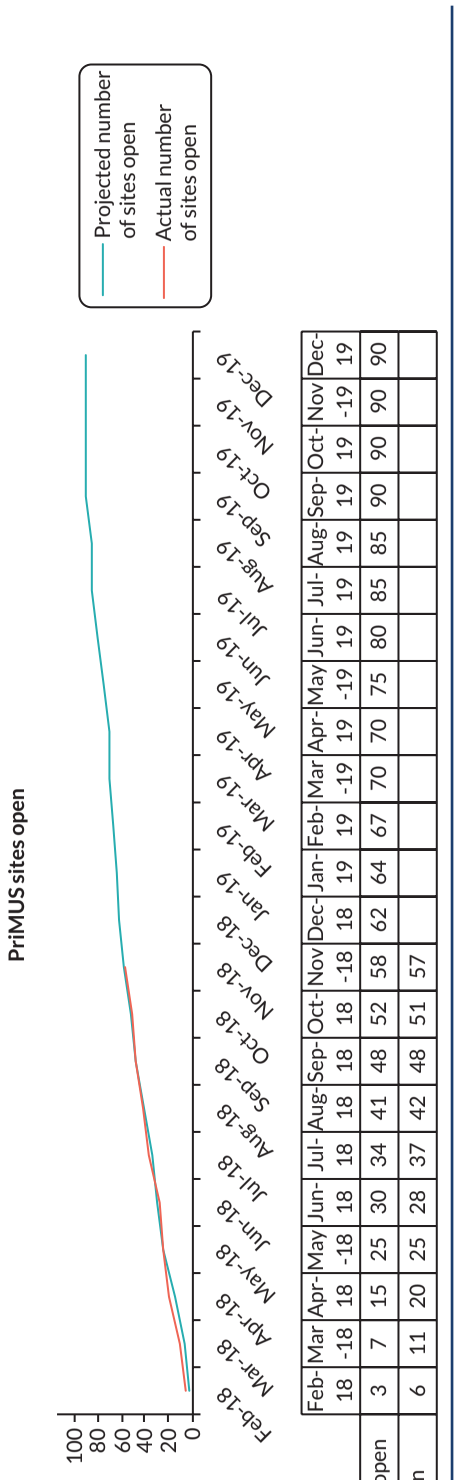
Reasons for patients being eligible but not recruited:

- Not willing to give informed consent for participation in study
- Patient changed mind Nursing home resident – not independently mobile
- Did not want to do urodynamics
- Declined investigation
- Under clinical review first
- Patient still considering

Forty per cent of sites have returned patient screening information, whereas 60% of sites have not responded. The study team have worked hard to address this issue, maintaining regular contact with GP practices, running monthly competitions and updating the screening log to better reflect the process. This has recently resulted in more accurate screening logs being returned from more sites.

Figure 30 shows our actual recruitment numbers in comparison to our revised recruitment projections.

Figure 31 provides a breakdown of these figures by recruitment centre.



	Feb-18	Mar -18	Apr-18	May -18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18	Nov -18	Dec-18	Jan-19	Feb-19	Mar -19	Apr-19	May -19	Jun-19	Jul-19	Aug-19	Sep-19	Oct-19	Nov -19	Dec-19
Projected number of sites open	3	7	15	25	30	34	41	48	52	58	62	64	67	70	70	75	80	85	85	90	90	90	90
Actual number of sites open	6	11	20	25	28	37	42	48	51	57													

FIGURE 28 Sites open graph.

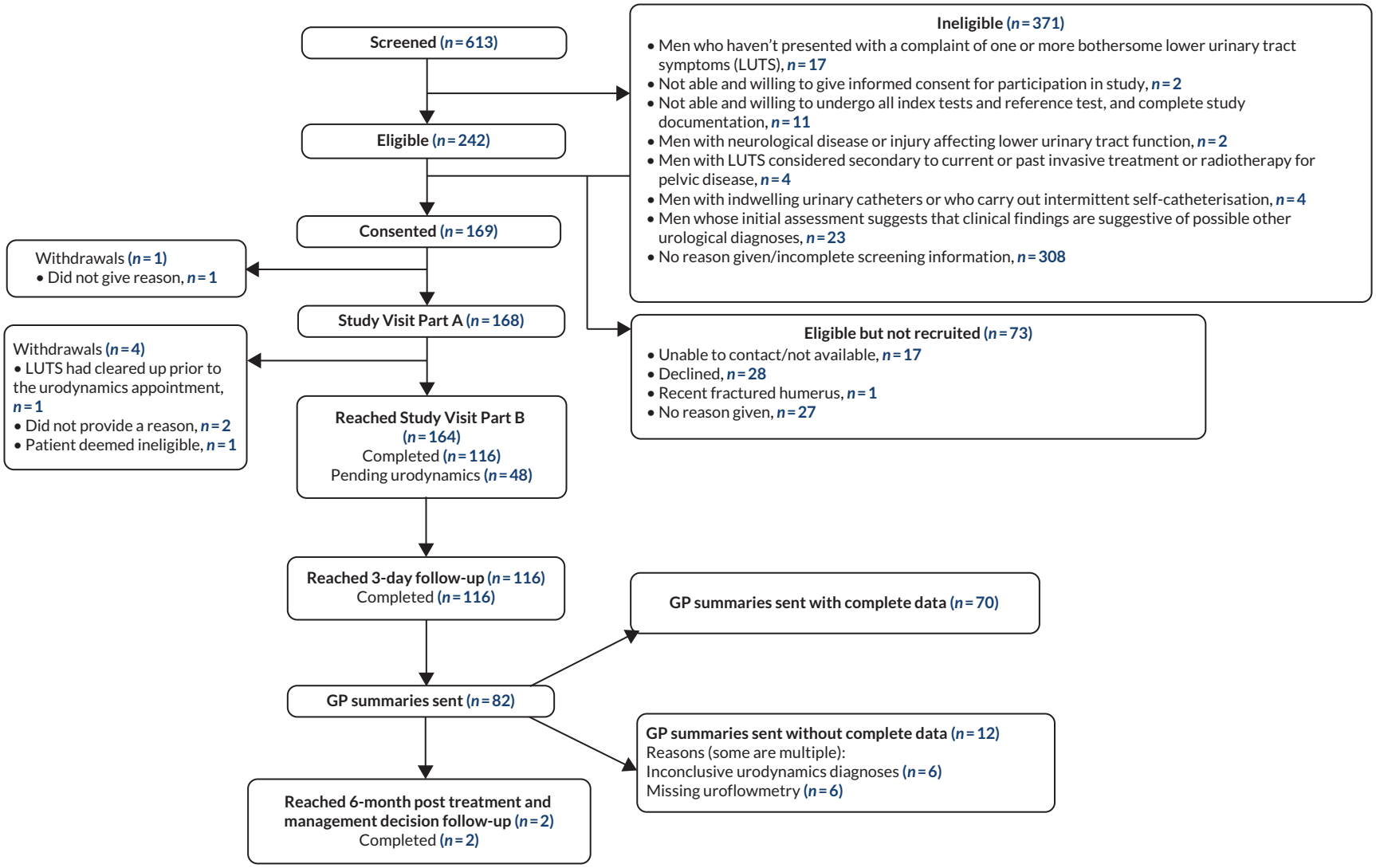
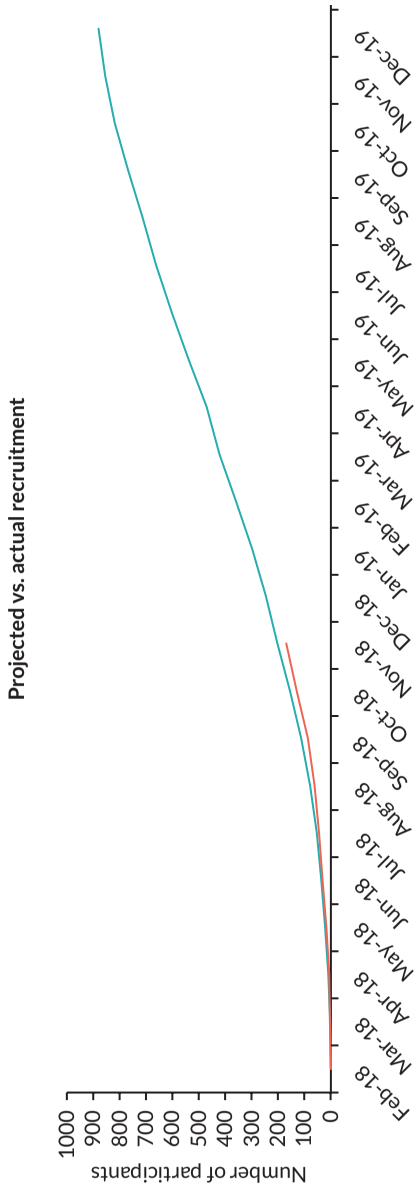


FIGURE 29 Screening, recruitment and data collection during interim pilot phase.



	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18	Nov-18	Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Sep-19	Oct-19	Nov-19	Dec-19
Proposed accumulative target	0	3	10	22	37	54	79	113	155	202	246	298	359	422	471	539	603	662	713	767	819	855	880
Actual accumulative accrual	0	4	6	17	31	45	62	87	130	169													

FIGURE 30 Projected vs. actual recruitment graph.

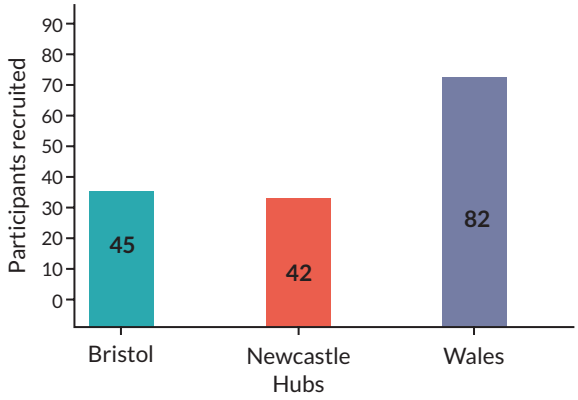


FIGURE 31 Recruitment during internal pilot phase of the PriMUS study.

Quality checking and analysis of urodynamic data

Overall, 116 urodynamic procedures have been successfully carried out (*Table 30*), out of the 169 participants who have been recruited. Nine urodynamic procedures have missing data (classed as missing Q_{max} and $P_{det}Q_{max}$).

One hundred and eleven urodynamic procedures have been quality checked out of the 116 carried out. Seventy-one of these gained consensus with the first reviewer. Forty out of these procedures have been escalated to a second reviewer due to no consensus on interpretation. A second reviewer receives the original nurses' data and the first reviewer's data, to allow them to come to an adjudication.

Collection of data

- Nine urodynamics procedures have missing data (classed as missing Q_{max} and $P_{det}Q_{max}$).
- Five withdrawals with no collection of urodynamic or uroflowmetry data.
- Five sets of uroflowmetry data missing.

The follow-up takes place 6 months after the treatment and management decision takes place. We calculate this timescale by adding 2× weeks after the GP summary has been generated. No follow-ups are due to be completed yet.

Adverse events

For the purpose of the PriMUS Study, only urodynamics-related adverse events are captured (*Table 31*). Six (5%) participants have experienced immediate adverse events after the urodynamic procedure and 17 (14%) participants have experienced adverse events reported within 3 days following the procedure.

Twenty-three participants out of 81 have experienced adverse events in relation to the urodynamic procedures.

We have received no reported SAEs.

Withdrawals

Five participants overall have withdrawn from the study.

