



Effective Health Care Program

Nonsurgical Treatments for Urinary Incontinence in Adult Women: Diagnosis and Comparative Effectiveness

Executive Summary

Background

Urinary incontinence (UI) is the involuntary loss of urine.¹ About 25 percent of young women,² 44 to 57 percent of middle-aged and postmenopausal women,³ and about 75 percent of older women experience some involuntary urine loss.⁴ UI can affect women's physical, psychological, and social well-being, and sometimes imposes significant lifestyle restrictions. The effects of UI range from slightly bothersome to debilitating.

The cost of incontinence care in the United States averaged \$19.5 billion in 2004.⁵ Six percent of nursing home admissions of older women are attributable to UI,⁵ and by one estimate, the annualized cost of women's nursing home admissions due to UI was \$3 billion.⁶

Nonpharmacological therapies target strengthening the pelvic floor and changing behaviors that influence bladder function, whereas pharmacological therapies address innervating the bladder and sphincter. The etiology of incontinence is multifactorial; risk factors include age, pregnancy, pelvic floor trauma after vaginal delivery, menopause, hysterectomy, obesity, urinary tract infections, functional and/or cognitive impairment, chronic cough, and

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at www.effectivehealthcare.ahrq.gov/reports/final.cfm.

constipation.⁷ Assessments of women complaining of UI begin with exclusion of underlying causes such as pelvic organ prolapse, urinary tract infection, and poor bladder emptying,⁸ all of which are beyond the scope of this review, as is neurogenic



Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Effective
Health Care

UI associated with spinal cord injury or stroke.⁹ We focus specifically on women with stress UI associated with sphincter function, and with urgency UI, often associated with overactive bladder (Table 1 in the full report).

Incontinence types are distinguished by their baseline mechanisms. Stress incontinence is associated with impaired sphincter function, and results in an inability to retain urine during coughing or sneezing.⁹ Urgency incontinence is defined as involuntary loss of urine associated with the sensation of a sudden compelling urge to void that is difficult to defer.⁹ Mixed UI is the term applied when both stress and urgency UI are present. These definitions reflect the consensus definitions developed by the International Urogynecological Association/International Continence Society.⁹ Overactive bladder is defined as urinary urgency with or without incontinence, usually accompanied by frequency and nocturia (the need to urinate at night).⁹ Approximately one-third of women with overactive bladder also experience urgency UI.

The types of UI imply different attendant risk factors and recommended treatments; however, UI etiology is frequently mixed.⁸ Stress UI is more common in younger women in association with pelvic floor trauma and uterine prolapse, both of which are often related to vaginal delivery and may require surgical treatments.⁷ Urgency and mixed UI are more common in older women in association with overactive bladders with or without sphincter dysfunction.^{1,7}

Although UI can be diagnosed based on patients' reports of involuntary urine leakage,⁷ researchers have also proposed clinical methods for objective diagnosis of different UI types. Urodynamic diagnosis of pure stress UI without detrusor overactivity has demonstrated usefulness for women undergoing surgery for stress UI.⁹ Diagnostic studies use multichannel urodynamics as a reference standard test to compare with noninvasive tests applicable to ambulatory care. However, researchers disagree on whether urodynamic examination represents the gold standard for UI diagnosis.⁸ Furthermore, urodynamic examination is not possible in ambulatory primary care. Previously published systematic reviews have reported a weak association between urodynamic test results and self-reported symptoms,¹⁰ but these reviews did not focus on the most appropriate methods to distinguish different types of UI in ambulatory care settings. The role of invasive diagnostic methods in predicting which patients will benefit from specific treatments for UI remains unclear.

Standard UI treatments for women include lifestyle changes, pelvic floor muscle training, and, for predominant

stress UI, surgical treatments.¹ In addition, several drugs have been approved for adults with overactive bladder, with or without urgency UI.¹ Clinical interventions to reduce the frequency of UI episodes in women have been extensively reviewed in recent years,^{8,11} but the reviews did not emphasize continence or women's perceptions of treatment success and satisfaction. Continence (complete voluntary control of the bladder) has been considered a primary goal in UI treatment^{8,12} and is the most important outcome associated with quality of life in women with UI;¹³ yet, it is rarely examined as a primary outcome in syntheses of evidence.¹⁴ Thus, we focus on continence and quality of life as primary outcomes for this Comparative Effectiveness Review.

While definitions of continence are similar, the definitions most commonly applied to improvement in UI vary and include different degrees of change in frequency and severity of symptoms.¹⁵ Furthermore, improvement in UI has been viewed very differently by women and by researchers. Women define improvement according to reduced lifestyle restrictions or improved overall perception of bladder symptoms, especially resolution of urine leakage, whereas researchers define improvement as a decrease in the amount of lost urine during pad tests, or any statistically significant decrease in the frequency of UI episodes.¹⁵ Treatments for overactive bladder aim to decrease the frequency and intensity of urgency sensations, as well as the frequency of urgency UI episodes. Previous reviews of treatments for overactive bladder have considered clinical success as any statistically significant decrease in the frequency of UI episodes and voiding, irrespective of whether women perceived improvement.¹⁴ Measurement of treatment outcomes should be patient centered and based on factors important to women, rather than on the results of invasive tests.¹² Thus, treatment success and failure should be evaluated according to what women report in validated questionnaires or scales. Ultimately, discussions of UI are complicated by the wide variety of measures used to describe the problem and its treatment outcomes. This review examines improvement thresholds of clinical importance in validated scales and checklists that can be applied to judge UI treatment success according to women's own perceptions.

This report synthesizes published evidence about diagnosis and management of UI in adult women. We focused on adult women in ambulatory care settings and on nonsurgical nonpharmacological treatments and pharmacological agents available in the United States. This report is intended as a companion piece to an earlier Evidence-based Practice Center report⁷ that examined a

wide range of treatment alternatives, including surgery. We focus on techniques appropriate to primary care ambulatory practice and nonsurgical interventions for women with refractory UI.

Our report also addresses the role of urodynamic testing, which is not typically performed in primary care. We include it here primarily as background information for primary care practitioners, and because it raises a conundrum. As we have emphasized, the primary outcome for UI should be patient-centered reports of the UI experience, especially the presence or absence of UI. Although we typically think of physiological testing as more objective than patient reports, these results are, at best, akin to intermediate outcomes. In the diagnostic context, physiological testing can inform in one of three ways: (1) establishing a diagnosis, (2) determining an etiology with therapeutic implications, and (3) generating a prognosis. In the case of UI, it is unclear whether physiological measures represent a gold standard against which other measures can be compared, or whether they should be viewed as information that may predict key patient-centered outcomes. Hence, we may be more interested in levels of agreement between physiological measures and patient outcomes but hard pressed to interpret differences between them. We examine the role of urodynamic testing in diagnosing and treating UI to provide insight into this conundrum.

Our systematic review is intended to help clinicians, consumers, and policymakers make clinical recommendations and informed decisions based on synthesized evidence and other relevant factors.

Objectives

We present a comprehensive synthesis of evidence regarding valid methods to diagnose UI in adult women and to monitor treatment benefits and harms. We evaluated the clinical efficacy and comparative effectiveness of pharmacological and nonsurgical treatments for UI in adult women following the principles from the Methods Guide for Effectiveness and Comparative Effectiveness Reviews from the Agency for Healthcare Research and Quality (AHRQ) (www.effectivehealthcare.ahrq.gov). We examined the following questions:

Key Question 1. What constitutes an adequate diagnostic evaluation for women in the ambulatory care setting on which to base treatment of urinary incontinence?

1. What are the diagnostic values of different methods—questionnaires, checklists, scales,

self-reports of UI during a clinical examination, pad tests, and ultrasound—when compared with multichannel urodynamics?

2. What are the diagnostic values of different methods—questionnaires, checklists, scales, self-reports of UI during a clinical examination, pad tests, and ultrasound—when compared with a bladder diary?
3. What are the diagnostic values of the methods listed above for different types of UI, including stress, urgency, and mixed incontinence?
4. What is the association between patient outcomes (continence, severity and frequency of UI, quality of life) and UI diagnostic methods?

Key Question 2. How effective is the pharmacological treatment of UI in women?

1. How do pharmacologic treatments affect continence, severity and frequency of UI, and quality of life when compared with no active treatment or with combined treatment modalities?
2. What is the comparative effectiveness of pharmacological treatments when compared with each other or with nonpharmacological treatments of UI?
3. What are the harms from pharmacological treatments when compared with no active treatment?
4. What are the harms from pharmacological treatments when compared with each other or with nonpharmacological treatments of UI?
5. Which patient characteristics, including age, type of UI, severity of UI, baseline disease that affects UI, adherence to treatment recommendations, and comorbidities, can modify the effects of the pharmacological treatments on patient outcomes, including continence, quality of life, and harms?

Key Question 3. How effective is the nonpharmacological treatment of UI in women?

1. How do nonpharmacological treatments affect incontinence, UI severity and frequency, and quality of life when compared with no active treatment?
2. How do combined modalities of nonpharmacological treatments with drugs affect incontinence, UI severity and frequency, and quality of life when compared with no active treatment or with monotherapy?
3. What is the comparative effectiveness of nonpharmacological treatments when compared with each other?

