Ultrasound bladder wall thickness and detrusor overactivity

Latthe, Pallavi; Middleton, Lee; Rachaneni, Suneetha; Mccooty, S.; Daniels, Jane; Coomarasamy, Aravinthan; Balogun, M.; Duckett, J.; Thakar, R.; Goranitis, Ilias; Roberts, Tracy; Deeks, Jonathan

DOI:
10.1111/1471-0528.14503

Document Version
Peer reviewed version

Citation for published version (Harvard):
Latthe, P, Middleton, L, Rachaneni, S, Mccooty, S, Daniels, J, Coomarasamy, A, Balogun, M, Duckett, J, Thakar, R, Goranitis, I, Roberts, T & Deeks, J 2017, 'Ultrasound bladder wall thickness and detrusor overactivity: A multicentre test accuracy study', BJOG: An International Journal of Obstetrics & Gynaecology, vol. 124, no. 9, pp. 1422–1429. https://doi.org/10.1111/1471-0528.14503

Ultrasound bladder wall thickness and Detrusor Overactivity: a multicentre test accuracy study

Pallavi Latthe, Senior Clinical Lecturer and Consultant Urogynaecologist[1,2], Lee Middleton, Statistician[3,] Suneetha Rachaneni, Clinical Research Fellow[1], Shanteela McCooty, Research Nurse[2], Jane Daniels, Senior Research Fellow[1,3], Arri Coomarasamy, Professor of Gynaecology[1,2], Moji Balogun, Consultant Radiologist[2], Jonathan Duckett, Consultant Urogynaecologist[4], Ranee Thakar, Consultant Urogynaecologist[5], Ilias Goranitis, Fellow in Health Economics[6], Tracy Roberts, Professor of Health Economics[6],' Jonathan Deeks, Professor of Biostatistics[3*]

On behalf of the BUS Collaborative Group

1. University of Birmingham, Institute of Metabolism and Systems Biology, Edgbaston, Birmingham, UK

2. Birmingham Women's NHS Foundation Trust, Birmingham, UK

3. University of Birmingham, Birmingham Clinical Trials Unit, UK

4. Medway NHS Foundation Trust, Medway Maritime Hospital, Gillingham, UK

5. Croydon Health Services NHS Trust, Croydon University Hospital, Croydon, UK

6. University of Birmingham, Health Economics Unit, Edgbaston, Birmingham, UK B15 2TT

*Corresponding author

Prof Jon Deeks

Birmingham Clinical Trials Unit (BCTU)

Institute of Applied Health Research

University of Birmingham

Birmingham B15 2TT UK

Phone: + 44 (0) 121-414-5100

Email: j.deeks@bham.ac.uk

Tweetable abstract

Bladder wall thickness is not a good replacement test for urodynamics in women with overactive bladder.

Running title

Ultrasound bladder wall thickness and detrusor overactivity

Abstract

Objective

Women with overactive bladder (OAB) often undergo urodynamics before invasive treatments are considered. Ultrasound measurement of bladder wall thickness (BWT) is a less invasive, less expensive and widely available test. It has the potential to diagnose the presence of Detrusor Overactivity (DO). We aimed to evaluate the accuracy of BWT in diagnosis of DO.

Design

Prospective cohort study

Setting

22 UK clinics (university and district general hospitals)

Methods

Consecutive eligible women with OAB symptoms had transvaginal ultrasound to estimate BWT (index test). The reference standard for the diagnosis of DO was urodynamic testing with multichannel subtracted cystometry.

Main outcome measures

The sensitivity, specificity and likelihood ratios using a BWT threshold of ≥5mm to indicate the presence of DO, and the area under the receiver operating characteristics (ROC) curve to give an overall estimate of BWT accuracy.

Results

Between March 2011-2013, 644/ 687 (94%) women recruited had both tests. The mean age was 52.7 years (standard deviation 13.9) and DO was diagnosed in 399/666 (60%) of women. BWT had a sensitivity of 43% (95% confidence interval (CI): 38-48%), specificity of 62% (95% CI: 55-68%), and likelihood ratios of 1.11 (95% CI: 0.92-1.35) and 0.93(95% CI: 0.82-1.06) for positive and negative tests respectively. The area under the ROC curve was 0.53 (95% CI: 0.48-0.57). Extensive sensitivity analyses and subgroup analyses were carried out, but did not alter the interpretation.

Conclusions

BWT is not a good replacement test for urodynamics in women with overactive bladder.

Keywords

Overactive bladder; bladder wall; ultrasound; urodynamics; diagnostic accuracy

Study Registration

The study protocol was registered as ISRCTN:46820623.

http://www.isrctn.com/ISRCTN46820623

Introduction

Overactive bladder (OAB) is a debilitating condition that affects 12% of world population and increases with age.[1] It is defined as a symptom complex of urinary urgency with or without incontinence, usually with increased urinary frequency, or nocturia, but in the absence of infection or other proven pathology.[2] It is associated with a considerable economic burden from both a societal and patient perspective.[3]

At present, invasive urodynamics is the gold standard test for assessment of OAB unresponsive to conservative management.[4] It is an intimate and invasive test, with urinary tract infection (UTI) rates following urodynamics reported to be between 3 to 20%.[5] A common pathology underlying OAB is detrusor overactivity (DO), observed in 54-58% of women with symptoms.[6] At present, National Institute of Health and Care Excellence (NICE) guidance recommends treatments such as Botulinum toxin A or neurostimulation only in patients diagnosed with DO on urodynamics.[JPD1][7]

Bladder wall thickness (BWT) measured by ultrasonography has been proposed as a less invasive alternative to urodynamics to identify DO. A systematic review of BWT noted that all existing studies were small and of variable quality.[8] The reported sensitivity varied between 40-84%, and specificity between 78-89%. The need for further evidence was identified as a priority in a NICE guideline[7] and a Health Technology Assessment report.[9]

We report results of a large prospective, multicentre test accuracy study undertaken to evaluate whether BWT measured by ultrasonography can accurately diagnose DO in

women with OAB symptoms. The assessment of reproducibility of the test and the cost-effectiveness analysis are reported elsewhere.[10]

Methods

Participants

Consecutive women scheduled for investigation of OAB symptoms were prospectively approached for recruitment and consent (Web appendix Figure S1). Women were asked to complete a bladder diary for the three days preceding their test appointment.

Women were eligible for inclusion if they provided written informed consent and satisfied all the following criteria: 1) urinary frequency of 9 voids or more in a 24-hour period for at least one day in a 3 day bladder diary; 2) mild to severe urgency recorded on at least two occasions in the bladder diary and 3) post void residual (PVR) volume <100 ml.

Exclusion criteria were: 1) symptoms of pure stress urinary incontinence (SUI) or stress predominant mixed incontinence; 2) current pregnancy or up to six weeks postpartum; 3) SUI surgery and/or intradetrusor Botulinum toxin A in the past six months; 4) urine dipstick positive for leucocytes or nitrites; 5) Greater than stage II (any compartment) on Pelvic organ prolapse quantification (POP-Q) system ; 6) previous urodynamics in the past six months; or 7) current or previous use of antimuscarinics for >6 months continuously.