TABLE 30 Detrusor overactivity, BOO and DU prevalence

Prevalence	No. of diagnoses present	No. with diagnoses absent	Estimated prevalence (%)	Actual prevalence (%)
DO	82	26	57	76
BOO	39	66	31	37
DU	42	62	16	40

TABLE 31 Adverse events

Immediate urodynamic adverse events	No. reported
Visible haematuria/urethral bleeding requiring intervention	3
Vasovagal episode (fainting)	3
3-Day follow-up phone call adverse events	**No. reported**
UTI	4
Visible haematuria	1
Bleeding from urethra	0
Urinary retention	0
Discomfort	15
Dysuria	2
Other	4

1. One participant was withdrawn from the study, as he felt his LUTS had cleared up prior to the urodynamics appointment.
2. Two other participants withdrew their consent from the study prior to the urodynamics but did not provide a reason.
3. One participant withdrew from the study before receiving any index tests or urodynamic test.
4. One participant was withdrawn, as upon further discussion with the patient they were actually deemed ineligible for the study. The GP had followed the eligibility criteria correctly and due to complexity of the patient, ineligibility only became apparent upon the patient's further declaration of symptoms.

Qualitative findings

Telephone interviews with a sample of patients and healthcare professionals from each study site are being conducted throughout the pilot phase to explore their experiences of the study (see *Table 32*).

TABLE 32 Telephone interview participants

Participant group		Target number	Completed to date
Patients participating in the main study	Newcastle	8	8
	Wales	6	4
	Bristol	6	0
Patients who declined to take part in the main study	Newcastle	2	2
	Wales	2	0
	Bristol	1	0
GPs	Newcastle	5	5
	Wales	5	0
	Bristol	5	0
Urodynamic nurses	Newcastle	1	0
	Wales	2	0
	Bristol	1	0

Main study recruitment process

In the majority of cases, patients were informed about the study and invited to take part by their GP during a routine appointment; Newcastle GPs suggested that this opportunistic face-to-face invitation was the most effective method of study recruitment. It was suggested that having an interested GP and administrator to run database searches regularly to identify potential participants was key to successfully using this pathway to recruitment. Several patients reported that they had mentioned their bladder symptoms while visiting their GP about something else, suggesting that publicising the study in GP waiting rooms is an effective strategy to increase recruitment uptake, emphasising the importance of displaying the PriMUS poster and study animation in practice waiting rooms.

One Newcastle GP mentioned that the recruitment process was overly lengthy as in some practices patients reply to the central Clinical Research Network (CRN) team to say that they want to take part, rather than contacting the GP directly. The GP then receives a message from the team asking them to make contact with the patient, which adds an extra step to the process. This highlights the need for additional communication and training with the CRN nurses to help simplify this process.

Patient motivation to take part

Patients gave three main reasons for deciding to take part in the study: (1) to help improve services in the future or to benefit others; (2) to reduce waiting time for diagnostic tests for themselves; and (3) to receive a comprehensive check-up. Some said that they wanted to help their GP surgery due to the good service they received. In line with this, GPs identified that a good relationship between the patient and GP had been a positive factor in encouraging men to take part in the study. GPs agreed that being able to have detailed tests motivated many patients to participate.

In line with the above, when asked what could be done to encourage others to take part, patients suggested GPs should emphasise that study participants will be able to access sophisticated diagnostic tests much more quickly than via usual care. GPs themselves reported that they 'sold' the study to patients by explaining that if they took part they would get hospital tests completed in a shorter time frame.

Both patients interviewed who declined to take part in the main study said this was because they did not want to go to a hospital (although they would not have needed to, highlighting the need for extra training with the CRN nurses in Newcastle to ensure patients are aware of this); however, one also said he did not want to undergo any invasive tests. One patient who took part in the main study said that he had preferred completing the tests at his local GP surgery and would not have taken part if he had to go to hospital for them, again suggesting that patients may prefer to be managed in primary care, rather than secondary care. Certainly, all patients interviewed agreed that they would prefer to have their bladder symptoms diagnosed and treated in primary rather than secondary care if possible. The main reasons for this were shorter waiting times, convenient location of their GP surgery, and familiarity (leading to better communication, patients feeling more comfortable and increased likelihood of seeing someone with knowledge of the patient's history). Patients felt confident in their GP to provide diagnosis and treatment, and several mentioned having a good relationship with their GP. The only cited advantage of secondary care was being able to access specialist advice or resources that may not always be available at the GP surgery. There was a general dislike of visiting hospitals among patients interviewed, who viewed this as 'something to be endured – it's not really somewhere you want to go is it?'

General practitioners identified that some patients from rural communities do not wish to take part as they do not want to travel to appointments, even in cases where all appointments would take place at their usual GP practice.

General practitioners reported that the invasive nature of urodynamics was the main factor that had stopped patients taking part in the study. It was suggested that some patients are not keen on having investigations in general and may agree to using a Flowtaker for example, but not to urodynamics. Some patients who were initially interested in participating dropped out after realising how much was involved.

Study information

Patient comments about the study information (both written materials and verbal explanation from healthcare professionals) were very positive: patients felt that it was easy to understand and that they had been fully prepared for the study. Some specifically mentioned that they had valued the lack of pressure to take part, together with checks

by healthcare staff that they were still happy to be involved at every stage. GPs agreed that the study information was straightforward to explain to patients, although it was identified that there is a lot of information for patients to take in, with one GP suggesting that it could be simplified by reducing it to an explanation of urodynamics and what will happen on the day. However, patients felt they had received the right amount of information. It was suggested that loose study information sheets would be helpful for GPs to have on their desks, rather than only having them as part of the study packs.

Study assessments

One GP reported that having a PSA test as part of the study screening is potentially problematic as results are not necessarily accurate, and this can sometimes raise more questions and lead to additional investigations which could be unnecessary. However, additional investigations raised from PSA testing are important to ensure a patient's eligibility in taking part in the study.

The majority of patients reported no issues in completing the bladder diary and said that it was clear what they had to do. Two mentioned that one of the statements in the diary ('if you had urgency but it passed away before you had to visit the toilet') did not accurately describe their own symptoms. Both explained that the urgency they felt did not pass away before they got to the toilet, but they could not pass urine when they got to the toilet. The bladder diary is a validated document, so the study team cannot make changes, or adapt this document to address the comment above.

Patients reported that the Flowtaker was easy to use but could be an 'inconvenience'. Some mentioned missed readings because they were out of the house a lot and could not take the jug with them as it was too large. One said that he tried to time going to the toilet around his working hours to ensure that he could use the Flowtaker. This highlights potential bias as he changed his voiding habits to fit in with study processes, emphasising the need for nurses to explain the purpose of the Flowtaker and iterate to patients that they need to use it in the same way as they use the bladder diary. Another patient said that the Flowtaker was a 'nuisance' to use through the night when he was half-asleep. One said that he had forgotten to use the Flowtaker several times but had tackled this by putting the toilet lid down and placing the jug on top as a reminder.

Patients reported experiencing mild discomfort when having the urodynamic test, but said that the procedure was 'fine'. The majority felt well-prepared and said that the test was explained fully, reporting that it was no better or worse than they had expected. Two felt it was better than they had expected, and none reported any problems. Several patients felt that the procedure was more invasive than they had anticipated, but felt that they had been given the right amount of information beforehand. Patients reported that the nurses who conducted the test were very professional, put them at ease and explained what was happening throughout the procedure.

General practitioners confirmed that patients appeared to have found study measures acceptable and have not reported any problems or concerns.

Study procedures

Study appointments were generally at patients' own GP surgery or a nearby surgery, and were scheduled at a convenient time for patients. None reported problems travelling to appointments, although one said he had to travel an extra 20 miles to a different surgery so would have preferred to have tests completed at his local GP practice. GPs also identified that it was advantageous for urodynamics to be conducted at the patients' local surgery.

General practitioners reported that it has been easy to stick to the study protocol, and that study procedures have been straightforward and have run smoothly. They felt well-informed about the study and identified that information from the study team has been clear and comprehensive.

The importance of getting patient results to GPs promptly was emphasised; in one case, this had been delayed. The study team have worked hard to manage patient and GPs' expectations of receiving the GP Summary, due to all the internal quality checks for the urodynamic data review. This is now being managed more effectively and patients and GPs have a realistic expectation of when they will receive the results.

General practitioner workload

It was reported that the study has been realistic in terms of workload and has matched expectations, as everything GPs need to do to prepare patients for the study are things they would be doing anyway, such as examinations, taking patient history, etc.

However, where patients have been recruited via a mail-out rather than at an appointment this has been problematic in terms of timescale, as there has been an expectation that all these preparatory things can be completed with quite a tight turnaround. In one case, a patient was offered a study appointment so quickly that the GP did not have sufficient time to have an initial meeting with the patient and complete the study forms, emphasising the importance of discussing the patient pathway with the Newcastle CRN.

One patient had completed all the preparatory processes necessary for the study; however, unfortunately they were a few days over the 6-month cut-off, so examinations needed to be repeated. This has highlighted a need to look at making the timescale requirement clearer in the study protocol.

Conclusion

Overall, patients reported that they were happy with their involvement in the study and felt that it was explained well. They valued being able to have a comprehensive check-up with minimal waiting time and were pleased that other patients may benefit from their involvement in the future.

All the patients reported that they would prefer to be managed in primary rather than secondary care, confirming that a decision support tool in primary care would be acceptable to patients.

The study team will not change study material, as patients have reported being happy with this. Although a GP suggested it could be simplified by reducing it to an explanation of urodynamics and what will happen on the day, the urodynamic information sheet appears to be important for the patient to be able to fully consider whether they want to take part or not.

The patients' views of the study assessments were helpful, particularly the ones completed at home, to establish whether they were working. The study team plan to look at ways to address the patient views of using the Flowtaker, to help reduce any potential bias. Patients have advised of their general acceptance of the urodynamic test; although this was an invasive procedure, the nurses appear to have been extremely professional in their ability to provide information and reassurance throughout.

All the GP views are from the Newcastle Hub, so we will continue with the acceptability interviews into the main recruitment phase, to collate feedback from all hubs, to continue to improve processes for site staff delivering the study and patients recruited into the study.

The Study Manager and the Newcastle CRN Facilitator have planned to look at ways of modifying the patient pathway in Newcastle to help simplify the patient pathway (perhaps suggesting Wales/Bristol pathway) and address some of the difficulties Newcastle have experienced. The study manager will also facilitate additional training for Newcastle CRN nurses, to ensure the processes are streamlined. The study team will continue to emphasise the importance of displaying the PriMUS poster and study animation in practice waiting rooms. This will hopefully encourage an uptake in patient recruitment numbers in Newcastle.

In conclusion, the qualitative interviews have shown that the study has been positively received by both GPs and study participants and highlighted any issues to improve study processes.

Discussion

Modifications to design and/or study processes

We hold monthly weekly Study Management Meetings, where we have discussed ways to improve recruitment rates. We have evaluated and improved a few recruitment processes for sites to help increase recruitment rates:

Changes to study documentation

- The study team have combined the mail-out packs to just one mail-out pack, reducing the number of steps sites needed to take to recruit patients via the database search pathway.
- The study team have updated the screening log to reflect the recruitment process. This ensures that sites are following up potential patients to see if they want to consent and allowing the sites to advise whether the patient was identified via database/opportunistic recruitment. This new version was rolled out in the August newsletter.

Training

- Study Manager compiled and delivered further primary care training to staff running Study Visit Part A (who were not the urodynamic nurses delivering this visit). This helped improve their understanding and knowledge of the urodynamics procedure, along with frequently asked questions, in order to give the patient as much information as they need to make an informed decision about consenting to the study.

Site engagement

- Study team are in regular contact with sites (particularly those who have not recruited) to find out why they have not recruited yet. The team make suggestions to improve recruitment (ensuring poster/animation displayed in practice, following up patients contacted via the mail-out pack), involving more staff at the practice, using the manuals and information available.
- Study team provide tailored support to practices, discussing any problems they may have which is impacting their recruitment rates and helping to address and resolve these issues.
- Study Team GP has recently been responsible for contacting the sites who have not recruited. We hope that they may be more likely to share issues/problems with recruitment to a peer (than a researcher) and help us to resolve any of these. The Study Team GP will also go and visit these sites if necessary.
- Study Team have been in contact with well recruiting sites, to collate feedback and share best practice with sites who are not performing as well in their recruitment.

