HEALTH TECHNOLOGY ASSESSMENT

VOLUME 20 ISSUE 7 JANUARY 2016
ISSN 1366-5278

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

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

National Institute for Health Research

DOI 10.3310/hta20070

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

This report should be referenced as follows:

Rachaneni S, McCooty S, Middleton LJ, Parker VL, Daniels JP, Coomarasamy A, *et al*. Bladder ultrasonography for diagnosing detrusor overactivity: test accuracy study and economic evaluation. *Health Technol Assess* 2016;**20**(7).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

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).

Abstract

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 j.p.daniels@bham.ac.uk

Background: Urodynamics (UDS) has been considered the gold standard test for detrusor overactivity (DO) in women with an overactive bladder (OAB). Bladder ultrasonography to measure bladder wall thickness (BWT) is less invasive and has been proposed as an alternative test.

Objectives: To estimate the reliability, reproducibility, accuracy and acceptability of BWT in women with OAB, measured by ultrasonography, in the diagnosis of DO; to explore the role of UDS and its impact on treatment outcomes; and to conduct an economic evaluation of alternative care pathways.

Design: A cross-sectional test accuracy study.

Setting: 22 UK hospitals.

Participants: 687 women with OAB.

Methods: BWT was measured using transvaginal ultrasonography, and DO was assessed using UDS, which was performed blind to ultrasonographic findings. Intraobserver and interobserver reproducibility were assessed by repeated measurements from scans in 37 and 57 women, respectively, and by repeated scans in 27 women. Sensitivity and specificity were computed at pre-specified thresholds. The smallest real differences detectable of BWT were estimated using one-way analysis of variance. The pain and acceptability of both tests were evaluated by a questionnaire. Patient symptoms were measured before testing and after 6 and 12 months using the International Consultation on Incontinence modular Questionnaire Overactive Bladder (short form) (ICIQ-OAB) questionnaire and a global impression of improvement elicited at 12 months. Interventions and patient outcomes were analysed according to urodynamic diagnoses and BWT measurements. A decision-analytic model compared the cost-effectiveness of care strategies using UDS, ultrasonography or clinical history, estimating the cost per woman successfully treated and the cost per quality-adjusted life-year (QALY).

Results: BWT showed very low sensitivity and specificity at all pre-specified cut-off points, and there was no evidence of discrimination at any threshold ($p = 0.25$). Extensive sensitivity and subgroup analyses did not alter the interpretation of these findings. The smallest detectable difference in BWT was estimated to be 2 mm. Pain levels following both tests appeared relatively low. The proportion of women who found the test 'totally acceptable' was significantly higher with ultrasonography than UDS (81% vs. 56%; $p < 0.001$). Overall, subsequent treatment was highly associated with urodynamic diagnosis ($p < 0.0001$). There was no evidence that BWT had any relationship with the global impression of improvement responses at 20 months ($p = 0.4$). Bladder ultrasonography was more costly and less effective than the other strategies. The incremental cost-effectiveness ratio (ICER) of basing treatment on the primary clinical presentation compared with UDS was £491,500 per woman successfully treated and £60,200 per QALY. Performing a UDS in those women with a clinical history of mixed urinary incontinence had an ICER of £19,500 per woman successfully treated and £12,700 per QALY compared with the provision of urodynamic to all women. For DO cases detected, UDS was the most cost-effective strategy.

Conclusion: There was no evidence that BWT had any relationship with DO, regardless of the cut-off point, nor any relationship to symptoms as measured by the ICIQ-OAB. Bladder ultrasonography has no diagnostic or prognostic value as a test in this condition. Furthermore, despite its greater acceptability, BWT measurement was not sufficiently reliable or reproducible.

Trial registration: Current Controlled Trials ISRCTN46820623.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 20, No. 7. See the NIHR Journals Library website for further project information.

Contents

List of tables

List of figures

List of abbreviations

ANOVA	analysis of variance
AUC	area under the curve
BTX-A	botulinum toxin serotype A
BMI	body mass index
BUS	Bladder Ultrasound Study
BWT	bladder wall thickness
CEAF	cost-effectiveness acceptability frontier
CI	confidence interval
DICOM	Digital Imaging and Communications in Medicine
DMC	Data Monitoring Committee
DO	detrusor overactivity
DWT	detrusor wall thickness
EQ-5D	European Quality of Life-5 Dimensions
EVPI	expected value of perfect information
GBS	Group B *Streptococcus*
GP	general practitioner
GUP	good urodynamic practice
HRG	Healthcare Resource Group
HTA	Health Technology Assessment
ICC	intraclass correlation coefficient
ICECAP-A	Investigating Choice Experiments CAPability measure for Adults
ICER	incremental cost-effectiveness ratio
ICIQ	International Consultation on Incontinence modular Questionnaire
ICIQ-OAB	International Consultation on Incontinence modular Questionnaire Overactive Bladder (short form)
IQR	interquartile range
LR	likelihood ratio
LUTS	lower urinary tract symptoms
MESA	Medical, Epidemiological, and Social Aspects of Ageing
MUI	mixed urinary incontinence
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
OAB	overactive bladder
OR	odds ratio
PFMT	pelvic floor muscle training
PNE	peripheral nerve evaluation
POP	pelvic organ prolapse
PSA	probabilistic sensitivity analysis
PTNS	percutaneous tibial nerve stimulation
PVR	post-void residual
QALY	quality-adjusted life-year
QoL	quality of life
RCT	randomised controlled trial
ROC	receiver operating characteristic
SD	standard deviation
SNS	sacral nerve stimulation
SOP	standard operating procedure
STAI-6	State–Trait Anxiety Inventory six-item short form
SUI	stress urinary incontinence
TSC	Trial Steering Committee
UDS	urodynamics
UI	urinary incontinence
USI	urodynamic stress incontinence
UTI	urinary tract infection
VAS	visual analogue scale
VD	voiding dysfunction
WTP	willingness to pay

Plain English summary

Overactive bladder (OAB) is a distressing condition that affects about 12% of the population. The symptoms are the frequent and urgent desire to urinate, sometimes with loss of control of the bladder. The underlying cause may be uncontrollable contractions of the bladder wall called detrusor overactivity. These contractions can be observed during an intimate test called urodynamics (UDS), in which the bladder is artificially filled with water. Drugs can then be prescribed to relax the bladder.

The contractions cause the bladder wall to thicken and the thickness can be measured by vaginal ultrasonography. If there is a good relationship between bladder wall thickness, contractions and symptoms, ultrasonography could replace UDS.

We undertook a study in over 600 women with OAB symptoms. Participants underwent both ultrasonography and UDS. Women preferred the ultrasonography test to UDS, but we found that the ultrasonography results did not relate to the UDS diagnosis, nor to the severity of symptoms. Repeating ultrasonography did not produce consistent measurements.

We also asked women about their symptoms and treatments at 6 and 12 months after testing. We found that doctors use UDS results to provide the most appropriate treatment, but only 57% of women reported symptom improvements. We also found UDS provides value for money.

The results of our study conclusively show that ultrasonography is not useful for diagnosis of OAB. More research is needed to compare whether women have better outcomes if treatment decisions are based on the urodynamic results or on symptoms only.

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 (≥ 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.

Chapter 1 Introduction

Overactive bladder

Definition and prevalence

Overactive bladder (OAB) is defined by the International Continence Society and the International Urogynaecology Association[1] as a symptom complex of urinary urgency (an intense, sudden desire to void) with or without incontinence, usually with increased urinary frequency or nocturia, but in the absence of infection or other proven pathology. Increased urinary frequency and urgency seem to be more common symptoms of OAB than urinary incontinence (UI). Incontinence may be the most distressing symptom of OAB, but affects only one-third of patients.[2]

Overactive bladder affects millions of people worldwide. In the epiLUTS study, OAB prevalence was found to be 12.8%.[3,4] Prevalence and severity of OAB are known to increase with age from 14.9% in the 18- to 29-year group, to 21.3% in the 30- to 39-year group, 32.9% in the 40- to 49-year group, 35.8% in the 50- to 59-year group and up to 39.8% in the 60- to 69-year group.[5] With the increase in longevity owing to advances in health care and the population growth, the burden of OAB is going to increase in the next few decades, with a 9% increase anticipated from 500 million globally in 2013 to 546 million by 2018.[6] Two-thirds of a predominantly female sample of people with OAB had sought treatment in the 6 months prior to a multinational survey.[7]

Moderate-to-severe symptoms may have an adverse impact on lifestyle. Affected women tend to cope by restricting fluid intake and 'toilet mapping' to control urgency and frequency.[8] Women may avoid sexual contact because of the risk of coital incontinence.[9] These coping strategies can have a deleterious effect on physical, social and emotional health. Low mood and depression due to social restriction and fear of embarrassment are also associated with OAB in women (with and without urgency incontinence), while OAB also has significant financial implications (e.g. cost of pads, prescriptions, time off work, job losses, effects on the family, etc.).[10,11] Urgency or nocturia in the elderly have been linked to a higher risk of falls and fractures.[12]

Mixed urinary incontinence

The International Continence Society and International Urogynecology Association describe mixed urinary incontinence (MUI) as a complaint of involuntary loss of urine associated with urgency and also with effort or physical exertion, or on sneezing or coughing.[1] MUI may be urgency-predominant or stress-predominant and accounts for 49% of UI in women.[13] In a study on the prevalence of individual symptoms of MUI, 29% of participants had stress-predominant MUI, 15% of participants had urgency-predominant MUI and 56% (912/1626) of participants had equal severity of urgency- and stress-related MUI.[13] Appropriate categorisation of women into urgency-predominant or stress-predominant MUI has been a matter of great debate. Previous studies have used the Medical, Epidemiological, and Social Aspects of Ageing (MESA) Questionnaire,[14] the Urogenital Distress (UI) Inventory[15] and visual analogue scale (VAS) scores alongside 7-day bladder diaries to categorise women with mixed incontinence based on the predominant symptom (stress or urgency).[16] However, in routine clinical practice, women are categorised based on which symptoms they consider are more bothersome.

Underlying pathology of overactive bladder

The pathology behind OAB symptoms has been found to be detrusor overactivity (DO) in 54–58% of the cases. The remaining 42–46% of the patients may have other pathologies causing OAB symptoms[17] (*Table 1*).

TABLE 1 Conditions that cause OAB

Lower urinary tract conditions	Mechanism of affect
Idiopathic DO	Involuntary detrusor contractions associated with urgency during filling cystometry when there is no other cause
Urinary tract infection	Inflammatory markers trigger activation of sensory afferent signalling pathways
Obstruction including prolapse	Detrusor muscle hypertrophy and urinary stagnation leading to DO
Impaired bladder contractility	Urinary retention from reduced contractility, causing frequency and urgency incontinence
Bladder abnormalities or inflammation	Intravesical abnormalities may precipitate urgency and urgency incontinence
Neurological conditions: cerebrovascular accidents, Alzheimer's disease, multi-infarct dementia, Parkinson's disease, multiple sclerosis, normal pressure hydrocephalus, benign tumours or secondaries in the spine	Higher cortical inhibition of the bladder is impaired, causing neurogenic DO
Oestrogen deficiency	Inflammation from bladder mucosal atrophy, atrophic vaginitis and urethritis (e.g. local irritation and risk of urinary tract infections)

The pathophysiology of the DO and other causes of OAB have not been understood completely. Enhanced afferent activity generated by the detrusor smooth muscle and the urothelium/lamina propria may be the main mechanism.[18] In vitro studies have shown that spontaneous contractile activity was seen more often in muscle strips from overactive than from normal bladders.[19]

According to the myogenic theory, DO may be caused by an intrinsic abnormality of the detrusor muscle rather than a disturbance of its neural control. Detrusor smooth muscle cells may become hyperexcitable, react to minor stimuli and result in untimely bladder contractions resulting in urgency. In isolated bladder preparations from patients with DO, there seems to be increased co-ordination leading to larger amplitude contractions, possibly reflecting changes in intercellular communication.[20]

In OAB patients, increase in intravesical pressure secondary to detrusor contractions may cause the symptom of urgency and prompt the woman to increase her urethral closure pressure using her urethral sphincter and pelvic floor musculature. Such isometric contractions against a closed bladder neck and competent sphincter may lead to the hypertrophy of the detrusor. A thickened bladder wall is suggestive of detrusor hypertrophy.[21]

Clinical history

Clinical history taking for urinary symptoms includes the type of incontinence (provoked by urgency or activity), duration and severity of symptoms, impact of symptoms on quality of life (QoL), exacerbating factors including diet, fluid and medications, co-existing medical, surgical or gynaecological conditions and the person's strategies for coping with symptoms. In a semisystematic review, mean sensitivity and specificity of clinical history compared with a reference standard diagnosis made by urodynamics (UDS) was 69% (range 35–96%) and 60% (range 21–97%), respectively, for OAB, and 51% (range 38–84%) and 66% (range 43–96%), respectively, for women with MUI.[22] Mean sensitivity for predicting DO in women with clinical symptoms of OAB was 76%, but the specificity was only 57%.[22] Meta-analytical averages were not presented.

However, in a more methodologically thorough review, Martin *et al.*[23] reported the pooled sensitivity of clinical history assessment of urgency symptoms, compared with a UDS diagnosis of DO, to be lower at 61% [95% confidence interval (CI) 57% to 65%], and the specificity to be higher at 87% (95% CI 85% to 89%).

History of urinary symptoms alone may not help in the differentiation of the underlying pathology.[24,25] Women go through a thorough clinical assessment to rule out other causes of OAB, for example urogenital atrophy, significant pelvic organ prolapse (POP), and should undergo post-void residual (PVR) ultrasonography to rule out incomplete bladder emptying.

Bladder diaries in the assessment of overactive bladder

Bladder diaries are useful tools in the investigation of lower urinary tract symptoms (LUTS)[3] and also to assess treatment response.[26] Episodes of urgency and sensation may also be recorded, along with the activities performed during or immediately preceding the involuntary loss of urine. Severity of incontinence in terms of leakage episodes and pad use may also be reported in bladder diaries. On evaluation of accuracy of bladder diaries in OAB patients against the urodynamic diagnosis of DO, sensitivity and specificity were found to be 0.88 and 0.83, respectively.[27] A high subjective score of urgency, frequent voiding and urgency incontinence episodes (over a 3-day diary period) were strongly associated with urodynamic DO in a multivariate analysis.[28]

Urodynamics

Urodynamics are used to assess the neuromuscular function of the urinary tract and understand its storage and evacuation.[29] There are two basic aims of the UDS test: (1) to reproduce the patient's symptoms and (2) to provide a pathophysiological explanation for the patient's problems.[2] At present, laboratory UDS remain the gold standard test for assessment of OAB. Multiple diagnoses can be given following UDS tests, which include DO, urodynamic stress incontinence (USI), a combination of DO and USI and voiding dysfunction (VD).

Urodynamics consists of uroflowmetry, which measures the flow rate during voiding, and multichannel cystometry, which evaluates the pressure–volume relationship in the bladder during both filling and voiding phases. It is usually performed in the sitting or supine position and takes approximately 30 minutes to perform. Ambulatory UDS utilises natural filling and provides a more physiological technique for continuous monitoring of bladder function under nearly natural conditions, but does require a longer period of observation and for the patient to carry a data storage device while catheterised. Standard multichannel UDS has a NHS tariff of £401.[30] There were approximately 39,792 UDS attendances in England and Wales in 2013,[31] although there is a 23-fold variation in uptake between localities.[32]

Detrusor overactivity and low compliance are urodynamic diagnoses. DO is the occurrence of involuntary detrusor contractions during the filling phase of UDS. These contractions, which may be spontaneous or provoked, produce a wave form on the cystometrogram of variable duration and amplitude.[33] Neurogenic DO is where there is DO and there is evidence of a relevant neurological disorder.[33] In a normal compliant bladder, the bladder accommodates large volumes without a significant rise in detrusor pressure, but if the elasticity of the bladder is reduced, pressure rises with filling and the bladder is said to have low compliance.

In women with OAB symptoms, 45% do not have a diagnosis DO on UDS. UDS may miss DO if DO is not present at all times during the filling phase, or because UDS is not a test mimicking normal physiology and hence is not able to capture DO at its occurrence.[34] In a study of 2737 women with UI symptoms, 1626 (59%) reported mixed UI, of whom 42% had USI, 25% had pure DO, 18% had both DO and USI and 15% had normal UDS.[35] In those with stress-predominant MUI, 64% had pure USI and in those with urgency-predominant MUI, only 47% had solely DO.[35]

Clinical use of urodynamics

National Institute for Health and Care Excellence (NICE) guidelines (CG 171)[36] advise against the use of multichannel, ambulatory UDS or video UDS prior to commencing conservative management for OABs. However, there are recommendations to perform UDS before proceeding to invasive treatments for DO, for example botulinum toxin serotype A (BTX-A) (Onabotulinum A, Botox™, Allergan Ltd.) or neurostimulation.

Standardisation of the urodynamic assessment

Urodynamics is a skilled procedure, which requires training in setting up the UDS equipment, calibration of the machine, interpreting the pressure/flow recordings and counselling patients. One of the difficulties often encountered during UDS is the ability to identify artefacts and interpret the results.[5] One of the key aspects of maintaining the accuracy of UDS is to ensure that the initial resting pressures are correct and recognised.[8] Previous studies showed that even under ideal test–retest conditions, reliability can be poor for many urodynamic parameters.[9,10] In clinical settings in which UDS is performed, this can be further compromised by inconsistencies in practice. The site-to-site variation in UDS procedure may have resulted from difference in equipment and training of staff.[3] Quality control is a process through which procedure quality is maintained or improved, errors reduced or eliminated and is a crucial element of the good urodynamic practice (GUP) guidelines, which were developed following poor quality control observed during review of UDS traces from multicentre trials by the International Continence Society in 2002.[6]

Acceptability of urodynamics

Women experience significant emotional distress in relation to diagnostic evaluation.[37] UDS is an intimate and invasive test[13] involving the catheterisation of the bladder and rectum/vagina and some provocative manoeuvres.[13] A significant number of people who undergo UDS find it embarrassing, painful or distressing.[38,39] These feelings can be relieved by appropriate interpersonal skills, communication skills, maintenance of privacy and confidence in the technical ability of the health-care professional.[40] Younger women and those with a history of anxiety or depression, and those receiving a diagnosis of OAB and painful bladder syndrome have been reported to have more negative experiences during UDS.[41] Female patients found that UDS was more embarrassing when carried out by a male examiner, although they felt it less painful than their male counterparts. Patients with higher 'bother' scores may not tolerate UDS as well as a patient who has a lower bother score.[38]

Bladder ultrasonography

Ultrasonography has been claimed to be a potentially accurate and reliable test of DO and definitely is a less invasive method of diagnosis of DO through direct measurement of bladder wall thickness (BWT), an increase of which has been shown to be associated with DO.[21,42] The bladder can be visualised by transabdominal, transperineal and transvaginal scanning. Transvaginal scanning is considered the optimal method of measuring BWT in women as the probe is closest to the bladder and captures high-quality images. Ultrasonographic measurement of BWT or detrusor wall thickness (DWT) both visualise and quantify bladder wall hypertrophy, but differ in their extent of measurement. BWT includes the detrusor, the mucosa and the adventitia of the bladder wall, while DWT includes only the detrusor. BWT values thus always exceed DWT values in the same patient, rendering them incomparable.[43]

In one study, female adults with normal UDS pressure patterns had a mean BWT of 3.9 mm [interquartile range (IQR) 3.4–4.5 mm].[44] Another study of 166 women without urinary symptoms found a mean BWT 3.04 mm [standard deviation (SD) 0.77 mm; range 1.2–7.6 mm] and a small positive correlation between the increase in age and BWT. The small increase in BWT with age could be related to detrusor hypertrophy or secondary to increased interstitial collagen deposition.[45] As the age-related increase in the thickness is small, and smaller than the likely measurement error of ultrasonography, correction with respect to age may not be required.[46]

The optimum bladder volume to measure BWT is still a matter of debate. In our clinical practice, we measured BWT at a bladder volume < 30 ml. BWT is known to be fairly constant when bladder volumes are measured in the range from 0 to 50 ml.[44] Bladder outline is difficult to visualise at higher bladder volumes. Transabdominal measurement of BWT needs higher bladder volumes of around 250 ml, leading to more stretching of the bladder wall. BWT decreases rapidly between 50 ml and 250 ml of bladder filling (or until 50% of bladder capacity) but reaches a plateau with only minor changes thereafter.[43]

Evidence for the accuracy of ultrasonographic measurement of bladder wall thickness in diagnosis of detrusor overactivity

There has been conflicting evidence in the literature on the diagnostic accuracy of BWT in diagnosing DO.[47,48] A systematic review identified five studies[49–53] of women with OAB symptoms, three of these studies[49,52,53] used transvaginal bladder ultrasonography but different UDS methods.[48] In one study of transvaginal bladder ultrasonography from which accuracy data could be extracted, a BWT cut-off point of 5 mm gave a sensitivity of 84% (95% CI 76% to 89%) and specificity of 89% (95% CI 90% to 96%) for identifying DO, compared with video cystourethrography and ambulatory UDS.[49] In another study, BWT varied significantly between UDS-defined lower urinary tract conditions – the mean BWT was 3.78 mm (SD 0.39 mm) for USI, 4.97 mm (SD 0.63 mm) for DO and 6.01 mm (SD 0.73 mm) for bladder outflow obstruction; $p < 0.0001$.[54] This study reported an area under the receiver operating characteristic (ROC) curve of 0.94 (95% CI 0.89 to 0.99) for the differentiation between USI and DO or bladder outflow obstruction and 0.87 (95% CI 0.78 to 0.97) to differentiate DO and bladder outflow obstruction, suggesting thresholds of 4.1 mm and 5.6 mm, respectively, for maximal diagnostic accuracy.[54] A subsequent update to the review identified seven further studies that have shown a significantly increased BWT in patients with DO than in patients without DO.[27,50,55–59] However, the heterogeneity in the methods, including the standardisation of bladder ultrasonography and of UDS, and poor reporting of the proportions of DO cases identified as well as missed, precludes a formal diagnostic meta-analysis. Other methodological weaknesses of prior studies include unclear description of how bladder ultrasonography and UDS results were blinded to each other, recruitment mainly from a single centre and retrospective design.

Acceptability of bladder ultrasonography

Transvaginal bladder ultrasonography is a less invasive technique than UDS. The acceptability and psychological impact of transvaginal scanning has been extensively studied. In a study of 755 pregnant women undergoing transvaginal scanning, 272 (36%) experienced some pain or physical discomfort,[60] the majority (92%) describing it as 'mild' or 'discomforting', but a small minority found the scan 'distressing' (5%), 'horrible' (3%) or 'excruciating' (1%). The level of psychological trauma, measured by the impact of event score (with ratings of symptoms of avoidance and intrusion) was low, with a mean of 4.3 out of a possible maximum score of 40. The majority of the women (86%) said they would definitely or probably have a repeat scan in the future.

Rationale for the study

Urinary symptoms alone can be unreliable in establishing the diagnosis of DO in women with symptoms of OAB, so clinical guidelines recommend UDS. UDS is invasive, poorly tolerated by patients and costly with an associated risk of urinary tract infections (UTIs). The mean BWT, as determined by bladder ultrasonography, appears to be higher in women with UDS defined DO and, therefore, may have a potential discriminatory role. Existing studies were small and of variable quality and further research into the role of bladder ultrasonography in diagnosis of DO in women with OAB symptoms is required.[48] If shown to be accurate, reproducible and cost-effective, bladder ultrasonography would reduce the need for UDS.

In the absence of comprehensive evidence on the accuracy of UDS and its role in influencing the appropriate treatment pathway, the necessity of UDS has increasingly started to be questioned. In women with uncomplicated stress urinary incontinence (SUI), evidence suggests that UDS is not a necessary or cost-effective component in the treatment pathway.[61–65] However, for women with predominant symptoms of OAB, the evidence is still inconclusive. On the one hand, studies conclude that urodynamic evaluation is essential in the management of women with symptoms of OAB,[29] which is not a reliable indicator of DO in women.[17] On the other hand, others conclude that an urodynamic observation of DO is not a good predictor of the outcome of a variety of treatments for OAB.[66] NICE recommends the use of UDS prior to invasive treatments for OAB,[36] but there has also been a call for further studies examining the role of bladder ultrasonography.[36]

Overview of the research

The Bladder Ultrasound study (BUS) was a prospective multicentre diagnostic accuracy study to evaluate the accuracy of BWT in diagnosing DO. The study compared BWT measurement derived from transvaginal bladder ultrasonography with a reference standard of multichannel UDS used to verify the presence or absence of DO and other UDS defined diagnoses. Consecutive women with OAB symptoms who satisfied the eligibility criteria were approached. Consenting women with OAB symptoms were characterised by their clinical history and frequency, severity and 'bother' of their symptoms using a bladder diary and validated questionnaire International Consultation on Incontinence modular Questionnaire Overactive Bladder (short form) (ICIQ-OAB).[67] The interobserver and intraobserver reproducibility of bladder ultrasonography was assessed in substudies (see *Chapter 4*). Pain, embarrassment and acceptability of the two tests were assessed. An economic evaluation would compare different diagnostic strategies and treatments, using study and published data (see *Chapter 7*), to determine the most cost-effective diagnostic route. Women were also followed for 12 months following the investigations and the relationships between UDS diagnosis and bladder ultrasonographic measurement with treatments and symptoms were assessed.

Methodology for determination of the accuracy of bladder ultrasonography

Evaluation of the accuracy of a diagnostic test involves comparing the findings a new test with a reference standard diagnosis, which may be based on one or several pieces of test information. Accuracy focuses on estimating rates of test errors: false negatives – those who have the condition but who wrongly receive a negative test result; and false positives – those who do not have the condition but wrongly receive a positive test result. Sensitivity describes the ability of the test to correctly identify the disease (i.e. not give false-negative results) and specificity the ability to identify those without the disease (i.e. to not give false-positive results). For the evaluation of bladder ultrasonography, the findings of UDS are used for the reference standard as it is the best test for diagnosing DO. For bladder ultrasonography to be considered as a test to replace UDS, it is necessary for the rates of false negatives and false positives to be low (i.e. sensitivity and specificity to be high).

There are many possible sources of bias in accuracy studies[68] and we report our study in accordance with the Standards for Reporting of Diagnostic Accuracy statement to ensure that the risk of bias can be assessed.[69] There are three main domains of bias: (1) selection of the sample, (2) verification of the reference standard and (3) completeness of the data. Selection bias may arise if the sample is not suitably representative of the population. This is likely to occur with use of non-consecutive or convenience sampling. The BUS sought to approach all consecutive eligible women. A related issue is that of spectrum bias whereby the accuracy of tests varies among study samples with differences in disease severity (a measurable characteristic). We planned subgroup analysis to explore the variation in test accuracy owing to spectrum composition.

Empirical studies have shown that studies with differential verification, whereby the reference standard use is dependent on the index test result, produce more biased estimates of accuracy than studies with complete verification by the preferred reference standard.[70] Some of the studies of bladder ultrasonography have mixed reference standards according to the results of the index test, which can lead to bias.[44,49,52] This has occurred through using ambulatory UDS in selected subsets of patients selected according to the results of bladder ultrasonography. Although ambulatory UDS is probably a more accurate reference standard than standard UDS, it was available in only a single recruiting centre, and is costly and inconvenient, and thus we could not use it in all centres. We did aim to include a subset of patients in whom ambulatory UDS had been carried out on and use this enhanced reference standard in a sensitivity analysis, but the primary analysis is based on standard UDS assessments. We mandated that both tests were completed to ensure that we had complete data to enable the accuracy assessment to see if bladder ultrasonography can replace UDS.[71]

Reproducibility of a test

The accuracy of a diagnostic test relates to its ability to detect differences in measurements between individuals related to disease state (the signal) against a background of variability in measurements caused by measurement error (the noise or analytical variability). Tests may fail when there is little signal or when the magnitude of the noise is high compared with the size of the signal. Studies of reliability and reproducibility provide estimates of analytical variability and allow assessment of the ability of a test to detect real differences of varying magnitude. For imaging studies, there are two core sources of analytical variability: first, relating to the interpretation of images, with variability caused by measurement error within an observer (intraobserver) and between observers (interobserver); and second, relating to the imaging technique.

Any newer diagnostic test developed to assess bladder function accurately should ideally be reliable and reproducible and easy to interpret. Reproducibility of BWT is of particular importance given the fact that the bladder is a distensible organ and its thickness is known to inversely correlate with the amount of urine present in the bladder.[46] Intraobserver and interobserver reproducibility is demonstrated by studying the difference between blinded observers/measurements when exposing the same patient to the technique independently at different points of time. Transabdominal and transperineal measurement of BWT was shown to have higher interobserver variation than transvaginal measurement in a study by Panayi *et al.*[72] Hence we chose transvaginal measurement of BWT to evaluate diagnostic accuracy in diagnosing DO.

Assessing the acceptability of bladder ultrasonography and urodynamics

Extreme anxiety disrupts and unsettles behaviour by lowering the individual's concentration and affecting their self-confidence and muscular control. Unsettled behaviour during intimate and invasive tests such as UDS and bladder ultrasonography may have an impact on the level of co-operation gained from the patient, the ability to complete the test and may have an adverse effect on interpretation of test results.[73,74]

There is no reported literature on the explicit quantification of the anxiety levels elicited by UDS and very little attention has been paid to the psychological impact of invasive diagnostic testing of lower urinary tract conditions.[26] We aimed to assess state anxiety, defined as a mood state associated with preparation for possible upcoming negative events[75] and consider this alongside pain during and shortly after each test.

Effect of tests on subsequent treatment pathway

Pharmacotherapy is considered first-line treatment of OAB, with or without the use of conservative interventions like bladder retraining, pelvic floor muscle training (PFMT) (with or without biofeedback), weight loss and fluid management. The motor nerve supply to the bladder is via the parasympathetic nervous system (via sacral nerves S2, S3, S4),[76–78] which stimulates detrusor muscle contraction. This is mediated by acetylcholine acting on muscarinic receptors at the level of the bladder. Cholinergic blockade may abolish or reduce the intensity of detrusor muscle contraction.[79] Various anticholinergic medications differ with respect to efficacy, tolerability and side effect profile. Women taking anticholinergic medications

frequently experience adverse effects such as dry mouth, headache, constipation, dizziness, decreased visual acuity and tachycardia. A pharmacological classification of bladder agents used in OAB is:

- Non-selective anticholinergics: tolterodine tartrate, trospium chloride, oxybutynin hydrochloride, propiverine hydrochloride, propantheline bromide.
- M2–M3 selective anticholinergic: solifenacin succinate (Vesicare®, Astellas).
- M3 selective antagonist: darifenacin hydrobromide (Emselex®, Merus).
- Beta3 receptor agonist: mirabegron (Betmiga®, Astellas).

In patients who are refractory to conservative management of OAB (because of either a lack of efficacy or troublesome adverse effects), BTX-A has been used for over a decade with successful outcomes. The majority of patients who commence treatment with BTX-A may require long-term repeat treatments.[80] There is evidence of sustained reductions in UI episodes and increase in volume/void as well as QoL in patients with neurogenic DO[81] on repeat (up to five) injections with BTX-A.

Percutaneous tibial nerve stimulation (PTNS) involves stimulation of the posterior tibial nerve in the ankle using a fine-gauge needle, given weekly for 12 weeks and topped up as required. Improvement in OAB symptoms using PTNS is comparable to the effect of antimuscarinics but with a better side effect profile.[82] The studies included in the published systematic review considered only short-term outcomes after initial treatment.[82] In order to recommend PTNS as a practical treatment option, long-term data and health economic analysis are needed.[82] The NICE guideline on UI recommends that PTNS for OAB can be offered only if there has been a multidisciplinary team review, conservative management including OAB drug treatment has not worked adequately and the woman does not want BTX-A or percutaneous sacral nerve stimulation (SNS).[36]

Sacral neuromodulation involves implantation of a permanent sacral nerve root stimulator, which is designed to stimulate the third sacral nerve root. It is recommended to patients with refractory OAB who have failed conservative measures (including drugs), who have not responded to BTX-A treatment and have voiding difficulties.[36,83] The surgical treatment for SUI is offered if the PFMT has been unsuccessful or declined and mainly consists of a mid-urethral sling.[36]

The impact of a test is whether or not it ultimately improves patient outcomes, by identifying those that need treatment, or differentiating between alternative diagnoses and directing an appropriate treatment. There are widely held concerns that although a UDS DO diagnosis is accurate, it is not clear that it leads to different patient management or predicts patient outcome. There are some studies to suggest that patient-related outcomes are similar whether or not there is an urodynamic diagnosis of DO in patients with OAB, following a variety of treatment options.[84–87]

A comparable situation exists for stress incontinence, whereupon non-invasive assessments alone were found to be not inferior to UDS for outcomes at 1 year in a randomised controlled trial (RCT).[61] A meta-analysis has concluded that pre-operative UDS does not influence the likelihood of subjective cure or post-operative complications in women without VD undergoing primary surgery for uncomplicated SUI and so should not be carried out.[88] In a RCT, urodynamic status could not predict treatment outcomes between patients treated with extended-release tolterodine tartrate or placebo.[84]

There is a necessity, therefore, to establish the role of UDS and its impact on treatment and patient outcomes in OAB as at present its role is unclear. The BUS provided a unique opportunity to address this question and so, midway through, we proposed an extension to the study. The extension aimed to establish if treatment pathways differed following confirmation of DO based on UDS. Moreover, we sought to assess if the UDS diagnosis had an effect on patient-reported severity and improvement at 6 and 12 months after testing, and whether or not receiving the most appropriate treatment according to the UDS diagnosis improved symptoms.

Economic evaluation of the alternative diagnostic strategies

In addition to evaluating the reproducibility, accuracy and acceptability of bladder ultrasonography, it is important to assess the cost-effectiveness of testing strategies involving bladder ultrasonography, UDS and based on the primary presenting symptom in clinical history alone. The BUS would enable comprehensive primary resource utilisation for bladder ultrasonography and UDS to be collected as part of the study and the extension provided the opportunity to collect outcome data for all women who having reported bladder problems, and the treatment they received, whether or not they had a UDS diagnosis of DO. These data, together with other estimates obtained from the literature,[89–92] were used to clarify whether or not the UDS test itself represents an appropriate and justifiable use of health service resources, given the doubt over its predictive ability.

Aims and objectives of the Bladder Ultrasound Study

The original primary research objective was to estimate the diagnostic accuracy of BWT, measured by transvaginal 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 UDS and bladder ultrasonography
3. to assess whether or not measurements of BWT made using transvaginal ultrasonography have adequate reliability and reproducibility to be likely to detect differences in BWT potentially indicative of disease.

We also aimed to investigate the value added by bladder ultrasonography to information already obtained from routinely used initial non-invasive tests (history, bladder diary, disease-specific QoL questionnaire), but this became redundant when the accuracy of bladder ultrasonography was found to be poor. Subsequently, a fifth objective was added to the BUS, namely to establish the role of UDS and its impact on treatment and patient outcomes in OAB and MUI. There were six key questions:

1. Does the UDS diagnosis affect treatment pathways?
2. What were the patient-reported outcomes in the cohort of women recruited in the BUS at 6 and 12 months after testing?
3. Does the diagnosis by UDS have any effect on symptoms after 6 and 12 months, that is, can UDS predict improvement in different patient groups?
4. Does receiving treatment concordant with the urodynamic diagnosis improve patients' symptoms, compared with not receiving a concordant treatment?
5. Are presenting symptoms related to outcomes at 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?

Chapter 2 Diagnostic accuracy of bladder wall thickness via bladder ultrasonography in the diagnosis of detrusor overactivity

Introduction

A cross-sectional test accuracy study was undertaken to assess the accuracy of BWT in diagnosing DO in women with OAB symptoms. Women were recruited with OAB or urgency-predominant MUI and BWT was measured from transvaginal ultrasound scans. DO status was judged from findings obtained from multichannel UDS, which was undertaken blind to the findings from the transvaginal ultrasonography. Test accuracy was estimated by comparing BWT measures (the index test) against diagnosis of DO (the target condition) obtained from UDS (the reference standard).

Oversight

The study plan was detailed in a protocol which received a favourable ethical opinion from the Nottingham Research ethics committee (MREC 10/H0408/57). NHS trust research governance approval was obtained for 22 recruiting hospitals in the UK, with the Birmingham Women's Hospital and University of Birmingham acting as sponsors. A detailed description of the independent oversight of the study is given in *Appendix 1*.

Methods

Study sample

Consecutive women attending urogynaecology or urology clinics at either specialised referral centres or district general hospitals were approached for consent to the study. They were eligible for inclusion in the study (*Figure 1*) if they satisfied the following criteria:

1. Frequency of nine or more voids in 24 hours as reported in a 3-day bladder diary (on at least on one of the days).
2. Urgency (cannot defer the urge to void) recorded on at least two occasions in the 3-day bladder diary.
3. PVR volume $\leq$ 100 ml on the screening bladder scan.
4. No stress incontinence surgery and/or BTX-A in the past 6 months.
5. Provided written informed consent.

Exclusion criteria were:

1. Current pregnancy or up to 6 weeks post partum.
2. Pure symptoms of stress incontinence or stress-predominant mixed incontinence.
3. Evidence of cystitis (dipstick positive for leucocytes/nitrites).
4. Voiding difficulties (e.g. PVR of > 100 ml).
5. Prolapse > grade II (any compartment, as defined by the Pelvic Organ Prolapse (POP) Quantification system).[93]
6. Previous UDS assessment in the past 6 months.
7. Use of antimuscarinics for more than 6 months continuously.
8. Current use of antimuscarinics (e.g. tolterodine, solifenacin, oxybutynin). If the woman was taking antimuscarinics at the point of consent, she was eligible if the medication was ceased immediately and there was a delay of at least 2 weeks before the index and reference tests were carried out.

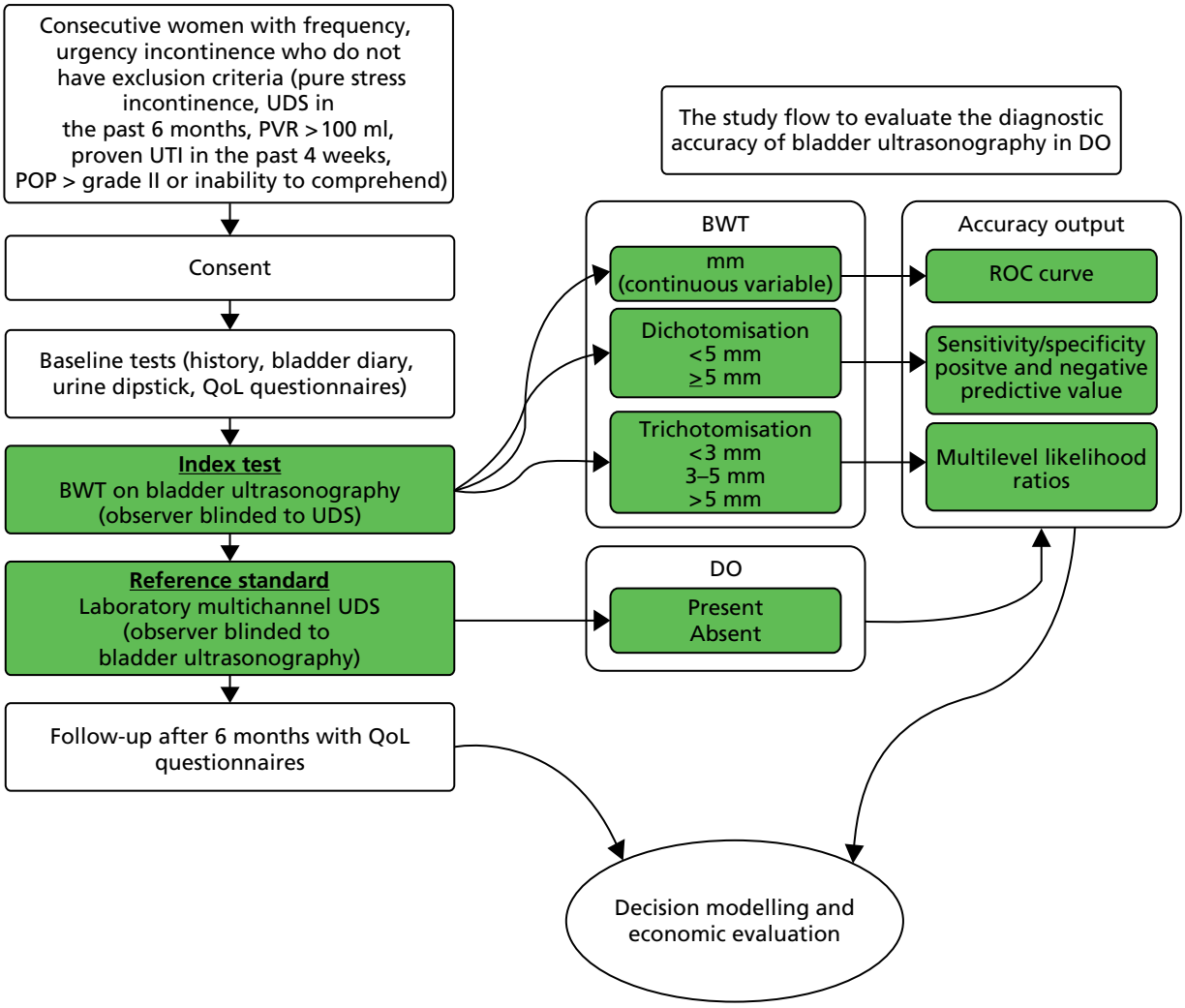

FIGURE 1 Study flow chart.

Given the need to fully inform each woman about the study (to provide time for her to consider participation and to avoid burdening her with information at the time of recruitment), a two-stage informed consent strategy was employed. The study information leaflets along with a sample consent form and bladder diaries were posted to all prospective participants with their clinic appointment letter. Research nurses and principal investigators in the recruiting hospitals were trained to reinforce the information provided and answer any questions that the women may have had. The collaborating teams approached patients for recruitment and consent at the time of consultation. We advised the recruiting centres to maintain a screening log of the eligible and ineligible participants was (identification and demographic details and any exclusion criteria).

Demographic and clinical history of the participants was collected prior to testing. This included medical and surgical history, previous treatments, if any, for bladder, bladder diary results and incontinence pad use. The ICIQ-OAB questionnaire[67] was also administered prior to testing and then at 6 and 12 months post-testing for use in the long-term follow-up study (see *Chapter 6*). A generic and preference-based health-related QoL questionnaire, the European Quality of Life-5 Dimensions (EQ-5D),[94] was administered at the same time points. The Investigating Choice Experiments CAPability measure for Adults (ICECAP-A) was given at baseline and 6 months only. Questionnaires on the acceptability of the testing were given immediately post-testing (see *Chapter 5* for output).

Setting of tests

Urodynamics was carried out by health-care professionals (doctors or nurses) who routinely carry out the procedure in clinical practice. The standard operating procedures (SOPs) (see *Appendix 5*) and quality assurance processes are described in *Chapter 3*. For bladder ultrasonography (see *Appendix 6*), hands-on training was delivered on site at each of the recruiting sites for the clinicians (doctors or sonographers) and two training workshops were carried out at the Birmingham Women's Hospital.

At centres where ultrasonography was undertaken in the UDS suite, the scan was performed at the same clinic visit but by an independent trained observer blinded to the UDS result. At centres where the scan was to be performed in the radiology department, both diagnostic tests were carried out within 4 weeks of each other. Clinicians performing each of the tests were blinded to the results of the other. The operator performing the UDS recorded the findings and the diagnosis in the UDS test pro forma and the operator performing bladder ultrasonography recorded the BWT measurements in the bladder ultrasonography test. The index test (bladder ultrasonography) was carried out in a scan suite and the UDS carried out in the UDS suite in the majority of the centres. Both data collection forms were then collected by a research nurse/research fellow co-ordinating the recruitment at each centre and were sent to the bladder ultrasonography trial office (Birmingham Clinical Trials Unit) in a sealed envelope.

Index test: bladder wall thickness via ultrasonography

The BWT was measured from a transvaginal scan with the end-firing transvaginal probe in the sagittal plane (midline) introduced 1 cm beyond the vaginal introitus in the midline. BWT measurement involved measuring the inner and outer hyperechoic areas along with the hypoechoic detrusor sandwiched in the middle.[49,95] When the inner and outer hyperechoic areas are excluded and the inner hypoechoic area only is measured, it is classed as DWT. We measured BWT via the transvaginal approach at a volume < 30 ml. The echo-poor central area of the urethra was used as a landmark for BWT scanning. BWT was measured in millimetres using tools provided on the ultrasonography machine at the following three sites perpendicular to the luminal surface of the bladder (*Figures 2* and *3*):

1. The thickest part of the trigone.
2. The dome of the bladder in the midline.
3. The anterior wall of the bladder.

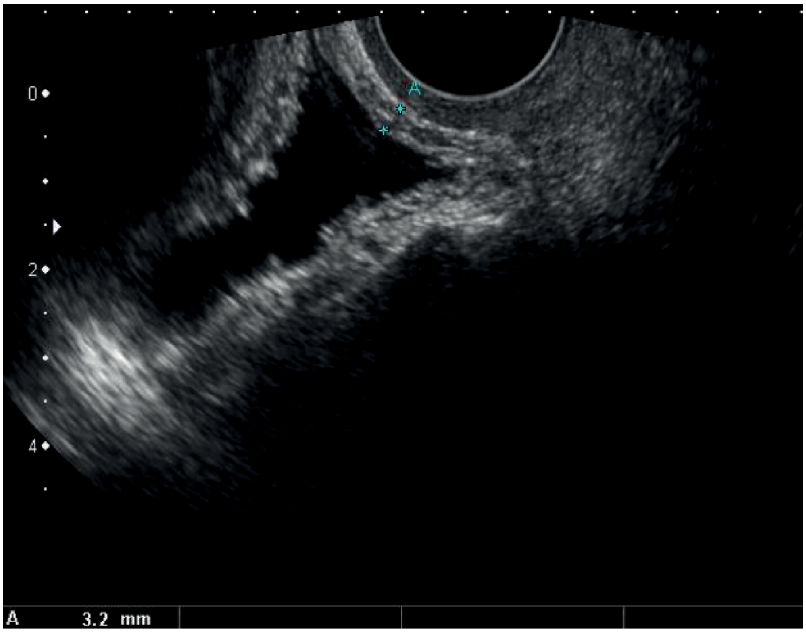

FIGURE 2 Bladder wall thickness measurement at trigone.

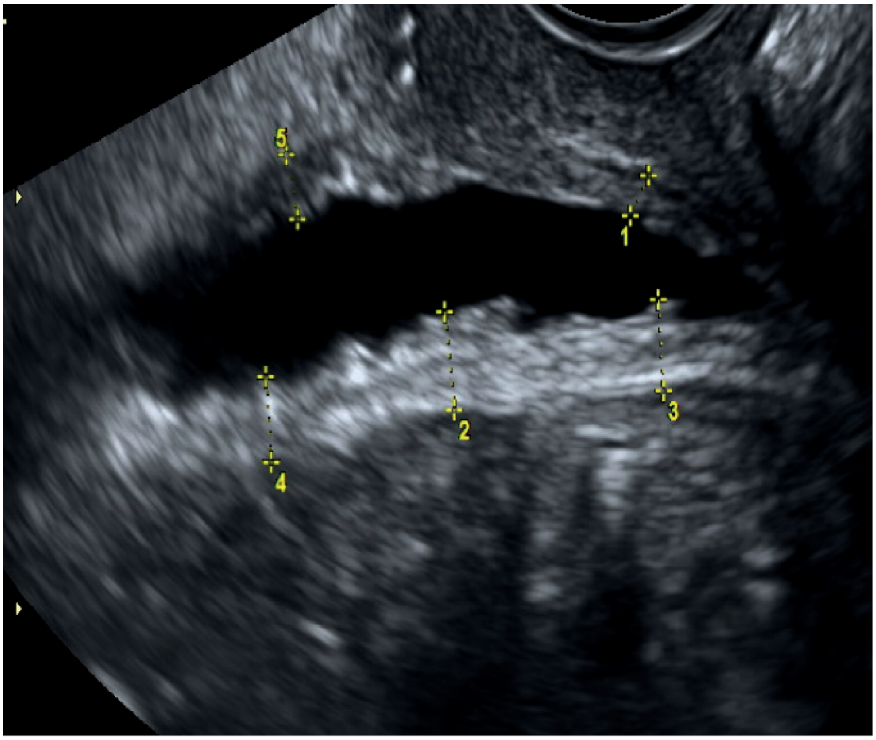

FIGURE 3 Dome midline and two lateral measurements on either side.

For the purposes of this study, BWT was to be calculated as the mean of these three measurements. In order to assess whether or not the mean of three values at the dome was similar to the value obtained on dome, anterior wall and trigone, two further measurements of DWT 1 cm on either side of the dome midline were also taken.

Reference standard: urodynamics

All women underwent UDS, using a standardised protocol in accordance with the GUP guidelines of the International Continence Society.[96] Participants attended the UDS clinic with a full bladder. Uroflowmetry was performed with the woman voiding in private and recorded on a gravimetric flow meter. Filling cystometry was then performed with the woman in sitting position. This was then followed by voiding cystometry with the pressure lines in situ (see *Appendix 5*).

Sample size

A target sample size of 600 participants was determined a priori. The computation was based on presuming a prevalence of 50% for DO,[17] providing 300 women for the estimate of sensitivity and 300 for the estimate of specificity. This allows estimation of sensitivities and specificities with 95% CIs of width 10% for sensitivity and specificity values between 70% and 95%, and narrower for higher values.

Data analysis

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 DO, $<$ 5 mm indicating absence of DO). BWT of 5 mm was chosen as the cut-off point for discriminating DO based on the evidence from previous studies[49,97] and was pre-specified in the protocol.

Likelihood ratios for the following three ordered categories of BWT were also pre-specified: $<$ 3 mm, $\geq$ 3 mm to $<$ 5mm, $\geq$ 5 mm.[98] 95% CIs were calculated using binomial exact methods. A ROC curve was constructed and the area under the curve (AUC) computed (with 95% CI) to give an overall estimate of BWT accuracy across all thresholds. Statistical significance was tested by comparing against the

uninformative model (i.e. for which AUC = 0.5) using a non-parametric approach.[99] The distributions of BWT measurements in groups with and without a DO diagnosis were depicted using box-and-whisker plots and mean values compared using a two-sample *t*-test.

A number of sensitivity analyses were performed on the primary population to test the robustness of the results to protocol deviations and missing data. ROC curves and associated AUC values were computed for each analysis. The analyses were:

1. Excluding those patients for whom it was revealed that the UDS test result was not blinded to the results of ultrasonography (to exclude any possible diagnostic review bias).
2. Excluding those patients for whom it was calculated to be more than 4 weeks between index and reference standard tests (to exclude any possible disease progression bias).
3. Including results of incomplete ultrasonographic measurements, that is when not all three components of BWT were recorded (in these cases, if one or two measurements were missing, the average of the remaining values was taken to be BWT; this was to exclude the impact of missing measures).
4. Replacing the original UDS diagnosis with that from the additional ambulatory UDS test when available (this happened only in 14 instances, all from one centre; this analysis was intended to assess the impact of this presumed more sensitive UDS than supine UDS).
5. Using the trigone measurement alone as BWT (to explore whether or not a single measurement was as accurate as the mean of three locations).
6. Excluding those who had 'provoked DO' (detrusor pressure rise on provocation testing – 187 cases; this was to assess the impact of iatrogenic DO).
7. Excluding those who had PVR of > 30 ml on BWT testing (34 cases; this was to explore the impact of those with minor degrees of incomplete emptying).
8. Taking the average of dome, 1 cm left of dome and 1 cm right of dome as BWT (to explore whether or not measurements at this location improve accuracy).