4. What are the harms from nonpharmacological treatments when compared with no active treatment?
5. What are the harms from nonpharmacological treatments when compared with each other?
6. Which patient characteristics, including age, type of UI, severity of UI, baseline disease that affects UI, adherence to treatment recommendations, and comorbidities, can modify the effects of the nonpharmacological treatments on patient outcomes, including continence, quality of life, and harms?

Methods

Input From Stakeholders

We developed research questions and an analytic framework after discussions with key informants and technical experts. Research questions for the systematic review were posted for public comment, based on which we identified interventions eligible for this review. Stakeholders recommended a focus on patient-centered outcomes and interventions most relevant for ambulatory care and not evaluated in previous systematic reviews. Stakeholders also recommended reviewing nonsurgical interventions relevant to women with refractory UI. Comprehensive information about all nonsurgical treatment choices can lead to evidence-based referral practices for women with refractory UI.

Candidates to serve as key informants, technical experts, and peer reviewers were approved by the Task Order Officer from AHRQ after disclosure of conflicts of interest. The protocol was developed with input from the Technical Expert Panel.

Data Sources and Selection

We sought studies from MEDLINE® via OVID and via PubMed®, the Cochrane Library, SCIRUS, Google Scholar, other databases, and manual searches of reference lists from systematic reviews. We identified studies published in English from 1990 through December 30, 2011.

Study Selection

Three investigators independently determined the eligibility of the studies. For Key Question 1, we included studies that evaluated different methods to diagnose UI in women that are applicable to ambulatory care settings. Index methods that are applicable to ambulatory care settings were compared in eligible studies with

urodynamic or clinical diagnosis of UI made by investigators in specialized clinics.

For Key Questions 2 and 3, we included randomized controlled trials (RCTs) that combined men and women if they reported outcomes in women separately or included more than 75 percent women. We excluded studies of men, children, or residents of long-term care facilities. We excluded studies of surgical treatments for UI or urogenital prolapse and studies of drugs not available in the United States. We analyzed harms regardless of how authors perceived the causality of treatments. We included observational studies with adjusted treatment estimates. We included observational studies of treatments not examined in RCTs.

Data Extraction

Evaluations of the studies, data extraction, and quality control were conducted by four researchers using a standardized form. We abstracted minimum datasets for diagnostic and therapeutic studies. We abstracted inclusion of minorities, inclusion of women who failed prior therapy for UI, inclusion of mixed UI, baseline daily UI, and presence of urogenital prolapse or hysterectomy in female participants. We focused on urgency UI in women with overactive bladder and did not analyze urgency, voiding frequency, or nocturia.

Quality Assessment

We evaluated the quality of studies and classified them by their designs. We evaluated studies for Key Question 1 with predefined criteria for assessing the quality of the diagnostic accuracy of studies. We evaluated the quality of therapeutic studies using predefined criteria to assess the risk of bias, which included randomization, adequacy of randomization and allocation concealment, masking of the treatment status, and intention-to-treat analyses. We examined sponsorship and conflict of interest but did not downgrade quality using this information. We incorporated quality in the synthesis of evidence, conducting meta-regression, subgroup, and sensitivity analysis for each quality criterion rather than for the overall quality score. Well-designed RCTs are believed to have a low risk of bias. We defined studies as having a medium or high risk of bias if one or more quality criteria were not met.

Applicability of the population was estimated by evaluating the selection of women in observational studies and clinical trials. For each study, we examined settings, including ambulatory care or specialized clinics, recruitment in the clinical settings or in the community, inclusion age and type of UI, and exclusion criteria.

Data Synthesis and Analysis

For Key Question 1, results of individual studies were summarized to analyze sensitivity, specificity, predictive values, diagnostic odds ratios, and predictive likelihood ratios for correct diagnosis of any, stress, and urgency UI. We focused on the predictive likelihood ratios of UI in women examined with index tests when compared to women who had urodynamic or clinical diagnosis. Ratios of 1 indicated that the tests likely do not provide accurate UI diagnosis. Ratios of more than 10 provided large and often conclusive increases in the likelihood of UI. We pooled diagnostic test data with random effects models using an inverse variance weighting method with Meta-Analyst software. Random effects meta-analyses incorporate heterogeneity by assuming a normal distribution of underlying effects. In cases of heterogeneity, we used bivariate pooling methods.

Following guidelines and recommendations from key informants and members of our Technical Expert Panel, we focused on patient-centered outcomes, including continence, improvement in UI, quality of life, adverse effects, and discontinuation due to adverse effects. Voiding frequency in women with overactive bladder had been reviewed previously and was outside of our scope. The methods to assess harms were not assessed for validity. For Key Questions 2 and 3, we calculated relative risk, absolute risk differences, number needed to treat, and the number of events attributable to active treatment per 1,000 persons treated for binary outcomes. We assessed missing data across studies, including loss to followup and dropout patterns, and forced intention-to-treat analyses using the number of randomized subjects for all calculations.

Meta-analysis was conducted when clinical populations, interventions, and outcomes were deemed sufficiently similar. We chose the random-effects inverse variance weights model to incorporate in the pooled analysis differences across trials in patient populations, baseline rates of the outcomes, dosage of drugs, and other factors. We analyzed adverse effects with drugs for urgency UI using double arcsine transformations of the event rates. We examined consistency in results across studies with Chi square tests and I square statistics. Using a standard preplanned algorithm, we explored heterogeneity with meta-regression, subgroup, and sensitivity analysis by clinical diversity, treatment dose and duration, and quality criteria of individual studies, and whether conflict of interest was disclosed by study authors. When exploring heterogeneity, we did not use subject-level variables to avoid an ecological fallacy. We calculated Bayesian odds

ratios with 95 percent credible intervals. All calculations were performed using Meta-Analyst and STATA (Statistics/Data analysis, 10.1) software at 95 percent confidence limits. We assumed publication bias, and did conduct formal statistical tests.

We assessed strength of evidence and judged it according to the domains of risk of bias, consistency, directness, and precision for each major outcome. We defined evidence as strong when several well-designed RCTs with a low risk of bias demonstrated consistent treatment effects. Significant dose-response association or large magnitude of treatment effects increased the level of evidence. We defined evidence as insufficient when only a single study examined treatment effects or associations.

Results

We identified and retrieved 5,185 references. We included 905 references for this review.

Diagnosis of UI

For Key Question 1, 99 studies of 81,043 women provided information on different methods for diagnosing UI. Described use of urodynamic testing as a reference standard test was very similar across the studies. Diagnostic methods to establish a clinical diagnosis of UI were described with different levels of detail and included patient history, physical and pelvic examination, urine culture, and other instrumental measures.

The majority of studies demonstrated that the tests had only small diagnostic value in distinguishing women with urodynamic stress or urgency UI (Table A). The diagnostic values were similar after random effects versus bivariate pooling methods. The quality of the studies did not explain statistical heterogeneity in pooled estimates.

Measuring Treatment Success

Urodynamic evaluation, which was used as a reference method in many diagnostic studies, detects the presence of UI but not the frequency and severity of UI episodes. Validated tools to measure UI treatment success based on meaningful changes in symptoms and quality of life for women include the Incontinence Severity Index; Patient Global Impression of Improvement and of Severity; Patient Perception of Bladder Condition; Urogenital Distress Inventory; Bladder Self-Assessment Questionnaire; International Consultation on Incontinence Modular Questionnaire-SF; Incontinence Impact Questionnaire; Urinary Incontinence-Specific Quality of Life Instrument; King's Health Questionnaire; and

Protection, Amount, Frequency, Adjustment, Body Image assessment tool.

A reduction in UI episode frequency assessed with a 3- to 7-day diary was the most common primary outcome in the included RCTs. Importantly, women with daily stress UI perceived important clinical benefit at reductions of approximately 50 percent and important incremental clinical value at reductions of 75 percent and 90 to 100 percent. Women reported improved quality of life and clinical success only when they experienced a greater than 70 percent reduction in urinary episode frequency assessed by a voiding diary. Smaller decreases (20 to 40 percent) in UI episode frequency were not clinically important when the results from a voiding diary were analyzed in association with the validated Incontinence Quality of Life questionnaire. The quality-of-life impact was similar for stress UI episode reductions of >40 percent to <70 percent. In the case of women with persistent urge, stress, or mixed UI, more than 60 percent reported complete treatment satisfaction on the Global Perception of Improvement and Incontinence Impact Questionnaire when they experienced more than 70 percent reduction in UI episodes according to voiding diaries.

The few RCTs that analyzed differences in outcomes depending on baseline urodynamic diagnosis versus self-reported symptoms of stress, urgency, or mixed UI suggested no advantage with urodynamic diagnosis. However, baseline urodynamic evaluation resulted in better prediction of harms from surgery for stress UI refractory to conservative treatments.

Evidence was insufficient for the superiority of urodynamic evaluation's prediction of nonsurgical treatment outcomes compared to diagnosis based on self-reported symptoms. Women's perceptions of treatment success depend upon clinically important differences in their voiding diaries, scales, questionnaires, and impressions of global improvement.