Branding and publicity

- Monthly newsletters are sent which include recruitment and data collection tips, friendly reminders and competitions to try and encourage practices to engage and thus maximise recruitment.
- The study has a presence on social media. The team circulate regular tweets to engage practices in recruitment.

Proposed modifications to design and/or trial processes

Patient pathway and screening

It is currently mandatory for the GP to complete the screening tests and therefore confirm eligibility, as we originally found that most nurses may not be trained in performing the DRE screening test. By attending practices and collating feedback from GP practices, it transpires that some practices have already trained nurse practitioners who can complete these screening tests. We suggest amending the protocol to permit already trained nurse practitioners to complete the screening tests, if this is available in the practice.

Additional sites

We have received interest from a consultant urologist in a fourth hub in Birmingham, who has liaised with his local CRN and offered his urodynamic nurse to carry out the procedures (along with the necessary equipment). We are currently liaising with the local CRN and Consultant Urologist, to assess feasibility of setting up a fourth hub in this area. If this becomes a possibility, we will have a bigger patient population to recruit from, which should increase our recruitment uptake. This would address the gap between our target and actual recruitment numbers.

One site in Newcastle reported that some elderly gentlemen were not able to travel to the hub. The CRN and Study Manager liaised to promote this site to a hub practice, so that in this scenario, the gentlemen could receive the urodynamics in their own practice.

Site engagement

The study manager and Newcastle CRN plan to address the GP views and corresponding recruitment numbers in Newcastle by re-assessing their patient pathway and simplifying the process. The study manager will also facilitate additional training with the CRN nurses, to ensure they are confident and competent in these processes and how they can support GP Practices to recruit.

Recommendations from Study Steering Committee end of pilot phase meeting

Following the SSC Meeting on 20 November 2018, the members have made the following decisions regarding the PriMUS Study:

- The Independent members are happy for the study to continue.
- All members were happy with the small number of adverse events. They advised that they would like to see other information including:
    - The percentages of patients that report adverse events (as a patient may experience more than one adverse event).
    - An average age of patients.
- Re-consideration of accrual recruitment figures as the figures for data completion are higher than expected.
- Consideration of resources for opening a fourth hub if it is believed sites currently open would become saturated.

Appendix 3 Model results from the sensitivity analyses

TABLE 33 Recruitment by study hub and GP practice

Study hub	GP practice	Frequency (%) *n* = 350
Wales Primary Care ('PiCRIS') Research Network	PRACTICE W1	7 (2.0)
	W2	2 (0.6)
	W3	24 (6.9)
	W4	1 (0.3)
	W5	7 (2.0)
	W6	6 (1.7)
	W7	6 (1.7)
	W8	3 (0.9)
	W9	4 (1.1)
	W10	3 (0.9)
	W11	5 (1.4)
	W12	6 (1.7)
	W13	2 (0.6)
	W14	3 (0.9)
	W15	9 (2.6)
	W16	1 (0.3)
	W17	3 (0.9)
	W18	3 (0.9)
	W19	1 (0.3)
	W20	2 (0.6)
	W21	5 (1.4)
	W22	2 (0.6)
	W23	10 (2.9)
	W24	3 (0.9)
	W25	0 (0.0)
	W26	6 (1.7)
	W27	2 (0.6)
	W28	1 (0.3)
	W29	0 (0.0)
	W30	0 (0.0)
	Subtotal	127 (36.3)

continued

TABLE 33 Recruitment by study hub and GP practice (*continued*)

Study hub	GP practice	Frequency (%) *n* = 350
North East England and North Cumbria Research Network	PRACTICE N1	8 (2.3)
	N2	1 (0.3)
	N3	0 (0.0)
	N4	7 (2.0)
	N5	5 (1.4)
	N6	7 (2.0)
	N7	2 (0.6)
	N8	10 (2.9)
	N9	3 (0.9)
	N10	1 (0.3)
	N11	7 (2.0)
	N12	0 (0.0)
	N13	5 (1.4)
	N14	0 (0.0)
	N15	3 (0.9)
	N16	5 (1.4)
	N17	3 (0.9)
	N18	0 (0.0)
	N19	1 (0.3)
	N20	1 (0.3)
	N21	0 (0.0)
	N22	0 (0.0)
	Subtotal	69 (19.7)
Western Research Network (Bristol)	B1	6 (1.7)
	B2	0 (0.0)
	B3	19 (5.4)
	B4	0 (0.0)
	B5	8 (2.3)
	B6	2 (0.6)
	B7	4 (1.1)
	B8	65 (18.6)
	B9	10 (2.9)
	B10	1 (0.3)
	B11	2 (0.6)

TABLE 33 Recruitment by study hub and GP practice (*continued*)

Study hub	GP practice	Frequency (%) *n* = 350
	B12	28 (8.0)
	B13	8 (2.3)
	B14	1 (0.3)
	B15	0 (0.0)
	B16	0 (0.0)
	Subtotal	154 (44.0)

Note
Practices anonymised to avoid potential identification of recruited participants.

TABLE 34 Participant demographics and clinical characteristics of the development cohort for diagnosis of BOO

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		BOO *n* = 163	No BOO *n* = 173	Indeterminate *n* = 14	Total *n* = 350
Age (years)	Median (IQR)	69.8 (64.0–74.9)	67.8 (58.9–74.7)	69.1 (62.0–75.9)	69.0 (61.8–74.9)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Height (cm)	Median (IQR)	175.0 (171.0–180.0)	177.0 (172.0–180.0)	177.5 (176.0–180.0)	176.0 (171.0–180.0)
	Missing, *n* (%)	0 (0.0)	1 (0.6)	0 (0.0)	1 (0.3)
Weight (kg)	Median (IQR)	82.0 (73.0–93.0)	85.0 (76.0–97.0)	78.5 (71.0–96.0)	83.5 (75.0–96.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Ethnicity, *n* (%)	White	158 (96.9)	169 (97.7)	14 (100.0)	341 (97.4)
	Asian (Pakistani)	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.3)
	Asian (Chinese)	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.3)
	Asian (Indian)	0 (0.0)	2 (1.2)	0 (0.0)	2 (0.6)
	Asian (Other)	2 (1.2)	0 (0.0)	0 (0.0)	2 (0.6)
	Black (Caribbean)	0 (0.0)	2 (1.2)	0 (0.0)	2 (0.6)
	Black (African)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Black (Other)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Other	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.3)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Smoking status, *n* (%)	Never	67 (41.1)	75 (43.4)	8 (57.1)	150 (42.9)
	Current	17 (10.4)	12 (6.9)	0 (0.0)	29 (8.3)
	Ex-smoker	79 (48.5)	87 (49.7)	6 (42.9)	171 (48.9)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

continued

TABLE 34 Participant demographics and clinical characteristics of the development cohort for diagnosis of BOO (*continued*)

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		BOO *n* = 163	No BOO *n* = 173	Indeterminate *n* = 14	Total *n* = 350
Medical history, *n* (%)	None	43 (26.4)	58 (33.5)	6 (42.9)	107 (30.6)
	Adrenal insufficiency	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Cerebrovascular accident	7 (4.3)	5 (2.9)	0 (0.0)	12 (3.4)
	Chronic heart failure	0 (0.0)	2 (1.2)	0 (0.0)	2 (0.6)
	Chronic venous stasis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Dependent oedema	1 (0.6)	1 (0.6)	1 (7.1)	3 (0.9)
	Diabetes insipidus	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus I	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.3)
	Diabetes mellitus II	19 (11.7)	22 (12.7)	0 (0.0)	41 (11.7)
	Hypercalcaemia	1 (0.6)	2 (1.2)	0 (0.0)	3 (0.9)
	Liver failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Obstructive sleep apnoea	3 (1.8)	9 (5.2)	1 (7.1)	13 (3.7)
	Polyuric renal failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Pyelonephritis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Sickle cell anaemia	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.3)
	Other	115 (70.6)	110 (63.6)	7 (50.0)	232 (66.3)

IQR, interquartile range.

TABLE 35 Participant demographics and clinical characteristics of the development cohort for diagnosis of DU

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		DU *n* = 141	No DU *n* = 193	Indeterminate *n* = 16	Total *n* = 350
Age (years)	Median (IQR)	69.2 (63.1–75.1)	68.5 (61.2–74.6)	69.1 (61.7–75.2)	69.0 (61.8–74.9)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Height (cm)	Median (IQR)	175.0 (170.0–180.0)	176.0 (172.0–180.0)	177.5 (174.5–180.0)	176.0 (171.0–180.0)
	Missing, *n* (%)	1 (0.7)	0 (0.0)	0 (0.0)	1 (0.3)
Weight (kg)	Median (IQR)	83.0 (74.0–95.0)	84.0 (76.0–96.0)	81.0 (73.0–97.5)	83.5 (75.0–96.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

TABLE 35 Participant demographics and clinical characteristics of the development cohort for diagnosis of DU (*continued*)

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		DU *n* = 141	No DU *n* = 193	Indeterminate *n* = 16	Total *n* = 350
Ethnicity, *n* (%)	White	138 (97.9)	187 (96.9)	16 (100.0)	341 (97.4)
	Asian (Pakistani)	0 (0.0)	1 (0.5)	0 (0.0)	1 (0.3)
	Asian (Chinese)	0 (0.0)	1 (0.5)	0 (0.0)	1 (0.3)
	Asian (Indian)	2 (1.4)	0 (0.0)	0 (0.0)	2 (0.6)
	Asian (Other)	0 (0.0)	2 (1.0)	0 (0.0)	2 (0.6)
	Black (Caribbean)	1 (0.7)	1 (0.5)	0 (0.0)	2 (0.6)
	Black (African)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Black (Other)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Other	0 (0.0)	1 (0.5)	0 (0.0)	1 (0.3)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Smoking status, *n* (%)	Never	64 (45.4)	77 (39.9)	9 (56.3)	150 (42.9)
	Current	6 (4.3)	22 (11.4)	1 (6.3)	29 (8.3)
	Ex-smoker	71 (50.4)	94 (48.7)	6 (37.5)	171 (48.9)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Medical history, *n* (%)	None	48 (34.0)	52 (26.9)	7 (43.8)	107 (30.6)
	Adrenal insufficiency	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Cerebrovascular accident	6 (4.3)	6 (3.1)	0 (0.0)	12 (3.4)
	Chronic heart failure	2 (1.4)	0 (0.0)	0 (0.0)	2 (0.6)
	Chronic venous stasis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Dependent oedema	1 (0.7)	1 (0.5)	1 (6.3)	3 (0.9)
	Diabetes insipidus	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus I	1 (0.7)	0 (0.0)	0 (0.0)	1 (0.3)
	Diabetes mellitus II	16 (11.4)	24 (12.4)	1 (6.3)	41 (11.7)
	Hypercalcaemia	2 (1.4)	1 (0.5)	0 (0.0)	3 (0.9)
	Liver failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Obstructive sleep apnoea	5 (3.5)	7 (3.6)	1 (6.3)	13 (3.7)
	Polyuric renal failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Pyelonephritis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Sickle cell anaemia	1 (0.7)	0 (0.0)	0 (0.0)	1 (0.3)
	Other	88 (62.4)	136 (70.5)	8 (50.0)	232 (66.3)

IQR, interquartile range.