In addition, some unplanned exploratory analyses were also performed to gauge the effect of changing the population of interest and also to see which parameters were associated with DO diagnosis. The populations explored were:

1. Women with urgency alone on clinical history (i.e. excluding those with mixed stress/urgency incontinence).
2. Women with 'pure' DO only (i.e. not alongside another diagnosis from UDS).
3. Women with 'wet' DO only (i.e. not including those with 'dry' DO).

Pre-planned subgroup analyses were also performed to compare test accuracy between subgroups. ROC curves were created for each subgroup and associated AUC values compared using a large sample chi-squared test for independent curves.[100] The subgroups used here to dichotomise patient groups were:

1. Previous treatment with antimuscarinics.
2. A clinical history suggesting mixed incontinence.
3. Presence of a UTI in the previous 12 months.
4. Voiding difficulties.
5. Previous incontinence surgery.
6. Body mass index (BMI) (< 25 kg/m^2, ≥ 25 kg/m^2).

Exploratory analyses were undertaken to assess variables associated with a diagnosis of DO using logistic regression. The above six subgroup variables were included along with pre-test International Consultation on Incontinence modular Questionnaire (ICIQ) score, BWT, age, duration of symptoms, ethnicity, number of vaginal deliveries, menopausal status, parity and previous POP surgery. Covariates were considered individually and then in a multivariable analysis. Three multivariable models were constructed: one using all possible explanatory variables, another using all possible explanatory variables but using a multiple

imputation approach to generate missing responses[101] and another using a backward-step process to eliminate unimportant variables (a level of $p = 0.1$ was used here as criteria for staying in the model). We also examined whether or not BWT had any relationship with baseline ICIQ-OAB score using a simple linear regression model.

Results

Recruitment

Recruitment of participants started in March 2011 and closed in March 2013. A total of 1310 women were approached. Six hundred and eighty seven women who were eligible and consented to participate were recruited into the study from 22 centres (*Table 2* and see *Figure 23*). This number was slightly higher than the agreed sample size calculation to compensate for a small number of study withdrawals and women without complete index and reference standard test results (*Figure 4*).

TABLE 2 Recruitment by centre

Recruiting centre	Number	Per cent
Birmingham Women's Hospital	254	37
Medway Maritime Hospital, Kent	109	16
Mayday University Hospital, Croydon	92	13
Basingstoke and North Hampshire Hospital	30	4
St Mary's Hospital, Manchester	26	4
Staffordshire General Hospital	26	4
Stepping Hill Hospital	25	4
Ormskirk and District General Hospital	19	3
New Cross Hospital, Wolverhampton	16	2
Royal Bournemouth General Hospital	16	2
The Alexandra Hospital, Redditch	14	2
City General Hospital (University Hospital of North Staffordshire)	10	1
Crosshouse Hospital, Ayrshire	9	1
Manor Hospital, Walsall	8	1
Northampton General Hospital	6	1
Royal Hallamshire Hospital, Sheffield	6	1
Derriford Hospital, Plymouth	5	1
Pinderfields General Hospital	4	1
St Mary's Hospital, Paddington	4	1
Southern General Hospital, Glasgow	3	< 1
The Royal London Hospital	3	< 1
Sandwell General Hospital, Birmingham	2	< 1
Total	687	100

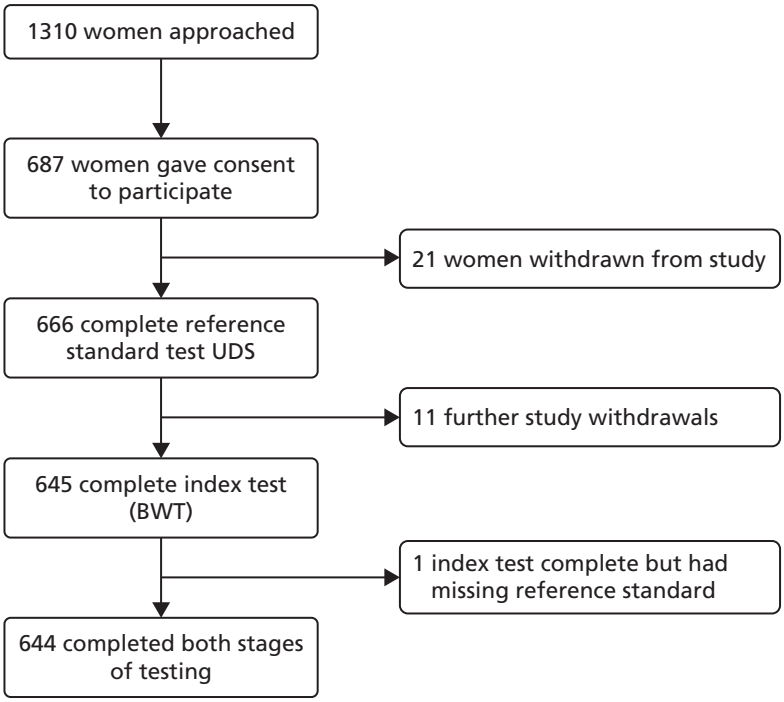

FIGURE 4 Participant flow diagram.

Characteristics of participants

Characteristics of women who consented to take part in the study are shown in *Table 3*. The mean age of women was 52.7 years (SD 13.9 years) and the average BMI was 30.6 kg/m^2 (SD 12.2 kg/m^2). A total of 55% (378/687) of the women were post menopausal. According to the clinical history, 33% (226/687) reported only urinary urgency without incontinence and 61% (419/687) had urgency-predominant MUI. The median duration of symptoms was 3.0 years (IQR 1.6–7.0 years).

TABLE 3 Characteristics of included participants ($n = 687$)

Characteristic	Category	Value
Age (years)	Mean (SD)	52.7 (13.9)
	Missing	0 (–)
Ethnicity, *n* (%)	White British/Irish/other	538 (78)
	Asian Pakistani/Indian/Bangladeshi/other	72 (10)
	Black Caribbean/African/other	49 (7)
	Mixed/other	18 (3)
	Not given/missing	10 (1)
Parity, *n* (%)	0	69 (10)
	1	90 (13)
	2	241 (35)
	3	152 (22)
	4	56 (8)
	> 4	63 (9)
	Missing	16 (2)

continued

TABLE 3 Characteristics of included participants ($n = 687$) (*continued*)

Characteristic	Category	Value
Post menopausal (last menstrual period > 1 year), *n* (%)	Yes	378 (55)
	No	293 (43)
	Missing	16 (2)
BMI (kg/m^2)	Mean (SD)	30.6 (12.2)
	Missing	28
Incontinence type, *n* (%)	MUI	419 (61)
	Urgency incontinence alone	226 (33)
	Stress incontinence alone	4 (1)
	Neither	19 (3)
	Missing	19 (3)
If mixed, which started first, *n* (%) ($N = 419$)	Urgency	226 (54)
	Stress	107 (26)
	Unsure	54 (13)
	Missing	32 (8)
Current or previous treatment with antimuscarinics, *n* (%)	Yes	226 (33)
	No	444 (65)
	Missing	17 (2)
Recurrent cystitis (three or more in last 12 months), *n* (%)	Yes	50 (7)
	No	606 (88)
	Missing	31 (5)
Voiding difficulties, *n* (%)	Yes	286 (42)
	No	374 (54)
	Missing	27 (4)
Vaginal birth, *n* (%)	Yes	561 (82)
	No	95 (14)
	Missing	31 (5)
Previous incontinence surgery, *n* (%)	Yes	36 (5)
	No	623 (91)
	Missing	28 (4)
Previous POP/UI surgery, *n* (%)	Yes	56 (8)
	No	603 (88)
	Missing	28 (4)

The reference standard: urodynamics

The number of participants with a complete reference standard diagnosis was 666/687 (97%). The other 21 (3%) decided to withdraw from the study before any testing could take place (see *Figure 4*). Details of the findings in these tests are given in *Table 4*. Of these, 399 (60%) were diagnosed with DO (95% CI 56% to 64%) (*Table 5*). Of the 399, 245 were given further subdiagnosis of 'wet' DO (61%) and 154 as 'dry' DO (39%). The participants also had their DO diagnosis subcategorised as phasic 'spontaneous' DO (detrusor contraction during the filling phase: 182/369, 49%; 30 observations missing), provoked DO (if the detrusor contraction occurred during or after provocative measures such as cough, running water or immersion of hands in cold water: 56/369, 15%) or both spontaneous and provoked (131/369, 36%).

TABLE 4 Measurements from UDS tests ($n = 666$)

Uroflowmetry	Number of women (%) for binary data, median (IQR) for continuous data, *n* values recorded
Patient had comfortably full bladder, yes, *n/N* (%)	502/655 (77)
Volume voided (ml), *n* (IQR)	129 (58 to 245), $n = 627$
PVR volume (ml), *n* (IQR)	10 (2 to 40), $n = 619$
Maximum flow rate (ml/seconds)	16 (9 to 25), $n = 596$
Filling cystometry	
Patient in recommended sitting position for test, yes, *n/N* (%)	457/664 (69)
Fill rate (ml/min), *n* (IQR)	100 (70 to 100), $n = 660$
First desire (ml), *n* (IQR)	135 (84 to 197), $n = 644$
Normal desire (ml), *n* (IQR)	200 (140 to 268), $n = 587$
Strong desire (ml), *n* (IQR)	272 (199 to 357), $n = 562$
Pain (if reported) (ml), *n* (IQR)	300 (203 to 395), $n = 154$
Leakage (if applicable) (ml), *n* (IQR)	10 (0 to 100), $n = 229$
Total volume in bladder at the end of filling (ml), *n* (IQR)	421 (314 to 498), $n = 639$
Rise in detrusor pressure upon filling, yes, *n/N* (%)	350/598 (59)
Detrusor pressure at start (cm H_2O), *n* (IQR)	0 (–1 to 1), $n = 638$
Detrusor pressure rise on filling to 500 ml (cm H_2O), *n* (IQR)	12 (6 to 21), $n = 576$
Detrusor pressure rise when complaint of urgency (cm H_2O), *n* (IQR)	12 (5 to 21), $n = 557$
Provocation test (when performed)	
Detrusor pressure rise with cough, yes, *n/N* (%)	101/517 (20)
Detrusor pressure rise with running tap, yes, *n/N* (%)	119/367 (32)
Detrusor pressure rise with exercise, *n/N* (%)	39/124 (31)
Voiding cystometry	
Peak flow rate (ml/seconds), *n* (IQR)	20 (15 to 28), $n = 624$
Maximum voiding pressure (cm H_2O), *n* (IQR)	41 (29 to 60), $n = 577$
Residual volume (ml), *n* (IQR)	0 (0 to 20), $n = 540$

TABLE 5 Summary of all urodynamic diagnoses

Urodynamic diagnosis	Number of women (%), *n* = 666
***Including DO* (n = *399*)**	
DO only	258 (39)
DO/USI	97 (15)
DO/VD	18 (3)
DO/VD/USI	12 (2)
DO/low compliance	8 (1)
DO/USI/low compliance	5 (1)
DO/VD/USI/low compliance	1 (< 1)
***Not including DO* (n = *267*)**	
Normal	124 (19)
USI only	78 (12)
Low compliance only	36 (5)
VD only	14 (2)
VD/USI	8 (1)
USI/low compliance	6 (1)
VD/low compliance	1 (< 1)

The index test: bladder wall thickness

The number of participants with all three BWT measurements (trigone, dome midline, anterior wall midline) available was 645 (94%). Eleven patients had withdrawn after having UDS but prior to ultrasonography (see *Figure 4*) and a further 10 (1%) had partial measurements recorded (nine with two out of the three measurements and one with one of the three measurements). Summary statistics and distribution of BWT are provided in *Table 6* and *Figure 5*.

TABLE 6 Bladder wall thickness (mm) summary statistics

Measurement	Mean (SD), *n*	Minimum, maximum
(a) Trigone	4.51 (1.50), 649	0.70, 9.90
(b) Dome midline	5.00 (1.67), 654	1.30, 11.90
(c) Anterior wall midline	4.85 (1.53), 651	1.00, 11.30
Average (a, b, c)	4.78 (1.34), 645	1.07, 9.60
(d) Dome 1 cm left	4.84 (1.69), 641	1.20, 12.10
Average (a, d, c)	4.73 (1.32), 631	0.97, 10.13
(e) Dome 1 cm right	5.00 (1.76), 639	1.10, 10.80
Average (a, e, c)	4.78 (1.36), 629	1.10, 10.50

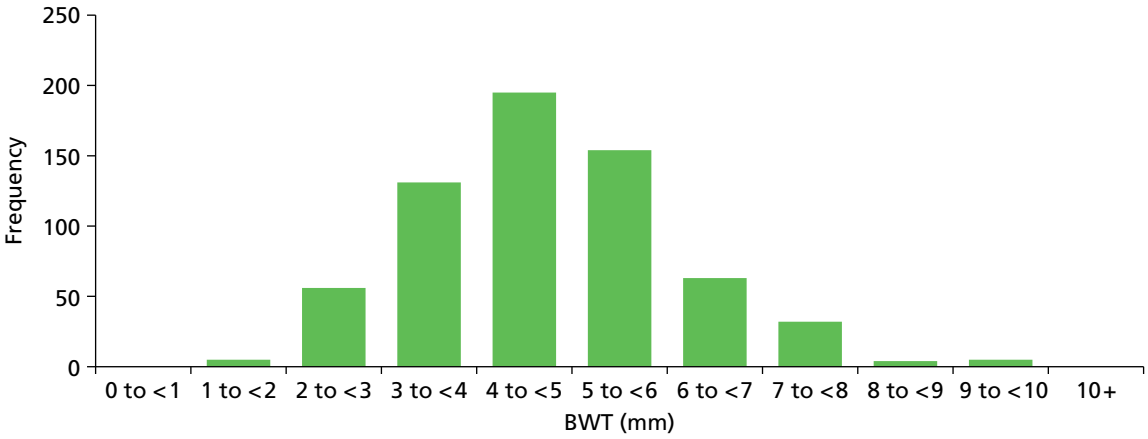

FIGURE 5 Histogram of BWT measurements (average: trigone/dome midline/anterior wall).

Timing and safety of tests

Six hundred and forty-four participants had both complete index and reference standard results (one had a complete index test but was missing their reference standard, see *Figure 4*). Of these, 439/644 (68%) had both BWT and UDS performed on the same day. Only a small proportion (26/644, 4%) were performed more than 4 weeks apart. Ninety-seven per cent of reference tests (616/632, 12 observations missing) were confirmed as being blind to the index test. No serious adverse events were reported following either test, although 49/479 (10%) of those responding reported having a UTI within 2 weeks of testing at a 6-month follow-up. Seventy five per cent (36/48, one observation missing) of these were diagnosed by a general practitioner (GP) or in a hospital and resulted in antibiotic use in 83% of cases (39/47, two observations missing).

Estimates of test accuracy

Estimation of the accuracy of BWT showed poor sensitivity, specificity and LRs at all pre-specified cut-off points of 5 mm, < 3 mm/3–5 mm/≥ 5 mm (*Tables 7–9*). The ROC curve (*Figure 6*) 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 DO-positive group than the DO-negative group: 4.85 mm (SD 1.36 mm) versus 4.70 mm (SD 1.29 mm); $p = 0.19$ (*Figure 7*) or that it had any relationship with ICIQ-OAB symptoms score when measured at presentation ($r = –0.01$; $p = 0.88$).

The planned and unplanned sensitivity analysis described above did not change the interpretation of these findings (see *Appendix 2*, *Figure 23* and *Appendix 3*, *Figures 24–36*). There was some evidence, albeit weak, that those diagnosed with 'wet' DO had higher BWT than those with 'dry' DO (wet 4.94 mm vs. dry 4.69 mm; $p = 0.08$) (see *Appendix 3*, *Figure 32*). However, when the BWT for the wet DO group was analysed alone, the AUC was only 0.55, 95% CI 0.50 to 0.59. There was no evidence that BWT performed any differently in any of the pre-specified subgroups (see *Table 45*).

TABLE 7 Comparison of index and reference standard results: dichotomised at 5 mm

		Reference standard (UDS)		
		DO	Non-DO	Total
Index test: BWT by ultrasonography	Positive result (≥ 5 mm)	165	98	263 (41%)
	Negative result (< 5 mm)	223	158	381 (59%)
	Total	388 (60%)	256 (40%)	644

TABLE 8 Estimates of BWT test accuracy: dichotomised at 5 mm

Accuracy parameter	Value	95% CI
Sensitivity	43%	38% to 48%
Specificity	62%	55% to 68%
PPV	63%	57% to 69%
NPV	41%	36% to 47%
LR+	1.11	0.92 to 1.35
LR–	0.93	0.82 to 1.06

NVP, negative predictive value; PPV, positive predictive value.

TABLE 9 Estimates of BWT test accuracy: trichotomised at 3 mm, 3–5 mm and ≥ 5 mm

		Reference standard (UDS)				
		DO	Non-DO	Total	LR	95% CI
Index test: BWT by ultrasonography	Result > 5 mm	165	98	263	1.11	0.92 to 1.35
	Result 3–5 mm	193	132	325	0.96	0.83 to 1.13
	Result < 3 mm	30	26	56	0.76	0.46 to 1.26
	Total	388 (60%)	256 (40%)	644		

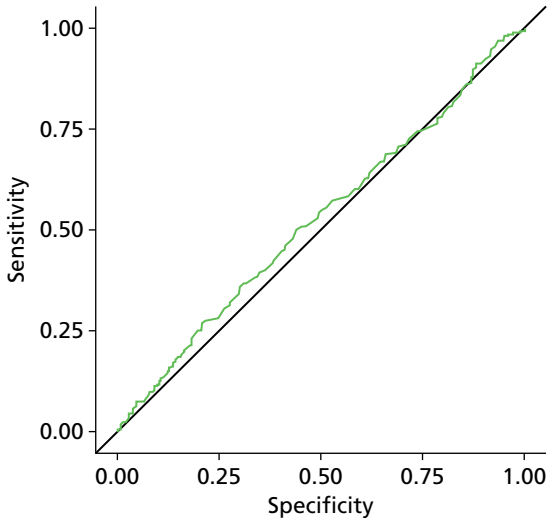

FIGURE 6 Receiver operating curve analysis for BWT. AUC = 0.5267.

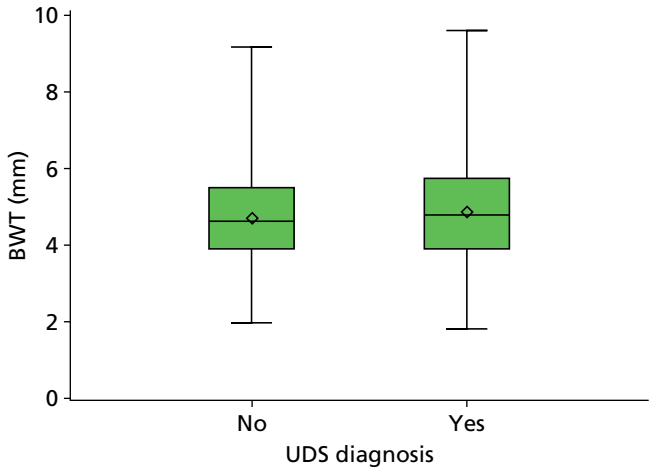

FIGURE 7 Box and whisker plot comparing BWT with DO diagnosis.

In the multivariable exploration of factors possibly associated with DO diagnosis, only higher baseline ICIQ score (i.e. worse symptoms) was associated with DO [odds ratio (OR) 1.21, 95% CI 1.13 to 1.29; $p < 0.0001$ from the model including all possible variables], that is, the odds of DO diagnosis were increased by 21% for every point increase in ICIQ score. Previous treatment with antimuscarinics and previous history of UTI in the previous 12 months also showed some relationship but these were of borderline statistical significance (see *Tables 46* and *47*). Despite the evidence of association between ICIQ score and DO diagnosis, ICIQ was not found to be an accurate predictor of DO with an AUC of 0.65 (95% CI 0.61 to 0.70).

Discussion

Summary of main findings

To date, BUS is the largest diagnostic accuracy study evaluating the diagnostic accuracy of BWT to diagnose DO in women with symptoms of OAB and using transvaginal ultrasonography with a near empty bladder. We could not find any evidence that BWT was of any clinical benefit in women with DO, indeed it appeared to be no better than chance at making this diagnosis with an AUC of 0.53 (95% CI 0.48 to 0.57). Extensive sensitivity analyses and subgroup analyses were carried out but did not alter the interpretation of these findings. Furthermore, BWT had no relationship to ICIQ score upon presentation, indicating that it has no relationship with symptom severity. ICIQ score was shown to have some relationship with DO diagnosis. Based on this evidence, we conclude that BWT is not a useful test in diagnosing DO.

Strengths and limitations of methods

The validity of our findings relied on the quality of the study. The protocol pre-specified key study methods and analyses and was peer reviewed. The study was undertaken with independent oversight with biannual meetings of independent Data Monitoring Committee (DMC) and Study Steering Committees. Blinding of operators performing bladder ultrasonography and UDS was ensured for 97% of the women recruited into the study. Verification bias was minimised by incorporating a complete verification design. Disease progression bias was minimised by conducting BWT and UDS within a short time span of each other, often within the same day. The spectrum variation (recruiting women with varying degree of severity of OAB or urgency-predominant MUI) was a strength of the design, assuring the applicability of the findings to NHS practice.

The study was powered to ensure that estimates of sensitivity and specificity would be made with adequate precision to draw robust conclusions and we recruited beyond the target. Participants in the study were recruited from several centres (university teaching and district general hospitals). Women were of mixed ages, ethnicities and social background, and were recruited from various parts of the UK.

The prevalence of DO in our study was 60%, which was similar to other studies (48% of 'OAB dry' and 58% of 'OAB wet' women were found to have DO on UDS).[102] The adverse events in our study were UTI in 10% of the population following the UDS procedure. The post UDS rate of UTI was very similar to the other studies so far.[103]

As bladder ultrasonography is a relatively new transvaginal scan technique, concerns may be raised on the quality of scan measurements in the study. However, the technique is straightforward to perform with the urinary bladder being an anterior and relatively superficial midline structure. To standardise the performance of bladder ultrasonography, we developed a SOP (see *Appendix 5*) for carrying out ultrasonography and provided hands-on training at individual recruitment sites to our co-investigators and organised two workshops on BWT measurements. The chief investigator and clinical research fellow were assessed by the main recruiting centre consultant radiologist of 14 years' experience in the technique of BWT measurement. The collaborators had to be signed off as competent by the chief investigator or clinical research fellow once they had completed at least five scans before recruiting patients into the study.

Good site-to-site reliability is essential for multicentre clinical trials using UDS. With the use of continuous quality improvement and training on standardised urodynamic testing procedures and interpretation guidelines, the technical quality of urodynamic findings were improved.[104] To improve reliability of our gold standard test UDS, we used various proactive measures such as standard urodynamic testing protocols, standard interpretation guidelines and auditing the traces centrally every 6 months to ensure ongoing quality assurance. We are therefore confident of the methods and hence the results.

Concerns about the reference standard

Accuracy measures express how the results of the test under evaluation agree with the outcome of the reference standard. Estimates of diagnostic accuracy are directly influenced by the quality of the reference standard.[105] When the accuracy of the reference standard is unknown, or known to be imperfect, estimates of the sensitivity and specificity or AUC for new diagnostic tests will be biased as misclassifications in the reference standard diagnosis will have been misattributed to errors made by the index test. Our reference standard, UDS, has been shown to have less than perfect reproducibility in previous studies in patients with OAB[106,107] and also in healthy women.[108,109] When estimating the test accuracy of BWT against an imperfect reference standard (UDS), the accuracy of BWT may have been biased to an unknown degree, or submerged in the 'noise' from the imperfect reference standard. However, the poor accuracy for BWT elicited in our study is unlikely to have been entirely caused by misclassifications made by UDS, as there was no significant relationship between BWT measurements and grades of DO severity or subsequent treatment responses (see *Chapter 6*). When the test values do not differ among those with varying grades of the target condition, it can be inferred that the lack of accuracy may be an inherent feature of the index test.

Evidence of the misclassification rates for UDS comes from several studies.[110,111] Homma *et al.*[107] undertook repeat UDS in DO patients within 2–4 weeks to study reproducibility. There was increase in the volume variables by 10–13% ($p < 0.01$), absence of involuntary contractions (10%) and a reduction in the maximum detrusor pressure by 18% during the repeat urodynamic test indicating poor reproducibility.[107] In a study of 59 healthy women without LUTS, UDS was repeated immediately after the initial test without removing the catheters and then again 1–5 months later. The mean difference was dispersed away from zero in both immediate and short-term reproducibility indicating poor reproducibility.[112]

Placing the results in the context of other research

In an update of the systematic review on the diagnostic accuracy of BWT in diagnosing DO (unpublished data: Suneetha Rachaneni, University of Birmingham, 2015) 21 studies[28,49–59,72,97,113–119] have been identified that have investigated the relationship between BWT measured by ultrasonography and DO, results from which for the sensitivity, specificity and AUC for bladder ultrasonography vary from 37% to 91%, 61% to 97% and 0.61–0.91, respectively (see *Appendix 7*). The studies draw mixed conclusions, ranging from

claims that bladder ultrasonography is highly diagnostic, to finding statistically significant but diagnostically weak relationships, to finding no relationship at all.[49,51,54,57,72,117] Our study is the most conclusively negative study of the diagnostic value of bladder ultrasonography to date.

Initial studies have shown transvaginal BWT to be an accurate diagnostic marker for DO.[44,49] For a mean BWT cut-off point of 5 mm, the specificity was calculated to be 89% (95% CI 78.8% to 96.11%) with a sensitivity of 84% (95% CI 75.8% to 89.7%).[49] However, this study used ambulatory UDS (in patients who had normal video UDS) as a secondary diagnostic test in 25% of the study population and this introduced workup bias that might have resulted in inaccurate estimation of sensitivity and specificity of BWT.

Many of the other studies have differences in the tests used and the populations studied and methodological weaknesses, which may explain why their results vary and differ from the finding of this prospective study. Poor reporting renders it difficult to make comparisons with findings in 10 studies that did not report estimates of sensitivity, specificity or area under the ROC curve. These studies typically compared the mean BWT between diagnostic groups and did not interpret the findings against a positivity threshold.

Eight of the studies did not use transvaginal ultrasonography,[28,50,51,56,57,114,116,119] but instead used transabdominal or translabial ultrasonography. Some transabdominal ultrasonography studies included men as well as women and usually were undertaken with a full bladder leading to much smaller measures of mean BWT.[116,119]

Ten studies made comparisons with healthy controls,[28,50,55–59,114–116] two others excluded patients with MUI[54,117] and one other enriched with women with equivocal UDS findings.[52] One was undertaken in patients with spinal injuries.[119] It is known that altering the spectrum of patients from that encountered in practice will influence estimates of sensitivity and specificity.[120] Exclusion of the MUI cases and inclusion of healthy controls will lead to overestimation of test accuracy; enrichment of difficult to diagnose cases will underestimate test accuracy.

A key difference between the BUS and all others is the focus on the accuracy of ultrasonography in women who do not have signs of pure SUI. Women with SUI diagnosed by symptoms and diary usually proceed to treatment without further diagnostic investigation, including UDS.[36] Thus, ultrasonography has no role in this group. The BUS focused on identifying women with DO among those who have urgency or mixed incontinence, for which UDS currently is used, to assess whether or not ultrasonography can replace UDS in this context. Women in whom SUI is diagnosed clinically, by symptoms and bladder diary responses, usually proceed to treatment without further diagnostic investigation, including UDS.[36] Thus, ultrasonography has no role in this group. Studies that have assessed the value of ultrasonography to differentiate between pure SUI and OAB have addressed a question that is no longer relevant to clinical practice.

In summary, our study was the largest prospective study to date, it is the only study that recruited a representative sample of women presenting with an urgency-dominant complaint (others have recruited other wider groups), and used service-based (but trained and quality assured) ultrasonography services from across many centres (whereas other studies were all single centre and often tertiary centre based).

Variation in technique of transvaginal bladder wall scanning

Our technique of measuring BWT in the sagittal plane with a transvaginal probe placed at the introitus was easy to learn and perform with visualisation of the urethra as the landmark. Previous studies have described parasagittal measurements of BWT at three places on the bladder wall, with the calculation of an average BWT.[41,109] In one study, DWT instead of BWT was measured by transperineal scanning.[51] The rest of the studies have used transabdominal scanning at various bladder volumes, but we rejected this technique on account of the greater reported interobserver variation.[71]

Interpretation of findings

Measuring BWT is not accurate enough to consistently identify those with DO and hence it will not be helpful in reducing the need for UDS. It is believed that spontaneous DO is secondary to pathology in detrusor muscle whereas provoked DO is caused by pathology in the bladder neck. Provoked DO, which was previously called urethral instability, could be caused by primary urethral aetiology or some unknown pathology.[121] As women with spontaneous DO are known to respond better to antimuscarinic treatment[122] than those with provoked DO or urethral instability, we carried out sensitivity analyses to evaluate diagnostic accuracy of BWT excluding those with provoked DO. Similar to findings in the study by Serati,[113] we did not find any difference in BWT in spontaneous and provoked DO groups.[113]

Implications for practice

Given the poor performance for BWT as tool for diagnosing DO, BWT cannot be recommended as a replacement test for UDS.

Recommendations for research

There is some emerging evidence in the literature that the response to invasive therapies may be similar in patients with frequency and urgency with or without urgency incontinence, with or without the observation of DO on UDS.[123–125] The necessity to diagnose DO on UDS and its role in improving patient-related outcome measures needs to be evaluated in future diagnostic RCTs.

Chapter 3 Quality control of the urodynamics

Introduction

Objectives

The objectives of this part of the study were as follows:

1. To audit the quality of UDS traces submitted to the study office at the University of Birmingham Clinical Trials Unit as the reference standard for the patients in the BUS.
2. To assess whether or not recommended changes in UDS practice had been implemented following an initial audit among the recruiting centres for the BUS.

Methods

Urodynamic studies

Urodynamic studies were to be performed in a standardised manner as per the GUP guidelines from the International Continence Society[6] and a SOP that had been produced for the study based on this (see *Chapter 2, The reference standard: urodynamics*). For patients at the main centre (Birmingham Women's Hospital), according to the study protocol, patients were offered ambulatory UDS if they had a normal result from standard multichannel UDS.

Assessment of protocol compliance

An assessment of compliance with the UDS SOP for the BUS and with GUP was performed on 64 (20%) of the first 302 UDS traces received at the study office, between May 2011 and May 2012.

An expert panel, comprising a consultant in urogynaecology and a urogynaecology research nurse from the lead centre, assessed the traces independently. They were blinded to the investigator's UDS diagnosis, the UDS operator and centre providing the trace, to minimise any bias in the assessment. All participating sites were audited regarding the urodynamic technique used throughout the study.

Traces were randomly selected by the BUS trial co-ordinator from traces supplied by each of the recruiting hospitals, as suggested by the DMC. The expert panel rereviewed UDS traces they had undertaken (the lead centre recruited 37% of all participants into the BUS) but were not aware of which traces were from the lead centre.

Following review, the requirements of the UDS SOP were reiterated to recruiting centres via e-mails and a newsletter and also discussed at BUS training days, which were held 3 months and 15 months into the recruitment period. A total of 6 months after the initial audit, a second audit (June 2012 to December 2012) was performed on a further 60 traces.

Audit standards for urodynamic studies

The traces were reviewed to assess the presence of the following criteria:

- adequate subtraction prior to filling cystometry (i.e. initially bladder and rectal catheters should be zero when open to atmosphere, at the level of upper symphysis)
- sitting position during filling cystometry (recommended as per study protocol)
- the cystometry filling rate was 100 ml/minute to begin with (then slowed as and if necessary)
- a cough pre-void
- a cough post void
- presence of one cough per minute to assess ongoing adequate subtraction of intravesical and abdominal pressures.

In addition, agreement of initial diagnosis between study investigator and expert panel was assessed. The expert panel determined a diagnosis of USI, DO, low compliance, VD or normal, or combinations of these diagnoses.

Results

Between May 2011 and December 2012, a total of 124 UDS traces were reviewed against the UDS SOP criteria, *Table 10* illustrates the result of the overall quality control check in UDS.

The initial urodynamic diagnoses given by the study investigator was consistent with the two assessors in the expert panel in 86% of cases, which improved, following the feedback given to all investigators, to 95% of cases. There was also improvement in the compliance with the repeated coughing, which was necessary to calibrate the traces. Despite reiterating that the sitting position is recommended for filling cystometry, compliance decreased following the second audit, although this could be a spurious finding from the small sample reviewed.

Out of 68 patients who had a normal UDS at Birmingham Women's Hospital, 14 accepted the option of having an additional test in the form of ambulatory UDS and, of these, seven had a DO diagnosis.

TABLE 10 Comparison between audit and reaudit compliance of UDS traces

Recommended actions in SOP	Audit May 2011 to May 2012 (%), $n = 64$	Reaudit June 2012 to December 2012 (%), $n = 60$
Adequate baseline zeroing pressures	60 (93.7)	57 (95)
Filling in sitting position	49 (76.5)	38 (63.3)
Cystometry filling rate 100 ml/minute	31 (48.4)	43 (71.6)
Cough pre-void	49 (70.3)	60 (100)
Cough post void	23 (35.9)	45 (75)
Cough per minute	49 (70.3)	58 (96.6)
Agreement between urodynamic diagnoses	55 (85.9)	57 (95)

Discussion

The results of this audit show that reiteration of SOP had a positive impact on the quality of UDS carried out within the BUS. Training and education through the use of e-mails, newsletters and study site visits significantly raised the compliance with the GUP guidelines and the study UDS SOP.

The strengths of this audit include anonymisation of the centre and UDS operator, availability of clear standards in the form of SOPs, predefined audit criteria and the closure of the audit loop via feedback. A larger audit, including all the UDS tests conducted in the BUS could have been carried out to provide greater reassurance, but the DMC were reassured that a representative sample of UDS tests was reviewed and the quality of the UDS was adequate.

Kraus *et al.*[126] have identified that clinical trials using UDS as outcomes require additional standardised procedures to ensure that intersite variability is limited and kept to a minimum. To maintain reliability of urodynamic data, there should be standardisation of urodynamic technique, interpretation and performance. Multicentre urodynamic studies require a continuous quality control process, with audits, multidisciplinary team meetings and refresher training to reinforce the initial training. Our audit results demonstrate that ongoing education and training is paramount in the implementation of clinical guidelines and SOPs for UDS, and via regular communication with the recruiting centres, discrepancies were minimised, resulting in improved reliability of UDS diagnosis. Currently, urodynamic training in the UK is varied and ranges from bedside training, a certificate course, to individual study days. All centres should share the same course content in order to achieve standardisation.

There are a number of factors that may influence the detection of DO.[127] For example, if a poor technique is implemented the results may become compromised or misleading. Emphasis should be made on the patient's position during the filling cystometry phase of the UDS procedure. It is known that most patients with OAB caused by DO complain about their symptoms while in the sitting or standing position.[127] A study by Arunkalaivanan *et al.*[128] found that in 96 women with LUTS, 55% ($n = 53/96$) of them were diagnosed with DO while in the sitting position. In contrast, only 9% ($n = 9/96$) had the same diagnosis while in the supine position. Filling is often performed between 50 ml/minute and 100 ml/minute[96] to improve the DO pick-up rate. In order to correlate symptoms and diagnosis, the filling rate specified within our SOP complied with the GUP guidelines.

Clinicians can maximise the chance of reproducing the patient's symptoms by using provocation measures such as coughing, running hands under water and jogging.[96,128]

Conclusion

The results of this audit have demonstrated a positive change in urodynamic practice as a direct result of reinforcing the importance of adhering to guidelines of clinical practice and procedures, and that the UDS process was judged as adequate by the DMC. The results of the audit and a reminder about the SOP and GUP recommendations were shared with the BUS investigators via newsletters, e-mails, site visits and training days. Reiterating information and results can increase the level of adherence in the technique of performing the test. Ongoing audits of the technique and interpretations of UDS are necessary to maintain the high standards and reliability of the test.

Chapter 4 Reproducibility of bladder wall ultrasonography

Aims

The aim of the repeatability and reproducibility studies was to assess whether or not measurements of BWT made using transvaginal ultrasonography were adequately reliable and reproducible, and thus likely to detect differences in BWT potentially indicative of disease. There were three key objectives:

1. To estimate the intraobserver measurement error in interpreting images by comparing blinded duplicate assessments of images by a single observer.
2. To estimate the interobserver measurement error in interpreting images by comparing blinded duplicate assessments of images by different observers.
3. To estimate the interobserver measurement error in the complete scanning and interpretation process by comparing measurements made by different observers using different scans made on women on two separate occasions.

Methods

Participants were a subset of those who were recruited to the BUS (for eligibility criteria see *Chapter 2*, *Methods*). On days in which a second observer was available, women who agreed to have two transvaginal scans by different operators were recruited into the study evaluating the reproducibility of scans. For the studies evaluating the reproducibility of the interpretation of scans, random selections of images were sent by the recruiting centres at the request of the trial co-ordinator.

Measurements of BWT at the trigone, dome midline and anterior wall midline were made as per a SOP using a two-dimensional transvaginal end-firing probe (detailed in *Chapter 2*, *Index test: bladder wall thickness via ultrasonography* and also see *Figure 3*). The reported BWT measurement was defined as the mean of the measurements made at the three locations [(trigone + dome midline + anterior wall midline)/3]. The process of measurement requires placing a calliper reference point on the image, using a mouse-operated cursor on the electronic image, at the interface between the bladder wall and the adjacent tissue or lumen. The computer software for the ultrasonography machine calculates the BWT using internal calibration algorithms and reports the thickness in millimetres. Images can be saved with and without the calliper points and reported thickness.

Three substudies were undertaken to address the three objectives.

In substudy A, BWT was measured on 37 ultrasound images from individual participants. Repeat measurements were made by the same observer on the same images 6–12 months later. All second measurements were made blind to the original measurement, using images without calliper marks. All images were from the Birmingham Women's Hospital and were measured on the scan machine. The measurement process was the same from the beginning to the end of the study and includes repeat measurements. All measurements complied with the SOP.

In substudy B, BWT was measured on ultrasound images from 57 individual participants by a single observer. Repeated measurements were made by one of a further five different observers on the same images, such that there were duplicate measures for each image (*Table 11*). All second measurements were made blind to the original measurement, using images without calliper marks. Images from the

TABLE 11 Details the experience of the observers in the substudies

Study	Observer and location	'Skill level' (years of experience in obstetrics and gynaecology)
Substudy A	Observer 1: Birmingham ($n = 37 \times 2$)	15
Substudy B	Observer 1: Birmingham ($n = 57$)	15
	Observer 2: Birmingham ($n = 34$)	10
	Observer 3: Birmingham ($n = 4$)	5
	Observers 4–6: Bournemouth ($n = 7$), St Mary's ($n = 4$), Medway ($n = 8$)	7–20
Substudy C	Observer 1: Birmingham ($n = 27$)	15
	Observer 2: Birmingham ($n = 16$)	15
	Observer 3: Birmingham ($n = 11$)	10

Birmingham Women's Hospital were measured on the scan machine and those performed elsewhere were measured using Digital Imaging and Communications in Medicine (DICOM) viewing software.

In substudy C, 27 women underwent two separate ultrasonography sessions undertaken by different observers. The second scan and measurements of BWT were made blind to measurements of the first observer. Three observers were used in total; all women received scans from observer 1 and also either observer 2 or observer 3. The experience of the ultrasonography operators is detailed in *Table 11*.

Statistical analysis

For each study, BWT measurements were analysed using one-way analysis of variance (ANOVA). One-way ANOVA decomposes the total variation observed (SD_T^2) into that originating from differences between women (SD_I^2 – individual variability) and that caused by the measurement process (SD_A^2 – analytical variability). The estimates are linked as $SD_T = \sqrt{SD_A^2 + SD_I^2}$. The SD for analytical variability, SD_A estimates the measurement error.

Two further statistics were computed from these values. The intraclass correlation coefficient (ICC) describes the fraction of the total variance in BWT measurements owing to individual rather than analytical variation (SD_I^2 / SD_T^2). ICC values lie between zero and one; measurements that are reliable have ICCs approaching one, as the signals (the individual variation) dominate the noise (the analytical variation).

The repeatability coefficient describes the smallest real difference that can be detected with a specified degree of certainty between two measurements and is computed as $\sqrt{2}Z\sqrt{SD_A^2}$ (in which Z takes the value of 1.96 for a difference which has 95% certainty of being a real effect and not measurement error). Smallest real difference values are given in the units of the original measurement.

The above analyses were all undertaken assuming exchangeability of observers, that is, that the ordering of the measurements has no relevance. Generalisability of these findings relies on the observers being presumed to be representative of those who would make the measurements in practice.

For the purpose of graphical display only, we also display the data in scatterplots and Bland–Altman plots to demonstrate the distribution of measurements and differences between measurements (see *Figures 8–13*). For these analyses assignment of measurements to particular observers is important. In substudy A, there is a logical choice for the first and second measurements, and the distribution of these differences is of interest. In substudies B and C, measurements made by observer 1 were arbitrarily taken as the first measurement (referred to in *Figures 10* and *12* as operator A1) and remaining observers 2 to 6 were taken as the second measurement (referred to in *Figures 10* and *12* as operators A2 or A3, respectively).

Results

A total of 121 women took part in the substudies. The distribution of BWT measures are shown in *Table 12*. The mean and SD of the BWT measures in each substudy were similar to that of the BUS cohort. Ranges in the substudies were lower, which is expected as ranges increase with sample size.

Substudy A: intraobserver repeatability of the same scans

Paired measurements were available for 37 women. The scatter of measurements is shown in *Figure 8* and the distribution of differences in measurements in *Figure 9*. The later measurements were, on average, higher than the earlier measurements by 0.35 mm (95% CI 0.19 mm to 0.51 mm; $p < 0.0001$) but without any evidence of a relationship between error and the mean BWT value. Differences in measurements of up to 1.5 mm were observed.

The SD for the analytical variation for intraobserver variability was estimated as 0.42 mm (*Table 13*). This level of variability compares with a SD of 1.04 mm between individual differences, thus 86% of the total variability observed is attributed to individual variability and 14% to measurement error. With this level of measurement error, differences of over 1.16 mm are 95% likely to be real for this single assessor.

TABLE 12 Comparison of distribution of BWT measures between the full bladder ultrasonography and the substudies

		BWT (mm)			
Study	Observer	*n*	Mean	SD	Range
BUS	Multicentre study	645	4.78	1.34	1.07–9.60
Substudy A	Measurement 1	37	5.60	1.14	3.40–7.73
	Measurement 2	37	5.95	1.08	4.00–8.03
Substudy B	Observer 1	57	5.05	1.30	2.10–7.53
	Observers 2–6	57	5.07	1.26	2.27–7.33
Substudy C	Observer 1	27	4.86	1.04	2.77–7.17
	Observers 2–3	27	4.73	1.38	1.83–7.43

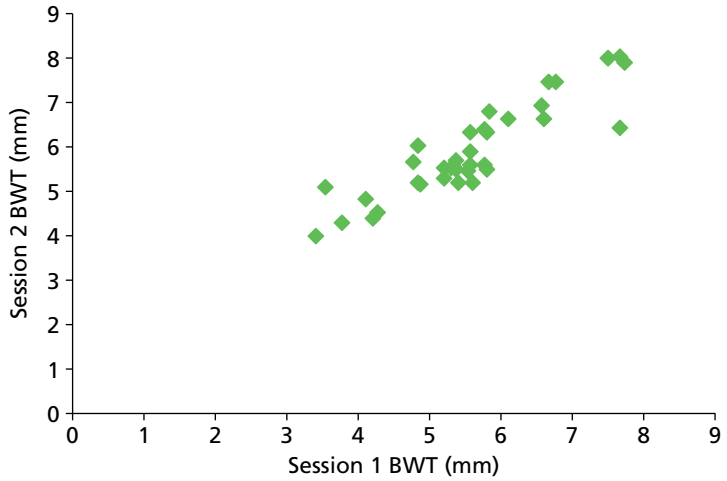

FIGURE 8 Scatterplot for substudy A.

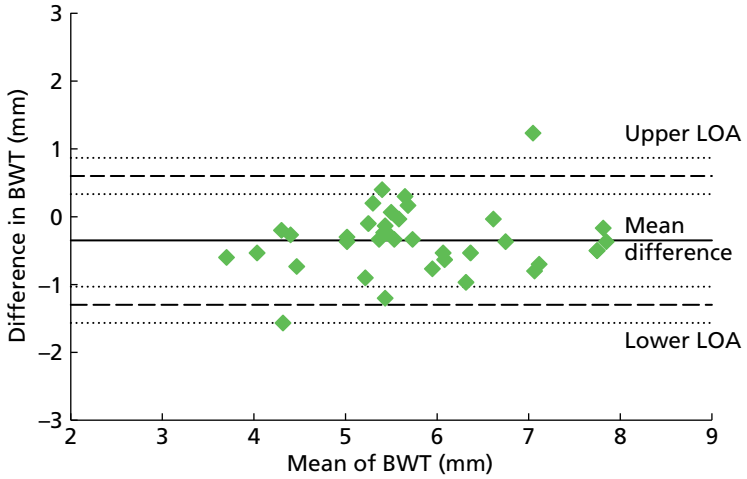

FIGURE 9 Bland–Altman analysis for substudy A (limits of agreement and 95% CIs shown). LOA, limit of agreement.

TABLE 13 Estimates of measures of analytical and individual variability

Substudy	Individual variability SD (mm)	Analytical variability SD (mm)	ICC (95% CI)	Smallest real difference (mm)
Substudy A (intraobserver interpretation of same scans)	1.04	0.42	0.86 (0.75 to 0.92)	1.16
Substudy B (interobserver interpretation of same scans)	1.23	0.35	0.93 (0.88 to 0.96)	0.97
Substudy C (interobserver measures from repeated scans)	0.95	0.76	0.61 (0.32 to 0.80)	2.11

Substudy B: interobserver repeatability of the same scans

Paired assessments were available for 57 women made by six different observers. The distribution of measurements and differences are shown in *Figures 10* and *11*. Differences as large as 2 mm were observed.

The SD for the analytical variation for intraobserver variability was estimated as 0.35 mm (see *Table 13*). This level of variability compares with a SD of 1.23 mm between individuals, thus 93% of the total variability observed is attributed to individual variability and 7% to measurement error. With this level of measurement error, differences made by assessors similar to these would need to be at least 0.97 mm to be 95% likely to be real.

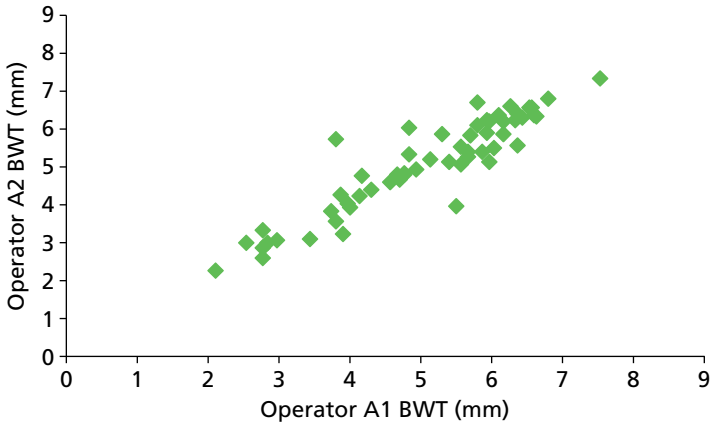

FIGURE 10 Scatterplot for substudy B.

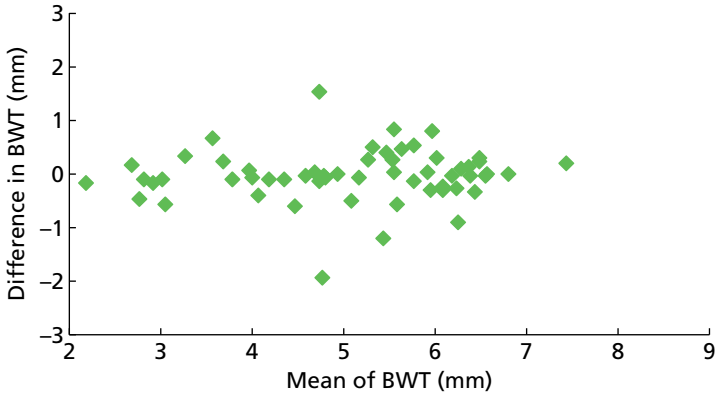

FIGURE 11 Bland–Altman analysis for substudy B.

Substudy C: interobserver repeatability of different scans

Paired measurements were made for 27 women using three different observers. The design of substudy C included estimation of variation occurring from repeated scans together with the interpretation of scans. The distribution of measurements is shown in *Figure 12* and distribution of differences in *Figure 13*. Maximum differences were again around 2 mm, but they were more common in this substudy than in previous substudies A and B.

The SD for the analytical variation for intravariability of repeated scans was estimated as 0.76 mm (see *Table 13*). This level of variability compares with a SD of 0.95 mm between individuals, thus 61% of the total variability observed is attributed to individual variability and 39% to measurement error. With this level of measurement error, differences made by assessors similar to these would need to be at least 2.11 mm to be 95% likely to be real.

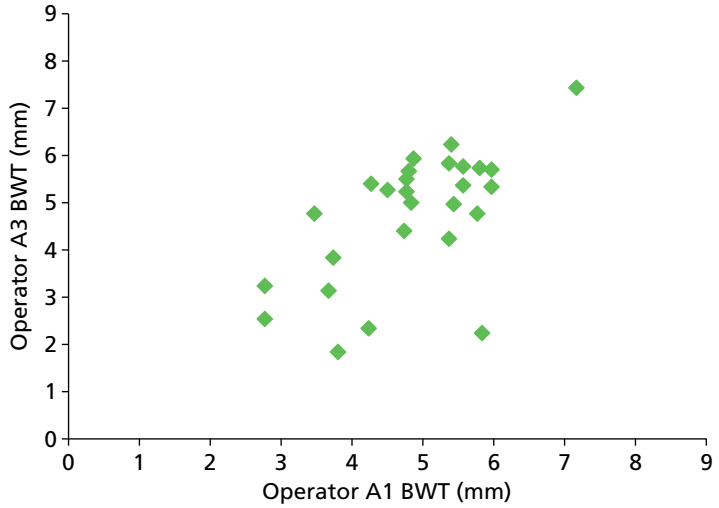

FIGURE 12 Scatterplot for substudy C.

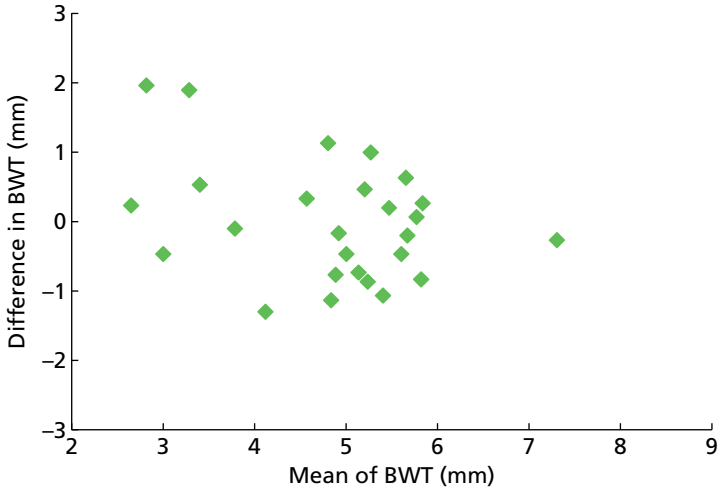

FIGURE 13 Bland–Altman analysis for substudy C.