Efficacy of Pharmacological Treatments

We synthesized the evidence of efficacy and comparative effectiveness of the drugs for predominant stress UI (including topical estrogen and serotonin-noradrenalin uptake inhibitors) and drugs for overactive bladder. Table B demonstrates how many studies were examined for each outcome, how many subjects participated in the studies, and what percentage of subjects experienced the outcomes. The last column indicates our level of confidence that the evidence reflects the true effect of the treatment and that future research is unlikely to change the estimate of

effect (Appendix Table F1 in the full report). Drugs were more effective than placebo in achieving continence and improving UI, but the magnitude of effect was low. The absolute risk difference in continence was less than 20 percent for all drugs. Pharmacological treatments resulted in fewer than 200 cases of continence attributable to the drugs per 1,000 treated. The studies had good quality with low risk of bias. Individual quality criteria and disclosure of conflict of interest were not associated with differences in the results.

Stress UI

Estrogen. Individual RCTs indicated greater continence and improvement in UI with vaginal estrogen formulations and worsening of UI with transdermal patches.

Duloxetine. Duloxetine did not resolve stress UI when compared to placebo (Table B). The risk of adverse effects was significantly higher with duloxetine than with placebo. Duloxetine resulted in improved UI in 75-140 women per 1,000 treated, while 129 women per 1,000 treated stopped taking duloxetine because of adverse effects.

Urgency UI

Oxybutynin. Oxybutynin increased continence rates and improved UI more often than placebo but also resulted in treatment discontinuation due to adverse effects. Oxybutynin resolved UI in 114 women per 1,000 treated (95% CI, 64 to 163), while 63 women per 1,000 treated (95% CI, 12 to 127) discontinued oxybutynin because of adverse effects.

Tolterodine. Tolterodine increased continence rates and significantly improved UI more often than placebo. Tolterodine resolved UI in 85 women per 1,000 treated (95% CI, 40 to 129), while 83 women per 1,000 treated (95% CI, 47 to 120) experienced adverse effects. Discontinuation of treatment due to adverse effects did not differ between tolterodine and placebo.

Darifenacin. Darifenacin significantly improved urgency UI and several domains of quality of life more often than placebo. Darifenacin improved UI in 117 women per 1,000 treated (95% CI 57 to 177), while 190 women per 1,000 treated (95% CI, 118 to 260) experienced adverse effects. Treatment discontinuation rates due to adverse effects did not differ between darifenacin and placebo.

Solifenacin. Solifenacin increased continence rates; higher doses resulted in greater benefits. Treatment discontinuation due to adverse effects was more common with solifenacin than with placebo. Solifenacin resolved UI in 107 women per 1,000 treated (95% CI, 58 to 156),

while 13 women per 1,000 (95% CI, 1 to 26) discontinued treatment because of adverse effects.

Fesoterodine. Fesoterodine increased continence rates. Significant improvement in UI with fesoterodine compared to placebo was dose responsive. Fesoterodine resulted in higher rates of adverse effects and discontinuation of treatment due to adverse effects than placebo. Fesoterodine resolved UI in 130 women per 1,000 treated (95% CI, 58 to 202), while 31 women per 1,000 (95% CI, 10 to 56) stopped treatment due to adverse effects.

Trospium. Trospium increased continence rates more often than placebo. Risk of adverse effects was greater with trospium than with placebo. Trospium resolved UI in 114 women per 1,000 treated (95% CI, 83 to 144), while 18 women per 1,000 (95% CI, 4 to 33) stopped treatment because of harmful adverse effects.

Comparative Effectiveness of Pharmacological Treatments

Evidence of the comparative effectiveness of different drugs was insufficient for the majority of comparisons. Oxybutynin and tolterodine had the same benefits, but tolterodine was safer. The numbers needed to treat (NNT) to achieve continence in one woman were similar across drugs. Treatment discontinuation due to adverse effects was greater than with placebo for all drugs, excluding darifenacin and tolterodine; NNT to achieve discontinuation due to adverse effects was highest with solifenacin (NNT=78) and lowest with oxybutynin (NNT=16). Several retrospective observational studies analyzed the long-term comparative effectiveness and safety of pharmacological treatments for UI. The evidence-based cost utility analysis reported that more than half of patients stop taking drugs for UI after 1 year of treatment. The lowest rates of treatment discontinuation were with 5 mg of solifenacin.¹⁶

Role of Patient Characteristics on Outcomes of Pharmacological Treatments

Age

Treatment response was similar across age groups. Solifenacin increased continence rates more often than placebo, regardless of age.

Oxybutynin, trospium, and darifenacin improved UI in older women. Oxybutynin reduced UI frequency and produced subjective benefits compared to placebo in frail community-dwelling older people. Darifenacin improved UI when compared to placebo in older women. The drug

needed to be given to eight older patients to achieve more than a 50 percent reduction in UI episodes in one person. Cognitive function changes did not differ between darifenacin and placebo in short-term (2-week) treatment. Trospium improved UI and quality of life in older subjects with overactive bladder. Solifenacin caused serious adverse effects less often than oxybutynin in older patients, with no differences between the drugs in younger patients.

Race

We found limited evidence about treatment responses in race subgroups. Only one study, of duloxetine, examined clinical outcomes in different race groups. Evidence was inconclusive about racial differences in the treatment effects of duloxetine in women with stress UI.

Comorbidities

One RCT examined the role of comorbidities. Duloxetine was no better than placebo in women with depression, diabetes, and chronic lung diseases. Trospium was effective in resolving UI regardless of body mass index in obese and normal weight women.

Baseline UI

Evidence was limited from which to conclude any differences in benefits by baseline frequency and severity of UI. Studies found no differences in outcomes between tolterodine and solifenacin in subjects with baseline mixed or pure urgency UI. Subjects with mixed UI may require a larger dose and longer treatment than women with urgency UI to achieve clinical benefits from solifenacin. Inclusion of women with mixed UI did not significantly modify the treatment benefits from oxybutynin and solifenacin across the studies in meta-regression and subgroup analyses.

The baseline frequency of UI did not dramatically modify the effects of the drugs on clinical outcomes. Subjects with more frequent UI had slightly greater benefits with solifenacin or fesoterodine than with placebo. In contrast, trospium was better than placebo at resolving UI only in subjects with fewer than five UI episodes per day. Trospium did not resolve UI in subgroups with more than five episodes of UI per day (relative risk [RR] 1.2, 95% CI, 0.93 to 1.56).

Prior Treatment Response

Solifenacin was effective regardless of the response to previous treatments; however, poor responders did not benefit from increasing the dose of the drug. We could not examine differences in the treatment response to other drugs among those who failed prior treatments because

the studies provided neither subgroup analyses within trials nor consistent reporting of the percentage of nonresponders for subgroup analyses across the trials.

Concomitant Treatments

Trospium reduced the number of urgency UI episodes irrespective of concomitant medications. Adverse effects were more common in those taking seven or more concomitant medications.

Efficacy of Nonpharmacological Treatments

Nonpharmacological treatments were better than no active treatment in achieving continence and improving UI, according to RCTs (Table B). The magnitude of effect was large. The majority of the studies included women with mixed UI. Inclusion of women with mixed UI did not dramatically modify the treatment effects in meta-regression and subgroup analyses. We examined the effects of the interventions on predominant stress or urgency UI when the authors reported that information. A summary of the evidence of effectiveness of all treatments, including strength of evidence, is found in Table B.

Stress UI

Pelvic Floor Muscle Training. Pelvic floor muscle training (PFMT) increased continence rates and improved UI more often than usual care. PFMT combined with bladder training increased continence rates and improved mixed UI. PFMT with biofeedback improved UI.

Vaginal Cones. Evidence was insufficient from which to draw valid conclusions. Uncontrolled high risk of bias studies of other intravaginal and intraurethral devices demonstrated that they improved UI but also resulted in high discontinuation rates and adverse effects.

Intravaginal Electrical Stimulation. Intravaginal electrical stimulation increased continence rates and improved UI more often than sham stimulation.

Magnetic Stimulation. Magnetic stimulation improved UI but did not increase continence more than sham stimulation.

Urgency UI

Bladder Training. Bladder training improved UI when compared to usual care.

Percutaneous Tibial Nerve Stimulation. Percutaneous tibial nerve stimulation improved UI. Individual RCTs indicated no difference in adverse effects and treatment discontinuation with active or sham stimulation.

Mixed UI

Specialized Continence Services. Studies indicated no consistently greater benefits for continence or improvement of UI with continence services implemented by specialized providers compared to usual care. Comparison across studies was difficult because of the variety of interventions that constituted complex continence services.

Weight Loss. Weight loss and exercise improved UI in obese women without evident harms.

Comparative Effectiveness of Nonpharmacological Treatments

Clinical outcomes of one nonpharmacological treatment versus another were reported in 54 RCTs, but these trials rarely compared the same treatment effects, which decreased the strength of evidence to low.

We found no differences in UI between supervised PFMT combined with bladder training and self-administered PFMT. Continence did not differ between bladder training combined with PFMT and bladder training alone.

Indirect comparison indicated the comparable effectiveness of nonpharmacological treatments on continence. Cases of continence achieved per 1,000 treated were 299 for PFMT, 162 for electrical stimulation, and 166 for PFMT combined with bladder training. Rates of continence were comparable with different treatments: 38 percent of women became continent with PFMT, 23 percent became continent with electrical stimulation, and 21 percent became continent with PFMT combined with bladder training.

Discussion

Our findings agree with those of previously published systematic reviews of diagnosis and treatment of UI by AHRQ, the Cochrane Collaborative Group, and the International Consultation on Incontinence. Our report offers a comprehensive analysis of patient-centered outcomes, including continence, improvement in UI, and harms from nonsurgical treatments for female UI that are available in the United States.