The clinical history of the participants was collected prior to the tests and included

previous treatments for bladder problems and the International Consultation on Incontinence Questionnaire-Overactive bladder (ICIQ-OAB) questionnaire.[11].

Procedures

Both the bladder ultrasound and urodynamics were done within four weeks of each other by different practitioners. The practitioner undertaking the urodynamics was unaware of the result (blinded) of the ultrasononography.

Index test - bladder wall thickness on ultrasonography

A standard operating procedure (SOP) for carrying out the transvaginal ultrasound was produced and clinicians were required to attend training provided by study team.

The PVR was measured by the following formula: length x width x depth in cm x 0.5223 in millilitres. The bladder wall was measured with a 7-9 MHz end-firing transvaginal probe in the sagittal plane introduced 1 cm beyond the vaginal introitus in the midline. The BWT was measured at three sites perpendicular to the luminal surface of the bladder (Web appendix Figure S2): the thickest part of the trigone, dome in the midline and the anterior wall. BWT was calculated as the mean of these 3 measurements in millimetres.

Reference standard – Urodynamics

For urodynamics, we developed SOP based on the Good Urodynamic Practice Guidelines of the International Continence Society.[12] Women attended the clinic with a full bladder for the uroflowmetry in privacy. Filling cystometry was performed with the woman in sitting position at the rate of 100mls/minute, followed by provocation manoeuvres and then voiding cystometry.

Detrusor Overactivity was detected when involuntary detrusor contractions were seen during filling cystometry. These contractions (spontaneous or provoked), could be of variable duration and amplitude, phasic or terminal, with or without urgency and/or urgency incontinence. Voiding dysfunction was defined as abnormally slow (<15mls/sec) flow and/or incomplete micturition (PVR >100 mls).

Quality assurance

The lead investigator assessed the competency of local investigators by reviewing at least five ultrasound scans prior to allowing them to enter patients into the study. A reproducibility assessment done for intra and interobserver variation is reported elsewhere.[10]

The quality of urodynamics was audited with anonymised traces every six months and comparing them to the interpretation guidelines[2] to ensure ongoing quality assurance.[13]

Statistical methods

Sample size

A minimum target sample size of 600 participants was pre-specified in order to obtain estimates of sensitivities and specificities with 95% confidence intervals (CI) of width 10% or less, anticipating sensitivity and specificity values between 70% and 95%. The computation was based on a prevalence of 50% for DO[4], providing 300 women each for the estimate of sensitivity and specificity.

Data analysis

Sensitivity, specificity and predictive values were calculated using a BWT ≥5mm indicating presence of DO as a pre-specified cut-off based on the evidence from previous studies. Likelihood ratios for categories of BWT were also pre-specified: <3mm; ≥3mm to <5mm and ≥5mm. CIs were calculated using binomial exact methods. A receiver operating characteristic (ROC) curve was constructed and the area under the curve (AUC) was 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. where AUC=0.5) using a non-parametric approach.[14] BWT measurements in groups with and without DO were compared using a two-sample t-test. The relationship of BWT with pre-test ICIQ-OAB score was tested using simple linear regression. Statistical significance was defined as $p<0.05$.

Sensitivity analyses were performed on the primary population to test the robustness of the results to protocol deviations (excluding those where tests were not blinded or undertaken >4 weeks apart), missing data (incorporating measurements where not all three components of the BWT were available) and more stringent inclusion criteria (excluding those with mixed stress/urge incontinence; and excluding those with post void residual urine volume >30ml). We also investigated the impact of using different BWT measures (trigone alone; mean of dome with measures 1cm to the left and right of the dome).

The accuracy of BWT was compared between subgroups according to 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 and 6) BMI (<25, ≥ 25). ROC curves were generated for each subgroup and the AUC compared using a large sample chi-squared test for independent curves.[15]

Exploratory analyses were undertaken to identify variables that predicted a diagnosis of DO using multivariable logistic regression. The variables considered were BWT measurement, ICIQ-OAB scores, age, duration of symptoms, ethnicity, vaginal birth, menopausal status, parity, previous POP surgery and the subgroups listed above.

Results

Participants

Six hundred and eighty seven women who were eligible and consented to participate were recruited from 22 centres between March 2011 - 2013. The study over-recruited to compensate for study withdrawals and women without complete index and reference standard test results (Figure 1).

The mean age of women was 52.7 years (standard deviation [SD]: 13.9) and the average BMI was 30.6 (SD: 12.2) (Table 1). Of the 687 women, 387 (55%) were postmenopausal. According to clinical history, 419 (61%) had urgency-predominant mixed incontinence and 226 (33%) reported only urinary urgency and frequency. The median duration of symptoms was 3.0 years (Inter Quartile Range: 1.6 to 7.0).

Test completion

Complete urodynamic diagnoses were obtained in 666/687(97%) of women (Figure 1).

Of these, 399 (60%) were diagnosed with DO (95% CI: 56-64%) (Web appendix Table S1) and 245 (61%) were given a further sub-diagnosis of DO incontinence (defined as detrusor pressure rise and leak).

All three BWT measurements (trigone, dome midline, anterior wall midline) were available in 645 (94%) women. The average BWT measurement was 4.78mm (SD 1.34) (Web appendix Table S2).

Of the 644 participants who had the two tests (Figure 1), both were performed on the same day in 439 (68%); and only 26 (4%) were performed more than four weeks apart. Ninety-seven percent of reference tests (616/632, twelve 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 at six month follow-up reported having urine infection within two weeks of the test. Of these, 36/48 (one observation missing) were diagnosed by a General Practitioner.

Estimates of test accuracy

BWT showed poor sensitivity (43%; 95% CI: 38-48), specificity (62%; 95% CI: 55-68), positive (63%; 95% CI: 57-69) and negative (41%; 95% CI: 36-47) predictive values for diagnosis of DO (Table 2).

Likelihood ratios were non-discriminatory at all pre-specified cut-offs of ≥ 5mm (1.11; 95% CI: 0.92-1.35), ≥3mm <5mm (0.96; 95% CI: 0.83-1.13) and <3mm (0.76: 95% CI: 0.46, 1.26) (Web appendix Table S3).

The ROC curve (Figure 2) showed no evidence of discrimination at any threshold

between those with and without DO (AUC 0.53; 95% CI: 0.48- 0.57; p=0.25). There was no evidence that the mean BWT measurements were any higher in the DO positive group than in the DO negative group (mean (SD): 4.85(1.36) mm versus 4.70(1.29) mm; p=0.19) (Web appendix Figure S3); or that BWT had any relationship with pre-test ICIQ-OAB symptoms score (*r*=-0.01; p=0.88). The AUC remained below 0.55 in all sensitivity analyses and in all pre-specified subgroups (Web appendix Figures S4-S13 and Table S4).

In the multivariable exploration of factors possibly associated with DO diagnosis, only higher baseline ICIQ score (worse symptoms) was associated with DO (OR: 1.21; 95% CI: 1.13- 1.29; p<0.0001), i.e. the odds of DO diagnosis were increased by 21% for every point increase in ICIQ score (Web appendix Tables S5-S6).

Discussion

Summary of Main Findings

Bladder wall ultrasonography appeared to be no better than chance at making the diagnosis of DO, with an AUC of 0.53 (95% CI: 0.48-0.57). 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. Based on this evidence, we conclude that BWT is not a useful test in diagnosing DO and should not be used in clinical practice.