Discussion

Main findings

We undertook three separate studies to investigate the reliability and repeatability of measurements of BWT using transvaginal ultrasonography. Our analyses have found that differences of < 2 mm in BWT cannot be safely interpreted as indicating real differences in BWT, such differences are in the realms of those attributable to analytical variability (measurement error). The substudies were also designed to identify the magnitude of the possible sources of the analytical variability. We observed that the process of interpreting scans introduces measurement error of around 1 mm, suggesting that the remaining 1 mm is attributable to a combination of the scanning process and biological variability.

We did not assess whether or not the differences in interpreting the scans arose because of within-observer or between-observer variability. Surprisingly, our estimate of intraobserver variation is greater than that of interobserver variation. As the study samples are not large, and different scans were assessed for these repeat intraobserver measurements, this observation potentially could be explained by the play of chance or confounding. However, problems were also experienced with the quality of recorded images used in substudies A and B from the Birmingham Women's hospital which were not saved using DICOM software, which may explain this observation and may also explain the finding that second reads of scans in substudy A gave, on average, BWT measures of 0.6 mm greater than the original.

Findings in the context of existing data

Six previous studies on reliability and reproducibility of BWT have used a variety of ultrasonography techniques, including transabdominal[56] and translabial[51] as well as transvaginal[44,56,72,129] (as summarised in *Appendix 8*). Several of these studies investigated women with different or a mixture of aetiologies, or used levels of bladder filling. Of the transvaginal ultrasonography studies, only two[56,72] included assessment of repeated scans as in our third substudy, evaluating 10 and 25 women, respectively. The other two studies[44,129] only evaluated the reproducibility of image interpretation (as in our first and second substudy) based on repeated assessment of 10 and 1544 images, respectively.

Comparison of findings between these previous studies and the BUS is complicated owing to the common inappropriate use of Bland–Altman analyses and Pearson's correlation coefficients. Neither of these methods directly estimates the degree of analytical variability, allowing assessment of the reproducibility of the measurement and the signal-to-noise ratio, although a pseudo estimate of analytical variability can be computed from the SD of the differences. Reporting of the study design and statistical analysis was often incomplete or ambiguous. Khullar *et al.*,[44] Panayi *et al.*[72] and Tubaro *et al.*[129] all reported Bland–Altman analyses from which pseudo estimates of measurement error have been computed (see *Appendix 8*).

Estimates of the SD of analytical variability vary between 0.3 mm and 1.3 mm for interobserver and intraobserver variation for image interpretation (compared with 0.3 mm and 0.4 mm for the BUS), and a SD of 0.4 mm for interobserver variation for repeated scans (compared with 0.8 mm for the BUS)[56] only reported Pearson's correlation coefficients from which no useful measures of reproducibility can be obtained.

Strengths and limitations of the study

The women in our study were recruited prospectively as part of the BUS, which involved good characterisation of symptoms and disease state. They were recruited from standard NHS incontinence clinics and are thus highly likely to be representative of women in whom BWT would be measured as part of the diagnosis of DO. The subsamples in each study appear to be representative of the larger cohort.

Bladder wall thickness measurements were made according to a standardised protocol implemented following a programme of rigorous investigator training implemented in the larger study, which will have minimised variability due to differences in technique.

Both the number of women and the number of assessors limit the precision of the estimates made, although the study was larger than many previously undertaken. The assessors who partook in the intraobserver studies generally had high levels of experience and expertise with the techniques, such that the estimates of operator-dependent analytical variability may be lower than those in standard practice.

In substudies A and B, the use of stored images of the original bladder transvaginal scan was problematic because those not stored using DICOM software were of poorer quality than the original images, as reported in previous interobserver variation studies using ultrasonography.[130] We found that the brightness of echogenic serosa and mucosa were reduced in the stored ultrasound images, making the bladder wall less distinct and reducing the ability to place the callipers accurately. This may have led to overestimation in analytical variation in substudies A and B, particularly in A as no images were stored using DICOM.

Interpretation

Ultrasonographic measurements of BWT have a high level of analytical variation arising from the scanning technique, underlying biological variability and interpretation of images, such that only differences > 2 mm should be interpreted as indicating real changes in BWT. The range of BWT measurements observed in the full cohort ranged from 1 mm to 10 mm. To illustrate the potential impact of measurement error of this magnitude, if a threshold of 5 mm is used to define test positives, those between 3 mm and 5 mm could be misclassified as test negatives through measurement error, and those between 5 mm and 7 mm could be misclassified as test positives. In the BUS cohort of 645 women, 326 had values between 3 mm and < 5 mm, and 217 had values between 5 mm and < 7 mm. These groups constitute 84% of the complete sample. Only 41 women (6%) had BWT measures of ≥ 7 mm and 61 (10%) had measurements of < 3 mm. Thus, for the majority of women included in the study there is a possibility that a transvaginal ultrasonographic measurement of BWT would misclassify them using a 5 mm threshold owing to analytical variation. Rates of potential misclassification for higher or lower thresholds would be lower, but still substantial.

Conclusion

In the presence of high levels of analytical variation for a relatively small measurement of BWT, it is unlikely that BWT measurement on a transvaginal ultrasound has sufficient reliability and reproducibility to be an accurate diagnostic test.

Chapter 5 A comparative evaluation of patient acceptability of bladder ultrasonography scanning and urodynamics

Introduction

Objectives

In this chapter we describe a comparative evaluation of the acceptability of performing both transvaginal bladder ultrasonography and UDS from the patient's perspective through the completion of self-reported questionnaires.

Methods

The participants in this substudy were the 687 patients who took part in the BUS (see *Chapter 2*). They underwent UDS and bladder ultrasonography in the participating centres and, when possible, these were carried out on the same day. If it was not possible for both tests to be performed on the same day, they had to be completed within a 4-week period. Immediately after each test, acceptability questionnaires were given to the participant for completion. Items included in the questionnaire were as follows:

- Pain measured using a VAS on a 0- (no recorded pain) to 100- (worst pain imaginable) point scale in relation to pain during and shortly after testing.[131]
- Acceptability of testing measured on ordinal response Likert scales.
- State–Trait Anxiety Inventory six-item short form (STAI-6) a six-item measure of generic state anxiety. This is a validated and widely accepted instrument used to assess the intensity of current feelings in relation to how you 'generally feel today'.[132] Scores range from 4 (most positive) to 24 (most negative). The short form was used to improve patient compliance, as opposed to the long form, and to accommodate time constraints in busy clinics.

Analysis

All responses were compared using paired (UDS vs. bladder ultrasonography) methods.[133] For VAS and STAI-6 scores, mean differences and 95% CIs were calculated with statistical significance determined by a paired *t*-test. Wilcoxon signed-rank test was used for ordinal responses and McNemar's test for binary responses. Analysis was performed using SAS software, version 9.2 (SAS Institute Inc., Cary, NC, USA).

Results

At least 94% (646/687) of the participants in the BUS responded to the questions on pain and acceptability in the two tests, and 87% (602/687) responded to the anxiety questionnaire.

Pain during and after the 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; these differences were statistically significant (*Table 14*).

TABLE 14 Pain scores during and after the UDS and bladder ultrasonography tests (minimum 0, maximum 100)

Time point	Mean UDS (SD), *n*	Mean ultrasonography (SD), *n*	Mean difference (95% CI)	*p*-value[a]
During	28.1 (28.2) 643	12.7 (19.4) 643	15.3 (13.1 to 17.6)	< 0.001
After	21.2 (26.8) 639	7.9 (16.1) 639	13.3 (11.2 to 15.4)	< 0.001

a From paired *t*-test.

Acceptability

Bladder ultrasonography was rated as more acceptable to the women than UDS ($p < 0.001$), with the proportion of women who found the test totally acceptable significantly higher (81% vs. 56%). The number reporting an unacceptable test was still relatively low following UDS (2%). More women found the exposure required for UDS more embarrassing than bladder ultrasonography (proportion reporting some embarrassment 64% vs. 48%; $p < 0.001$). Fewer women felt that they would recommend UDS to a friend than bladder ultrasonography (86% vs. 96%; $p < 0.001$) and have the same test again (88% vs. 97%; $p < 0.001$) (*Table 15*).

Anxiety

Anxiety levels associated with both tests appeared moderate, although the scores were slightly higher with UDS (*Table 16*).

TABLE 15 Acceptability of UDS and transvaginal bladder ultrasonography tests

Question	Response	UDS ($n = 645$) frequency (%)	Ultrasonography ($n = 645$) frequency (%)	*p*-value[a]
Procedure acceptability	Totally	360 (56%)	519 (81%)	< 0.001
	Generally	271 (42%)	123 (19%)	
	Unacceptable	12 (2%)	1 (< 1%)	
	Missing response	2	2	
Exposure for test embarrassing?	Extremely	54 (8%)	21 (3%)	< 0.001
	Moderately	125 (19%)	63 (10%)	
	A little	234 (36%)	227 (35%)	
	No	232 (36%)	334 (52%)	
	Missing response	–	–	
Recommend test to a friend?	Yes	551 (86%)	621 (96%)	< 0.001
	No	93 (14%)	23 (4%)	
	Missing response	1	1	
Have same test again?	Yes	564 (88%)	623 (97%)	< 0.001
	No	78 (12%)	19 (3%)	
	Missing response	3	3	

a From paired tests (see *Chapter 5, Analysis*).

TABLE 16 State–Trait Anxiety Inventory six-item short form scores following each test

Mean pain score for UDS (SD, *n*)	Mean pain score for ultrasonography (SD, *n*)	Mean difference in pain scores (95% CI)	*p*-value[a]
12.9 (3.8, 584)	12.7 (3.8, 584)	0.3 (0.1 to 0.5)	0.02

a From paired *t*-tests

Discussion

Our results show that bladder ultrasonography was more acceptable and less embarrassing and painful than UDS. Despite this, a high proportion of women would recommend the UDS test to a friend (86%) and also have it repeated (88%). Anxiety scores were also higher for UDS than bladder ultrasonography, but the mean difference appeared small (0.3 points on a 4- to 24-point scale). To our knowledge, this is the first formal evaluation of the comparison of tolerability and acceptability of various diagnostic procedures to evaluate bladder function.

The question is whether or not UDS is likely to be deemed acceptable in terms of having it repeated in view of its clinical importance. Women were aware that their treatment plans were based on the UDS diagnoses and not that of bladder ultrasonography. This awareness of importance may have contributed to the improved acceptability of having this test repeated again if necessary. One patient made the comment below regarding the acceptability of UDS:

> *Although BUS [bladder ultrasonography] was a generally more acceptable test, I felt that because UDS physically replicated my symptoms, it allowed me a better understanding of my condition.*

Pain was higher with UDS although the average score was only 28 on a scale of 0–100. The origin of pain may be multifactorial. Pain could be caused by physical components such as the urethral and rectal catheterisation and the artificial filling of the bladder. Following UDS, participants often expressed urethral pain. The exact mechanism of pain perception after the UDS is not known; however, slight trauma to the urethra may be considered to be an aetiological factor.[134]

Elevated anxiety levels during UDS may also have contributed to patient's perception of pain. Women's concerns with invasive procedures include whether or not the examination will be painful and/or uncomfortable.[132] Expectation of pain is associated with greater anxiety in dental patients[135] and minor surgery patients.[136] Pain and discomfort associated with medical procedures may be mediated by fear of pain.[137] Fear of pain refers to trait-like fear responses to painful situations and is a key component in medical fears;[138] this may also influence willingness for invasive testing and follow-ups.[139] The perceived level of pain was strongly correlated with the level of apprehension and embarrassment during different steps of UDS. Younger age and apprehension were found to be significant risk factors for the heightened perception of pain on multivariate linear regression analyses.[140]

Our results on pain/anxiety and embarrassment provoked are similar to those mentioned in the literature. In a prospective study of 208 patients, although UDS was only associated with minor complications, it was perceived to be painful, worrying and traumatic.[38] In a study of pre-test and post-test evaluation of anxiety with UDS, severe pre-test anxiety was reported in only a small fraction of women (4.6%) undergoing UDS. Following completion of UDS, 77.5% of women indicated that they were not at all or slightly anxious, not at all or slightly embarrassed (84.1%) and experienced no or slight physical discomfort (75.5%).[141] Younger age, history of anxiety or depression and a diagnosis of OAB and painful bladder syndrome may lead to more negative experiences during UDS.[41]

Strengths and limitations

A strong component of the study was that a large number of patients were recruited and responded to the questionnaires in the study (at least 87% of the 687 recruited). The population were derived from several geographical areas and various ethnicities within the UK (see *Chapter 2* for full details). Data collected from multicentre studies may be more applicable and generalisable than that collected from a single centre. Our study provides information to assist in counselling women who may be apprehensive or anxious regarding an invasive test such as UDS.

The study had some limitations. We administered the instrument to measure anxiety only after each test procedure. We could have measured the difference between pre-test and post-test questionnaires to know the anxiety provoked by each procedure. This would have yielded valid data about fluctuations in anxiety state before and after each test and a comparison of the difference would have been ideal.

Participants were also aware of the fact that the diagnoses and management plan were made on the basis of the information gained during UDS. This knowledge of the test may have introduced bias and influenced the participant's decision to recommend the test or have it repeated again if required. Bladder ultrasonography was used for assessment only and did not aid in the management of a bladder diagnosis.

We used the short-form anxiety questionnaire in this substudy, which only measures state and not trait anxiety. Trait-anxious individuals tend to respond to stressful situations with increases in state anxiety. The higher the levels of trait anxiety, the more likely it is that an individual will experience anxiety in a variety of situations, relative to individuals low in trait anxiety.[142] Anxiety and urgency incontinence appeared to exacerbate each other.[143] High anxiety score on the STAI-6 was a predisposing factor for UI.[144] UI is associated with a reduced QoL and increased anxiety among community-dwelling elderly women.[145] Given the strong relationship between state and trait anxiety, women with higher trait anxiety levels might have experienced greater state anxiety with UDS than bladder ultrasonography.

In addition, we have not studied the impact of the information given to the women before the procedure and how well we prepared the women for each diagnostic test. We sent out information leaflets about both UDS and bladder ultrasonography along with the appointment letters. During the clinic visit, we reinforced the patient's understanding of each procedure. This is our routine practice based on previous evidence that women were likely to find the test less distressing when they felt they had been given adequate information about it.[134]

Interpretation of findings

Our findings of increased pain and embarrassment perceived during UDS highlight the importance of clarifying the role of UDS in the management of women with OAB. The results may also be used by those preparing information material for women undergoing UDS and/or transvaginal ultrasonography, enabling them to give women a realistic picture of how they might feel.

Conclusion

The UDS procedure had statistically significant higher levels of pain and a lower rate of acceptability than the bladder ultrasonography procedure. In spite of this, the majority of the women would have repeat UDS if needed.

Recommendations for future research

The elevated anxiety levels elicited by invasive diagnostic testing on a background of increased trait anxiety due to LUTS need careful evaluation and interpretation. This may improve the support patients receive during invasive diagnostic testing and subsequent patient satisfaction.

Chapter 6 The impact of urodynamics on treatment and outcomes

Introduction

There is much debate about the role of UDS in the workup of patients with incontinence symptoms. There is uncertainty about (1) its diagnostic accuracy (e.g. does it accurately identify those with DO?), (2) its position in a diagnostic algorithm (e.g. should it be a used to screen all patients, or be reserved for selected use, such as patients scheduled to have surgery?) and (3) diagnostic impact (e.g. are patients better off as a result of this test?). NICE guidelines[36] advise against the use of multichannel cystometry, ambulatory UDS or video UDS prior to commencing conservative management for OAB. However, there are recommendations to perform multichannel cystometry and offer BTX-A and sacral neuromodulation-like invasive treatments in the presence of DO.[36]

There are some studies to suggest that patient-related outcomes (i.e. diagnostic impact) are similar whether or not there is urodynamic diagnosis of DO in patients with OAB who undergo various therapies.[83–86,146] There is a necessity to establish the role of UDS and its impact on treatment and patient outcomes in OAB as at present its role is unclear.

The aim was to establish if treatment pathways and outcomes differed following findings on UDS. More specifically, the following objectives were addressed:

- Does the UDS diagnosis affect treatment pathways (therapeutic impact)?
- What were the patient-reported outcomes in the cohort of patients recruited into the BUS, as measured by a global impression of improvement question (have your bladder problems improved since the tests?) and the ICIQ-OAB questionnaire, at 6 and 12 months after testing?
- Does the diagnosis by UDS have any effect on symptoms after 6 and 12 months, that is, can UDS predict improvement in different patient groups?
- Does receiving a medical or surgical treatment concordant with the UDS diagnosis improve patients' symptoms, compared with not receiving a concordant treatment?
- Are presenting symptoms related to outcomes at 6 and 12 months?
- Does transvaginal ultrasonographic measurement of BWT have any prognostic value?

Methods

Patients recruited to the BUS completed symptom questionnaire booklets pre-test and then were sent booklets at 6 and 12 months after the tests. If no response was received within 4 weeks, a reminder was sent. Patients who did not respond within 6–8 weeks of the initial request were contacted by telephone and questionnaires were completed in a telephone interview with a member of the research team.

In order to establish if a diagnosis by UDS had an effect on subsequent treatment management and long-term patient outcomes, the data were simplified using the following criteria:

- Diagnoses from UDS testing were categorised as (1) DO + USI, (2) DO alone, (3) USI alone and (4) patients with no demonstrable findings were defined as normal UDS.
- Mid-urethral slings, bladder neck injection and colposuspension are all categorised as 'USI surgery'. These interventions were considered concordant with a UDS diagnosis including USI, and discordant otherwise.
- Nerve modulation (PTNS or sacral neuromodulation) and BTX-A toxin injections into the bladder are categorised as 'DO surgery'. These interventions were considered concordant with a UDS diagnosis including DO, and discordant otherwise.
- The use of antimuscarinics and mirabegron are referred to as 'bladder relaxants'. These medical treatments were considered concordant with a UDS diagnosis including DO or normal UDS, and discordant otherwise.
- No treatment was considered concordant with a UDS diagnosis of normal UDS, and discordant otherwise.
- The effect of bladder relaxant tablets was assumed to provide relief from symptoms at only the time point it was reported, whereas the effect of surgery was assumed to be longer term (i.e. those reporting surgery at 6 months were presumed to still have had surgery at 12 months even if not reported again).
- The number of patients reporting having had both surgery and bladder relaxants as a proportion of those having surgery was in the minority (15% at 6 months, 26% at 12 months). To simplify the analysis, this dual effect was ignored in the analysis of concordant treatments (i.e. for women categorised as having SUI or DO surgery this may or may not include the use of bladder relaxants). Surgery was assumed to supersede the use of bladder relaxants.

The relationship between UDS diagnosis (DO + USI/DO/USI/normal UDS) and subsequent treatment (DO surgery/USI surgery/bladder relaxants/no treatment) was examined using a multinomial logistic regression model[147] with the treatment group variable as the outcome and UDS diagnosis as the explanatory variable. The overall importance of this variable was determined by chi-squared test with results presented alongside estimates of OR and 95% CIs with treatment group; 'no treatment' used as the reference variable.

Frequencies and percentages are presented for the results of the global impression of improvement question measured at 20 months, with means and SDs presented for ICIQ scores [assessment of symptoms of OAB on QoL, scores on a 0 (best) to 16 (worst) scale] at 7 and 20 months. Mean change 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 (over the responses at baseline, 7 and 20 months) regression models[148] for the global impression of improvement and ICIQ-OAB scores, to examine the effect of the following variables: UDS diagnosis (as listed above), concordant treatment (binary: yes/no), presenting symptoms (binary: urgency + SUI/urgency +/– urgency incontinence alone) and BWT (continuous). The statistical importance of these explanatory variables was determined by chi-squared and *t*-tests. Comparative ORs and mean differences between these subgroups were calculated along with 95% CIs using the standard error taken from the respective regression models. Interaction between the concordant treatment and BWT variables with the UDS diagnosis and presenting symptoms variables was examined by including the corresponding variables and interaction variable terms in the regression models. The statistical importance of the interaction terms was determined by *F*-test. In a similar fashion interaction between the aforementioned variables and time with respect to ICIQ scores was also examined.

Results

Follow-up

The 6- and 12-month follow-up questionnaires were received at a median time of 7 and 20 months (IQRs 6–8 and 15–24 months, respectively) after tests. Results are henceforth referred to as at 7 and 20 months. The study recruited a total of 687 women and questionnaire booklet booklets were returned by 489 (71%) and 475 (69%) participants at the 7- and 20-month time points (*Figure 14*).

Urodynamics diagnosis

Of the total number recruited, 666 (97%) had complete UDS testing. On performing UDS, 43% ($n = 284$) were diagnosed with DO alone, 17% ($n = 115$) had DO + USI, 14% ($n = 92$) were diagnosed with USI alone and 26% ($n = 175$) had normal UDS. *Table 17* summarises these frequencies by presenting symptoms.

Effect of urodynamics on the management offered

Over the whole follow-up period, the majority of women reported some treatment (292/467, 63%). Seventy per cent of these treatments (205/292) were reported as bladder relaxants only, 20% as USI surgery (57/292) and 10% as DO surgery (30/292). Surgery was used in isolation in 66% of cases (57/87) and in combination with bladder relaxants in the other cases. Further details are given in *Table 18* (if bladder relaxants were used in combination with surgery these are indicated in brackets, otherwise surgery was used in isolation).

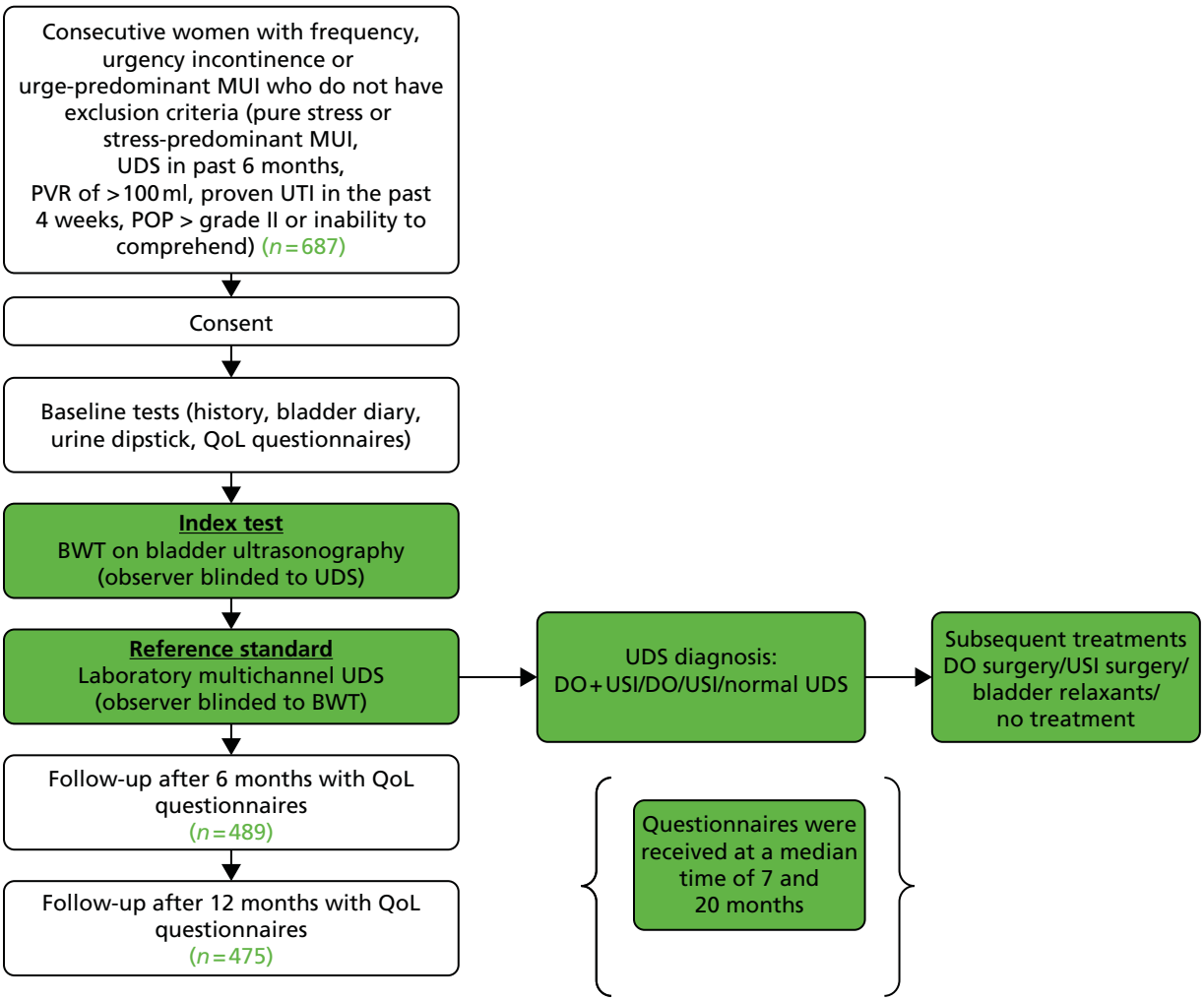

FIGURE 14 Study flow chart of follow-up of participants after 6 and 12 months.

TABLE 17 A UDS diagnosis by presenting symptoms (as indicated through clinical history)

	Presenting symptoms (from clinical history), *n* (%)		
UDS diagnosis	Urgency + stress	Urgency alone	Total
DO + USI	96 (23)	18 (8)	114
DO	160 (38)	120 (50)	280
USI	80 (19)	11 (5)	91
Normal UDS	83 (20)	91 (38)	174
Total	419	240	659[a]

a Seven patients had missing presenting symptoms.

TABLE 18 Reported interventions over the whole follow-up period by UDS diagnosis

	Treatment frequency, *n* (%)				
UDS diagnosis	DO surgery (including bladder relaxants)	USI surgery (including bladder relaxants)	Bladder relaxants only	No treatment	Total
DO + USI	3 (3)	27 (11)	29	23	82
DO	19 (5)	6 (3)	119	57	201
USI	2 (0)	18 (6)	16	25	61
Normal UDS	6 (1)	6 (1)	41	70	123
Total	30	57	205	175	467[a]

a Eight participants returned follow-up forms but did not complete treatment information.

Overall, subsequent treatment was highly associated with diagnosis group ($p < 0.0001$) suggesting that the clinicians and patients appeared to be guided in part by UDS diagnoses in selecting treatment options. For example, the odds of having USI surgery were increased by 15-fold if UDS diagnosis was DO + USI and the odds of DO surgery or bladder relaxants were increased by threefold if UDS diagnosis was DO (*Table 19*).

Long-term symptom responses in the overall group

Fifty-three per cent (248/470) of the participants thought that their bladder problems had improved since going for the tests at 20 months (global impression of improvement question). Analysis of ICIQ responses showed improving scores reflecting declining severity of patients' symptoms (*Table 20*). The reduction in the mean ICIQ scores over time from baseline is depicted in *Figure 15*.

TABLE 19 Odds of intervention vs. no treatment over the whole period using the normal UDS diagnosis group as a reference

	Treatment		
Diagnosis effect	DO surgery, OR (95% CI)	USI surgery, OR (95% CI)	Bladder relaxants, OR (95% CI)
DO vs. normal UDS	2.6 (1.2 to 5.4)	1.0 (0.3 to 2.7)	3.3 (2.4 to 4.7)
DO + USI vs. normal UDS	0.8 (0.2 to 2.8)	14.9 (6.6 to 33.8)	2.0 (1.2 to 3.2)
USI vs. normal UDS	1.0 (0.3 to 3.2)	8.2 (3.5 to 19.3)	1.1 (0.7 to 1.9)

TABLE 20 ICIQ scores from baseline to 20 months

Time point	Mean ICIQ (SD), *n*	ICIQ change from baseline: mean (95% CI), *p*-value
Baseline	9.3 (2.7), 637	
7 months	7.3 (3.3), 469	–1.9 (–2.2 to –1.6), $p < 0.0001$
20 months	7.0 (3.5), 460	–2.2 (–2.5 to –1.9), $p < 0.0001$

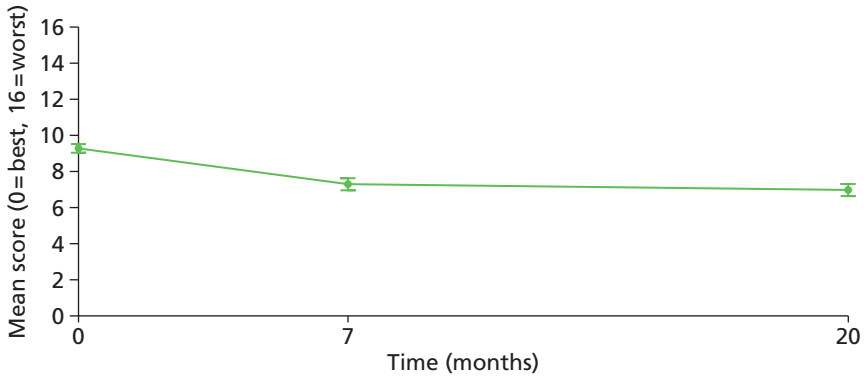

FIGURE 15 Overall ICIQ scores over time (95% CIs are shown at each time point).

Long-term responses by urodynamics diagnosis

Positive responses to the global impression of improvement question were higher in the USI (35/56, 63%) and DO + USI (48/83, 58%) diagnosis groups than the DO (104/205, 51%) and normal UDS (60/125, 48%) groups, although the importance of diagnosis group was not statistically significant overall ($p = 0.2$). ICIQ scores were reduced from baseline in all groups at 7 and 20 months ($p < 0.001$ or less; *Table 21* and *Figure 16*). There was some evidence that ICIQ responses varied between diagnosis groups overall ($p = 0.02$) and pairwise comparisons between them indicated that the DO + USI group have a greater reduction than the DO group (–1.1 points, 95% CI –1.7 points to –0.4 points; $p = 0.002$) over both time points. There were no statistically significant differences between the other groups.

TABLE 21 ICIQ scores by diagnosis groups

	UDS diagnosis			
Time point	DO + USI	DO	USI	Nothing
Baseline mean (SD), *n*	9.8 (2.5), 92	9.8 (2.7), 236	8.9 (2.7), 67	8.2 (2.3), 144
7 months mean (SD), *n*	6.8 (3.5), 81	8.1 (3.4), 207	7.1 (3.3), 57	6.5 (2.8), 124
20 months mean (SD), *n*	6.6 (3.6), 82	7.7 (3.8), 199	6.2 (3.3), 55	6.5 (3.1), 123

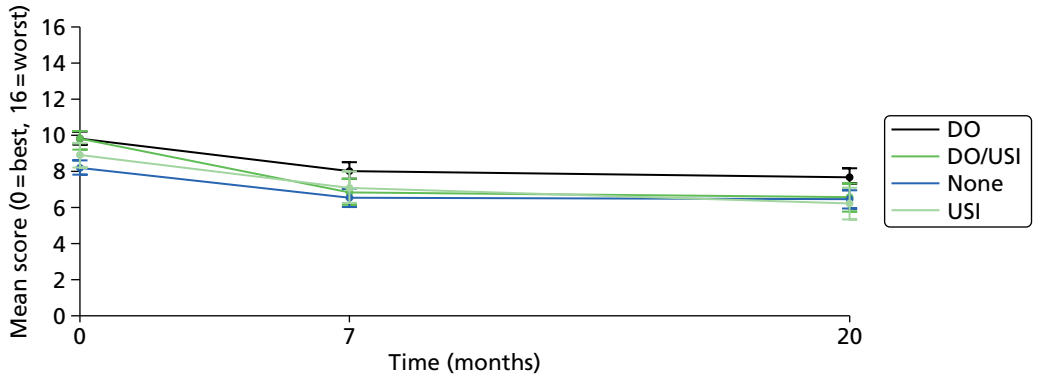

FIGURE 16 The ICIQ scores over time by diagnosis group (95% CIs are shown at each time point).

Effect of receiving a medical or surgical treatment concordant with urodynamics diagnosis

Sixty-two per cent and 66% of patients had received a medical or surgical treatment concordant with their UDS diagnosis by 7 and 20 months, respectively (*Table 22*). At 20 months, 57% (168/296) of patients who had received a concordant treatment responded positively to the global impression of improvement question compared with 45% (69/152) of patients whose treatment was discordant with their diagnosis (OR 1.6, 95% CI 1.1 to 2.3; $p = 0.02$). There was no overall evidence that this varied by UDS diagnosis ($p = 0.1$), although positive responses appeared higher in the DO + USI and USI groups (*Table 23*). ICIQ scores were reduced at 7 and 20 months overall (*Table 24* and *Figure 17*) and the reduction was significantly greater at both time points in those who had received a concordant treatment than those who had received a discordant treatment (–0.5, 95% CI –0.9 to –0.1; $p = 0.02$). There was no evidence that the effect of receiving a concordant treatment varied by UDS group ($p = 0.3$; *Table 25*).

TABLE 22 Proportion of patients who had received a medical or surgical treatment concordant with their UDS diagnosis at 7 and 20 months

UDS diagnosis	Concordant medical or surgical treatment, *n*/*N* (%)
7 months	
DO + USI	40/71 (56)
DO	113/194 (58)
USI	6/53 (11)
Normal UDS	111/118 (94)
Total	270/436 (62)
20 months	
DO + USI	56/82 (68)
DO	121/201 (60)
USI	18/61 (30)
Normal UDS	111/123 (90)
Total	306/467 (66)

TABLE 23 Proportion of patients reporting improvement in symptoms at 20 months by UDS diagnosis and whether a medical or surgical treatment concordant with this diagnosis had been received

UDS diagnosis	Concordant medical or surgical treatment = no, *n*/*N* (%)	Concordant medical or surgical treatment = yes, *n*/*N* (%)	OR (95% CI)
DO + USI	8/25 (32)	39/55 (71)	5.2 (1.9 to 14.4)
DO	36/79 (46)	63/118 (53)	1.4 (0.8 to 2.4)
USI	24/41 (59%)	10/14 (71)	1.8 (0.5 to 6.6)
Normal UDS	1/7 (14)	56/109 (51)	6.3 (0.7 to 54.4)
Total	69/152	168/296	

TABLE 24 The ICIQ score responses by whether a medical or surgical treatment concordant with UDS diagnosis had been received

	Concordant medical or surgical treatment = no		Concordant medical or surgical treatment = yes	
Time point	Mean (SD), *n*	Change from baseline: mean (95% CI), *p*-value	Mean (SD), *n*	Change from baseline: mean (95% CI), *p*-value
Baseline	9.3 (2.5)		9.0 (2.8)	
7 months	7.7 (3.6), 159	–1.6 (–2.1 to –1.1), $p < 0.0001$	7.0 (3.2), 261	–2.1 (–2.4 to –1.7), $p < 0.0001$
20 months	7.3 (3.6), 150	–1.9 (–2.4 to –1.3), $p < 0.0001$	6.8 (3.5), 289	–2.4 (–2.8 to –2.0), $p < 0.0001$

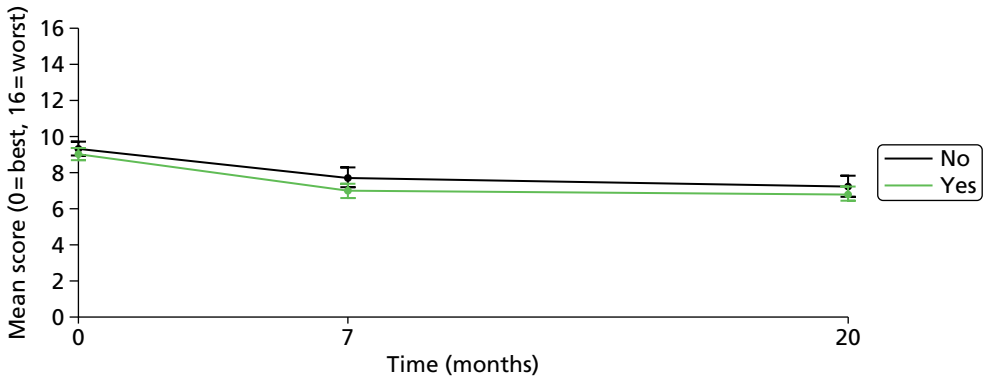

FIGURE 17 The ICIQ scores over time by medical or surgical treatment concordant with UDS diagnosis (95% CIs are shown at each time point).

TABLE 25 The ICIQ scores by UDS diagnosis and whether a medical or surgical treatment concordant with this diagnosis had been received

	Concordant medical or surgical treatment = no		Concordant medical or surgical treatment = yes	
UDS diagnosis	Mean (SD), *n*	Change from baseline: mean (95% CI), *p*-value	Mean (SD), *n*	Change from baseline: mean (95% CI), *p*-value
DO + USI				
Baseline	9.4 (2.1)		9.5 (2.7)	
7 months	6.9 (3.2), 28	–2.7 (–4.0 to –1.4), $p = 0.0003$	6.2 (3.7), 38	–3.4 (–4.5 to –2.3), $p < 0.0001$
20 months	7.4 (3.6), 26	–1.8 (–3.3 to –0.4), $p = 0.02$	6.0 (3.4), 53	–3.6 (–4.6 to –2.6), $p < 0.0001$
DO				
Baseline	9.6 (2.8)		9.9 (2.8)	
7 months	8.6 (3.6), 79	–1.0 (–1.6 to –0.4), $p = 0.002$	7.8 (3.3), 111	–2.2 (–2.8 to –1.6), $p < 0.0001$
20 months	7.6 (3.8), 77	–1.6 (–2.3 to –0.8), $p < 0.0001$	7.9 (3.8), 115	–2.3 (–3.0 to –1.6), $p < 0.0001$
USI				
Baseline	8.9 (2.4)		9.3 (4.0)	
7 months	7.1 (3.4), 45	–1.8 (–2.8 to –0.8), $p = 0.0006$	6.5 (4.1), 6	–2.8 (–7.8 to 2.1), $p = 0.2$
20 months	6.4 (3.3), 40	–2.6 (–3.8 to –1.3), $p = 0.0001$	5.8 (3.1), 14	–2.3 (–4.3 to –0.2), $p = 0.03$
Normal UDS				
Baseline	9.9 (2.0)		8.0 (2.3)	
7 months	6.3 (3.5), 7	–3.6 (–6.9 to –0.2), $p = 0.04$	6.5 (2.8), 106	–1.5 (–1.9 to –1.0), $p < 0.0001$
20 months	8.1 (3.2). 7	–1.7 (–5.1 to 1.8), $p = 0.3$	6.3 (3.1), 107	–1.9 (–2.4 to –1.3), $p < 0.0001$

Relationship between presenting symptoms and long-term outcomes

Positive responses to the global impression of improvement question at 20 months were slightly higher in the urgency + SUI group (56%, 159/286) than the urgency alone group (48%, 86/181) but this difference was not statistically significant ($p = 0.09$). There was also no evidence that the effect of these presenting symptoms varied by whether or not a treatment concordant with UDS diagnosis had been received ($p = 0.2$). ICIQ scores were lower at both time points in both groups (*Table 26* and *Figure 18*); however, there was some evidence of the urgency + SUI group improving more than the urgency alone group by 20 months (0.60 points, 95% CI 0.00 points to 1.20 points) but not at 7 months (–0.03 points, 95% CI –0.56 points to 0.51 points).

Relationship between bladder wall thickness and long-term outcomes

There was no evidence that BWT had any relationship with the global impression of improvement responses at 20 months ($p = 0.4$) or ICIQ scores ($p = 0.8$) over 7 and 20 months (correlation coefficients $r = 0.01$ at both time points); *Figure 19* shows the distribution of scores at 7 months. There was also no evidence of BWT having any interaction effect with (1) UDS diagnosis (global impression of improvement: $p = 0.9$; ICIQ: $p = 0.7$) or (2) presenting symptoms (global impression of improvement: $p = 0.2$; ICIQ: $p = 0.5$).

TABLE 26 Relationship between ICIQ scores and clinical symptoms

Clinical symptoms	Mean (SD), *n*	Change from baseline mean (95% CI)	*p*-value
Urgency + stress			
Baseline	9.4 (2.6)		
7 months	7.4 (3.4), 286	–1.9 (–2.3 to –1.6)	< 0.0001
20 months	6.9 (3.6), 278	–2.5 (–2.9 to –2.1)	< 0.0001
Urgency alone			
Baseline	9.1 (2.8)		
7 months	7.3 (3.2), 181	–1.8 (–2.2 to –1.4)	< 0.0001
20 months	7.2 (3.5), 179	–1.8 (–2.3 to –1.3)	< 0.0001

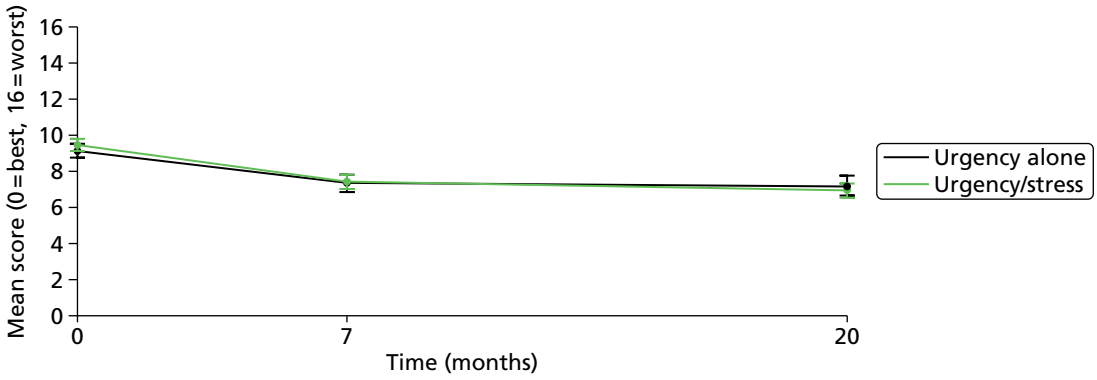

FIGURE 18 The ICIQ scores over time by presenting symptoms (95% CIs are shown at each time point).

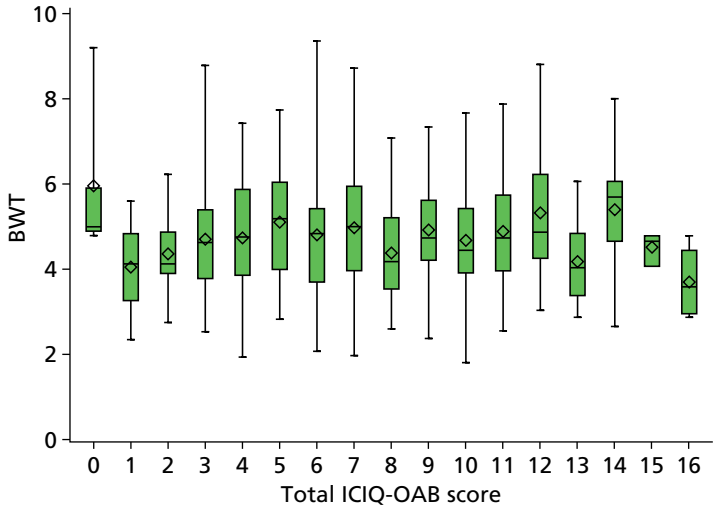

FIGURE 19 Seven months ICIQ scores vs. BWT ($r = 0.01$).

Discussion

Summary of main findings

Our analysis demonstrated that UDS diagnosis appeared to have an effect on subsequent treatment. Women with DO were three times more likely to have had bladder relaxants than no treatment compared with women with normal UDS. This might mean that those who were shown to have DO were more likely to be prescribed bladder relaxant tablets or that patient compliance with taking the treatment improved. Women with a diagnosis of DO + USI were 15 times more likely to have USI surgery than no treatment, which may at least partly explain the improved ICIQ scores and global impression of improvement over patients with pure DO.

Numerous studies have demonstrated that a diagnosis of OAB cannot be accurately made based on symptomatology alone.[29,35,149,150] Clement *et al.*[151] concluded in their Cochrane review that urodynamic testing did change clinical decision-making in women treated for UI (relative risk 5.07, 95% CI 1.87 to 13.74) and more women received drugs in the UDS groups. Contrary to the finding of the Cochrane review, that women were not more likely to undergo surgery, we found that more women with DO + USI had received surgery by the 20-month follow-up.[151] Confirmation of the concurrent pathophysiology of DO + USI may have resulted in more clinicians offering USI surgery when bladder relaxants alone failed to provide improvements.

Over half (53%) of the study population reported an improvement in symptoms and disease-specific scores by a margin that appears clinically meaningful. The ICIQ-OAB provides a brief and robust measure to assess the impact of symptoms of OAB on QoL. In the follow-up of the BUS participants, scores were significantly reduced from baseline by 2.2 points ($p < 0.0001$). The findings also support 'responsiveness' of ICIQ questionnaire to therapy as the change in ICIQ scores was two-thirds of a SD and most studies would accept a difference of half a SD to be a clinically important change.[152]

Patients who received a medical or surgical treatment concordant with their UDS findings were more likely to report an improvement in bladder symptoms at 20 months than those who did not (57% vs. 45%; $p = 0.02$). The improvement reported by those who did not receive a concordant treatment could be for several reasons, such as natural fluctuation of disease state, regression to the mean and Hawthorne effects.[153] The experience of UDS may have helped women to understand their condition better and improve compliance, including with lifestyle measures.[154]

The ICIQ scores were reduced at both time points regardless of whether or not the women received a treatment concordant with UDS findings; however, patients receiving a concordant treatment reported a slightly greater reduction (–0.5 points; $p = 0.02$). There was also some evidence that ICIQ responses were greater in the DO + USI diagnosis group than the DO group (–1.1 points; $p = 0.002$); that is, women with MUI treated based on UDS diagnoses appear to have greater reductions in symptoms.

There was no evidence that BWT had any relationship with the global impression of improvement responses or ICIQ scores. This concurred with the findings from the BUS which indicated that BWT is not a useful test in the management of these patients.

Strengths and limitations

The patients were followed up for more than 12 months and validated questionnaires were used. This study is one of the few reporting on prospective follow-up in patients with urgency-predominant MUI and suggestive of better prognosis in the MUI group than the DO group.

The response rate for continued follow-up of the BUS cohort was 69% which, although not high, is superior to other studies in the field at over 1-year follow-up. We could only ascertain whether or not women had ever having taken bladder relaxants and could not determine whether or not women were

currently taking bladder relaxants. In addition, the number of women having both bladder relaxants and surgery was small and, therefore, could not be reliably distinguished from those who had surgery alone in the analysis. We also did not collect data on therapies, such as supervised intensive PFMT, bladder retraining, lifestyle changes, etc., but we presumed that the conservative treatment was already exhausted before patients were referred for UDS.

As the study is an observational study, it is not possible to directly infer the clinical effectiveness of the therapies from the comparison of whether treatment options were compliant or non-compliant with the UDS diagnosis. There may be other reasons why patients whose treatment did not follow from the diagnosis have different outcomes from those who did, for example reasons related to health states, patient preferences and characteristics.

Interpretations of findings

Urodynamics appears to influence treatment decisions made by clinicians in determining treatment pathways in women presenting with UI. Women treated with medical or surgical interventions based on UDS diagnoses appear to have greater reductions in symptoms than who were not. This was more evident in women with MUI on UDS. In the overall population, half the women reported long-term improvement in symptoms. BWT is not associated with patient management or treatment responses, either global or ICIQ score changes and has no prognostic value as a test in this condition.

Recommendations for research

There is need for randomised trials involving women with OAB or urgency-predominant MUI, comparing treatment based on UDS with treatment based on clinical symptoms and examination alone.[155]

Further work needs to be carried out to establish the ‘minimally important clinical difference’ in ICIQ scores to enable its use to quantify response to treatment objectively in clinical practice. This could be done by assessing the changes in the bladder diary variables and the individual ICIQ items.[156]

Chapter 7 Economic evaluation of alternative diagnostic strategies

Introduction

Objectives

The initial objective of the primary study was to evaluate the accuracy of bladder ultrasonography in the diagnosis of DO and to investigate the added value of using the test results alongside the information obtained from routinely used non-invasive tests. These objectives were revised [with agreement of the National Institute for Health Research (NIHR) Health Technology Assessment (HTA)] because bladder ultrasonography was shown in the accuracy study to be of limited diagnostic value. The focus switched instead to the investigation of the accuracy and cost-effectiveness of UDS in the investigation of OAB-like presentations. However, given that the initial objective was to evaluate bladder ultrasonography and given the extent of data collected on this test, for completeness it has been included in the economic analysis.

National Institute for Health and Care Excellence recommends the use of UDS only prior any invasive intervention for OAB,[36] based on evidence indicating its limited role in determining the outcome of conservative treatments.[66] Thus, the objective of the model-based economic evaluation of this chapter is to compare the relative cost-effectiveness of undertaking three alternative diagnostic strategies based on (1) UDS, (2) a bladder ultrasonography test or (3) clinical history for women with predominant symptoms of OAB for whom conservative treatments were not effective. Each strategy contains both the test and the subsequent treatment options made based on the results of the test. For brevity, we will refer to the strategies by the test name, but in all instances we consider a strategy as consisting of the test plus the treatment choices, which are made dependent on the test result obtained.

Methods

In the economic evaluation, three principal diagnostic pathways are compared. In the first, treatment is based on UDS and this represents the way treatment pathways are determined in current practice. In the second, treatment is based on the bladder ultrasonography test, which can be used to determine the appropriate treatment based on the measurement of BWT. In the third, treatment is based on clinical history and this represents the clinical pathway that a woman would follow if no diagnostic test was available and treatment relied on a patient's history only. In a secondary analysis combinations of these strategies are also explored.

The primary symptoms of women with OABs are urgency and often urgency-related incontinence. These symptoms can be accompanied by another type of incontinence that occurs as a result of increased abdominal but not detrusor pressure and this is known as SUI. Where they coexist, the condition is referred to as MUI. Women who experience predominant symptoms of OAB are suspected to most likely have either DO or MUI. Sometimes, these women can have other syndromes – SUI, low compliance, VD – or, on the basis of diagnosis, are referred to as normal despite the existence of symptoms.

As far as treatment is concerned, all women are treated conservatively in the first instance. Conservative treatments typically comprise bladder training and oral treatment with antimuscarinic drugs. If these treatments are not effective, women undergo further investigation and tests before more invasive treatment is considered. The benefit of UDS compared with the other two diagnostic strategies is that it provides a more clear-cut diagnosis of the underlying syndrome. More specifically, UDS can identify (1) if the cause of the predominant symptoms of OAB is a DO, (2) if the symptoms observed are because of other syndromes (MUI, USI, low compliance, VD) or (3) if women are considered to have a normal bladder despite the existence of symptoms.

The hypothesis used in the primary study was that the bladder ultrasonography test predicts DO if BWT is above a predefined threshold. In our model we follow this using data on accuracy from the primary study and presume that bladder ultrasonography does not discriminate between MUI, USI, low compliance, VD and normal otherwise. For brevity, we will call this set of diagnoses MUI. Finally, clinical history according to whether women have only urinary urgency, with or without urgency incontinence, or had urgency-predominant MUI was used to indicate diagnoses either of OAB, most likely caused by DO, or of MUI. Despite the theoretical advantage of UDS, previous evidence has indicated – contrary to the findings reported in *Chapter 6* – that an urodynamic observation rarely has an impact on the type of intervention undertaken or on outcomes.[66] Therefore, a model-based economic evaluation was required to synthesise data on test accuracy and evaluate the costs and outcomes incurred by women following any particular treatment pathway based on the test result.

Model structure

In order to maintain patient history and given the short-term nature of the decision problem, the appropriate model structure to describe the options being compared and their treatment pathways is a decision tree. The model was developed in TreeAge Pro 2014 software (TreeAge Software, Inc., Williamstown, MA, USA) and the structure was informed by clinical input and NICE guidelines on the management of UI.[36] Women with a mean age of 55 years (95% CIs 39 years to 71 years) enter the model if they present predominant symptoms of OAB and conservative treatments have not been effective. The term 'predominant symptoms of OAB' refers to symptoms of urgency or urgency-related incontinence, possibly accompanied by stress incontinence. Women who have not undergone conservative treatments and women for whom conservative treatments were effective in treating the urgency symptom but the stress symptom remained were not included in the analysis.

