Comparisons of approaches to pelvic floor muscle training for urinary incontinence in women (Review)

Hay-Smith EJC, Herderschee R, Dumoulin C, Herbison GP

THE COCHRANE COLLABORATION®

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2011, Issue 12

http://www.thecochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Comparisons of approaches to pelvic floor muscle training for urinary incontinence in women

E. Jean C Hay-Smith[1], Roselien Herderschee[2], Chantale Dumoulin[3], G Peter Herbison[4]

[1]Rehabilitation Teaching and Research Unit, Department of Medicine, University of Otago, Wellington, New Zealand. [2]Department of Obstetrics & Gynaecology Academic Medical Centre, University of Amsterdam, Amsterdam, Netherlands. [3]School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Canada. [4]Department of Preventive & Social Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

Contact address: E. Jean C Hay-Smith, Rehabilitation Teaching and Research Unit, Department of Medicine, University of Otago, Wellington, New Zealand. jean.hay-smith@otago.ac.nz.

Editorial group: Cochrane Incontinence Group.
Publication status and date: New, published in Issue 12, 2011.
Review content assessed as up-to-date: 11 July 2011.

Citation: Hay-Smith EJC, Herderschee R, Dumoulin C, Herbison GP. Comparisons of approaches to pelvic floor muscle training for urinary incontinence in women. *Cochrane Database of Systematic Reviews* 2011, Issue 12. Art. No.: CD009508. DOI: 10.1002/14651858.CD009508.

ABSTRACT

Background

Pelvic floor muscle training is the most commonly recommended physical therapy treatment for women with stress urinary incontinence. It is also sometimes recommended for mixed and, less commonly, urge urinary incontinence. The supervision and content of pelvic floor muscle training programmes are highly variable, and some programmes use additional strategies in an effort to increase adherence or training effects.

Objectives

To compare the effects of different approaches to pelvic floor muscle training for women with urinary incontinence.

Search methods

We searched the Cochrane Incontinence Group Specialised Trials Register, which contains trials identified from the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE and CINAHL, and handsearching of journals and conference proceedings (searched 17 May 2011), and the reference lists of relevant articles.

Selection criteria

Randomised trials or quasi-randomised trials in women with stress, urge or mixed urinary incontinence (based on symptoms, signs or urodynamics). One arm of the study included pelvic floor muscle training. Another arm was an alternative approach to pelvic floor muscle training, such as a different way of teaching, supervising or performing pelvic floor muscle training.

Data collection and analysis

We independently assessed trials for eligibility and methodological quality. We extracted then cross-checked data. We resolved disagreements by discussion. We processed data as described in the *Cochrane Handbook for Systematic Reviews of Interventions* (version 5.2.2). We subgrouped trials by intervention.

Main results

We screened 574 records for eligibility and included 21 trials in the review. The 21 trials randomised 1490 women and addressed 11 comparisons. These were: differences in training supervision (amount, individual versus group), in approach (one versus another, the effect of an additional component) and the exercise training (type of contraction, frequency of training). In women with stress urinary incontinence, 10% of those who received weekly or twice-weekly group supervision in addition to individual appointments with the therapist did not report improvement post-treatment compared to 43% of the group who had individual appointments only (risk ratio (RR) for no improvement 0.29, 95% confidence interval (CI) 0.15 to 0.55, four trials). Looking at this another way, 90% of those who had combined group and individual supervision reported improvement versus 57% of women receiving individual supervision only. While women receiving the combination of frequent group supervision and individual supervision of pelvic floor muscle training were more likely to report improvement, the confidence interval was wide, and more than half of the 'control' group (the women who did not get the additional weekly or twice-weekly group supervision) reported improvement. This finding, of subjective improvement in both active treatment groups, with more improvement reported by those receiving more health professional contact, was consistent throughout the review.

We feel there are several reasons why caution is needed when interpreting the results of the review: there were few data in any comparison; a number of trials were confounded by comparing two arms with multiple differences in the approaches to pelvic floor muscle training; there was a likelihood of a relationship between attention and reporting of more improvement in women who were not blind to treatment allocation; some trials chose interventions that were unlikely to have a muscle training effect; and some trials did not adequately describe their intervention.

Authors' conclusions

This review found that the existing evidence was insufficient to make any strong recommendations about the best approach to pelvic floor muscle training. We suggest that women are offered reasonably frequent appointments during the training period, because the few data consistently showed that women receiving regular (e.g. weekly) supervision were more likely to report improvement than women doing pelvic floor muscle training with little or no supervision.

PLAIN LANGUAGE SUMMARY

Comparisons of approaches to pelvic floor muscle training for urinary incontinence in women

Involuntary urine leakage (or incontinence) is a widespread condition experienced by about a quarter of women. Exercise for the pelvic floor muscles is often the first treatment women are offered. Improving the strength, endurance and co-ordination of the pelvic floor muscles can help decrease the urine leakage. This review included 21 studies in 1490 women and looked at whether one way of teaching, supervising or performing these exercises was better than another. Women who had regular and repeated contact with the person who taught them to do the exercises and monitored their progress were more likely to report they were improved after treatment. Further research is needed because there were problems interpreting the studies, which meant we could not draw any firm conclusions about many of the other possible ways of teaching, supervising or performing these exercises.

BACKGROUND

Description of the condition

Urinary incontinence is a common problem amongst adults living in the community. It is more frequent in women, increasing with age, and is particularly common amongst those in residential care (Hunskaar 2002). For a variety of reasons (such as difference in study populations, definitions and measurement) estimates of urinary incontinence prevalence differ widely. A review of 36 general population studies included in the 4^{th} International Consultation on Incontinence, found that most studies reported a prevalence of 'any' urinary incontinence in the range of 25% to 45% for women; this estimate comes from studies in which symptoms of urinary

incontinence were reported as 'ever', 'any' or 'at least once in the past 12 months' (Milsom 2009). Urinary incontinence can be a debilitating condition with a large negative impact on quality of life (Bartoli 2010).

The two most common types of urinary incontinence in women are stress urinary incontinence (SUI) and urgency urinary incontinence (UUI). SUI is typically experienced as involuntary urine leakage with cough, sneeze and other types of physical exertion, while UUI is characterised by involuntary urine leakage associated with urgency (that is, the sudden and compelling need to urinate). Some women experience both SUI and UUI; these women are said to have mixed urinary incontinence.

The mechanisms that cause the involuntary leakage in SUI and UUI are different. In SUI the bladder outlet (urethra) is not closed off properly during exertion, resulting in leakage. The lack of closure pressure in the urethra is due to anatomic changes in the bladder and urethra (for instance, the bladder has prolapsed or 'dropped down') and muscles (for example, the pelvic floor muscles are weak and do not lift the bladder or squeeze the urethra shut). In UUI the problem is that the bladder muscle (the detrusor muscle) contracts so hard that the increased bladder pressure overwhelms the urethral closure pressure and this results in leakage. Thus, in UUI the problem is one of an overactive bladder muscle.

A wide range of treatments has been used in the management of urinary incontinence, including conservative interventions such as:

- pelvic floor muscle training (Dumoulin 2010);
- vaginal cones (Herbison 2002);
- lifestyle interventions;
- bladder training (Wallace 2004);
- anti-incontinence devices (Shaikh 2006);
- pharmaceutical interventions (for example, anticholinergics (Nabi 2006);
- surgery (for example, minimally invasive sling operations (Ogah 2009); or
- absorbent products (Fader 2007; Fader 2008).

The underlying reason for leakage is different for different types of incontinence, therefore the choice of therapy can also differ. For example, women with UUI are commonly offered an anticholinergic drug which reduces the overactive detrusor muscle contractions. Women with SUI might be offered surgery that lifts the bladder neck and increases urethral pressure. One of the most commonly offered therapies for SUI, UUI and MUI is pelvic floor muscle training.

In a prior Cochrane systematic review Dumoulin and Hay-Smith (2010) concluded that there was support for the widespread recommendation that pelvic floor muscle training (PFMT) is offered as first-line conservative management programmes for women with stress, urge or mixed urinary incontinence (Dumoulin 2010). For a more comprehensive background to the rationale for treatment of urinary incontinence with PFMT please see Dumoulin 2010.

As PFMT is more effective than no treatment, placebo or inactive control treatments for women with urinary incontinence, then the question 'What is the most effective approach to performing, teaching and supervising PFMT?' arises. For example, the need to identify the optimal frequency and duration of supervised PFMT was recently identified as an important research question by Buckley and colleagues (Buckley 2009). This research question was an outcome of a study to develop a methodology (using the James Lind criteria process) in which patients and clinicians worked together to identify and prioritise important urinary incontinence research questions through consensus (Buckley 2009).

Description of the intervention

When making a recommendation about PFMT for an individual woman, the clinician will balance a number of factors that may include the findings from assessment of the pelvic floor muscles (for example, muscle strength, endurance and co-ordination), the symptoms, the goal of treatment, the woman's lifestyle and preferences, and the resources available within the health setting. While this might lead to individualised PFMT to some extent, there are some core variables the clinician might consider, such as the following.

- The type of PFMT, for example direct PFMT (e.g. voluntary pelvic floor muscle contraction) versus indirect PFMT (e.g. pelvic floor muscle contraction facilitated through abdominal muscle contraction), or strength versus co-ordination or behavioural training (e.g. use of a voluntary pelvic floor muscle contraction in anticipation of and during a cough or sneeze to reduce leakage, described by Miller and colleagues as 'The Knack' (Miller 1998)).
- The PFMT exercise parameters, for example the number of contractions, the length of hold and rests, the speed of contraction and the amount of contraction effort.
- The addition of resistance to training (e.g. intravaginal resistance devices).
- The amount of contact with the health professional supervising the PFMT programme.
- The type of teaching and supervision, such as group versus individual, face to face versus written instructions.
- The use of strategies to enhance adherence to training, such as alarms, diaries and other psychological or educational interventions.

Why it is important to do this review

The purpose of this review is to summarise the existing trials comparing different approaches to PFMT for urinary incontinence in women to inform further research to address the existing uncertainty regarding optimal training.

This review is one of a set of Cochrane reviews which address various approaches to the conservative treatment of urinary incontinence in women. Others include:

1. a specific investigation of the effect of adding feedback and biofeedback to PFMT (Herderschee 2011), to determine if this adjunct might be part of the optimal protocol;
2. PFMT versus no treatment, or inactive control treatments, for urinary incontinence in women (Dumoulin 2010);
3. PFMT for the prevention and management of urinary incontinence in antenatal and postnatal women (Hay-Smith 2008).

OBJECTIVES

We tested the following hypothesis:

There are differences in the effects of alternative approaches to pelvic floor muscle training in the management of urinary (stress, urge, mixed) incontinence in women.

We made the following comparisons.

1. More versus less contact with health professionals
2. Group versus individual supervision of PFMT
3. Direct versus indirect methods of PFMT
4. Individualised versus generic PFMT
5. Near maximal versus submaximal contractions
6. Daily versus three times per week PFMT
7. Upright and supine versus supine exercise positions alone
8. Strength and motor learning versus motor learning PFMT alone
9. PFMT and abdominal muscle exercise versus PFMT alone
10. PFMT with intravaginal resistance device versus PFMT alone
11. PFMT and adherence strategy versus PFMT alone
12. 'More intensive' versus 'less intensive' PFMT programmes

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised controlled trials and quasi-randomised trials (for example, allocation to group according to date or month of birth). We excluded other forms of controlled clinical trials.

Types of participants

All women with urinary incontinence diagnosed as having stress, urge or mixed incontinence on the basis of symptoms, signs or urodynamic evaluation, as defined by the trialists. Thus, we excluded studies of women without urine leakage (prevention studies).

We excluded studies of women with urinary incontinence whose symptoms might be due to significant factors outside the urinary tract, for example neurological disorders, cognitive impairment, lack of independent mobility. We also excluded studies investigating nocturnal enuresis in women.

We excluded studies that specifically recruited antenatal or postnatal women. Given the physiological changes of pregnancy and postpartum period it is possible that the effect of PFMT might differ in this group. PFMT for the prevention and management of urinary incontinence in antenatal and postnatal women is addressed in another Cochrane review (Hay-Smith 2008).