Strengths and Limitations

The BUS study is the biggest test accuracy study on the subject; it was designed to minimize bias and ensure that the results would be applicable to imaging services available in routine clinical practice. We attempted to recruit consecutive women fulfilling the eligibility criteria. Women were of mixed ages, ethnicities and were recruited from multiple centres across the UK. The prevalence of DO in our study was 60%, which was similar to other studies.[6]

As transvaginal bladder ultrasound is a relatively new technique, concerns may be raised on the quality of scan measurements. However, the technique has been easy to teach and learn (personal experience of the authors) with the urinary bladder being an anterior and relatively superficial midline structure and previously has been reported to be reproducible .[16] Both ultrasononography and urodynamics were undertaken in 94% of women, and blinding of test results ensured for 97% of them. Risk of disease progression bias was minimized by conducting the tests within 4 weeks of each other in 96% of women. All analyses and cut-offs were pre-specified in the protocol. 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. We have undertaken multiple sensitivity analyses to investigate the impact of excluding women with variations in presentation such as urgency predominant mixed incontinence in the study, and in all cases these additional analyses have shown no discrimination.

One area of concern is misclassification made by the reference standard.[17] Urodynamics is known to be less than 100% reproducible in previous studies of patients with OAB and

healthy volunteers.[18] Errors in the reference standard typically lead to underestimation of sensitivity and specificity but misclassification rates would have to be extreme for no relationship to be observed at all. The poor accuracy for BWT elicited in our study is thus likely to mainly be explained by the fact that BWT bears no relation to DO. This is reinforced by no relationship being observed between BWT and disease severity measured by the ICIQ-OAB questionnaire, whereas there was a strong relationship between DO status from urodynamics and the ICIQ-OAB.

Interpretation in light of other evidence

The initial studies suggested a greater BWT to be an accurate diagnostic marker for DO. For a mean BWT cut-off of 5mm, the specificity was reported to be 89% (95% CI 79- 96) with a sensitivity of 84% (95% CI 76-90). [16] We identified further studies[19-30], which investigated the relationship between BWT measured by ultrasonography and DO, with estimates of sensitivity, specificity and AUC for bladder ultrasound varying from 37-91%, 61-97% and 0.61-0.91 respectively.[8]

The published studies are mostly from single centre (often tertiary) and with multiple reasons to have concerns about the validity and applicability of findings. Some studies added ambulatory urodynamics to the reference standard, if the patients had normal video-urodynamics, as a tie-breaker test.[16] Some studies[19, 23-26] made comparisons with healthy controls, some[28] excluded patients with mixed urinary incontinence and one enriched with women with equivocal urodynamics findings.[30] Altering the spectrum of patients from that encountered in practice influences the estimates of sensitivity and specificity[17]; exclusion of the mixed urinary incontinence cases and inclusion of healthy

controls will lead to overestimation of test accuracy; enrichment with difficult to diagnose cases will underestimate test accuracy. Some studies used transabdominal[23, 25, 27] or translabial ultrasound with higher interobserver variability.[26]

Our study focused on recruiting women with urgency or urgency predominant mixed incontinence undergoing urodynamics to identify DO. Some of the previous studies that compare pure SUI with OAB have assessed the value of ultrasonography to differentiate between SUI and DO: the higher observed accuracy may well reflect that BWT is higher in those with DO than with SUI,[21] but this is not of direct relevance to the clinical role that ultrasonography could play.

There is some emerging evidence in literature that the response to invasive therapies might be similar in patients with frequency and urgency with or without incontinence, regardless of the observation of DO. Everything seems to be more complicated when there is mixed incontinence involved, and urodynamics might have value there.[10] Robustly designed randomized controlled trials are required to evaluate whether patients are more likely to benefit from decisions to use invasive therapies based on urodynamic findings of DO versus just clinical evaluation alone.[10]

Conclusion

Bladder wall thickness ultrasonography cannot identify women with Detrusor Overactivity and hence cannot be used to reduce the need for invasive urodynamics.

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 Trial Steering Committee (TSC) and Data Monitoring Committee (DMC) for their assistance throughout the project: Professor D Tincello (Professor of Urogynaecology and Honorary Consultant Gynaecologist, University of Leicester) Chair of TSC; Jean Perks (Patient Representative, Birmingham); Dr Patrick Chein, (Consultant Obstetrician and Gynaecologist, University of Dundee) Chair of DMC; Dr Jonathan Cook (University of Oxford) Independent Statistician; Dr Karen Ward (Consultant Urogynaecologist, Central Manchester Hospitals) Independent Clinician

We would also like to thank the co-applicants of the study: Professor Khalid Khan, Mr Philip Toozs-Hobson, Mr Dharmesh Kapoor, Mr Abu-Bakr-El-Mardi, Mr Zaki Almallah and Karen Powell.

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 Brookes who was the study coordinator, and Elizabeth Lawson, Leanne Fulcher, Laura Gennard, Lisa Leighton for the support in data management. We also thank Nicholas Hilken for designing and developing the study database and Versha Cheed who provided statistical support. We acknowledge the help of Dr Anuradha Sajja, Dr Tina Verghese, 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 the British Society for Urogynaecology for endorsing this study.

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

BUS Collaborators

Birmingham Women's Hospital, Birmingham (*254*). A Sajja, C Burton, E Millership, K Perkins, M Parsons, **P Latthe**, P Toozs-Hobson, T. Verghese, Dr. Balogun, S McCooty, S Rachaneni; City General Hospital, North Staffordshire(*10*). A Clark, F O'Mahony, I Siddiqui, **J Cooper**, J Stokes, S Banerjee, S Jerreat; Crosshouse Hospital, Ayrshire(*9*). D Gilmore, L Oliver, M Henry, M Riad, **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**, Mr. Afifi, 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 & 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 Methewally, **S Radley**, S Duffy; Sandwell Hospital, West Midlands(*2*). M Baptiste, **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, Y Jacobs.

Bold type indicates the principal investigator for each hospital. The number in parentheses indicates the number of participants recruited at each hospital.

Contributors

PL conceived the study, secured funding and was the chief investigator. PL, LJM, JPD, AC, JJD contributed to the study design, data analysis and interpretation. SR, SM, JD, RT contributed to patient recruitment and data collection. MB contributed to the study design and training of collaborators. IG and TR contributed to the study design, economic evaluation and data interpretation. All authors participated in the preparation, review, and crucial revision of the report, which has been approved by each author. PL and JJD are guarantors.

Ethical approval

Nottingham Research Ethics Committee (reference 10/H0408/57, first approved) and NHS trust research governance approvals were obtained for the study. The University of Birmingham was the study's sponsor.

Competing interests none

Dr Latthe has received financial support from Pfizer and Astellas to attend international urogynaecology conferences.

All other authors have completed the ICMJE uniform disclosure form at and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work."

Ethical approval

Nottingham Research Ethics Committee (reference 10/H0408/57) provided the initial favourable ethical opinion on 17th August 2010. NHS trust research governance approvals for each participating hospital were subsequently obtained.

Funding

The study was funded by the National Institute for Health Research Health Technology Assessment programme, project number 09/22/122. The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Reference List

1. 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(1): 96-100.

