Bladder ultrasonography for diagnosing detrusor overactivity: test accuracy study and economic evaluation

Suneetha Rachaneni,[1] Shanteela McCooty,[2] Lee J Middleton,[3] Victoria L Parker,[3] Jane P Daniels,[1,3*] Arri Coomarasamy,[1,2] Tina S Verghese,[1,2] Moji Balogun,[2] Ilias Goranitis,[4] Pelham Barton,[4] Tracy E Roberts,[4] Jonathan J Deeks[3,5] and Pallavi Latthe[1,2] on behalf of the Bladder Ultrasound Study (BUS) Collaborative Group

[1]School of Clinical and Experimental Medicine, University of Birmingham, Birmingham, UK
[2]Birmingham Women's NHS Foundation Trust, Birmingham, UK
[3]Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK
[4]Health Economics Unit, University of Birmingham, Birmingham, UK
[5]Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK

*Corresponding author

Declared competing interests of authors: Dr Latthe has received financial support from Pfizer and Astellas to attend international urogynaecology conferences.

Published January 2016
DOI: 10.3310/hta20070

Scientific summary

Bladder ultrasonography for diagnosing detrusor overactivity

Health Technology Assessment 2016; Vol. 20: No. 7
DOI: 10.3310/hta20070

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Overactive bladder (OAB) is a symptom complex of urinary urgency (intense, sudden desire to void) with or without incontinence, increased urinary frequency or nocturia in the absence of infection or other proven pathology. Detrusor overactivity (DO) is involuntary detrusor contractions associated with urgency observed during the filling phase of a bladder test called urodynamics (UDS). The pathology behind OAB symptoms may be DO in 54–58% of patients.

Urodynamics has been considered the gold standard test for DO, but is intimate and invasive, with a risk of urinary infection. Transvaginal ultrasonography to measure bladder wall thickness (BWT) has been proposed as a potentially less invasive method of diagnosis of DO.

Objectives

The original primary research objective was to estimate the diagnostic accuracy of BWT measured by bladder ultrasonography, in the diagnosis of DO.

The original secondary research objectives were:

1. to conduct a decision-analytical model-based economic evaluation comparing the cost-effectiveness of various care pathways (including pathways that incorporate bladder ultrasonography)
2. to investigate the acceptability of both tests
3. to assess whether or not measurements of BWT made using transvaginal ultrasonography have adequate reliability and reproducibility to be likely to detect differences potentially indicative of disease.

We also aimed to investigate the value added by bladder ultrasonography to information already obtained from routine initial non-invasive tests (e.g. history, bladder diary, disease-specific quality-of-life questionnaire).

Subsequently, a fifth objective was added to the Bladder Ultrasound Study (BUS), namely to establish the role of UDS and its impact on treatment and patient outcomes in OAB and mixed urinary incontinence (MUI). There were six key questions:

1. Does the urodynamic diagnosis affect treatment pathways?
2. What were the patient-reported outcomes in the cohort of women recruited into the BUS 6 and 12 months after testing?
3. Can diagnoses by UDS predict improvement in different patient groups?
4. Does receiving treatment concordant with the urodynamic diagnoses improve patients' symptoms, compared with not receiving a concordant treatment?
5. Are presenting symptoms related to outcomes after 6 and 12 months?
6. Does ultrasonographic measurement of BWT have any prognostic value?
7. What is the cost-effectiveness of UDS in the diagnosis of DO?

Methods

A cross-sectional test accuracy study recruited 687 women with OAB symptoms from 22 UK hospitals. BWT was measured using transvaginal ultrasonography. DO status was determined by multichannel UDS, undertaken blind to the ultrasonography findings. Test accuracy was estimated by comparing BWT measured by bladder ultrasonography (the index test) with diagnosis of DO (the target condition) obtained from UDS (the reference standard). The primary analysis involved calculations of sensitivity, specificity, predictive values and likelihood ratios (LRs) using a BWT of 5 mm as a cut-off point ($\geq$ 5 mm indicating presence of detrusor activity). A receiver operating characteristic (ROC) curve was constructed and the area under the curve (AUC) computed [with 95% confidence interval (CI)] to give an overall estimate of ultrasonography accuracy across all thresholds of BWT. Statistical significance was tested by comparing with the uninformative model using a non-parametric approach. Multiple sensitivity analyses were performed to test the robustness of the results to protocol deviations, missing data and different subpopulations based on clinical presentation.

Intraobserver and interobserver reproducibility of BWT measurements were assessed by repeated measurement from scans made in subsets of 37 and 57 women, respectively, and by repeated scans in 27 women. The intraclass correlations and smallest real difference were derived using one-way analysis of variance.

The acceptability of transvaginal bladder ultrasonography and UDS from the patient's perspective was evaluated through the completion of a self-reported questionnaire containing visual analogue scales for pain, ordinal scales for acceptability and a generic state anxiety measure. Mean differences and 95% CIs were calculated with statistical significance determined by a paired *t*-test for pain and anxiety scores. Wilcoxon signed-rank test was used for acceptability responses and McNemar's test for binary responses.