In the primary analysis of the model-based economic evaluation, women are assumed to follow one of three alternative strategies for the treatment of their symptoms: (1) based on UDS observation alone, (2) based on bladder ultrasonography test alone and (3) based on clinical history. These are presented in the branches to the right of the decision node (square symbol) in *Appendix 9* (see *Figure 43*). For completeness, pathways for strategies that represent other combinations in which clinical history and diagnostic tests can in theory be used together are also considered. These are illustrated in *Appendix 9* (see *Figure 44*).

Once women enter the model, it is assumed that they proceed from the least invasive (and more common) interventions to the most invasive ones. For the purpose of the model, it is assumed that further conservative treatment is not provided. However, in real life, conservative treatments are ongoing and can be complementary to the more invasive interventions.

As a first-line surgical intervention for the treatment of OAB caused by DO (diagnosed as DO by UDS, or a positive bladder ultrasonography test, or as urgency with or without urgency incontinence, or clinical history) women go through either BTX-A injections or PTNS depending on patient and physician preferences. If these interventions are not effective, the second-line intervention is determined through a peripheral nerve evaluation (PNE), which is the first stage of the SNS. The outcome then determines whether or not a permanent implantation of a stimulation device (stage II) will be given. If this is not indicated, women are assumed to undergo either BTX-A injections or PTNS depending on which of the two has not been applied earlier in the treatment pathway.

The treatment of SUI (USI diagnosed only by UDS) involves a sling surgery as a first-line intervention and Burch colposuspension as a second-line surgical treatment.

In cases of MUI (diagnosed as DO + USI by UDS, a negative bladder ultrasonography test, or urgency-dominant MUI clinical history) women can opt for BTX-A injections prior to the sling surgery and Burch colposuspension as a second-line surgical treatment.

Women who are diagnosed with a normal bladder, despite the existence of symptoms, can be identified by UDS only. These women are assumed to remain symptomatic without further interventions.

For the strategies bladder ultrasonography test and clinical history, when treatment either for OAB or MUI is initiated, women are assumed to receive invasive interventions as a result of a misdiagnosis. The same is assumed for women who have low compliance only or VD only despite the predominant symptoms of OAB. The treatment pathways for each syndrome in the UDS strategy are illustrated in *Appendix 9* (see *Figures 45–47*). The pathways for the diagnostic strategies of bladder ultrasonography test and clinical history are shown in *Appendix 9* (see *Figures 48–51*).

Model assumptions

To carry out the model-based analysis some further pragmatic assumptions were required. These are presented below.

- Diagnostic tests may be repeated once in each treatment pathway before the second-line interventions and intervention strategies will change if the second test contradicts the original test.
- In the UDS and bladder ultrasonography strategies, treatment is assumed to be initiated 3 months after the diagnostic test, but instantly in the case of clinical assessment. Three months was selected as an arbitrary cut-off point below the maximum of 18 weeks used in the NHS[157] to reflect the waiting time for a subsequent visit for a first-line intervention.
- Symptoms of VD and low compliance that can accompany DO, MUI or SUI are assumed to not have an impact on the clinical effectiveness of interventions.
- Women diagnosed with VD require self-catheterisation training.
- A woman can have up to three BTX-A injections at yearly intervals.
- Subjective cure from BTX-A entails the possibility of acquiring voiding difficulties that will require self-catheterisation training.
- PTNS is offered in 12 sessions, a week apart, in the final 3 months. If it is successful, a monthly session is required for the rest of the model period in order for its clinical effectiveness to be maintained.
- In the case of SNS, a proportion of women will require a revision of the surgery and maintenance or removal of the implanted neurostimulator device, depending on the time between SNS surgery and the end of the model period. These were assumed to take place at the end of the model period.
- De novo DO or urgency UI is a possible outcome of an antistress incontinence surgery. It is described as the development of a new OAB-related symptom in women who did not demonstrate DO in their pre-operative evaluation. Given that women entering the model have predominant symptoms of OAB, de novo DO or urgency UI is not modelled.
- If a woman becomes subjectively cured, improvements are assumed to be maintained throughout the model period.

Clinical data

The prevalence and accuracy data used in the model are based on the results of the primary study using UDS as a reference standard. These are presented in *Tables 27–30*. The probability of each intervention being clinically effective is presented in *Table 31*. Data on the effectiveness of PTNS in women with DO,[82] the clinical effectiveness of sling surgery in women with SUI[160] and MUI,[161] as well as the clinical effectiveness of colposuspension in women with SUI[164] were drawn from meta-analyses. The probability of SNS being effective in women with DO was taken from a systematic review.[89] Observational studies were used to inform the probability of three repetitive BTX-A injections being effective in women with DO[158] and the cure rate of colposuspension in women with MUI.[163] From the former observational study, the proportion of women that become subjectively cured after three BTX-A injections (56.8%), the disaggregation of the overall effectiveness across the three injections and the drop-out rate after each injection were used in the model (*Table 32*). In the absence of relevant literature, the disaggregation of the overall effectiveness across the three injections and the drop-out rate after each injection were assumed to be the same if BTX-A was offered in women with other syndromes.

TABLE 27 Prevalence data used in the model

Description	Proportion	Distribution (parameter values)
Women with DO	0.507	Dirichlet (106, 19, 35, 36, 9, 3, 1)
Women with SUI	0.091	
Women with MUI	0.167	
Women that had a normal UDS	0.172	
Women with low compliance only	0.043	
Women with VD only	0.014	
Women with low compliance and VD	0.005	

TABLE 28 Accuracy data used in the model for UDS

Description	Proportion	Distribution (parameter values)
Women with DO that had a positive UDS for DO	0.964	Dirichlet (107, 1, 1, 1, 1)
Women with DO that had a positive UDS for SUI	0.009	
Women with DO that had a positive UDS for MUI	0.009	
Women with DO that had a normal UDS	0.009	
Women with DO that had 'other' UDS	0.009	
Women with SUI that had a positive UDS for SUI	0.833	Dirichlet (20, 1, 1, 1, 1)
Women with SUI that had a positive UDS for DO	0.042	
Women with SUI that had a positive UDS for MUI	0.042	
Women with SUI that had a normal UDS	0.042	
Women with SUI that had 'other' UDS	0.042	
Women with MUI that had a positive UDS for MUI	0.900	Dirichlet (36, 1, 1, 1, 1)
Women with MUI that had a positive UDS for SUI	0.025	
Women with MUI that had a positive UDS for DO	0.025	
Women with MUI that had a normal UDS	0.025	
Women with MUI that had 'other' UDS	0.025	

TABLE 29 Accuracy data used in the model for the bladder ultrasonography test

Description	Proportion	Distribution (parameter values)
Women with DO that had a positive bladder ultrasonography	0.377	Beta(40,66)
Women with DO that had a negative bladder ultrasonography	0.623	Remainder from above
Women with MUI that had a positive bladder ultrasonography	0.514	Beta(18,17)
Women with MUI that had a negative bladder ultrasonography	0.486	Remainder from above
Women with SUI that had a positive bladder ultrasonography	0.474	Beta(9,10)
Women with SUI that had a negative bladder ultrasonography	0.526	Remainder from above
Women that had a normal UDS that had a positive bladder ultrasonography	0.444	Beta(16,20)
Women that had a normal UDS that had a negative bladder ultrasonography	0.556	Remainder from above
Women with 'other' that had a positive bladder ultrasonography	0.462	Beta(6,7)
Women with 'other' that had a negative bladder ultrasonography	0.538	Remainder from above

TABLE 30 Accuracy data used in the model for clinical history

Description	Proportion	Distribution (parameter values)
Women with DO that had a clinical history of OAB	0.481	Beta(51,55)
Women with DO that had a clinical history of MUI	0.519	Remainder from above
Women with MUI that had a clinical history of OAB	0.114	Beta(4,31)
Women with MUI that had a clinical history of MUI	0.886	Remainder from above
Women with SUI that had a clinical history of OAB	0.105	Beta(2,17)
Women with SUI that had a clinical history of MUI	0.895	Remainder from above
Women that had a normal UDS that had a clinical history of OAB	0.500	Beta(18,18)
Women that had a normal UDS that had a clinical history of MUI	0.500	Remainder from above
Women with 'other' that had a clinical history of OAB	0.692	Beta(9,4)
Women with 'other' that had a clinical history of MUI	0.308	Remainder from above

TABLE 31 Probabilities for the different interventions in the model

Description	Mean (95% CI)	Distribution (parameter values)[a]	References
Clinical effectiveness of BTX-A in women with DO[b]	0.568 (0.464 to 0.669)[c]	Beta(50,38)	Dowson *et al.*[158]
Clinical effectiveness of BTX-A in women with MUI[b]	0.329 (0.200 to 0.483)	Beta(13.42,27.36)	Expert opinion[d]
Clinical effectiveness of BTX-A in women with SUI[b]	0.143 (0.133 to 0.375)	Beta(4.32,25.86)	Expert opinion[d]
Clinical effectiveness of PTNS in women with DO	0.606 (0.558 to 0.653)[c]	Beta(245,159)	Burton *et al.*[82]
Clinical effectiveness of PTNS in women with MUI or SUI	0.314 (0.283 to 0.567)	Beta(12.36,27)	Expert opinion[d]
Clinical effectiveness of SNS surgery in women with DO	0.675 (0.633 to 0.715)[c]	Beta(338,163)	Brazzelli *et al.*[89]
Clinical effectiveness of SNS surgery in women with MUI or SUI	0.271 (0.233 to 0.550)	Beta(7.74,20.82)	Expert opinion[d]
Clinical effectiveness of sling surgery in women with DO	0.310 (0.250 to 0.350)	Beta(5.84,13)	Jorgensen *et al.*[159] in Weber and Walters (2000)[90]
Clinical effectiveness of sling surgery in women with SUI	0.868 (0.841 to 0.894)[c]	Beta(547,83)	Latthe *et al.*[160]
Clinical effectiveness of sling surgery in women with MUI	0.560 (0.534 to 0.579)[c]	Beta(1050,837)	Jain *et al.*[161]
Clinical effectiveness of colposuspension in women with DO	0.163 (0.153 to 0.386)	Beta(5.95,30.56)	Expert opinion[d]
Clinical effectiveness of colposuspension in women with SUI	0.690 (0.612 to 0.762)[c]	Beta(100,45)	Dean *et al.*[162]
Clinical effectiveness of colposuspension in women with MUI	0.489 (0.381 to 0.595)[c]	Beta(40,42)	Kulseng–Hanssen *et al.*[163]

a When mean and effective number of patients were available, these were used to derive the parameters for the distribution. In other cases, the distribution was fitted to the mean and CIs.
b Clinical effectiveness relies on three BTX-A injections at yearly intervals.
c Information in the source consisted of point estimates and effectively treated number of patients. 95% CIs have been derived from this information.
d The source of the expert opinion is described in *Clinical data*.

TABLE 32 Disaggregation of the effectiveness and drop-out rate in each BTX-A injection

Description	First injection	Second injection	Third injection	Reference
Proportion of the overall cure	0.34	0.52	0.14	Dowson *et al.*[158]
Drop-out rate[a]	0.20	0.08	0.00	Dowson *et al.*[158]

a Drop-out rate is based on only BTX-A related reasons. Other reasons (e.g. lost to follow-up or moved out of the area) were not taken into consideration.

A literature search was undertaken to identify the clinical effectiveness of interventions modelled in a situation of a misdiagnosis. Apart from the clinical effectiveness of a sling surgery in women with DO identified in a published economic evaluation,[90] there was limited robust evidence to inform the model about the clinical effectiveness of BTX-A in women with MUI and SUI or the clinical effectiveness of PTNS and SNS in women with MUI or SUI. For this reason, these values were elicited from the anonymous expert opinion of eight study collaborators on 18 June 2014. Expert opinion is a legitimate source of information in decision modelling when other information is not available. Given that the parameters of interest apply to very rare circumstances, it was appropriate to decide in advance to use a beta distribution and use a simple approach to elicitation, asking for most likely value as well as lowest and highest, interpreting these as mean and lower and upper limits of a 95% CI. Use of more elaborate elicitation techniques[165] was judged to be unlikely to make an appreciable difference to the modelling.

Other data used in the model are presented in *Table 33*. The proportion of women that proceed to an implantation of a neurostimulator device (SNS stage II) after a PNE was informed by a systematic review.[88] SNS often requires a revision of the surgery and this was assumed to vary across time. More specifically, 9% of women undergoing a SNS surgery are assumed to require a revision of the surgery within 2 years from the surgery and 33% after 3–5 years. These values were taken from two systematic reviews.[90,166] The probability of removal of the implanted device and the probability that a maintenance surgery will be required $\geq$ 2 years after surgery were also taken from these systematic reviews. In addition, VD is a possible adverse event after a subjective cure from BTX-A injections and the proportion of women acquiring VD was taken from a RCT.[91] Finally, the proportion of women with a diagnosis of MUI that choose BTX-A injections prior to a sling surgery and women with a diagnosis of DO (or OAB) of that choose BTX-A injections instead of PTNS were informed by the anonymous collaborators of this study.

TABLE 33 Additional model parameters

Description	Proportion (95% CI)	Distribution (parameter values)[a]	References
Women choosing BTX-A before sling surgery	0.314 (0.275 to 0.683)	Beta(5.74,12.53)	Expert opinion[b]
Women choosing BTX-A instead of PTNS	0.750 (0.421 to 0.963)[c]	Beta(6,2)	Expert opinion[b]
Women acquiring voiding difficulties after BTX-A	0.086 (0.042 to 0.143)[c]	Beta(10,106)	Tincello *et al.* (2012)[91]
Women with DO who are suitable for a SNS stage II after a PNE	0.670 (0.450 to 0.880)	Beta(11.23,5.53)	Brazzelli *et al.* (2006)[89]
Women requiring revision of the SNS stage II between 2 and 5 years after the surgery	0.330 (0.299 to 0.362)[c]	Beta(282,573)	Brazzelli *et al.* (2006)[89]
Women requiring revision of the SNS stage II within 2 years of surgery	0.090 (0.064 to 0.119)[c]	Beta(36,366)	Siddiqui *et al.* (2010)[166]
Women requiring maintenance after 2 years of the SNS stage II surgery	0.150 (0.111 to 0.195)[c]	Beta(42,237)	Brazzelli *et al.* (2006)[89]
Women requiring removal of the SNS stage II-implanted device	0.107 (0.068 to 0.152)[c]	Beta(22,184)	Siddiqui *et al.* (2010)[166]

a Where mean and effective number of patients were available, these were used to derive the parameters for the distribution. In other cases, the distribution was fitted to the mean and 95% CIs.
b The source of the expert opinion is described in *Clinical data*.
c Information in the source consisted of point estimates and effectively treated number of patients. 95% CIs have been derived from this information.

Cost data

Information from the NHS Reference Costs 2012–13[30] and Curtis[92] was used to parameterise the cost component of the decision model. Costs were calculated in 2012–13 UK Great British pounds (£). These costs of different interventions modelled are presented in *Table 34*. Costs selected from the NHS Reference Costs have been calculated based on the weighted average value of elective inpatient and day-case costs and the proportion of patients in each group. An exception is in the case of the evaluation stage of the SNS, which is offered as a day case. Unit costs from the urology category were selected instead of the average across different medical specialties, apart from the case of bladder ultrasonography for which only a total Healthcare Resource Group (HRG) cost was available. The cost of three BTX-A injections has been calculated by multiplying the unit cost of each injection (£912, 95% CI £704 to £1060) by the proportion of women undergoing each injection. This has been calculated based on the proportions of women becoming subjectively cured and dropping out after each BTX-A injection, which combines information from *Tables 31* to *32*. Information about the mean cost and 95% CIs for three BTX-A injections across the different syndromes is illustrated in *Table 35*.

TABLE 34 Cost data used in the model (£, 2012/13 prices)

Item	NHS Reference Cost code	Unit cost (95% CI)	Distribution (parameter values)[a]	Reference
UDS	LB42 A (Urology)	401 (216 to 462)	Gamma(40.65,9.86)	NHS Reference Costs[30]
Bladder ultrasonography	RA23Z[b]	51	Gamma(1.00,51.07)[c]	NHS Reference Costs[30]
BTX-A	LB14Z (Urology)	912 (704 to 1060)	Gamma(100.67,9.06)	NHS Reference Costs[30]
PTNS	AA21F (Urology)	2221 (1274 to 2838)	Gamma(30.81,72.08)	NHS Reference Costs[30]
Sling surgery	LB59Z (Urology)	3917 (2599 to 5309)	Gamma(31.93,122.69)	NHS Reference Costs[30]
SNS (evaluation stage)	AA21F (Urology)[d]	1162 (1010 to 1293)	Gamma(258.88,4.49)	NHS Reference Costs[30]
SNS (implantation stage)	AB07Z (Urology)	6530 (4966 to 8347)	Gamma(57.14,114.28)	NHS Reference Costs[30]
SNS (removal/maintenance)	AB04Z (Urology)	4160 (2960 to 5831)	Gamma(32.09,129.65)	NHS Reference Costs[30]
Burch colposuspension	LB59Z (Urology)	3917 (2599 to 5309)	Gamma(31.93,122.69)	NHS Reference Costs[30]
Training for self-catheterisation[e]	–	84	Gamma(1.00,84.00)[c]	Curtis[92]

a Unless otherwise indicated, distributions were fitted based on mean value and CIs.
b Based on total HRGs.
c Distributions were fitted by the method of moments assuming variance is equal with the mean cost.
d As a day case.
e Assuming an hour contact with a nurse.

TABLE 35 Undiscounted cost of three BTX-A injections for each syndrome

Syndrome	Mean	Lower 95% CI (£)	Upper 95% CI (£)
DO	£1669	1288	1939
MUI	£1930	1490	2244
SUI	£2134	1647	2480
Normal or 'other'[a]	£2290	1768	2662

a The cost shown for this category is based on the assumption of zero clinical effectiveness.

Outcomes

To examine whether or not a diagnostic test is a cost-effective component in the treatment pathway for women with predominant symptoms of OAB, three different outcomes were considered: (1) women successfully treated, as determined by subjective symptoms, (2) DO cases detected and (3) quality-adjusted life-years (QALYs) based on the best available QoL data in the literature. The third combines quantity with QoL, which is measured using utility weights. A review of the Cost-Effectiveness Analysis Registry[167] using the terms 'overactive bladder', 'urinary incontinence' and 'detrusor overactivity' was performed to identify relevant utility weights for the outcomes experienced at a time point beyond the primary study end point. Eighteen studies[168–186] were identified and utility weights from one study[168] were selected to represent the QoL for subjective cure. This study was selected because the utility values were obtained with a sound theoretical approach (time trade-off) and were relevant to the symptoms of OAB and the type of interventions modelled.[169] Utility weights from other studies were not considered, either because they were focusing on a population with OAB that was undergoing conservative treatments rather than invasive[170–176] or because they were focusing on women with SUI.[177–182] In four more studies,[183–186] the values of 0.95 and 0.73 were used to represent the utility of continent and incontinent state for a relevant population and interventions, but these were not considered appropriate because the former value in reality represents the utility weight of people with no chronic condition[187] and the latter is based on women with SUI.[188] In the model, women who remain symptomatic were assumed to maintain their initial QoL, which was informed by the primary study. Thus, no QoL decrements were allowed to occur after diagnostic tests or after surgeries, even when they were the outcome of a misdiagnosis. The QoL scores used in the model are shown in *Table 36*. QALYs were estimated combining the utility weights with estimates of the duration of different health states.

TABLE 36 Quality-of-life data used in the model

Utilities	Mean	Lower 95% CI	Upper 95% CI	Distribution (parameter values)[a]	References
Initial utility					
DO	0.600	0.532	0.668	Beta(8.96,5.98)	BUS
SUI	0.660	0.514	0.807	Beta(18.92,9.74)	BUS
MUI	0.718	0.637	0.799	Beta(49.12,19.29)	BUS
Normal	0.656	0.558	0.753	Beta(22.13,11.63)	BUS
'Other'	0.744	0.547	0.942	Beta(11.75,4.03)	BUS
Utility for subjective cure[b]					
Without side effects	0.920	0.710	0.990	Beta(10.69,0.93)	Chen *et al.*[168]
With side effects	0.870	0.830	0.900	Beta(304,45.43)	Chen *et al.*[168]

a Distributions were fitted based on mean value and CIs.
b For women that remain symptomatic it was assumed that initial QoL is maintained.

Analysis

The decision model was constructed to investigate the cost-effectiveness of UDS compared with the diagnostic strategies, the bladder ultrasonography test and clinical history alone. Two separate economic analyses are carried out. The primary analysis provides a comparative evaluation of the costs and benefits of the UDS, bladder ultrasonography and clinical history diagnostic strategies. In a secondary analysis, all the different ways in which clinical history and a diagnostic test (UDS, bladder ultrasonography test) can be used together are explored. This allows the exploration of whether or not a diagnostic test in selective subgroups of women with predominant symptoms of OAB is a more cost-effective strategy than the three strategies explored in the primary analysis.

In both primary and secondary analyses, subjective cure is used for the identification of women successfully treated and UDS has been used as a reference standard. In the UDS arm, women with 'normal' or 'other' observations are assumed to remain symptomatic throughout the model period. However, in the case of bladder ultrasonography and clinical history, these women receive treatment for OAB or MUI as a result of a misdiagnosis. In such a situation, women were assumed to have a probability of becoming asymptomatic mainly as a result of a placebo effect. In the absence of robust evidence to inform this probability, the smallest figure of *Table 31* (0.143, 95% CI 0.133 to 0.375) was selected.

The analyses were carried out from the perspective of the UK NHS and the primary outcome is in terms of cost per woman successfully treated. Other outcomes included in the analysis are cost per DO case detected and cost per QALY. The latter is the recommended outcome for economic evaluations in the UK.[189] Results presented in terms of QALYs represent an additional analysis and were not part of the principal objective of the study. Results are presented in terms of incremental cost-effectiveness ratios (ICERs) and cost-effectiveness acceptability frontiers (CEAFs). A 5-year time horizon was considered appropriate to reflect all key differences, in terms of costs and benefits, for the options compared. Costs and QALYs accruing beyond 12 months were discounted at a rate of 3.5% per year. According to NICE,[36] based on existing evidence,[66] UDS should not be applied prior to the initiation of conservative treatments but only when these are not effective and women are expected to proceed to more invasive interventions. For this reason, the decision model is focusing on the subgroup population of the primary study that had been taking conservative treatment before enrolling into the study and had complete accuracy data on the three strategies modelled ($n = 209$).

Deterministic and probabilistic sensitivity analyses (PSAs) are performed to explore the effects of the inherent uncertainty in parameter estimates on model results. In deterministic sensitivity analysis, one or more parameters are varied while keeping the remaining at their baseline value. Although deterministic sensitivity analyses can be helpful to identify which model inputs are important in driving a decision or identify threshold values, comprehensive representation can be obtained by undertaking a PSA, in which the uncertainty around a parameter is represented with a probability distribution.[190,191] In the PSA, using 10,000 repeated random draws from those distributions, a Monte Carlo simulation of the model provides an indication of how variation in the model parameters leads to a variation in the results generated. Beta and Dirichlet distributions were used for binomial and multinomial data, respectively (see *Tables 27–33* and *Table 36*) and a Gamma distribution for costs (see *Table 34*). The parameters of each distribution are shown in the corresponding tables.

Deterministic sensitivity analyses

A number of deterministic sensitivity analyses were conducted in both the primary and secondary analyses. These deterministic analyses include both univariate and multivariate analyses to assess the impact of any uncertainty in model parameters on the final results. Three univariate analyses and four multivariate analyses were conducted based on the following justifications:

1. Univariate analyses

    i. Reducing the cost of UDS from £401 (95% CI £216 to £462) to £173. The base-case value was calculated based on the weighted average value of elective inpatient and day-care costs for a urological intervention and the proportion of patients in each group. The value used in this sensitivity analysis represents the weighted average value of total HRGs.[30]
    ii. Increasing the cost of sling surgery by 50% to account for possible adverse events such as bladder injury, vaginal erosion or groin pain.[160]
    iii. Lowering the utility weight of women subjectively cured from 0.92 to 0.84. This value is drawn from the BUS and represents the EQ-5D QoL score of those women that scored below four in the ICIQ (low bother).

2. Multivariate analyses

    i. Changing the accuracy data. The main analyses use UDS as a reference standard, which possibly leads to an overestimation of its diagnostic accuracy. In addition, urinary diaries can complement clinical history and provide a more accurate clinical assessment. The accuracy of UDS was reduced using evidence from an economic evaluation,[64] which reports 14% probability of false-positive diagnoses for MUI when DO or SUI are the true conditions and 25% probability of false-positive diagnoses for SUI when MUI is the true condition. Similarly, a systematic review[150] has reported that clinical history and information from diaries detect 53.4% of women with DO, increasing the accuracy of clinical history by five percentage points.
    ii. Changing the rates of clinical effectiveness elicited from expert opinion to their lowest value.
    iii. Changing the rates of clinical effectiveness elicited from expert opinion to their highest value.
    iv. Assuming one diagnostic test is offered prior to the initiation of the first-line interventions in line with NICE guidelines.[36]

Results

Primary analysis

The costs and outcomes of each strategy in the primary analysis are shown in *Table 37*. In terms of cost, UDS is the least expensive strategy (£4524) with clinical history and bladder ultrasonography costing an additional £1278 (£5801) and £1424 (£5947), respectively. In terms of outcomes, treatment based on clinical history appeared to be the most effective strategy apart from the number of DO cases detected, for which UDS was more effective (as DO is an urodynamic observation rather than a clinical symptom, a DO diagnosis is never made based on clinical history).

It is important to note that the total cost per patient (from £4524 to £5947) greatly exceeds the cost of the initial tests (£401 for UDS or £51 for bladder ultrasonography) and thus the major determinants of the costs are the treatment selections made based on the results of the tests.

In the deterministic analysis, the bladder ultrasonography test strategy was dominated by the other two strategies as it was the most expensive and least effective strategy across all outcomes explored in the analysis. In the remainder of this subsection, the results of the incremental analysis are described with more information provided about the cost for an additional unit of outcome between the strategies compared. Dominated strategies, such as bladder ultrasonography, are not described as a part of the incremental analysis.

TABLE 37 Deterministic analyses results for primary and secondary analysis across the three outcome measures

Strategy	Cost (£)	Clinical effectiveness		
		Women successfully treated	QALYs	DO cases detected
Primary analysis				
UDS	4524	0.6149	3.6693	0.4870
Clinical history	5801	0.6175	3.6905	0.2440
Bladder ultrasonography	5947	0.6145	3.6211	0.2086
Secondary analysis				
UDS	4524	0.6149	3.6693	0.4870
Clinical history and UDS in MUI	5126	0.6458	3.7169	0.4967
Clinical history and UDS in OAB	5198	0.5866	3.6429	0.2343
Clinical history and bladder ultrasonography in MUI	5768	0.6540	3.6891	0.3523
Clinical history	5801	0.6175	3.6905	0.2440
Bladder ultrasonography	5947	0.6145	3.6211	0.2086
Clinical history and bladder ultrasonography in OAB	5965	0.5957	3.6361	0.1004

Results of primary analysis for the outcome of women successfully treated

As shown in *Table 38*, the diagnostic strategy of clinical history leads to an additional 26 cases per 10,000 women successfully treated compared with UDS at an additional cost of £1278 per woman. This results in an ICER of £491,100 per woman successfully treated, which means that an additional £491,100 is required for each additional woman to be successfully treated as a result of a diagnosis by clinical history compared with a diagnosis using UDS. The scatterplot in *Appendix 10* (see *Figure 52*) shows the modelled uncertainty in the cost and clinical effectiveness between UDS and clinical history from 10,000 Monte Carlo simulations. The result of each simulation is plotted on the cost-effectiveness plane providing information about the joint density of the differences in cost and clinical effectiveness between the two strategies. It is evident that even though UDS is almost certainly a less costly diagnostic strategy, it is uncertain whether or not it is more effective than clinical history. It is thus uncertain whether or not the strategy based on UDS dominates clinical history or whether it provides savings for one woman less successfully treated.

TABLE 38 Deterministic results for non-dominated strategies of the primary analysis for women successfully treated

Strategy	Cost (£)	Incremental cost (£)	Clinical effectiveness	Incremental effectiveness	ICER (£)
UDS	4524	–	0.6149	–	–
Clinical history	5801	1278	0.6175	0.0026	491,100

The results of the PSA based on Monte Carlo simulation can also be used to illustrate, with a CEAF, the probability that the optimal strategy, in terms of maximising the net benefit, is cost-effective under current uncertainty at different levels of decision-makers' willingness to pay (WTP) per additional woman successfully treated. As illustrated in the CEAF for the comparison between UDS and clinical history (see *Appendix 10*, *Figure 53*) for any value of WTP per woman successfully treated below £100,000, the probability that UDS is cost-effective exceeds 61%.

Given that there is not a pre-specified threshold of WTP for an additional woman successfully treated, as in the case of QALYs for which £20,000 to £30,000 are the recommended cut-off points by NICE,[192] the identification of the probability of UDS being cost-effective is less straightforward and subject to uncertainty owing to the number of assumptions required. Evidence suggests that the QoL of a woman successfully treated is approximately 0.92[173] and that improvements could be maintained for 5 years,[161,166,193] which results in 4.6 QALYs. If a woman remained symptomatic for these years then she would have 3.2 QALYs. This is a multiplication of the weighted average of the initial QoL of each syndrome (see *Table 36*) and the number of women with each syndrome (see *Table 27*), which leads to a value of 0.69, with the number of years. The QALY gain from becoming subjectively cured is thus 1.4, and if the WTP for a QALY is £20,000, £28,000 for a woman subjectively cured could be an acceptable WTP threshold. At this value, there is a probability of 86% that UDS is the optimal strategy.

Results of primary analysis for the outcome of quality-adjusted life-years

In terms of QALYs, the diagnostic strategy of clinical history results in an additional 0.0212 QALYs gained per woman compared with UDS. Given the additional cost of £1278 per woman, the mean ICER for clinical history compared with UDS was estimated at £60,200 per QALY (*Table 39*). The results of the PSA show that UDS is likely to be the most cost-effective diagnostic strategy for the commonly used £20,000–30,000 threshold of WTP for a QALY[192] with 72% and 61% probability, respectively. If decision-makers are willing to invest more than £60,200 for an additional QALY, clinical history becomes the optimal strategy with a probability ranging from 52% to 57% as the WTP per QALY increases from £60,200 to £100,000. The incremental cost-effectiveness scatterplot and the CEAF are shown in *Appendix 10* (see *Figures 54* and *55*).

Results of primary analysis for the outcome of detrusor overactivity cases detected

As evident in *Table 37*, UDS dominates clinical history as it is less expensive and detects a greater number of cases of DO. However, the outcome of DO cases detected should be used with caution given that DO constitutes a UDS observation rather than a clinical symptom, which consequently favours UDS. The dominance of UDS is also depicted in a scatterplot and CEAF (see *Appendix 10*, *Figures 56* and *57*).

TABLE 39 Deterministic results for non-dominated strategies of the primary analysis for QALYs

Strategy	Cost (£)	Incremental cost (£)	Clinical effectiveness	Incremental effectiveness	ICER (£)
UDS	4524	–	3.6693	–	–
Clinical history	5801	1278	3.6905	0.0212	60,200

Secondary analysis

In this analysis, apart from the main strategies explored in the primary analysis, four further strategies that use a diagnostic test as an adjunct to the clinical history have been included. In these strategies, a diagnostic test (UDS, bladder ultrasonography) is conducted when clinical history has indicated either OAB or MUI. This analysis aims to provide a further insight into the cost-effectiveness of performing a diagnostic test only in selective subgroups of the population modelled. When UDS is used as an adjunct to the clinical history if OAB or MUI has been indicated from patient history, the strategies have been termed as *clinical history and UDS in OAB* and *clinical history and UDS in MUI,* respectively. Similarly, when the bladder ultrasonography test complements clinical history, the terms *clinical history and BUS in OAB* as well as *clinical history and BUS in MUI* are used.

The mean estimated costs and outcomes for each of the seven strategies compared in the secondary analysis are presented in *Table 37*. In terms of cost, UDS is again the least expensive strategy with a mean cost of £4524. In terms of outcomes, conducting a diagnostic test when clinical history suggests MUI appears to be more effective in all three outcomes considered in the analysis. More specifically, performing bladder ultrasonography if the patient's history indicates MUI is the most effective strategy when the outcome is measured in women successfully treated, while conducting UDS if MUI is indicated by the clinical history is the most effective strategy when the outcome is measured in terms of QALYs and DO cases detected.

Results of secondary analysis for the outcome of women successfully treated

Appendix 10, *Figure 58*, illustrates the mean cost and clinical effectiveness in terms of women successfully treated for all strategies included in the secondary analysis. The line connecting the estimates for UDS, to that for the strategy of clinical history and UDS in MUI, then to that for clinical history and BUS in MUI, creates a cost-effectiveness frontier. These strategies, as indicated in *Figure 58*, have been dominated. The deterministic results of secondary analysis for the non-dominated strategies are presented in *Table 40*. The results show that performing UDS when clinical history indicates MUI means there will be an additional 309 successfully treated women per 10,000 and an additional cost of £603 per woman. This leads to an ICER of £19,500 per woman successfully treated. Performing a bladder ultrasonography test when clinical history indicates MUI requires an additional £641 per woman and leads to an additional 82 cases per 10,000 women successfully treated compared with UDS, which gives an ICER of £78,600 for an additional woman successfully treated.

The scatterplot in *Appendix 10*, *Figure 59*, depicts the overall uncertainty in the cost and clinical effectiveness (women successfully treated) when 10,000 random draws from parameter distributions in a Monte Carlo simulation process are used instead of mean estimates. Given the degree of overlap in the results obtained from the different strategies compared, it is evident that there is uncertainty as to which diagnostic strategy is optimal. This uncertainty has been graphed in the CEAF in *Appendix 10*, *Figure 60*. As shown in *Figure 60*, if the WTP for a woman successfully treated is below £19,500, UDS is the optimal strategy under current uncertainty with a probability of being cost-effective above 52%. Above this WTP threshold, clinical history with UDS in MUI becomes the optimal strategy. At WTP values between £19,500 and £78,600 per woman successfully treated, there is a probability of 37–48% that clinical history and UDS in MUI is the strategy that maximises the net benefit. If a decision-maker is willing to invest more than £78,600 per woman successfully

TABLE 40 Deterministic results for non-dominated strategies of the secondary analysis for women successfully treated

Strategy	Cost (£)	Incremental cost (£)	Clinical effectiveness	Incremental effectiveness	ICER (£)
UDS	4524	–	0.6149	–	–
Clinical history and UDS in MUI	5126	603	0.6458	0.0309	19,500
Clinical history and bladder ultrasonography in MUI	5768	641	0.6540	0.0082	78,600

treated, clinical history with bladder ultrasonography in MUI becomes the optimal choice. At the WTP threshold of £28,000 per woman successfully treated, which, according to the corresponding section of the primary analysis, was considered to be an acceptable WTP threshold for this outcome, there is a probability of almost 50% that clinical history with UDS in MUI is the optimal strategy.

Results of secondary analysis for the outcome quality-adjusted life-years

In terms of QALYs, the only non-dominated strategies are the UDS and the UDS as an adjunct to the clinical history in women with patient history of MUI (see *Appendix 10*, *Figure 61*). As shown in *Table 41*, the latter strategy leads to an additional 0.476 QALYs gained per woman at an additional cost of £603. This means that the strategy of performing UDS in women with a clinical history of MUI requires an additional £12,700 for an additional QALY compared with UDS, which is considered to be cost-effective. The results of the PSA indicate that at commonly cited ceiling ratios between £20,000 and £30,000 per QALY, clinical history and UDS in MUI is likely to be the optimal strategy with a probability of being cost-effective at 72% and 78%, respectively. The graphs representing the uncertainty in the estimated costs and QALYs and the uncertainty around the optimal strategy across a range of possible values of WTP for an additional QALY are shown in *Appendix 10* (see *Figures 62* and *63*).

Results of secondary analysis for the outcome of detrusor overactivity cases detected

According to the evidence of the deterministic analysis presented in *Table 42*, having UDS as an adjunct to the clinical history in women with a patient history of MUI is £603 (£5126) more expensive and more effective, leading to an ICER of £62,100 for an additional DO case detected compared with UDS. According to the results of the PSA, if decision-makers' WTP for a DO case detected is below £62,100, UDS is the optimal strategy. As the WTP for this outcome increases from £62,100 to £100,000, the strategy of clinical history and UDS in women with a clinical history of MUI becomes more cost-effective with a probability ranging from almost 50% to 75%. As in the outcome of women successfully treated, the identification of the optimal strategy and the probability of this strategy to be cost-effective for an additional DO case detected is not straightforward as there is not a predefined threshold of WTP. However, it was shown that the probability of a DO case to become subjectively cured is approximately 60%. In this situation, an increase in QoL from 0.6, which is the QoL of a woman with DO (see *Table 36*), to 0.92 would be expected to occur. That means that the weighted average of the QoL for a DO case detected is 0.79. A DO case detected could thus lead to a 0.95 QALY gain. As a consequence, a plausible WTP threshold for a DO case detected could be £19,000. For this value, there is a 99% probability that UDS is the optimal choice. The cost-effectiveness plane, the scatterplot and the CEAF for this outcome are shown in *Appendix 10* (see *Figures 64–66*).

TABLE 41 Deterministic results for non-dominated strategies of the secondary analysis for QALYs

Strategy	Cost (£)	Incremental cost (£)	Clinical effectiveness	Incremental effectiveness	ICER (£)
UDS	4524	–	3.6693	–	–
Clinical history and UDS in MUI	5126	603	3.7169	0.0476	12,700

TABLE 42 Deterministic results for non-dominated strategies of the secondary analysis for DO cases detected

Strategy	Cost (£)	Incremental cost (£)	Clinical effectiveness	Incremental effectiveness	ICER (£)
UDS	4524	–	0.4870	–	
Clinical history and UDS in MUI	5126	603	0.4967	0.0097	62,100

Sensitivity analyses for the primary and secondary analysis: univariate sensitivity analyses

Reducing the cost of urodynamics

The cost of UDS used in main analyses (£401) represents the weighted average value between elective inpatient and day-case costs and the proportion of patients in each of the two groups for the urology service description. In this sensitivity analysis, the average weighted value across all HRGs (£173) is used.[30] Lowering the value of UDS would lead to an increase in the ICER of any strategy compared with this diagnostic test. As evident in *Table 43*, the ICER of all strategies increased in all three outcomes compared with the deterministic results of the main analysis without any impact on the conclusions drawn.

TABLE 43 Results of univariate sensitivity analyses for primary and secondary analysis

	ICER (£)		
Strategy	Women successfully treated	QALYs	DO cases detected
Reducing the cost of UDS (primary analysis)			
UDS	–	–	–
Clinical history	602,800	73,900	Dominated
Reducing the cost of UDS (secondary analysis)			
UDS	–	–	–
Clinical history and UDS in MUI	23,200	15,100	74,000
Clinical history and bladder ultrasonography in MUI	100,000	Dominated	Dominated
Increasing the cost of sling surgery by 50% (primary analysis)			
UDS	–	–	–
Clinical history	706,800	86,600	Dominated
Increasing the cost of sling surgery by 50% (secondary analysis)			
UDS	–	–	–
Clinical history and UDS in MUI	17,600	11,400	56,000
Clinical history and bladder ultrasonography in MUI	119,100	Dominated	Dominated
Lowering the utility weight of women subjectively cured (primary analysis)			
UDS	–	–	–
Clinical history	491,000	Dominated	Dominated
Lowering the utility weight of women subjectively cured (secondary analysis)			
UDS	–	–	–
Clinical history and UDS in MUI	19,500	22,500	62,200
Clinical history and bladder ultrasonography in MUI	78,600	Dominated	Dominated

Increasing the cost of sling surgery by 50%

In this sensitivity analysis, the cost of sling surgery was increased from £3917 to £5876 with the purpose of capturing other possible adverse events, such as bladder injury, vaginal erosion or groin pain,[160] that were not modelled. An increase of the cost of sling surgery led to an increase of the ICER of clinical history because of the larger number of women with DO (or OAB) undergoing surgeries for MUI or SUI as a result of a misdiagnosis. Having UDS as an adjunct to the clinical history when MUI is indicated by the patient does, in theory, mean fewer women with DO (or OAB) undergoing sling surgery owing to the larger sample size on which diagnostic accuracy is applied. This explains the decrease of the ICER of clinical history and UDS in MUI strategy. The results obtained from the main analyses remained unchanged.

Lowering the utility weight of women subjectively cured

This sensitivity analysis explores the impact that a lower QoL score for women becoming subjectively cured would have on model results. Given that more women become subjectively cured from any strategy containing clinical history than with UDS and bladder ultrasonography test alone, lowering the utility weight would result in an increase of the ICER for the different strategies of clinical history. As shown in *Table 43*, lowering the utility weight of women becoming subjectively cured from 0.92 to 0.84 makes clinical history dominated by UDS in the primary analysis and leads to an increase in the ICER of clinical history and UDS in MUI from £12,700 to £22,500 in secondary analysis. Again, the conclusions drawn from the main analyses remained unchanged.

Sensitivity analyses for the primary and secondary analysis: multivariate sensitivity analyses

Changing the accuracy data

In the main analyses, UDS was used as a reference standard. This possibly led to an overestimation of its diagnostic accuracy and subsequently an overestimation of the results obtained from the economic evaluation. In addition, clinical history can be complemented by urinary diaries (clinical assessment) and possibly provide a more accurate diagnosis of the underlying syndrome. For this reason, this sensitivity analysis explores the impact on the results generated from the main analysis if lower accuracy is assumed for UDS and higher for clinical history. A 14% probability of false-positive diagnoses for MUI when DO or SUI are the true conditions and 25% probability of false-positive diagnoses for SUI when MUI is the true condition were assumed for UDS.[64] For clinical history, a 53.4% accuracy of detecting DO from OAB predominant symptoms was assumed.[150] As seen in *Table 44*, this has significantly improved the ICERs but without a major impact on the conclusions from the primary analysis. The ICER of clinical history from the primary analysis was reduced from £60,204 to £31,000, which is slightly over the upper WTP threshold for a QALY. The only change in the decision was in the case of DO cases detected, in which the clinical history and UDS in MUI strategy became cost-effective, as opposed to dominated, considering the ICER found from the sensitivity analysis (£14,100) is below £19,000, which was considered a plausible value of WTP for a DO case detected. With this change, clinical history and UDS in MUI is the optimal choice across the three outcomes used in the analysis.

Using the lowest rates of clinical effectiveness elicited from expert opinion

In this sensitivity analysis, the lowest rates of clinical effectiveness elicited from expert opinion are used to examine the impact of lower effectiveness in cases of a misdiagnosis on the model results (see *Clinical data*). For this analysis, the clinical effectiveness of BTX-A, when the true conditions are MUI and SUI, were decreased from 0.329 and 0.143 to 0.200 and 0.133, respectively. The clinical effectiveness of PTNS and SNS in women without DO was also reduced from 0.314 and 0.271 to 0.283 and 0.233, respectively. Finally, the clinical effectiveness of colposuspension in women with DO was reduced from 0.163 to 0.153. Reducing the effectiveness of interventions in situations of a misdiagnosis was expected to lead to an increase in the ICER of the strategies compared with UDS. This is evident in the results of *Table 44*. The results obtained from the sensitivity analysis did not impact on the conclusions drawn from the main analyses.

TABLE 44 Results of multivariate sensitivity analyses for primary and secondary analysis

Strategy	ICER (£)		
	Women successfully treated	QALYs	DO cases detected
Changing the accuracy data (primary analysis)			
UDS	–	–	–
Clinical history	63,700	31,000	Dominated
Changing the accuracy data (secondary analysis)			
UDS	–	–	–
Clinical history and UDS in MUI	16,500	10,500	14,100
Clinical history and bladder ultrasonography in MUI	39,600	Dominated	Dominated
Using the lowest rates of clinical effectiveness elicited from expert opinion (primary analysis)			
UDS	–	–	–
Clinical history	Dominated	66,900	Dominated
Using the lowest rates of clinical effectiveness elicited from expert opinion (secondary analysis)			
UDS	–	–	–
Clinical history and UDS in MUI	20,900	13,100	62,300
Using the highest rates of clinical effectiveness elicited from expert opinion (primary analysis)			
UDS	–	–	–
Clinical history	30,700	21,000	Dominated
Bladder ultrasonography	34,800	Dominated	Dominated
Using the highest rates of clinical effectiveness elicited from expert opinion (secondary analysis)			
UDS	–	–	–
Clinical history and bladder ultrasonography in MUI	14,000	Dominated	Dominated
Clinical history and UDS in MUI	Dominated	11,500	63,000
Clinical history	Dominated	83,800	Dominated
Assuming one diagnostic test is given in the treatment pathway (primary analysis)			
UDS	–	–	–
Clinical history	555,000	69,500	Dominated
Assuming one diagnostic test is given in the treatment pathway (secondary analysis)			
UDS	–	–	–
Clinical history and UDS in MUI	22,300	14,400	78,600

Using the highest rates of clinical effectiveness elicited from expert opinion

In this sensitivity analysis, effectiveness rates elicited from expert opinion were taken to their highest value (see *Clinical data*). More specifically, the effectiveness of BTX-A, when the true conditions are MUI and SUI, were increased to 0.483 and 0.375, respectively. For the effectiveness of PTNS and SNS in women without DO the values 0.567 and 0.555 were used. Finally, the effectiveness of colposuspension in women with DO was increased to 0.386. According to the results of *Chapter 5*, *Table 16*, this sensitivity analysis affected the initial decision in one outcome of the base-case and secondary analysis. More specifically, in the base-case analysis, clinical history became cost-effective in terms of QALYs with an ICER of £21,000 compared with UDS. In the secondary analysis, clinical history and bladder ultrasonography in MUI became cost-effective with an ICER of £14,000 for a woman successfully treated compared with UDS.

Assuming one diagnostic test is given in the treatment pathway

This sensitivity analysis explores the impact on the results of the main analyses if one diagnostic test was performed in the treatment pathway prior to the initiation of more invasive interventions, as indicated in the 2006 NICE report.[36] As shown in *Table 44*, the ICER of all strategies compared with UDS was increased without any impact on the conclusions from the main analyses. This finding implies that at the level of accuracy data used in the model, having one UDS is more cost-effective.

Discussion

Principal findings

The results of the deterministic analysis suggest that a patient management strategy based on using bladder ultrasonography to select treatment options is more costly and less effective than the strategies based on UDS and clinical history. Given that the results of the main study show that bladder ultrasonography is of limited diagnostic value, the uncertainty around the deterministic results of this strategy was not explored in a PSA.

With regard to the other two strategies, treating according to the primary symptoms in clinical history is more effective than UDS, leading to an additional 26 cases per 10,000 women successfully treated and 0.02 QALYs gained per woman. This comes at an additional cost of approximately £1300 per woman arising from the costs of the treatments used. Given current acceptable thresholds, this additional cost would not be justified by the clinical gain. In terms of DO cases detected, UDS detects more cases than clinical history at a lower cost. Thus, the results of the primary analysis suggest that UDS would be the preferred strategy on cost-effectiveness grounds for women with predominant symptoms of OAB. This result holds for all three outcome measures used in the primary analysis. This finding has arisen because, although the UDS test itself adds to the cost of clinical history, basing treatment decisions on the UDS leads to more efficient use of expensive interventions than basing decisions on clinical history alone.

In the secondary analysis, the strategy of giving UDS to women with a clinical history of MUI is more effective than universal UDS. The selective use was shown to result in an additional 309 cases per 10,000 women successfully treated and the QALY gain per woman was 0.05. The additional cost per woman was £600. This strategy would be considered value for money according to acceptable thresholds in the UK. Based on cost per DO case detected, UDS was not deemed cost-effective for these women given the ICER of £62,000. This is not a surprising finding as DO is an urodynamic observation and not a clinical symptom.

Univariate and multivariate sensitivity analyses were performed to explore the robustness of the model results to changes in the assumptions made. Conclusions drawn from the main analysis remained robust to all sensitivity analyses apart from the scenario in which the highest effectiveness rates were assumed in cases of misdiagnosis.

Strengths and limitations of the study

The prevalence and accuracy data were estimated in the primary study and the main clinical pathways were parameterised with clinical effectiveness data mostly from systematic reviews and meta-analyses. In addition, costs were drawn from national sources and the development of the model relied on national guidelines. These are likely to enhance the generalisability of the study's findings. Finally, all assumptions used in the model were agreed a priori and key assumptions have been tested in sensitivity analyses.

However, there are a few weaknesses. Instead of clinical history, the best alternative strategy to compare the cost-effectiveness of UDS with could have been clinical assessment. Urinary diaries, for example, are used to complement clinical history and provide a joint decision based on clinicians' assessment rather than on patient's history alone. Furthermore, ambulatory UDS has been shown to be more sensitive in detecting DO,[194] or generally the underlying syndrome of UI,[195,196] than conventional UDS and thus the results of the

economic evaluation may overstate the cost-effectiveness of UDS, as it has been used as a reference standard. Ambulatory UDS was used as part of the primary study in situations of a negative UDS, but in the subgroup population used for the model, the number of observations were not enough to be tested in a sensitivity analysis ($n = 7$).

A further limitation relates to the QoL estimates used for the outcomes experienced beyond the primary study end point. In the absence of robust utility weights, it was assumed that women maintain their QoL if an invasive intervention is not successful and that all women have a utility score of 0.92 once they become subjectively cured. In reality, however, utility decrements can occur after invasive and poorly tolerated diagnostic strategies, such as UDS or surgeries. In addition, utility gains tend to differ between different syndromes and severity of symptoms. Despite the low accuracy of the QALY estimates used in the model, there was a significant consistency between the results obtained from the primary outcome (cost per woman successfully treated) and the results obtained using QALYs.

Furthermore, in the absence of available literature on the clinical effectiveness of the different interventions modelled in situations of misdiagnosis, values had to be elicited from expert opinion (see *Clinical data*). Probabilistic and deterministic sensitivity analyses were undertaken to reduce the uncertainty around these parameter estimates and explore their impact on conclusions drawn. Another limitation relates to the fact that the order of the interventions modelled for the treatment of women with predominant symptoms of OAB might be slightly different in practice. For example, after an unsuccessful sling surgery, instead of going through a Burch colposuspension, women may repeat the sling surgery. However, such differences are not expected to significantly impact on the results of the model. Finally, the syndromes of OAB, SUI and MUI are often associated with substantial personal costs. These costs are not included in economic analyses conducted from a health-care perspective and thus economic evaluations from a societal perspective could shed further light onto the cost-effectiveness of the alternative diagnostic strategies for women with predominant symptoms of OAB.

Strengths and limitations in relation to other studies

Economic evaluations that have explored the role of UDS in the treatment of women with predominant symptoms of SUI[65] or genuine SUI[64,89] have concluded that UDS is not a cost-effective diagnostic test. However, for women with predominant symptoms of OAB, relevant economic evidence on the role of UDS is lacking, while clinical evidence remains inconclusive.[17,29,66]

The results of the economic evaluation of this chapter suggest that UDS has a role to play in the treatment pathway of women with OAB. An investigation into whether or not a diagnostic test is more cost-effective when performed in only specific subgroups of women concluded that the most cost-effective diagnostic strategy is to perform UDS in women with a clinical history of mixed incontinence. This finding seems to be further supported by evidence from clinical studies indicating that in women with OAB only an urodynamic observation of DO does not have an impact on the outcome of more invasive interventions,[84,86,197] but also from evidence indicating that mixed urinary symptoms tend to be more common than mixed UI on UDS.[13,198] The magnitude of DO appears to be close to 50%,[198] as it was also found in the primary study (55%).