TABLE 36 Participant demographics and clinical characteristics of the development cohort for diagnosis of DO

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		DO *n* = 253	No DO *n* = 87	Indeterminate *n* = 10	Total *n* = 350
Age (years)	Median (IQR)	70.2 (64.0–75.6)	64.5 (56.6–70.7)	67.6 (64.5–71.8)	69.0 (61.8–74.9)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Height (cm)	Median (IQR)	176.0 (171.0– 180.0)	176.0 (171.0– 180.0)	178.5 (173.0–180.0)	176.0 (171.0–180.0)
	Missing, *n* (%)	0 (0.0)	1 (1.1)	0 (0.0)	1 (0.3)
Weight (kg)	Median (IQR)	83.0 (73.0–95.0)	85.0 (78.0–99.0)	86.3 (76.0–96.0)	83.5 (75.0–96.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Ethnicity, *n* (%)	White	248 (98.0)	83 (95.4)	10 (100.0)	341 (97.4)
	Asian (Pakistani)	1 (0.4)	0 (0.0)	0 (0.0)	1 (0.3)
	Asian (Chinese)	0 (0.0)	1 (1.2)	0 (0.0)	1 (0.3)
	Asian (Indian)	1 (0.4)	1 (1.2)	0 (0.0)	2 (0.6)
	Asian (Other)	2 (0.8)	0 (0.0)	0 (0.0)	2 (0.6)
	Black (Caribbean)	1 (0.4)	1 (1.2)	0 (0.0)	2 (0.6)
	Black (African)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Black (Other)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Other	0 (0.0)	1 (1.2)	0 (0.0)	1 (0.3)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Smoking status, *n* (%)	Never	109 (43.1)	36 (41.4)	5 (50.0)	150 (42.9)
	Current	22 (8.7)	7 (8.1)	0 (0.0)	29 (8.3)
	Ex-smoker	122 (48.2)	44 (50.6)	5 (50.0)	171 (48.9)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Medical history, *n* (%)	None	73 (27.7)	33 (37.9)	4 (40.0)	107 (30.6)
	Adrenal insufficiency	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Cerebrovascular accident	10 (4.0)	2 (2.3)	0 (0.0)	12 (3.4)
	Chronic heart failure	1 (0.4)	1 (1.2)	0 (0.0)	2 (0.6)
	Chronic venous stasis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Dependent oedema	2 (0.8)	0 (0.0)	1 (10.0)	3 (0.9)
	Diabetes insipidus	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus I	1 (0.4)	0 (0.0)	0 (0.0)	1 (0.3)
	Diabetes mellitus II	30 (11.9)	11 (12.6)	0 (0.0)	41 (11.7)
	Hypercalcaemia	2 (0.8)	1 (1.2)	0 (0.0)	3 (0.9)
	Liver failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Obstructive sleep apnoea	9 (3.6)	3 (4.6)	1 (10.0)	13 (3.7)

TABLE 36 Participant demographics and clinical characteristics of the development cohort for diagnosis of DO (*continued*)

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		DO *n* = 253	No DO *n* = 87	Indeterminate *n* = 10	Total *n* = 350
	Polyuric renal failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Pyelonephritis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Sickle cell anaemia	1 (0.4)	0 (0.0)	0 (0.0)	1 (0.3)
	Other	178 (70.4)	49 (56.3)	5 (50.0)	232 (66.3)

IQR, interquartile range.

TABLE 37 Participant treatment history based on diagnosis of BOO for the development cohort

Treatment history, *n* (%)		Diagnosis			
		BOO *n* = 163	No BOO *n* = 173	Indeterminate *n* = 14	Total *n* = 350
None		22 (13.5)	28 (16.2)	3 (21.4)	53 (15.1)
Uroselective medications	Alpha-blockers	48 (29.4)	49 (28.3)	5 (35.7)	102 (29.1)
	5-alpha reductase inhibitors	9 (5.5)	13 (7.5)	3 (21.4)	25 (7.1)
	Anticholinergics	4 (2.5)	10 (5.8)	0 (0.0)	14 (4.0)
	Beta-3-agonists	2 (1.2)	4 (2.3)	1 (7.1)	7 (2.0)
Calcium channel blockers		29 (17.8)	36 (20.8)	0 (0.0)	65 (18.6)
Diuretics	Thiazides and related diuretics	7 (4.3)	11 (6.3)	0 (0.0)	18 (5.1)
	Loop diuretics	2 (1.2)	5 (2.9)	0 (0.0)	7 (2.0)
SSRI antidepressants		13 (8.0)	13 (7.5)	0 (0.0)	26 (7.4)
Other		116 (71.2)	135 (78.0)	11 (78.6)	262 (74.9)

TABLE 38 Participant treatment history based on diagnosis of DU for the development cohort

Treatment history, *n* (%)		Diagnosis			
		DU *n* = 141	No DU *n* = 193	Indeterminate *n* = 16	Total *n* = 350
None		19 (13.5)	30 (15.5)	4 (25.0)	53 (15.1)
Uroselective medications	Alpha-blockers	38 (27.0)	58 (30.1)	6 (37.5)	102 (29.1)
	5-alpha reductase inhibitors	12 (8.5)	10 (5.2)	3 (18.8)	25 (7.1)
	Anticholinergics	6 (4.3)	8 (4.1)	0 (0.0)	14 (4.0)
	Beta-3-agonists	0 (0.0)	6 (3.1)	1 (6.3)	7 (2.0)
Calcium channel blockers		19 (13.5)	45 (23.3)	1 (6.3)	65 (18.6)
Diuretics	Thiazides and related diuretics	8 (5.7)	10 (5.2)	0 (0.0)	18 (5.1)
	Loop diuretics	5 (3.5)	2 (1.0)	0 (0.0)	7 (2.0)
SSRI antidepressants		8 (5.7)	17 (8.8)	1 (6.3)	26 (7.4)
Other		106 (75.2)	144 (74.6)	12 (75.0)	262 (74.9)

TABLE 39 Participant treatment history based on diagnosis of DO for the development cohort

Treatment history, *n* (%)		Diagnosis			
		DO *n* = 253	No DO *n* = 87	Indeterminate *n* = 11	Total *n* = 350
None		37 (14.6)	13 (14.9)	3 (30.0)	53 (15.1)
Uroselective medications	Alpha-blockers	79 (31.2)	20 (23.0)	3 (3.0)	102 (29.1)
	5-alpha reductase inhibitors	16 (6.3)	8 (9.2)	1 (10.0)	25 (7.1)
	Anticholinergics	12 (4.7)	2 (2.3)	0 (0.0)	14 (4.0)
	Beta-3-agonists	6 (2.4)	0 (0.0)	1 (10.0)	7 (2.0)
Calcium channel blockers		51 (20.2)	14 (16.1)	0 (0.0)	65 (18.6)
Diuretics	Thiazides and related diuretics	12 (4.7)	6 (6.9)	0 (0.0)	18 (5.1)
	Loop diuretics	6 (2.4)	1 (1.1)	0 (0.0)	7 (2.0)
SSRI antidepressants		20 (7.9)	6 (6.9)	0 (0.0)	26 (7.4)
Other		190 (75.1)	65 (74.7)	7 (70.0)	262 (74.9)

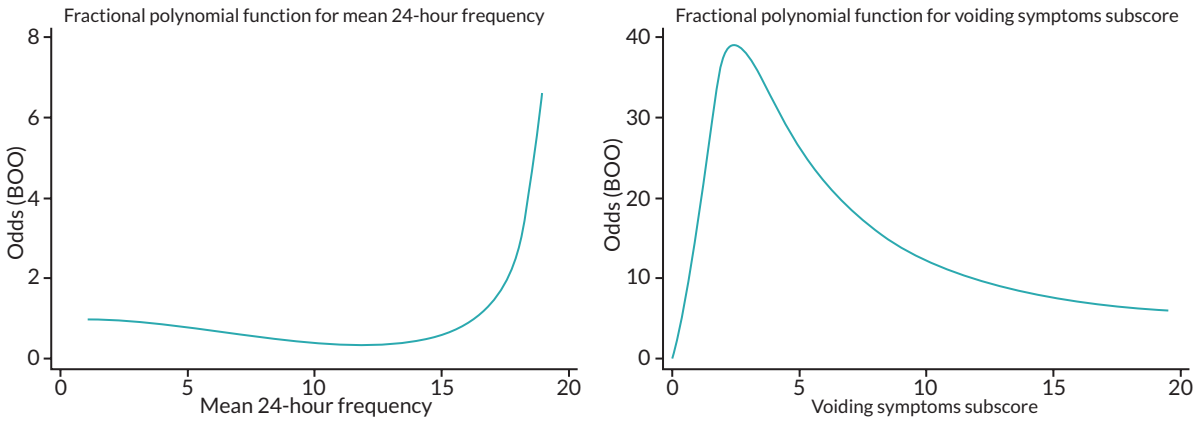


FIGURE 32 Fractional polynomial functions for mean 24-hour frequency (left) and voiding symptoms subscore (right) in BOO model 2.

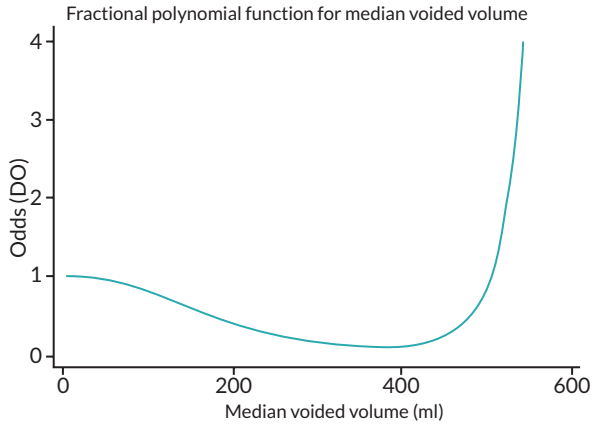


FIGURE 33 Fractional polynomial function for median voided volume in DO model.

Sensitivity analysis 1

TABLE 40 Apparent coefficients and odds ratios for predictors included in the SA1 model for BOO

	Apparent		
Predictors and intercept	**Coefficient (95% CI)**	**Odds ratio (95% CI)**	***p*-value**
Age_1 (years)	−815.23 (−1411.12 to 219.35)	0.00 (0.00 to 5.48e-96)	0.007
Age_2 (years)	644.29 (193.99 to 1094.60)	6.5e+ 279 (1.77e+ 84)[a]	0.005
PSA result (ng/ml)	0.19 (−0.01 to 0.39)	1.20 (0.99 to 1.47)	0.066
Voiding symptoms subscore_1 (ICIQ–MLUTS questionnaire)	2.73 (1.08 to 4.37)	15.26 (2.96 to 78.76)	0.001
Voiding symptoms subscore_2 (ICIQ–MLUTS questionnaire)	1.34 (0.48 to 2.20)	3. 82 (1.62 to 9.02)	0.002
Median maximum flow rate (ml/second)	−0.38 (−0.47 to −0.28)	0.69 (0.62 to 0.76)	0.001
Median voided volume (ml)	0.011 (0.005 to 0.016)	1.011 (1.005 to 1.016)	0.001
Mean 24-hour frequency_1 (voids per day)	−1.01 (−1.85 to −0.17)	0.36 (0.16 to 0.84)	0.018
Mean 24-hour frequency_2 (voids per day)	2.09 (0.38 to 3.80)	8.09 (1.46 to 44.66)	0.017
Intercept	−8.71 (−14.77 to −2.64)	0.0002 (3.83e-07 to 0.07)	0.005

$$Age\_1 = \left(\frac{age}{10}\right)^{-2}$$

$$Age\_2 = \left(\frac{age}{10}\right)^{-2} \ln\left(\frac{age}{10}\right)$$

$$Voiding\ symptoms\ \text{sub-}score\_1 = \left(\frac{Vss + 1}{10}\right)^{-1}$$

$$Voiding\ symptoms\ \text{sub-}score\_2 = \left(\frac{Vss + 1}{10}\right)^{-1} \ln\left(\frac{Vss + 1}{10}\right)$$

$$Mean\ 24\text{-}hour\ frequency\_1 = \left(\frac{MF}{10}\right)^{3}$$

$$Mean\ 24\text{-}hour\ frequency\_2 = \left(\frac{MF}{10}\right)^{3} \ln\left(\frac{MF}{10}\right)$$

MF, mean 24-hour frequency; *VSS*, voiding symptoms subscore.
a Upper bound of the CI is extremely large.