Patient symptoms were measured before testing and after 6 and 12 months, using the validated International Consultation on Incontinence modular Questionnaire Overactive Bladder (short form) (ICIQ-OAB) questionnaire and a global impression of improvement elicited at 12 months. The relationship between UDS diagnosis and subsequent treatment was examined using a multinomial logistic regression model. The overall importance of this variable was determined by chi-squared test, with results presented alongside estimates of odds ratios (ORs) and 95% CI, with no treatment used as the reference variable.

Frequencies and percentages are presented for the results of the global impression of improvement question measured at follow-up. Mean change of ICIQ-OAB scores from baseline and 95% CIs were calculated, with a paired *t*-test used to test statistical significance of the change. Further statistical analysis was completed using logistic and linear repeated measures regression models.

In the economic evaluation, three diagnostic pathways were compared. The first was based on UDS and represented the way treatment pathways are determined in current practice. The second was based on the results of the bladder ultrasound, and the third on the clinical history. In a secondary analysis, strategies using a diagnostic test as an adjunct to clinical history were additionally explored.

A decision-analytic model was constructed to estimate the cost-effectiveness of the different diagnostic strategies. The analyses were carried out from the perspective of the UK NHS and the primary outcome was in terms of cost per woman successfully treated. Secondary outcomes included in the analyses were cost per quality-adjusted life-year (QALY) and cost per DO case detected. A 5-year time-horizon was selected and costs and QALYs accruing beyond 12 months were discounted at an annual rate of 3.5%. Results were presented in terms of incremental cost-effectiveness ratios (ICERs) and cost-effectiveness acceptability frontiers. Deterministic and probabilistic sensitivity analyses were performed to explore the effects of the inherent uncertainty in parameter estimates on model results.

Results

Main findings of test accuracy and reproducibility study

The mean age of the 687 women recruited to the study was 52.7 years [standard deviation (SD) 13.9 years] and the average body mass index was 30.6 kg/m^2 (SD 12.2 kg/m^2). Fifty-six per cent (378/687) of the women were post menopausal. According to the clinical history, 61% (419/687) had urgency-predominant MUI and 33% (226/687) reported only urinary urgency along with increased frequency. The median duration of symptoms was 3.0 years [interquartile range (IQR) 1.6–7.0]. A total of 644 participants (94%) had both complete index and reference standard results.

Estimation of the accuracy of BWT showed very low sensitivity, specificity and LRs at all pre-specified cut-off points. The ROC curve showed no evidence of discrimination at any threshold between those with and without DO ($p = 0.25$); the AUC was 0.53 (95% CI 0.48 to 0.57). Furthermore, there was no evidence that the mean BWT measurements were any higher in the detrusor overactive group than the no overactivity group [4.85 mm (SD 1.36 mm) vs. 4.70 mm (SD 1.29 mm); $p = 0.19$] or that it had any relationship with ICIQ-OAB symptoms score when measured at presentation ($r = –0.01$; $p = 0.88$). Extensive sensitivity analyses and subgroup analyses were carried out, including exclusion of those with a history of mixed incontinence and those with 'dry' OAB, but did not alter the interpretation of these findings.

Analyses of intraoperator and interoperator variability found that differences of less than 2 mm in BWT cannot be safely interpreted as indicating real differences, as such differences are in the realms of those attributable to analytical variability (measurement error). We observed that the process of interpreting scans introduces a measurement error of around 1 mm, suggesting that the remaining 1 mm is attributable to a combination of the scanning process and biological variability.

Main findings of acceptability

A total of 646 (94%) participants in the study responded to the acceptability questionnaire following both tests. Pain levels following both tests appeared relatively low, although scores during and shortly after UDS were higher than the corresponding scores during and after bladder ultrasonography, and these differences were statistically significant. The proportion of women who found the test 'totally acceptable' was significantly higher with ultrasonography than UDS (81% vs. 56%; $p < 0.001$). Fewer women felt that they would recommend UDS to a friend than ultrasonography (86% vs. 96%; $p < 0.001$) and have the same test again (88% vs. 97%; $p < 0.001$). Anxiety levels associated with both tests appeared moderate (12.6 for ultrasonography and 12.9 for UDS), although the scores were only slightly higher with UDS (0.3 points difference on a 4- to 24-point scale, 95% CI 0.1 to 0.5; $p = 0.02$).