Diagnosis of predominant stress or urgency UI in ambulatory care settings includes clinical history and evaluation, voiding diary, and validated scales.¹⁷ Urodynamic diagnosis is more invasive and not applicable to ambulatory settings. Although it more sensitively distinguishes UI mechanisms, including detrusor overactivity, this added sensitivity did not better predict

treatment benefits for patients undergoing nonsurgical UI treatments. It did, however, better predict harms from surgery for women with refractory UI by identifying women with detrusor overactivity, which is associated with greater risk of postsurgical urgency UI, an important quality-of-life outcome.¹⁸ Studies of pharmacological treatments for urgency UI included women treated surgically for stress UI but did not distinguish treatment effects within this subpopulation.¹⁹

Outcome evaluations for treatments of female UI address issues that women consider important: continence, 50 to 70 percent or more reduction in UI episode frequency, meaningful changes in scales measuring quality of life, and treatment satisfaction.²⁰ However, previous reviews of drugs for overactive bladder have focused on other outcomes, such as reduction in frequency of both urgency micturition and urgency UI episodes.^{14,21,22} The majority of drug RCTs were designed to test differences in the frequency of UI episodes. Medical and statistical reviews by the Food and Drug Administration also focused on reduction in the frequency of UI. Based on women's definitions of clinical success, we focused on clinical outcomes, including continence and quality of life.

Policymakers should consider patient-centered outcomes when making regulatory decisions. Research based on patient-centered outcomes provides patients and clinicians the necessary information for effective and informed decisions about health care services.²³ Prescription drugs for UI all demonstrated more effectiveness than placebo in some women. The magnitude of the association was not strong, with fewer than 200 attributable cases of continence per 1,000 patients treated. Adverse effects were common with all drugs and varied between the drugs. Nonpharmacological treatments for UI showed clinically significant benefit with a large magnitude of effect and very few adverse effects.

Direct evidence for the comparative effectiveness of nonpharmacological treatments and drugs was insufficient. However, the few RCTs that compared clinical outcomes between nonpharmacological treatments and drugs found similar effectiveness but better safety with nondrug interventions. This finding is significant, considering that side effects from drugs were common and frequently bothersome enough to negatively affect treatment compliance and continuation. The synthesis of evidence was hampered by differences in definitions of improvement in UI, quality of life, and treatment-related adverse effects. Valid comparisons of benefits and harms with different treatments were possible only for studies that used similar definitions of the outcomes.

While the comparative safety of UI drugs could inform clinical decisions, information on long-term comparative safety was rarely available in RCTs, despite high discontinuation rates suggesting that there were adverse effects. Continuous monitoring of the drugs' adverse effects in clinical practice could provide information about long-term comparative safety. For example, continuous prescription-event monitoring as a part of postmarketing surveillance has provided valuable information about the unfavorable long-term effects of tolterodine, which has been shown to have a significantly higher risk of hallucinations than 10 drugs of other therapeutic classes.²⁴

Additionally, RCTs have not yet examined the role of concurrent treatments, but postmarketing surveillance could address the long-term safety of UI drugs when combined with other medications. For instance, relative risks of ventricular arrhythmias (adjusted RR 5.5, 95% CI, 1.3 to 22.3) or sudden death (adjusted RR 21.5, 95% CI, 5.2 to 88.3) were very high among older people using UI medications in combination with antihistamine/cytochrome inhibitors.²⁵

Meanwhile, very few studies provided evidence for individualized treatment decisions. Evidence of aggregate treatment effects may not be applicable to individuals with specific characteristics.²⁶ An average treatment effect in a clinically diverse population may not reflect the actual effect for a specific group.²⁷ Yet few existing studies examined the role of clinical predictors of treatment failure and success in patient subpopulations.²⁸ Patient comorbidity and baseline severity of UI were associated with differences in treatment benefits. The direction and magnitude of the association varied. Benefits from solifenacin and fesoterodine were greater in those with more than two or three daily episodes of UI; trospium was not better than placebo in those with frequent baseline UI (>5 episodes/day). Which factors are associated with differences in harms remains unclear.

Adherence to UI treatments is poor. Treatment discontinuation due to adverse effects of drugs is common. Yet, very few studies have addressed adherence to treatment, pharmacological or nonpharmacological. Observational economic drug evaluations^{29,30} have demonstrated greater absolute rates of treatment discontinuation due to adverse effects or treatment failure than have been demonstrated in RCTs. One possible explanatory factor for poor adherence is that polypharmacy or previous use of the drugs for urinary tract infections was associated with adherence to the drugs for overactive bladder in California Medicaid program beneficiaries.³¹ Cost-effectiveness analyses^{29,32} that should incorporate

comparative effectiveness, safety, and adherence to treatments were beyond the scope of our review. High discontinuation rates also apply to nonpharmacologic treatments such as PFMT and bladder training. Reasons for poor adherence are not well established.

The nonsurgical treatments included in this review are applicable to ambulatory care settings. Appropriately trained continence nurses and physical therapists can provide high quality UI care for women; women were satisfied with care provided by continence nurses.³³⁻³⁵ A large cross-sectional community survey by mail of women with UI in France, Germany, Spain, and the United Kingdom found that many women actually prefer to be treated for UI by primary care providers, despite easy access to specialized services.³⁶ However, adherence to evidence-based recommendations by ambulatory care providers is not satisfactory and should be improved.^{37,38}

The quality of most drug RCTs was good. The majority of drug studies were double blind with adequate randomization and clear reporting of planned intention-to-treat analysis. Benefits and harms with drugs did not differ by individual quality criteria. We concluded that there was a low risk of bias in the drug studies.

Most nonpharmacological RCTs had good quality. Baseline data demonstrated the adequacy of randomization in the majority of RCTs. Double or single blinding was reported in approximately half of the RCTs. The quality of the studies, including intention-to-treat analysis and adequacy of allocation concealment, did not demonstrate significant modification of the association between treatments and patient outcomes. We concluded that there was a moderate risk of bias in the nonpharmacological studies.

Our review has limitations. We restricted our review to English-language studies published in journals, presented at scientific meetings, reviewed by the Food and Drug Administration,³⁹ or reported on the ClinicalTrials.gov Web site. Even after such an exhaustive review of evidence, we do not know how many funded and unregistered studies we missed in our review. Evidence was insufficient for individualized treatment recommendations by age, race, comorbidity, and baseline UI. Evidence was also insufficient regarding women whose prior treatments had failed. However, previous research has demonstrated that women with stress UI whose conservative treatments failed may benefit from a tension-free vaginal tape procedure.⁴⁰ For women with urgency UI whose conservative treatments failed, percutaneous tibial nerve stimulation,⁴¹ sacral neuromodulation,⁴² and

botulinum toxin injections⁴³ may be of benefit. Invasive treatments, including midurethral slings, sacral nerve stimulation, and radiofrequency ablation, were beyond our scope. We were unable to explain why drug efficacy studies reported substantially different outcome rates for the same comparator placebo treatments. Therefore, we avoided making indirect comparisons of drugs never tested in head-to-head RCTs.

Our report has implications for future research. Such research should clarify which characteristics of women, including age, race, genitourinary characteristics, and comorbidities, are associated with greater treatment benefits and adherence and fewer adverse events. Future studies should assess treatment success with primary outcomes centered on women, including long-term continence, reduction of 50 to 70 percent or more in UI episodes, and clinically important improvement in scales of severity and quality of life. All harms should be analyzed, regardless of investigator judgment about possible association with tested treatments. Nonsurgical treatments for predominant stress UI are limited to PFMT, with very few ongoing studies of bulking agents and devices. Future research should explore new treatment options for women with stress UI. The results from all studies, including 25 closed and 124 ongoing registered studies, should be made available for future reviews of evidence. A comparison of different methods of delivery of nonpharmacological interventions—Internet-based, group-based, and self-management—is also a possible area of future research, with great applicability for ambulatory care populations. Future research should address which factors might increase adherence to UI treatments. Finally, the preventive effects of PFMT, bladder training, and electrical stimulation in premenopausal women should be examined, and future large well-designed head-to-head randomized trials should examine whether combined drug and nonpharmacological treatment modalities are superior to mono-drug therapy.

Key Findings

Diagnosis

- Clinical evaluation with validated tools for diagnosis of UI, its type, frequency, severity, and impact on quality of life informs nonsurgical treatment decisions.
- Compared with diagnosis by patients' symptom reports, multichannel urodynamics did not better predict which patients would benefit from nonsurgical treatments.

Measuring Treatment Success

- Women with daily stress UI perceived important clinical benefit from reductions of approximately 50 percent in UI frequency and important incremental clinical value from reductions of 75 percent and 90 to 100 percent.
- Women reported improved quality of life and clinical success only when they experienced a greater than 70 percent reduction in UI episode frequency assessed by a voiding diary.
- More than 60 percent of women with persistent urgency, stress, or mixed UI reported complete treatment satisfaction when they experienced more than 70 percent reduction of UI episodes. Validated tools have been used to assess threshold values of clinical importance for evaluating treatment success in women.