TABLE 41 Apparent model coefficients and odds ratios for predictors included in the SA1 model for DU

	Apparent		
Predictors and intercept	**Coefficient (95% CI)**	**Odds ratio (95% CI)**	***p*-value**
Post-void residual urine (ml)	0.003 (0.001 to 0.005)	1.003 (1.001 to 1.005)	0.001
Median maximum flow rate (ml/second)	−0.08 (−0.12 to −0.04)	0.92 (0.88 to 0.96)	0.002
Intercept	0.41 (−0.24 to 1.06)	1.51 (0.79 to 2.89)	0.213

TABLE 42 Apparent model coefficients and odds ratios for predictors included in the SA1 model for DO

Predictors and intercept	Apparent		
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value
Age (years)	0.05 (0.03 to 0.08)	1.05 (1.03 to 1.08)	0.001
Post-void residual urine (ml)	−0.003 (−0.005 to −0.001)	0.997 (0.995 to 0.999)	
Median voided volume_1 (ml)	−0.19 (−0.32 to −0.07)	0.82 (0.73 to 0.93)	0.002
Median voided volume_2 (ml)	0.13 (0.04 to 0.21)	1.13 (1.04 to 1.23)	0.004
Intercept	−1.30 (−3.05 to 0.46)	0.27 (0.05 to 1.58)	0.147

$$Median\ voided\ volume\_1 = \left(\frac{MVV}{100}\right)^3$$

$$Median\ voided\ volume\_2 = \left(\frac{MVV}{100}\right)^3 \ln\left(\frac{MVV}{100}\right)$$

Sensitivity analysis 2

TABLE 43 Apparent model coefficients and odds ratios and optimism-corrected coefficients for predictors included in the SA2 model for BOO

Predictors and intercept	Apparent			Optimism-corrected coefficient
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value	
Age (years)	0.01 (−0.01 to 0.04)	1.01 (0.99 to 1.04)	0.523	0.01
PSA test result (ng/ml)	0.23 (0.06 to 0.40)	1.26 (1.06 to 1.50)	0.004	0.19
Maximum flow rate (ml/second)	−0.15 (−0.22 to −0.07)	0.86 (0.80 to 0.93)	0.001	−0.12
Voided volume (ml)	−2.19 (−4.33 to −0.05)	0.11 (0.01 to 0.95)	0.003	−1.81
Intercept[a]	3.39 (−0.73 to 7.52)	29.71 (0.48 to 1838.62)	0.302	2.80

$$Voided\ volume = \left(\frac{VV + 294.95}{1000}\right)^{-0.5}$$

a Upper bound of the CI is extremely large.

Note
Optimism-corrected coefficients multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.83.

TABLE 44 Apparent model coefficients and odds ratios and optimism-corrected coefficients for predictors included in the SA2 model for DU

Predictors and intercept	Apparent			Optimism-corrected coefficient
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value	
Maximum flow rate (ml/second)	−0.13 (−0.20 to −0.07)	0.87 (0.82 to 0.94)	0.001	−0.11
Voided volume (ml)	0.004 (0.001 to 0.007)	1.004 (1.001 to 1.007)	0.006	0.003
Mean 24-hour fluid intake (ml)	−0.0006 (−0.0010 to −0.0002)	0.9994 (0.9990 to 0.9998)	0.006	−0.0004
Intercept	1.22 (0.44 to 2.00)	3.39 (1.56 to 7.37)	0.002	1.18

Note
Optimism-corrected coefficients multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.80.

TABLE 45 Apparent model coefficients and odds ratios and optimism-corrected coefficients for predictors included in the SA2 model for DO

Predictors and intercept	Apparent			Optimism-corrected coefficient
	Coefficient (95% CI)	Odds ratio (95% CI)	*p*-value	
Age (years)	0.05 (0.02 to 0.07)	1.05 (1.02 to 1.07)	0.001	0.03
Storage symptoms subscore (IPSS questionnaire)	0.14 (0.06 to 0.22)	1.15 (1.06 to 1.24)	0.001	0.10
Intercept	−3.09 (−4.78 to −1.41)	0.05 (0.01 to 0.24)	0.001	−2.05

Note
Optimism-corrected coefficients multiplied with a shrinkage factor (obtained from the bootstrapping procedure) of 0.74.

Appendix 4 Model performance from the sensitivity analyses

Sensitivity analysis 1

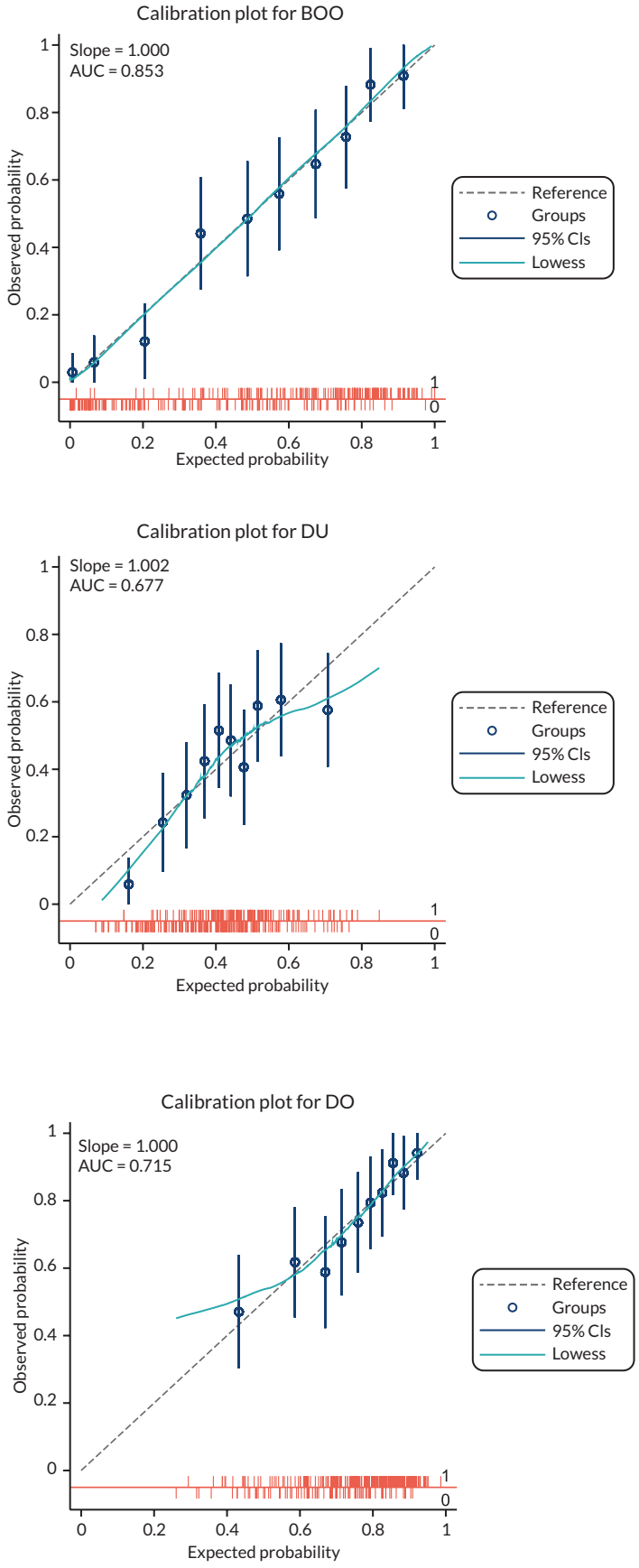


FIGURE 34 Calibration plots comparing observed and predicted probabilities for BOO, DU and DO, by decile of risk (SA1). AUC, area under the curve.

Sensitivity analysis 2

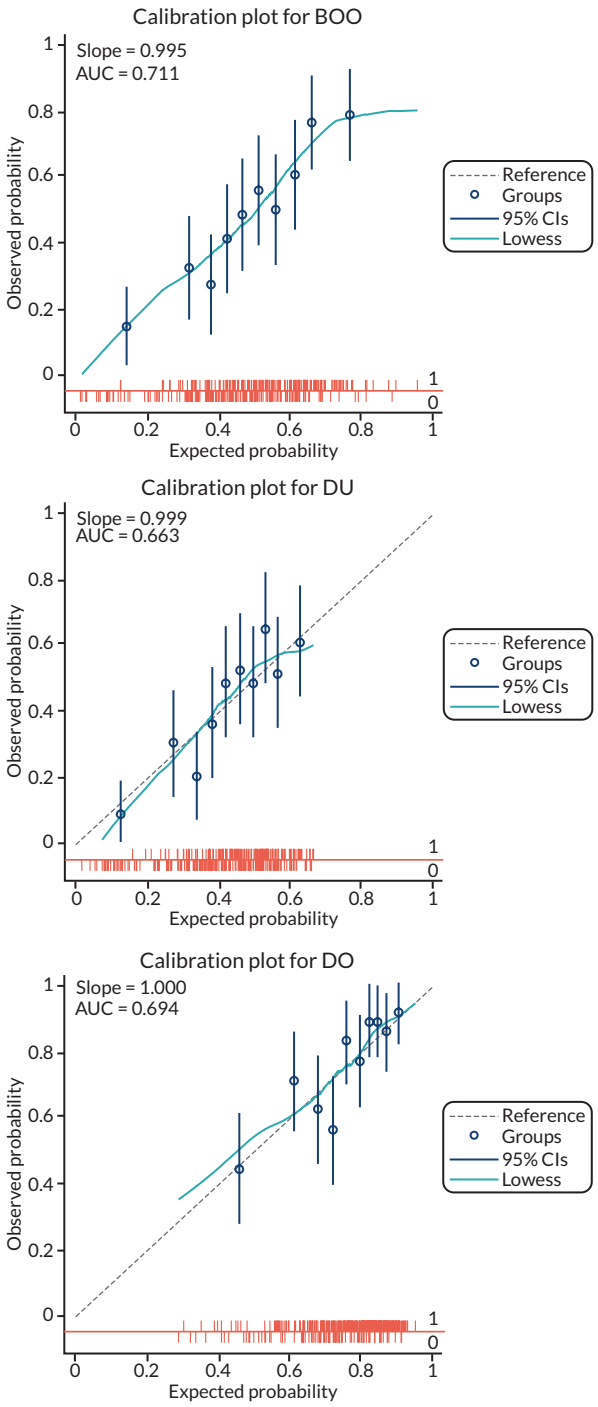


FIGURE 35 Calibration plots comparing observed and predicted probabilities for BOO, DU and DO, by decile of risk (SA2). AUC, area under the curve.