Types of interventions

At least two arms of all trials included the use of PFMT to treat the symptoms of urine leakage with some difference in the PFMT between the two arms. PFMT was defined as any programme of repeated voluntary pelvic floor muscle contractions, or 'indirect' voluntary pelvic floor muscle contraction irrespective of variations in purpose and training parameters. 'Direct' PFMT is that in which the woman is asked to focus specifically on a voluntary contraction of the pelvic floor muscles. With 'indirect' PFMT, the premise is that a pelvic floor muscle contraction may be facilitated or enhanced through co-contraction of another related muscle group (e.g. abdominal or hip or gluteal muscles). Where the focus of contraction is 'other' muscle group(s) in order to facilitate/enhance or substitute for a direct pelvic floor muscle contraction we have called this 'indirect' PFMT.

Aside from direct versus indirect training, we were also interested in other differences in approach such as:

- different exercise parameters, such as differences in the type (near maximal, submaximal, 'The Knack', with or against gravity), number (per set, per day or per week), or duration of contractions;
- addition of resistance to contractions, such as the use of intravaginal resistance devices (e.g. air or water filled balloon catheters, spring-loaded speculum);
- differences in the type of instruction, such as verbal, written, online/web-based, face-to-face, individual or group instruction;
- differences in the amount and type of health professional supervision of training;

- addition of adjuncts to enhance adherence (for instance, an alarm or text messaging).

We excluded the following interventions:

- PFMT with adjunctive biofeedback unless the same biofeedback intervention was given in both arms. Another Cochrane review specifically investigates the effect of adding biofeedback to PFMT (Herderschee 2011).
- PFMT combined with lifestyles or fluid management advice (such as weight loss or reduction in caffeine intake respectively) unless the same advice was given in both arms. Another Cochrane review specifically investigates the effectiveness of lifestyles interventions (Imamura 2010a).
- PFMT combined with another 'stand alone' conservative therapy (such as bladder training (i.e. a scheduled voiding regimen), electrical stimulation, vaginal cones), or drug therapy (for example, an anticholinergic).

Types of outcome measures

A subcommittee (Outcome Research in Women) of the Standardisation Committee of the International Continence Society suggested that research investigating the effect of therapeutic interventions for women with urinary incontinence consider five outcome categories (Lose 1998):

- the woman's observations (symptoms);
- quantification of symptoms (for example, urine loss);
- the clinician's observations (anatomical and functional);
- quality of life and socioeconomic measures.

We chose one or more outcomes of interest from each domain for the review. We chose the woman's observations of change in symptoms as a primary outcome. We chose condition-specific quality of life as the other primary outcome measure to reflect the findings of a recent study of women with urinary incontinence who identified this as the most important outcome to be measured in incontinence research (Herbison 2009).

Primary outcomes

The primary outcomes of interest were:

1. symptomatic cure or improvement as reported by the woman (measured as urinary incontinence not cured or improved);
2. condition-specific quality of life assessment (for example, Incontinence Impact Questionnaire, King's Health Questionnaire).

Secondary outcomes

Secondary outcomes of interest were:

1. number of leakage episodes;
2. measures of leakage severity (for example, pad usage);
3. micturition frequency;
4. symptom impact (that is, measures of symptom impact or distress other than those measured with validated incontinence-specific quality of life instruments);
5. measures of pelvic floor muscle function (for example, electromyography, vaginal squeeze pressure);
6. other health status or quality of life measures (for example, Short Form-36);
7. formal economic analysis (for example, cost-effectiveness, cost utility).

Other outcomes of interest were:

8. treatment adherence;
9. any of the primary or secondary outcomes in the longer term (that is 12 months or more);
10. adverse events;
11. any other outcome not pre-specified, but judged important when performing the review.

Search methods for identification of studies

We did not impose any restrictions, for example language or publication status, on the searches.

Electronic searches

This review drew on the search strategy developed for the Cochrane Incontinence Group. We identified relevant trials from the Cochrane Incontinence Group Specialised Trials Register. For more details of the search methods used to build the Specialised Register please see the Group's module in *The Cochrane Library*. The register contains trials identified from the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE and CINAHL, and handsearching of journals and conference proceedings. The trials in the Cochrane Incontinence Group Specialised Register are also contained in CENTRAL. The date of the last search was: 17 May 2011.

The terms used to search the Incontinence Group Specialised Register are given below:

(({DESIGN.CCT*}
OR {DESIGN.RCT*}) AND ({INTVENT.PHYS.PFMT*} OR {INTVENT.PHYS.BIOFEED*}) AND {TOPIC.URINE.INCON*})

(All searches were of the keyword field of Reference Manager 12, Thomson Reuters).

Searching other resources

We searched the references lists of relevant articles.

Data collection and analysis

We conducted data collection and analysis in accordance with the *Cochrane Handbook for Systematic Reviews of Interventions* (version 5.2.2) (Higgins 2009).

Selection of studies

Two review authors (principally JHS and RH) independently evaluated records of all studies retrieved by the Trials Search Co-ordinator for eligibility without prior consideration of the results. We immediately excluded studies that were ineligible on the basis of title or abstract alone (for example, trial participants were men or children) after cross-checking. We retrieved the full text of all remaining records. We independently evaluated full-text records (principally JHS and RH) and cross-checked decisions about eligibility. Excluded studies are listed with reasons for their exclusion in the Characteristics of excluded studies table.

We resolved all disagreements through discussion. Where one of the review authors (JHS, CD or PH) was an author of a study identified by the search, that review author had no involvement in the decision about eligibility.

Data extraction and management

We adapted the data extraction form from a previous review of PFMT (Dumoulin 2010) and two review authors (JHS and RH) tested it. These two review authors undertook and cross-checked data extraction independently. We used a third person for any trial in which JHS was involved as a researcher. We resolved any disagreements as previously described.

Where trial data were possibly collected but not reported, or data were reported in a form that could not be used in the formal comparisons, we sought further clarification from the trialists. In addition, where the reported data were clearly incomplete (that is, data from abstracts for ongoing trials) we contacted trialists for data from the completed trial. Ten trialists responded with further information (see Acknowledgements) or to confirm that the data were no longer retrievable. If additional data were provided, we stated this in the additional data tables. Data entry was carried out by RH and cross-checked by JHS.

Assessment of risk of bias in included studies

Two review authors assessed risk of bias independently by using the Cochrane 'Risk of bias' assessment tool. This included: random sequence generation, allocation concealment, blinding of participants, therapists and outcome assessors, completeness of outcome data, selective outcome reporting and other potential sources of bias. JHS and RH assessed these domains (or RH and another review author where JHS was a researcher in the trial being assessed) and we resolved any disagreement by consensus or discussion with PH. No review author assessed risk of bias for a trial in which they had been a researcher. The assessment of risk of bias is summarised in the 'Risk of bias' table.

Measures of treatment effect

For categorical outcomes such as self reported cure we related the numbers reporting an outcome to the numbers at risk in each group to derive a risk ratio. We dichotomised ordinal data (such as Likert scales for symptom improvement) and managed them as a categorical outcome. For continuous variables such as quality of life score we used means and standard deviations to derive mean differences. We treated count data (such as leakage episodes, which were considered a relatively common event) as continuous data.

Unit of analysis issues

The primary analyses were per women randomised.

Dealing with missing data

We carried out the data analysis on an intention-to-treat basis as far as possible. We made attempts to obtain missing data from the trialists. If additional data were provided, we made this clear in additional data tables.

Assessment of heterogeneity

We assessed statistical heterogeneity in three ways: visual examination of the forest plots; Chi^2 test ($P <= 0.10$) for heterogeneity and I^2 statistics. An I^2 statistic measurement greater than 50% was taken to indicate substantial heterogeneity. We sought and discussed plausible explanations for statistically significant heterogeneity.

Assessment of reporting biases

Due to of the difficulty in detecting and correcting for publication bias and other reporting biases, the review authors aimed to minimise their potential impact by ensuring a comprehensive search for eligible studies and by being alert for duplication of data.

Data synthesis

We analysed and displayed the data from primary studies in the forest plots using fixed-effect models unless otherwise stated.

If there were enough trials, we combined results for meta-analysis. The direction of benefit is clearly marked and labelled on the forest plots. If meta-analysis was not considered appropriate we discussed the findings of studies in a narrative synthesis.

If outcome measurements were reported in such a way that the data could not be combined (e.g. data reported as mean without a measure of dispersion), we used 'other data' tables to present the results.

Subgroup analysis and investigation of heterogeneity

It was intended to conduct subgroup analysis by type of incontinence. The rationale for PFMT is different for the two main types of urinary incontinence (stress and urgency) therefore it is plausible to expect a difference in the outcome of PFMT on the basis of the type of incontinence. It is commonly believed that PFMT is most effective for women with stress urinary incontinence and that it may be effective, in combination with behavioural interventions, for women with mixed urinary incontinence. In the past, PFMT has rarely been the first-choice treatment for women with urgency urinary incontinence alone.

Sensitivity analysis

We planned sensitivity analysis with respect to risk of bias as there is evidence that this may have an impact on the findings of meta-analysis (Moher 1998; Pildal 2007). In the event, there were insufficient trials to do a sensitivity analysis within any of the comparisons.

RESULTS

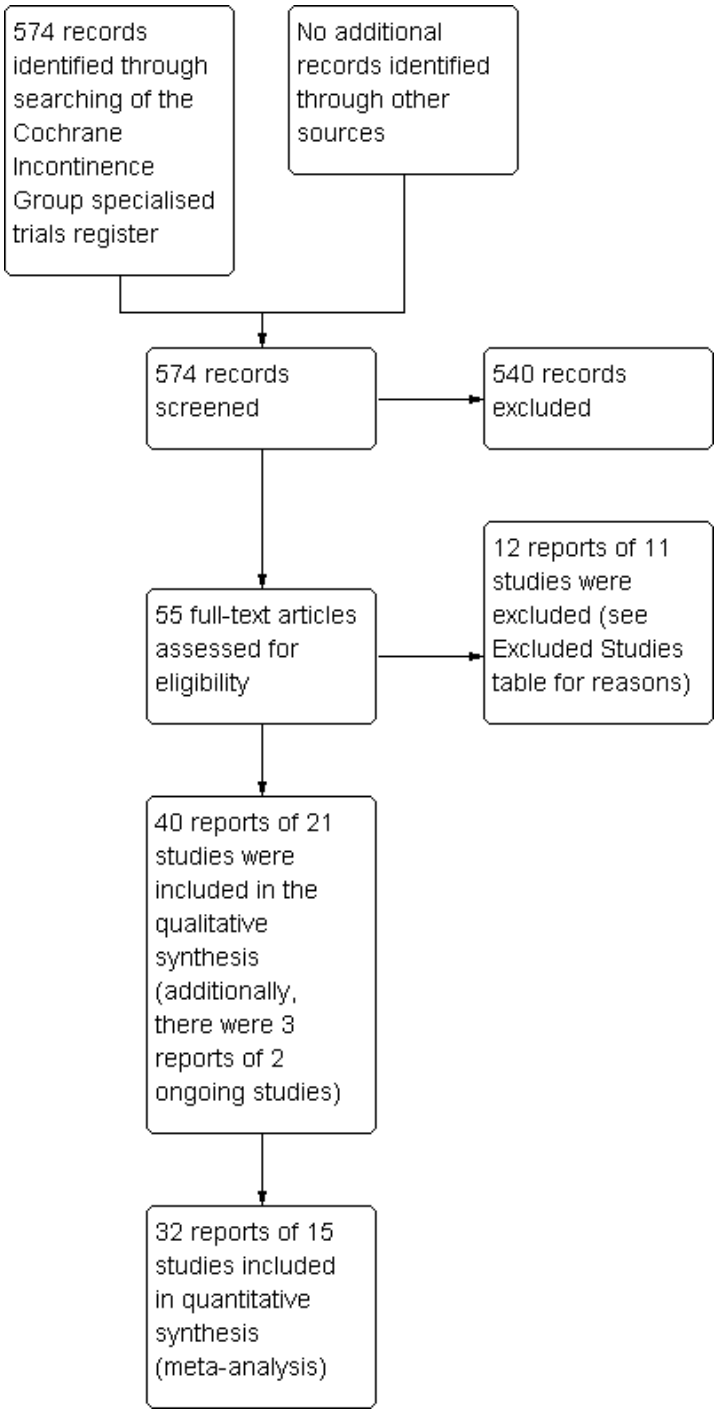
Description of studies

See: Characteristics of included studies; Characteristics of excluded studies; Characteristics of ongoing studies.