Pharmacological Treatments

- All anticholinergic medications were more effective than placebo in achieving continence and improving UI, but the degree of benefit was low for all drugs, with fewer than 200 cases of continence attributable to treatment per 1,000 patients treated (absolute risk difference with placebo <20 percent).
- Treatment benefits, including continence, were achieved with antimuscarinic drugs, including trospium, solifenacin, fesoterodine, tolterodine, and oxybutynin.
- Drugs for urgency UI demonstrated similar effectiveness. Treatment discontinuation due to adverse effects was most common with oxybutynin and least common with solifenacin.
- Pharmacological treatments for stress UI, including off-label use of low-dose topical estrogen formulations, may improve stress UI in postmenopausal women.
- Duloxetine has an unfavorable balance between improvement in stress UI and treatment discontinuation due to adverse effects.
- Compliance rates for prescription drugs are low; discontinuation due to side effects is common. Dry mouth, constipation, and blurred vision were among the most frequent adverse effects.
- Evidence is insufficient for the long-term safety of pharmacological treatments.
- Women with urgency UI whose prior treatments failed may benefit from solifenacin; however, poor responders would not benefit from increasing the dose of the drug.

- Oxybutynin, trospium, and darifenacin improved UI in older women.

Nonpharmacological Treatments

- Nonpharmacological treatments result in significant clinical benefit with a low risk of adverse effects. The magnitude of benefit is large, with more than 100 percent relative difference in continence rates.
- Women with stress UI can achieve continence performing PFMT. Continence rates are similar between those who undergo PFMT with and without biofeedback.

Glossary

| | |
|------|--|
| AHRQ | Agency for Healthcare Research and Quality |
| CI | Confidence interval |
| NNT | Number needed to treat |
| PFMT | Pelvic floor muscle training |
| RCT | Randomized controlled trial |
| RR | Relative risk |
| UI | Urinary incontinence |

References

1. Abrams P. Incontinence: 4th International Consultation on Incontinence, Paris, July 5-8, 2008. 4th ed. [Paris]: Health Publications Ltd. 2009, Committee 1 Epidemiology of Urinary (UI) and Fecal (FI) Incontinence and Pelvic Organ Prolapse (POP).
2. Carls C. The prevalence of stress urinary incontinence in high school and college-age female athletes in the midwest: implications for education and prevention. *Urol Nurs*. 2007 Feb;27(1):21-4. 39. PMID 17390923.
3. Kinchen KS, Lee J, Fireman B, et al. The prevalence, burden, and treatment of urinary incontinence among women in a managed care plan. *J Womens Health (Larchmt)*. 2007 Apr;16(3):415-22. PMID 17439386.
4. Boyington JE, Howard DL, Carter-Edwards L, et al. Differences in resident characteristics and prevalence of urinary incontinence in nursing homes in the southeastern United States. *Nurs Res*. 2007;56:97-107. PMID 17356440.
5. Morrison A, Levy R. Fraction of nursing home admissions attributable to urinary incontinence. *Value Health*. 2006 Jul-Aug;9(4):272-4. PMID 16903997.
6. Anger JT, Saigal CS, Madison R, et al. Increasing costs of urinary incontinence among female Medicare beneficiaries. *J Urol*. 2006 Jul;176(1):247-51; discussion 51. PMID 16753411.
7. Shamliyan T, Wyman J, Bliss DZ, et al. Prevention of urinary and fecal incontinence in adults. *Evid Rep Technol Assess (Full Rep)*. 2007 Dec(161):1-379. PMID 18457475.
8. Abrams P. Incontinence: 4th International Consultation on Incontinence, Paris, July 5-8, 2008. 4th ed. [Paris]: Health Publications Ltd.; 2009.

9. 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 Urodyn*. 2010;29(1):4-20. PMID 19941278.
10. Holroyd-Leduc JM, Tannenbaum C, Thorpe KE, et al. What type of urinary incontinence does this woman have? *JAMA*. 2008 Mar 26;299(12):1446-56. PMID 18364487.
11. Abrams P. Incontinence: 4th International Consultation on Incontinence, Paris, July 5-8, 2008: Health Publications Ltd: 2009. Committee 12. Adult Conservative Management.
12. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, et al. Draft Guidance for Industry and FDA Staff - Clinical Investigations of Devices Indicated for the Treatment of Urinary Incontinence. Rockville, MD 20852: Food and Drug Administration, 5630 Fishers Lane, Room 1061; 2008. www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070852.htm. Accessed August 2009.
13. Coyne KS, Sexton CC, Kopp ZS, et al. The impact of overactive bladder on mental health, work productivity and health-related quality of life in the UK and Sweden: results from EpiLUTS. *BJU Int*. 2011 Mar 3. PMID 21371240.
14. Hartmann KE, McPheeters ML, Biller DH, et al. Treatment of overactive bladder in women. *Evid Rep Technol Assess (Full Rep)*. 2009 Aug(187):1-120, v. PMID 19947666.
15. Shamliyan TA, Kane RL, Wyman J, et al. Systematic review: randomized, controlled trials of nonsurgical treatments for urinary incontinence in women. *Ann Intern Med*. 2008 Mar 18;148(6):459-73. PMID 18268288.
16. Cardozo L, Thorpe A, Warner J, et al. 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 Feb 3. PMID 20132203.
17. Martin JL, Williams KS, Sutton AJ, et al. Systematic review and meta-analysis of methods of diagnostic assessment for urinary incontinence. *Neurourol Urodyn*. 2006;25:674-83. PMID 17016795.
18. Dmochowski RR, Blaivas JM, Gormley EA, et al. Update of AUA guideline on the surgical management of female stress urinary incontinence. *J Urol*. 2010 May;183(5):1906-14. PMID 20303102.
19. National Collaborating Centre for Women's and Children's Health. Urinary incontinence: The management of urinary incontinence in women Commissioned by the National Institute for Health and Clinical Excellence. October 2006. www.nice.org.uk/nicemedia/pdf/CG40NICEguideline.pdf.
20. Ghoniem G, Stanford E, Kenton K, et al. Evaluation and outcome measures in the treatment of female urinary stress incontinence: International Urogynecological Association (IUGA) guidelines for research and clinical practice. *International Urogynecology Journal*. 2008 Jan;19(1):5-33. PMID 18026681.
21. Campbell JD, Gries KS, Watanabe JH, et al. Treatment success for overactive bladder with urinary urge incontinence refractory to oral antimuscarinics: a review of published evidence. *BMC Urol*. 2009;9:18. PMID 19930578.
22. McDonagh MS, Selover D, Santa J, et al. Drug class review on agents for overactive bladder: Final report Oregon Health & Science University. Dec 2005.
23. Agency for Healthcare Research and Quality. HHS Awards \$473 Million in Patient-Centered Outcomes Research Funding. Agency for Healthcare Research and Quality, Rockville, MD; 2010.
24. Layton D, Pearce GL, Shakir SA. Safety profile of tolterodine as used in general practice in England: results of prescription-event monitoring. *Drug Saf*. 2001;24(9):703-13. PMID 11522122.
25. Wang PS, Levin R, Zhao SZ, et al. Urinary antispasmodic use and the risks of ventricular arrhythmia and sudden death in older patients. *J Am Geriatr Soc*. 2002 Jan;50(1):117-24. PMID 12028256.
26. Ioannidis JPA, Lau J. Heterogeneity of the baseline risk within patient populations of = clinical trials: a proposed evaluation algorithm. *Am J Epidemiol*. 1998;148:1117-26.
27. Arends LR, Hoes AW, Lubsen J, et al. Baseline risk as predictor of treatment benefit: three clinical = meta-re-analyses. *Stat Med*. 2000;19:3497-518.
28. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ*. 1994;309:1351-5.
29. Ko Y, Malone DC, Armstrong EP. Pharmacoeconomic evaluation of antimuscarinic agents for the treatment of overactive bladder. *Pharmacotherapy*. 2006 Dec;26(12):1694-702. PMID 17125433.
30. Woodward A, Matthews C, Lamb E, et al. Prescribing antimuscarinics for overactive bladder; how many chances do we have to get it right? *Neurourology and Urodynamics*; 2010; Joint Meeting of the International Continence Society and the International Urogynecological Association, Toronto, Canada, 23-27 August 2010:29.
31. Yu YF, Nichol MB, Yu AP, et al. Persistence and adherence of medications for chronic overactive bladder/urinary incontinence in the California Medicaid program. *Value Health*. 2005 Jul-Aug;8(4):495-505. PMID 16091027.
32. Perfetto EM, Subedi P, Jumadilova Z. Treatment of overactive bladder: a model comparing extended-release formulations of tolterodine and oxybutynin. *Am J Manag Care*. 2005 Jul;11(4 Suppl):S150-7. PMID 16161388.
33. Laurant M, Reeves D, Hermens R, et al. Substitution of doctors by nurses in primary care. *Cochrane Database Syst Rev*. 2005(2):CD001271. PMID 15846614.
34. Festen L, Duggan P, Coates D. Improved quality of life in women treated for urinary incontinence by an authorised continence nurse practitioner. *Int Urogynecol J Pelvic Floor Dysfunct*. 2008 Apr;19(4):567-71. PMID 17898919.
35. Jeffery S, Doumouchtsis SK, Fynes M. Patient satisfaction with nurse-led telephone follow-up in women with lower urinary tract symptoms. *J Telemed Telecare*. 2007;13(7):369-73. PMID 17958940.
36. O'Donnell M, Viktrup L, Hunskaar S. The role of general practitioners in the initial management of women with urinary incontinence in France, Germany, Spain and the UK. *Eur J Gen Pract*. 2007;13(1):20-6. PMID 17366290.

