CRD summary
The review assessed the effectiveness and cost-effectiveness of diagnostic methods for urinary incontinence, specifically urodynamic stress incontinence and detrusor overactivity. The authors put forward a number of conclusions, which are likely to be reliable.

Authors' objectives
To determine the effectiveness and cost-effectiveness of diagnostic methods for urinary incontinence, specifically urodynamic stress incontinence (USI) and detrusor overactivity (DO). Only clinical effectiveness is considered in this abstract.

Searching
MEDLINE (1966 to 2002), CINAHL (1982 to 2002) and EMBASE (1980 to 2002) were searched for relevant papers published in the English language; the search terms were reported.

Study selection
Study designs of evaluations included in the review
Studies with a comparison group were included in the review.

Specific interventions included in the review
Studies comparing one or more of the following methods of diagnosing urinary incontinence were eligible for inclusion: simple investigations (e.g. validated scales, urinary diaries and pad tests), invasive investigations (e.g. urodynamics), or clinical history-taking.

Reference standard test against which the new test was compared
The reference standard selected was multichannel urodynamics. Single-channel urodynamics, ambulatory urodynamics, and history were also included as reference standards in the review. Additional reference standards (such as validated scales and pad tests) were also included.

Participants included in the review
Studies of adults (older than 19 years) with urinary incontinence were eligible for inclusion. Individuals with USI, DO, mixed urinary incontinence, any urinary incontinence, or any leakage were included in the review. Participants from a variety of settings (primary care, secondary care and clinical trials) were included in the review.

Outcomes assessed in the review
The outcomes included sensitivity, specificity and positive predictive values. Studies investigating interventional procedures where diagnostic tests were used as outcome measures were excluded.

How were decisions on the relevance of primary studies made?
One reviewer selected potentially relevant studies on the basis of titles and/or abstracts; all of the abstracts were classified as relevant, not relevant, or unclear. A second reviewer independently assessed a sample of these papers: 20% of the relevant records, 10% of the not relevant records, and 100% of the unclear records. The agreement between reviewers was 98%. One reviewer then assessed full copies of papers identified as potentially relevant and classified them as relevant, irrelevant, or unclear; a second reviewer read 20% of the relevant and not relevant papers and 100% of the unclear papers. Any disagreements were discussed. The agreement between reviewers was 96%.

Assessment of study quality

The quality of the primary papers was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. Additional instructions were added to ensure consistency between assessors. Seven reviewers were involved in the validity assessment; two different reviewers assessed a sample (16 of 121) of the papers to test inter-reader agreement. The agreement between reviewers ranged from 0.50 to 1.00. Any disagreements were resolved by consensus.

Data extraction
Members of the project team extracted details including size, gender and age of sample, as well as care setting and country. Where additional information was required, the authors of the primary studies were contacted.

Methods of synthesis
How were the studies combined?
Studies of the same diagnostic procedure using the same cut-off threshold were pooled using random-effects models. Summary estimates were presented as sensitivity, specificity and diagnostic odds ratio (DOR), along with their associated 95% confidence intervals (CIs). The positive likelihood ratio was calculated from the pooled specificity and sensitivity. Where a quantitative analysis was not appropriate, a narrative synthesis was presented.

How were differences between studies investigated?
Statistical heterogeneity was assessed using chi-squared tests. For each type of test, the studies were grouped by type of urinary incontinence (USI or DO) and gender.

Results of the review
One hundred and twenty-nine studies were included in the review.

Of the included papers: 84% reported an appropriate reference standard; 86% reported that individuals received the same reference standard regardless of the index test result; 77% reported that the two diagnostic tests were independent of each other; 91% reported that the whole sample or a random selection of the whole sample received verification using a reference standard of diagnosis; 39% did not adequately report the selection criteria used; it was unclear whether 79% of the reference tests and 83% of the index tests were interpreted without knowledge of the other test; 67% did not report withdrawals; 64% did not report the duration between the two diagnostic tests; 64% included a representative spectrum of patients; 64% of the index tests and 59% of the reference tests were described sufficiently for replication; and 79% provided the same clinical data as would be available in clinical practice.

Clinical history when compared with urodynamics (presence or absence of stress incontinence symptoms) for diagnosing USI in women was found to have a pooled sensitivity of 0.92 (95% CI: 0.91, 0.93) and a pooled specificity of 0.56 (95% CI: 0.53, 0.60), based on 15 studies. Significant heterogeneity was found for both analyses (p<=0.001 and p<=0.001). The positive likelihood ratio was 2.09 (95% CI: 1.83, 2.35) and the area under the curve (AUC) for the receiver operating characteristic (ROC) curve for the pooled DOR was 0.83 (95% CI: 0.71, 0.95). For diagnosing DO in women, a pooled specificity of 0.61 (95% CI: 0.57, 0.65) and a pooled sensitivity of 0.87 (95% CI: 0.85, 0.89) were found, based on 8 studies. Significant heterogeneity was found for both analyses (p<=0.001 and p<=0.001). The positive likelihood ratio was 4.69 (95% CI: 4.05, 5.33) and the AUC for the ROC curve for the pooled DOR was 0.83 (95% CI: 0.69, 0.97). Three studies considered clinical history for diagnosing DO in men; the sensitivity ranged from 0.50 to 1.00 and the specificity from 0.50 to 0.77.