Results of the search

The search produced 574 study records. From this we identified 55 potentially eligible reports of studies. We included 40 reports of 21 studies, three reports of two studies were ongoing, and we excluded 12 reports of 11 studies (please see the Characteristics of excluded studies table with the reasons for exclusion). The PRISMA diagram showing the flow of literature through the search and assessment process can be seen in Figure 1.

Figure 1. Study flow diagram.



Included studies

One of the 21 trials had three arms (and two comparisons) that met the criteria for the review. To differentiate the two comparisons these are labelled Sriboonreung 2011a and Sriboonreung 2011b. More detail of the three trial arms, and two comparisons, are found in the Characteristics of included studies.

None of the included trials was large; the 21 trials randomised 1490 women. Nearly two-thirds (13 of 21 trials) had more than 20 and fewer than 50 participants per comparison group. Four trials were small, with fewer than 20 per comparison group (Ferguson 1990; Johnson 2001; Konstantinidou 2007; Savage 2005) and one of these was reported as a pilot study (Savage 2005). The largest trial randomised about 120 women to each of two trial arms (Liebergall 2009). Eleven trials reported an a priori power calculation to estimate sample size (Felicissimo 2010; Gallo 1997; Ghoniem 2005; Hay-Smith 2002; Hung 2010; Johnson 2001; Konstantinidou 2007; Liebergall 2005; Liebergall 2009; Ng 2008; Sriboonreung 2011a; Sriboonreung 2011b) and a 12th reported power post hoc (de Oliveira 2009).

Sample characteristics

The full inclusion and exclusion criteria for each trial are described in the Characteristics of included studies. Based on the reported demographics, the observed pattern was that the trials generally recruited only women with stress urinary incontinence or predominant stress urinary incontinence who were on average about 50 years of age, had up to four leakage episodes per day and a mean duration of symptoms of at least five years. A summary of diagnosis, age and symptom duration is given below.

Diagnosis

The sample populations in the trials were described as having the following.

- Only urodynamic stress urinary incontinence (Bø 1990; de Oliveira 2009; Diniz Zanetti 2007; Felicissimo 2010; Ferguson 1990; Gallo 1997; Johnson 2001; Konstantinidou 2007). Note: urodynamic stress incontinence is the term used when urodynamic studies demonstrate involuntary loss of urine during increased intra-abdominal pressure, but the leakage is not caused by a contraction of the detrusor muscle (bladder smooth muscle).
- Urodynamic stress urinary incontinence or stress urinary incontinence (based on signs or symptoms) (Ghoniem 2005).
- Only stress urinary incontinence (based on signs or symptoms) (Borello-France 2006; Liebergall 2009; Ramsay 1990; Savage 2005; Sriboonreung 2011a; Sriboonreung 2011b; Sugaya 2003).
- Either stress urinary incontinence or mixed urinary incontinence (where stress incontinence was the predominant symptom) (Delgado 2009; Hay-Smith 2002).
- Either stress incontinence or mixed urinary incontinence (Hung 2010; Liebergall 2005; Wells 1999).
- Only mixed urinary incontinence (Ng 2008).

Age

Some trials set age limits, either a lower or an upper limit or both. Ten trials set an upper limit that excluded:

- the young-old (more than 65 years, Hung 2010; Johnson 2001; Liebergall 2005; Liebergall 2009; Sriboonreung 2011a; Sriboonreung 2011b; more than 70 years, Borello-France 2006); and
- old-old (more than 75 years, de Oliveira 2009; Ghoniem 2005; more than 80 years Gallo 1997).

The trial that recruited the youngest participants had a mean age of 36 years (Ferguson 1990), although most trials recruited women who were on average somewhat older. The oldest mean age for a sample population was about 60 years (Gallo 1997). One trial (Ramsay 1990) did not report participant age. In general, on the basis of median or mean age, the trials could be grouped as follows:

- up to 45 years (Bø 1990; Ferguson 1990);
- 45 to 49 years (Hay-Smith 2002; Hung 2010; Konstantinidou 2007; Liebergall 2009);
- 50 to 54 years (Borello-France 2006; de Oliveira 2009; Delgado 2009; Felicissimo 2010; Johnson 2001; Liebergall 2005; Ng 2008; Savage 2005; Sriboonreung 2011a; Sriboonreung 2011b);
- 55 years or more (Diniz Zanetti 2007; Gallo 1997; Ghoniem 2005; Sugaya 2003; Wells 1999).

This distribution of mean or median age is congruent with stress urinary incontinence as the most common diagnostic category for the included trials. Urgency and urgency urinary incontinence become more prevalent in older age.

Leakage episodes

Under half (nine of 21 trials) gave data on baseline incontinence frequency. Liebergall 2005 reported that about 44% of their participants experienced leakage once or more per day. The approximate median or mean number of leakage episodes per day in the other eight trials was fewer than one (Borello-France 2006; de Oliveira 2009; Hung 2010), one to two (Hay-Smith 2002; Konstantinidou 2007), or three to four per day (Ghoniem 2005; Johnson 2001; Sugaya 2003).

Duration of leakage symptoms

About half (12 of 21 trials) reported some information about how long women had experienced incontinence symptoms prior to enrolling in the trial. Wells 1999 stated that 68% of women had symptoms for more than one year. The approximate mean or median duration of symptoms in the other 11 trials was five years (de Oliveira 2009; Delgado 2009; Diniz Zanetti 2007; Felicissimo 2010) or up to 10 years (Bø 1990; Hay-Smith 2002; Hung 2010; Konstantinidou 2007; Savage 2005; Sugaya 2003).

The comparisons

Some trials were designed so that it was very clear what was being tested. In 13 instances the 'experimental' group received some additional intervention over and above a PFMT intervention that was common to both groups, specifically:

- more PFMT supervision in the form of an exercise group (Bø 1990; Felicissimo 2010; Konstantinidou 2007), individual appointments (Diniz Zanetti 2007) or phone calls (Ng 2008);
- an intravaginal resistance device (Delgado 2009; Ferguson 1990; Wells 1999);
- a cue to exercise (Gallo 1997; Sugaya 2003);
- two more exercise positions (Borello-France 2006);
- a strength training programme (Hay-Smith 2002);
- an abdominal muscle exercise programme (Sriboonreung 2011b).

In another four instances, the trialists kept all aspects of the intervention the same in all respects except one. These were direct comparisons of:

- 'indirect' versus 'direct' PFMT (Ghoniem 2005; Ramsay 1990; Savage 2005): in Ghoniem 2005 and Ramsay 1990 the 'indirect' training group were asked to cross their ankles and do isometric hip abductor contractions, and in Savage 2005 the 'indirect' training group were doing a Pilates exercise programme;
- submaximal versus near maximal pelvic floor muscle contractions (Johnson 2001);
- PFMT daily versus PFMT three times a week (Sriboonreung 2011a).

In the remaining three trials there were multiple differences between the intervention groups, such as differences in both the PFMT programmes and the amount of health professional contact. These trials contributed to more than one comparison in the analysis. It was difficult to be sure how to attribute any differences in outcome between the trial arms because the comparison was potentially confounded by some other intervention variable. The three trials compared:

- 'indirect' PFMT for 15 to 45 minutes per day with weekly individual supervision versus 'direct' PFMT of 15 minutes per day with weekly group supervision (Liebergall 2005; Liebergall 2009); the 'indirect' PFMT was the 'Paula' method (see the trial reports for a description of the intervention);
- 'indirect' PFMT and fortnightly clinic visits versus 'direct' PFMT and no clinic visits (Hung 2010); the 'indirect' PFMT was a breathing, abdominal and PFM rehabilitation programme described by Sapsford 2004;
- 'standard' PFMT with twice-weekly group supervision versus 'individualised' PFMT with twice-weekly individual supervision (de Oliveira 2009).

PFMT interventions

The PFMT interventions are described in detail in Table 1; Table 2; Table 3; Table 4; Table 5; Table 6; Table 7; Table 8; Table 9; Table 10; Table 11. As the purpose of the review was to compare different approaches to PFMT the considerable variation in interventions was not surprising. We have summarised some key contributors to effective training below, namely the confirmation of a correct voluntary pelvic floor muscle contraction, training duration and the exercise 'dose'.

Confirmation of a correct voluntary pelvic floor muscle contraction

Fourteen trials confirmed a correct voluntary pelvic floor muscle contraction in both comparison groups (Bø 1990; Borello-France 2006; de Oliveira 2009; Delgado 2009; Felicissimo 2010; Gallo 1997; Hay-Smith 2002; Hung 2010; Johnson 2001; Konstantinidou 2007; Savage 2005; Sriboonreung 2011a; Sriboonreung 2011b; Sugaya 2003; Wells 1999) and in three it was not clear if this was done (Diniz Zanetti 2007; Ferguson 1990; Ng 2008). The remaining four trials made a comparison between 'direct' and 'indirect' PFMT methods (Ghoniem 2005; Liebergall 2005; Liebergall 2009; Ramsay 1990). Ghoniem 2005 clearly stated that a correct contraction was confirmed in the 'direct' training group only, but it was not clear whether a correct contraction was confirmed in either group in the other three trials.

Duration of PFMT

The duration of training varied considerably, although about half (10 of 21 trials) had an intervention period of about 12 weeks (Borello-France 2006; de Oliveira 2009; Diniz Zanetti 2007; Ghoniem 2005; Konstantinidou 2007; Liebergall 2005; Liebergall 2009; Ramsay 1990; Savage 2005; Sriboonreung 2011a; Sriboonreung 2011b). Five trials had shorter training durations (four to six weeks(Gallo 1997); six weeks (Ferguson 1990; Johnson 2001); and eight weeks (Felicissimo 2010; Sugaya 2003)) and six trials had interventions longer than 12 weeks (16 weeks (Delgado 2009); 18 weeks(Hung 2010); 20 weeks(Hay-Smith 2002; Wells 1999); and 24 weeks (Bø 1990; Ng 2008)).

The exercise 'dose'

The PFMT interventions are described in detail in Table 1; Table 2; Table 3; Table 4; Table 5; Table 6; Table 7; Table 8; Table 9; Table 10; Table 11. There is one table per comparison so that all the available information about the interventions in any comparison is in the same place.

PFMT may be prescribed to:

- increase **strength** (the maximum force generated by a muscle in a single contraction). Characteristic features of strength training include low numbers of repetitions with high loads. One way to increase 'load' is to increase the amount of voluntary effort with each contraction;
- increase **endurance** (ability to contract repetitively, or sustain a single submaximal contraction over time). Endurance training is characterised by high numbers of repetitions or prolonged contractions with low to moderate loads;
- **co-ordinate** muscle activity to reduce leakage or to suppress urge. These learned behaviours use a voluntary pelvic floor muscle contraction in response to a specific situation to improve co-ordination of a contraction, in particular with raised intra-abdominal pressure ('The Knack') or detrusor contraction (urge suppression); or
- a combination of these.

Strength training requires short duration, high load (near maximal intensity) contractions, and endurance training light loads (submaximal effort) with high repetitions (ACSM 2009). There is not an absolute dividing line that differentiates strength from endurance-type exercise programmes; it is common for both strength and endurance to improve in response to an exercise programme, although one may be affected more than another.

In each table we categorised each exercise programme as a strength, endurance or behavioural training programme, or a combination of these. If the description of the PFMT programme was not sufficiently detailed to be sure we have categorised the purpose of training as uncertain.

Excluded studies

Two completed trials were yet to be fully reported and were classified as ongoing studies (Kincade 2005; von der Heide 2003). There were 11 further exclusions. These were:

- two trials where the intervention included bladder training in addition to PFMT (Hill 2007; Hui 2006);
- one trial that included an adjunctive device in one arm but it was not clear if the primary purpose of this device was biofeedback or intravaginal resistance (Klinger 1995);
- one trial in which the participants were postnatal women with persistent stress incontinence. While some of the participants in this study were more than six months post-delivery we decided to exclude this study because it included postnatal women (Dumoulin 2003);
- seven trials that had collected data on one or more of the primary and secondary outcomes of interest in the review, but the trial reports did not contain data that could be used in the analysis (Crothers 2001; de Jong 2006; Nygaard 1996; Orelle NatraTone 2008; Taylor 1986; Wong 1997; Yoon 1999).

Details of the difficulties with the reported data and other reasons for exclusion are given in the Characteristics of excluded studies.

Risk of bias in included studies

A summary of the risk of bias is provided in Figure 2 and Figure 3. Due to brevity of reporting it was difficult to assess one trial that was available only as a conference abstract (Ramsay 1990).