Meaning of the study

The results of the economic evaluation showed that UDS is a cost-effective diagnostic strategy for women with predominant symptoms of OAB for whom conservative treatments were not effective. The primary analysis suggested that UDS is the most cost-effective strategy across all outcomes providing significant cost-savings for a small reduction in outcome. The secondary analysis suggested that selective UDS, restricting its use to women with a clinical history of MUI, is a cost-effective strategy.

Unanswered questions and future research

In this study, UDS was used as the reference standard, which may have led to an overestimation of its diagnostic accuracy. Evidence suggests that ambulatory UDS is more sensitive in detecting DO[194] and other syndromes of UI.[195,196] Furthermore, a more appropriate comparator for UDS would have been the clinical assessment, which would provide an overall clinical diagnosis based on the clinical history (validated questionnaires) and urinary diaries. Thus, a comparative evaluation of the costs and benefits of UDS and clinical assessment using ambulatory UDS as a reference standard would help gain further insight into the cost-effectiveness of UDS in women with predominant symptoms of OAB.

Furthermore, there are significant inconsistencies in the instruments used in the literature to capture QoL in women with urinary symptoms. Disease-specific instruments, such as the incontinence-specific QoL questionnaire and other non-preference-based measures that are often used in women with OAB symptoms, offer limited usefulness as outcome measures in cost-effectiveness analyses;[199] although, mapping algorithms have now started to be developed.[200] Nevertheless, for conditions such as OAB, in which symptoms can also be associated with non-health impacts (e.g. embarrassment), instruments that go beyond health and capture wider changes in individual's QoL may be deemed more pertinent. An example is the ICECAP-A measure,[201] which incorporates capability attributes on Attachment, Security, Role, Enjoyment and Control. Further research is required in the use of such instruments in women with urinary symptoms.

Finally, it appears that further research is needed to resolve current decision uncertainty around the cost-effectiveness of UDS in all women with predominant symptoms of OAB or in only those with a clinical history of MUI. A measure of the maximum possible value of any such research is the expected value of perfect information (EVPI). EVPI is estimated from the difference between the expected net benefit with perfect and current information, and multiplied by the size of the population that could benefit from additional information. Based on estimates of prevalence of women with predominant symptoms of OAB in the UK[202–204] and an incidence rate of 6% (54,000 women),[205] while assuming a time horizon of 10 years and an annual discount rate of 3.5%, the EVPI was estimated at almost £36M and £77M at a WTP threshold of £28,000 per additional woman successfully treated for the primary and secondary analyses, respectively (see *Appendix 10*, *Figures 67* and *68*). These figures reflect the large expected opportunity loss associated with current decision uncertainty and the need for further primary research.

Chapter 8 Discussion

Introduction

We have completed five distinct studies that address the following aims in this HTA project:

1. To determine the accuracy (sensitivity, specificity, predictive values, ROC curves, AUC) of transvaginal BWT as the index test and UDS as the reference standard to investigate DO in women with OAB.
2. To determine the reproducibility of the index test BWT and audit the quality of index test and the reference standard.
3. To compare the acceptability of bladder ultrasonography with UDS.
4. To determine the outcome of treatment responses based on the UDS diagnoses at 6 and 12 months.
5. To determine the cost and cost-effectiveness of UDS in the management pathway of women with OAB using a decision-analytic model.

Each of these studies has been described in detail and the main findings and conclusions reported in the four preceding chapters. This chapter focuses on the key findings and limitations.

Principal findings

Contrary to findings from previous studies, we found no evidence that transvaginal BWT had any useful role in the diagnosis of DO, regardless of cut-off point (AUC 0.53, 95% CI 0.48 to 0.57). Extensive sensitivity analyses and subgroup analyses were carried out without any evidence of an increase in the performance of BWT; these included excluding the population with a history of MUI and also excluding those with 'dry' OAB. Furthermore, BWT had no relationship to symptoms as measured by ICIQ-OAB score, either on presentation or in the long term, so has no predictive or prognostic value as a test in this condition.

The accuracy study is the largest and only multicentre study evaluating bladder ultrasonography and was undertaken in a pragmatic manner to reflect NHS practice. The fact that we picked up DO in 60% of the OAB/urgency-predominant MUI population and this is similar to other studies on DO prevalence in OAB[17] indicates that our sampling was representative. However, our study differed in that we excluded cases of pure SUI as these are accurately diagnosed on the basis of symptoms and examination without need for UDS or ultrasonography. We focused on the value of ultrasonography in those who do not have a clear clinical diagnosis of SUI, as this is the place in the course of the diagnostic pathway where ultrasonography would be used (if accurate) to replace UDS.

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 in BWT 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. In spite of 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 model-based economic evaluation concluded that UDS is a cost-effective strategy for women with predominant symptoms of OAB for whom conservative treatments were not effective. Further analyses concluded that this may be a more pertinent strategy in the population of women with mixed urinary continence. Bladder ultrasonography was not a cost-effective test.

In current practice, UDS diagnosis appeared to have an effect 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 USI if given a combined diagnosis of DO and USI than those with normal UDS results). Women diagnosed with DO were three times more likely to have reported taking bladder relaxants in the follow-up period up to 2 years post test. This raises a question as to whether or not UDS can increase clinicians' willingness to prescribe bladder relaxants or patients' compliance and perseverance with antimuscarinics. NICE[36] and other guidelines around the world[206] recommend that UDS is not necessary before conservative management, including pharmacotherapy. However, our study findings raise the possibility that a UDS diagnosis of DO may have a role in helping women to continue treatment with antimuscarinics.

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 may play a part in improving outcomes of subsequent treatment for women with suspected OAB.

Strengths and limitations

The robust design and execution of our test accuracy study provides confidence that our estimates of diagnostic accuracy are valid. We complied with, and reported all, criteria for a high-quality test accuracy evaluation.[207] A very high proportion of patients completed index test verification with the reference standard [97% (644/666)], ensuring a near complete verification design. Other strengths include a recruitment of a large sample of 644 women with both index and reference standard results, in excess of our original sample size calculation of 600, ensuring we had sufficient power to exclude the possibility of BWT being a clinically important test. Recruitment was prospective and consecutive in a multicentre setting of women with varying ages, parity, ethnicity and disease severity, increasing the generalisability of the results. We minimised the risk of bias by ensuring that the index tests and reference standard were performed independently and interpreted blind to each other. We ensured that tests were undertaken by trained NHS staff (ensuring generalisability) but according to standard protocols for both index and reference standard testing with quality assurance checks to ensure quality control. A pre-specified statistical plan was followed, and comprehensive independent oversight was maintained via an independent Trial Steering Committee (TSC).

The study recruited women of varying ages, parity, ethnicity, disease severity and from diverse regions across the NHS. The large sample also gave the opportunity to rigorously analyse the acceptability of both the index test and reference standard and to provide data for the concurrent health economic evaluation. The follow-up of patients up to 2 years post test provided invaluable data with regard to the subsequent treatment and clinical outcomes in this group.

In spite of very careful planning of our study design, we acknowledge several limitations of our study. Inclusion of urgency-predominant MUI was based on the patient's history, which revealed whether or not stress or urgency incontinence was most bothersome at the time of presentation. Other studies have utilised the MESA questionnaire subscale scores for urgency and SUI,[208] and categorised women who had higher MESA scores for urgency incontinence (than their MESA subscale scores for SUI) as urgency-predominant MUI along with urogenital distress inventory scores, bladder diaries and UDS.[16] Although we have not objectively assessed the severity of each type of incontinence before deciding if the patient suffered from urgency-predominant MUI, we have utilised direct patient questioning of which is the most bothersome type of incontinence on history taking, which is a true reflection of routine clinical practice.

Although our reference standard, UDS, is the accepted 'gold standard' for diagnosis, it is known for its uncertain reproducibility. In previous urodynamic trials that evaluated the reproducibility of UDS in healthy patients and patients with OAB, inconsistency between serial urodynamic procedures were a common finding.[106–109] In a multicentre study with serial UDS in patients with OAB, there was increased variability in pressure measurements than volume measurements.[209] Using an imperfect test as a reference standard may underestimate the test accuracy of our index test, but not to the point of finding no relationship at all, as occurred in the BUS. Our investigation of the long-term follow-up results showed prognostic relationships with UDS, indicating that it does have discriminatory value.

We did not address a study objective to examine the add-on value of bladder ultrasonography to the information already gained from non-invasive tests such as bladder diaries, clinical history and the disease-specific QoL questionnaires. As we did not find bladder ultrasonography to be an accurate test for detection of DO on its own, this objective became redundant. However, we did examine which clinical history variables were associated with DO diagnosis, and baseline ICIQ-OAB score appeared to be highly associated. We were unable to perform ambulatory UDS as a secondary verification test despite its inclusion in the protocol as it was not available at all centres. Sensitivity analyses for accuracy of BWT of a small subgroup of women ($n = 14$) who underwent ambulatory UDS at the main recruiting centre did not show any difference.

The follow-up of this cohort of women is based on selected treatments and thus does not provide evidence of the clinical effectiveness of different interventions owing to the inevitable confounding by indication. We did not have enough power to differentiate between the effect of the different types of treatment; instead, we combined them into a 'meta-effect' of treatment concurrent with diagnosis.

Implications for practice

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, particularly in the MUI subgroup once conservative treatments were exhausted. Offering UDS earlier on in the management of mixed incontinence may help achieve a greater degree of patient satisfaction and save valuable money and time for the NHS.

Recommendations for research

Diagnostic accuracy of individual components of office evaluation of OAB may be different to the composite test accuracy of all the components of office evaluation. Studying composite test accuracy of various components in this context may be a highly complicated exercise. Further studies need to be planned to look into composite test accuracy of office evaluation with or without UDS in OAB.

In women with OAB/urgency-predominant MUI, RCTs comparing treatment based on UDS diagnoses compared with clinical diagnoses (history and examination alone), and related health economic evaluations for these interventions, are required to consolidate the role of UDS in the management of OAB/MUI women, as already carried out for SUI.[87]

Acknowledgements

We gratefully thank all who worked very hard as research and nursing staff at the participating centres. We also thank the nursing managers from each of the trusts for their support.

We thank the members of the TSC and DMC for their assistance throughout the project: Professor D Tincello (Professor of Urogynaecology and Honorary Consultant Gynaecologist, University of Leicester) chairperson of TSC; Jean Perks (Patient Representative, Birmingham Women's Hospital); Dr Patrick Chein, (Consultant Obstetrician and Gynaecologist, University of Dundee) Chairperson of DMC; Associate Professor Jonathan Cook (Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford) independent statistician; and Dr Karen Ward (Consultant Urogynaecologist, Central Manchester Hospitals) independent clinician.

The BUS Trial was co-ordinated by Birmingham Clinical Trials Unit at the University of Birmingham and we acknowledge all of the hard work of all the staff involved in the study, Victoria Parker (trial co-ordinator), Elizabeth Lawson, Leanne Fulcher, Laura Gennard and Lisa Leighton. We thank Dr Moji Balogun for training the investigators in bladder ultrasonography. We also thank Nicholas Hilken for designing and developing the study database and Versha Cheed who was the study statistician. We acknowledge the help of Dr Anuradha Sajja, Kal Perkins, Emma Millership, Kelly Hard, Balvir Sehjal, Lorna Smith and Joan Crook at the Birmingham Women's Hospital. We are grateful to UK Continence Society and British Society of Urogynaecology for endorsing this study.

Contribution of authors

Dr Suneetha Rachaneni (Clinical Research Fellow) recruited patients, provided training to the study investigators, contributed to delivery and interpretation of the accuracy and reproducibility studies, and produced the first draft of the final report.

Ms Shanteela McCooty (Urogynaecology Research Nurse) contributed to the patient recruitment, UDS testing of the accuracy and acceptability studies and performed the UDS audit.

Mr Lee J Middleton (Senior Statistician) contributed to the statistical analysis and interpretation of the accuracy, reproducibility, acceptability and outcomes assessment studies of BUS.

Dr Victoria L Parker (Senior Trial Co-ordinator) was responsible for the day-to-day management and delivery of all study components.

Dr Jane P Daniels (Deputy Director of BCTU and coapplicant) contributed to the design, delivery, and interpretation of BUS and overall editing of the final peport.

Professor Arri Coomarasamy (Professor of Gynaecology and coapplicant) contributed to the design and interpretation of BUS.

Dr Tina S Verghese (Clinical Research Fellow) contributed to the data collection, delivery and interpretation of the outcomes assessment.

Dr Moji Balogun (Consultant Radiologist and coapplicant) contributed to the design of the accuracy study and delivery of the training to the study investigators.

Dr Ilias Goranitis (Economics Research Fellow), **Dr Pelham Barton** (Reader in Mathematical Modelling) and **Professor Tracy E Roberts** (Professor of Health Economics and coapplicant) contributed to the design, analysis and interpretation of the economic evaluation.

Professor Jon J Deeks (Professor of Medical Statistics and coapplicant) contributed to the design, analysis and interpretation of BUS.

Dr Pallavi Latthe (Consultant Urogynaecologist and lead applicant) conceived the idea for the study, contributed to the design, delivery, analysis and interpretation of BUS, recruited patients and, as chief investigator, had overall responsibility for BUS.

All authors contributed to the writing of their respective chapters and overall conclusions of the final report.

Bladder Ultrasound Study collaborators

Number of participants are shown in brackets. The local principal investigator is shown in bold.

Birmingham Women's Hospital, Birmingham (254). A Sajja, C Burton, E Millership, K Perkins, M Parsons, **P Latthe**, P Toozs-Hobson, S McCooty, S Rachaneni; City General Hospital, North Staffordshire (10). A Clark, F O'Mahony, I Siddiqui, **J Cooper**, J Stokes, S Banergee, S Jerreat; Crosshouse Hospital, Ayrshire (9). D Gilmore, L Oliver, M Henry, M Raed, **W Agur**; Derriford Hospital, Plymouth (5). B Cree, H Hollands, M Vij, **R Freeman**; Manor Hospital, West Midlands (8). N Ensor-Coton, **P Jain**; Mayday Hospital, Surrey (92). A Sultan, F Lone, M Naidu, M Lawrence, **R Thakar**; Medway Hospital, Kent (109). A Patil, D Chakani, **J Duckett**, M Eaton; New Cross Hospital, West Midlands (16). **A Elnaqa**, E Virgilio, N Denyer, S Kempson; North Hampshire Hospital, Hampshire (30). G Harding, C Rowe Jones, **C Phillips**, I Mahfouz, J Moody, L Rush; S Singh, T Sayer; Northampton General Hospital, Northamptonshire (6). **A Shulkla**, D Hunt, R Hitchcock; Ormskirk and District General Hospital, Lancashire (19). L Baldwin, **S Aleem**, Z Haslam; Pinderfields General Hospital, West Yorkshire (4). A Hayton-Bott, B Taylor, **S Sundaram**; Royal Bournemouth General Hospital, Dorset (16). D Abbott, **D Kapoor**, L Wilkinson; Royal Hallamshire Hospital, Sheffield (6). D Gopinath, J Metherall, **S Radley**, S Duffy; Sandwell Hospital, West Midlands (2). M Baptiest, **V Olagundoye;** Southern General Hospital, Glasgow (3). C Archibald, **K Guerrero;** St Mary's Hospital London, Greater London (4). A Derpapas, A Kabia, C Hendricken, **V Khullar**; St Mary's Hospital, Manchester (26). A Bryant, **F Reid**, L Dwyer, M Hockey; Staffordshire General Hospital, Staffordshire (26). **A Elmardi**, T Harrison; Stepping Hill Hospital, Cheshire (25). **A Yassin**, G Clarke; The Alexandra Hospital, Worcestershire (14). **D Sinha**, K Powles; The Royal London Hospital, Greater London (3). A Tirlapur, **C Chaliha**, M Gabas, Y Jacobs.

Finally we thank all of the women who consented to participation. The study would not have been possible without them.

Publications

Rachaneni S, Balogun M, Latthe P. Bladder-wall-thickness ultrasound scan in the investigation of LUTS in women: challenges and limitations. *Int Urogynecol J* 2013;**24**:725–8.

McCooty S, Latthe P. Quality control in urodynamics. *Nurs Stand* 2013;**27**:35–8.

Rachaneni S, Arya P, Latthe P. Urinary nerve growth factor: a biomarker of detrusor overactivity? A systematic review. *Int Urogynecol J* 2013;**24**:1603–9.

Rachaneni S, Latthe P. Does preoperative urodynamics improve outcomes for women undergoing surgery for stress urinary incontinence? A systematic review and meta-analysis. *BJOG* 2015;**122**:8–16.

References

1. Haylen BT, de Ridder D, Freeman RM, Swift SE, Berghmans B, Lee J, *et al.* An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction. *Neurourol Urodynam* 2010;**29**:4–20. http://dx.doi.org/10.3109/9781439807217-111
2. Abrams P. Describing bladder storage function: overactive bladder syndrome and detrusor overactivity. *Urology* 2003;**62**:28–37. http://dx.doi.org/10.1016/j.urology.2003.09.050
3. 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. http://dx.doi.org/10.1111/j.1464-410X.2009.08427.x
4. Irwin DE, Milsom I, Hunskaar S, Reilly K, Kopp Z, Herschorn S, *et al.* Population-based survey of urinary incontinence, overactive bladder, and other lower urinary tract symptoms in five countries: results of the EPIC study. *Eur Urol* 2006;**50**:1306–15. http://dx.doi.org/10.1016/j.eururo.2006.09.019
5. Coyne KS, Sexton CC, Bell JA, Thompson CL, Dmochowski R, Bavendam T, *et al.* The prevalence of lower urinary tract symptoms (LUTS) and overactive bladder (OAB) by racial/ethnic group and age: results from OAB-POLL. *Neurourol Urodynam* 2013;**32**:230–7. http://dx.doi.org/10.1002/nau.22295
6. 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. http://dx.doi.org/10.1111/j.1464-410X.2010.09993.x
7. Jimenez-Cidre M, Costa P, Ng-Mak D, Sahai A, Degboe A, Smith CP, *et al.* Assessment of treatment-seeking behavior and healthcare utilization in an international cohort of subjects with overactive bladder. *Curr Med Res Opin* 2014;**30**:1557–64. http://dx.doi.org/10.1185/03007995.2014.918028
8. 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. http://dx.doi.org/10.1111/j.1464-410X.2005.05889.x
9. Jha S, Strelley K, Radley S. Incontinence during intercourse: myths unravelled. *Int Urogynecol J* 2012;**23**:633–7. http://dx.doi.org/10.1007/s00192-011-1583-0
10. Freeman RM, Adekanmi OA. Overactive bladder. *Best Pract Res Clin Obstet Gynaecol* 2005;**19**:829–41. http://dx.doi.org/10.1016/j.bpobgyn.2005.08.002
11. Irwin DE, Mungapen L, Milsom I, Kopp Z, Reeves P, Kelleher C. The economic impact of overactive bladder syndrome in six Western countries. *BJU Int* 2009;**103**:202–9. http://dx.doi.org/10.1111/j.1464-410X.2008.08036.x
12. Mobley DF, Baum N. Etiology, evaluation, and management of nocturia in elderly men and women. *Postgrad Med* 2014;**126**:147–53. http://dx.doi.org/10.3810/pgm.2014.03.2751
13. Digesu GA, Salvatore S, Fernando R, Khullar V. Mixed urinary symptoms: what are the urodynamic findings? *Neurourol Urodynam* 2008;**27**:372–5. http://dx.doi.org/10.1002/nau.20530
14. Herzog AR, Diokno AC, Brown MB, Normolle DP, Brock BM. Two-year incidence, remission, and change patterns of urinary incontinence in noninstitutionalized older adults. *J Gerontol* 1990;**45**:M67–74. http://dx.doi.org/10.1093/geronj/45.2.M67

15. Shumaker SA, Wyman JF, Uebersax JS, McClish D, Fantl JA. Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program in Women (CPW) Research Group. *Qual Life Res* 1994;**3**:291–306. http://dx.doi.org/10.1007/BF00451721

16. Brubaker L, Lukacz ES, Burgio K, Zimmern P, Norton P, Leng W, *et al.* Mixed incontinence: comparing definitions in non-surgical patients. *Neurourol Urodyn* 2011;**30**:47–51. http://dx.doi.org/10.1002/nau.20922

17. Hashim H, Abrams P. Is the bladder a reliable witness for predicting detrusor overactivity? *J Urol* 2006;**175**:191–4. http://dx.doi.org/10.1016/S0022-5347(05)00067-4

18. Andersson KE. Bladder activation: afferent mechanisms. *Urology* 2002;**59**:43–50. http://dx.doi.org/10.1016/S0090-4295(01)01637-5

19. Kinder RB, Mundy AR. Pathophysiology of idiopathic detrusor instability and detrusor hyper-reflexia. An in vitro study of human detrusor muscle. *Br J Urol* 1987;**60**:509–15. http://dx.doi.org/10.1111/j.1464-410X.1987.tb05031.x

20. Brading AF. A myogenic basis for the overactive bladder. *Urology* 1997;**50**:57–67. http://dx.doi.org/10.1016/S0090-4295(97)00591-8

21. Serati M, Salvatore S, Cattoni E, Siesto G, Soligo M, Braga A, *et al.* Female urinary incontinence at orgasm: a possible marker of a more severe form of detrusor overactivity. Can ultrasound measurement of bladder wall thickness explain it? *J Sex Med* 2011;**8**:1710–16. http://dx.doi.org/10.1111/j.1743-6109.2011.02245.x

22. Colli E, Artibani W, Goka J, Parazzini F, Wein AJ. Are urodynamic tests useful tools for the initial conservative management of non-neurogenic urinary incontinence? A review of the literature. *Eur Urol* 2003;**43**:63–9. http://dx.doi.org/10.1016/S0302-2838(02)00494-3

23. Martin JL, Williams KS, Sutton AJ, Abrams KR, Assassa RP. Systematic review and meta-analysis of methods of diagnostic assessment for urinary incontinence. *Neurourol Urodynam* 2006;**25**:674–83. http://dx.doi.org/10.1002/nau.20340

24. Jackson SL. Prevalence and natural history of female incontinence. *J Womens Health* 1998;**7**:472–3.

25. James M, Jackson S, Shepherd A, Abrahms P. Pure stress leakage symptomatology: is it safe to discount detrusor instability? *Br J Obstet Gynaecol* 1999;**106**:1255–8. http://dx.doi.org/10.1111/j.1471-0528.1999.tb08178.x

26. Coyne KS, Kaplan SA, Chapple CR, Sexton CC, Kopp ZS, Bush EN, *et al.* Risk factors and comorbid conditions associated with lower urinary tract symptoms: EpiLUTS. *BJU Int* 2009;**103**:24–32. http://dx.doi.org/10.1111/j.1464-410X.2009.08438.x

27. Parsons M, Amundsen CL, Cardozo L, Vella M, Webster GD, Coats AC, *et al.* Bladder diary patterns in detrusor overactivity and urodynamic stress incontinence. *Neurourol Urodynam* 2007;**26**:800–6. http://dx.doi.org/10.1002/nau.20406

28. Chung SD, Chui B, Kuo HC, Chuang YC, Wang CC, Guan Z, *et al.* Transabdominal ultrasonography of detrusor wall thickness in women with overactive bladder. *BJU Int* 2010;**5**:668–72. http://dx.doi.org/10.1002/nau.21057

29. Digesu GA, Khullar V, Cardozo L, Salvatore S. Overactive bladder symptoms: do we need urodynamics? *Neurourol Urodyn* 2003;**22**:105–8. http://dx.doi.org/10.1002/nau.10099

30. Department of Health. *NHS Reference Costs 2012–13*. London: Department of Health; 2013.

31. Health and Social Care Information Centre. *Hospital Outpatient Activity 2013–14*. London: Health and Social Care Information Centre; 2014.

32. Public Health England. *The NHS Atlas of Variation in Diagnostic Services*. London: Public Health England; 2013.

33. Haylen BT, de Ridder D, Freeman RM, Swift SE, Berghmans B, Lee J, *et al.* An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction. *Int Urogynecol J* 2010;**21**:5–26. http://dx.doi.org/10.1007/s00192-009-0976-9

34. Digesu GA, Hutchings A, Salvatore S, Selvaggi L, Khullar V. Reproducibility and reliability of pressure flow parameters in women. *BJOG* 2003;**110**:774–6. http://dx.doi.org/10.1111/j.1471-0528.2003.02081.x

35. Digesu GA, Khullar V, Panayi D, Calandrini M, Gannon M, Nicolini U. Should we explain lower urinary tract symptoms to patients? *Neurourol Urodyn* 2008;**27**:368–71. http://dx.doi.org/10.1002/nau.20527

36. NICE. *Urinary Incontinence In Women. Clinical Guideline Cg171*. London: NICE; 2013.

37. Pacheco-Unguetti AP, Acosta A, Callejas A, Lupianez J. Attention and anxiety: different attentional functioning under state and trait anxiety. *Psychol Sci* 2010;**21**:298–304. http://dx.doi.org/10.1177/0956797609359624

38. Ku JH, Kim SW, Kim HH, Paick JS, Son H, Oh SJ. Patient experience with a urodynamic study: a prospective study in 208 patients. *J Urology* 2004;**171**:2307–10. http://dx.doi.org/10.1097/01.ju.0000125144.82338.0c

39. Shaw C, Williams K, Assassa PR, Jackson C. Patient satisfaction with urodynamics: a qualitative study. *J Adv Nurs* 2000;**32**:1356–63. http://dx.doi.org/10.1046/j.1365-2648.2000.01627.x

40. Shaw C, Williams KS, Assassa RP. Patients' views of a new nurse-led continence service. *J Clin Nurs* 2000;**9**:574–82. http://dx.doi.org/10.1046/j.1365-2702.2000.00414.x

41. Yeung JY, Eschenbacher MA, Pauls RN. Pain and embarrassment associated with urodynamic testing in women. *Int Urogynecol J* 2014;**25**:645–50. http://dx.doi.org/10.1007/s00192-013-2261-1

42. Cartwright R, Afshan I, Derpapas A, Vijaya G, Khullar V. Novel biomarkers for overactive bladder. *Nat Rev Urol* 2011;**8**:139–45. http://dx.doi.org/10.1038/nrurol.2011.7

43. Oelke M. International Consultation on Incontinence-Research Society (ICI-RS) report on non-invasive urodynamics: the need of standardization of ultrasound bladder and detrusor wall thickness measurements to quantify bladder wall hypertrophy. *Neurourol Urodyn* 2010;**29**:634–9. http://dx.doi.org/10.1002/nau.20834

44. Khullar V, Salvatore S, Cardozo L, Bourne TH, Abbott D, Kelleher C. A novel technique for measuring bladder wall thickness in women using transvaginal ultrasound. *Ultrasound Obst Gynecol* 1994;**4**:220–3. http://dx.doi.org/10.1046/j.1469-0705.1994.04030220.x

45. Elbadawi A, Diokno AC, Millard RJ. The aging bladder: morphology and urodynamics. *World J Urol* 1998;**16**:S10–34. http://dx.doi.org/10.1007/PL00014134

46. Hakenberg OW, Linne C, Manseck A, Wirth MP. Bladder wall thickness in normal adults and men with mild lower urinary tract symptoms and benign prostatic enlargement. *Neurourol Urodyn* 2000;**19**:585–93. http://dx.doi.org/10.1002/1520-6777(2000)19:5<585::AID-NAU5>3.0.CO;2-U

47. Kuhn A, Bank S, Robinson D, Klimek M, Kuhn P, Raio L. How should bladder wall thickness be measured? A comparison of vaginal, perineal and abdominal ultrasound. *Neurourol Urodyn* 2010;**29**:1393–6. http://dx.doi.org/10.1002/nau.20876

48. Latthe PM, Champaneria R, Khan KS. Systematic review of the accuracy of ultrasound as the method of measuring bladder wall thickness in the diagnosis of detrusor overactivity. *Int Urogynecol J* 2010;**21**:1019–24. http://dx.doi.org/10.1007/s00192-010-1144-y

49. Khullar V, Cardozo LD, Salvatore S, Hill S. Ultrasound: a noninvasive screening test for detrusor instability. *Br J Obstet Gynaecol* 1996;**103**:904–8. http://dx.doi.org/10.1111/j.1471-0528.1996.tb09910.x

50. Minardi D, Piloni V, Amadi A, El Asmar Z, Milanese G, Muzzonigro G. Correlation between urodynamics and perineal ultrasound in female patients with urinary incontinence. *Neurourol Urodyn* 2007;**26**:176–82. http://dx.doi.org/10.1002/nau.20327

51. Lekskulchai O, Dietz HP. Detrusor wall thickness as a test for detrusor overactivity in women. *Ultrasound Obstet Gynecol* 2008;**32**:535–9. http://dx.doi.org/10.1002/uog.5370

52. Robinson D, Anders K, Cardozo L, Bidmead J, Toozs-Hoson P, Khullar V. Can ultrasound replace ambulatory urodynamics when investigating women with irritative urinary symptoms? *BJOG* 2002;**109**:145–8. http://dx.doi.org/10.1111/j.1471-0528.2002.01021.x

53. Yang JM, Huang WC. Bladder wall thickness on ultrasonographic cystourethrography: affecting factors and their implications. *J Ultrasound Med* 2003;**22**:777–82.

54. Kuhn A, Genoud S, Robinson D, Herrmann G, Gunthert A, Brandner S, *et al.* Sonographic transvaginal bladder wall thickness: does the measurement discriminate between urodynamic diagnoses? *Neurourol Urodyn* 2011;**30**:325–8. http://dx.doi.org/10.1002/nau.20997

55. Ibrahim S, Najdy M. *Measuring Bladder Wall Thickness in Cases of Detrusor Instability Using Transvaginal Ultrasound*. Proceedings of the International Continence Society (ICS) 2011 Annual Meeting: Glasgow; 2011.

56. Kuo HC. Measurement of detrusor wall thickness in women with overactive bladder by transvaginal and transabdominal sonography. *Int Urogynecol J* 2009;**20**:1293–9. http://dx.doi.org/10.1007/s00192-009-0946-2

57. Ozturk H, Aydur E, Irkilata HC, Seckin B, Dayanc M. *The Diagnostic Role of Transabdominal Ultrasonographic Bladder Wall Thickness in Female Urinary Incontinence*. Glasgow: Proceedings of the International Continence Society (ICS) 2011 Annual Meeting; 2011. http://dx.doi.org/10.1016/j.urology.2011.07.1250

58. Soligo M, Salvatore S, Khullar V, Luppino G, Arcari V, Milani R. *Overactive Bladder Definition and Ultrasound Measurement of Bladder Wall Thickness: The Right Way Without Urodynamics*. Heidelberg: Proceedings of the International Continence Society (ICS) 2002 Annual Meeting; 2002.

59. Yang JM, Huang WC. Discrimination of bladder disorders in female lower urinary tract symptoms on ultrasonographic cystourethrography. *J Ultrasound Med* 2002;**21**:1249–55.

60. Clement S, Candy B, Heath V, To M, Nicolaides KH. Transvaginal ultrasound in pregnancy: its acceptability to women and maternal psychological morbidity. *Ultrasound Obstet Gynecol* 2003;**22**:508–14. http://dx.doi.org/10.1002/uog.893

61. Nager CW, Brubaker L, Litman HJ, Zyczynski HM, Varner RE, Amundsen C, *et al.* A randomized trial of urodynamic testing before stress-incontinence surgery. *N Engl J Med* 2012;**366**:1987–97. http://dx.doi.org/10.1056/NEJMoa1113595

62. Van Leijsen S, Kluivers K, Mol B, Broekhuis S, Milani A, Bongers M, *et al.* Can preoperative urodynamic investigation be omitted in women with stress urinary incontinence? A non-inferiority randomized controlled trial. *Neurourol Urodyn* 2012;**31**:1118–23. http://dx.doi.org/10.1002/nau.22230

63. van Leijsen SAL, Kluivers KB, Mol BWJ, Milani AL, Roovers J-PW, den Boon J, *et al.* Value of urodynamics before stress urinary incontinence surgery: a randomized controlled trial. *Obstet Gynecol* 2013;**121**:999–1008. http://dx.doi.org/10.1097/AOG.0b013e31828c68e3

64. Weber A, Taylor R, Wei J, Lemack G, Piedmonte M, Walters M. The cost-effectiveness of preoperative testing (basic office assessment vs urodynamics) for stress urinary incontinence in women. *BJU Int* 2002;**89**:356–63. http://dx.doi.org/10.1046/j.1464-4096.2001.01687.x

65. Geisler JP, Drenchko R. Cost-effectiveness of urodynamics testing in women with predominant stress incontinence symptoms. *Obstet Gynecol* 2014;**123**:197S. http://dx.doi.org/10.1097/01.AOG.0000447227.13819.5d

66. Bosch J, Cardozo L, Hashim H, Hilton P, Oelke M, Robinson D. Constructing trials to show whether urodynamic studies are necessary in lower urinary tract dysfunction. *Neurourol Urodyn* 2011;**30**:735–40. http://dx.doi.org/10.1002/nau.21130

67. Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourol Urodyn* 2004;**23**:322–30. http://dx.doi.org/10.1002/nau.20041

68. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;**174**:469–76. http://dx.doi.org/10.1503/cmaj.050090

69. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;**326**:41–4. http://dx.doi.org/10.1136/bmj.326.7379.41

70. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;**140**:189–202. http://dx.doi.org/10.7326/0003-4819-140-3-200402030-00010

71. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;**11**:88–94. http://dx.doi.org/10.1177/0272989X9101100203

72. Panayi DC, Digesu GA, Tekkis P, Fernando R, Khullar V. Ultrasound measurement of vaginal wall thickness: a novel and reliable technique. *Int Urogynecol J* 2010;**21**:1265–70. http://dx.doi.org/10.1007/s00192-010-1183-4

73. Lang EV, Benotsch EG, Fick LJ, Lutgendorf S, Berbaum ML, Berbaum KS, *et al.* Adjunctive non-pharmacological analgesia for invasive medical procedures: a randomised trial. *Lancet* 2000;**355**:1486–90. http://dx.doi.org/10.1016/S0140-6736(00)02162-0

74. Lang EV, Hatsiopoulou O, Koch T, Berbaum K, Lutgendorf S, Kettenmann E, *et al.* Can words hurt? Patient–provider interactions during invasive procedures. *Pain* 2005;**114**:303–9. http://dx.doi.org/10.1016/j.pain.2004.12.028

75. Barlow DH. *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic.* 2nd edn. New York, NY: Guilford Press; 2001.

76. Abrams P, Freeman R, Anderstrom C, Mattiasson A. Tolterodine, a new antimuscarinic agent: as effective but better tolerated than oxybutynin in patients with an overactive bladder. *Br J Urol* 1998;**81**:801–10. http://dx.doi.org/10.1046/j.1464-410x.1998.00717.x

77. Ouslander JG, Kane RL, Abrass IB. Urinary incontinence in elderly nursing home patients. *JAMA* 1982;**248**:1194–8. http://dx.doi.org/10.1001/jama.1982.03330100032026

78. Ouslander JG, Sier HC. Drug therapy for geriatric urinary incontinence. *Clin Geriatr Med* 1986;**2**:789–807.

79. Madhuvrata P, Cody JD, Ellis G, Herbison GP, Hay-Smith EJC. Which anticholinergic drug for overactive bladder symptoms in adults. *Cochrane Database Syst Rev* 2012;**1**:CD005429. http://dx.doi.org/10.1002/14651858.cd005429.pub2

80. Tincello DG, Rashid T, Revicky V. Emerging treatments for overactive bladder: clinical potential of botulinum toxins. *Res Rep Urol* 2014;**6**:51–7. http://dx.doi.org/10.2147/rru.s44665

81. Kennelly M, Dmochowski R, Ethans K, Karsenty G, Schulte-Baukloh H, Jenkins B, *et al.* Long-term efficacy and safety of onabotulinumtoxinA in patients with urinary incontinence due to neurogenic detrusor overactivity: an interim analysis. *Urology* 2013;**81**:491–7. http://dx.doi.org/10.1016/j.urology.2012.11.010

82. Burton C, Sajja A, Latthe P. Effectiveness of percutaneous posterior tibial nerve stimulation for overactive bladder: a systematic review and meta-analysis. *Neurourol Urodyn* 2012;**31**:1206–16. http://dx.doi.org/10.1002/nau.22251

83. Smith A, Bevan D, Douglas HR, James D. Management of urinary incontinence in women: summary of updated NICE guidance. *BMJ* 2013;**347**:f5170. http://dx.doi.org/10.1136/bmj.f5170

84. Malone-Lee JG, Al-Buheissi S. Does urodynamic verification of overactive bladder determine treatment success? Results from a randomized placebo-controlled study. *BJU Int* 2009;**103**:931–7. http://dx.doi.org/10.1111/j.1464-410X.2009.08361.x

85. Groenendijk PM, Lycklama a Nyeholt AA, Heesakkers JP, van Kerrebroeck PE, Hassouna MM, Gajewski JB, *et al.* Urodynamic evaluation of sacral neuromodulation for urge urinary incontinence. *BJU Int* 2008;**101**:325–9. http://dx.doi.org/10.1111/j.1464-410X.2007.07258.x

86. Rovner ES, Goudelocke CM. Urodynamics in the evaluation of overactive bladder. *Curr Urol Rep* 2010;**11**:343–7. http://dx.doi.org/10.1007/s11934-010-0130-8

87. Vandoninck V, van Balken MR, Finazzi Agro E, Petta F, Micali F, Heesakkers JP, *et al.* Percutaneous tibial nerve stimulation in the treatment of overactive bladder: urodynamic data. *Neurourol Urodyn* 2003;**22**:227–32. http://dx.doi.org/10.1002/nau.10111

88. Rachaneni S, Latthe P. Does preoperative urodynamics improve outcomes for women undergoing surgery for stress urinary incontinence? A systematic review and meta-analysis. *BJOG* 2015;**122**:8–16. http://dx.doi.org/10.1111/1471-0528.12954

89. Brazzelli M, Murray A, Fraser C. Efficacy and safety of sacral nerve stimulation for urinary urge incontinence: a systematic review. *J Urol* 2006;**175**:835–41. http://dx.doi.org/10.1016/S0022-5347(05)00326-5

90. Weber AM, Walters MD. Cost-effectiveness of urodynamic testing before surgery for women with pelvic organ prolapse and stress urinary incontinence. *Am J Obstet Gynecol* 2000;**183**:1338–47. http://dx.doi.org/10.1067/mob.2000.111251

91. Tincello DG, Kenyon S, Abrams KR, Mayne C, Toozs-Hobson P, Taylor D, *et al.* Botulinum toxin a versus placebo for refractory detrusor overactivity in women: a randomised blinded placebo-controlled trial of 240 women (the RELAX study). *Eur Urol* 2012;**62**:507–14. http://dx.doi.org/10.1016/j.eururo.2011.12.056

92. Curtis L. *Unit Costs of Health and Social Care 2013*. University of Kent, Canterbury: Personal Social Services Research Unit; 2013.

93. Bump RC, Mattiasson A, Bo K, Brubaker LP, DeLancey JO, Klarskov P, *et al.* The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. *Am J Obstet Gynecol* 1996;**175**:10–7. http://dx.doi.org/10.1016/S0002-9378(96)70243-0

94. Drummond M, Sculpher M. Common methodological flaws in economic evaluations. *Med Care* 2005;**43**(Suppl. 7):5–14. http://dx.doi.org/10.1097/01.mlr.0000170001.10393.b7

95. Rachaneni S, Balogun M, Latthe P. Bladder-wall-thickness ultrasound scan in the investigation of LUTS in women: challenges and limitations. *Int Urogynecol J* 2013;**24**:725–8. http://dx.doi.org/10.1007/s00192-013-2048-4

96. Schafer W, Abrams P, Liao LM, Mattiasson A, Pesce F, Spangberg A, *et al.* Good urodynamic practices: Uroflowmetry, filling cystometry, and pressure-flow studies. *Neurourol Urodyn* 2002;**21**:261–74. http://dx.doi.org/10.1002/nau.10066

97. Khullar V, Salvatore S, Cardozo L, Bourne TH, Abbott D, Kelleher C. A novel technique for measuring bladder wall thickness in women using transvaginal ultrasound. *Ultrasound Obstet Gynecol* 1994;**4**:220–3. http://dx.doi.org/10.1046/j.1469-0705.1994.04030220.x

98. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;**329**:168–9. http://dx.doi.org/10.1136/bmj.329.7458.168

99. Xiao-Hua Z, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. 2nd edn. Hoboken, NJ: Wiley; 2011.

100. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford: Oxford University Press; 2004.

101. Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall; 1997. http://dx.doi.org/10.1201/9781439821862

102. Hashim H, Abrams P. Pharmacological management of women with mixed urinary incontinence. *Drugs* 2006;**66**:591–606. http://dx.doi.org/10.2165/00003495-200666050-00002

103. 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. http://dx.doi.org/10.1002/14651858.cd008224.pub2

104. Nager CW, Albo ME, FitzGerald MP, McDermott SM, Kraus S, Richter HE, *et al.* Process for development of multicenter urodynamic studies. *Urology* 2007;**69**:63–7. http://dx.doi.org/10.1016/j.urology.2006.08.1118

105. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6. http://dx.doi.org/10.1001/jama.282.11.1061

106. Broekhuis SR, Kluivers KB, Hendriks JCM, Massolt ET, Groen J, Vierhout ME. Reproducibility of same session repeated cystometry and pressure-flow studies in women with symptoms of urinary incontinence. *Neurourol Urodyn* 2010;**29**:428–31. http://dx.doi.org/10.1002/nau.20783

107. Homma Y, Kondo Y, Takahashi S, Kitamura T, Kawabe K. Reproducibility of cystometry in overactive detrusor. *Eur Urol* 2000;**38**:681–5. http://dx.doi.org/10.1159/000020362

108. Mortensen S, Lose G, Thyssen H. Repeatability of cystometry and pressure-flow parameters in female patients. *Int Urogynecol J Pelvic Floor Dysfunct* 2002;**13**:72–5. http://dx.doi.org/10.1007/s001920200018

109. Sorensen SS, Nielsen JB, Norgaard JP, Knudsen LM, Djurhuus JC. Changes in bladder volumes with repetition of water cystometry. *Urol Res* 1984;**12**:205–8. http://dx.doi.org/10.1007/BF00256804

110. Brostrom S, Jennum P, Lose G. Short-term reproducibility of cystometry and pressure-flow micturition studies in healthy women. *Neurourol Urodyn* 2002;**21**:457–60. http://dx.doi.org/10.1002/nau.10019

111. Bellucci CH, Wollner J, Gregorini F, Birnbock D, Kozomara M, Mehnert U, *et al.* Neurogenic lower urinary tract dysfunction – do we need same session repeat urodynamic investigations? *J Urol* 2012;**187**:1318–23. http://dx.doi.org/10.1016/j.juro.2011.11.107

112. Gupta A, Defreitas G, Lemack GE. The reproducibility of urodynamic findings in healthy female volunteers: results of repeated studies in the same setting and after short-term follow-up. *Neurourol Urodyn* 2004;**23**:311–16. http://dx.doi.org/10.1002/nau.20039

113. Serati M, Salvatore S, Cattoni E, Soligo M, Cromi A, Ghezzi F. Ultrasound measurement of bladder wall thickness in different forms of detrusor overactivity. *Int Urogynecol J* 2010;**21**:1405–11. http://dx.doi.org/10.1007/s00192-010-1194-1

114. Chan L, The S, Titus J, Tse V. P14.02: The value of bladder wall thickness measurement in the assessment of overactive bladder syndrome. *Ultrasound Obstet Gynecol* 2005;**26**:460. http://dx.doi.org/10.1002/uog.2546

115. Parsons M, Vella M, Cardozo L. *Does Bladder Wall Thickness Discriminate Detrusor Overactivity From Overactive Bladder and Normality*. Montreal: Proceedings of the International Continence Society (ICS) 2002 Annual Meeting; 2005.

116. Blatt AH, Titus J, Chan L. Ultrasound measurement of bladder wall thickness in the assessment of voiding dysfunction. *J Urol* 2008;**179**:2275–8. http://dx.doi.org/10.1016/j.juro.2008.01.118

117. Abou-Gamrah A, Fawzy M, Sammour H, Tadros S. Ultrasound assessment of bladder wall thickness as a screening test for detrusor instability. *Arch Gynecol Obstet* 2014;**289**:1023–8. http://dx.doi.org/10.1007/s00404-013-3092-5

118. Otsuki EN, Araujo Junior E, Oliveira E, Sartori MG, Girao MJ, Jarmy-Di Bella ZI. Ultrasound thickness of bladder wall in continent and incontinent women and its correlation with cystometry. *Sci World J* 2014;**2014**:684671. http://dx.doi.org/10.1155/2014/684671

119. Silva JA, Gonsalves Mde C, de Melo RT, Carrerette FB, Damiao R. Association between the bladder wall thickness and urodynamic findings in patients with spinal cord injury. *World J Urol* 2015;**33**:131–5. http://dx.doi.org/10.1007/s00345-014-1265-x

120. van der Schouw YT, Verbeek AL, Ruijs SH. Guidelines for the assessment of new diagnostic tests. *Invest Radiol* 1995;**30**:334–40. http://dx.doi.org/10.1097/00004424-199506000-00002

121. Ulmsten U, Falconer C. Connective tissue in female urinary incontinence. *Curr Opin Obstet Gynecol* 1999;**11**:509–15. http://dx.doi.org/10.1097/00001703-199910000-00017

122. Salvatore S, Serati M, Digesu GA, Triacca P, Uccella S, Khullar V, *et al.* Efficacy of tolterodine in relation to different urodynamic findings of detrusor overactivity. *Int Urogynecol J Pelvic Floor Dysfunct* 2008;**19**:701–4. http://dx.doi.org/10.1007/s00192-007-0512-8

123. Rovner ES. Complications of anterior compartment vaginal surgery. *Curr Urol Rep* 2007;**8**:405–12. http://dx.doi.org/10.1007/s11934-007-0039-z

124. Vandoninck V, Van Balken MR, Finazzi Agro E, Petta F, Caltagirone C, Heesakkers JP, *et al.* Posterior tibial nerve stimulation in the treatment of urge incontinence. *Neurourol Urodyn* 2003;**22**:17–23. http://dx.doi.org/10.1002/nau.10036

125. South MM, Romero AA, Jamison MG, Webster GD, Amundsen CL. Detrusor overactivity does not predict outcome of sacral neuromodulation test stimulation. *Int Urogynecol J Pelvic Floor Dysfunct* 2007;**18**:1395–8. http://dx.doi.org/10.1007/s00192-007-0351-7

126. Kraus SR, Dmochowski R, Albo ME, Xu L, Klise SR, Roehrborn CG. Urodynamic standardization in a large-scale, multicenter clinical trial examining the effects of daily tadalafil in men with lower urinary tract symptoms with or without benign prostatic obstruction. *Neurourol Urodyn* 2010;**29**:741–7. http://dx.doi.org/10.1002/nau.20825

127. Al-Hayek S, Belal M, Abrams P. Does the patient's position influence the detection of detrusor overactivity? *Neurourol Urodyn* 2008;**27**:279–86. http://dx.doi.org/10.1002/nau.20503

128. Arunkalaivanan AS, Mahomoud S, Howell M. Does posture affect cystometric parameters and diagnoses? *Int Urogynecol J Pelvic Floor Dysfunct* 2004;**15**:422–4; discussion 4. http://dx.doi.org/10.1007/s00192-004-1195-z

129. Tubaro A, Khullar V, Oelke M, Wijkstra H, Tretter R, Stow B, *et al.* Intra- and inter-reader variability of transvaginal ultrasound bladder wall thickness measurements: results from the shrink study. *Neurourol Urodyn* 2013;**32**:711–12.