2. Haylen BT, de Ridder D, Freeman RM, 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(1): 4-20.

3. Coyne KS, Wein A, Nicholson S, Kvasz M, Chen CI, Milsom I. Economic burden of urgency urinary incontinence in the United States: a systematic review. *J Managed Care Pharm* 2014; 20(2): 130-40.

4. Homma Y, Kondo Y, Takahashi S, Kitamura T, Kawabe K. Reproducibility of cystometry in overactive detrusor. *European Urology* 2000; 38(6): 681-5.

5. 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; CD008224(10).

6. Hashim H, Abrams P. Is the bladder a reliable witness for predicting detrusor overactivity? *J Urology* 2006; 175(1): 191-4.

7. Smith A, Bevan D, Douglas HR, James D. Management of urinary incontinence in women: summary of updated NICE guidance. *BMJ* 2013; 347.

8. 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(8): 1019-24.

9. 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(7): 674-83.

10. Rachaneni S, McCooty S, Middleton LJ, et al. Accuracy and economic evaluation of Bladder Ultrasound in the diagnosis of Detrusor Overactivity: a study to evaluate if ultrasound can reduce the need for urodynamics. *Heath Technol Assess* 2016; 20 (7).

11. 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(4): 322-30.

12. Schafer W, Abrams P, Liao L, et al. Good urodynamic practices: uroflowmetry, filling cystometry, and pressure-flow studies. *Neurourol Urodyn* 2002; 21(3): 261-74.

13. McCooty S, Latthe P. Quality control in urodynamics. *Nursing Standard* 2013; 27(43): 35-8.

14. Xiao-Hua Z, Obuchowski NA, McClish DK. Statistical Methods in Diagnostic Medicine. 2nd ed. Hoboken, NJ; Wiley; 2011.

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

16. Khullar V, Cardozo LD, Salvatore S, Hill S. Ultrasound: a noninvasive screening test for detrusor instability. *Br J Obstet Gynaecol* 1996; 103(9): 904-8.

17. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282(11): 1061-6.

18. 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 Urodynam* 2010; 29(3): 428-31.

19. 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.

20. 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(5): 1023-8.

21. Kuhn A, Genoud S, Robinson D, et al. Sonographic Transvaginal Bladder Wall Thickness: Does the Measurement Discriminate Between Urodynamic Diagnoses? *Neurourol Urodynam* 2011; 30(3): 325-8.

22. Panayi DC, Tekkis P, Fernando R, Hendricken C, Khullar V. Ultrasound measurement of bladder wall thickness is associated with the overactive bladder syndrome. *Neurourol Urodyn* 2010; 29(7): 1295-8.

23. Kuo HC. Measurement of detrusor wall thickness in women with overactive bladder by transvaginal and transabdominal sonography. *Int*

Urogynecol J Pelvic Floor Dysfunct 2009; 20(11): 1293-9.

24. 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(11): 1405-11.

25. Chung SD, Chiu B, Kuo HC, et al. Transabdominal ultrasonography of detrusor wall thickness in women with overactive bladder. *BJU Int* 2010; 105(5): 668-72.

26. 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 Urodynam* 2007; 26(2): 176-82.

27. 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(4): 460-.

28. Yang JM, Huang WC. Discrimination of bladder disorders in female lower urinary tract symptoms on ultrasonographic cystourethrography. *J Ultrasound Med* 2002; 21(11): 1249-55.

29. Yang JM, Huang WC. Bladder wall thickness on ultrasonographic cystourethrography: affecting factors and their implications. *J Ultrasound Med* 2003; 22(8): 777-82.

30. 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(2): 145-8.

Legends

Figure 1: Participant flow diagram

Figure 3: Participant flow diagram

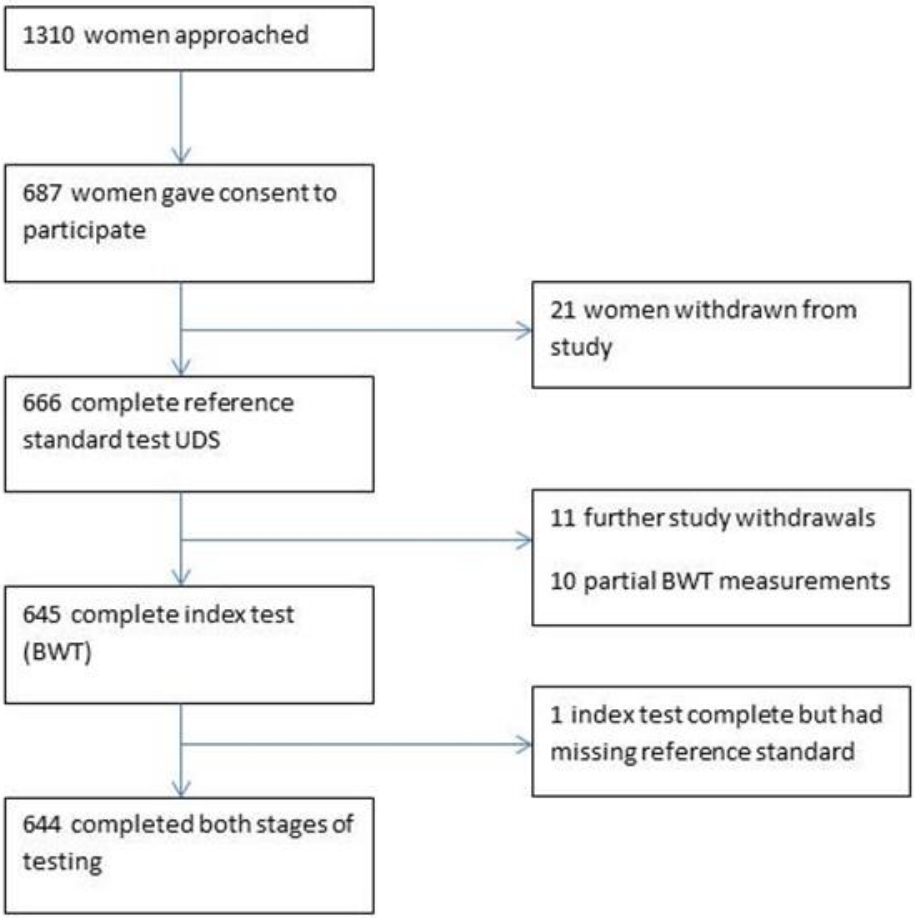

Figure 2: Receiver operating curve of transvaginal bladder wall thickness scan in the diagnosis