Figure 2.

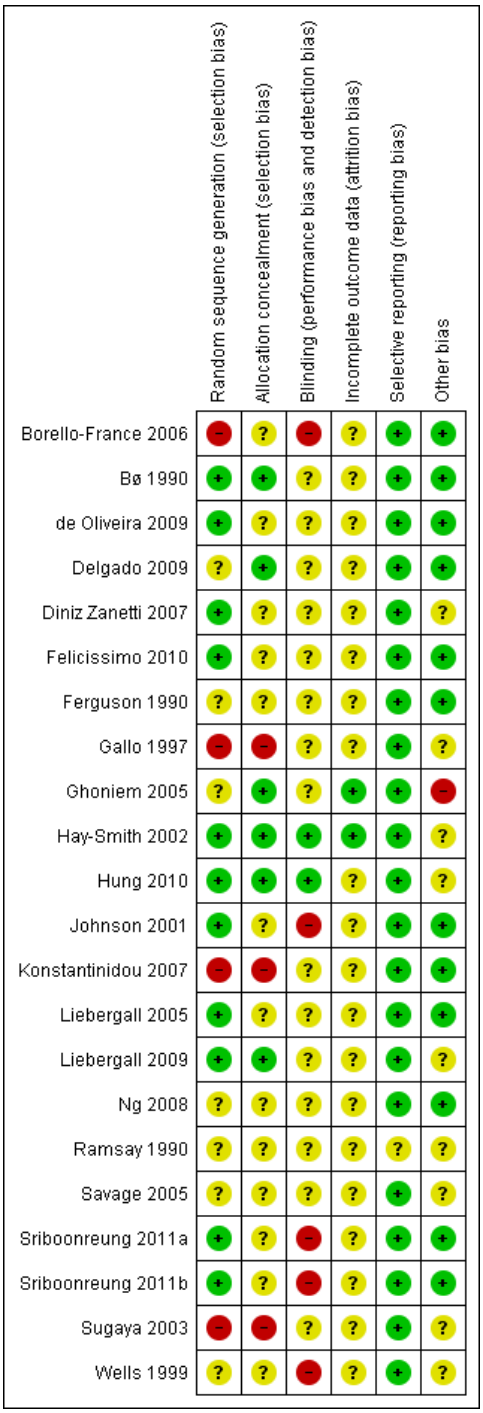
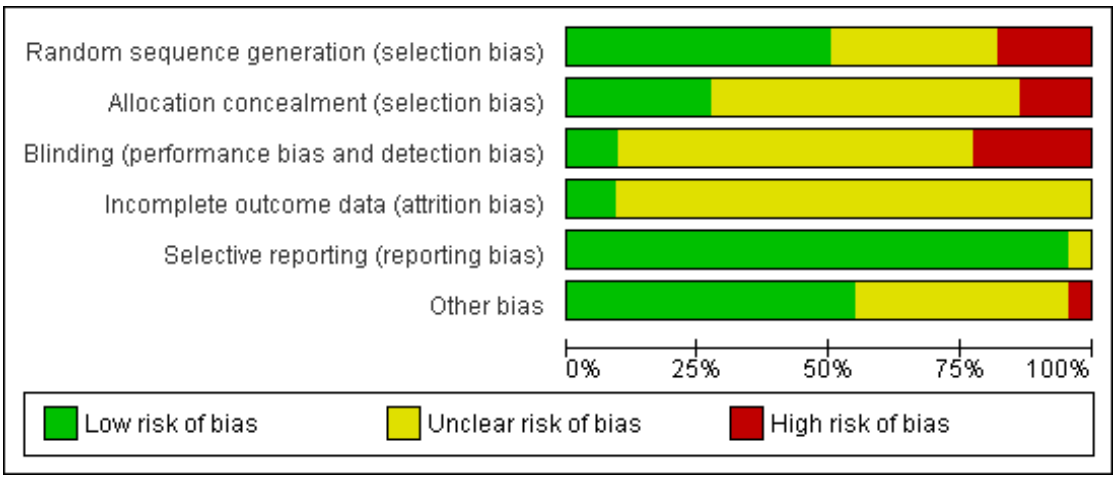


Figure 3.



Allocation

Ten trials provided sufficient detail of their methods of generating a random sequence to be sure this was genuinely random (Bø 1990; de Oliveira 2009; Diniz Zanetti 2007; Felicissimo 2010; Hay-Smith 2002; Hung 2010; Johnson 2001; Liebergall 2005; Liebergall 2009; Sriboonreung 2011a; Sriboonreung 2011b) and in three this was clearly not random (Gallo 1997; Konstantinidou 2007; Sugaya 2003).

Six trials provided sufficient detail to be sure that allocation was concealed (Bø 1990; Delgado 2009; Ghoniem 2005; Hay-Smith 2002; Hung 2010; Liebergall 2009) and in three allocation was clearly not concealed (every other patient randomly assigned, Gallo 1997; consecutive alternative fashion, Konstantinidou 2007; divided in order of presentation, Sugaya 2003). One trial stated that allocation was at "random" and then "allocation was adjusted to balance for age and severity" (Borello-France 2006). We decided this trial did not have adequate allocation concealment.

Overall, with regard to random allocation and concealment, we considered:

- six trials were at low risk of bias (Bø 1990; Delgado 2009; Ghoniem 2005; Hay-Smith 2002; Hung 2010; Liebergall 2009);
- four were at high risk of bias (Borello-France 2006; Gallo 1997; Konstantinidou 2007; Sugaya 2003); and
- for the remainder the risk of bias was unclear.

Blinding

It is difficult to blind participants to an intervention such as PFMT. Two trials attempted this (Ghoniem 2005; Ramsay 1990) although given the widespread availability of information about PFMT it is not clear if women were truly blind or not. Further, it is probably impossible to blind those providing instruction in PFMT. We considered that none of the trials was able to blind treatment providers, and it was likely few if any participants were blind to treatment allocation. Therefore we were presented with two options; either to categorise all trials as being at high risk of bias with regard to these two aspects of blinding, or decide that because all the trials had a similar problem this was not a source of increased bias in the review. We took the latter option and rated all the trials as being a low risk of bias for blinding of participants and providers. Thus the rating given for performance and detection bias in the 'Risk of bias' tables is based solely on blinding of outcome assessment (or detection bias).

Ideally, all trials could blind outcome assessment of some or all outcomes, although the primary outcomes in this review were self reported and therefore could not be blinded. Only three trials clearly stated that outcome assessment was blinded for one or more of the outcomes of interest in the review (Hay-Smith 2002; Hung 2010; Savage 2005) and two trials stated that a lack of blind outcome assessment was a limitation of their study (Sriboonreung 2011a; Sriboonreung 2011b; Wells 1999).

Incomplete outcome data

One trial had no loses to follow-up (Ferguson 1990) and two trials did not clearly state whether there were any loses to follow-up or not (Diniz Zanetti 2007; Ramsay 1990), although in these latter two trials it seems as though there were none. Of the rest, the proportion of losses to follow-up ranged from about 2% to 45%:

- 1% to 10% (Bø 1990; de Oliveira 2009; Felicissimo 2010; Ghoniem 2005; Hay-Smith 2002; Hung 2010; Liebergall 2005; Savage 2005);
- 11% to 20% (Borello-France 2006; Gallo 1997; Johnson 2001; Sriboonreung 2011a; Sriboonreung 2011b; Sugaya 2003);
- 21% to 30% (Delgado 2009; Konstantinidou 2007; Liebergall 2009; Ng 2008);
- 31% or higher (Wells 1999).

Four trials did not clearly state the number of losses to follow-up by treatment group (de Oliveira 2009; Delgado 2009; Ghoniem 2005; Johnson 2001). In the rest, the proportion of dropouts did not usually differ by treatment group. The trials in which there were slight differences were Bø 1990 (none from the home PFMT group, five from the weekly exercise group), Gallo 1997 (9/43 home PFMT group and 2/43 home PFMT with audiotape) and Liebergall 2009 (36/123 'direct' PFMT and 21/117 'indirect' PFMT).

For analysis according to the full intention-to-treat principle, we used two criteria. First, participants were analysed in the group to which they were originally assigned. Second, if there were any missing data, the effect of this was assessed in a reasonable way, such as 'multiple imputation' (Lane 2008). None of the included trials met both criteria for a full intention-to-treat analysis.

Two trials clearly stated that women were analysed in the group to which they were assigned (Ghoniem 2005; Hay-Smith 2002). We decided that these trials were are low risk of bias providing the proportion of dropouts was 10% or under, and there was no evidence of differential dropout rates between the comparison groups. All other trials were assessed as being at unclear risk of bias.

Five trials reported how they managed missing data in their intention-to-treat analysis (Borello-France 2006; Ghoniem 2005; Hung 2010; Liebergall 2009; Ng 2008) and none of these approaches met the second criterion. In four, missing data were dealt with by carrying the last outcome or baseline value forward (Borello-France 2006; Ghoniem 2005; Hung 2010; Ng 2008) and in the fifth, a best-case/worst-case sensitivity analysis was used (Liebergall 2009).

Selective reporting

Overall, we did not find anything to suggest that the included trials selectively reported their data. A more common problem was incomplete data reporting. For example, sometimes raw data were given for one comparison group but not the other for an outcome. On occasion we saw mention in the methods section that participants were asked about adverse events, but the trialists did not mention adverse events in the results (either that there were or were not any).

Other potential sources of bias

Two trials were funded by medical or pharmaceutical companies (Gallo 1997; Ghoniem 2005) and one had support 'in-kind' from a medical company in addition to public funding (Johnson 2001). Ghoniem 2005 and colleagues also made a conflict of interest statement in which all but the first author of this paper declared a financial or other relationship with the two companies funding the research.

Ten trials declared funding support from one or more public grant bodies (Bø 1990; Borello-France 2006; Ferguson 1990; Hay-Smith 2002; Hung 2010; Liebergall 2005; Liebergall 2009; Ng 2008; Sugaya 2003; Wells 1999). One trial stated it had been conducted "independently of company input" (Delgado 2009) and another that no funding was received (Diniz Zanetti 2007). Seven papers, all recent publications, stated the authors had no conflict of interest (de Oliveira 2009; Diniz Zanetti 2007; Felicissimo 2010; Konstantinidou 2007; Liebergall 2009; Savage 2005; Sriboonreung 2011a; Sriboonreung 2011b).

Effects of interventions

Comparison 1. More versus less contact with health professionals

Six trials contributed to this comparison, in three subgroups.

Three trials had the same home pelvic floor muscle training (PFMT) programme in both trial arms and investigated the effect of adding group supervision of exercise (weekly 45-minute exercise class, Bø 1990; twice-weekly 50-minute exercise class, Felicissimo 2010; weekly group session, Konstantinidou 2007) in addition to the individual supervision of participants that was the same in both treatment arms. The comparison arm had only individual supervision, therefore these three trials are also considered in Comparison 2 (individual versus group supervision), but we considered these trials were principally a comparison of less versus more health professional contact. A fourth trial also added twice-weekly 45-minute exercise sessions (Diniz Zanetti 2007) although it was not clear if this was individually or group supervised exercise. We grouped the trial by Diniz Zanetti 2007 and colleagues with those of Bø 1990, Felicissimo 2010 and Konstantinidou 2007 for analysis because of the similarity in the amount of time and frequency of extra health professional contact. All four trials recruited women with urodynamic stress incontinence. These four trials formed one subgroup in this comparison, called 'additional group supervision'.

The fifth trial also had the same home PFMT programme in both trial arms and investigated the effect of adding twice-weekly phone calls (Ng 2008) after the initial period of face-to-face contact with a

health professional. This trial recruited women with mixed urinary incontinence. This was the only trial in the second subgroup, 'additional phone calls'.

The sixth trial had differences the amount of health professional contact and also differences in the PFMT programme between groups (Hung 2010). One treatment group was given advice and instruction in a home PFMT programme, with no further health professional contact. The other treatment group completed a structured 16-week programme of 'indirect' PFMT (in that this was a combination of diaphragmatic, transversus abdominus and combined transversus abdominus/PFM contraction) with fortnightly clinic visits for four months. The 'indirect' PFMT group were asked not to perform isolated voluntary PFM contractions during the intervention period. The trial recruited women with stress or mixed urinary incontinence. This trial was the only trial in the third subgroup, 'individual supervision versus no supervision (difference in PFMT)'.