130. Amer SAKS, Li TC, Bygrave C, Sprigg A, Saravelos H, Cooke ID. An evaluation of the inter-observer and intra-observer variability of the ultrasound diagnosis of polycystic ovaries. *Hum Reprod* 2002;**17**:1616–22. http://dx.doi.org/10.1093/humrep/17.6.1616

131. Price DD, Mcgrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain* 1983;**17**:45–56. http://dx.doi.org/10.1016/0304-3959(83)90126-4

132. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *Br J Clin Psychol* 1992;**31**:301–6. http://dx.doi.org/10.1111/j.2044-8260.1992.tb00997.x

133. Agresti A. Score and pseudo-score confidence intervals for categorical data analysis. *Stat Biopharm Res* 2011;**3**:163–72. http://dx.doi.org/10.1198/sbr.2010.09053

134. Gorton E, Stanton S. Women's attitudes to urodynamics: a questionnaire survey. *Br J Obstet Gynaecol* 1999;**106**:851–6. http://dx.doi.org/10.1111/j.1471-0528.1999.tb08408.x

135. Arntz A, Vaneck M, Heijmans M. Predictions of dental pain – the fear of any expected evil, is worse than the evil itself. *Behav Res Ther* 1990;**28**:29–41. http://dx.doi.org/10.1016/0005-7967(90)90052-K

136. Terry R, Brodie EE, Niven CA. Exploring the phenomenology of memory for pain: Is previously experienced acute pain consciously remembered or simply known? *J Pain* 2007;**8**:467–75. http://dx.doi.org/10.1016/j.jpain.2006.12.006

137. Bradley P, Kennedy Y. Confronting pain face-to-face at NSU. *Todays FDA* 2008;**20**:37–40.

138. McNeil DW, Berryman ML. Components of dental fear in adults? *Behav Res Ther* 1989;**27**:233–6. http://dx.doi.org/10.1016/0005-7967(89)90041-7

139. Denberg TD, Melhado TV, Coombes JM, Beaty BL, Berman K, Byers TE, *et al.* Predictors of nonadherence to screening colonoscopy. *J Gen Intern Med* 2005;**20**:989–95. http://dx.doi.org/10.1111/j.1525-1497.2005.00164.x

140. Yiou R, Audureau E, Loche CM, Dussaud M, Lingombet O, Binhas M. Comprehensive evaluation of embarrassment and pain associated with invasive urodynamics. *Neurourol Urodyn* 2015;**34**:156–60. http://dx.doi.org/10.1002/nau.22521

141. Neustaedter EG, Milne J, Shorten K, Weckman B, Tse A, Tange S. How well informed are women who undergo urodynamic testing? *Neurourol Urodynam* 2011;**30**:572–7. http://dx.doi.org/10.1002/nau.21014

142. Spielberger CDC. *Manual for the State-Trait Anxiety Inventory (Form Y).* Menlo Park, CA: Mind Garden Inc.; 1983.

143. Perry S, McGrother CW, Turner K. An investigation of the relationship between anxiety and depression and urge incontinence in women: development of a psychological model. *Br J Health Psychol* 2006;**11**:463–82. http://dx.doi.org/10.1348/135910705X60742

144. Tinetti ME, Inouye SK, Gill TM, Doucette JT. Shared risk-factors for falls, incontinence, and functional dependence – unifying the approach to geriatric syndromes. *JAMA* 1995;**273**:1348–53. http://dx.doi.org/10.1001/jama.1995.03520410042024

145. Heidrich SM, Wells TJ. Effects of urinary incontinence: psychological well-being and distress in older community-dwelling women. *J Gerontol Nurs* 2004;**30**:47–54. http://dx.doi.org/10.3928/0098-9134-20040501-10

146. Saleem A. In women with urinary incontinence how necessary is cystometry? *J Pak Med Assoc* 2010;**60**:356–9.

147. Lemeshow H. *Applied Logistic Regression*. 2nd edn. New York, NY: John Wiley and Sons, Inc. 2000.

148. Verbeke G MG. *Linear Mixed Models For Longitudinal Data*. New York, NY: Springer; 2000.

149. Jackson S. The patient with an overactive bladder–symptoms and quality-of-life issues. *Urology* 1997;**50**(Suppl. 6A):18–22; discussion 3–4. http://dx.doi.org/10.1016/S0090-4295(97)00580-3

150. Martin J, Williams K, Abrams KR, Turner DA, Sutton AJ, Chapple C, *et al.* Systematic review and evaluation of methods of assessing urinary incontinence. *Neurourol Urodyn* 2006;**25**:674–83. http://dx.doi.org/10.1002/nau.20340

151. Clement KD, Lapitan MC, Omar MI, Glazener CM. Urodynamic studies for management of urinary incontinence in children and adults: A short version Cochrane systematic review and meta-analysis. *Neurourol Urodyn* 2015;**34**:407–12. http://dx.doi.org/10.1002/nau.22584

152. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;**41**:582–92. http://dx.doi.org/10.1097/01.MLR.0000062554.74615.4C

153. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. *BMC Med Res Methodol* 2007;**7**:30. http://dx.doi.org/10.1186/1471-2288-7-30

154. Majumdar A, Latthe P, Toozs-Hobson P. Urodynamics prior to treatment as an intervention: a pilot study. *Neurourol Urodyn* 2010;**29**:522–6. http://dx.doi.org/10.1002/nau.20810

155. Nager CW, Brubaker L, Daneshgari F, Litman HJ, Dandreo KJ, Sirls L, *et al.* Design of the Value of Urodynamic Evaluation (ValUE) trial: a non-inferiority randomized trial of preoperative urodynamic investigations. *Contemp Clin Trials* 2009;**30**:531–9. http://dx.doi.org/10.1016/j.cct.2009.07.001

156. Gotoh M, Homma Y, Yokoyama O, Nishizawa O. Responsiveness and minimal clinically important change in overactive bladder symptom score. *Urology* 2011;**78**:768–73. http://dx.doi.org/10.1016/j.urology.2011.06.020

157. Department of Health. *The Handbook to the NHS Constitution (2012 edition)*. London: Department of Health; 2012.

158. Dowson C, Watkins J, Khan MS, Dasgupta P, Sahai A. Repeated botulinum toxin type A injections for refractory overactive bladder: medium-term outcomes, safety profile, and discontinuation rates. *Eur Urol* 2012;**61**:834–9. http://dx.doi.org/10.1016/j.eururo.2011.12.011

159. Jorgensen L, Lose G, Mortensen SO, Pedersen LM, Kristensen JK. The Burch colposuspension for urinary incontinence in patients with stable and unstable detrusor function. *Neurourol Urodyn* 1988;**7**:435–41. http://dx.doi.org/10.1002/nau.1930070504

160. Latthe PM, Foon R, Toozs-Hobson P. Transobturator and retropubic tape procedures in stress urinary incontinence: a systematic review and meta-analysis of effectiveness and complications. *BJOG* 2007;**114**:522–31. http://dx.doi.org/10.1111/j.1471-0528.2007.01268.x

161. Jain P, Jirschele K, Botros SM, Latthe PM. Effectiveness of midurethral slings in mixed urinary incontinence: a systematic review and meta-analysis. *Int Urogynecol J* 2011;**22**:923–32. http://dx.doi.org/10.1007/s00192-011-1406-3

162. Dean N, Herbison P, Ellis G, Wilson D. Laparoscopic colposuspension and tension-free vaginal tape: a systematic review. *BJOG* 2006;**113**:1345–53. http://dx.doi.org/10.1111/j.1471-0528.2006.01101.x

163. Kulseng-Hanssen S, Berild GH. Subjective and objective incontinence 5 to 10 years after Burch colposuspension. *Neurourol Urodyn* 2002;**21**:100–5. http://dx.doi.org/10.1002/nau.10011

164. Novara G, Artibani W, Barber MD, Chapple CR, Costantini E, Ficarra V, *et al.* Updated systematic review and meta-analysis of the comparative data on colposuspensions, pubovaginal slings, and midurethral tapes in the surgical treatment of female stress urinary incontinence. *Eur Urol* 2010;**58**:218–38. http://dx.doi.org/10.1016/j.eururo.2010.04.022

165. Leal J, Wordsworth S, Legood R, Blair E. Eliciting expert opinion for economic models: an applied example. *Value Health* 2007;**10**:195–203. http://dx.doi.org/10.1111/j.1524-4733.2007.00169.x

166. Siddiqui NY, Wu JM, Amundsen CL. Efficacy and adverse events of sacral nerve stimulation for overactive bladder: A systematic review. *Neurourol Urodyn* 2010;**29**(Suppl. 1):18–23. http://dx.doi.org/10.1002/nau.20786

167. Tufts Medical Center. *Cost-Effectiveness Analysis Registry*. URL https://research.tufts-nemc.org/cear4/ (accessed June 2014).

168. Chen HW, Bercik RS, Werner EF, Thung SF. Cost-effectiveness of percutaneous tibial nerve stimulation versus extended release tolterodine for overactive bladder. *J Urol* 2012;**187**:178–84. http://dx.doi.org/10.1016/j.juro.2011.09.052

169. Wu JM, Fulton RG, Amundsen CL, Knight SK, Kuppermann M. Patient preferences for different severities of and treatments for overactive bladder. *Female Pelvic Med Reconstr Surg* 2011;**17**:184–9. http://dx.doi.org/10.1097/SPV.0b013e318223c8ad

170. Arlandis-Guzman S, Errando-Smet C, Trocio J, Arumi D, Rejas J. Cost-effectiveness analysis of antimuscarinics in the treatment of patients with overactive bladder in Spain: A decision-tree model. *BMC Urol* 2011;**11**:9. http://dx.doi.org/10.1186/1471-2490-11-9

171. Cardozo L, Thorpe A, Warner J, Sidhu M. The cost-effectiveness of solifenacin vs fesoterodine, oxybutynin immediate-release, propiverine, tolterodine extended-release and tolterodine immediate-release in the treatment of patients with overactive bladder in the UK National Health Service. *BJU Int* 2010;**106**:506–14. http://dx.doi.org/10.1111/j.1464-410X.2009.09160.x

172. Hakkaart L, Verboom P, Phillips R, Al MJ. The cost utility of solifenacin in the treatment of overactive bladder. *Int Urol Nephrol* 2009;**41**:293–8. http://dx.doi.org/10.1007/s11255-008-9448-2

173. Herschorn S, Vicente C, Piwko C. Canadian cost-effectiveness analysis of solifenacin compared to oxybutynin immediate-release in patients with overactive bladder. *J Med Econ* 2010;**13**:508–15. http://dx.doi.org/10.3111/13696998.2010.509244

174. Kobelt G, Jönsson L, Mattiasson A. Cost-effectiveness of new treatments for overactive bladder: the example of tolterodine, a new muscarinic agent: A Markov model. *Neurourol Urodyn* 1998;**17**:599–611. http://dx.doi.org/10.1002/(SICI)1520-6777(1998)17:6<599::AID-NAU4>3.0.CO;2-J

175. Pradelli L, Iannazzo S. Solifenacin in the treatment of overactive bladder syndrome in Italian patients: pharmacoeconomic evaluation. *J Med Econ* 2009;**12**:25–35. http://dx.doi.org/10.3111/13696990902767800

176. Speakman M, Khullar V, Mundy A, Odeyemi I, Bolodeoku J. A cost–utility analysis of once daily solifenacin compared to tolterodine in the treatment of overactive bladder syndrome. *Curr Med Res Opin* 2008;**24**:2173–9. http://dx.doi.org/10.1185/03007990802234829

177. Dumville J, Manca A, Kitchener H, Smith A, Nelson L, Torgerson D. Cost-effectiveness analysis of open colposuspension versus laparoscopic colposuspension in the treatment of urodynamic stress incontinence. *BJOG* 2006;**113**:1014–22. http://dx.doi.org/10.1111/j.1471-0528.2006.01036.x

178. Jacklin P, Duckett J, Renganathan A. Analytic model comparing the cost utility of TVT versus duloxetine in women with urinary stress incontinence. *Int Urogynecol J* 2010;**21**:977–84. http://dx.doi.org/10.1007/s00192-010-1132-2

179. Kilonzo M, Vale L, Stearns SC, Grant A, Cody J, Glazener C, *et al.* Cost effectiveness of tension-free vaginal tape for the surgical management of female stress incontinence. *Int J Technol Assess Health Care* 2004;**20**:455–63. http://dx.doi.org/10.1017/S0266462304001357

180. Manca A, Sculpher MJ, Ward K, Hilton P. A cost–utility analysis of tension-free vaginal tape versus colposuspension for primary urodynamic stress incontinence. *BJOG* 2003;**110**:255–62.

181. Mihaylova B, Pitman R, Tincello D, Van Der Vaart H, Tunn R, Timlin L, *et al.* Cost-effectiveness of Duloxetine: The Stress Urinary Incontinence Treatment (SUIT) study. *Value Health* 2010;**13**:565–72. http://dx.doi.org/10.1111/j.1524-4733.2010.00729.x

182. Wu JM, Visco AG, Weidner AC, Myers ER. Is Burch colposuspension ever cost-effective compared with tension-free vaginal tape for stress incontinence? *Am J Obstet Gynecol* 2007;**197**:62.e1–5. http://dx.doi.org/10.1016/j.ajog.2007.02.039

183. Arlandis S, Castro D, Errando C, Fernández E, Jiménez M, González P, *et al.* Cost-effectiveness of sacral neuromodulation compared to botulinum neurotoxin a or continued medical management in refractory overactive bladder. *Value Health* 2011;**14**:219–28. http://dx.doi.org/10.1016/j.jval.2010.08.006

184. Leong RK, de Wachter SG, Joore MA, van Kerrebroeck PE. Cost-effectiveness analysis of sacral neuromodulation and botulinum toxin A treatment for patients with idiopathic overactive bladder. *BJU Int* 2011;**108**:558–64. http://dx.doi.org/10.1111/j.1464-410X.2010.09905.x

185. Siddiqui NY, Amundsen CL, Visco AG, Myers ER, Wu JM. Cost-effectiveness of sacral neuromodulation versus intravesical botulinum A toxin for treatment of refractory urge incontinence. *J Urol* 2009;**182**:2799–804. http://dx.doi.org/10.1016/j.juro.2009.08.031

186. Wu JM, Siddiqui NY, Amundsen CL, Myers ER, Havrilesky LJ, Visco AG. Cost-effectiveness of botulinum toxin a versus anticholinergic medications for idiopathic urge incontinence. *J Urol* 2009;**181**:2181–6. http://dx.doi.org/10.1016/j.juro.2009.01.037

187. Schultz SE, Kopec JA. Impact of chronic conditions. *Health Rep* 2003;**14**:41–53.

188. Subak LL, Brubaker L, Chai TC, Creasman JM, Diokno AC, Goode PS, *et al.* High costs of urinary incontinence among women electing surgery to treat stress incontinence. *Obstet Gynecol* 2008;**111**:899–907. http://dx.doi.org/10.1097/AOG.0b013e31816a1e12

189. NICE. *Guide to the Methods of Technology Appraisal*. London: NICE; 2013.

190. Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Health Technol Assess* 2009;**13**(29). http://dx.doi.org/10.3310/hta13290

191. Briggs AH, Claxton K, Sculpher MJ. *Decision Modelling For Health Economic Evaluation*. Oxford: Oxford University Press; 2006.

192. NICE. *Social Value Judgements: Principles for the Development of NICE Guidance*. 2nd edn. London: NICE; 2008.

193. Clark AL, Gregory T, Smith VJ, Edwards R. Epidemiologic evaluation of reoperation for surgically treated pelvic organ prolapse and urinary incontinence. *Am J Obstet Gynecol* 2003;**189**:1261–7. http://dx.doi.org/10.1067/S0002-9378(03)00829-9

194. Radley SC, Rosario DJ, Chapple CR, Farkas AG. Conventional and ambulatory urodynamic findings in women with symptoms suggestive of bladder overactivity. *J Urol* 2001;**166**:2253–8. http://dx.doi.org/10.1016/S0022-5347(05)65545-0

195. Dokmeci F, Seval M, Gok H. Comparison of ambulatory versus conventional urodynamics in females with urinary incontinence. *Neurourol Urodyn* 2010;**29**:518–21. http://dx.doi.org/10.1002/nau.20821

196. Gorton E, Stanton S. Ambulatory urodynamics: do they help clinical management? *BJOG* 2000;**107**:316–9. http://dx.doi.org/10.1111/j.1471-0528.2000.tb13224.x

197. Rovner E, Kennelly M, Schulte-Baukloh H, Zhou J, Haag-Molkenteller C, Dasgupta P. Urodynamic results and clinical outcomes with intradetrusor injections of onabotulinumtoxina in a randomized, placebo-controlled dose-finding study in idiopathic overactive bladder. *Neurourol Urodyn* 2011;**30**:556–62. http://dx.doi.org/10.1002/nau.21021

198. Bump RC, Norton PA, Zinner NR, Yalcin I, Group DUIS. Mixed urinary incontinence symptoms: urodynamic findings, incontinence severity, and treatment response. *Obstet Gynecol* 2003;**102**:76–83. http://dx.doi.org/10.1016/S0029-7844(03)00376-4

199. Gray A, Clarke PM, Wolstenholme JL, Wordsworth S. *Applied Methods of Cost-Effectiveness Analysis in Healthcare.* Oxford: Oxford University Press; 2010.

200. Kay S, Tolley K, Colayco D, Khalaf K, Anderson P, Globe D. Mapping EQ-5D Utility scores from the Incontinence Quality of Life Questionnaire among patients with neurogenic and idiopathic overactive bladder. *Value Health* 2013;**16**:394–402. http://dx.doi.org/10.1016/j.jval.2012.12.005

201. Coast J, Flynn TN, Natarajan L, Sproston K, Lewis J, Louviere JJ, *et al.* Valuing the ICECAP capability index for older people. *Soc Sci Med* 2008;**67**:874–82. http://dx.doi.org/10.1016/j.socscimed.2008.05.015

202. Office of National Statistics. *UK Population Estimates 2013*. URL: www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk–england-and-wales–scotland-and-northern-ireland/2013/sty-population-estimates.html (accessed 26 June 2014).

203. Milsom I, Abrams P, Cardozo L, Roberts RG, Thuroff J, Wein AJ. How widespread are the symptoms of an overactive bladder and how are they managed? A population-based prevalence study. *BJU Int* 2001;**87**:760–6. http://dx.doi.org/10.1046/j.1464-410x.2001.02228.x

204. Stewart WF, Van Rooyen JB, Cundiff GW, Abrams P, Herzog AR, Corey R, *et al.* Prevalence and burden of overactive bladder in the United States. *World J Urol* 2003;**20**:327–36.

205. Thuroff JW, Abrams P, Andersson KE, Artibani W, Chapple CR, Drake MJ, *et al.* [EAU guidelines on urinary incontinence.] *Actas Urol Esp* 2011;**35**:373–88. http://dx.doi.org/10.1016/j.acuroe.2011.03.002

206. Farrell SA, Epp A, Flood C, Lajoie F, MacMillan B, Mainprize T, *et al.* The evaluation of stress incontinence prior to primary surgery. *J Obstet Gynaecol Can* 2003;**25**:313–24.

207. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Chem Lab Med* 2003;**41**:68–73. http://dx.doi.org/10.1515/CCLM.2003.012

208. Ogunyemi TO, Siadat MR, Arslanturk S, Killinger KA, Diokno AC. Novel application of statistical methods to identify new urinary incontinence risk factors. *Adv Urol* 2012;**2012**:276501. http://dx.doi.org/10.1155/2012/276501

209. Frenkl TL, Railkar R, Palcza J, Scott BB, Alon A, Green S, *et al.* Variability of urodynamic parameters in patients with overactive bladder. *Neurourol Urodyn* 2011;**30**:1565–9. http://dx.doi.org/10.1002/nau.21081

210. Grant AM, Altman DG, Babiker AG, Campbell MK, Clemens F, Darbyshire JH, *et al.* A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet* 2005;**365**:711–22. http://dx.doi.org/10.1016/S0140-6736(05)70939-9

211. Medical Research Council. *MRC Guidelines for Good Clinical Practice for Clinical Trials*. London: Medical Research Council; 1998.

212. Daniels J, Gray J, Pattison H, Roberts T, Edwards E, Milner P. Rapid testing for group B streptococcus during labour: a test accuracy study with evaluation of acceptability and cost-effectiveness. *Health Technol Assess* 2009;**13**(42). http://dx.doi.org/10.3310/hta13420

213. Macskassy SA, Provost F, Rosset S, editors. *ROC Confidence Bands: An Empirical Evaluation*. Proceedings of the 22nd international conference on Machine Learning. Bonn: Association for Computing Machinery; 2005. http://dx.doi.org/10.1145/1102351.1102419

214. Pannek J, Bartel P, Gocking K, Frotzler A. Clinical usefulness of ultrasound assessment of detrusor wall thickness in patients with neurogenic lower urinary tract dysfunction due to spinal cord injury: urodynamics made easy? *World J Urol* 2013;**31**:659–64. http://dx.doi.org/10.1007/s00345-012-0970-6

Appendix 1 Independent monitoring of the Bladder Ultrasound Study incorporating a novel approach for developing stopping criteria using the interim assessment of test accuracy

Background

Independent oversight for the BUS was requested by the funding body (NIHR HTA programme); however, how this oversight should be arranged or what it should entail was unclear as the study was not considered to be distinct from that of more commonly funded study types, that is, RCTs.

Numerous recommendations are available regarding the establishment and operation of TSCs and DMCs[210,211] but there is a lack of guidance for studies of test accuracy. In a rare example, Daniels *et al.*[212] reported their experiences in the monitoring of a test accuracy study in another HTA funded study – Group B *Streptococcus* (GBS) Study. Here, they make some recommendations about how they think such studies should be monitored including monitoring by combining the functions of the TSC and DMC, provided care pathways were not being altered as a result of evaluating new tests. More specifically, the independent members of the TSC were to form a subgroup DMC in charge of reviewing recruitment, disease prevalence, study quality, safety and the need for additional analyses.

One interesting component of the recommendations in the GBS study was that interim assessment of test accuracy (e.g. sensitivity, specificity) should be made available to the DMC if possible during the study. However, there is no guidance on how these estimates should be used to contribute towards the decision-making by the DMC about whether the study should be closed prematurely, modified or continued without change. For example, for the recommendation of 'the study should be closed immediately and completely', the associated rationale is based on only potential harm to the patient through the test or through new convincing external evidence.

In this section we briefly outline the independent monitoring arrangements for the BUS, which were similar to the aforementioned Daniels *et al.*[212] study. We then describe the novel arrangements proposed for making decisions about whether or not to modify the study based on interim estimates of sensitivity and specificity.

Arrangements for oversight

There was a working assumption by the co-investigators that the arrangements for oversight would work in a similar fashion to the aforementioned GBS study[212] (*Figure 20*).

Given this assumption, five independent members were initially appointed to the TSC with no separate DMC group proposed. These members were Professor Doug Tincello (Professor of Urogynaecology, University of Leicester) chairperson; Dr Patrick Chien (Consultant Gynaecologist, University of Dundee); Dr Jonathan Cook (statistician and Associate Professor, University of Oxford); Dr Karen Ward (Consultant Urogynaecologist, Central Manchester University Hospitals); and Mrs Jean Perks (lay representative). Dr Ward had experience in the clinical area and Dr Cook had experience in diagnostic research methods. Non-independent members consisted of the chief investigator (Dr Pallavi Latthe) and other co-investigators (Professor Arri Coomarasary, Professor Jon Deeks, Dr Jane Daniels and Miss Victoria Brooks). Mr Lee Middleton was the study statistician.

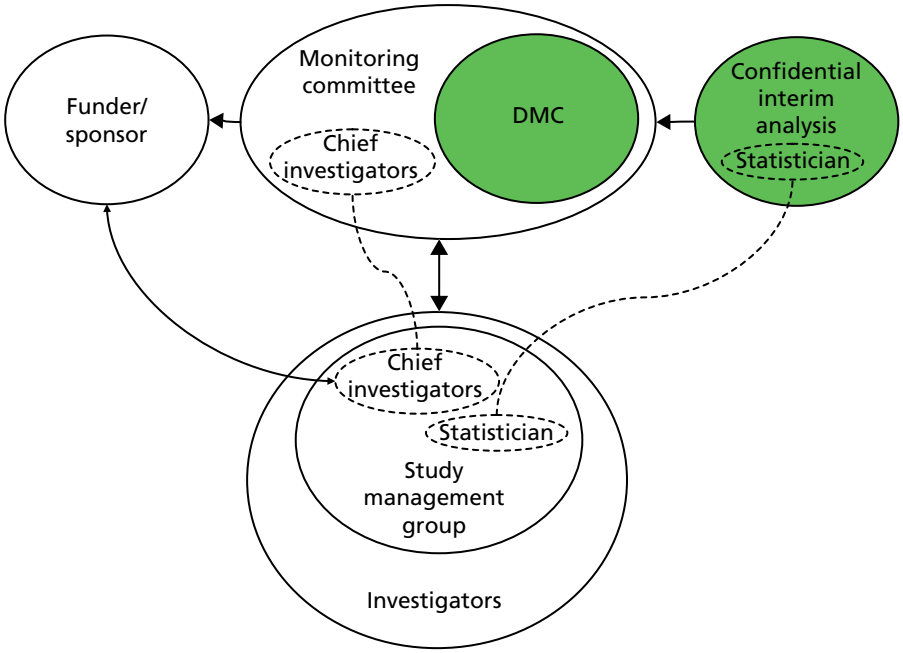

FIGURE 20 Flow chart of arrangements for oversight.

At an initial pre-recruitment meeting of all members, the group felt the combined oversight model similar to the GBS study was indeed appropriate as the proposed index test was low risk and the results of the index test were not being used in patient care. In addition, patients with suspected DO would receive the reference standard test as part of routine care and thus they were not being subjected to any additional testing apart from the additional index test.

One change from the GBS approach was proposed. Although the DMC committee was to be formed as a subgroup from the independent TSC members, it was decided that the chairperson of the TSC would not be part of this subgroup. The rational for this was to make sure there was a level of separation between DMC discussions and the larger TSC. This left three independent members on the subgroup DMC (Professor Patrick Chien, Dr Jonathan Cook and Dr Karen Ward).

Discussions took place about whether or not interim estimates of the index test would be used to make any decisions about the continuation or modification of the study. Given the likelihood of harm to the patient though testing was considered to be low, one option considered was to not use them at all and to base any decision-making on other factors such as recruitment, patient safety and new external evidence. Conversely, it was debated whether or not it would be appropriate to carry on the study if it was shown at interim analysis that BWT was overwhelmingly good or bad as a diagnostic test. Factors here included saving time, expense and the good will of the study participants. The independent members of the committee agreed on the latter argument and decided that the subgroup DMC members, study statisticians Jon Deeks and Lee Middleton, should meet to decide on a sensible way to proceed with this dilemma (see *Development of a stopping rule*) as there was no methodological work in this area that would be useful for guidance.

Future meetings of the TSC and subgroup DMC were to take place on the same day on a 6-monthly basis. The subgroup DMC and study statistician, Lee Middleton, were to meet prior to the full group to discuss the interim analysis and were to feedback any recommendations based on this analysis to the wider TSC group. Only the independent members of the DMC would be eligible to vote on any recommendations. These recommendations could consist of: no action needed, study continues as planned; the study should be closed immediately and completely; the study should be closed to a particular subgroup on the grounds of safety or accuracy; the study should be closed owing to futility with respect to study recruitment; the study should be modified.

Development of a stopping rule

The members of the DMC agreed that any potential stopping rule should be stringent and provide convincing evidence in test-positive and -negative cases. If there was convincing evidence, then the DMC may consider stopping the study prematurely. However, it was considered that other factors, such as safety data and external evidence, should be taken into account before making this decision.

It was suggested that the clinical members of the group should consider what levels of post-test probabilities would be considered conclusive, that is what value would be so high in test-positive cases that bladder ultrasonography diagnosis of DO would be considered conclusively convincing or, alternatively, so low in test-negative cases that a non-DO diagnosis was convincing. Post-test probabilities were suggested as they focus the attention on probability of disease given test result in an individual patient as opposed to thresholds of sensitivity and specificity, which are population-based measures.[98] The two clinical members suggested anything over 90% post-test probability would be convincing for test-positive cases, with less than 60% not convincing. For test-negative cases, anything under 10% would be convincing, with greater than 40% not convincing.

A further complication was that a pre-planned cut-off point for BWT that optimises sensitivity and specificity was not known; a cut-off point of 5 mm was suggested in the statistical analysis plan, with further pre-planned analysis using ROC curves to investigate where the optimal point might be. For this reason, it was considered that the focus of any interim analysis output should be the ROC curve rather than any pre-prescribed cut-off point. It was suggested that plots of ROC curves of interim results could be visually compared with the post-test probabilities suggested by the clinical members if these post-test probabilities were converted to LRs and plotted on the same graph. If CIs for the ROC curves were also included then these could be used as thresholds of convincing evidence; 99% CIs were suggested as a conservative measure.

An example of this suggested method was drawn up as per *Figure 21*. The LRs are calculated using the formula: LR = post-test probability/[pre-test odds × (1 – post-test probability)].[98] The pre-test odds were taken from the current literature for DO which suggested 50% in this population, that is the pre-test odds in the case was equal to one [0.5/(1–0.5)]. If the lower CI of the ROC curve were to enter the shaded green area (i.e. the test is overwhelmingly convincing at ruling in/out disease) or similarly the upper CI of the ROC curve were to enter the green shaded area (i.e. the test is overwhelmingly unconvincing at ruling in/out disease) then the members could consider stopping the study prematurely.

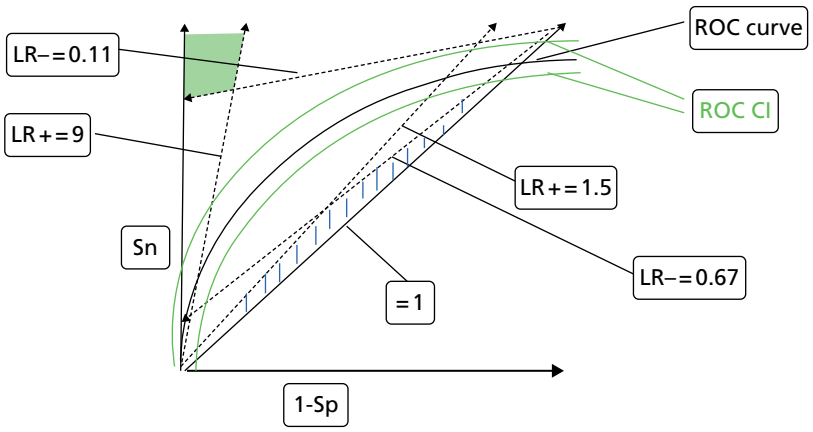

FIGURE 21 Plots of ROC curves.

Meetings

First meeting to assess interim data, 31 July 2012: 369 participants recruited, 300 complete sets of data available for analysis.

The members reviewed the following items, which did not raise any concerns: recruitment, participant demographics and withdrawal from the study, data completeness and study organisational issues. No serious adverse events had been reported in relation to the tests.

The current estimate of prevalence of DO was estimated to be 62% (95% CI 57% to 68%), higher than the 50% estimate used in the sample size calculation at the outset. This caused an unexpected complication as it was also higher than the lower threshold for post-test probability that the DMC suggested would be unconvincing in test-positive cases (60%). It was also apparent (*Figure 22*) that BWT was performing much worse than expected as a test and that it was difficult to determine where an optimal cut-off point would be, given the poor overall performance. The DMC decided to remain with the LRs agreed originally for poor performance at ruling in or out DO (1.50 and 0.67, respectively). They also agreed that 99% CIs for the LRs at the 5 mm cut-off point – which had been suggested a priori as an optimal point – would be used to compare against these thresholds to determine if the study would stop or continue. The positive LR at the 5 mm cut-off point was 1.01 (99% CI 0.65 to 1.57) and the negative likelihood was 1.00 (99% CI 0.81 to 1.23). The latter did not contain the original threshold (0.67) but the former did (1.5). It was agreed with all the independent members that recruitment should carry on but further analysis should be completed once complete data were available on 450 women to give the data further chance to mature but to be able to retain the option to stop the study early.

Second meeting to assess interim data, 12 December 2012: 546 participants recruited, 456 complete sets of data available for analysis.

The members again reviewed recruitment, demographics, withdrawals, data completeness, safety and study organisational issues. No concerns were raised. The positive LR at the 5 mm cut-off point was 1.09 (99% CI 0.76 to 1.56) and the negative LR was 0.96 (99% CI 0.81 to 1.14). As the positive LR contained 1.5, it was decided that the study should continue to recruit until its scheduled end date; end of March 2013. It was decided that recruitment should definitely not extend beyond this time period even if the estimated disease prevalence was higher than originally estimated (and, therefore, different from the original sample size assumptions). There would be no more interim analyses.

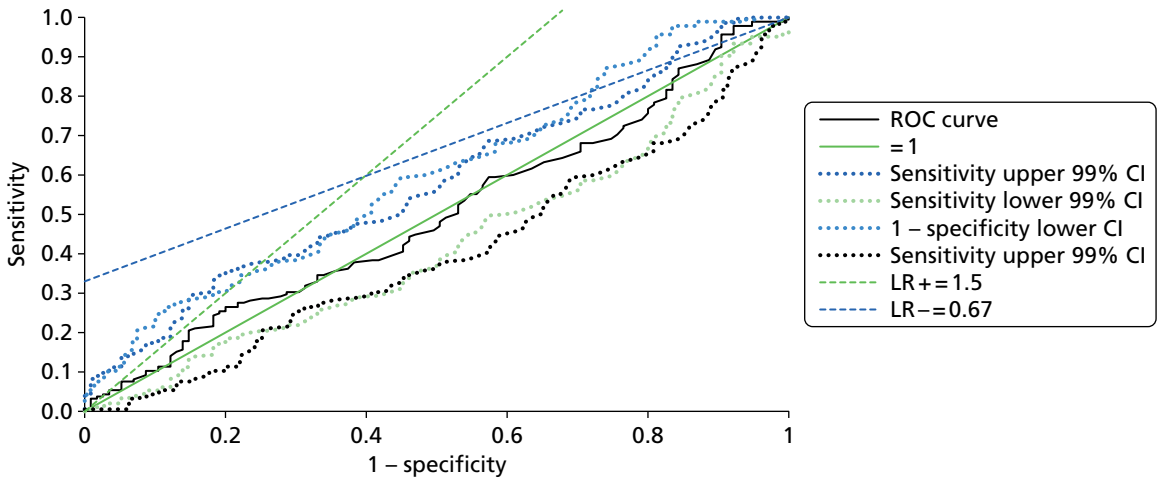

FIGURE 22 The ROC curve for BWT at DMC meeting.

Summary

In this section we described the arrangements for oversight in the BUS. This oversight had been requested by the funding body but with limited guidance that did not take into account the distinct nature of a diagnostic study as opposed to a RCT. A TSC was formed with a subgroup, which agreed to perform the role of a DMC.

Given that the index and reference standard test results were available within a month of patient recruitment, it was decided that interim analyses could be used to drive a decision about whether or not to stop the study prematurely. This was based on the rationale that if the main study question had been answered early then it would be unethical to continue on the grounds of saving time, money and to not waste the good will of the participants. Results of the index test were not available to clinicians and were not changing care pathways and hence this was not a consideration.

The members derived a stopping rule by considering levels of post-test probability that would be considered conclusive in test-positive and -negative cases. These were then converted to positive and negative LRs using an estimate of the pre-test odds. It was these LRs that were used as thresholds to compare against interim estimates and CIs. As an optimum cut-off point for the index test was not known, it was suggested that plots of LRs could be compared visually against ROC curves. This later proved difficult and a simpler approach using the estimates of LR from the best-known estimate of cut-off point was used.

Similar to any RCT there exists some uncertainty in the initial sample size calculation in any diagnostic study. This means that there is potential that a test could be conclusively proven to be useful or useless prior to the full target sample being recruited. The approach adopted here, using ROC curves or simple cut-off point estimate of LR, is straightforward and could be adapted by any large diagnostic study for which results are available relatively quickly during the recruitment period.

While considering post-test probabilities to inform stopping thresholds has the advantage of being a more interpretable measure than sensitivity and specificity, one needs a decent estimate of the pre-test odds to convert these to a LR. This may not be available at all or it may be inaccurate as in this study. If there is likely to be large uncertainty around the estimate of pre-test odds, it may be advisable to skip this step and decide the stopping thresholds based on LRs or sensitivity/specificity estimates. For example, you could set a rule to stop the study prematurely if you were 99% confident of the LRs being greater than 10 (i.e. the lower bound of the CI greater than 10). Another option would be to base stopping rules on CIs for an AUC.

A further complication with incorporating CIs for ROC curves is that they are not very straightforward to produce owing to their two-dimensional nature;[213] SAS has no function or macro available to produce them and the only bespoke macro available in Stata (version 13; StataCorp LP, College Station, TX, USA) is out of date and, currently, not being supported. For the interim analysis in BUS we resorted to overlaying two separate plots – one with 99% CIs for sensitivity and a similar one for specificity.

In hindsight, the DMC might reflect that they were being too conservative suggesting the use of 99% CIs. Indeed, if a 95% CI had been used then the upper limit for the positive LR would have been 1.41 (less than the 1.5 limit used) at the first review of data in July 2012. They may have been more inclined to suggest stopping the study prematurely if this had been what they agreed originally but clearly there was no strong desire to do so. We can speculate that if the index test had been more invasive to women (ultrasonography was almost unanimously rated as a very acceptable test by participants) or was directing patient care, then they may have felt more inclined to recommend stopping the study.

In conclusion, despite the extra work need to produce interim analyses during the study, the methodology used here could be adapted for other large diagnostic studies for which results are available quickly. Side benefits of producing interim analysis also include the advantage of a closer examination of the data, which may uncover issue like the quantity of missing data.

Appendix 2 Study accrual

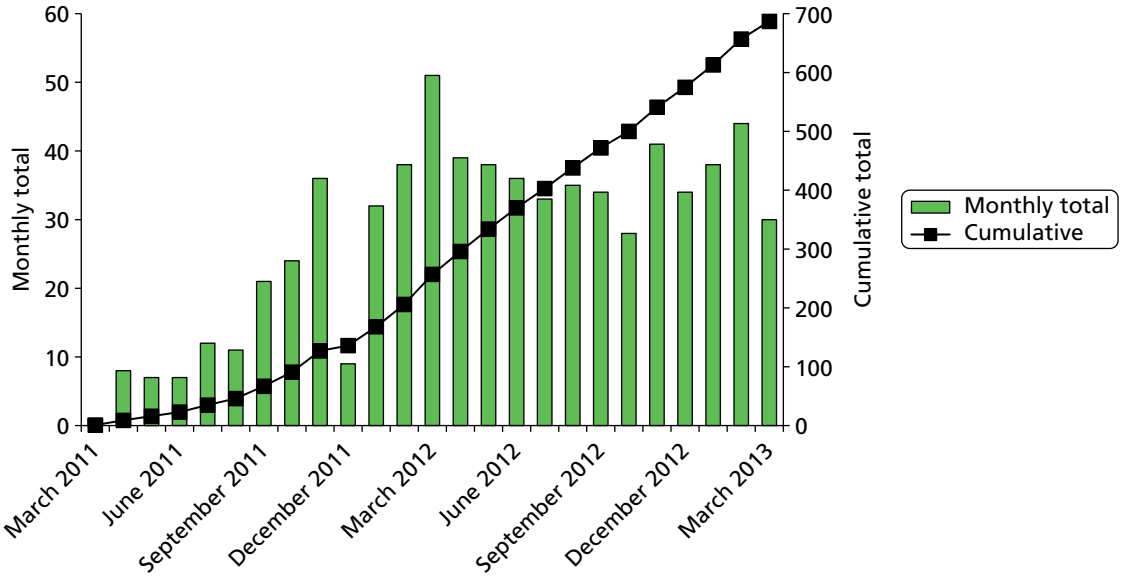

FIGURE 23 Graph showing the recruitment rate per month through the study period.

Appendix 3 Sensitivity analyses

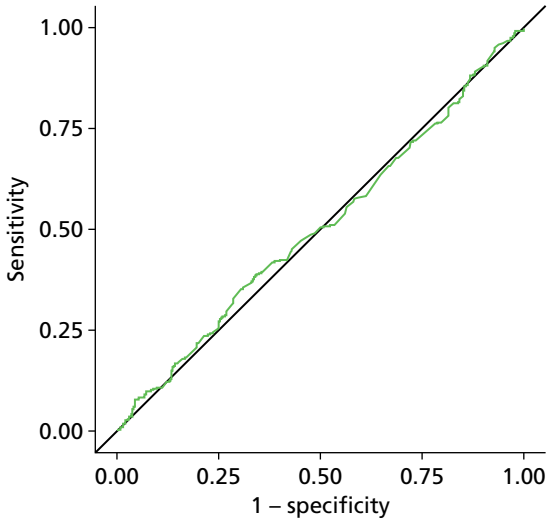

FIGURE 24 The ROC curve from sensitivity analysis using measurement 1 cm left of dome (AUC 0.507, 95% CI 0.461 to 0.553; $p = 0.76$ compared with AUC = 0.50).

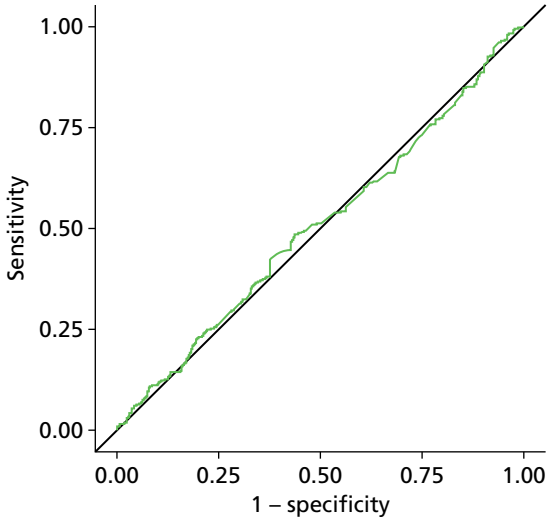

FIGURE 25 The ROC curve from sensitivity analysis using measurement 1 cm right of dome (AUC 0.502, 95% CI 0.456 to 0.548; $p = 0.92$ compared with AUC = 0.50).

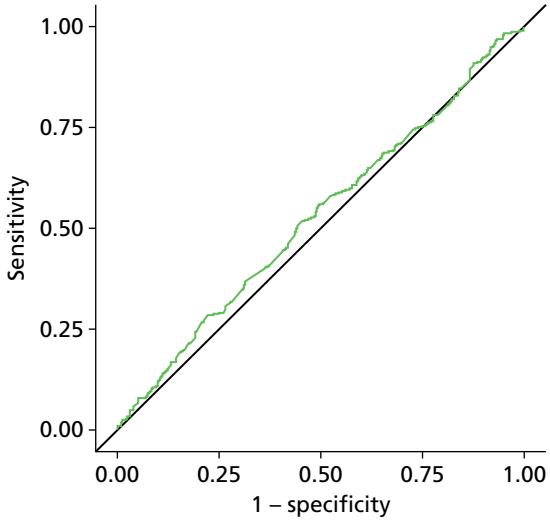

FIGURE 26 The ROC curve from sensitivity analysis excluding those results for which the UDS test was not blind to the results of the ultrasonography test [16/632 women (3%); AUC 0.528, 95% CI 0.480 to 0.575; $p = 0.25$ compared with AUC = 0.50].

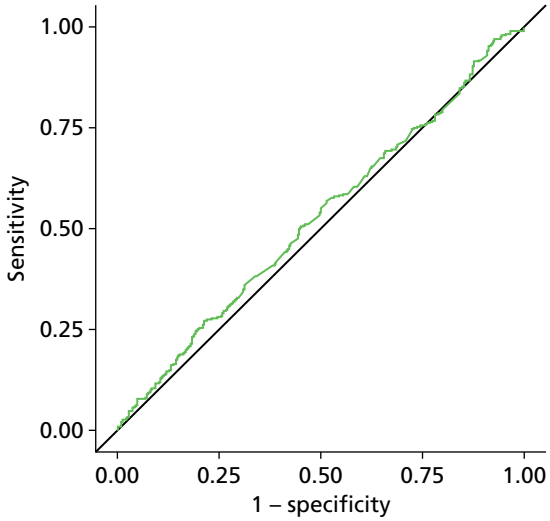

FIGURE 27 The ROC curve from sensitivity analysis excluding those results for which there was more than 4 weeks between the tests [26/660 women (4%); AUC 0.526, 95% CI 0.479 to 0.572; $p = 0.28$ compared with AUC = 0.50).

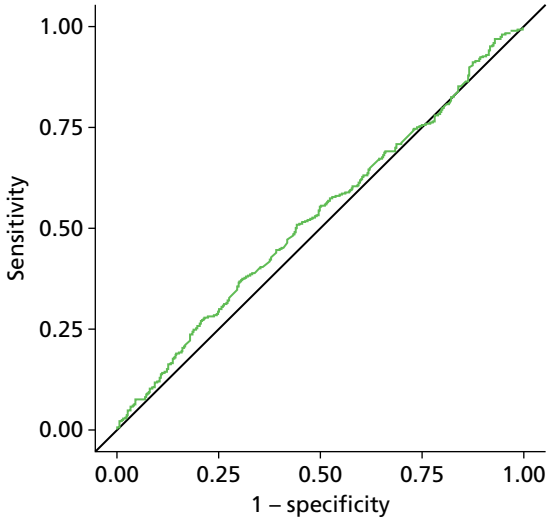

FIGURE 28 The ROC curve from sensitivity analysis incorporating incomplete ultrasonographic measurements (10 observations – average of remaining one or two measurements used; AUC 0.529, 95% CI 0.484 to 0.574; $p = 0.21$ compared with AUC = 0.50).

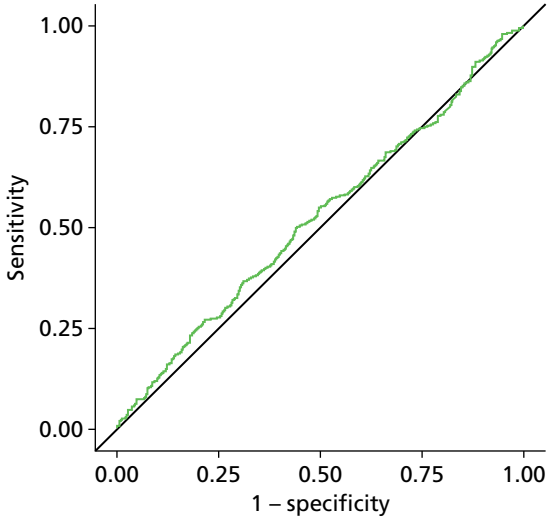

FIGURE 29 The ROC curve from sensitivity analysis using ambulatory UDS diagnosis where available (14 participants; AUC 0.520, 95% CI 0.475 to 0.566; $p = 0.39$ compared with AUC = 0.50).

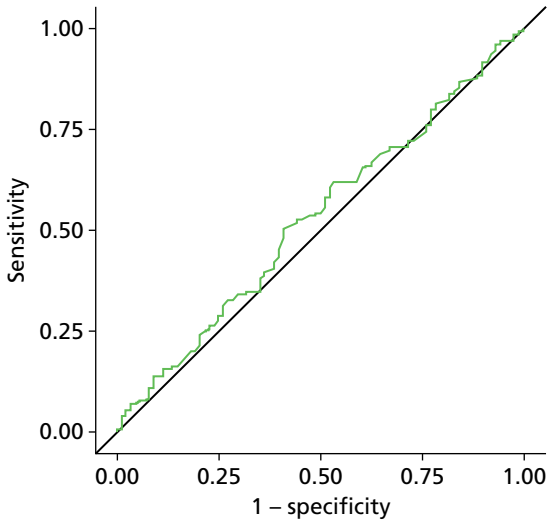

FIGURE 30 The ROC curve from exploratory analysis including the urgency alone group (as per clinical history; excluding mixed stress/urgency incontinence group: 217 patients; AUC 0.530, 95% CI 0.452 to 0.609; $p = 0.45$ compared with AUC = 0.50).

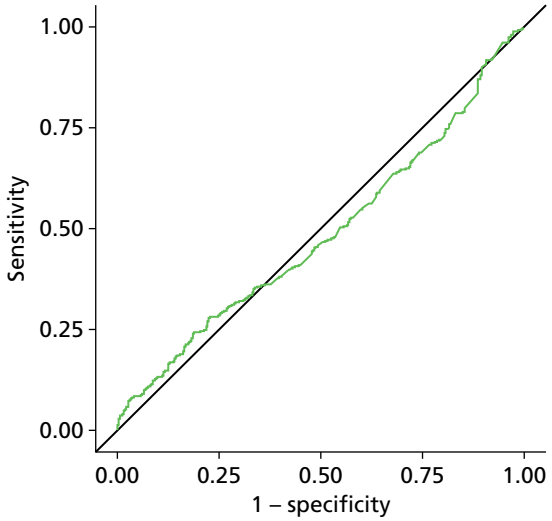

FIGURE 31 The ROC curve from exploratory analysis including the 'pure' DO group only [diagnosis of DO/low compliance/DO plus low compliance, excluding 'mixed' DO (DO with another diagnosis of USI or VD); AUC 0.489, 95% CI 0.440 to 0.531; $p = 0.54$ compared with AUC = 0.50].

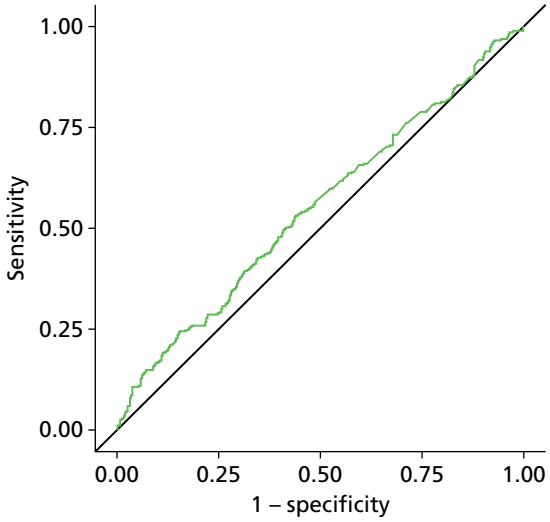

FIGURE 32 The ROC curve from exploratory analysis including the 'wet' DO group only (excluding 'dry' DO; AUC 0.548, 95% CI 0.501 to 0.594; $p = 0.05$ compared with AUC = 0.50).

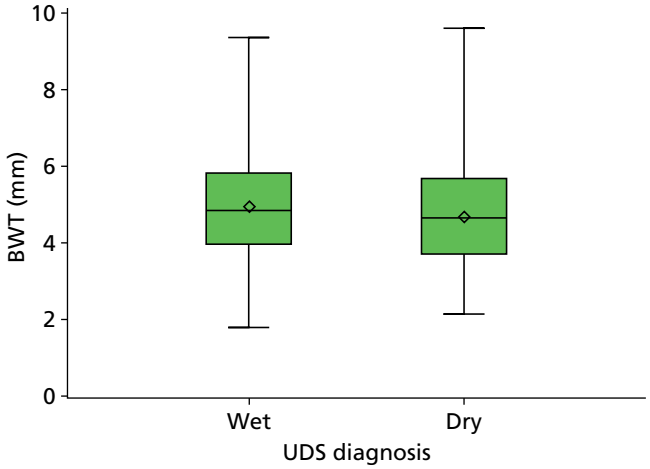

FIGURE 33 Box and whisker plot comparing BWT for 'wet' and 'dry' DO.

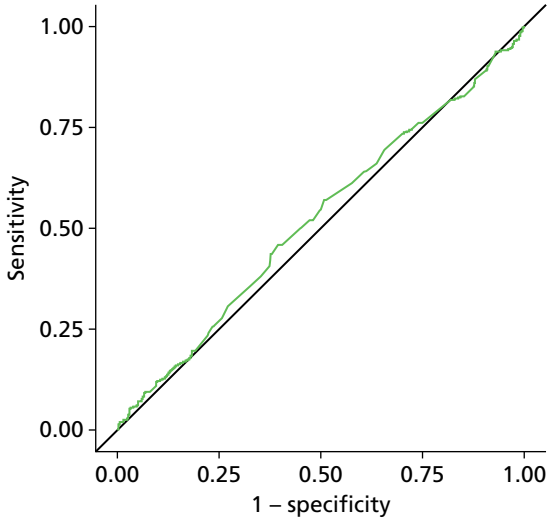

FIGURE 34 The ROC curve from exploratory analysis using the trigone measurement alone for BWT (AUC 0.519, 95% CI 0.473 to 0.564; $p = 0.42$ compared with AUC = 0.50).

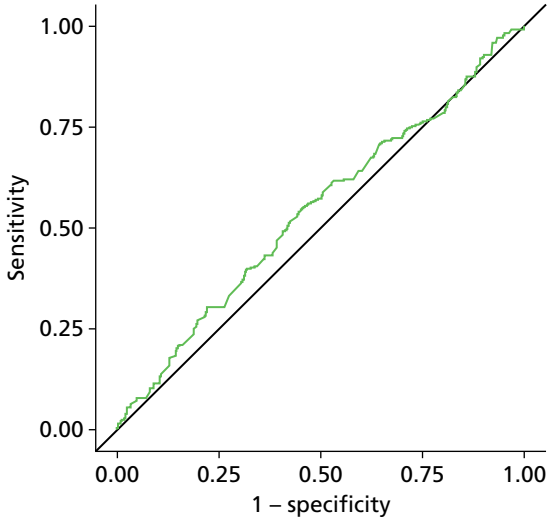

FIGURE 35 The ROC curve from exploratory analysis excluding those who had a detrusor pressure rise on provocation testing 'provoked DO' (187 cases; AUC 0.541, 95% CI 0.487 to 0.595; $p = 0.14$ compared with AUC = 0.50).

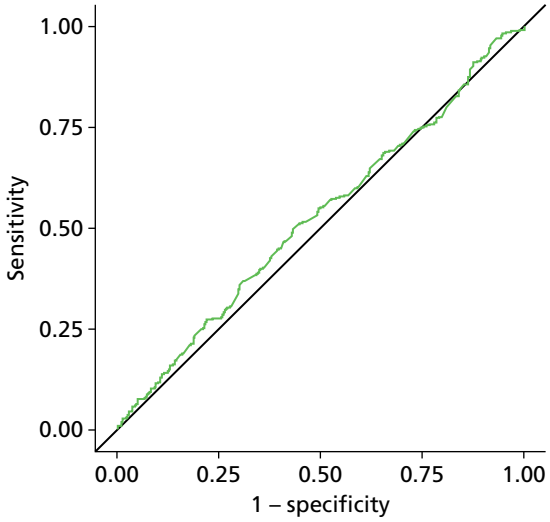

FIGURE 36 The ROC curve from exploratory analysis excluding those who had PVR of > 30 ml on testing (34 cases AUC 0.526, 95% CI 0.479 to 0.572; $p = 0.28$ compared with AUC = 0.50).

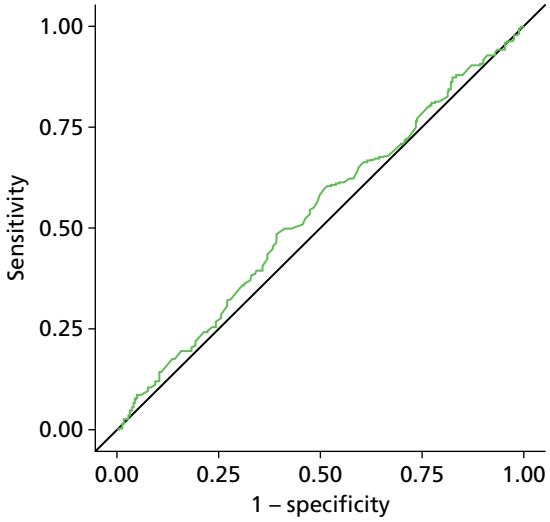

FIGURE 37 The ROC curve from exploratory analysis using the average of dome, 1 cm left of dome, 1 cm right of dome (AUC 0.537, 95% CI 0.491 to 0.582; $p = 0.12$ compared with AUC = 0.50).