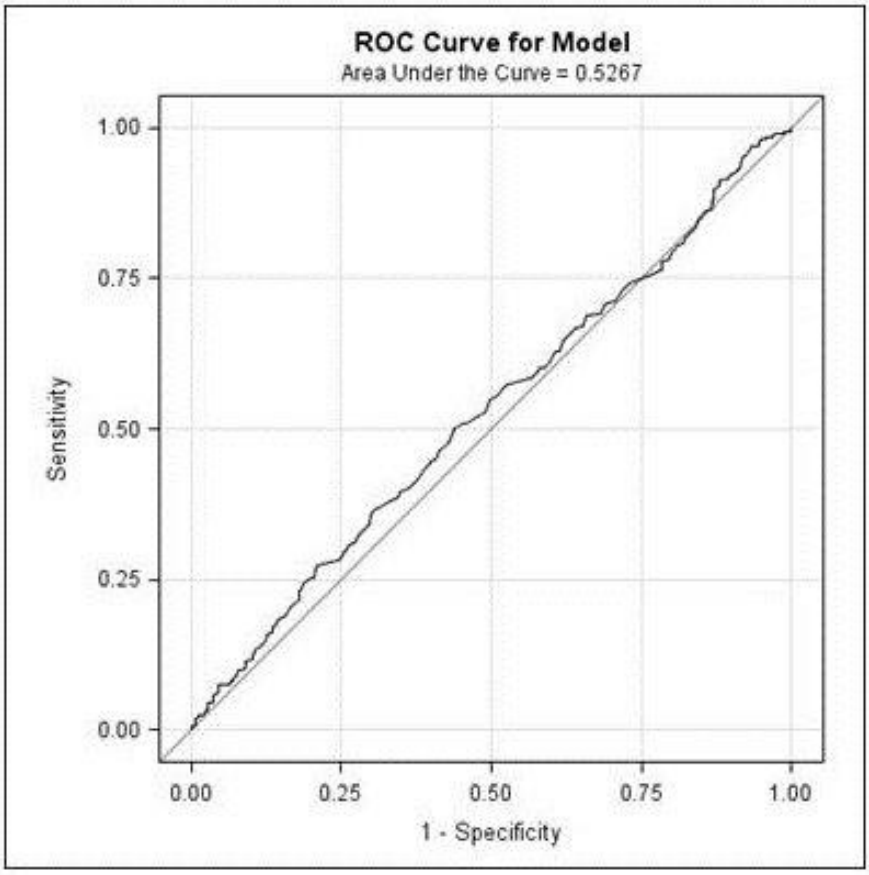

Table 1: Characteristics of included participants (n=687)

Age (years)	Mean (SD)	52.7 (13.9)
	Missing	0 (-)
Ethnicity	White British/Irish/Other	538 (78%)
	Asian Pak/Ind/Bang/other	72 (10%)
	Black Carrib/African/other	49 (7%)
	Mixed/other	18 (3%)
	Not given/missing	10 (1%)
Parity	0	69 (10%)
	1	90 (13%)
	2	241 (35%)
	3	152 (22%)
	4	56 (8%)
	>4	63 (9%)
	Missing	16 (2%)
Post-menopausal (lmp>1 year)	Yes	378 (55%)
	No	293 (43%)
	Missing	16 (2%)
BMI (kg/m^2)	Mean (SD)	30.6 (12.2)
	Missing	28
Incontinence type	Mixed urinary incontinence	419 (61%)
	Urgency incontinence alone	226 (33%)
	Stress incontinence alone	4 (1%)
	Neither	19 (3%)
	Missing	19 (3%)
If mixed, what started first (n=419)?	Urgency	226 (54%)
	Stress	107 (26%)
	Unsure	54 (13%)
	Missing	32 (8%)
Current or previous treatment with anti-muscarinics	Yes	226 (33%)
	No	444 (65%)
	Missing	17 (2%)
Recurrent cystitis (3 or more in last 12 months)	Yes	50 (7%)
	No	606 (88%)
	Missing	31 (5%)
Voiding difficulties	Yes	286 (42%)
	No	374 (54%)
	Missing	27 (4%)
Vaginal birth	Yes	561 (82%)
	No	95 (14%)
	Missing	31 (5%)
Previous incontinence surgery	Yes	36 (5%)
	No	623 (91%)
	Missing	28 (4%)
Previous POP/UI surgery	Yes	56 (8%)
	No	603 (88%)
	Missing	28 (4%)

Table 2: Accuracy of bladder wall thickness ≥5mm for diagnosis of Detrusor Overactivity

	Estimate in %	95% CI
Sensitivity	43% (165 TPs of 388 with DO)	38 to 48%
Specificity	62% (158 TNs of 256 without DO)	55 to 68%
Positive predictive value	63% (165 TPs of 263 with BWT≥5mm)	57 to 69%
Negative predictive value	41% (158 TNs of 381 BWT<5mm)	36 to 47%
Positive likelihood ratio	1.11	0.92 to 1.35
Negative likelihood ratio	0.93	0.82 to 1.06

Figure S1: Study flow chart

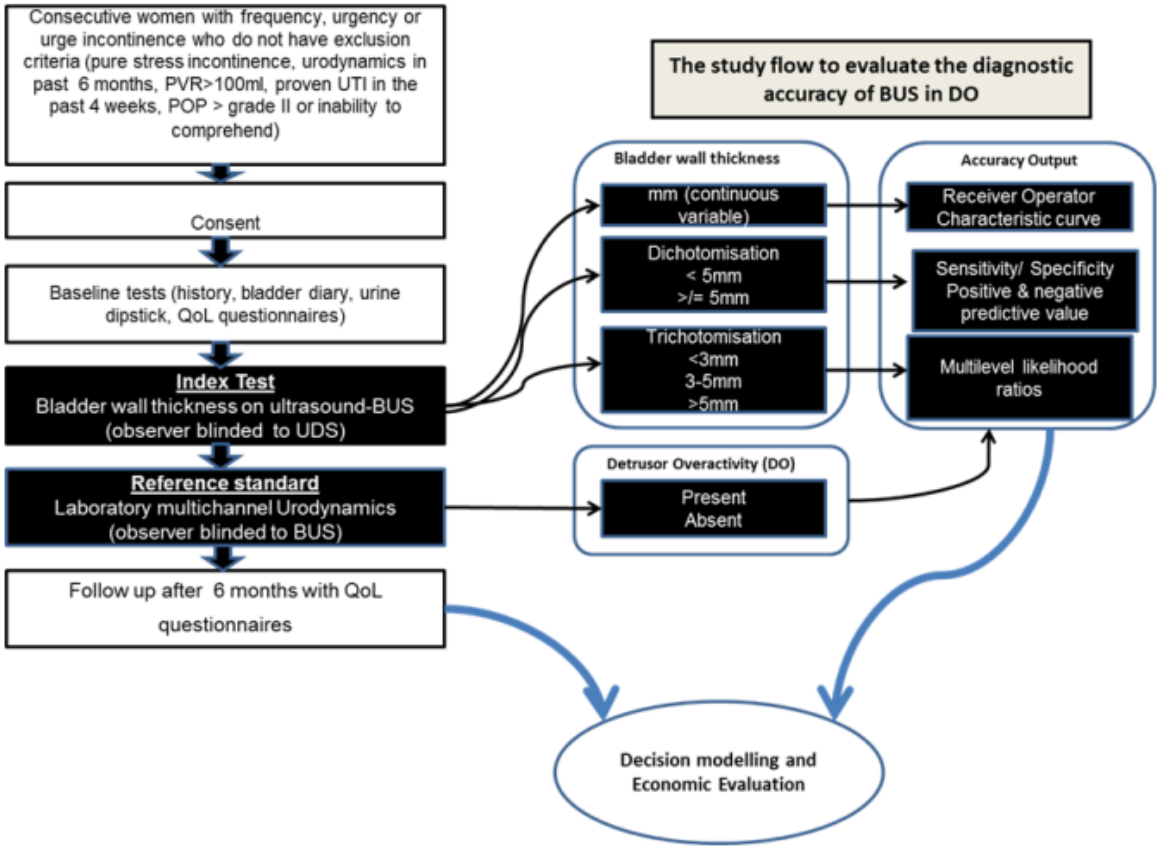

Figure S2: Transvaginal scan showing measurements of the trigone, anterior wall and dome of the bladder