TABLE 46 Recruitment by study hub and GP practice

Study hub	GP practice	Frequency (%) *n* = 251
Wales Primary Care ('PiCRIS') Research Network	Ely Bridge Surgery	9 (3.6)
	Roath House Surgery	2 (0.8)
	Greenmeadow Surgery (CaeTeg)	13 (5.2)
	Clifton Surgery	1 (0.4)
	Vauxhall Surgery	1 (0.4)
	The Practice of Health	14 (5.6)
	Bellevue Surgery	1 (0.4)
	Ashgrove Surgery	1 (0.4)
	Stanwell Surgery	0 (0.0)
	Llan Healthcare	5 (2.0)
	Cynon Vale Medical Practice	1 (0.4)
	North Celynen	1 (0.4)
	Llandaff and Pentyrch Surgery	0 (0.0)
	Cwm Gwyrdd Medical Centre	0 (0.0)
	Penygraig Surgery	2 (0.8)
	Keir Hardie Health Park	0 (0.0)
	The Foundry Town Clinic	0 (0.0)
	Oaklands Surgery	0 (0.0)
	St John's Medical Practice	0 (0.0)
	Llandaff North Medical Practice	0 (0.0)
	Pont Newydd Medical Centre	0 (0.0)
	Malpas Brook Health Centre	1 (0.4)
	St David's Clinic	6 (2.4)
	Nantgarw Road Medical Centre	1 (0.4)
	Whitchurch Village Practice	1 (0.4)
	Rumney Primary Care Centre	4 (1.6)
	Llwyncelyn/Hollybush Practice	0 (0.0)
	St Pauls Clinic	0 (0.0)
	St Andrews Surgery	1 (0.4)
	Oaktree Surgery	2 (0.8)
	Subtotal	**67 (26.7)**
North East England and North Cumbria Research Network	Hub – Swarland Surgery	2 (0.8)
	Hub – Redburn Park Medical	0 (0.0)
	Spoke – 49 Marine Avenue	5 (2.0)

continued

TABLE 46 Recruitment by study hub and GP practice (*continued*)

Study hub	GP practice	Frequency (%) *n* = 251
	Spoke – Park Road Medical	3 (1.2)
	Spoke – Alnwick MG	0 (0.0)
	Spoke – Well Close	0 (0.0)
	Spoke – Belford	0 (0.0)
	Hub – Cheviot Medical Group	1 (0.4)
	Spoke – Glendale Medical	1 (0.4)
	Hub – Village Medical Group	0 (0.0)
	Hub – Corbridge	3 (1.2)
	Spoke – Burn Brae	1 (0.4)
	Spoke – Branch End	6 (2.4)
	Spoke – Humshaugh & Wark	1 (0.4)
	Spoke – White Medical Group	8 (3.2)
	Hub – Pelton & Fellrose	1 (0.4)
	Spoke – West Farm Surgery	0 (0.0)
	Spoke – Bellingham Practice	1 (0.4)
	Spoke – Sele Medical Practice	0 (0.0)
	Spoke – Prudhoe Medical Group	8 (3.2)
	Haydon Bridge and Allendale Medical Practice	1 (0.4)
	Freeman Hospital	22 (8.8)
	Subtotal	**64 (25.5)**
Western Research Network	Hub – Kingswood Health Centre	2 (0.8)
	Hub – Clevedon Medical	6 (2.4)
	Hub – Westbury-on-Trym	3 (1.2)
	Spoke – (Harbourside Family practice) Portishead Surgery	1 (0.4)
	Spoke – Greenway Community	0 (0.0)
	Spoke – Fallodon Way	0 (0.0)
	Spoke – Monks Park	0 (0.0)
	Hub – Tyntesfield Medical	43 (17.1)
	Hub – West Walk Surgery	29 (11.6)
	Spoke – Courtside surgery	0 (0.0)
	Spoke – Frome Valley Surgery	0 (0.0)
	Hub – Heart of Bath Medical	20 (8.0)
	Mendip Vale Medical Practice	8 (3.2)
	Hub – Fishponds Family Practice	1 (0.4)
	Spoke – Beechwood Medical Practice	1 (0.4)
	Nightingale Valley Practice	6 (2.4)
	Subtotal	**120 (47.8)**

TABLE 47 Participant demographics and clinical characteristics of the validation cohort for diagnosis of BOO

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		BOO *n* = 112	**No BOO** *n* = 125	**Indeterminate** *n* = 14	**Total** *n* = 251
Age (years)	Median (IQR)	69.0 (63.5–74.6)	67.5 (57.0–73.0)	61.2 (53.9–70.5)	67.6 (60.6–74.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Height (cm)	Median (IQR)	176.5 (171.0–180.0)	176.0 (172.0–182.0)	177.0 (172.0–183.0)	176.0 (172.0–181.0)
	Missing, *n* (%)	3 (1.8)	3 (2.4)	0 (0.0)	5 (2.0)
Weight (kg)	Median (IQR)	83.0 (76.0–94.0)	88.4 (78.0–104.4)	88.5 (78.0–98.0)	86.0 (76.7–97.5)
	Missing, *n* (%)	3 (2.7)	4 (3.2)	0 (0.0)	7 (2.8)
Ethnicity, *n* (%)	White	106 (94.6)	118 (94.4)	13 (92.9)	237 (94.4)
	Asian (Pakistani)	1 (0.9)	0 (0.0)	1 (7.1)	2 (0.8)
	Asian (Chinese)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Asian (Indian)	0 (0.0)	3 (2.4)	0 (0.0)	3 (1.2)
	Asian (Other)	1 (0.9)	0 (0.0)	0 (0.0)	1 (0.4)
	Black (Caribbean)	0 (0.0)	1 (0.8)	0 (0.0)	1 (0.4)
	Black (African)	2 (1.8)	0 (0.0)	0 (0.0)	2 (0.8)
	Black (Other)	0 (0.0)	1 (0.8)	0 (0.0)	1 (0.4)
	Other	1 (0.9)	1 (0.8)	0 (0.0)	2 (0.8)
	Missing	1 (0.9)	1 (0.8)	0 (0.0)	2 (0.8)
Smoking status, *n* (%)	Never	59 (52.7)	66 (52.8)	7 (50.0)	132 (52.6)
	Current	3 (2.7)	11 (8.8)	3 (21.4)	17 (6.8)
	Ex-smoker	48 (42.9)	46 (36.8)	4 (28.6)	98 (39.0)
	Missing	2 (1.8)	2 (1.6)	0 (0.0)	4 (1.6)
Medical history, *n* (%)	None	28 (25.0)	38 (30.4)	4 (28.6)	70 (27.9)
	Adrenal insufficiency	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Cerebrovascular accident	6 (5.4)	10 (8.0)	1 (7.1)	17 (6.8)
	Chronic heart failure	4 (3.6)	1 (0.8)	1 (7.1)	6 (2.4)
	Chronic venous stasis	1 (0.9)	1 (0.8)	0 (0.0)	2 (0.8)
	Dependent oedema	1 (0.9)	1 (0.8)	0 (0.0)	2 (0.8)
	Diabetes insipidus	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus II	10 (8.9)	16 (12.8)	2 (14.3)	28 (11.2)

continued

TABLE 47 Participant demographics and clinical characteristics of the validation cohort for diagnosis of BOO (*continued*)

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		BOO *n* = 112	No BOO *n* = 125	Indeterminate *n* = 14	Total *n* = 251
	Hypercalcaemia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Liver failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Obstructive sleep apnoea	6 (5.4)	6 (4.8)	0 (0.0)	12 (4.8)
	Polyuric renal failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Pyelonephritis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Sickle cell anaemia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Other	77 (68.8)	81 (64.8)	9 (64.3)	167 (66.5)

IQR, interquartile range.

TABLE 48 Participant demographics and clinical characteristics of the validation cohort for diagnosis of DU

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		DU *n* = 87	No DU *n* = 152	Indeterminate *n* = 12	Total *n* = 251
Age (years)	Median (IQR)	69.2 (59.7–76.5)	67.9 (62.2–73.3)	60.7 (49.4–63.0)	67.6 (60.6–74.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Height (cm)	Median (IQR)	174.5 (171.0–180.0)	177.0 (172.0–181.0)	177.0 (174.0–183.0)	176.0 (172.0–181.0)
	Missing, *n* (%)	1 (1.2)	4 (2.6)	0 (0.0)	5 (2.0)
Weight (kg)	Median (IQR)	84.3 (75.0–95.0)	86.0 (77.0–102.5)	90.0 (81.8–101.9)	86.0 (76.7–97.5)
	Missing, *n* (%)	3 (3.5)	4 (2.6)	0 (0.0)	7 (2.8)
Ethnicity, *n* (%)	White	81 (93.1)	145 (95.4)	11 (91.7)	237 (94.4)
	Asian (Pakistani)	1 (1.2)	0 (0.0)	1 (8.3)	2 (0.8)
	Asian (Chinese)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Asian (Indian)	1 (1.2)	2 (1.3)	0 (0.0)	3 (1.2)
	Asian (Other)	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.4)
	Black (Caribbean)	1 (1.2)	0 (0.0)	0 (0.0)	1 (0.4)
	Black (African)	0 (0.0)	2 (1.3)	0 (0.0)	2 (0.8)
	Black (Other)	1 (1.2)	0 (0.0)	0 (0.0)	1 (0.4)
	Other	2 (2.3)	0 (0.0)	0 (0.0)	2 (0.8)
	Missing	0 (0.0)	2 (1.3)	0 (0.0)	2 (0.8)
Smoking status, *n* (%)	Never	48 (55.2)	78 (51.3)	6 (50.0)	132 (52.6)
	Current	4 (4.6)	10 (6.6)	3 (25.0)	17 (6.8)
	Ex-smoker	35 (40.2)	60 (39.5)	3 (25.0)	98 (39.0)
	Missing	0 (0.0)	4 (2.6)	0 (0.0)	4 (1.6)

TABLE 48 Participant demographics and clinical characteristics of the validation cohort for diagnosis of DU (*continued*)

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		DU *n* = 87	No DU *n* = 152	Indeterminate *n* = 12	Total *n* = 251
Medical history, *n* (%)	None	22 (25.3)	44 (29.0)	4 (33.3)	70 (27.9)
	Adrenal insufficiency	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Cerebrovascular accident	5 (5.8)	11 (7.2)	1 (8.3)	17 (6.8)
	Chronic heart failure	3 (3.5)	2 (1.3)	1 (8.3)	6 (2.4)
	Chronic venous stasis	1 (1.2)	1 (0.7)	0 (0.0)	2 (0.8)
	Dependent oedema	2 (2.3)	0 (0.0)	0 (0.0)	2 (0.8)
	Diabetes insipidus	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus II	11 (12.6)	15 (9.9)	2 (16.7)	28 (11.2)
	Hypercalcaemia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Liver failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Obstructive sleep apnoea	4 (4.6)	8 (5.3)	0 (0.0)	12 (4.8)
	Polyuric renal failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Pyelonephritis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Sickle cell anaemia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Other	61 (70.1)	99 (65.1)	7 (58.3)	167 (66.5)

IQR, interquartile range.