The interventions are described in more detail in Table 1. The only pre-specified outcome for which there were no data was health status or generic quality of life measures.

Primary outcome measures

Patient's perception of change in incontinence

The patient's perception of change in incontinence symptoms was measured in a variety of ways. Trialists reported data in the 'positive', that is whether participants were better. So long as we could categorise the data into 'cure' or 'improved' we calculated the inverse (i.e. not cured/incontinent, not improved) and entered these data into the meta-analysis, regardless of what instrument was used. We did not include any data for this outcome where the definition of cure or improvement was based on something other than the patient's perception of their own urinary incontinence (e.g. pad test cure, or no leakage episodes in a urinary diary, or clinician's perception). The trials reported the patient's perception of cure and improvement as follows:

Trial	Instrument	Trialists' definition	Our categorisation
Bø 1990	5-point Likert scale (worse to continent)	"continent"	cured
		"some improvement" and "almost continent"	improved (combined data)
Diniz Zanetti 2007	Satisfied with improvement and did not want further therapy	"yes"	improved
Felicissimo 2010	4-point Likert scale (cured to worse)	"cured"	cured
		"better"	improved
Konstantinidou 2007	Patient global assessment of improvement	"yes"	improved
Hung 2010	4-point Likert scale (worse to cured)	"cured"	cured
		"improved"	improved

Not cured (Analysis 1.1)

Fewer women were still incontinent in the group that received additional group supervision of PFMT (risk ratio (RR) for no cure 0.89, 95% confidence interval (CI) 0.78 to 1.03, Analysis 1.1.1; Bø 1990; Felicissimo 2010), although the difference is not statistically significant. Hung 2010 did not find any statistically significant difference between the supervised and unsupervised

groups (RR for no cure 0.86, 95% CI 0.73 to 1.02, Analysis 1.1.3). Ng 2008 reported the odds ratio (OR) (95% CI) of the difference between groups for their responses to two items from the Bristol Female Lower Urinary Tract Symptoms questionnaire (BFLUTS) (whether they did or did not have symptoms of stress incontinence or urgency incontinence). For both items the group receiving phone calls had reduced odds of stress or urgency incontinence (OR 0.49, 95% CI 0.31 to 0.76 and 0.40, 0.24 to 0.66 respectively, Ng 2008).

Not improved (Analysis 1.2)

Ten per cent (9 of 87) of those who received weekly or twice-weekly group supervision in addition to individual appointments with the therapist did not report improvement post-treatment compared to 43% (39 of 90) of the group who had individual appointments only (RR for no improvement 0.29, 95% CI 0.15 to 0.55, Analysis 1.2.1; Bø 1990; Diniz Zanetti 2007; Felicissimo 2010; Konstantinidou 2007). Looking at this another way, 90% of those who had combined group and individual supervision reported improvement versus 57% of women receiving individual supervision only. Thus, women receiving additional group supervision were more likely to report their incontinence was improved. Similarly, Hung 2010 found women in the supervised group were more likely to improve (RR for no improvement 0.10, 95% CI 0.01 to 0.71, Analysis 1.2.3).

Incontinence-specific quality of life (Analysis 1.3)

Two trials used validated measures of incontinence-specific quality of life (I-QoL, Diniz Zanetti 2007; ICIQ-SF, Felicissimo 2010). A third used a single item measure of unknown origin (Konstantinidou 2007), which we classified as a measure of symptom impact (see Secondary outcomes). The difference between the randomised groups could not be statistically evaluated in these trials (Analysis 1.3).

Secondary outcome measures

Symptom impact (Analysis 1.4)

Bø 1990 reported the Social Activity Index, which was developed for the trial. Konstantinidou 2007 reported "quality of life" that we re-classified as a measure of symptom impact because it was not clear if this was a validated measure of quality of life. Both Ng 2008 and Hung 2010 used translations of the Symptom Impact Index (Black et al 1996). The trialists tended to report the data from these various instruments by item rather than total or domain scores, although not all items were reported in all instances and not all data were complete (Analysis 1.4). It was therefore difficult to assess any consistent pattern of effect.

Frequency of leakage

Leakage episodes in 24 hours (Analysis 1.5)

One of the four trials investigating the effect of additional group supervision measured leakage episodes (Konstantinidou 2007). The women receiving additional supervision had fewer leakage episodes per day (mean difference (MD) -1.38, 95% CI -2.04 to -0.72, Analysis 1.5). Hung 2010 also measured leakage episodes and found no differences between the supervision and no supervision groups (median 0 leaks per day, IQR 0 to 0.3, in both groups).

Other measures (Analysis 1.6)

Bø 1990 reported the Leakage Index, which was developed for the trial. There are some items in the Index that favour the group receiving additional supervision, but the pattern of benefit was not consistent for all items (Analysis 1.6).

Amount of leakage

Pad, paper towel and cough tests (Analysis 1.7)

All four trials investigating the addition of group supervision used pad tests (90 seconds, Bø 1990; one-hour, Diniz Zanetti 2007; 24-hour, Felicissimo 2010 and Konstantinidou 2007). Hung 2010 used a 20-minute pad test. By and large it was not possible to estimate differences between the treatment groups (due to missing data), and where it was possible to calculate differences these were not statistically significantly different (Analysis 1.7).

Other measures (Analysis 1.8)

Konstantinidou 2007 reported the number of pad changes in 24 hours and how many women experienced wet underwear. Women in the group receiving additional supervision used fewer pads and were less likely to complain of wet underwear (Analysis 1.8).

Voiding frequency (Analysis 1.9)

Three trials reported the number of voids in 24 hours or per day (Diniz Zanetti 2007; Konstantinidou 2007; Hung 2010) and a fourth asked women about urinary frequency using an item from the BFLUTS (Hung 2010). Two trials reported whether women had nocturia or not (Hung 2010; Konstantinidou 2007). There were few data and some missing data, so it was difficult to observe any pattern in the data (Analysis 1.9).

Pelvic floor muscle (PFM) performance (Analysis 1.10)

Three trials, investigating the addition of group supervision, measured PFM performance (vaginal squeeze pressure, Bø 1990; Oxford scale, Felicissimo 2010; Konstantinidou 2007). The findings were not consistent (Analysis 1.10); one trial found a difference in favour of the additional superversion group (Bø 1990) and the other two did not (Felicissimo 2010; Konstantinidou 2007). Hung 2010, who used vaginal squeeze pressure, did not find a difference between the supervision and no supervision groups.

Treatment adherence (Analysis 1.11)

Neither Bø 1990 nor Felicissimo 2010 appeared to find that the group receiving additional superversion were more adherent to exercise (Analysis 1.11). Ng 2008, who investigated the added benefit of phone calls, said that a lack of adherence measurement was a limitation of their study.

Follow-up data (Analysis 1.12)

Only Bø 1990 followed trial participants longer term and this follow-up was confined to the group that received additional supervision. Thus no comparisons can be drawn between the two treatment groups beyond the end of the intervention period (Analysis 1.12).

Adverse events

Only one trial in this comparison made an explicit statement about adverse training events (Hung 2010), reporting that there were none.

Comparison 2. Group versus individual supervision of PFMT

We have included six trials (in two subgroups) in this comparison but all the trials have at least one other main difference between the trial arms, which means some caution is needed in interpreting these data.

The first subgroup was three of the six trials that were included in Comparison 1 (more versus less health professional contact). While these trials are included in Comparison 2 we considered Comparison 1 was the 'primary' comparison of these three trials. All three had the same PFMT programme in both trial arms and group supervision predominated in one trial arm (in terms of frequency and total time) whereas the other arm had only individual supervision (Bø 1990; Felicissimo 2010; Konstantinidou 2007). We did not include the trial by Diniz Zanetti 2007 here because it was not clear if the additional supervision in one of the trial arms was individually or in a group. This first subgroup was 'individual supervision only versus individual and group supervision', and all three trials recruited only women with stress incontinence.

The second subgroup comprised three trials that had differences in both supervision and the PFMT programmes in the treatment groups. de Oliveira 2009 randomised women to either individualised PFMT with twice-weekly 30-minute clinic visits or a generic PFMT programme with twice-weekly 45-minute group exercise classes. Liebergall 2005 and Liebergall 2009 randomised women to either 'indirect' PFMT (Paula method) with weekly 45-minute clinic visits or 'direct' PFMT at home with four weeks of weekly 30-minute group exercise classes. We called this second subgroup 'individual supervision only versus group supervision only (with difference in PFMT)'. These trials recruited women with stress incontinence (de Oliveira 2009; Liebergall 2009) or stress and mixed incontinence (Liebergall 2005).

The interventions are described in more detail in Table 2. The only pre-specified outcome for which there were no data was health status or generic quality of life measures.

Primary outcome measures

The trials reported the patient's perception of cure and improvement as follows:

Trial	Instrument	Trialists' definition	Our categorisation
Bø 1990	5-point Likert scale (worse to continent)	"continent"	cured
		"some improvement" and "almost continent"	improved (combined data)
de Oliveira 2009	How patient felt after treatment - satisfied or dissatisfied	"subjective SUI cure"	improved
Felicissimo 2010	4-point Likert scale (cured to worse)	"cured"	cured

(*Continued*)

		"better"	improved
Konstantinidou 2007	Patient global assessment of improvement	"yes"	improved

Although de Oliveira 2009 categorised the response "satisfied" as a cure, we felt with only two response options this question probably did not differentiate women who were cured from those who were improved. We chose to categorise these data as improved.

Patient perception of change in incontinence

Not cured (Analysis 2.1)

There was no statistically significant difference in the number of women reporting they were still incontinent between the group and individual supervision arms (RR for no cure 0.89, 95% CI 0.78 to 1.03, Analysis 2.1.1; Bø 1990; Felicissimo 2010) in the first subgroup. These were the same two trials that contributed data to the same outcome in Comparison 1.

Not improved (Analysis 2.2)

Women receiving group supervision were more likely to report that their incontinence was improved (RR for no improvement 0.16, 95% CI 0.05 to 0.46, Analysis 2.2.1; Bø 1990; Felicissimo 2010; Konstantinidou 2007). In the second subgroup, de Oliveira 2009 found no difference between group and individual supervision (RR 1.20, 95% CI 0.61 to 2.34, Analysis 2.2.2).

Incontinence-specific quality of life (Analysis 2.3)

In the first subgroup, one trial used a validated measure of incontinence-specific quality of life (ICIQ-SF, Felicissimo 2010) but did not report data for the group supervision arm (Analysis 2.3).
In the second subgroup all three trials used a validated incontinence-specific quality of life measure (King's Health Questionnaire (KHQ), de Oliveira 2009; I-QoL, Liebergall 2005; Liebergall 2009). de Oliveira 2009 found no difference between the trial arms for overall incontinence impact, and neither Liebergall 2005 nor Liebergall 2009 found a difference in the change in total score between arms.

Secondary outcome measures

Symptom impact (Analysis 2.4)

Bø 1990 reported the Social Activity Index, which was developed for the trial. Konstantinidou 2007 reported "quality of life" that we re-classified as a measure of symptom impact because it was not clear if this was a validated measure of quality of life. There were too few data (some of which were incomplete) to draw any conclusions (Analysis 2.4).

Frequency of leakage

Leakage episodes in 24 hours (Analysis 2.5)

Two trials measured leakage episodes, one with (de Oliveira 2009) and one without (Konstantinidou 2007) a difference in the PFMT programmes in addition to the group versus individual supervision comparison. In the first there was no difference between the trial arms (MD 0.10, 95% CI -0.16 to 0.36, Analysis 2.5.2; de Oliveira 2009) and in the second women receiving group supervision had fewer leakage episodes per day (MD -1.38, 95% CI -2.04 to -0.72, Analysis 2.5.1; Konstantinidou 2007).

Other measures of leakage frequency (Analysis 2.6)

Bø 1990 reported the Leakage Index, which was developed for the trial. There are some items in the Index that favour the group receiving additional supervision, but the pattern of benefit is not consistent for all items (Analysis 2.6). Liebergall 2009 did not find any difference between the groups in the number of women with incontinence episodes weekly or more often.

Amount of leakage

Pad, paper towel and cough tests (Analysis 2.7)

All three trials in the first subgroup used pad tests ((90 seconds), Bø 1990; 24-hour, Felicissimo 2010 and Konstantinidou 2007). Where it was possible to calculate differences these were not statistically significantly different (Analysis 2.7). All three trials in the second subgroup used one-hour pad tests; one of these used a standardised bladder volume (de Oliveira 2009). The overall pattern was one of no difference between the groups.