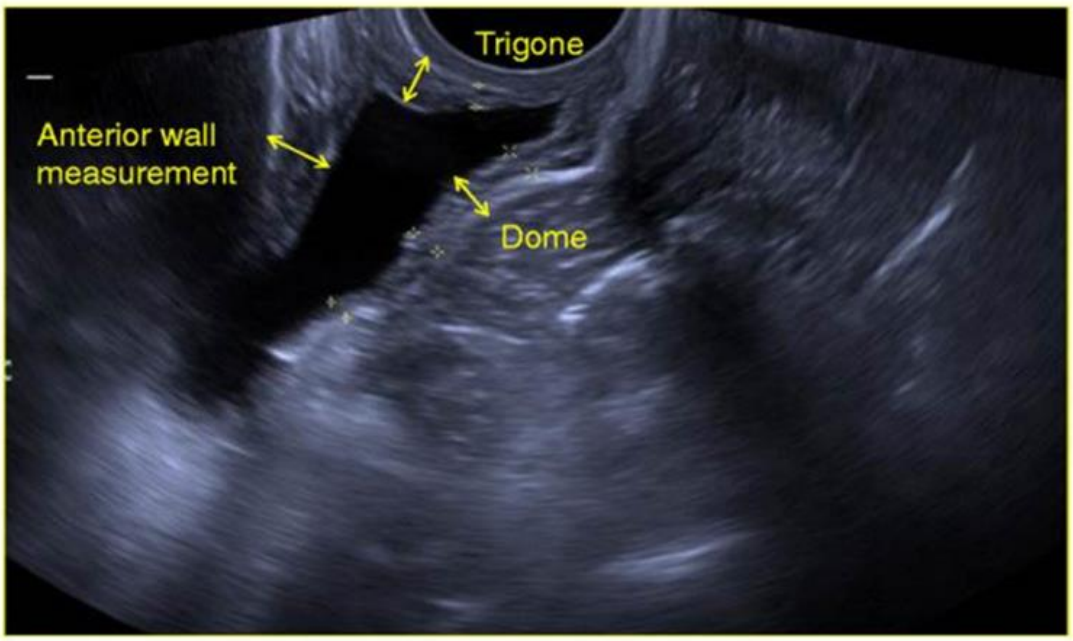

Figure S3: Box and whisker plot comparing bladder wall thickness with detrusor overactivity diagnosis

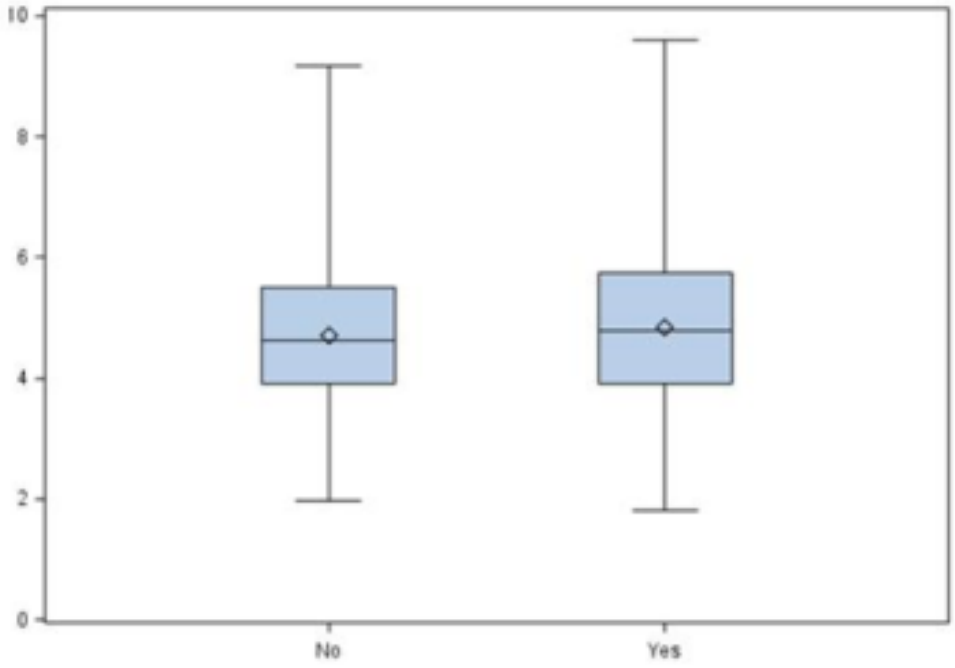

Figure S4: ROC curve from sensitivity analysis excluding those results where the urodynamics test was not blind to the results of the ultrasound test (16/632 women (3%); AUC: 0.528, 95%CI: 0.480, 0.575; p=0.25)

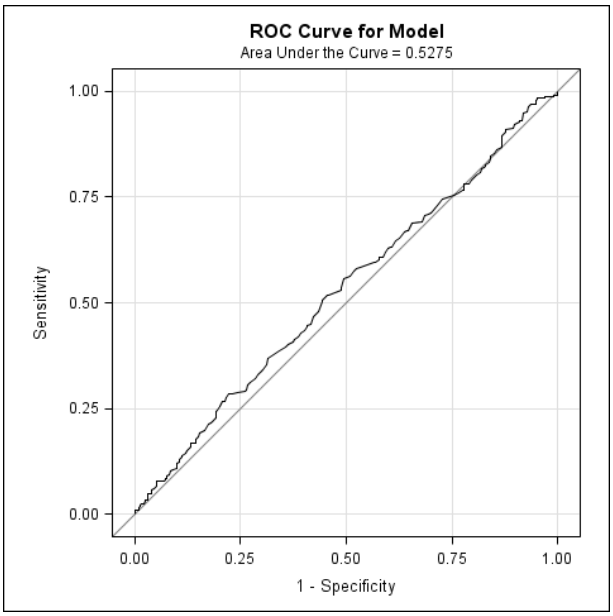

Figure S5: ROC curve from sensitivity analysis excluding those results where there was more than four weeks between the tests (26/660 women (4%); AUC: 0.526, 95%CI: 0.479, 0.572; p=0.28)