TABLE 49 Participant demographics and clinical characteristics of the validation cohort for diagnosis of DO

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		DO *n* = 166	No DO *n* = 78	Indeterminate *n* = 7	Total *n* = 251
Age (years)	Median (IQR)	68.7 (62.4–74.7)	65.5 (56.4–71.5)	60.7 (59.2–63.1)	67.6 (60.6–74.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Height (cm)	Median (IQR)	177.0 (172.0–181.0)	176.0 (171.0–180.0)	176.0 (172.0–183.0)	176.0 (172.0–181.0)
	Missing, *n* (%)	3 (1.8)	2 (2.6)	0 (0.0)	5 (2.0)
Weight (kg)	Median (IQR)	86.1 (76.3–99.8)	83.6 (77.0–94.0)	91.0 (69.8–98.0)	86.0 (76.7–97.5)
	Missing, *n* (%)	4 (2.4)	3 (3.9)	0 (0.0)	7 (2.8)
Ethnicity, *n* (%)	White	158 (95.2)	72 (92.3)	7 (100.0)	237 (94.4)

continued

TABLE 49 Participant demographics and clinical characteristics of the validation cohort for diagnosis of DO (*continued*)

Participant characteristic		Diagnosis based on reference standard (urodynamics)			
		DO *n* = 166	No DO *n* = 78	Indeterminate *n* = 7	Total *n* = 251
	Asian (Pakistani)	0 (0.0)	2 (2.6)	0 (0.0)	2 (0.8)
	Asian (Chinese)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Asian (Indian)	1 (0.6)	2 (2.6)	0 (0.0)	3 (1.2)
	Asian (Other)	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.4)
	Black (Caribbean)	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.4)
	Black (African)	2 (1.2)	0 (0.0)	0 (0.0)	2 (0.8)
	Black (Other)	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.4)
	Other	0 (0.0)	2 (2.6)	0 (0.0)	2 (0.8)
	Missing	2 (1.2)	0 (0.0)	0 (0.0)	2 (0.8)
Smoking status, *n* (%)	Never	89 (53.6)	39 (50.0)	4 (57.1)	132 (52.6)
	Current	9 (5.4)	7 (9.0)	1 (14.3)	17 (6.8)
	Ex-smoker	66 (39.8)	30 (38.5)	2 (28.6)	98 (39.0)
	Missing	2 (1.2)	2 (2.6)	0 (0.0)	4 (1.6)
Medical history, *n* (%)	None	43 (25.9)	26 (33.3)	1 (14.3)	70 (27.9)
	Adrenal insufficiency	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Cerebrovascular accident	16 (9.6)	1 (1.3)	0 (0.0)	17 (6.8)
	Chronic heart failure	5 (3.0)	1 (1.3)	0 (0.0)	6 (2.4)
	Chronic venous stasis	2 (1.2)	0 (0.0)	0 (0.0)	2 (0.8)
	Dependent oedema	0 (0.0)	2 (2.6)	0 (0.0)	2 (0.8)
	Diabetes insipidus	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Diabetes mellitus II	20 (12.1)	7 (9.0)	1 (14.3)	28 (11.2)
	Hypercalcaemia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Liver failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Obstructive sleep apnoea	9 (5.4)	3 (3.9)	0 (0.0)	12 (4.8)
	Polyuric renal failure	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Pyelonephritis	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Sickle cell anaemia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Other	113 (68.1)	49 (62.8)	5 (71.4)	167 (66.5)

IQR, interquartile range.

TABLE 50 Participant treatment history based on diagnosis of BOO for the validation cohort

Treatment history, *n* (%)		Diagnosis			
		BOO *n* = 112	No BOO *n* = 125	Indeterminate *n* = 14	Total *n* = 251
None		16 (14.3)	21 (16.8)	3 (21.4)	40 (15.9)
Uroselective medications	Alpha-blockers	39 (34.8)	42 (33.6)	2 (14.3)	83 (33.1)
	5-alpha reductase inhibitors	15 (13.4)	14 (11.2)	1 (7.1)	30 (12.0)
	Anticholinergics	6 (5.4)	9 (7.2)	2 (14.3)	17 (6.8)
	Beta-3-agonists	2 (1.8)	8 (6.4)	0 (0.0)	10 (4.0)
Calcium channel blockers		24 (21.4)	25 (20.0)	1 (7.1)	50 (19.9)
Diuretics	Thiazides and related diuretics	4 (3.6)	5 (4.0)	0 (0.0)	9 (3.6)
	Loop diuretics	2 (1.8)	4 (3.2)	1 (7.1)	7 (2.8)
SSRI antidepressants		8 (7.1)	11 (8.8)	1 (7.1)	20 (8.0)
Other		82 (73.2)	87 (69.6)	10 (71.4)	179 (71.3)

TABLE 51 Participant treatment history based on diagnosis of DU for the validation cohort

Treatment history, *n* (%)		Diagnosis			
		DU *n* = 87	No DU *n* = 152	Indeterminate *n* = 12	Total *n* = 251
None		13 (14.9)	25 (16.4)	2 (16.7)	40 (15.9)
Uroselective medications	Alpha-blockers	31 (35.6)	51 (33.6)	1 (8.3)	83 (33.1)
	5-alpha reductase inhibitors	8 (9.2)	21 (13.8)	1 (8.3)	30 (12.0)
	Anticholinergics	6 (6.9)	9 (5.9)	2 (16.7)	17 (6.8)
	Beta-3-agonists	6 (6.9)	4 (2.6)	0 (0.0)	10 (4.0)
Calcium channel blockers		18 (20.7)	31 (20.4)	1 (8.3)	50 (19.9)
Diuretics	Thiazides and related diuretics	4 (4.6)	5 (3.3)	0 (0.0)	9 (3.6)
	Loop diuretics	3 (3.4)	3 (2.0)	1 (8.3)	7 (2.8)
SSRIs antidepressants		8 (9.2)	11 (7.2)	1 (8.3)	20 (8.0)
Other		67 (77.0)	103 (67.8)	9 (75.0)	179 (71.3)

TABLE 52 Participant treatment history based on diagnosis of DO for the validation cohort

Treatment history, *n* (%)		Diagnosis			
		DO *n* = 166	No DO *n* = 78	Indeterminate *n* = 7	Total *n* = 251
None		20 (12.0)	20 (25.6)	0 (0.0)	40 (15.9)
Uroselective medications	Alpha-blockers	56 (33.7)	26 (33.3)	1 (14.3)	83 (33.1)
	5-alpha reductase inhibitors	19 (11.4)	10 (12.8)	1 (14.3)	30 (12.0)
	Anticholinergics	9 (5.4)	5 (6.4)	3 (42.9)	17 (6.8)
	Beta-3-agonists	6 (3.6)	4 (5.1)	0 (0.0)	10 (4.0)
Calcium channel blockers		33 (19.9)	16 (20.5)	1 (14.3)	50 (19.9)
Diuretics	Thiazides and related diuretics	9 (5.4)	0 (0.0)	0 (0.0)	9 (3.6)
	Loop diuretics	5 (3.0)	2 (2.6)	0 (0.0)	7 (2.8)
SSRI antidepressants		14 (8.4)	4 (5.1)	2 (28.6)	20 (8.0)
Other		124 (74.7)	50 (64.1)	5 (71.4)	179 (71.3)

Appendix 5

International Prostate Symptom Score (IPSS)

Over the past month, how often have you experienced...	Not at all	Less than 1 time in 5	Less than half the time	About half the time	More than half the time	Almost always	Your score
Incomplete emptying	0	1	2	3	4	5	
Frequency	0	1	2	3	4	5	
Intermittency	0	1	2	3	4	5	
Urgency	0	1	2	3	4	5	
Weak stream	0	1	2	3	4	5	
Straining	0	1	2	3	4	5	

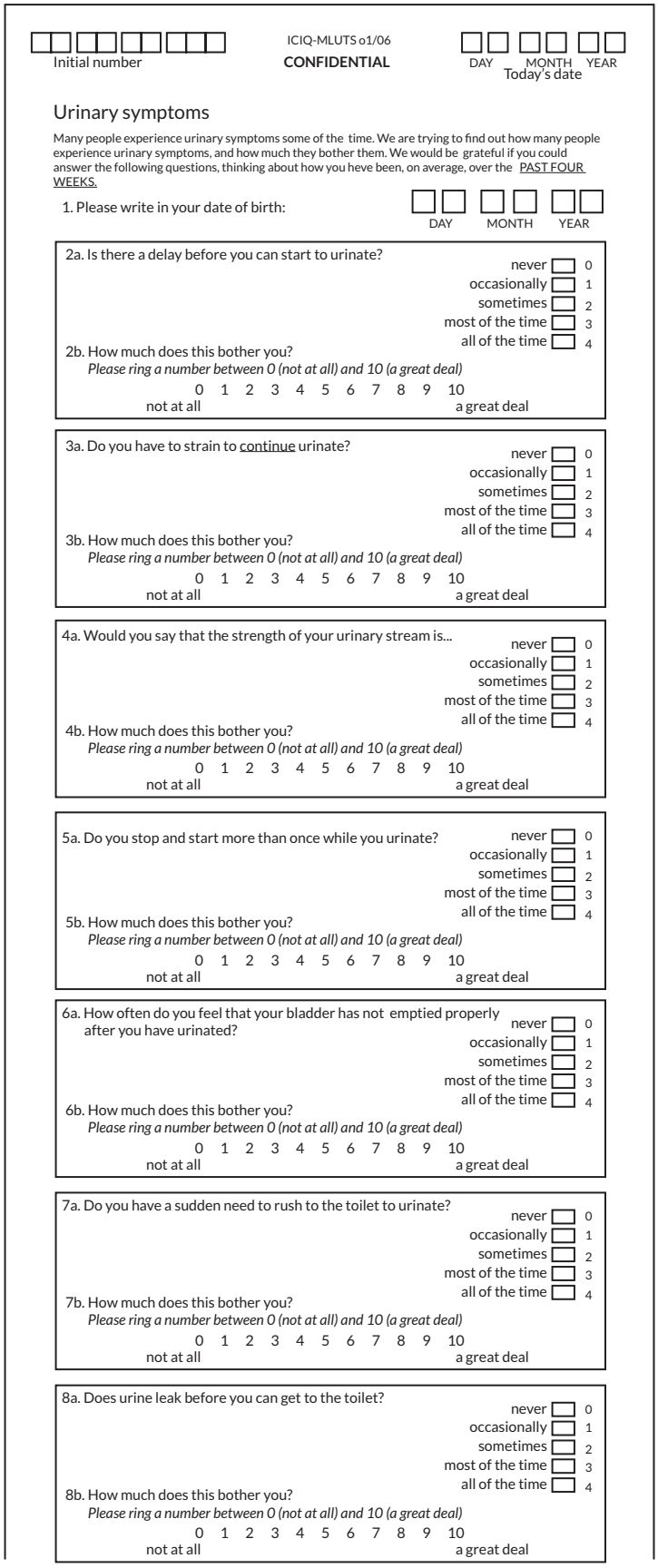
Nocturia	None	1 time	2 times	3 times	4 times	5 times or more	**Your score**
Over the past month, how many times did you most typically get up to urinate from the time you went to bed to the time you got up?	0	1	2	3	4	5	

Total IPSS score	

Quality of life due to urinary symptoms	Delighted	Pleased	Mostly satisfied	Equally satisfied and dissatisfied	Mostly dissatisfied	Unhappy	Terrible
If you were to spend the rest of your life with your urinary condition the way it is now, how would you feel?	0	1	2	3	4	5	6

Total score: 0—7 mildly symptomatic; 8—19 moderately symptomatic; 20—35 severely symptomatic.

FIGURE 36 International Prostate Symptom Score Scale.



Initial number

ICIQ-MLUTS o1/06

CONFIDENTIAL

DAY MONTH YEAR
Today's date

Urinary symptoms

Many people experience urinary symptoms some of the time. We are trying to find out how many people experience urinary symptoms, and how much they bother them. We would be grateful if you could answer the following questions, thinking about how you heve been, on average, over the PAST FOUR WEEKS.

1. Please write in your date of birth: DAY MONTH YEAR

2a. Is there a delay before you can start to urinate?
never 0
occasionally 1
sometimes 2
most of the time 3
all of the time 4

2b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)
0 1 2 3 4 5 6 7 8 9 10
not at all — a great deal

3a. Do you have to strain to continue urinate?
never 0
occasionally 1
sometimes 2
most of the time 3
all of the time 4

3b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)
0 1 2 3 4 5 6 7 8 9 10
not at all — a great deal

4a. Would you say that the strength of your urinary stream is...
never 0
occasionally 1
sometimes 2
most of the time 3
all of the time 4

4b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)
0 1 2 3 4 5 6 7 8 9 10
not at all — a great deal

5a. Do you stop and start more than once while you urinate?
never 0
occasionally 1
sometimes 2
most of the time 3
all of the time 4

5b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)
0 1 2 3 4 5 6 7 8 9 10
not at all — a great deal

6a. How often do you feel that your bladder has not emptied properly after you have urinated?
never 0
occasionally 1
sometimes 2
most of the time 3
all of the time 4

6b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)
0 1 2 3 4 5 6 7 8 9 10
not at all — a great deal

7a. Do you have a sudden need to rush to the toilet to urinate?
never 0
occasionally 1
sometimes 2
most of the time 3
all of the time 4

7b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)
0 1 2 3 4 5 6 7 8 9 10
not at all — a great deal

8a. Does urine leak before you can get to the toilet?
never 0
occasionally 1
sometimes 2
most of the time 3
all of the time 4

8b. How much does this bother you?
Please ring a number between 0 (not at all) and 10 (a great deal)
0 1 2 3 4 5 6 7 8 9 10
not at all — a great deal

FIGURE 37 International Consultation on Incontinence Questionnaire MLUTS.