Validated scales (based on question 3 of the Urogenital Distress Inventory) compared with urodynamics were found to have a pooled sensitivity of 0.87 (95% CI: 0.82, 0.92) and a pooled specificity of 0.60 (95% CI: 0.51, 0.69), based on 2 studies. No significant heterogeneity was found.

Seven studies compared pad tests with multichannel urodynamics. The sensitivity ranged from 0.86 to 0.94 and the specificity from 0.44 to 0.72; estimates were not pooled because there were insufficient studies comparing the same pad tests or reporting adequate data.

Urinary diary compared with urodynamics for diagnosing DO in women was found to have a sensitivity of 0.88 (95%
Imaging by ultrasound to determine leakage in the diagnosis of USI in women was found to have a pooled sensitivity of 0.89 (95% CI: 0.84, 0.93), based on 4 studies; significant heterogeneity was found (p=0.004). The pooled specificity was 0.82 (95% CI: 0.73, 0.89; 4 studies); no significant heterogeneity was found.

X-ray to image bladder neck descent compared with multichannel urodynamics for diagnosing USI was found to have a sensitivity of 0.79 (95% CI: 0.67, 0.88) and a specificity of 0.55 (95% CI: 0.43, 0.66), based on 2 studies. Significant heterogeneity was found for specificity only (p<0.001). The positive likelihood ratio was 1.76 (95% CI: 0.90, 2.61).

Six studies compared stress test with multichannel urodynamics, of which three were pooled. A sensitivity of 0.85 (95% CI: 0.78, 0.91) and a specificity of 0.83 (95% CI: 0.74, 0.90) were found for diagnosing USI in women in secondary care settings using the supine stress test compared with multichannel urodynamics. No significant heterogeneity was found. The positive likelihood ratio was 5.00 (95% CI: 3.79, 6.21), and the AUC for the ROC curve corresponding to the pooled DOR was 0.87 (95% CI: 0.69, 1.00).

Additional results for other comparators were also presented. No studies met the inclusion criteria for the diagnosis of bladder outlet obstruction.

**Cost information**
Urinary diary had the lowest cost-effectiveness ratio (between £35 and £77) per extra unit of effectiveness, or case diagnosed, of three primary care tests (diary, pad tests and validated scales) in addition to clinical history.

**Authors' conclusions**
Clinical history alone can correctly diagnose a large proportion of women with USI. Ultrasound imaging may offer a valuable alternative to urodynamic investigation. Of the urodynamic procedures, multichannel urodynamics is likely to give the most accurate result in a secondary care setting. The clinical stress test is effective in the diagnosis of USI. There is a lack of studies on diagnosing incontinence in men.

**CRD commentary**
The review question was supported by broad inclusion criteria. A number of relevant electronic databases were searched. However, the search strategy was restricted by language, there appears to have been no attempt to locate unpublished material, and no assessment of publication bias was reported. The procedures undertaken to select primary papers and assess their quality were likely to have minimised error or bias. Where possible the studies were synthesised quantitatively using generally appropriate methods, and narratively when this was not. The authors highlighted that reporting in the primary studies was poor, and that clinical interpretation was often difficult as few studies could be synthesised. The author's conclusions are likely to be reliable.

**Implications of the review for practice and research**
Practice: The authors stated that most diagnostic methods can be undertaken in primary or secondary care settings, and that simple investigations (e.g. pad test or urinary diary) may provide useful information on severity, which combined with clinical history may provide adequate information to commence low cost and low risk primary care treatment.

Research: The authors stated the need for large-scale, high-quality primary studies in primary care settings to verify or refute the findings of this review. These studies should assess both the clinical effectiveness, including quality of life, and the cost-effectiveness of diagnostic methods of urinary incontinence. In addition, the authors stated that the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) guidelines should be followed to ensure the accuracy and completeness of reporting the design and results.

**Funding**
NHS R&D Health Technology Assessment (HTA) Programme, project number 99/29/02.
Bibliographic details

PubMedID
16487456

Original Paper URL
http://www.hta.ac.uk/project.asp?PjtId=1213

Other publications of related interest

Indexing Status
Subject indexing assigned by NLM

MeSH
Diagnostic Tests, Routine /economics; Evaluation Studies as Topic; Great Britain; Humans; Meta-Analysis as Topic; Urinary Incontinence /diagnosis /etiology

AccessionNumber
12006008291

Date bibliographic record published
30/06/2007

Date abstract record published
30/06/2007

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.