37. Steel N, Bachmann M, Maisey S, et al. Self reported receipt of care consistent with 32 quality indicators: national population survey of adults aged 50 or more in England. *BMJ*. 2008;337:a957. PMID 18703659.
38. Wagg A, Potter J, Peel P, et al. National audit of continence care for older people: management of urinary incontinence. *Age Ageing*. 2008 Jan;37(1):39-44. PMID 18033776.
39. Wieseler B, McGauran N, Kaiser T. Finding studies on reboxetine: a tale of hide and seek. *BMJ*. 2010;341:c4942. PMID 20940211.
40. Imamura M, Abrams P, Bain C, et al. Systematic review and economic modelling of the effectiveness and cost-effectiveness of non-surgical treatments for women with stress urinary incontinence. *Health Technol Assess*. 2010 Aug;14(40):1-188, iii-iv. PMID 20738930.
41. National Institute for Health and Clinical Excellence. Percutaneous posterior tibial nerve stimulation for overactive bladder syndrome: guidance. Oct 27 2010. www.nice.org.uk/nicemedia/live/12412/51304/51304.pdf.
42. Goode PS, Burgio KL, Richter HE, et al. Incontinence in older women. *JAMA*. 2010 Jun 2;303(21):2172-81. PMID 20516418.
43. Thuroff JW, Abrams P, Andersson KE, et al. EAU Guidelines on Urinary Incontinence. *Eur Urol*. 2011 Mar;59(3):387-400. PMID 21130559.

Full Report

This executive summary is part of the following document: Shamliyan T, Wyman J, Kane RL. Nonsurgical Treatments for Urinary Incontinence in Adult Women: Diagnosis and Comparative Effectiveness. Comparative Effectiveness Review No. 36. (Prepared by the University of Minnesota Evidence-based Practice Center under Contract No. HHS 290-2007-10064-I.) AHRQ Publication No. 11(12)-EHC074-EF. Rockville, MD. Agency for Healthcare Research and Quality. April 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

For More Copies

For more copies of Nonsurgical Treatments for Urinary Incontinence in Adult Women: Diagnosis and Comparative Effectiveness: Comparative Effectiveness Review Executive Summary No. 36 (AHRQ Pub. No. 11(12)-EHC074-1), please call the AHRQ Publications Clearinghouse at 1-800-358-9295 or email ahrqpubs@ahrq.gov.

Table A. Diagnostic value of the test for urinary incontinence (UI) in women (pooled with random effects models and bivariate pooling)

| Type of Incontinence | Method Index | Reference Standard | # of Studies # of Subjects | Sensitivity/ Bivariate Pooling | Specificity/ Bivariate Pooling | Positive Likelihood Ratio ¹ | Negative Likelihood Ratio ¹ | Positive Predictive Value | Negative Predictive Value |
|------------------------------------|-----------------------------------|--------------------|-------------------------------|---|---|--|--|---------------------------------|---------------------------------|
| Urodynamic stress UI | Symptoms of stress UI | Urodynamic test | 27 5,780 | 0.93 ² (0.90 to 0.95) 0.94 (0.91 to 0.96) | 0.41 ² (0.34 to 0.49) 0.41 (0.31 to 0.51) | 1.54 (1.40 to 1.7) | 0.20 (0.14 to 0.27) | 0.74 (0.68 to 0.80) | 0.74 (0.67 to 0.81) |
| Detrusor overactivity | Symptoms of urgency UI | Urodynamic test | 23 5,485 | 0.82 ² (0.76 to 0.87) 0.82 (0.75 to 0.88) | 0.51 ² (0.44 to 0.59) 0.52 (0.40 to 0.65) | 1.54 (1.38 to 1.73) | 0.39 (0.30 to 0.50) | 0.56 (0.48 to 0.63) | 0.80 (0.73 to 0.86) |
| Detrusor overactivity | Symptoms of urgency | Urodynamic test | 9 6,418 | 0.84 ² (0.59 to 0.95) 0.82 (0.70 to 0.92) | 0.39 ² (0.17 to 0.67) 0.39 (0.24 to 0.55) | 1.36 (1.18 to 1.58) | 0.47 (0.33 to 0.67) | 0.48 (0.39 to 0.57) | 0.75 (0.67 to 0.81) |
| Detrusor overactivity ³ | Symptoms of urgency UI | Urodynamic test | 17 3,924 | 0.84 ² (0.78 to 0.89) 0.84 (0.79 to 0.90) | 0.43 ² (0.36 to 0.50) 0.44 (0.34 to 0.54) | 1.48 (1.31 to 1.66) | 0.40 (0.29 to 0.54) | 0.33 (0.26 to 0.41) | 0.89 (0.83 to 0.93) |
| Detrusor overactivity ³ | Symptoms of urgency | Urodynamic test | 6 1,598 | 0.86 (0.83 to 0.89) 0.86 (0.80 to 0.90) | 0.31 ² (0.24 to 0.39) 0.31 (0.20 to 0.45) | 1.21 (1.11 to 1.32) | 0.523 (0.41 to 0.67) | 0.27 (0.17 to 0.40) | 0.86 (0.76 to 0.93) |
| Mixed UI | Symptoms of stress and urgency UI | Urodynamic test | 11 2,767 | 0.73 ² (0.61 to 0.82) 0.72 (0.58 to 0.83) | 0.53 ² (0.40 to 0.66) 0.53 (0.34 to 0.72) | 1.45 (1.27 to 1.67) | 0.61 (0.52 to 0.71) | 0.26 (0.20 to 0.34) | 0.89 (0.85 to 0.92) |
| Urodynamic stress UI | Pad test | Urodynamic test | 3 574 | 0.84 (0.76 to 0.90) 0.83 (0.75 to 0.91) | 0.77 (0.72 to 0.82) 0.77 (0.17 to 0.97) | 3.62 (2.88 to 4.57) | 0.22 (0.15 to 0.32) | 0.82 (0.77 to 0.86) | 0.78 (0.73 to 0.83) |
| Detrusor overactivity | Pad | Urodynamic test | 2 469 | 0.72 ² (0.30 to 0.94) | 0.56 ² (0.38 to 0.72) | 1.56 (0.62 to 3.90) | 0.47 (0.10 to 2.33) | 0.32 (0.04 to 0.83) | 0.88 (0.83 to 0.91) |

Table A. Diagnostic value of the test for urinary incontinence (UI) in women (pooled with random effects models and bivariate pooling) (continued)

| Type of Incontinence | Method Index | Reference Standard | # of Studies # of Subjects | Sensitivity/ Bivariate Pooling | Specificity/ Bivariate Pooling | Positive Likelihood Ratio ¹ | Negative Likelihood Ratio ¹ | Positive Predictive Value | Negative Predictive Value |
|-----------------------|-----------------------------------|--------------------|-------------------------------|---|---|--|--|---------------------------------|---------------------------------|
| Urodynamic stress UI | Symptoms of stress UI | Clinical diagnosis | 5 947 | 0.88 ² (0.68 to 0.96) 0.86 (0.70 to 0.96) | 0.67 ² (0.54 to 0.78) 0.67 (0.51 to 0.81) | 2.35 (1.97 to 2.81) | 0.19 (0.09 to 0.41) | 0.80 (0.66 to 0.89) | 0.75 (0.58 to 0.87) |
| Detrusor overactivity | Symptoms of urgency UI | Clinical diagnosis | 4 735 | 0.82 ² (0.73 to 0.89) 0.82 (0.73 to 0.90) | 0.67 ² (0.53 to 0.79) 0.67 (0.45 to 0.86) | 2.52 (1.81 to 3.50) | 0.26 (0.18 to 0.38) | 0.72 (0.48 to 0.88) | 0.79 (0.54 to 0.92) |
| Mixed UI | Symptoms of stress and urgency UI | Clinical diagnosis | 3 654 | 0.65 ² (0.36 to 0.86) 0.64 (0.38 to 0.85) | 0.54 ² (0.21 to 0.84) 0.52 (0.06 to 0.94) | 1.57 (0.68 to 3.59) | 0.74 (0.28 to 1.95) | 0.36 (0.27 to 0.47) | 0.80 (0.43 to 0.96) |
| Urodynamic stress UI | Q-tip test | Urodynamic test | 3 267 | 0.62 (0.53 to 0.70) 0.62 (0.49 to 0.74) | 0.60 ² (0.40 to 0.78) 0.58 (0.00 to 1.00) | 1.70 (0.89 to 3.23) | 0.60 (0.31 to 1.17) | 0.58 (0.26 to 0.85) | 0.67 (0.34 to 0.89) |

¹Clinical interpretations of likelihood ratios:

Likelihood ratio Interpretation

- >10 Large and often conclusive increase in the likelihood of disease
- 5-10 Moderate increase in the likelihood of disease
- 2-5 Small increase in the likelihood of disease
- 1-2 Minimal increase in the likelihood of disease
- 1 No change in the likelihood of disease