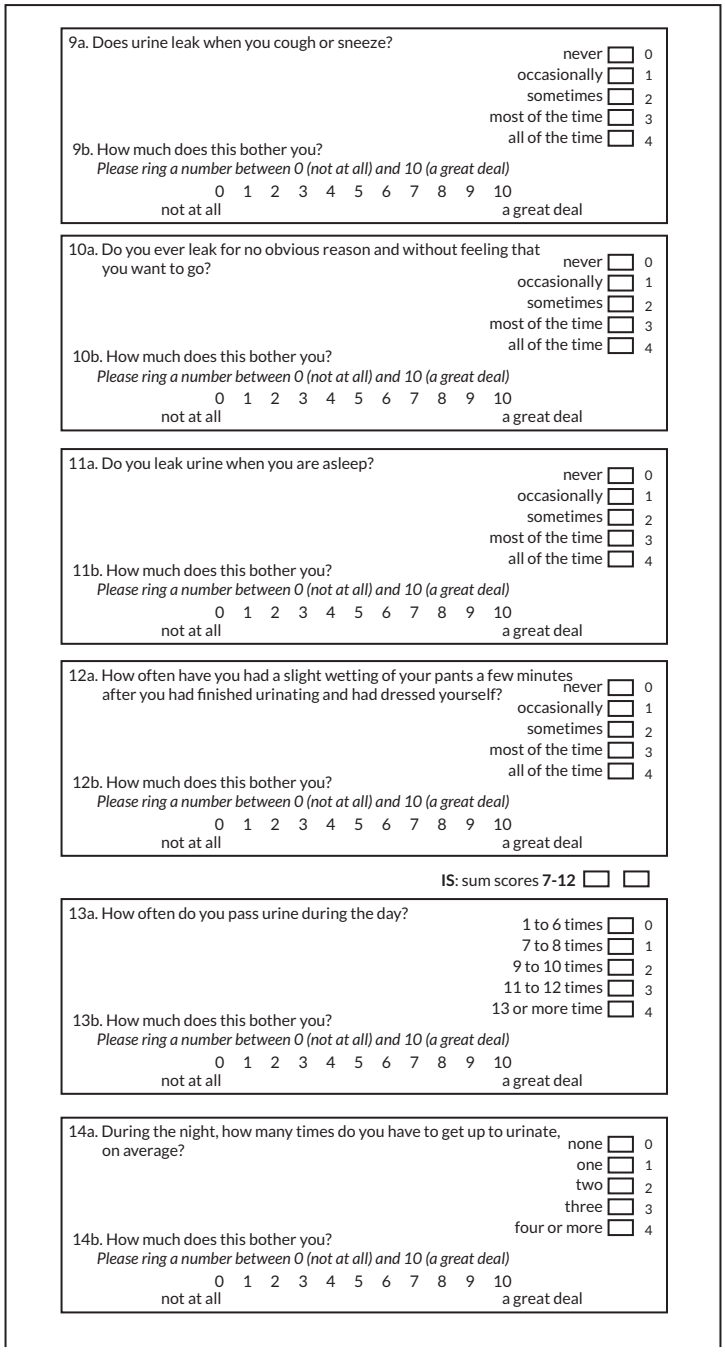


FIGURE 37 International Consultation on Incontinence Questionnaire MLUTS. Continued

Appendix 6 Protocol changes

TABLE 53 Protocol changes

Amendment	Detail	Protocol version and date
Substantial Amendment 1 **3 October 2017**	• Tightening up of screening and consent process (mandating three index tests in GP screening visit to ensure eligibility of patients) • Addition of patient-facing materials • Amendments to patient-facing material based on PPI feedback • Amendment and clarification to the wording in the exclusion criteria, including exclusion of men with any contraindications to urodynamics • Additional safety time point (RN call to patient 3 days following the urodynamic procedure) • Addition of IPSS questionnaire • Amended definition of adverse events, to only collect those adverse events related to the study • Amended training requirements for nurses	V2.0 11 September 2017
Substantial Amendment 2 **10 January 2018**	• Section 15. Safety Reporting (Adverse Events) - Study-Specific Adverse Events Updated to reflect PriMUS CRF - Expected events table updated • Typographical errors updated • Frequency Volume Chart changed to Bladder Diary (a version of frequency volume chart that collects more data) • Comment to include that urodynamic nurses will not be blinded to index tests, as is urodynamics not an isolated test in standard practice • Section 12. Withdrawal - Amended to more concise wording, to allow patients to withdraw consent of using data already collected	V3.0 20 December 2017
Non-Substantial Amendment 1 **R&D Approval Received 19 January 2018 HRA Approval Received 25 January 2018**	• Acceptability Interview Patient Information Sheet and Consent Form V1.1 15 January 2018 – Removal of sentence 'The lead person at your local hospital, also known as the local Principal Investigator is Dr. (PERSON) who can be contacted on (TELEPHONE)' • Potential Participant Letter V1.1 15 January 2018 – Removal of 'your GP surgery to make an appointment with your doctor. Please inform the receptionist that you would like to discuss taking part in the PriMUS study' replaced with 'the research nurse on <Tel> or complete response slip'	V3.0 20 December 2017
Non-Substantial Amendment 2 **R&D Approval Received 19 February 2018 HRA Approval Received 20 February 2018**	• Addition of two English Sites: 1. All GP Practices within North East and North Cumbria CRN 2. All GP Practices within NIHR West of England CRN	V3.0 20 December 2017
Non-Substantial Amendment 3 **R&D and HRA Approval Received 5 March 2018**	• Addition of one English Site: 1. North Bristol NHS Trust	V3.0 20 December 2017

TABLE 53 Protocol changes (*continued*)

Amendment	Detail	Protocol version and date
Substantial Amendment 3 20 April 2018	• Section 14.2 Qualitative - Amended consensus plans one-to-one interviews and clarified aims and processes within this, along with some updates to acceptability qualitative work • Eligibility Criteria: Eligibility Criteria updated to clarify for GPs – removal of renal impairment. Retention updated to 'retention e.g. palpable bladder after voiding' • Database search strategy for sites updated from additional three times to additional five times during the study to reflect service support costs • Section 7. Site and Investigator Selection - Documents required for a site to open to recruitment has been updated to reflect the site opening procedure - Confirmation of capacity and capability added (English sites) - GP agreement replaced study agreement - Nurse Delegation Log replaced roles and responsibilities • Section 15.5.2 Safety Reporting - Update terminology to reflect NIHR terminology for non-CTIMPS (USRE/SRE)	V4.0 9 April 2018
Non-Substantial Amendment 4 **R&D and HRA Approval Received 18 October 2018**	• Addition of one Wales Site: 1. Abertawe Bro Morgannwg University Health Board	V4.0 9 April 2018
Substantial Amendment 4 26 February 2019	• Changed wording of 'decision aid' to consistent referral of 'clinical decision support tool' • Amended name of PriMUS Administrator • Amended references to Data Protection Act to General Data Protection Regulations (GDPR) • Section 9. Patient Identification – added a sentence for practice GPs to use the same read codes to ensure patients can be identified in database searches • Section 9. Patient Identification – clarified the patient information in the mail-out packs • Section 9. Patient Identification – already trained site nurse practitioners can facilitate screening. They must be on the delegation log and be GCP trained • Section 10.4 Compliance – clarified the monitoring procedure during the pilot phase and the monitoring plans for the main recruitment phase • Section 11. Patient Pathway Flow chart updated to a simpler version, as more comprehensive flow charts are available in Research Nurse and GP Manuals • Section 11. Clarifications to patient pathway wording • Section 15. Safety adverse events and serious adverse events wording amended to be less ambiguous, with additional guidance on what constitutes an 'intervention' • Section 17.3 Amended title to Health Economic Modelling • Section 19. Added Substudies and provided information on SWAT • Section 26. Amended Milestones according to study timeline changes	V5.0 8 February 2019
Substantial Amendment 5 6 June 2019 **R&D and HRA Approval 6 June 2019**	• Section 14. Qualitative i. Development of Management Recommendations to inform the clinical decision support tool ◦ Due to the number of scenarios to work through. we have increased the urologist sample size, so the urologists will only be presented with half of the scenarios to work through. We have also included the option for urologists to complete an online survey instead of a telephone interview, which we hope urologists will be able to complete if they do not have time for a telephone interview ii. User-testing of the prototype clinical decision support tool (HCPs and patients) • Section 15. Safety Reporting – updated the SAE Flow chart and SAE definition table based on SAE Terminology changes in previous amendment	V6.0 13 May 2019

continued

TABLE 53 Protocol changes (*continued*)

Amendment	Detail	Protocol version and date
Substantial Amendment 6 6 September 2019 **R&D and HRA Approval 24 September 2019**	• Primary and Secondary Objectives and Measures – changes to the wording to match the Statistical Analysis Plan • Section 9.1 Participant identification – addition of identifying patients who have been referred to secondary care but have not yet been seen by a urologist • Section 14. Qualitative – user-testing of the prototype clinical decision support tool Patients In the previous amendment, this piece of work was changed to a workshop with study PPI representatives. However, we will need more participants to take part in the workshop to help facilitate and elicit deeper discussion. We are therefore going to recruit individuals approached to take part in the main study, to take part in the workshop (whether they enter the main study or not). PPI representatives will be workshop facilitators, along with members of the study team. • Addition of PriMUS PPI Workshop Consent to Contact Form V1.0 27 August 2019 • Addition of PI Workshop Information Sheet and Consent Form V1.0 27 August 2019	V7.0 27 August 2019
Non-Substantial Amendment 5 **HRA and R&D Approval Received 3 December 2019**	• Extend the end of recruitment date to 30 April 2020 (to reflect end of study date at the time)	V7.0 27 August 2019
Non-Substantial Amendment 6 **R&D and HRA Approval Received 5 February 2020**	• Addition of PICS	
Substantial Amendment 7 **R&D and HRA Approval** 23 March 2020	• Reduce sample size to 750 based on a lower-than-expected attrition rate • Secondary Objectives and Outcomes – removal of CPRD and percentage change objective/ outcome • Section 7 – the addition of Patient Identification Referral Centres • Section 13.1 – addition of new recruitment target graph • Section 14 Qualitative – removal of patient workshop for user-testing instead focusing on the completion of GP interviews • Section 17. Statistical Section – updates to the wording to match the Statistical Analysis Plan • Section 17.4 Referral Rates – no longer calculating the referral rates and comparing with CPRD • Section 27 Milestones – updated the milestones to reflect extension of the study	V8.0 10 March 2020
Non-Substantial Amendment 7 **Sponsor Approval Required Only 2 April 2020**	• Officially Pause Recruitment in PriMUS due to COVID-19	V8.0 10 March 2020
Substantial Amendment 8 25 August 2020 **R&D and HRA Approval 25 August 2020**	• COVID-19 adaptations to study processes • Addition of medication review for recruitment pathways • Consent and Study Visit Part A can take place remotely	V9.0 6 July 2020
Non-sub 9 (14 January 2021)	• Change of PI at an existing site	N/A
NSA 10 (19 May 2021)	• Extension of study end date to 30 November 2021 from 31 May 2021	N/A
NSA 11 (25 June 2021)	• Change of PI at Bellevue	N/A
NSA 12 (5 November 2021)	• Extension of study end date to 30 November 2022	N/A
NSA 13 (27 July 2022)	• Extension of study end date to 30 March 2023	N/A

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library