Main findings of long-term follow-up after 6 and 12 months

The question about value added by BWT on transvaginal ultrasonography became redundant as its diagnostic accuracy was found to be poor. Follow-up data were available for 489 (71%) and 475 (69%) participants, at a median time of 7 and 20 months (IQR 6–8 months and 15–24 months, respectively) post-testing, respectively. Over the whole follow-up period, the majority of women reported some treatment (292/467 providing information, 63%). Overall, subsequent treatment was highly associated with diagnosis group ($p < 0.0001$) suggesting that the clinicians and patients appeared to be guided by diagnoses from UDS in selecting treatment options. A total of 53% of the participants (248/470 providing information) thought that their bladder problems had improved at 20 months. There was no evidence that BWT had any relationship with the global impression of improvement responses at 20 months ($p = 0.4$) or ICIQ-OAB scores ($p = 0.8$) over 7 and 20 months (correlation coefficients Pearson's $r = 0.01$ at both time points). There was some evidence that ICIQ-OAB responses varied between diagnosis groups overall ($p = 0.02$) and pairwise comparisons between them indicated that the DO with urodynamic stress incontinence group have a greater reduction than the DO group (–1.1 points, 95% CI –1.7 to –0.4; $p = 0.002$) over both time points. At 20 months, 57% (168/296) of patients who had received a treatment concordant with their UDS diagnosis responded positively to the global impression of improvement question compared with 45% (69/152) of patients who had not (OR 1.6, 95% CI 1.1 to 2.3; $p = 0.02$).

Main findings of economic evaluation

Following National Institute for Health and Care Excellence recommendations on the use of UDS if conservative treatments have not been effective, the economic evaluation focused on women who were receiving antimuscarinic therapy before enrolling into the study. In the primary analysis, treatment based on clinical history was found to be more effective than UDS, leading to an additional 26 cases per 10,000 women successfully treated and 0.02 QALYs gained per woman. These came at an additional cost of approximately £1300 per woman. Bladder ultrasonography was more costly and less effective than the other strategies. The ICER of clinical history compared with UDS was estimated at £491,100 per woman successfully treated and £60,200 per QALY. In the secondary analysis, the strategy of giving UDS to women with a clinical history of MUI was found to be more effective than universal UDS. The selective use was shown to result in an additional 309 cases per 10,000 women successfully treated and 0.05 QALYs gained per woman at an additional cost of £600 per woman. This led to an ICER of £19,500 per woman successfully treated and £12,700 per QALY. In terms of DO cases detected, UDS was the most cost-effective diagnostic strategy in both analyses.

Conclusion

There was no evidence that transvaginal BWT had any relationship with DO, regardless of cut-off point. Extensive sensitivity analyses and subgroup analyses were carried out without any evidence of an increase in the performance. Furthermore, BWT had no relationship to symptoms as measured by ICIQ-OAB score either on presentation or in the long term. BWT thus has no predictive or prognostic value as a test in this condition.

We undertook three separate studies to investigate the intraobserver and interobserver variation of BWT using transvaginal ultrasonography and concluded that it was unlikely that this measurement would be sufficiently reliable or reproducible to be an accurate diagnostic test in routine clinical practice. Only differences > 2 mm could be safely interpreted as real change, meaning that for the vast majority of women (84%), there could be some possibility of misclassification when using a cut-off point of 5 mm.

Transvaginal ultrasonography was more acceptable as well as less painful than UDS. Despite this, a high proportion of women said that they would recommend the UDS test to a friend (88%) and also have it repeated (86%).

The urodynamic diagnosis appeared to have an affect on the subsequent treatment received by the patients who were followed up. These included increased chance of having surgical treatment (e.g. 15 times greater odds of having surgery for urinary stress incontinence if given a combined diagnosis with DO than those with normal urodynamic results). Women diagnosed with DO were three times more likely to have reported taking bladder relaxants in the 2 years post-test. The long-term follow-up also found that patients were more likely to report improvement in symptoms if they received a medical or surgical treatment in concordance with their urodynamic diagnoses. This suggests UDS did play a part in the subsequent treatment of women with suspected OAB.

In the economic evaluation, UDS was found to be a cost-effective diagnostic strategy. Carrying out UDS on women with predominant symptoms of OAB once conservative treatment strategies were exhausted resulted in significant cost-savings for a small reduction in effectiveness. A further investigation into whether or not a diagnostic test would be more cost-effective if performed only in specific subgroups of women concluded that the most cost-effective diagnostic strategy is to perform UDS in women with a clinical history of MUI only.

Implications for health care

Transvaginal ultrasonographic measurement of BWT is not an accurate method of diagnosing DO. UDS, the gold standard test for lower urinary tract conditions, was found to be the most cost-effective test in the management of OAB but especially in the subgroup of MUI once the conservative treatments were exhausted. Offering UDS earlier on in the management of mixed incontinence may increase patient satisfaction and save NHS resource.

Recommendations for research

In women with OAB or urgency-predominant MUI, adequately powered randomised controlled trials comparing treatment based on urodynamic diagnoses compared with treatment based on clinical assessment (history and examination alone) and related economic evaluation for these diagnostic interventions are required to consolidate the role of UDS in the management of these women, as has been done for stress urinary incontinence.

Trial registration

This trial is registered as ISRCTN46820623.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index.

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

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports 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

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

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.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 09/22/122. The contractual start date was in January 2011. The draft report began editorial review in October 2014 and was accepted for publication in April 2015. 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' report 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 report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).