²Significant heterogeneity

³Pure type

**Table B. Clinical outcomes with treatments for UI
(pooled with random effects estimates from head-to-head RCTs)**

| Treatments | Outcomes | Number of Studies | Patients | Rate, % Active/Control | Relative Risk (95% CI) | Absolute Risk Difference* (95% CI) | Number Needed to Treat (95% CI) | Attributable Events (95% CI) | Effect in Relative/Absolute Scale | Evidence |
|--|--|-------------------|----------|------------------------|------------------------|------------------------------------|---------------------------------|------------------------------|-----------------------------------|----------|
| Pharmacological Treatments for Stress UI | | | | | | | | | | |
| Duloxetine vs. placebo | Continence | 2 | 736 | 38/40 | 0.92 (0.86 to 0.99) | -0.03 (-0.12 to 0.06) | | | ↓/NS | Low |
| Duloxetine vs. placebo | Improved UI | 4 | 1,138 | 37/29 | 1.68 (0.94 to 3.00) | 0.08 (0.01 to 0.14) | 13 (7 to 143) | 75 (7 to 142) | NS/↑ | High |
| Duloxetine vs. placebo | Discontinuation due to adverse effects | 9 | 3,252 | 16/3 | 4.4 (3.24 to 5.86) | 0.13 (0.06 to 0.19) | 8 (5 to 16) | 129 (64 to 193) | ↑ | High |
| Pharmacological Treatments for Urgency UI | | | | | | | | | | |
| Darifenacin vs. placebo | Improved UI | 3 | 1,011 | 48/33 | 1.3 (1.2 to 1.5) | 0.12 (0.06 to 0.17) | 9 (6 to 18) | 117 (57 to 177) | ↑ | High |
| Darifenacin vs. placebo | Discontinuation due to adverse effects | 7 | 3,138 | 5/3 | 1.2 (0.8 to 1.8) | 0.00 (-0.01 to 0.02) | | | NS | High |
| Darifenacin vs. placebo | Discontinuation due to failure | 4 | 1,280 | 1/2 | 0.6 (0.2 to 1.7) | -0.01 (-0.02 to 0.01) | | | NS | Moderate |
| Fesoterodine vs. placebo | Continence | 2 | 2,465 | 61/48 | 1.3 (1.1 to 1.5) | 0.13 (0.06 to 0.20) | 8 (5 to 17) | 130 (58 to 202) | ↑ | Low |
| Fesoterodine vs. placebo | Improved UI | 2 | 1,896 | 42/32 | 1.3 (1.2 to 1.5) | 0.10 (0.06 to 0.15) | 10 (7 to 18) | 100 (56 to 145) | ↑ | High |
| Fesoterodine vs. placebo | Adverse effects | 4 | 4,145 | 51/38 | 1.4 (1.2 to 1.6) | 0.16 (0.11 to 0.20) | 6 (5 to 9) | 156 (112 to 200) | ↑ | High |
| Fesoterodine vs. placebo | Discontinuation due to adverse effects | 4 | 4,433 | 6/3 | 2.0 (1.3 to 3.1) | 0.03 (0.01 to 0.06) | 33 (18 to 102) | 31 (10 to 56) | ↑ | High |

Table B. Clinical outcomes with treatments for UI (pooled with random effects estimates from head-to-head RCTs) (continued)

| Treatments | Outcomes | Number of Studies | Patients | Rate, % Active/Control | Relative Risk (95% CI) | Absolute Risk Difference* (95% CI) | Number Needed to Treat (95% CI) | Attributable Events (95% CI) | Effect in Relative/Absolute Scale | Evidence |
|--------------------------|--|-------------------|----------|------------------------|------------------------|------------------------------------|---------------------------------|------------------------------|-----------------------------------|----------|
| Fesoterodine vs. placebo | Discontinuation due to failure | 2 | 1,896 | 2/3 | 0.6 (0.2;2.5) | -0.01 (-0.03 to 0.02) | | | NS | Moderate |
| Oxybutynin vs. placebo | Continence | 4 | 992 | 27/16 | 1.7 (1.3 to 2.1) | 0.11 (0.06 to 0.16) | 9 (6 to 16) | 114 (64 to 163) | ↑ | High |
| Oxybutynin vs. placebo | Improved UI | 9 | 1,244 | 53/32 | 1.5 (1.2 to 1.9) | 0.17 (0.10 to 0.24) | 6 (4 to 11) | 167 (95 to 240) | ↑ | Moderate |
| Oxybutynin vs. placebo | Discontinuation due to adverse effects | 5 | 1,483 | 10/5 | 1.7 (1.1 to 2.5) | 0.06 (0.01 to 0.13) | 16 (8 to 86) | 63 (12 to 127) | ↑ | High |
| Propiverine vs. placebo | Continence | 2 | 691 | 53/37 | 1.4 (1.2 to 1.7) | 0.16 (0.09 to 0.24) | 6 (4 to 12) | 163 (86 to 239) | ↑ | Low |
| Propiverine vs. placebo | Improved UI | 3 | 985 | 55/35 | 1.6 (1.3 to 2.0) | 0.19 (0.13 to 0.25) | 5 (4 to 8) | 192 (132 to 252) | ↑ | Moderate |
| Propiverine vs. placebo | Discontinuation due to adverse effects | 2 | 1,401 | 5/2 | 2.6 (1.4 to 5.00) | 0.03 (0.01 to 0.06) | 29 (16 to 77) | 34 (13 to 61) | ↑ | Low |
| Solifenacin vs. placebo | Continence | 5 | 6,304 | 39/28 | 1.5 (1.4 to 1.6) | 0.11 (0.06 to 0.16) | 9 (6 to 17) | 107 (58 to 156) | ↑ | High |
| Solifenacin vs. placebo | Improved UI | 2 | 1,507 | 60/42 | 1.5 (1.0 to 2.1) | 0.18 (0.10 to 0.26) | 6 (4 to 10) | 180 (97 to 263) | ↑ | Low |
| Solifenacin vs. placebo | Adverse effects | 3 | 1,713 | 52/36 | 1.7 (1.2 to 2.4) | 0.18 (0.09 to 0.27) | 6 (4 to 12) | 177 (85 to 267) | ↑ | High |
| Solifenacin vs. placebo | Discontinuation due to adverse effects | 7 | 9,080 | 5/4 | 1.3 (1.1 to 1.7) | 0.01 (0.00 to 0.03) | 78 (39 to 823) | 13 (1 to 26) | ↑ | High |
| Solifenacin vs. placebo | Discontinuation due to failure | 4 | 2,812 | 2/1 | 1.0 (0.5 to 1.8) | 0.00 (-0.01 to 0.01) | | | NS | Moderate |

**Table B. Clinical outcomes with treatments for UI
(pooled with random effects estimates from head-to-head RCTs) (continued)**

| Treatments | Outcomes | Number of Studies | Patients | Rate, % Active/Control | Relative Risk (95% CI) | Absolute Risk Difference* (95% CI) | Number Needed to Treat (95% CI) | Attributable Events (95% CI) | Effect in Relative/Absolute Scale | Evidence |
|------------------------------|--|-------------------|----------|------------------------|------------------------|------------------------------------|---------------------------------|------------------------------|-----------------------------------|----------|
| Tolterodine vs. placebo | Continence | 4 | 3,404 | 53/44 | 1.2 (1.1 to 1.4) | 0.09 (0.04 to 0.13) | 12 (8 to 25) | 85 (40 to 129) | ↑ | High |
| Tolterodine vs. placebo | Improved UI | 7 | 6,119 | 45/37 | 1.3 (1.1 to 1.4) | 0.10 (0.04 to 0.15) | 10 (7 to 24) | 96 (42 to 149) | ↑ | High |
| Tolterodine vs. placebo | Adverse effects | 12 | 4,162 | 45/38 | 1.2 (1.1 to 1.3) | 0.08 (0.05 to 0.12) | 12 (8 to 21) | 83 (47 to 120) | ↑ | High |
| Tolterodine vs. placebo | Discontinuation due to adverse effects | 10 | 4,466 | 4/3 | 1.0 (0.6 to 1.7) | 0.01 (-0.01 to 0.03) | | | NS | High |
| Tolterodine vs. placebo | Discontinuation due to failure | 5 | 4,049 | 1/2 | 0.5 (0.2 to 0.9) | -0.01 (-0.01 to 0.00) | | | NS | High |
| Trospium vs. placebo | Continence | 4 | 2,677 | 28/17 | 1.7 (1.5 to 2.0) | 0.11 (0.08 to 0.14) | 9 (7 to 12) | 114 (83 to 144) | ↑ | High |
| Trospium vs. placebo | Improved UI | 2 | 1,176 | 32/25 | 1.1 (0.6 to 2.0) | 0.08 (-0.10 to 0.25) | | | NS | Low |
| Trospium vs. placebo | Adverse effects | 5 | 2,967 | 41/29 | 1.4 (1.2 to 1.7) | 0.12 (0.09 to 0.16) | 8 (6 to 11) | 123 (88 to 159) | ↑ | Moderate |
| Trospium vs. placebo | Discontinuation due to adverse effects | 6 | 3,936 | 6/4 | 1.5 (1.1 to 1.9) | 0.02 (0.00 to 0.03) | 56 (30 to 228) | 18 (4 to 33) | ↑ | High |
| Fesoterodine vs. tolterodine | Continence | 2 | 3,312 | 61/56 | 1.10 (1.04 to 1.16) | 0.06 (0.02 to 0.09) | 18 (11 to 48) | 55 (21 to 88) | ↑ | Low |
| Fesoterodine vs. tolterodine | Improved UI | 3 | 4,425 | 44/35 | 1.06 (1; 1.2) | 0.03 (0; 0.06) | 36 (17 to 1,000) | 28 (1 to 57) | ↑/↑ | High |
| Fesoterodine vs. tolterodine | Discontinuation due to adverse effects | 4 | 4,440 | 5/4 | 1.54 (1.21 to 1.97) | 0.02 (0.01 to 0.03) | 58 (33 to 206) | 17 (5 to 31) | ↑ | Moderate |