Other measures (Analysis 2.8)

Konstantinidou 2007 reported the number of pad changes in 24 hours and how many women experienced wet underwear. Women in the group receiving additional supervision used fewer pads and were less likely to complain of wet underwear (Analysis 2.8). Liebergall 2009 did not find any difference between the groups in the number of women with self reported moderate amounts of leakage or more.

Voiding frequency (Analysis 2.9)

One trial reported the number of voids in 24 hours (Konstantinidou 2007) and two trials reported whether women had nocturia or not (Konstantinidou 2007; Liebergall 2009). There were no differences between the groups (Analysis 2.9).

PFM performance (Analysis 2.10)

All three trials in the first subgroup measured PFM performance (vaginal squeeze pressure, Bø 1990; Oxford scale, Felicissimo 2010; Konstantinidou 2007). The findings were not consistent; one trial found a difference in favour of the additional superversion group (Bø 1990) and the other two did not (Felicissimo 2010; Konstantinidou 2007) (Analysis 2.10).
In the second subgroup, de Oliveira 2009 used the Oxford scale and Liebergall 2005 measured vaginal squeeze pressure. In the former women receiving individual supervision with an individualised PFMT programme had greater palpable strength than group supervised women doing a generic training programme. In the latter trial there was no difference between groups.

Treatment adherence (Analysis 2.11)

Neither Bø 1990 nor Felicissimo 2010 appeared to find that women receiving group superversion were more adherent to exercise (Analysis 2.11). Liebergall 2009 found that fewer women who were individually supervised attended less than 50% of their appointments, but there was no difference in the proportion who documented exercising at home.

Follow-up data (Analysis 2.12)

Only Bø 1990 followed trial participants in the longer term and this follow-up was confined to the group that received additional supervision. Thus no comparisons can be drawn between the two treatment groups beyond the end of the intervention period (Analysis 2.12).

Adverse events

Only one trial in this comparison made an explicit statement about adverse training events (Liebergall 2005), reporting that there were none.

Comparison 3. 'Direct' versus 'indirect' methods of PFMT

There were six trials, in four subgroups within this comparison.
The first comprised two trials in women with stress incontinence that compared PFMT versus sham or imitation PFMT (Ghoniem 2005; Ramsay 1990). Both trials used the same sham PFMT programme, which was strong hip abduction with the ankles crossed. We categorised the sham PFMT as 'indirect' PFMT exercise because electromyography (EMG), dynamometric and magnetic resonance imaging (MRI) studies show that both hip abduction and external rotation give synergistic contraction of the PFM (Bø 1994; Dumoulin 2006; Morin 2004). Both trials used the same exercise parameters in both trial arms with regard to duration of contraction and rest, number of contractions per day and so on.
The second subgroup was the two trials comparing PFMT versus the 'Paula method' in women with stress (Liebergall 2009) or stress and mixed incontinence (Liebergall 2005). The 'Paula method' did include some direct PFM contractions and also contraction of other 'sphincters' (such as contraction of muscle closing the mouth and eyes). With a large part of the programme dedicated to contraction of 'other' muscle groups, we categorised the 'Paula method' as 'indirect' PFMT. In addition to the differences in PFMT these two trials had differences in supervision (group versus individual). Liebergall 2005 and Liebergall 2009 randomised women to either 'indirect' PFMT (Paula method) with weekly 45-minute clinic visits or 'direct' PFMT at home with four weeks of weekly 30-minute group exercise classes. These trials had differences in both PFMT programmes and type of supervision, so are also included in Comparison 2.
Third was the single trial that compared PFMT with the 'Sapsford' approach in women with stress or mixed incontinence (Hung 2010). The 'Sapsford' approach comprised a combination of diaphragmatic, transversus abdominus and co-contraction of transversus abdominus/PFM manoeuvres. The women activated and co-contracted their PFM through a contraction of abdominal muscle; participants in this intervention were asked not to perform isolated voluntary PFM contractions during the intervention period. Thus, we categorised the 'Sapsford' approach as 'indirect'

PFMT. This trial also had differences in the amount of health professional contact between groups, and was also considered in Comparison 1. The 'direct PFMT' group was given advice and instruction in a home PFMT programme, with no further health professional contact. The other group completed a structured 16-week programme of 'indirect' PFMT with fortnightly clinic visits. The fourth subgroup was also a single trial, a pilot study, that compared PFMT versus Pilates in women with stress incontinence (Savage 2005). The Pilates group did a range of exercises involving the muscles around the pelvis and hip, but no isolated PFM contractions so we classified the Pilates as 'indirect' PFMT. Both trial arms had the same number of individual health professional contacts, and one group was asked to do PFMT daily at home and the Pilates group were asked to exercise for 10 to 15 minutes on alternate days. The Pilates group were asked not to perform isolated PFM contractions although women did some co-contraction of transversus abdominus/PFM as part of their Pilates programme. The interventions are described in more detail in Table 3. There were no data for two pre-specified outcomes of interest: health status or generic quality of life measures and longer-term follow-up.

Primary outcome measures

The trials reported the patient's perception of cure and improvement as follows:

Trial	Instrument	Trialists' definition	Our categorisation
Ghoniem 2005	Patient Global Impression of Improvement Scale	"a little better", "much better", "very much better"	improved
Hung 2010	4-point Likert scale (worse to cured)	"cured"	cured
		"improved"	improved
Ramsay 1990	"subjective"	"improvement"	improved

Patient perception of change in incontinence

Not cured (Analysis 3.1)

A single trial contributed to this comparison (Hung 2010) and did not find any difference between 'direct' PFMT and the 'Sapsford' approach (RR 1.16, 95% CI 0.98 to 1.36, Analysis 3.1.1).

Not improved (Analysis 3.2)

The pooled data from the two trials comparing PFMT with imitation PFMT did not show a statistically significant difference in favour of either group (RR 0.69, 95% CI 0.47 to 1.02, Analysis 3.2.1; Ghoniem 2005; Ramsay 1990) although this was close to being in favour of more improvement in the 'direct' PFMT group. In contrast, Hung 2010 found that women in the 'indirect' PFMT group (with more health professional contact) were more likely to improve (RR for no improvement 10.33, 95% CI 1.42 to 75.41, Analysis 3.2.3) but the trial was small and the confidence interval wide.

Savage 2005 did not report improvement, but did measure satisfaction with treatment. There was no difference between the groups in the percentage satisfaction (MD -12.90, 95% CI -45.22 to 19.42, Savage 2005).

Incontinence-specific quality of life (Analysis 3.3)

Four trials used a validated incontinence-specific quality of life measure (KHQ, Savage 2005; I-QoL, Ghoniem 2005; Liebergall 2005; Liebergall 2009). Neither the data from the pilot study by Savage 2005 or the trial by Ghoniem 2005 and colleagues could be used to calculate an effect estimate. Neither Liebergall 2005 nor Liebergall 2009 found a difference in the change in total score between the two groups (Analysis 3.3).

Secondary outcome measures

Symptom impact (Analysis 3.4)

Hung 2010 used a translation of the Symptom Impact Index and reported only two item scores (presumably the ones that showed the greatest difference) rather than the total score (Analysis 3.4).

Frequency of leakage

Leakage episodes in 24 hours

Hung 2010 (median 0 leaks per day, interquartile range (IQR) 0 to 0.3, in both groups) found no difference between groups.

Other measures (Analysis 3.5)

Ghoniem 2005 did not find a difference in the number of women in each group who had a less than 50% reduction in leakage frequency, and Liebergall 2009 did not find any difference between the groups in the number of women with incontinence episodes weekly or more often (Analysis 3.5).

Amount of leakage

Pad, paper towel and cough tests (Analysis 6)

Four trials reported pad test data (20-minute, Hung 2010; one-hour, Liebergall 2005; Liebergall 2009; type not known, Ramsay 1990). A difference between groups could not be calculated based on the reported data from either Hung 2010 or Ramsay 1990. Neither Liebergall 2005 nor Liebergall 2009 found a difference between groups for the mean change in pad weight gain (baseline to post-intervention) (Analysis 3.6).

Other measures (Analysis 3.7)

Liebergall 2009 did not find any difference between the groups in the number of women with self reported moderate amounts of leakage or more (Analysis 3.7).

Voiding frequency (Analysis 3.8)

Two trials reported whether women had nocturia or not (Hung 2010; Liebergall 2009). There were too few useable data to measure any pattern (Analysis 3.8).

PFM performance (Analysis 3.9)

Both Hung 2010 and Liebergall 2005 measured vaginal squeeze pressure and neither trial found a difference between groups. Savage 2005 measured both endurance (in seconds) and strength (using the Oxford scale) and there was no difference between the groups for either post-treatment or change scores (Analysis 3.9).

Treatment adherence (Analysis 3.10)

Neither Ghoniem 2005 nor Ramsay 1990 appeared to find any important differences between the groups for the proportion of prescribed contractions performed or the number of exercise sessions per week recorded in a diary. Liebergall 2009 found that the women in the 'indirect' PFMT group (with individual supervision) were more likely to attend more than 50% of their appointments, but there was no difference between 'direct' and 'indirect' groups in the proportion who documented exercising at home (Analysis 3.10).

Adverse events

Two trials made an explicit statement about adverse training events (Hung 2010; Liebergall 2005), reporting that there were none.

Comparison 4. 'Individualised versus generic PFMT

In one small trial (de Oliveira 2009) randomised women to either individualised PFMT with twice-weekly 30-minute clinic visits or a generic PFMT programme with twice-weekly 45-minute group exercise classes. Thus, this trial had differences in both PFMT programmes and supervision. It is also included in Comparison 2. The interventions are described in more detail in Table 4. There were no data for one of the pre-specified primary outcomes of interest (self reported cure) or five other secondary outcomes of interest (health status or generic quality of life measures, symptom impact, voiding frequency, longer-term follow-up and adverse events).

Primary outcome measures

One small trial (de Oliveira 2009) asked women if they were satisfied or dissatisfied after treatment and categorised women who were satisfied as a "subjective SUI cure". We felt with only two response options this question probably did not differentiate women who were cured from those who were improved. We chose to categorise these data as improved.

Patient perception of change in incontinence

Not improved (Analysis 4.1)

One small trial (de Oliveira 2009) found no difference between generic and individualised exercise groups (RR 0.83, 95% CI 0.43 to 1.63, Analysis 4.1.1).

Incontinence-specific quality of life (Analysis 4.2)

One small trial (de Oliveira 2009) used a validated incontinence-specific quality of life measure (King's Health Questionnaire) and found no difference between the trial arms for overall incontinence impact (Analysis 4.2).

Secondary outcome measures

Frequency of leakage

Leakage episodes in 24 hours (Analysis 4.3)

There was no difference between the generic and individualised exercise groups (MD -0.10, 95% CI -0.36 to 0.16, Analysis 4.3; de Oliveira 2009) in one small trial.

Amount of leakage

Pad, paper towel and cough tests (Analysis 4.4)

de Oliveira 2009 used a one-hour pad test and found no difference between the groups (Analysis 4.4).

PFM performance (Analysis 4.5)

de Oliveira 2009 used the Oxford scale and found that women receiving individual supervision with an individualised PFMT programme had greater palpable strength than group supervised women doing a generic training programme (Analysis 4.5).

Treatment adherence (Analysis 4.6)

A high level of adherence was reported in both groups, but it was not clear what was measured (Analysis 4.6).

Comparison 5. Near-maximal versus submaximal contractions

In one small trial, Johnson 2001 compared two PFMT programmes that were the same in all respects except one group did near-maximal contractions and the other submaximal PFM contractions. Both groups had the same amount of contact with health professionals. The women in the study had stress incontinence. The interventions are described in more detail in Table 5. There were no data for any of the pre-specified primary outcomes of interest (self reported cure, self reported improvement, incontinence-specific quality of life) or six other secondary outcomes of interest (health status or generic quality of life measures, symptom impact, voiding frequency, treatment adherence, longer-term follow-up and adverse events).

Secondary outcome measures

Frequency of leakage

Leakage episodes in 24 hours (Analysis 5.1)

There was no statistically significant difference between the exercise groups for leakage episodes in 24 hours (MD -0.36, 95% CI -1.85 to 1.13, Analysis 5.1).