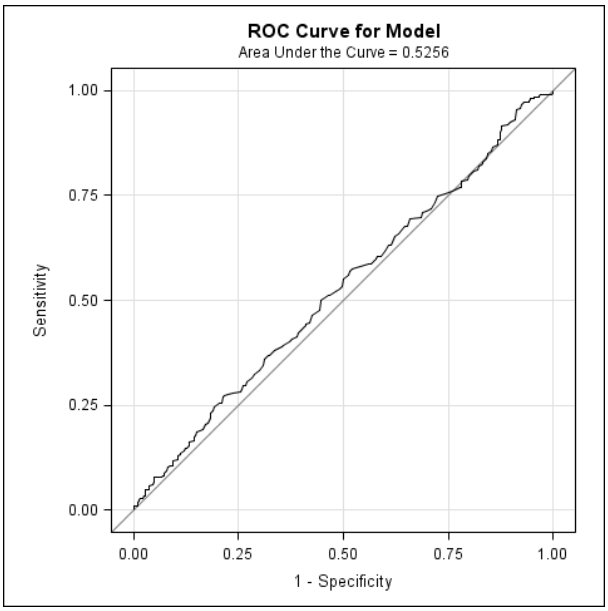

Figure S6: ROC curve from sensitivity analysis incorporating incomplete ultrasound measurements (10 observations – average of remaining one or two measurements used; AUC: 0.529, 95%CI: 0.484, 0.574; p=0.21)

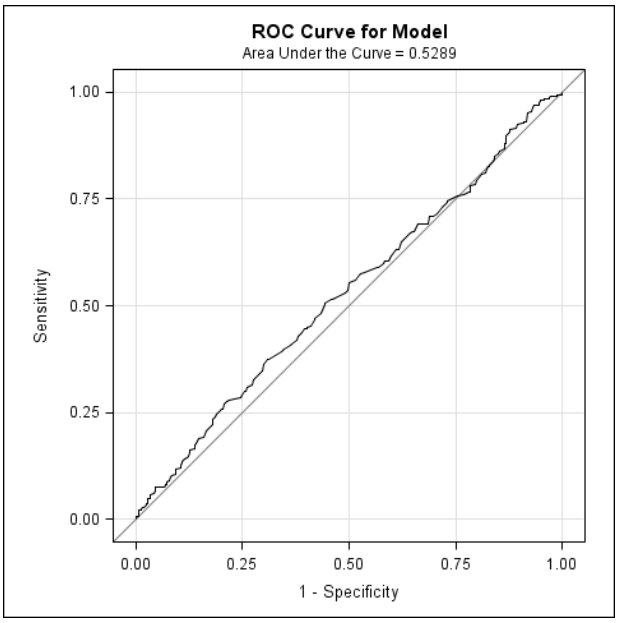

Figure S7: ROC curve from exploratory analysis including the urgency alone group (as per clinical history; excluding mixed stress/urge incontinence group: 217 patients; AUC: 0.530, 95%CI: 0.452, 0.609; p=0.45)

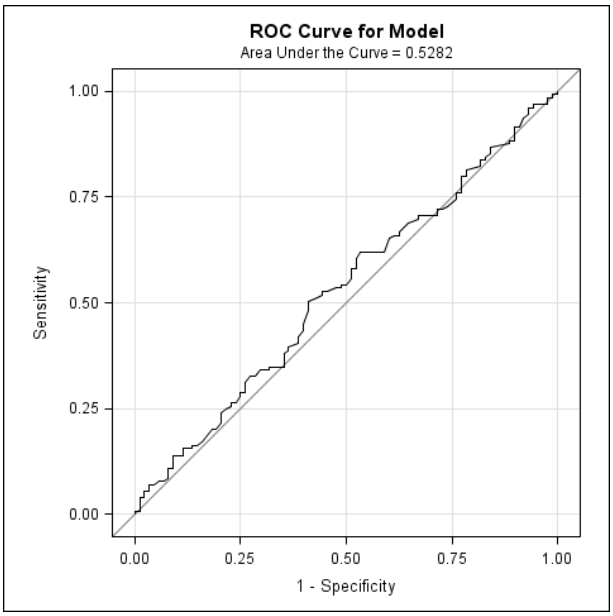

Figure S8: ROC curve from exploratory analysis excluding those who had PVR>30ml upon testing (34 cases AUC: 0.526, 95%CI: 0.479, 0.572; p=0.28)

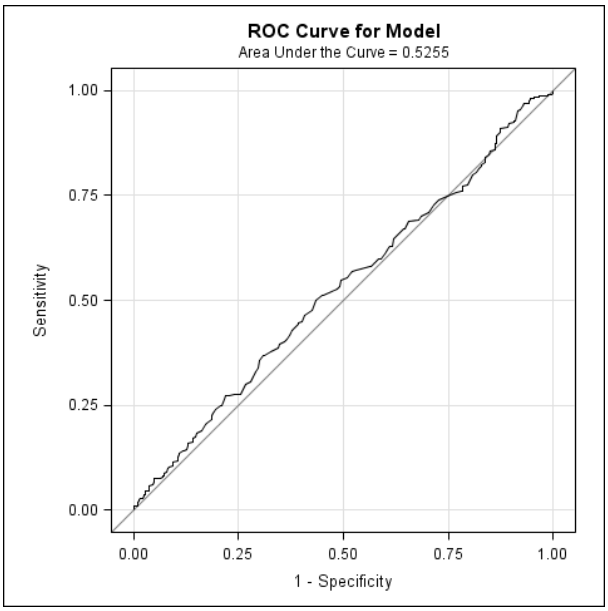

Figure S9: ROC curve from exploratory analysis using the trigone measurement alone for BWT (AUC: 0.519, 95%CI: 0.473, 0.564; p=0.42)

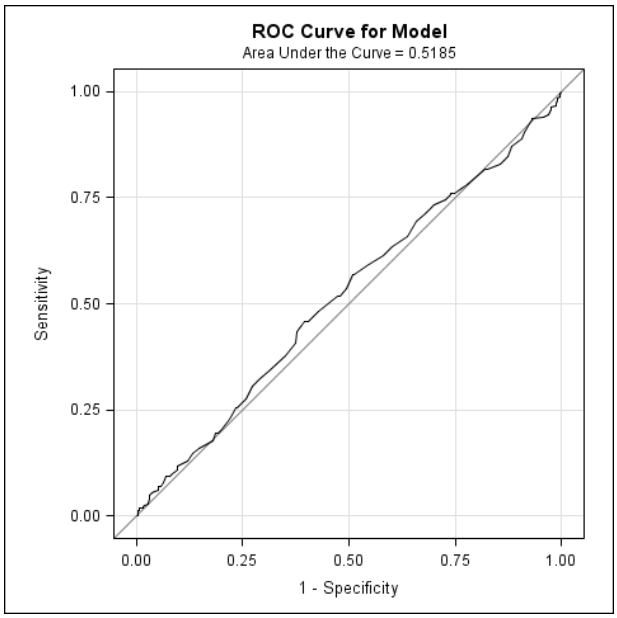
ROC Curve for Model
Area Under the Curve = 0.5185
Sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 - Specificity

Figure S10: ROC curve from exploratory analysis using the average dome, 1cm left of dome, 1cm right of dome (AUC: 0.537, 95%CI: 0.491, 0.582; p=0.12)