**Table B. Clinical outcomes with treatments for UI
(pooled with random effects estimates from head-to-head RCTs) (continued)**

| Treatments | Outcomes | Number of Studies | Patients | Rate, % Active/Control | Relative Risk (95% CI) | Absolute Risk Difference* (95% CI) | Number Needed to Treat (95% CI) | Attributable Events (95% CI) | Effect in Relative/Absolute Scale | Evidence |
|--|--|-------------------|----------|------------------------|------------------------|------------------------------------|---------------------------------|------------------------------|-----------------------------------|----------|
| Oxybutynin vs. tolterodine | Improved UI | 3 | 947 | 50/45 | 1.11 (0.94 to 1.31) | 0.05 (-0.03 to 0.13) | | | NS | Moderate |
| Oxybutynin vs. tolterodine | Discontinuation due to adverse effects | 6 | 2,323 | 13/6 | 1.9 (1.1 to 3.3) | 0.07 (0.01 to 0.15) | 14 (7 to 145) | 72 (7 to 154) | ↑ | High |
| Solifenacin vs. tolterodine | Discontinuation due to adverse effects | 3 | 2,755 | 4/3 | 1.28 (0.86 to 1.91) | 0.01 (0.00 to 0.03) | | | NS | Moderate |
| Trospium vs. oxybutynin | Discontinuation due to adverse effects | 2 | 2,015 | 5/7 | 0.75 (0.52; 1.1) | 0.00 (-0.03 to 0.05) | | | NS | Low |
| Nonpharmacological Treatments | | | | | | | | | | |
| Bladder training vs. no active treatment | Improved UI | 2 | 283 | 61.4/19.2 | 3.22 (2.25 to 4.60) | 0.43 (0.28 to 0.59) | 2 (2 to 4) | 430 (275 to 585) | ↑ | Low |
| Continence service vs. no active treatment | Continence | 3 | 3,939 | 29/20 | 1.6 (1.1 to 2.3) | 0.30 (-0.01 to 0.60) | | | ↑/NS | Moderate |
| Continence service vs. no active treatment | Improved UI | 2 | 4,038 | 62.6/53.5 | 1.33 (1.06 to 1.68) | 0.20 (-0.01 to 0.41) | | | ↑/NS | Low |
| Electrical stimulation vs. no active treatment | Continence | 7 | 420 | 23/8 | 2.9 (1.6 to 5.2) | 0.16 (0.06 to 0.26) | 6 (4 to 16) | 162 (64 to 259) | ↑ | High |

**Table B. Clinical outcomes with treatments for UI
(pooled with random effects estimates from head-to-head RCTs) (continued)**

| Treatments | Outcomes | Number of Studies | Patients | Rate, % Active/Control | Relative Risk (95% CI) | Absolute Risk Difference* (95% CI) | Number Needed to Treat (95% CI) | Attributable Events (95% CI) | Effect in Relative/Absolute Scale | Evidence |
|---|-----------------|--------------------------|-----------------|-------------------------------|-------------------------------|---|--|-------------------------------------|--|-----------------|
| Electrical stimulation vs. no active treatment | Improved UI | 8 | 582 | 31.7/15.1 | 2.01 (1.28 to 3.15) | 0.16 (0.08 to 0.23) | 6 (4 to 12) | 156 (84 to 228) | ↑ | High |
| Magnetic stimulation vs. no active treatment | Improved UI | 3 | 153 | 46.8/21.2 | 2.30 (1.43 to 3.71) | 0.27 (0.11 to 0.42) | 4 (2 to 9) | 265 (112 to 417) | ↑ | Moderate |
| Magnetic stimulation vs. no active treatment | Continence | 3 | 171 | 30.7/17.8 | 1.22 (0.78 to 1.88) | 0.09 (-0.01 to 0.18) | | | NS | Moderate |
| Percutaneous electrical stimulation vs. no active treatment | Improved UI | 3 | 405 | 40/20 | 1.9 (1.1 to 3.2) | 0.31 (0.04 to 0.58) | 3 (2 to 25) | 308 (40 to 577) | ↑ | Moderate |
| PFMT vs. no active treatment | Continence | 10 | 959 | 38/12 | 3.8 (2.1 to 6.8) | 0.30 (0.19 to 0.41) | 3 (2 to 5) | 299 (188 to 410) | ↑ | High |
| PFMT vs. no active treatment | Improved UI | 6 | 510 | 56.9/14.7 | 5.44 (1.57 to 8.83) | 0.41 (0.17 to 0.65) | 2 (2 to 6) | 412 (174 to 649) | ↑ | High |
| PFMT with bladder training vs. no active treatment | Continence | 5 | 1,369 | 21/12 | 3.8 (1.5 to 9.3) | 0.17 (0.06 to 0.27) | 6 (4 to 16) | 166 (63 to 268) | ↑ | High |

**Table B. Clinical outcomes with treatments for UI
(pooled with random effects estimates from head-to-head RCTs) (continued)**

| Treatments | Outcomes | Number of Studies | Patients | Rate, % Active/Control | Relative Risk (95% CI) | Absolute Risk Difference* (95% CI) | Number Needed to Treat (95% CI) | Attributable Events (95% CI) | Effect in Relative/Absolute Scale | Evidence |
|--|-----------------|--------------------------|-----------------|-------------------------------|-------------------------------|---|--|-------------------------------------|--|-----------------|
| PFMT with bladder training vs. no active treatment | Improved UI | 4 | 1,171 | 53.3/22.5 | 4.13 (1.58 to 0.78) | 0.39 (0.17 to 0.60) | 3 (2 to 6) | 387 (171 to 603) | ↑ | High |
| PFMT with biofeedback vs. no active treatment | Continence | 2 | 185 | 42/2 | 11.2 (2.2 to 56.4) | 0.49 (-0.10 to 1.08) | | | | Low |
| PFMT with biofeedback vs. no active treatment | Improved UI | 4 | 383 | 60.1/18.6 | 3.93 (1.00 to 5.49) | 0.39 (0.17 to 0.61) | 3 (2 to 6) | 390 (170 to 610) | ↑ | High |
| Weight Loss vs. no active treatment | Improved UI | 2 | 386 | 42.8/20.8 | 2.17 (1.26 to 3.76) | 0.27 (0.06 to 0.50) | 4 (2 to 18) | 273 (57 to 490) | ↑ | Moderate |
| PFMT + bladder training vs. bladder training | Continence | 2 | 271 | 21/21 | 1 (0.4 to 2.8) | 0.001 (-0.2 to 0.2) | | | NS | High |
| PFMT vs. electrical stimulation | Continence | 3 | 99 | 24/29 | 0.85 (0.45 to 1.61) | -0.04 (-0.20 to 0.11) | | | NS | Moderate |
| PFMT vs. electrical stimulation | Improved UI | 4 | 136 | 31/45 | 0.97 (0.62 to 1.51) | -0.01 (-0.17 to 0.16) | | | NS | Moderate |
| PFMT vs. vaginal cone | Continence | 3 | 320 | 22/27 | 0.78 (0.58 to 1.06) | -0.11 (-0.26 to 0.04) | | | NS | Moderate |

Table B. Clinical outcomes with treatments for UI (pooled with random effects estimates from head-to-head RCTs) (continued)

| Treatments | Outcomes | Number of Studies | Patients | Rate, % Active/Control | Relative Risk (95% CI) | Absolute Risk Difference* (95% CI) | Number Needed to Treat (95% CI) | Attributable Events (95% CI) | Effect in Relative/Absolute Scale | Evidence |
|--------------------------------|-------------|-------------------|----------|------------------------|------------------------|------------------------------------|---------------------------------|------------------------------|-----------------------------------|----------|
| PFMT vs. vaginal cone | Improved UI | 4 | 440 | 41/41 | 1.02 (0.91 to 1.14) | 0.01 (-0.08 to 0.09) | | | NS | Moderate |
| PFMT with biofeedback vs. PFMT | Continence | 6 | 542 | 30/25 | 1.27 (0.88 to 1.85) | 0.08 (-0.03 to 0.19) | | | NS | High |
| Supervised PFMT vs. self-PFMT | Continence | 4 | 300 | 35/22 | 1.92 (0.87 to 4.23) | 0.20 (-0.03 to 0.43) | | | NS | High |
| Supervised PFMT vs. self-PFMT | Improved UI | 4 | 283 | 50/33 | 1.51 (0.85 to 2.67) | 0.14 (-0.05 to 0.32) | | | NS | Moderate |

CI = confidence interval; PFMT = pelvic floor muscle training; NS = not significant; RCT = randomized controlled trial; UI = urinary incontinence; ↑ = effect of active drug is greater than control; ↓ = effect of active drug is lower than control

*Risk differences for drug adverse effects were calculated using arcsine transformation.