Other measures (Analysis 5.2)

There was no difference in the proportion of women in each group who experienced leakage in the last week of the study (Analysis 5.2).

Amount of leakage

Pad, paper towel and cough tests (Analysis 5.3)

There was no difference in pad weight between the groups in a 10-hour pad test (Analysis 5.3).

PFM performance (Analysis 5.4)

Johnson 2001 reported three measures of PFM performance. There were no differences between the groups for maximal vaginal EMG activity, vaginal squeeze pressure with a maximal contraction, or endurance (seconds) (Analysis 5.4).

Comparison 6. Daily versus three times per week PFMT

In one small trial, Sriboonreung 2011a compared two PFMT programmes that were the same in all respects except one group were asked to exercise daily and the other group three times a week. Both groups had the same amount of contact with health professionals. The women in the study had stress incontinence. The interventions are described in more detail in Table 6. There were no data for one of the pre-specified primary outcomes of interest (incontinence-specific quality of life) or six other secondary outcomes of interest (health status or generic quality of life measures, symptom impact, frequency of leakage, voiding frequency, treatment adherence and longer-term follow-up).

Primary outcome measures

Sriboonreung 2011a asked women about their "satisfaction with their incontinence condition" at the end of treatment on a five-point scale (worse to continent). We categorised "continent" as cure, and "improved" and "almost continent" as improved.

Patient perception of change in incontinence

Not cured (Analysis 6.1)

There was no statistically significant difference between the groups in the number of women who were still incontinent (RR for no cure 1.18, 95% CI 0.84 to 1.65, Analysis 6.1).

Not improved (Analysis 6.2)

All of the women in both groups reported some improvement in incontinence (Analysis 6.2).

Secondary outcome measures

Amount of leakage

Pad, paper towel and cough tests (Analysis 6.3)

There was no difference between the groups on a one-hour pad test (Analysis 6.3).

PFM performance (Analysis 6.4)

Women in the daily PFMT group had a statistically significantly greater change in vaginal squeeze pressure than the three times a week PFMT group, although there was no difference in maximal squeeze pressure (Analysis 6.4).

Adverse events

This trial stated that no adverse events were reported.

Comparison 7. Upright and supine versus supine exercise positions alone

In one small trial, Borello-France 2006 compared two PFMT programmes that were the same in all respects except one group alternated exercise sets between supine, sitting and standing (labelled upright) and the other group were asked to exercise only in the supine position. Both groups had the same amount of contact with health professionals. The women in the study had stress incontinence. The interventions are described in more detail in Table 7. There were no data for two of the pre-specified primary outcomes of interest (self report of cure or improvement) or five other secondary outcomes of interest (health status or generic quality of life measures, symptom impact, voiding frequency, longer-term follow-up and adverse events).

Primary outcome measures

Incontinence-specific quality of life (Analysis 7.1)

Borello-France 2006 used a validated incontinence-specific quality of life measure (Incontinence Impact Questionnaire) and found no difference between the trial arms for change in incontinence quality of life (Analysis 7.1).

Secondary outcome measures

Frequency of leakage

Leakage episodes in 24 hours (Analysis 7.2)

There was no statistically significant difference between the exercise groups for leakage episodes in 24 hours (MD 0.20, 95% CI -0.24 to 0.64, Analysis 7.2).

Amount of leakage

Pad, paper towel and cough tests (Analysis 7.3)

There was no difference between the groups for the mean change in pad weight gain on one-hour pad test (Analysis 7.3).

PFM performance (Analysis 7.4)

There was no difference between the groups for the mean change in vaginal palpation (Brink) score (Analysis 7.4).

Treatment adherence (Analysis 7.5)

There was no difference in the mean number of clinic visits completed in each group (Analysis 7.5).

Comparison 8. Strength and motor learning versus motor learning PFMT alone

In the one trial which addressed this comparison, Hay-Smith 2002 compared two motor learning PFMT programmes that were the same in all respects, but one group had additional strengthening PFMT. Both groups had the same amount of contact with health professionals. The women in the study had stress or mixed urinary incontinence. The interventions are described in more detail in Table 8. There were no data for five pre-specified outcomes of interest: health status or generic quality of life measures, symptom impact, PFM performance, treatment adherence and longer-term follow-up.

Primary outcome measures

Women were asked to rate treatment outcome on a six-point Likert scale (cured to much worse). We categorised "cured" as cured, and "much better" and "somewhat improved" as improved.

Patient perception of change in incontinence

Not cured (Analysis 8.1)

There was no statistically significant difference between the groups in the number of women who reported they were not cured (RR 1.05, 95% CI 0.98 to 1.13, Analysis 8.1).

Not improved (Analysis 8.2)

There was no difference between the groups in the number of women who reported they were not improved (RR 0.65, 95% CI 0.31 to 1.40, Analysis 8.2).

This trial considered three other outcomes indicative of self reported improvement and found no difference between the combination strength and motor learning versus motor learning groups for satisfaction with treatment (RR 1.15, 95% CI 0.74 to 1.81) or being comfortable to continue with training (RR 0.94, 95% CI 0.83 to 1.08), although women in the motor learning PFMT group were less likely to want further treatment (RR 0.72, 95% CI 0.53 to 0.96).

Incontinence-specific quality of life (Analysis 8.3)

Hay-Smith 2002 used a validated incontinence-specific quality of life measure (King's Health Questionnaire) and found no difference between the trial arms for overall incontinence impact (Analysis 8.3).

Secondary outcome measures

Frequency of leakage

Leakage episodes in 24 hours (Analysis 8.4)

There was no statistically significant difference between the groups for the number of leakage episodes per day (MD -0.20, 95% CI -0.55 to 0.15, Analysis 8.4).

Amount of leakage

Pad, paper towel and cough tests (Analysis 8.5)

There was no difference between the groups on either a paper towel test, or a 24-hour pad test (Analysis 8.5).

Other measures (Analysis 8.6)

There was no difference in the proportion of women per group making one or more pad changes in 24 hours (Analysis 8.6).

Voiding frequency (Analysis 8.7)

There was no difference between the groups for the average number of voids in 24 hours (Analysis 8.7).

Adverse events

The trial stated that no adverse events were reported by participants.

Comparison 9. PFMT and abdominal muscle exercise versus PFMT alone

In one small trial, Sriboonreung 2011b compared two PFMT programmes that were the same in all respects except one group were asked to do additional abdominal muscle exercise (not further described). Both groups had the same amount of contact with health professionals. The women in the study had stress incontinence. The interventions are described in more detail in Table 9. There were no data for one of the pre-specified primary outcomes of interest (incontinence-specific quality of life) or six other secondary outcomes of interest (health status or generic quality of life measures, symptom impact, frequency of leakage, voiding frequency, treatment adherence and longer-term follow-up).

Primary outcome measures

Sriboonreung 2011b asked women about their "satisfaction with their incontinence condition" at the end of treatment on a five-point scale (worse to continent). We categorised "continent" as cure, and "improved" and "almost continent" as improved.

Patient perception of change in incontinence

Not cured (Analysis 9.1)

There was no statistically significant difference between the groups in the number of women who were still incontinent (RR for no cure 0.90, 95% CI 0.63 to 1.29, Analysis 9.1).

Not improved (Analysis 9.2)

All of the women in both groups reported some improvement in incontinence (Analysis 9.2).

Secondary outcome measures

Amount of leakage

Pad, paper towel and cough tests (Analysis 9.3)

There was no difference between the groups on a one-hour pad test (Analysis 9.3).

PFM performance (Analysis 9.4)

Women in the PFMT with abdominal muscle exercise group had a statistically significantly greater change in vaginal squeeze pressure than the PFMT group, although there was no difference in maximal pressure (Analysis 9.4).

Adverse events

This trial stated that no adverse events were reported.

Comparison 10. PFMT with intravaginal resistance device versus PFMT alone

Three small trials compared two PFMT programmes that were the same in all respects except one group were asked to exercise with the addition of an intravaginal device to resist the PFM contraction (Delgado 2009; Ferguson 1990; Wells 1999). The resistance devices were: a spring-loaded device with two limbs (Delgado 2009); an intravaginal balloon (Ferguson 1990); and a vaginal dilator (Wells 1999). In all three trials, both arms had the same amount of contact with health professionals. The women in the study had stress incontinence only (Ferguson 1990), or stress or mixed incontinence (Delgado 2009; Wells 1999).

The interventions are described in more detail in Table 10. There were no data for one of the pre-specified primary outcomes of interest (incontinence-specific quality of life) or three other secondary outcomes of interest (health status or generic quality of life measures, voiding frequency and treatment adherence).

Primary outcome measures

The trials reported the patient's perception of cure and improvement as follows:

Trial	Instrument	Trialists' definition	Our categorisation
Delgado 2009	5-point Likert scale (never to all of time) for question 11a of the ICIQ-FLUTS (Does urine leak when you are physically active, exert yourself, cough or sneeze?)	"never"	cured
		"Improvement" is positive change by 2 or more points on the scale	improved
Wells 1999	10-point VAS (no leakage to a lot of leakage)	"cure" is "not post-treatment wetting"	cure
		"better" is 2 or more points lower	improved

Patient perception of change in incontinence

Not cured (Analysis 10.1)

There was no statistically significant difference between the groups in the number of women who were not cured (RR 1.07, 95% CI 0.96 to 1.20, Analysis 10.1; Delgado 2009; Wells 1999).

Not improved (Analysis 10.2)

There was no statistically significant difference between the groups in the number of women who were not improved (RR 0.86, 95% CI 0.62 to 1.20, Analysis 10.2; Delgado 2009; Wells 1999).

Secondary outcome measures

Symptom impact (Analysis 10.3)

Equal proportions of women in the two arms of one trial (Delgado 2009) reported moderate or a lot of leakage on physical activity (Analysis 10.3).

Frequency of leakage

Leakage episodes in 24 hours (Analysis 10.4)

Wells 1999 found no statistically significant difference between the two treatment groups (MD 0.20, 95% CI -0.13 to 0.53, Analysis 10.4).

Other measures (Analysis 10.5)

Wells 1999 did not find any difference in the proportion of women who were the same or worse based on the number of leakage episodes in the urinary diary after treatment (Analysis 10.5).

Amount of leakage

Pad, paper towel and cough tests (Analysis 10.6)

Three trials reported pad test data (30-minute, Ferguson 1990; 24-hour, Ferguson 1990; unspecified duration, Wells 1999). There were no statistically significant differences between the treatment groups (Analysis 10.6).

Other measures (Analysis 10.7)

Wells 1999 did not find any difference between the two groups for the average score on a visual analogue scale (VAS) for leakage (none to a lot) (Analysis 10.7).

PFM performance (Analysis 10.8)

Two trials measured PFM performance (vaginal squeeze pressure, Ferguson 1990; vaginal palpation (Brink) score, Wells 1999; EMG, Wells 1999). The findings were consistent in that none of the measures showed a statistically significant difference between the groups (Analysis 10.8).

Follow-up data

Ferguson 1990 followed participants up at 12 to 24 months post intervention and 19 of 20 trial participants responded. The data were not reported by group assignment. Nine women were exercising and 10 were not. None rated their symptoms as worse than prior to the trial, three had gynaecological surgery and two had return of symptoms when they stopped exercising (one of whom controlled her symptoms with resumption of exercise).

Adverse events

One trial made an explicit statement about adverse training events (Delgado 2009), reporting that there were none.

Comparison 11. PFMT and adherence strategy versus PFMT alone

Gallo 1997 and Sugaya 2003 aimed to improve exercise adherence by giving women an audiotape to play in the car or at home, or a small chiming (alarm) device respectively. The PFMT programmes were (Sugaya 2003) or appeared to be (Gallo 1997) the same both trial arms. Both trial arms in both trials had the same amount of contact with health professionals. The women in both trials had stress incontinence.

The interventions are described in more detail in Table 11. There were no data for two of the pre-specified primary outcomes of interest (self reported cure, incontinence-specific quality of life) or six other secondary outcomes of interest (health status or generic quality of life measures, symptom impact, voiding frequency, PFM performance, longer-term follow-up and adverse events).

Primary outcome measures

Sugaya 2003 reported a "Quality of life index for urination" using a seven-point Likert-type scale for responses (delighted to terrible). It was not clear if this was a genuine attempt (using an unvalidated instrument) to measure quality of life or was more a measure of satisfaction with treatment outcome. We decided to categorise three responses (delighted, pleased, mostly satisfied) as the patient's perception of improvement.