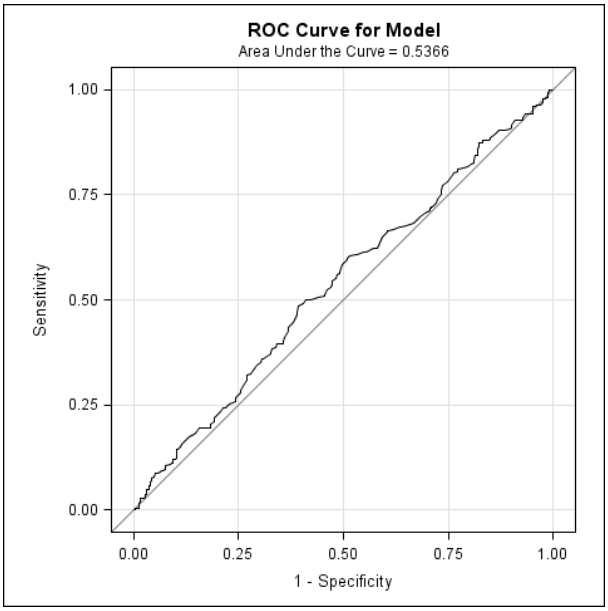

Figure S11: ROC curve from exploratory analysis redefining those who had detrusor pressure rise upon provocation testing 'provoked DO' as DO negative (187 cases; AUC: 0.541, 95%CI: 0.487, 0.595; p=0.14)

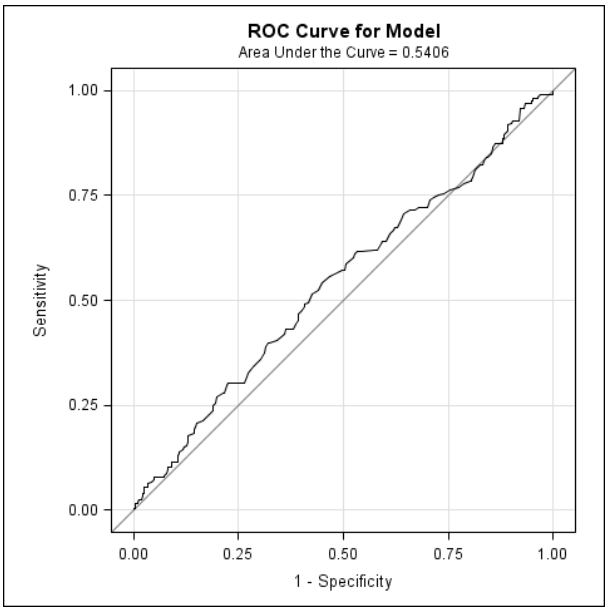

Figure S12: ROC curve from exploratory analysis redefining those with 'mixed' DO (DO with another diagnosis of USI or voiding dysfunction) as do negative; AUC: 0.489, 95%CI: 0.440, 0.531; p=0.54)

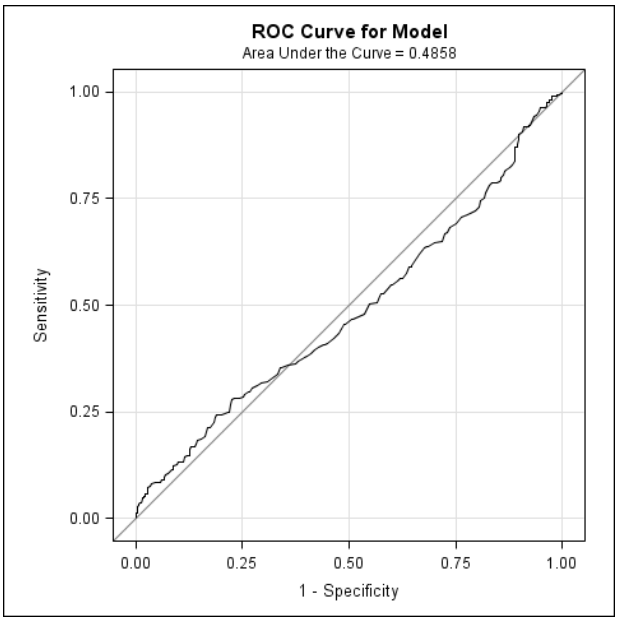

Figure S13: ROC curve from exploratory analysis redefining those with 'dry' DO as DO negative; AUC: 0.548, 95%CI: 0.501, 0.594; p=0.05)

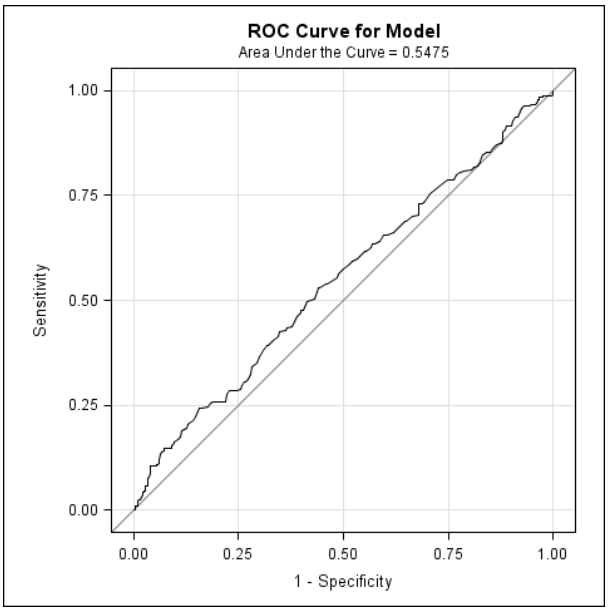

Table S1: Summary of all urodynamic diagnoses

	Number of women (%) n=666
Diagnoses that include DO (n=399):	
DO only	258 (39%)
DO/USI	97 (15%)
DO/voiding dysfunction	18 (3%)
DO/voiding dysfunction/USI	12 (2%)
DO/low compliance	8 (1%)
DO/USI/low compliance	5 (1%)
DO/voiding dysfunction/USI/low compliance	1 (<1%)
Diagnoses that do not include DO (n=267):	
Normal	124 (19%)
USI only	78 (12%)
Low compliance only	36 (5%)
Voiding dysfunction only	14 (2%)
Voiding dysfunction/USI	8 (1%)
USI/low compliance	6 (1%)
Voiding dysfunction/low compliance	1 (<1%)

Table S2: Bladder wall thickness (mm) summary statistics

	Overall group		**DO positive**		**DO negative**	
Measurement	**Mean (SD), n**	**Min, max**	**Mean (SD), n**	**Min, max**	**Mean (SD), n**	**Min, max**
Trigone	4.51 (1.49), 648	1.20, 9.90	4.55 (1.52), 391	1.20, 9.50	4.46 (1.44), 257	1.20, 9.90
Dome midline	5.01 (1.67), 653	1.30, 11.90	5.15 (1.68), 394	1.60, 10.80	4.79 (1.62), 259	1.30, 11.90
Anterior wall	4.86 (1.52), 650	1.60, 11.30	4.86 (1.54), 391	1.60, 11.30	4.85 (1.50), 259	1.60, 9.30
Average of above 3	**4.79 (1.33), 644**	**1.80, 9.60**	**4.85 (1.36), 388**	**1.80, 9.60**	**4.70 (1.29), 256**	**1.97, 9.17**

Table S3: Likelihood ratios for bladder wall thickness for diagnosis of Detrusor Overactivity

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

Table S4: Results of Receiver Operator Characteristics (ROC) curve analysis in pre-specified subgroupings

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

AUC- area under the curve

Table S5: Results of univariate analysis exploring factors possibly associated with Detrusor Overactivity (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, 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, 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, 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		

Table S6: Results of multivariable analysis exploring factors possibly associated with Detrusor Overactivity diagnosis

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