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

Scientific summary

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

Health Technology Assessment

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