Appendix 4 Subgroup analyses

TABLE 45 Results of ROC curve analysis in pre-specified subgroupings

Variable	Value	AUC	95% CI	*p*-value for difference between AUCs
Previous treatment with antimuscarinics	No	0.536	0.481 to 0.592	0.48
	Yes	0.501	0.420 to 0.582	
Clinical history suggested mixed incontinence	No	0.534	0.460 to 0.608	0.73
	Yes	0.518	0.460 to 0.575	
Presence of UTI in the last 12 months	No	0.530	0.482 to 0.578	0.53
	Yes	0.586	0.417 to 0.755	
Patients with voiding difficulties	No	0.533	0.472 to 0.594	0.84
	Yes	0.524	0.454 to 0.593	
Previous incontinence surgery	No	0.526	0.479 to 0.573	0.76
	Yes	0.493	0.294 to 0.693	
BMI	<25	0.519	0.424 to 0.614	0.95
	≥ 25	0.523	0.471 to 0.575	

TABLE 46 Results of univariate analysis exploring factors possibly associated with DO diagnosis

Variable	Data type	*p*-value	OR (95% CI) if statistically important	Frequencies (binary/categorical data)
ICIQ score (best = 0, worst = 16)	Continuous	<0.0001	1.23 (1.15 to 1.31)	
BWT, mm	Continuous	0.19		
Age, years	Continuous	0.66		
Duration of symptoms, years	Continuous	0.45		
BMI, kg/m^2	Continuous	0.38		
Ethnicity (white/black/Asian/other)	Categorical	0.59		
Vaginal birth, yes	Binary	0.64		
Clinical history suggests mixed incontinence, yes	Binary	0.40		
If clinical history suggests mixed incontinence, which came first (stress/urge/unsure/NA)	Categorical	0.66		
Previous treatment with antimuscarinics, yes	Binary	0.001	1.74 (1.24 to 2.44)	68% (152/222) DO when = yes; 56% (245/441) DO when = no
Previous UTI in last 12 months, yes	Binary	0.08	0.60 (0.34 to 1.07)	48% (24/50) DO when = yes; 61% (363/599) DO when = no
History of voiding difficulties, yes	Binary	0.16		
Post menopausal, yes	Binary	0.67		
Parity (0/1/2/3/4+)	Categorical	0.27		
Previous incontinence surgery, yes	Binary	0.59		
Previous POP surgery, yes	Binary	0.32		

NA, not applicable.

TABLE 47 Results of multivariable analysis exploring factors possibly associated with DO diagnosis

Model	Significant variables	*p*-value	OR (95% CI) if significant
Backward selection ($p = 0.1$ to stay in model)	ICIQ score	< 0.0001	1.21 (1.13 to 1.29)
	Previous UTI in last 12 months	0.04	0.51 (0.27 to 0.97)
All variables included	ICIQ score	< 0.0001	1.21 (1.13 to 1.29)
	Previous UTI in last 12 months	0.06	0.53 (0.27 to 1.03)
All variables included, multiple imputation used for missing data	ICIQ score	< 0.0001	1.23 (1.15 to 1.31)
	Previous treatment with antimuscarinics	0.02	1.57 (1.09 to 2.28)
	Previous UTI in last 12 months	0.07	0.57 (0.31 to 1.06)

Appendix 5 Standard operating procedure for urodynamics

The UDS should be performed with aseptic precautions, counselling and verbal consent and according to the Good Urodynamics Practice Guidelines (Schafer *et al.*[96]).

The equipment needed for the running of the urodynamic clinic include:

- catheter pack
- filling catheter
- abdominal and bladder pressure catheters
- instillagel/sterile lubricant gel
- Four × 3-way taps (depending on the type of transducers being used)
- Two × fluid-filled domes
- One 500 ml bag of normal saline used for irrigation
- One pump infusion set
- One set guard
- Two × giving sets
- Two × 100 ml bag normal saline to flush the domes
- Three × pieces of tape (micropore, etc., to attach once catheters inserted; ensuring they stay in place during filling)
- One pair of sterile gloves
- non-sterile gloves
- Two × incontinence pads (one used for the floor and one for the patient to sit on)
- paper roll to cover the couch
- sharps box
- plastic apron
- towel or cover for the patient
- clean trolley with antiseptic wipes.

Please note: the above items may vary depending on the type and make of equipment used in each clinic, supplies used at the trust and also in accordance to infection control and hospital policies.

- Ensure all equipment is set up and the person performing the test has not also undertaken ultrasonography on the patient.
- Ensure the UDS test form is to hand and the patient registration number is entered.
- If, for any reason, the test had to be abandoned, note this on the test form.

Uroflowmetry (initial voiding test)

- The patient is asked to attend clinic with a comfortably full bladder.
- The patient should be encouraged to sit in order to void into the voiding flow/volume transducer funnel mounted under the commode.
- The patient should be instructed to dispose of any tissues/wipes into the bin/bag provided and not into the flow metre.
- The utmost privacy must be maintained during the test and the patient should be made to feel comfortable and relaxed, enabling a usual voided pattern to be established.
- The maximum void flow rate and volume should then be recorded.
- The PVR volume should then be recorded using a drainage catheter and measuring container.

Filing cystometry (catheterisation)

It is essential that the machine is calibrated, set at zero at atmospheric pressure and a reference level for pressures should be established.

- Ideally the patient should be in the sitting position for the test. If this is not possible it should be recorded on the test form. A sheet should be provided for covering, maintaining dignity.
- Under aseptic technique, introduce catheters up through the urethra into the bladder and one into the rectum.
- Prior to filling, ask the patient to cough so that the traces can be observed. The spikes on the intravesical and intra-abdominal lines should be identical. Any necessary adjustments should be made and the cough repeated.
- Fill rate should be recorded on the test form, but is recommended as 100 ml per minute.
- Ask the patient to cough every minute to ensure continued subtraction. If the lines slip, then stop the filling and rectify the problem.
- Complete test form with the number of millilitres at which the patient reports first, normal and strong desire to urinate, pain and volume leaked (if applicable).
- Total volume in the bladder at end of filling should be recorded.
- Detail any rise in detrusor pressure with or without urgency.
- At the end of filling, the large catheter used for filling the bladder is removed. The small catheter remains in the bladder to record voiding pressures (if using two separate catheters in the bladder).

Provocation test (while bladder is still filled)

While the intravesical and intra-abdominal lines are in situ, the patient should stand up on the incontinence sheet provided and the provocation tests such as running taps, coughing, etc., should be performed.

- Complete methods used and observations on the test form.

Voiding cystometry

- Allow patient to void into commode, recording peak flow rate, maximum void pressure and residual volume with the pressure lines still in. During this voiding phase, the patient's dignity and privacy must be maintained and staff should leave the room if necessary.
- Ask the patient to cough pre-void and post void to ensure adequate subtraction.

Diagnosis

On completion of the investigation, the results may be explained to the patient and fluid advice should be given.

- Record diagnosis in red section of test form.
- Any 'optional'/additional tests undertaken should be noted at the end of the form.
- If video UDS is being done, then it is recorded at the end of the form.

Optional tests

- If patient is scheduled for ambulatory UDS, then please give this information on the form.

Advice for patients

All women who have undertaken the test should be advised to expect some dysuria for up to 72 hours, their fluid intake should be increased during this time.

The occurrence of systemic symptoms, pyrexia and malaise should be advised as an indication to seek medical advice, that is, from their GP.

A contact number should be provided if problems occur.

Appendix 6 Bladder Ultrasound Study standard operating procedure for scanning

To be used in conjunction with training video.

Clinician preparation

- The clinician performing the bladder ultrasonography should be different from the clinician performing the UDS, to ensure blinding between the two tests.
- If, for any reason, this is not possible, bladder ultrasonography should be performed **before** the UDS.

Patient preparation

- The patient may be seen in various settings:
  - clinic (patient may need other assessment)
  - scan department (may require renal tract assessment and attends with full bladder).
- Ensure patient empties bladder **before** assessment of BWT and PVR.
- Important to stress need to void as completely as possible.

Machine and probe preparation

- Ensure scanner is capable of measurement in millimetres (mm).
- Set to the scanner to the gynaecological preset.
- Use a transvaginal probe (not a rectal probe)
  - multifrequency – use between 7 and 9 MHz for optimal image (no lower than 5 MHz).
  - Prepare the probe:
    - clean
    - put gel into probe cover, excluding air
    - put gel onto tip of probe.

Timescales

- Ultrasonography should ideally be completed in a one-stop clinic with the UDS test.
- The two tests should be undertaken by different clinicians to ensure blinding of results.
- If it is not possible to hold one-stop clinics, ultrasonography and UDS tests should be undertaken no more than 4 weeks apart.
- The second test can be undertaken up to a further 4 weeks after this cut-off point (8 weeks in total from the first test), but the data collected will be classed as a protocol violation.
- If the patient is happy to have one or both tests retaken (effectively constituting a second 'set' of tests), they can do so, as long as they have not become ineligible in the interim (i.e. begun medication).
- An interval of more than 8 weeks between the two tests will be designated as breach of protocol and no per patient payment will be provided for these patients.

Ultrasonography assessment

- The patient should be in the supine position (stirrups or pad under pelvis as appropriate).
- The transvaginal probe should be inserted into the introitus in the longitudinal plane.
- Position should then be assessed on screen.
- Identify the urethra in sagittal orientation.
- Position probe such that the vesicoureteric junction is close to top of screen (*Figure 38*).
- Ensure image sizing is appropriate for screen, between 5 cm and 7 cm depth.
- Focal zone positioned at region of interest; multiple focal zones if possible to give good definition at various levels.

The PVR and BWT should be measured before any other assessment, as in patients who have filled their bladder beforehand, it may fill during scan.

Post-void residual

- Identify entire bladder in sagittal plane, measure longest anteroposterior dimension and then cranio–caudal dimension perpendicular to this (*Figure 39*).
- Rotate probe through 90° and measure axial dimension (*Figure 40*).
- Most machines are now able to generate volumes automatically (need to select this before starting with callipers).
- If not available, use the following to standardise calculations: cranio–caudal (H) × anterior–posterior (D) × transverse diameter (W) × 0.5233 (prolate ellipsoid) = PVR volume.

Eligibility: proceeding to bladder wall thickness measurement

- If PVR $\leq$ 30 ml, proceed with BWT measurement.
- If PVR > 30 ml, ask patient to revoid.
- If, after re-void, PVR is > 30 ml but < 100 ml measure the BWT.
- If PVR is > 100 ml, exclude patient from bladder ultrasonography.

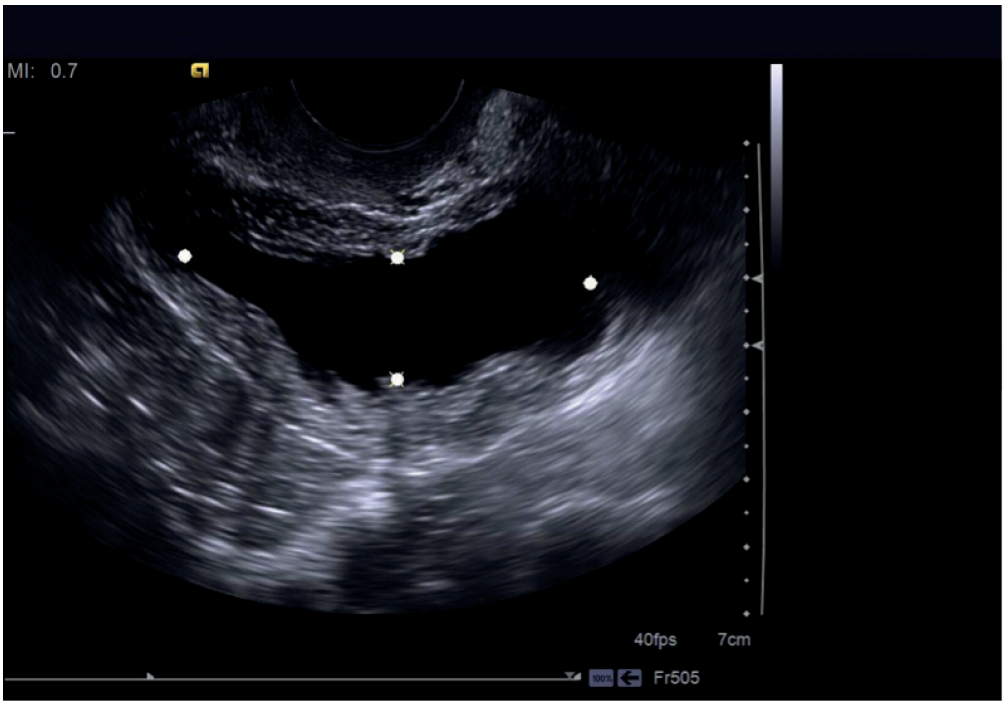

FIGURE 38 Image showing urethra and bladder neck near top left.

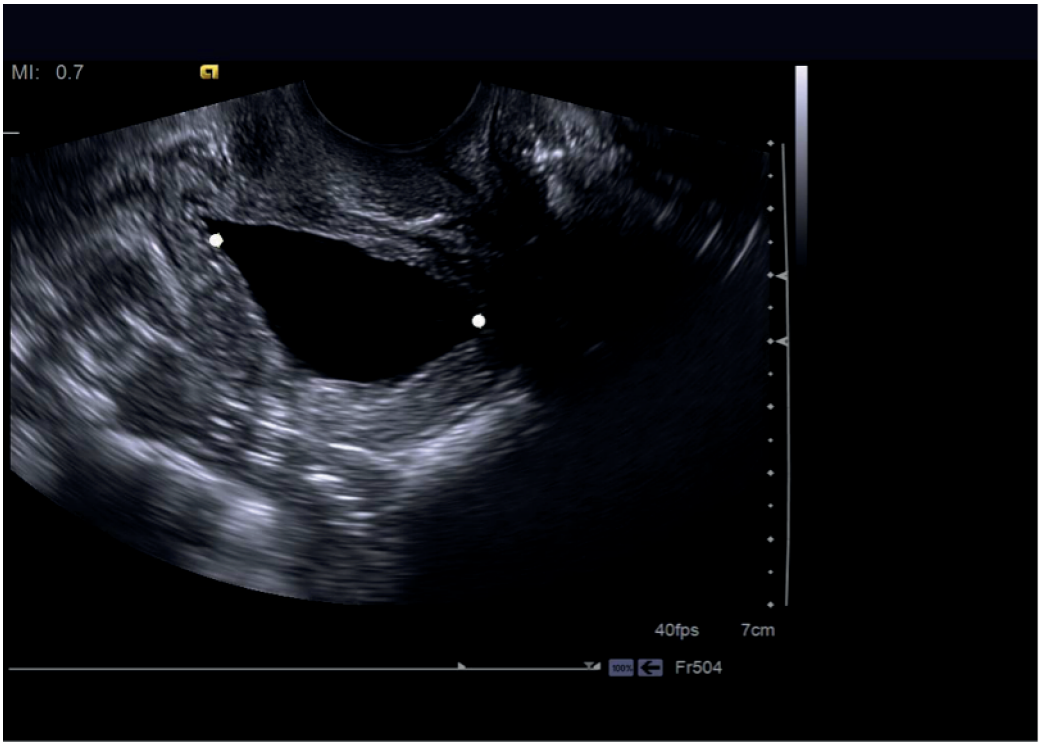

FIGURE 39 Transvaginal scan in sagittal section to measure the anteroposterior and cranio–caudal dimensions of the bladder.

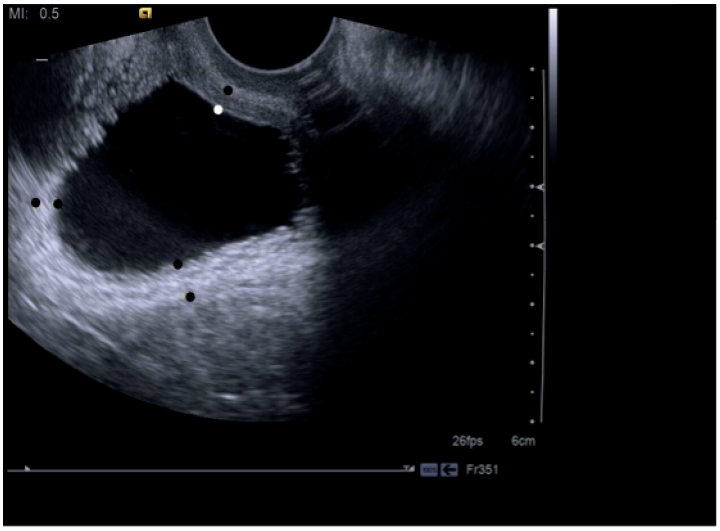

FIGURE 40 Transvaginal ultrasound in coronal section to measure the axial dimension of the bladder.

Bladder wall thickness assessment

- Measurement should be obtained in sagittal plane with vesicoureteric junction on margin of screen (probe may need to be introduced slightly further).
- Three × measurements: at the dome, anterior wall and trigone need to be obtained. Then also measure BWT at dome 1 cm to the left and right of midline (*Figures 41* and *42*).
- The same image should be stored twice: with and without callipers.
- Any focal areas of thickening need separate assessment and evaluation.
- It is not always possible to see the anterior wall or trigone well, as many women with bladder problems often have an irregular bladder outline, but it is always possible to see the dome.
- If the whole of the bladder is not visible on one image (as is preferable), there may be the need to angle anteriorly or posteriorly and measure the trigone or anterior wall thickness.
- Patients with significant cystocoeles may find complete emptying difficult and trigone may be difficult to visualise adequately.
- It is important to document any additional findings (e.g. diverticuli, cystocoeles, focal masses).

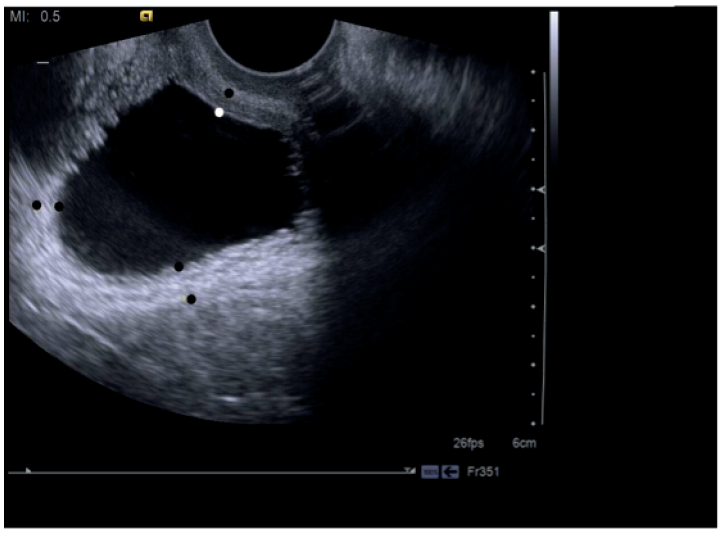

FIGURE 41 Transvaginal scan with callipers showing measurements on the dome, anterior wall and trigone.

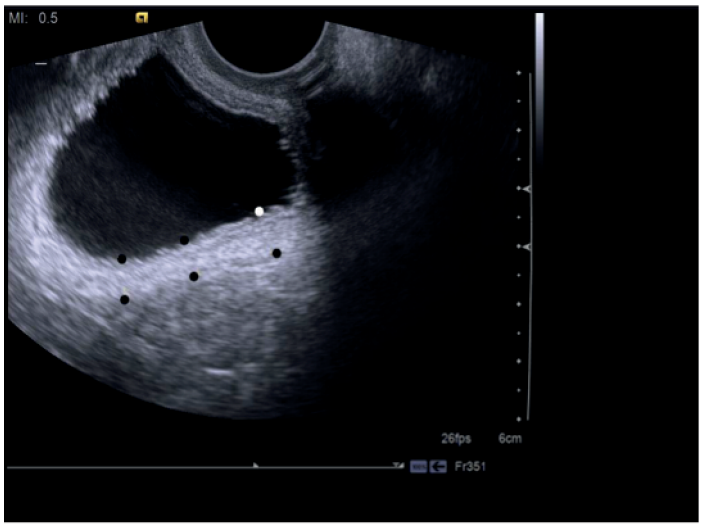

FIGURE 42 Transvaginal scan showing the three dome measurements.

Providing electronic images

Anonymised electronic images for each patient should be provided, labelled clearly with study identifier as the only identifier. These images should be as follows:

1. PVR image 1 WITH callipers.
2. PVR image 1 WITHOUT callipers.
3. PVR image 2 WITH callipers.
4. PVR image 2 WITHOUT callipers.
5. BWT image 1 WITH callipers.
6. BWT image 1 WITHOUT callipers.
7. BWT image 2 (optional) WITH callipers.
8. BWT image 2 (optional) WITHOUT callipers.

For optimal quality these should be sent to Birmingham Clinical Trials Unit on disc or memory stick, ideally in DICOM format.

Failing this JPEG (Joint Photographic Experts Group) images will be accepted.

If downloadable images are not possible, please supply a print out of each image listed above, ensuring that the print quality is optimal (what is seen on screen).

Summary

- Prepare patient (ensure voiding takes place just prior to scan and measure in supine position).
- Prepare machine (use vaginal probe and frequency 7–9 MHz for optimal image).
- Urgency severity scores assessment
    - Important to obtain good-quality images.
    - Callipers placed **on** margin of bladder wall for thickness; within bladder lumen (on wall) for volumes.
    - Store images **with** and **without** callipers for cross-referencing and evaluation.
    - Measure if PVR is ≤ 30 ml, revoid if PVR is > 30 ml and measure if PVR is > 30 ml but < 100 ml, exclude if PVR is > 100 ml.

Appendix 7 Summary of studies included in the updated systematic review on accuracy of bladder ultrasonography for diagnosing detrusor overactivity

TABLE 48 Study characteristics

Study, date, design	Population	Test	Reference standard	Mean (±2 SD or 95% CI) among those with and without DO in mm/specificity and sensitivity % at cut-off point > 5mm
Khullar 1994,[97] prospective	45 women with LUTS and attending UDS clinic; 19 detrusor instability; 20 stress incontinence	Transvaginal ultrasonography; residual volume < 20 ml; measurements perpendicular to the luminal surface at the thickest part of the trigone, at the dome and at the anterior wall. Operator performing BWT was blinded to UDS diagnosis	Video UDS in supine position with a fill rate of 100 ml/minute	Mean (SD) BWT: 6.7 mm (0.6 mm) in DO; 3.5 mm (0.6 mm) in SUI; difference between two conditions $p < 0.001$ on Mann–Whitney *U*-test
Khullar 1996;[49] prospective	184 women attending clinic owing to urinary symptoms, of whom 4 were excluded due to excessive PVR volume. Videocystourtherography UDS diagnosis: 43 detrusor instability; 52 SUI; 43 MUI; 34 normal UDS; 5 voiding difficulty; 3 sensory urgency)	Transvaginal ultrasonography; residual volume < 50 ml; measurements perpendicular to the luminal surface at the thickest part of the trigone, at the dome and at the anterior wall. Operator performing BWT was blinded to UDS diagnosis	Video UDS in supine position with a fill rate of 100 ml/minute; provocative tests used. BWT < 3.5 mm or with > 5 mm BWT but no DO also had ambulatory UDS ($n = 42$)	Median (IQR) BWT: 6.3 mm (5.3–7.7 mm) in DO; 3.9 mm (3.4–4.5 mm) in other groups; difference between two conditions $p < 0.0001$ on Mann–Whitney U-test. At BWT > 5 mm, 84% (75.8–89.7%) sensitivity; 89% (78.8–96.11%) specificity
Robinson 2002;[52] prospective	128 women with OAB with normal or equivocal UDS referred for ambulatory UDS (21 detrusor instability; 43 stress incontinence; 26 mixed incontinence; 37 normal; 1 other)	Transvaginal ultrasonography; residual volume < 50 ml; measurements perpendicular to the luminal surface at the thickest part of the trigone, at the dome and at the anterior wall. Operator performing BWT was blinded to UDS diagnosis	Ambulatory UDS; clinician performing UDS was blinded to BWT thickness	Mean (95% CI) BWT: 6.7 mm (6.0 mm to 7.4 mm) in DO; 5.1 mm (4.6 mm to 5.6 mm) in normal; 4.8 mm (4.4 mm to 5.3 mm) in USI; 5.8 mm (5.1 mm to 6.5 mm) in MUI; difference between four conditions. One-way ANOVA $p = 0.0001$
Soligo 2002;[58] prospective (conference abstract)	161 women with urinary symptoms; 70 with OAB symptoms (46 stable; 24 unstable bladder); 91 without OAB symptoms (76 stable; 15 unstable bladder)	Does not indicate which type of ultrasonography; sites of BWT measurement not elaborated. Mean BWT was calculated	UDS performed in all women	Mean (95% CI) BWT: 5.0 mm (4.6 mm to 5.3 mm) OAB with DO; 3.6 mm (3.4 mm to 3.9 mm) OAB no DO; difference between two groups $p < 0.001$ one-way ANOVA. Combination BWT ≥ 5 mm and OAB symptoms, PPV 83.3%; NPV 83.2%
Yang 2002;[59] retrospective with healthy controls	1049 women with LUTS (190 detrusor instability; 764 stress incontinence; 95 hypersensitive bladder); additional 36 healthy controls	Transvaginal ultrasonography; measurements at dome and trigone; residual volume < 50 ml; cystourethrography	Uroflowmetry, filling and voiding phase cystometry, and a urethral pressure profile at both resting and Valsalva manoeuvre	Mean (SD) BWT: 5.8 mm ± 1.9 mm with DO; 6.0 mm ± 2.4 mm with SUI; 5.3 mm ± 1.9 mm hypersensitive bladder; 4.9 mm ± 2.1 mm controls. Pairwise comparisons adjusted for multiple testing, all significant ($p < 0.05$) except DO vs. SUI

Study, date, design	Population	Test	Reference standard	Mean (± 2 SD or 95% CI) among those with and without DO in mm/specificity and sensitivity % at cut-off point > 5mm
Yang 2003;[53] retrospective	492 women with LUTS with normal urinalysis findings, negative urine culture results, or both. (UDS diagnosis: 38 DO; 248 SVI; 39 MUI; 35 hypersensitive bladder; 42 voiding difficulties; 90 normal UDS)	Transvaginal ultrasonography; measurement at dome and trigone; residual volume < 50 ml; cystourethrography	UDS at a filling rate of 80 ml/minute with patient sitting upright in a birthing chair	BWT measurements only reported in a figure
Chan 2005;[114] prospective (conference abstract)	86 women with VD; 22 DO; 22 sensory urgency; 42 normal	Transabdominal ultrasonography; volume 200 ml; measurement of anterior wall	No details of UDS given	Mean BWT: 1.7 mm in DO; 1.6 mm in sensory urgency; 1.7 mm in normal. No significant difference $p = 0.18$
Parsons 2005;[115] prospective with healthy controls (conference abstract)	250 women. 194 women with troublesome urinary tract symptoms (31 DO; 33 normal OAB; 26 withdrew; 104 not mentioned); 61 asymptomatic controls	Transvaginal ultrasonography; residual volume < 50 ml; measurements of the trigone, at the dome and at the anterior wall	Video UDS in all women	Mean BWT: 4.862 mm in DO; 4.085 mm in normal OAB; 3.92 mm in controls; *t*-test DO vs. normal OAB $p = 0.007$; *t*-test DO vs. control $p = 0.007$
Minardi 2007;[50] prospective with healthy controls	80 women, including 66 referals for meantinence and 14 healthy controls. UDS diagnosis of symptomatic women: (36 SUI; 30 urgency incontinence)	Translabial or introital ultrasonography; no mention of residual volume; measurement of DWT at bladder dome; performed without knowledge of UDS results but by same assessor	UDS using Duet MultiP (Medtronic, Minneapolis, MN, USA) according to International Continence Society criteria	Mean (unstated measure of variation) DWT: 7.1 mm (1.6 mm) in urgency incontinence; 4.1 mm (1.1 mm) in stress incontinence; 3.9 mm (1.9 mm) in controls. ANOVA for differences between three groups $p = 0.019$
Blatt 2008;[116] prospective	180 patients with non-neurogenic VD including 107 women (34 had DO; 39 had BOO; 38 had increased bladder sensation; 69 normal)	Transabdominal BWT measurements; 200 ml filling; measurements of the anterior bladder wall, 1 cm apart in the midline	Video UDS in all patients included uroflowmetry, filling cystometry at the rate of 50 ml per minute and pressure flow measurements	Mean (unstated measure of variation) BWT: 2.0 mm (0.53 mm) normal; 2.1 mm (0.47 mm) BOO; 1.9 mm (0.43 mm) DO; 1.8 mm (0.48 mm) increased bladder sensation. No significant difference was found between the groups (ANOVA $p = 0.064$)
Lekskulchai 2008;[51] retrospective	686 women attending a tertiary UDS service; four different UDS diagnoses were made, more than one diagnosis being made in several women. 184 DO; 135 sensory urgency. Number not stated for stress incontinence; number not stated for VD	Translabial ultrasonography; detrusor thickness measured at the bladder dome. Mean of three separate measurements was taken; measurements after voiding	Multichannel UDS	Mean (SD) DWT: 4.7 mm (1.9 mm) for DO; 4.1 mm (1.6 mm) for not DO; *t*-test $p < 0.001$. For 5 mm cut-off point, sensitivity 37%; specificity 79%; PPV 40%; NPV 70%; AUC 0.606 (95% CI 0.56 to 0.65)

continued

TABLE 48 Study characteristics (*continued*)

Study, date, design	Population	Test	Reference standard	Mean (±2 SD or 95% CI) among those with and without DO in mm/specificity and sensitivity % at cut-off point > 5mm
Kuo 2009;[56] prospective with healthy controls	92 women in total. 28 OAB (dry) 25 OAB (wet) 28 normal controls and 11 controls with renal stones, lower back pain, inguinal hernia and/or turbid urine complaints. On video UDS of the women with OAB and normal controls, 22 had DO, 32 had hypersensitive bladder and 27 had a normal UDS	Transvaginal DWT scan on empty bladder and transabdominal DWT at bladder capacity. After uroflowmetry, transvaginal DWT was measured at bladder neck, bladder base, anterior and posterior wall. Transabdominal DWT was measured on anterior wall at three sites and an average obtained	All women underwent video UDS	Transvaginal DWT: no significant difference at the bladder neck, anterior wall, posterior wall and bladder base among women with between normal, hypersensitive bladder and DO. Transabdominal DWT was greater in DO group at maximum capacity. A transabdominal DWT of 0.75 mm at bladder capacity had: sensitivity 73%; specificity 67%; AUC of 0.776 (95% CI 0.643 to 0.909) by natural filling
Panayi 2010;[72] prospective (conference abstract)	182 women reporting symptoms; numbers per UDS diagnosis group are not given	Transvaginal ultrasonography; residual volume < 50 ml; measurements of the trigone, at the dome and at the anterior wall; clinician measuring BWT was blinded to UDS	UDS were performed in all women	Results are not presented grouped by the UDS diagnoses but by the presenting symptoms
Serati 2010;[113] prospective	247 women who attended the urogynaecology unit between 2005 and 2008. Diagnosis following UDS: 66 UDS SUI; 72 MUI; 75 isolated DO; 34 normal	Transvaginal ultrasonography; residual volume < 50 ml; measurements of the trigone, at the dome and at the anterior wall; clinician measuring BWT was blinded to UDS	UDS in all women	Mean (SD) BWT: 5.22 mm (1.17 mm) DO; 4.09 mm (0.86 mm) SUI; 4.73 mm (1.27 mm) MUI; 4.19 mm (1.14 mm) normal. At a cut-off point of 5 mm BWT: 50.34% sensitivity; 85.0% specificity; AUC 0.704 (95% CI 0.64 to 0.77)
Chung 2011;[28] prospective with healthy controls	122 women presenting with LUTS (83 from clinics with wet or dry OAB, 39 normal controls); by UDS: 28 normal; 30 increased bladder sensation; 30 DO. In addition, 39 untested by UDS presumed to be normal controls	Transabdominal DWT; measured with full bladder	88 of 122 women underwent video UDS	Mean (SD) DWT: 0.95 mm (0.42 mm) in DO; 0.85 mm (0.31 mm) in increased bladder sensation; 0.85 mm (0.31 mm) in normal. No significant difference of DWT (no test result presented)

Study, date, design	Population	Test	Reference standard	Mean (± 2 SD or 95% CI) among those with and without DO in mm/specificity and sensitivity % at cut-off point > 5mm
Kuhn 2011;[54] prospective	122 women with lower urinary symptoms; MUI was excluded (59 had SUI; 40 had DO; 24 had obstruction)	Transvaginal BWT; residual volume < 50 ml; technique not described.; clinician measuring BWT was blinded to UDS	UDS performed in all women in sitting position	Mean BWT (variance measure not stated): 3.78 mm (0.39 mm) in SUI; 4.97 mm (0.63 mm) in DO; 6.01 mm (0.73 mm) with obstruction; ($p < 0.0001$). AUC calculated was 0.87 mm (95% CI 0.78 mm to 0.97 mm; $p < 0.0001$). At a cut-off point of 4.4 mm to diagnose overactive or obstructive incontinence; 90.6% sensitivity; 96.6% specificity
Ibrahim 2011;[55] prospective case–control study (conference abstract)	60 women (30 cases who had detrusor instability; 30 controls without LUTS)	Transvaginal BWT; measurements of trigone, dome and anterior wall; clinician measuring BWT was blinded to UDS result	UDS undertaken in the 30 cases only	Mean (variance measure not stated) BWT: 5.00 mm (1.09 mm) in DO; 4.17 mm (0.91 mm) in controls. At a cut-off point of 5 mm; 53.3% sensitivity; 86.7% specificity
Ozturk 2011;[57] prospective with healthy control (conference abstract)	82 women from outpatient clinic (39 with DO; 43 with SUI), 31 controls	Transabdominal BWT; 200 ml volume; measurements of anterior wall, right and left lateral wall	UDS	Mean BWT not presented: at a cut-off point of 4.88 mm; 87.1% sensitivity; 60.8% specificity
Abou-Gamrah 2014;[117] prospective	100 women with urinary symptoms, mixed incontinence was excluded (50 with detrusor instability; 50 with stress incontinence)	Transvaginal ultrasonography; residual volume < 50 ml; measurements of the trigone, at the dome and at the anterior wall; measurement blind to UDS	UDS	Mean BWT not presented: at a cut-off point of 4.78 mm; 90% sensitivity; 78% specificity; AUC 0.905
Otsuki 2014;[118] prospective with continent controls	91 women [30 stress incontinence; 30 DO; 31 continent (other gynaecological conditions)]	Transvaginal ultrasonography; residual volume < 50 ml; measurements of the trigone, at the dome and at the anterior wall; measurement blind to UDS	UDS	–
Silva 2014;[119] prospective	213 men and 59 women with neurogenic LUTS due to spinal injury (153 with DO or detrusor sphincter dyssynergia; 119 without detrusor sphincter dyssynergia)	Suprapubic ultrasonography; full bladder; single measurement of anterior bladder wall	Multichannel UDS in supine position	Mean (SD) BWT: 4.2 mm (1.3 mm) in NDO/DSD; 3.6 mm (1.2 mm) in not DSD. Difference between two conditions: $p < 0.001$ *t*-test; AUC 0.624 (95% CI 0.530 to 0.718)

BOO, bladder outflow obstruction; NDO/DSD, neurogenic detrusor overactivity associated with detrusor spincter dyssynergia; NVP, negative predictive value; PPV, positive predictive value.

Appendix 8 Interobserver and intraobserver variation studies

Study	Patients and study design	Technique and route of scan	Results presented	Comment on results
Khullar 1994[44]	10 women each received one scan which was interpreted twice by each of two readers	Transvaginal BWT	Intraobserver difference –0.02 mm, 95% CI[a] –0.22 mm to 0.18 mm; interobserver difference 0.02 mm, 95% CI[a] –0.32 mm to 0.35 mm	Not possible to ascertain the analytical variability, the smallest real difference or the ICC from the data presented. An approximate (under) estimate of the analytical variability can be obtained by dividing the SD of differences by the square root of 2, i.e. 0.3 mm for intra and 0.5 mm for inter observers
Lekskulchai 2008[51]	67 women each had one scan read once by two different readers	Translabial and DWT at dome	ICC estimate of ICC = 0.82, 95% CI 0.63 to 0.91	No estimates of analytical variability or smallest real difference can be computed
Kuo 2009[56]	10 women received two scans two weeks apart	Transvaginal and transabdominal DWT measurement	Pearson's correlation coefficients are reported for transvaginal measures: bladder base 0.833 ($p = 0.020$), anterior wall 0.759 ($p = 0.05$), posterior wall 0.599 ($p = 0.155$), bladder neck 0.768 ($p = 0.044$)	No estimates of analytical variability, smallest real difference or ICCs can be computed
Panayi 2010[72]	25 women each had two scans by two different operators on the same day	Transvaginal BWT at dome, anterior wall, and trigone	Mean difference and 95% CI for the three locations are: 0.13 mm (0.08 mm to 0.33 mm); 0.10 mm (–0.12 mm to 0.31 mm); –0.22 mm (–0.41 mm to 0.01 mm)	Not possible to ascertain the analytical variability, the smallest real difference or the ICC from the data presented. An approximate (under) estimate of the analytical variability can be obtained by dividing the SD of differences by the square root of 2, i.e. 0.4 mm for all three measures
Pannek 2013[214]	10 women had two measurements made by the same observer (and implies these were from different scans which were repeated immediately)	Transabdominal DWT at three different sites of the bladder	States that interobserver coefficient of variability was +14.78%, and the correlation (not stated whether Pearson's or ICC) was 0.984	The mean DWT is not reported in the paper, thus it is not possible to deduce the analytical variability, the smallest real difference and the ICC
Tubaro 2013[129]	40 women each had one scan which was interpreted twice by each of three readers. A further 1504 images were assessed twice by different readers	Transvaginal BWT	Data were analysed using the Bland–Altman method and mean differences and CIs within and between readers presented. SDs of differences between pairs of readers were 1.1 mm, 1.7 mm and 1.8 mm	Not possible to ascertain the analytical variability, the smallest real difference or the ICC from the data presented. An approximate (under) estimate of the analytical variability can be obtained by dividing the SD of differences by the square root of 2, i.e. 0.8 mm to 1.3 mm

a Paper states that the CI is computed as two SD either side of the mean, not two standard error.

Appendix 9 Economic evaluation decision trees

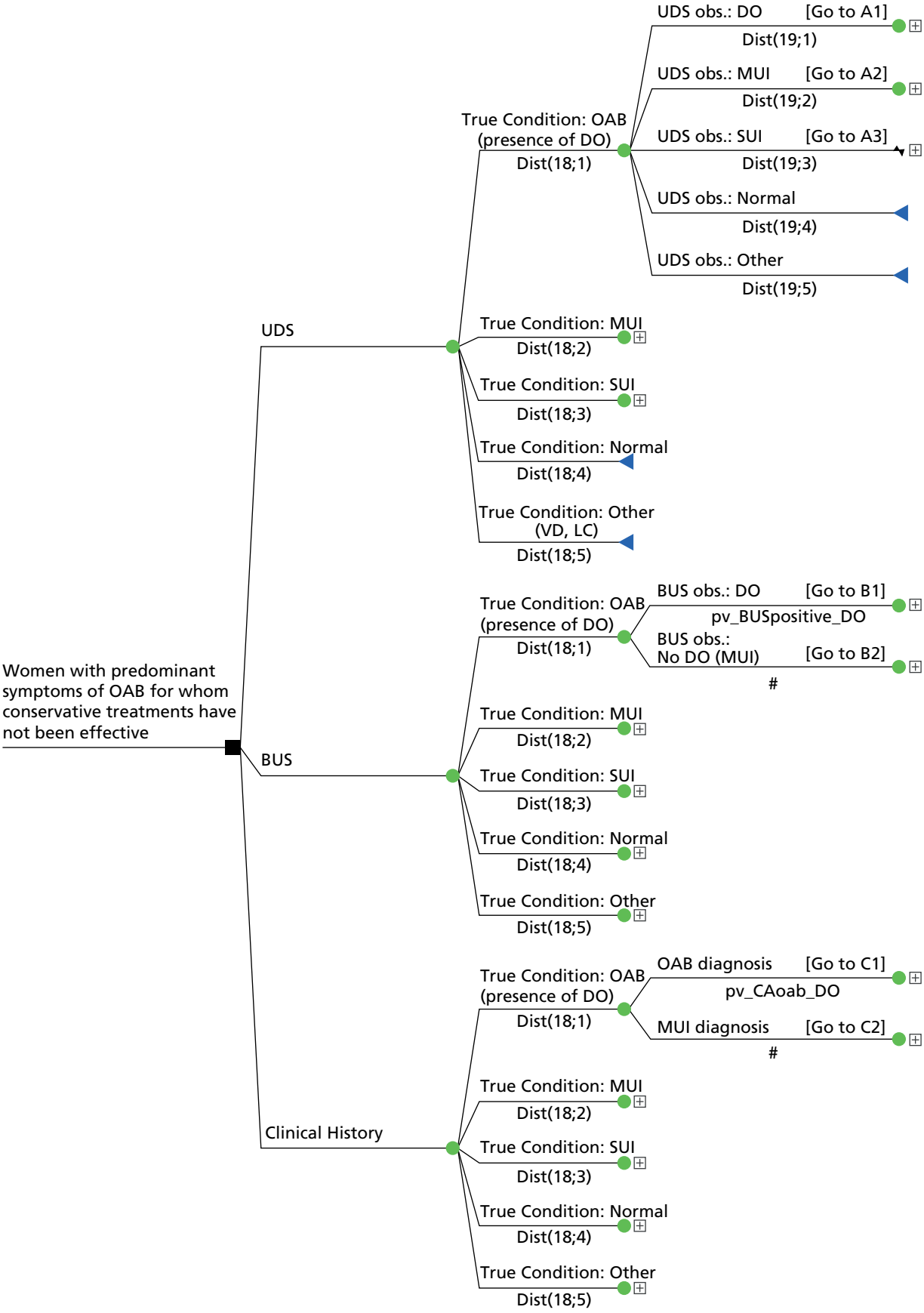

FIGURE 43 Basic structure of the decision tree (primary analysis). BUS, bladder ultrasonography.

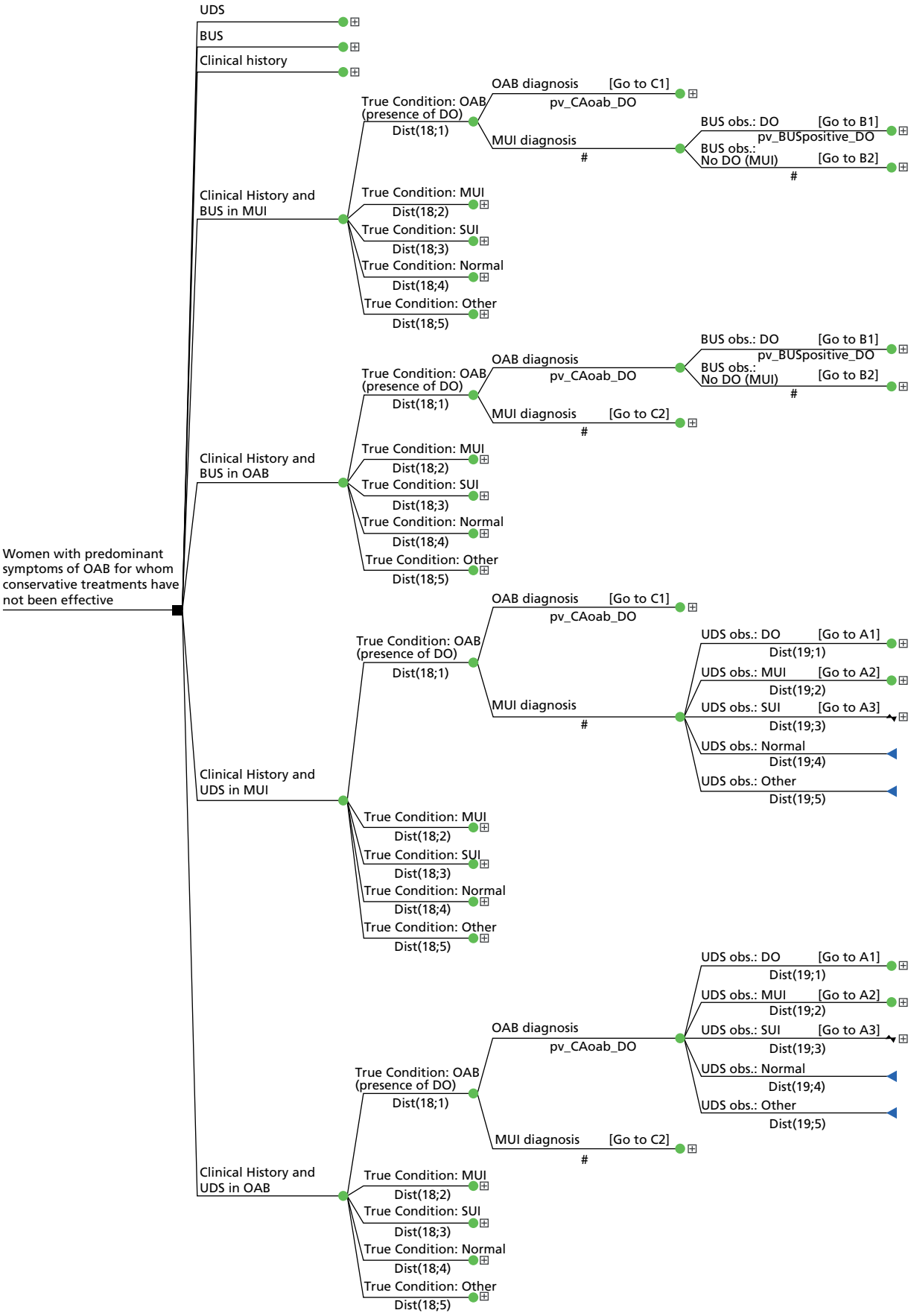

FIGURE 44 Basic structure of the decision tree (secondary analysis). BUS, bladder ultrasonography.

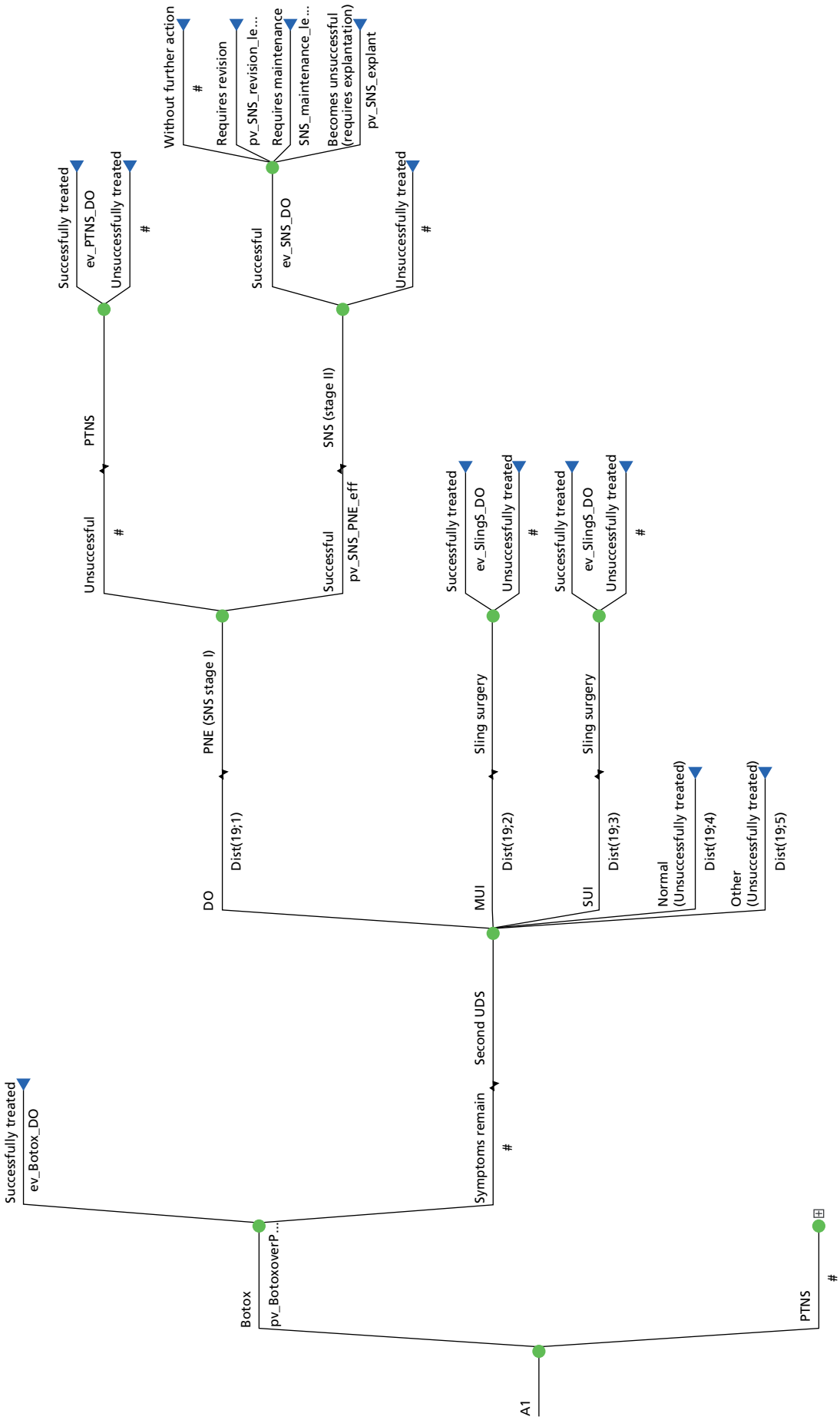

FIGURE 45 Treatment pathway for women with a diagnosis of DO in the UDS strategy.

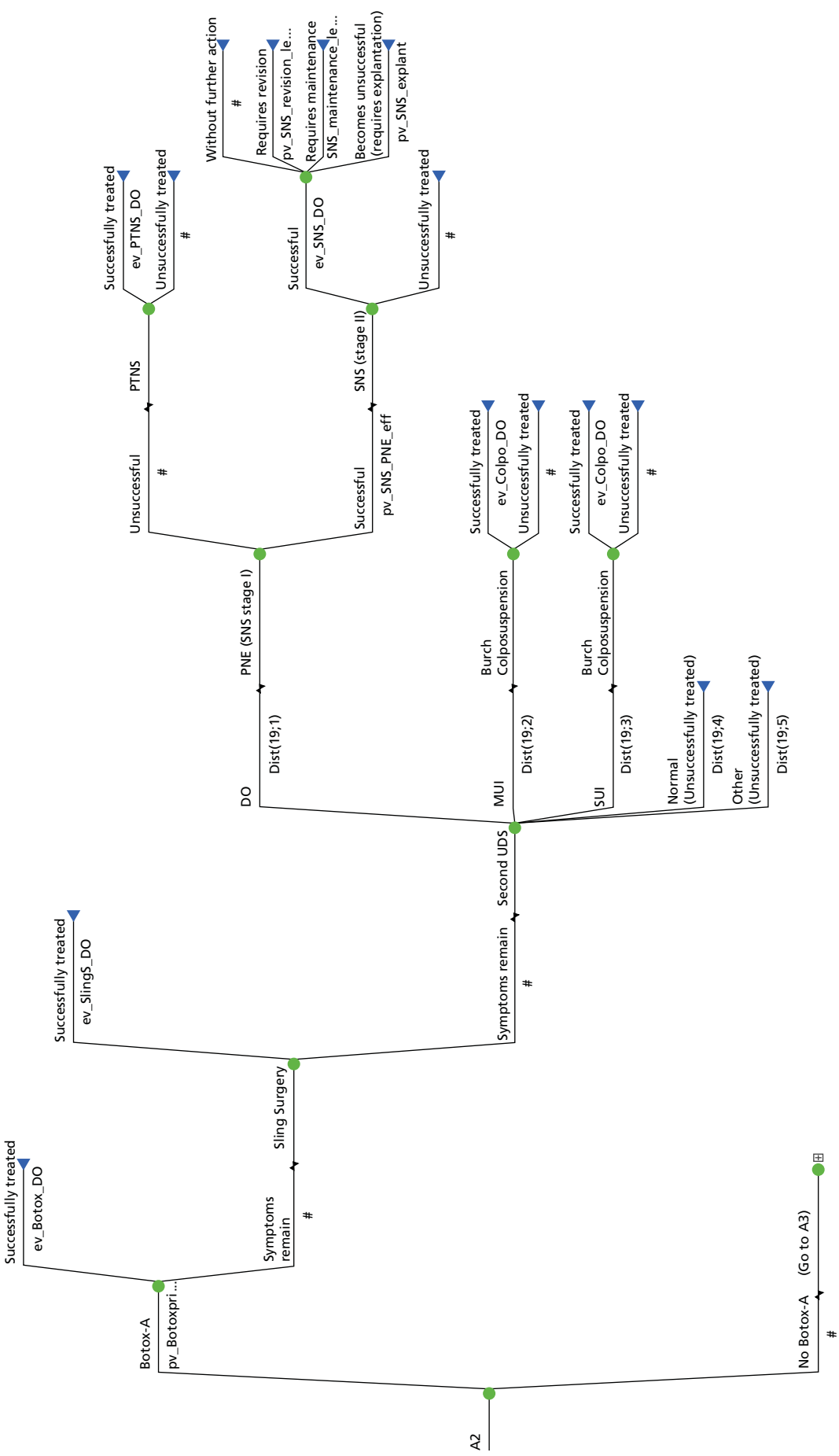

FIGURE 46 Treatment pathway for women with a diagnosis of MUI in the UDS strategy.

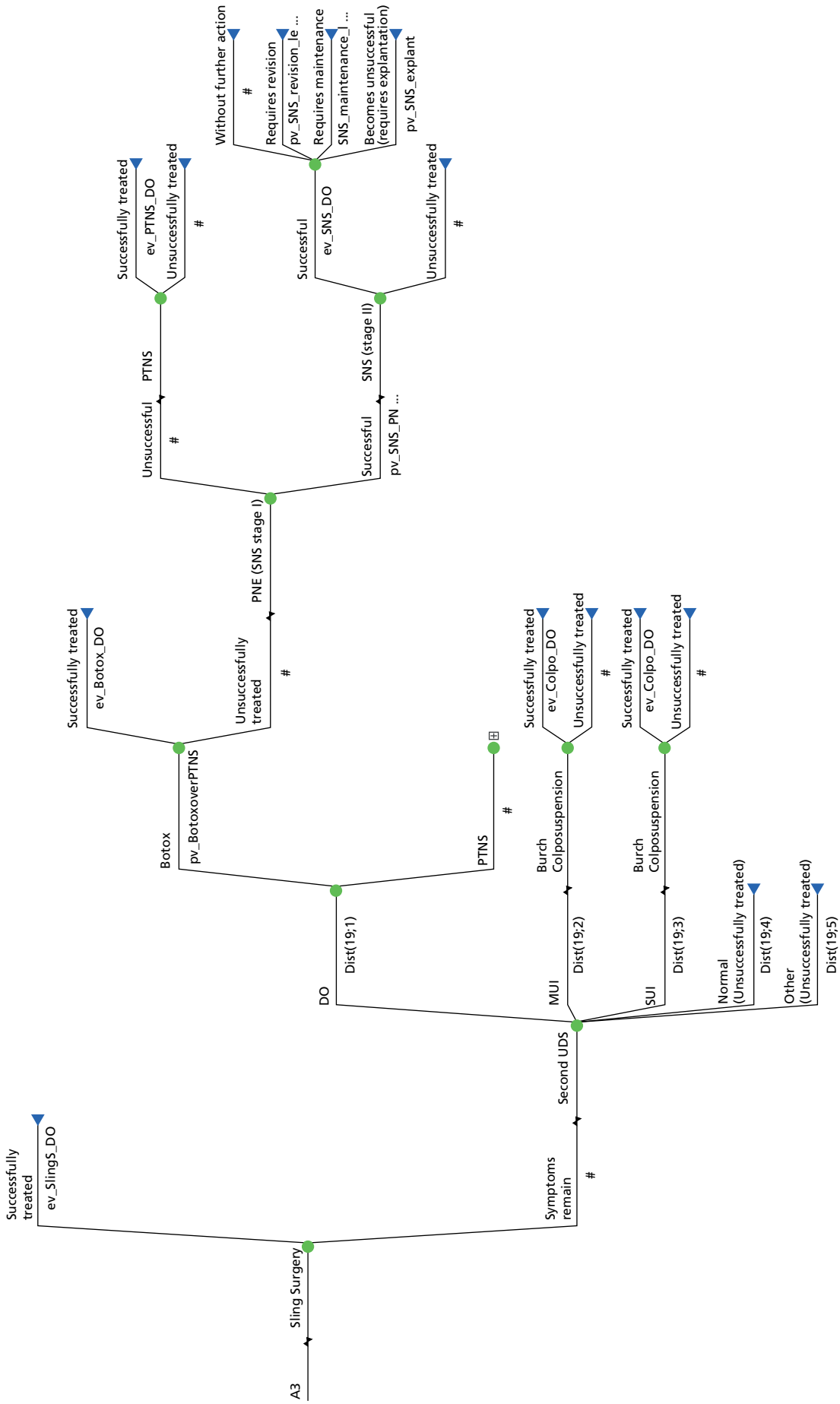

FIGURE 47 Treatment pathway for women with a diagnosis of SUI in the UDS strategy.

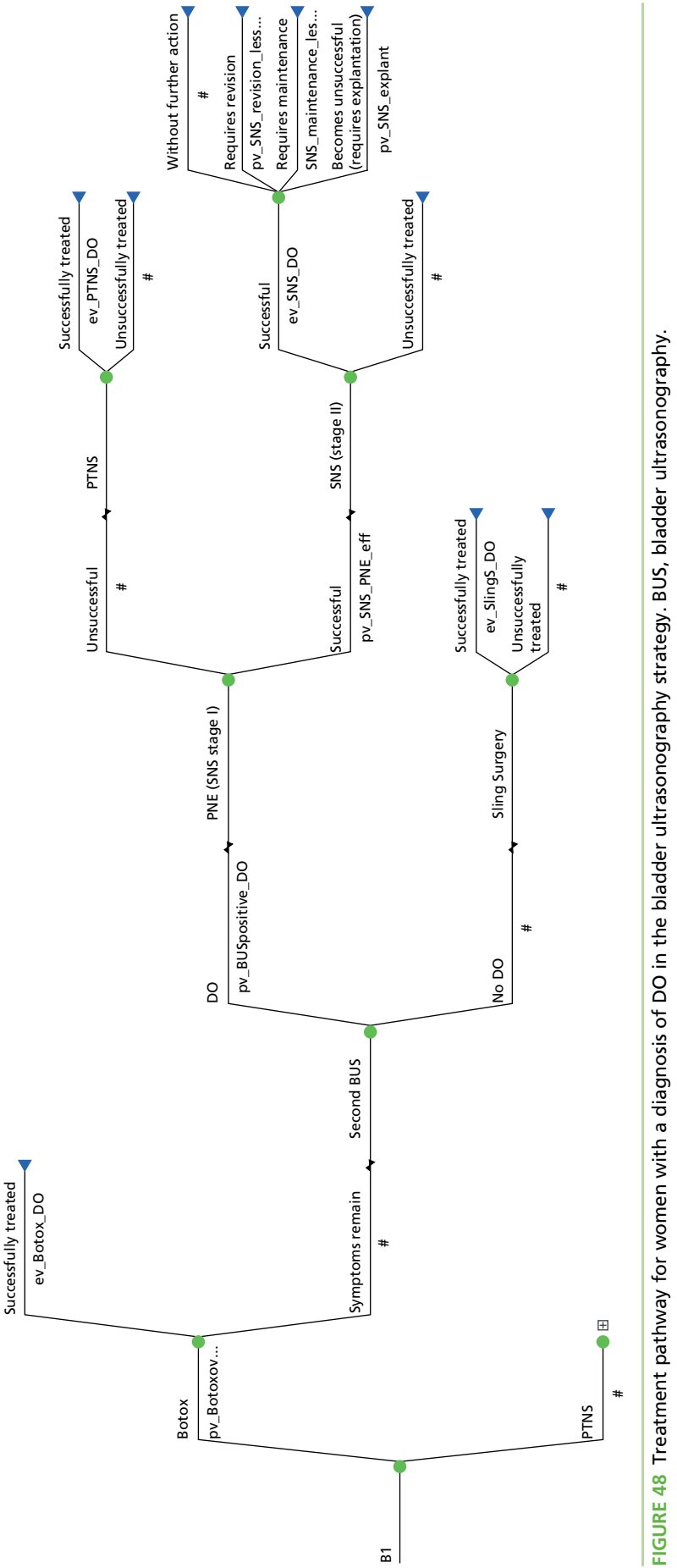

FIGURE 48 Treatment pathway for women with a diagnosis of DO in the bladder ultrasonography strategy. BUS, bladder ultrasonography.

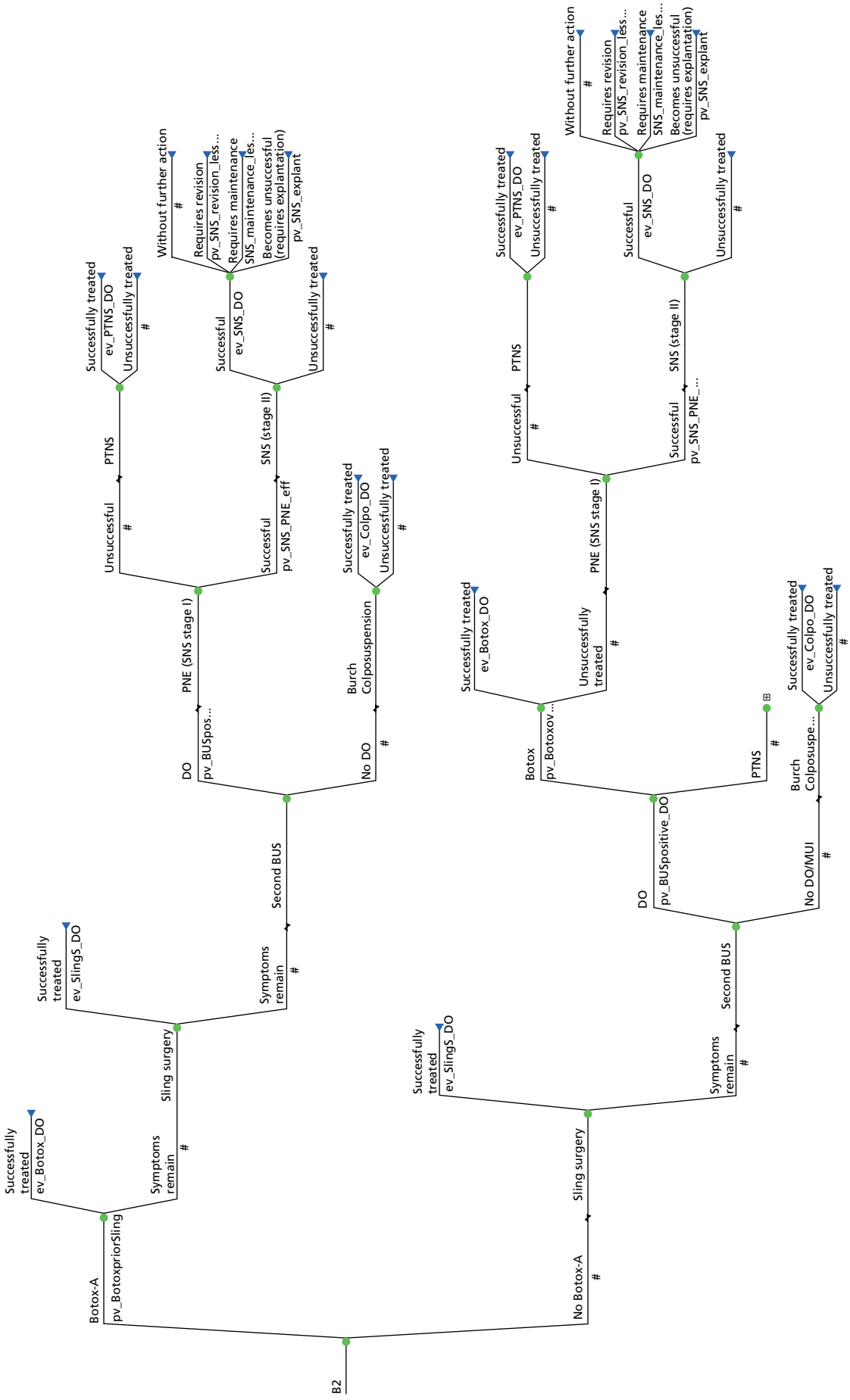

FIGURE 49 Treatment pathway for women with a diagnosis of no DO (MUI) in the bladder ultrasonography strategy. BUS, bladder ultrasonography.

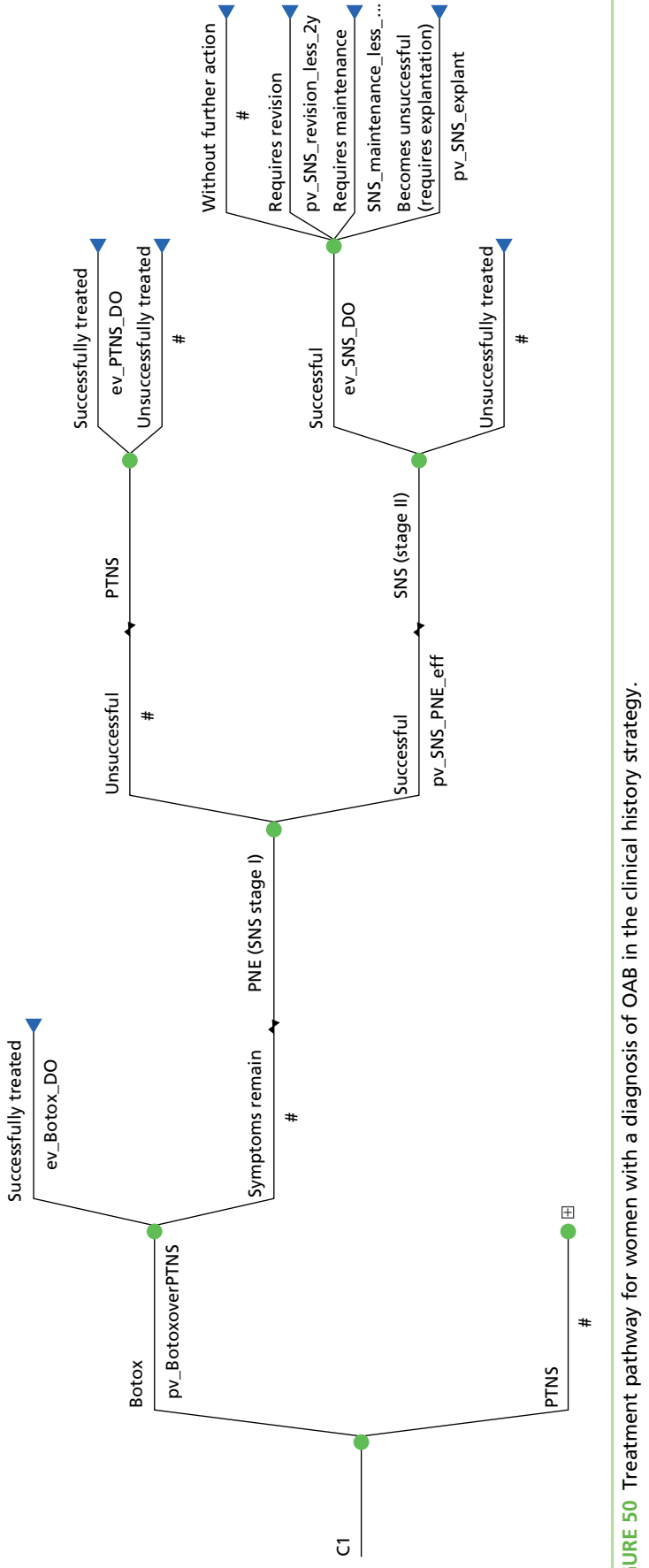

FIGURE 50 Treatment pathway for women with a diagnosis of OAB in the clinical history strategy.

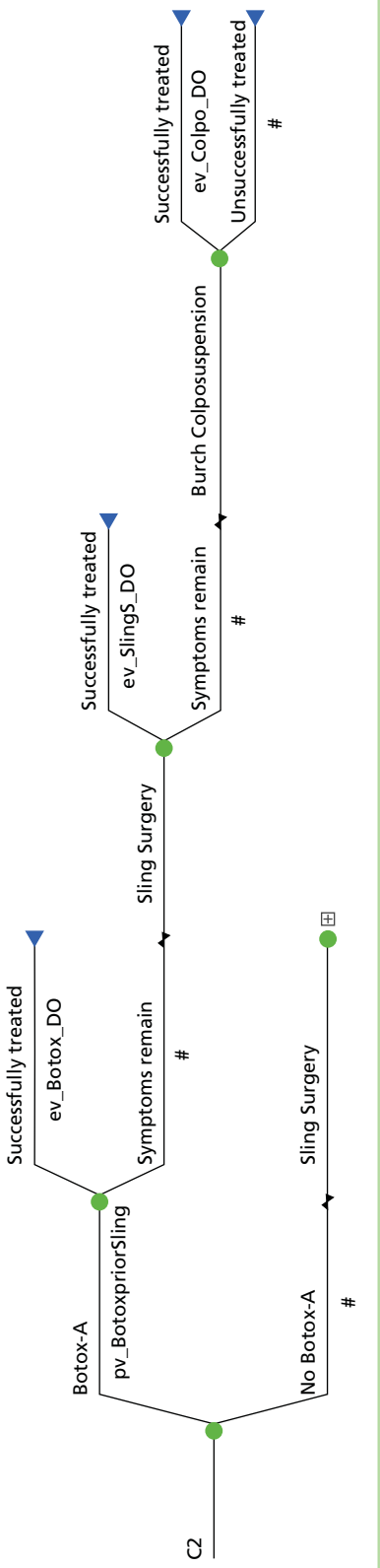

FIGURE 51 Treatment pathway for women with a diagnosis of MUI in the clinical history strategy.

Appendix 10 Incremental cost-effectiveness scatterplots and cost-effectiveness acceptability figures

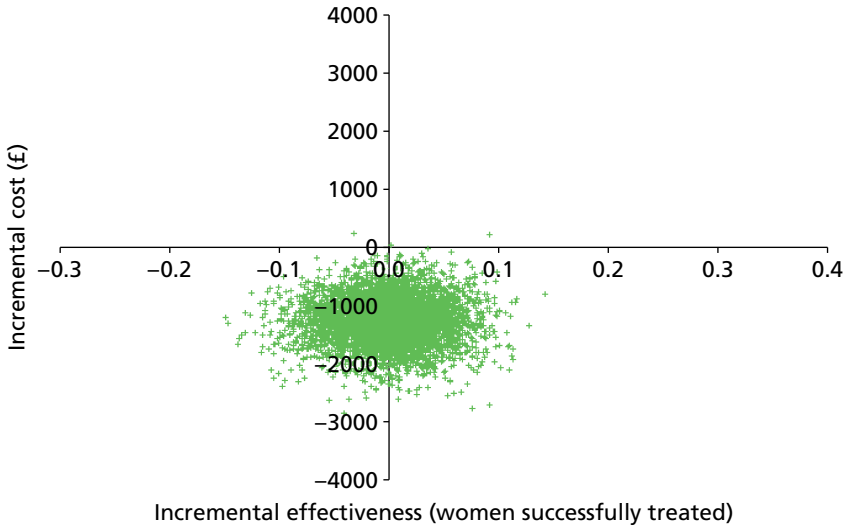

FIGURE 52 Incremental cost-effectiveness scatterplot of UDS vs. clinical history for women successfully treated.

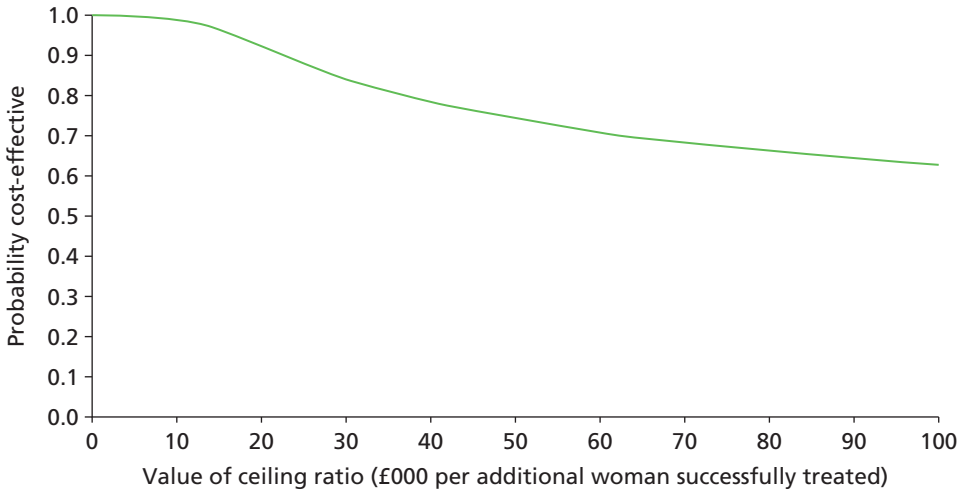

FIGURE 53 Cost-effectiveness acceptability frontier for the comparison between UDS and clinical history for the case of women successfully treated.

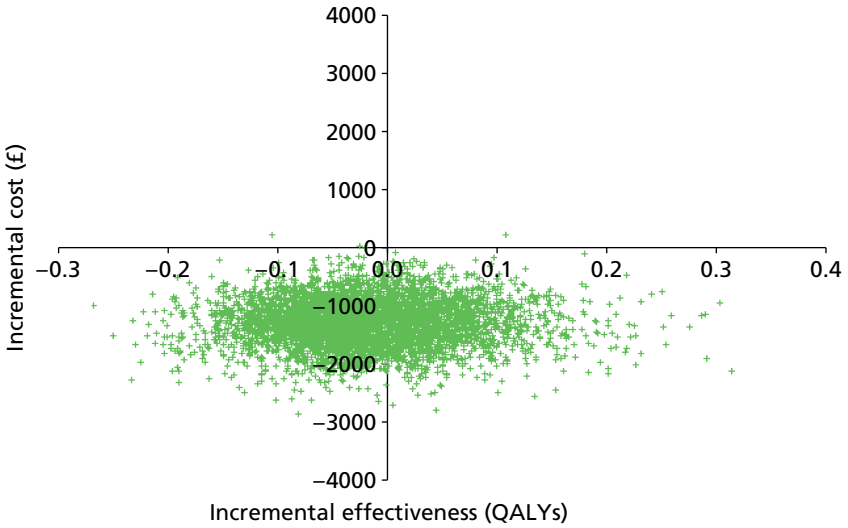

FIGURE 54 Incremental cost-effectiveness scatterplot of UDS vs. clinical history for QALYs.

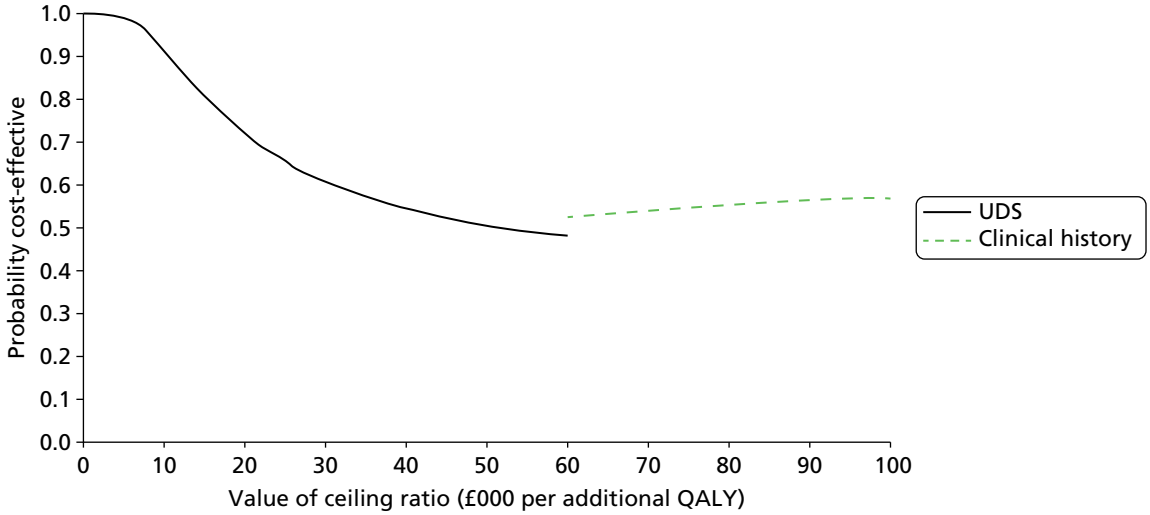

FIGURE 55 Cost-effectiveness acceptability frontier for the comparison between UDS and clinical history for the case of QALYs.

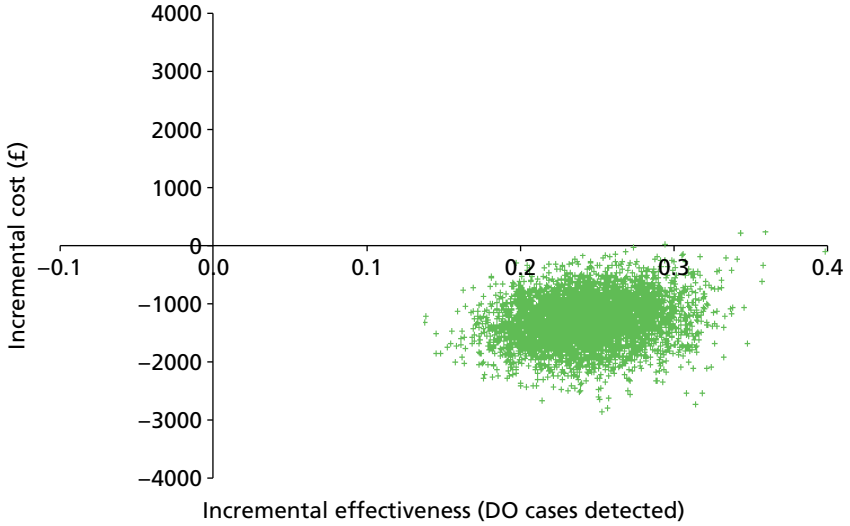

FIGURE 56 Incremental cost-effectiveness scatterplot of UDS vs. clinical history for DO cases detected.

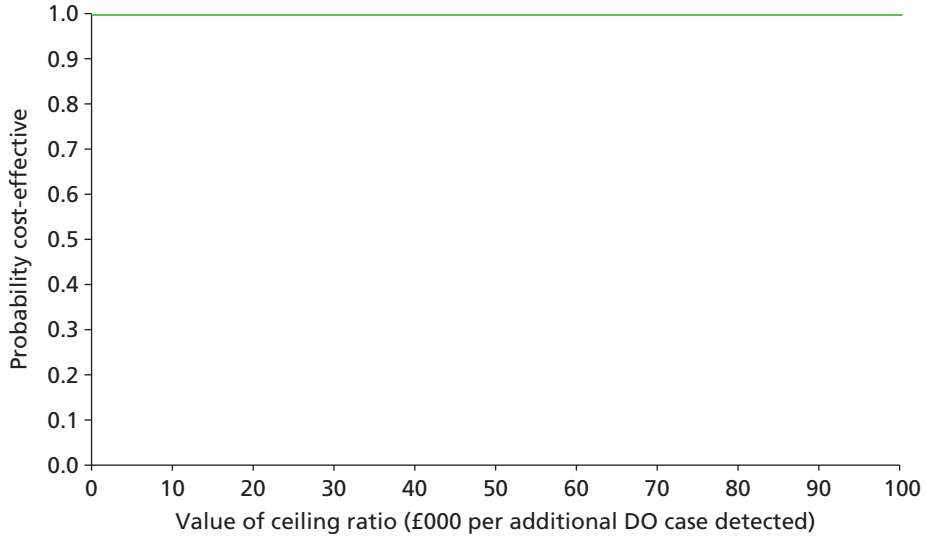

FIGURE 57 Cost-effectiveness acceptability frontier for the comparison between UDS and clinical history for the case of DO cases detected.

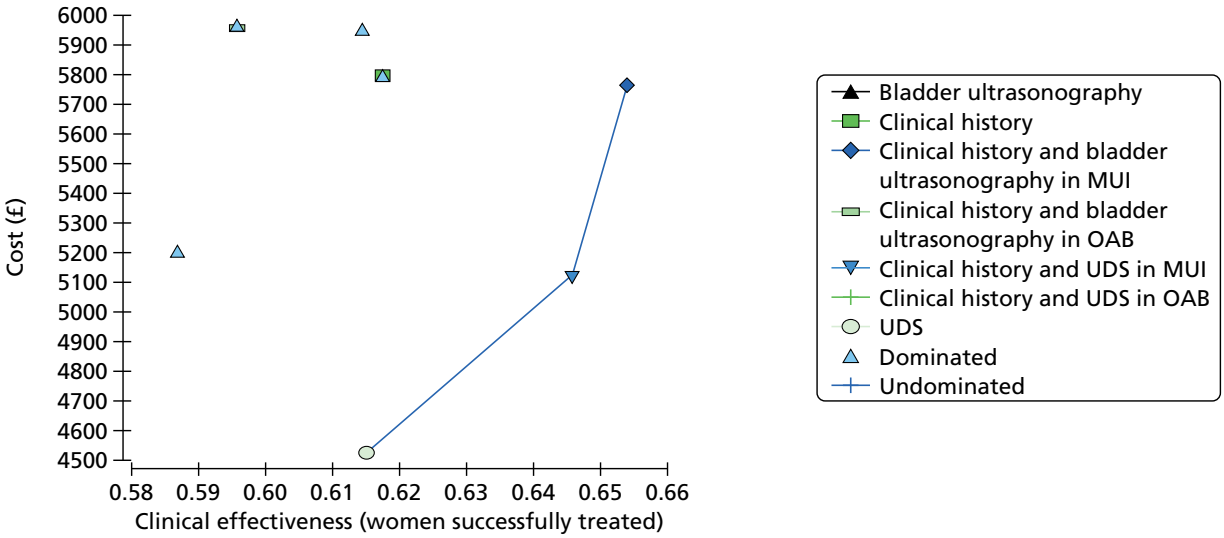

FIGURE 58 Cost-effectiveness plane showing the mean cost and clinical effectiveness (women successfully treated) for all strategies.

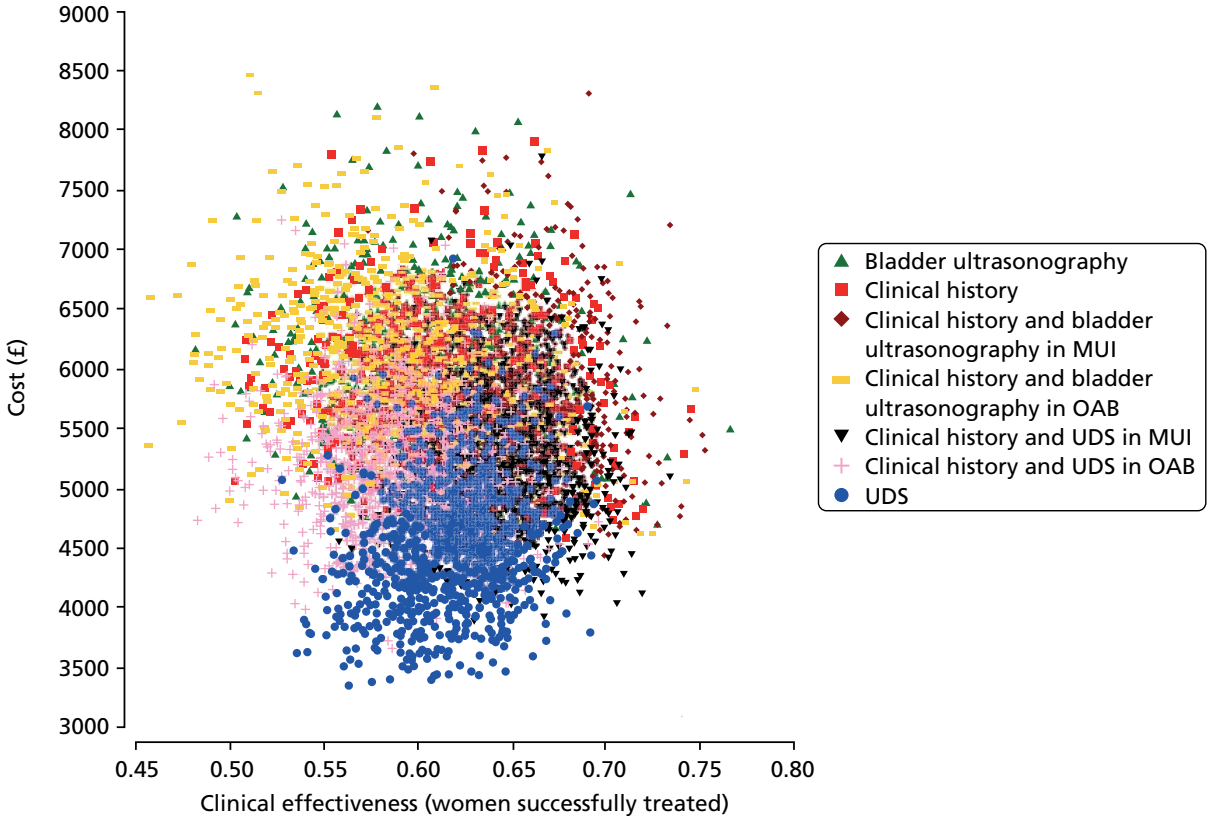

FIGURE 59 Scatterplot showing the uncertainty in costs and effectiveness (women successfully treated) for all strategies.

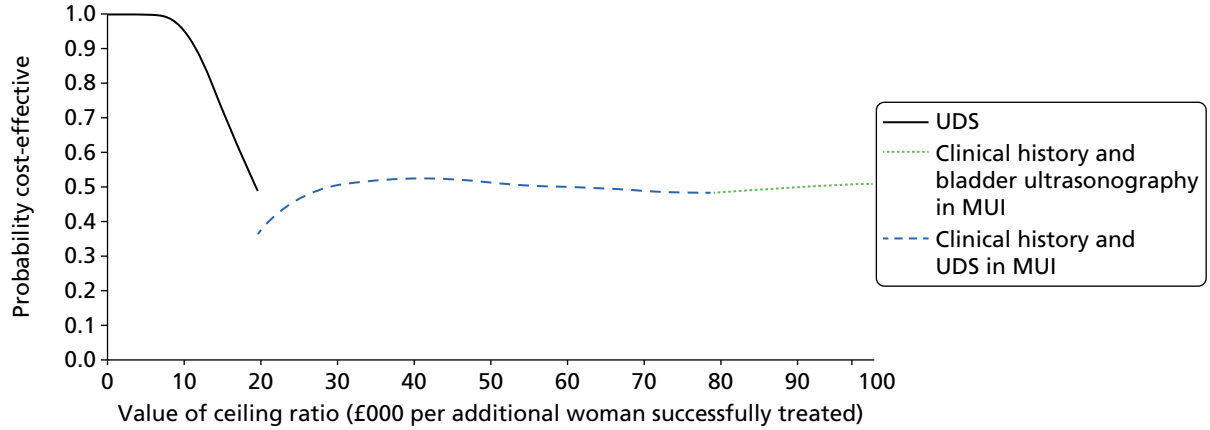

FIGURE 60 Cost-effectiveness acceptability frontier for the comparison between UDS, clinical history and bladder ultrasonography in MUI and clinical history and UDS in MUI for the case of women successfully treated.

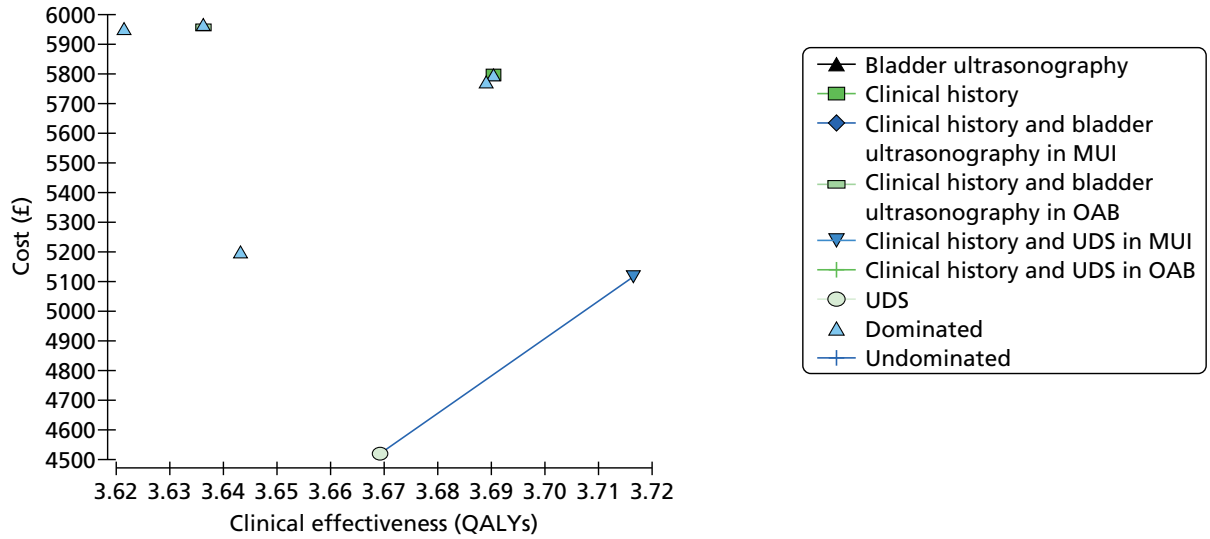

FIGURE 61 Cost-effectiveness plane showing the mean cost and clinical effectiveness (QALYs) for all strategies.

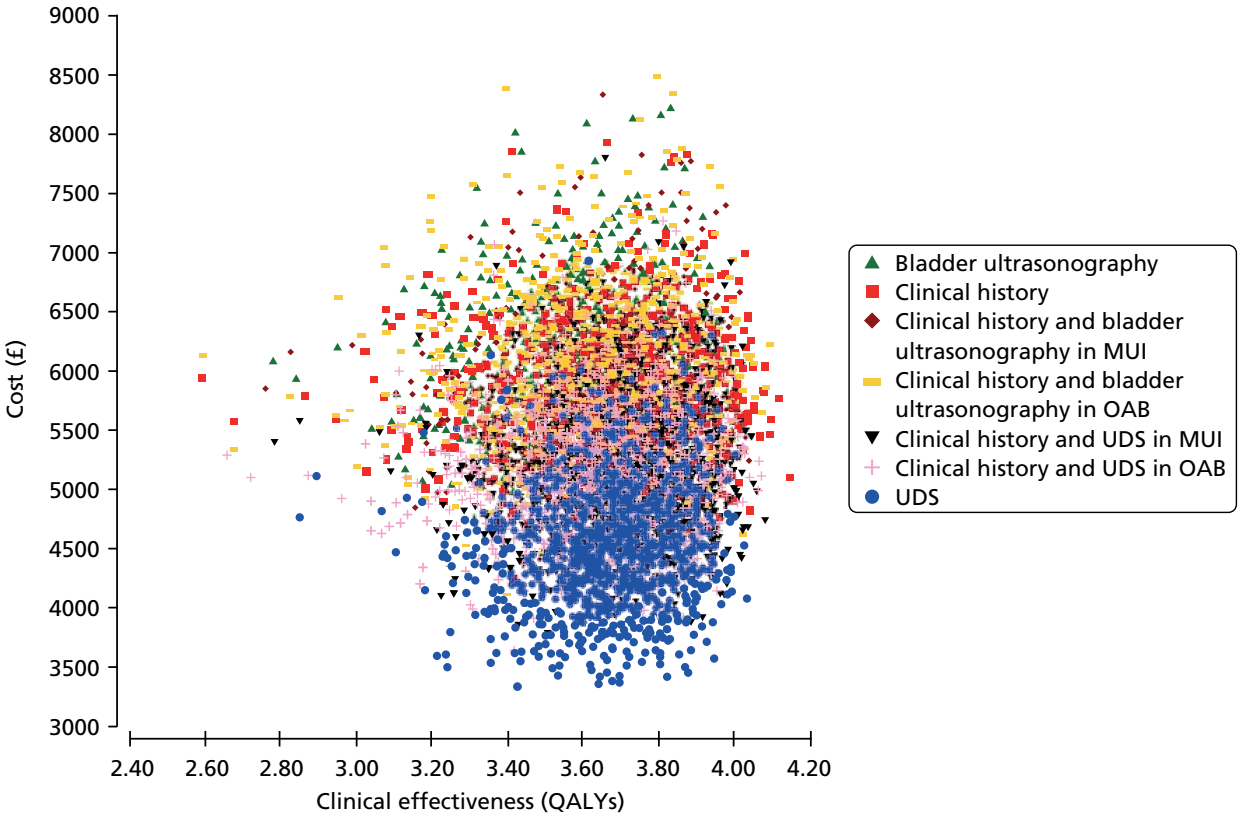

FIGURE 62 Scatterplot showing the uncertainty in costs and effectiveness (QALYs) for all strategies.

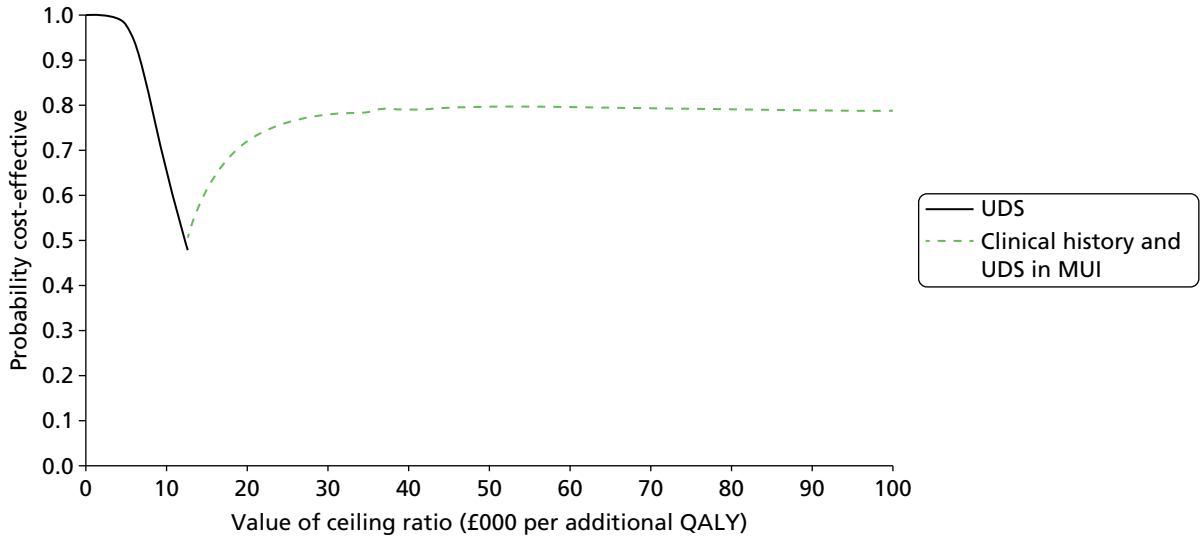

FIGURE 63 Cost-effectiveness acceptability frontier for the comparison between UDS, clinical history and bladder ultrasonography in MUI and clinical history and UDS in MUI for the case of QALYs.

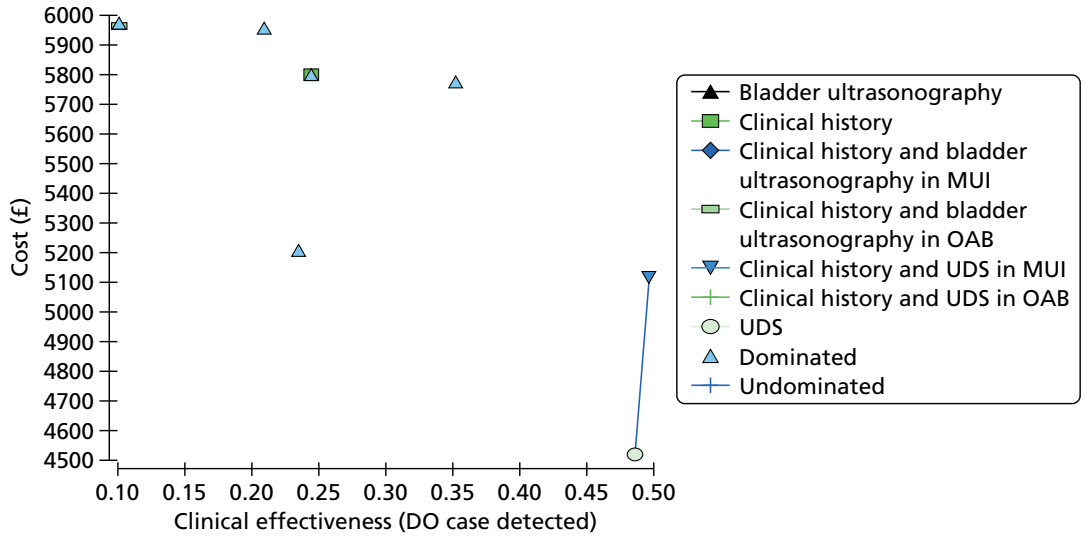

FIGURE 64 Cost-effectiveness plane showing the mean cost and clinical effectiveness (DO cases detected) for all strategies.

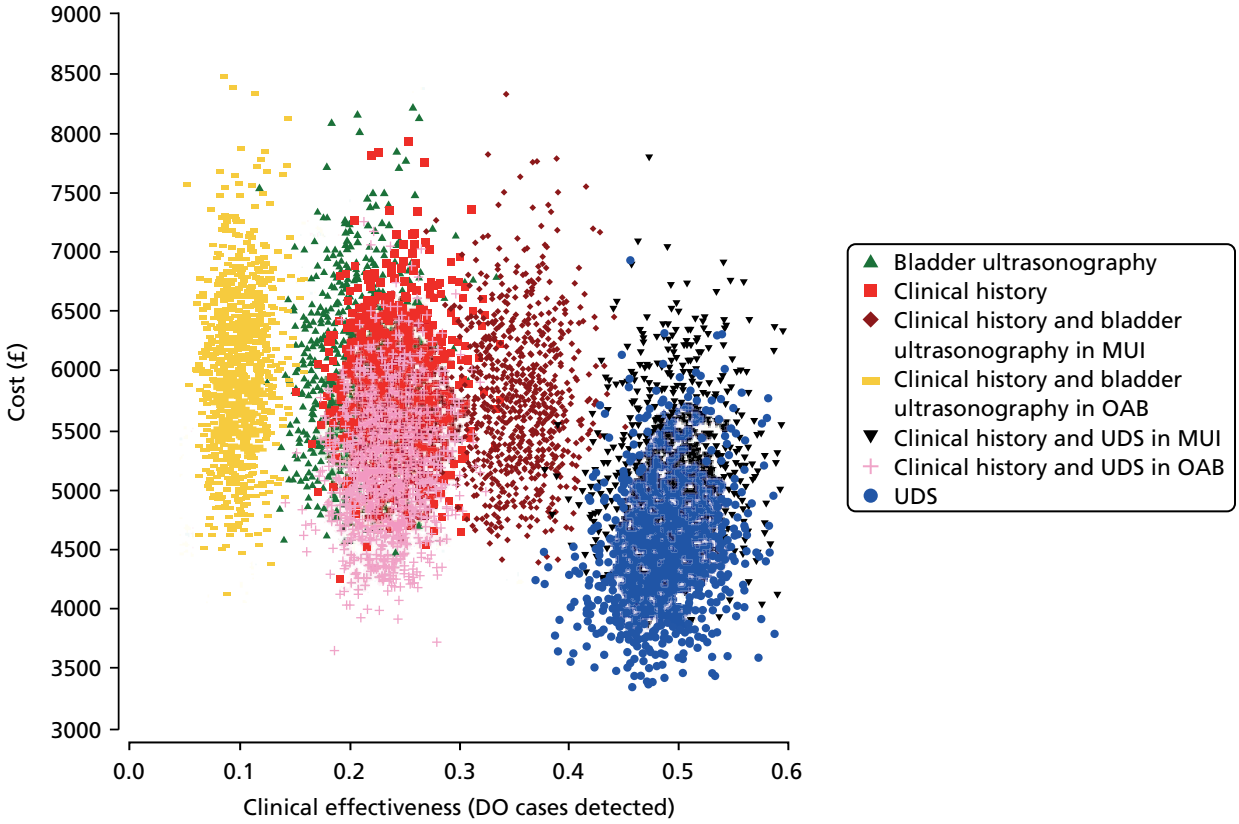

FIGURE 65 Scatterplot showing the uncertainty in costs and clinical effectiveness (DO cases detected) for all strategies.

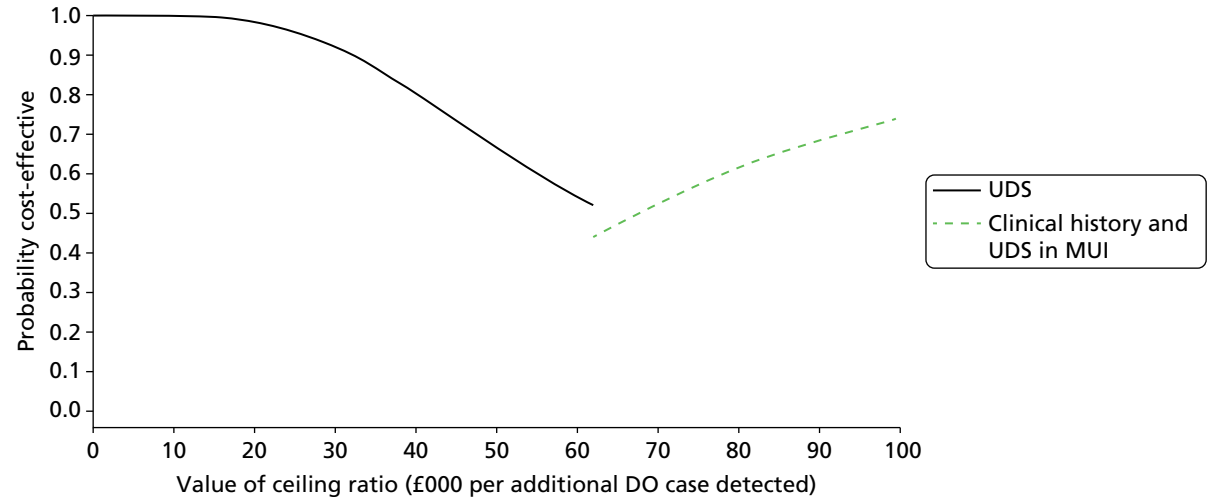

FIGURE 66 Cost-effectiveness acceptability frontier for the comparison between UDS, clinical history and bladder ultrasonography in MUI and clinical history and UDS in MUI for the case of DO cases detected.

FIGURE 67 Population EVPI for the main outcome measure (primary analysis).

FIGURE 68 Population EVPI for the main outcome measure (secondary analysis).

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library