Patient perception of change in incontinence

Not improved (Analysis 11.1)

The group using the alarm reminder device were more likely to improve after treatment (RR for no improvement, 0.56, 95% CI 0.34 to 0.91, Analysis 11.1).

Secondary outcome measures

Frequency of leakage

Leakage episodes in 24 hours (Analysis 11.2)

Sugaya 2003 found no statistically significant difference between the groups for the number of leakage episodes in 24 hours (MD -0.50, 95% CI -1.55 to 0.55, Analysis 11.2).

Amount of leakage

Pad, paper towel and cough tests (Analysis 11.3)

The average amount of leakage on one-hour pad test was statistically significantly less (12 g) after treatment in the group using the device (MD -12 g, 95% CI -21 to -2, Analysis 11.3; Sugaya 2003).

Other measures (Analysis 11.4)

However, there was no statistically significant difference between groups in the average number of pads used per day (Sugaya 2003) (Analysis 11.4).

Treatment adherence (Analysis 11.5)

Sugaya 2003 was able to use the device to record adherence to the prescribed exercise programme; the group using the device completed 25% to 100% of the recommended exercise. There was no measure of adherence in the other group (Analysis 11.5).
Gallo 1997 found that women allocated to the group with the audiotape to support exercise at home were much more likely to do "routine" PFMT and much more likely to be exercising twice per day as recommended (Analysis 11.5).

Subgroup analysis

Having planned subgroup analysis by diagnostic category (Subgroup analysis and investigation of heterogeneity) we grouped the trials accordingly:

- 15 trials in women with only stress urinary incontinence (Bø 1990; Borello-France 2006; de Oliveira 2009; Diniz Zanetti 2007; Felicissimo 2010; Ferguson 1990; Gallo 1997; Ghoniem 2005; Johnson 2001; Konstantinidou 2007; Liebergall 2009; Ramsay 1990; Savage 2005; Sriboonreung 2011a; Sriboonreung 2011b; Sugaya 2003).
- Five trials in women with only stress or mixed urinary incontinence (Delgado 2009; Hay-Smith 2002; Hung 2010; Liebergall 2005; Wells 1999).
- One trial in women with only mixed urinary incontinence (Ng 2008).

We started our analysis using these subgroups, but stopped because it was not worthwhile. Few trials contributed data to any single comparison in the review. Then, because the bulk of the trials were in the same subgroup ('stress incontinence only'), this was often the only subgroup within the comparison.
Instead we decided, post hoc, to use subgroup analysis to manage the heterogeneity in the interventions within each comparison. For example, in Comparison 3 ('direct' versus 'indirect' approaches to PFMT) we were able to subgroup the trials according to the type of 'indirect' training used. Using this approach, we were also able to subgroup studies within a comparison to alert readers to other potentially confounding variables such as a difference in both amount of supervision and content of PFMT programme (for example, see Comparison 1: More versus less contact with health professionals).

Comparison 12: More intensive versus less intensive PFMT programmes

At the end of the process our impression was that outcomes tended to favour the most intensive intervention within any comparison. Post hoc, we decided to perform an 'all in one analysis' of the trials according to the 'contrast' between the two interventions (i.e. 'high' contrast, 'moderate' contrast and 'low' contrast). This required us to decide how much contrast there was. Our categorisation is summarised in the table below. In making our decision the two criteria were the **amount of face-to-face health professional contact** and the **exercise intensity**.

Our final categorisation	Study	Contact intensity	Exercise intensity
'High' contrast	Bø 1990	High contrast: 6 individual and 24 group contacts versus 6 individual contacts	No contrast: same PFMT

(*Continued*)

	Felicissimo 2010	High contrast: 1 individual and 16 group contacts versus 1 individual contact	No contrast: same PFMT
	Konstantinidou 2007	High contrast: 3 individual and 12 group contacts versus 3 individual contacts	No contrast: same PFMT
	Hung 2010	High contrast: 8 individual versus no contacts	Moderate contrast: 'indirect' PFMT (Sapsford, no direct PFMT) versus 'direct' PFMT (no detail)
	Ghoniem 2005	No contrast: same amount of contact	High contrast: 'indirect PFMT (cross leg sham, no direct PFMT) versus 'direct' PFMT
	Ramsay 1990	No contrast: same amount of contact	High contrast: 'indirect PFMT (cross leg sham, no direct PFMT) versus 'direct' PFMT
	Savage 2005	No contrast: same amount of contact	High contrast: 'indirect PFMT (Pilates, no direct PFMT) versus 'direct' PFMT
'Moderate' contrast	Ng 2008	Moderate contrast: 8 individual and 40 phone contacts versus 8 individual contacts	No contrast: same PFMT
	Diniz Zanetti 2007	Moderate contrast: 3 individual and 6 group contacts versus 3 individual contacts	No contrast: same PFMT
	Liebergall 2005	Moderate contrast: 12 individual versus 4 group contacts	Moderate contrast: 'indirect' PFMT (Paula, with some direct PFM contractions) versus 'direct' PFMT
	Liebergall 2009	Moderate contrast: 12 individual versus 6 group contacts	Moderate contrast: 'indirect' PFMT (Paula, with some direct PFM contractions) versus 'direct' PFMT
'Low' contrast	de Oliveira 2009	Low contrast: 12 individual versus 12 group contacts	Low contrast: individualised versus generic PFMT
	Sriboonreung 2011b	No contrast: no difference in contact	Low contrast: daily PFMT versus 3x weekly PFMT
	Johnson 2001	No contrast: no difference in contact	Low contrast: near maximal versus submaximal PFM contractions

(*Continued*)

	Hay-Smith 2002	No contrast: no difference in contact	Low contrast: strengthening with motor relearning PFMT versus motor relearning alone
	Borello-France 2006	No contrast: no difference in contact	Low contrast: upright and supine PFMT versus supine PFMT
	Sriboonreung 2011a	No contrast: no difference in contact	Low contrast: PFMT plus abdominal muscle exercises versus PFMT
	Delgado 2009	No contrast: no difference in contact	Low contrast: PFMT plus intravaginal resistance device versus PFMT
	Ferguson 1990	No contrast: no difference in contact	Low contrast: PFMT plus intravaginal resistance device versus PFMT
	Wells 1999	No contrast: no difference in contact	Low contrast: PFMT plus intravaginal resistance device versus PFMT
	Gallo 1997	No contrast: no difference in contact	Low contrast: PFMT plus adherence strategy versus PFMT
	Sugaya 2003	No contrast: no difference in contact	Low contrast: PFMT plus adherence strategy versus PFMT

For the contact intensity to be 'high' contrast there was five times more face-to-face contact in one arm compared to the other. 'Low' intensity contrast in contact was where the number of face-to-face contacts was the same but the mechanism (e.g. individual versus group) was the same. Any other differences in contact were classified as 'moderate' contrast.

Trials that compared a 'direct' PFMT programme with an exercise programme that had no direct contractions of the pelvic floor muscles were classified as 'high' contrast comparisons for exercise intensity. 'Low' intensity contrasts in exercise intensity were those in which direct PFMT was used in both arms, with some difference in type of contraction (e.g. near maximal versus maximal), frequency of exercise (e.g. daily versus three times a week) or an additional element (such as intravaginal resistance, or adherence strategy).

To gather as much data as possible to investigate this we created three forest plots with all the available trial data for three outcomes (not cured, not improved and leakage episodes in 24 hours). We used a random-effects model for the pooled data, as there was considerable heterogeneity in the trials, and a random-effects model gives a more conservative estimate of effect.

Patient perception of change in incontinence - not cured

In the 'High' contrast comparison 83% of women (69 of 83) receiving the most intensive therapy were not cured versus 95% of the less intensive therapy group (87 of 92), a difference of 8% in favour of more intensive PFMT. This was a statistically significant difference (RR for no cure 0.89, 95% CI 0.80 to 0.98, Analysis 12.1.1, Bø 1990; Felicissimo 2010; Hung 2010) although the confidence interval was moderately wide. All three trials in this subgroup had substantively more health professional contact in the more 'intensive' treatment arm.

The reverse was observed in the 'Low' contrast comparison, where 92% of the more intensive therapy group (148 of 161) were not cured versus 88% of the less intensive therapy group (126 of 143), a difference of 4% in favour of the less intensive PFMT. This was close to a statistically significant difference (RR for no cure 1.06, 95% CI 1.00 to 1.13, Analysis 12.1.3, Delgado 2009; Hay-Smith 2002; Sriboonreung 2011a; Sriboonreung 2011b; Wells 1999). None of the trials contributing data to the 'Low' contrast subgroup had a difference in the amount of health professional contact between the trial arms.

Patient perception of change in incontinence - not improved

There was a statistically significant difference in each subgroup in favour of the more intensive PFMT, with greater treatment effect where there was more contrast between the interventions.

In the 'High' contrast subgroup 17% of women (29 of 166) receiving the most intensive therapy were not improved versus 40% of the less intensive therapy group (68 of 169), a difference of 23% in favour of more intensive PFMT. This was a statistically significant difference (RR for no improvement 0.37, 95% CI 0.17 to 0.84, random-effects, Analysis 12.2.1, Bø 1990; Felicissimo 2010; Ghoniem 2005; Hung 2010; Konstantinidou 2007; Ramsay 1990) although the confidence interval is wide. There was statistically significant heterogeneity (I^2 = 61%, Analysis 12.2.1) in this subgroup comparison. The two trials that contributed most to the heterogeneity were those by Ghoniem 2005 and Ramsay 1990, both of which were comparisons of PFMT with sham PFMT, with the same health professional contact in both groups. The other trials in this comparison all had a high contrast in the amount of health professional contact between the trial arms. When the two trials by Ghoniem 2005 and Ramsay 1990 are removed from the subgroup analysis the likelihood of cure increased and there was no statistically significant heterogeneity (RR for no improvement 0.17, 95% CI 0.06 to 0.44, random-effects, I^2 = 0%) and the width of the confidence interval decreased somewhat.

There was only one small trial in the 'Moderate' contrast subgroup (RR for no improvement 0.34, 95% CI 0.17 to 0.71, Analysis 12.2.2). In the 'Low' contrast subgroup 28% of women (59 of 212) receiving the most intensive therapy were not improved versus 40% of the less intensive therapy group (78 of 193), a difference of 12% in favour of more intensive PFMT. This was a statistically significant difference (RR for no improvement 0.75, 95% CI 0.59 to 0.95, random-effects, Analysis 12.2.3, de Oliveira 2009; Delgado 2009; Hay-Smith 2002; Sriboonreung 2011a; Sriboonreung 2011b; Sugaya 2003; Wells 1999). There was no statistically significant heterogeneity observed in the 'Moderate' (Diniz Zanetti 2007) or 'Low' contrast subgroups (de Oliveira 2009; Delgado 2009; Hay-Smith 2002; Sriboonreung 2011a; Sriboonreung 2011b; Sugaya 2003; Wells 1999).

Frequency of leakage - leakage episodes in 24 hours

Fewer trials contributed data to this comparison. In the 'High' contrast subgroup a single trial (Konstantinidou 2007) found the intensive intervention group had, on average, four fewer leakage episodes every three days compared to the less intensive therapy group. This was a statistically significant difference (MD -1.38, 95% CI -2.04 to -0.72, Analysis 12.3.1). In the 'Low' contrast subgroup, there was no important difference in the number of leakage episodes between more and less intensive interventions (MD -0.03, 95% CI -0.19 to 0.14, Analysis 12.3.3, Borello-France 2006; de Oliveira 2009; Hay-Smith 2002; Johnson 2001; Sugaya 2003; Wells 1999). Only Konstantinidou 2007 had a difference in the number of health professional contacts between the groups. In the other trials the amount of contact was the same (Borello-France 2006; Hay-Smith 2002; Johnson 2001; Sugaya 2003; Wells 1999) or the same frequency of contact provided either in a group or individually (de Oliveira 2009).

DISCUSSION

This review is the third in a series of reviews of pelvic floor muscle training (PFMT) for urinary incontinence in women, and it should be viewed in that context. This review considers whether one approach to PFMT is better than another. Prior reviews considered whether PFMT is better than no treatment, placebo, sham or non-active control treatments (Dumoulin 2010), and whether the addition of feedback or biofeedback adds benefit to PFMT (Herderschee 2011). Future reviews will consider whether PFMT is better than other treatments and if PFMT adds benefit to other treatments.

The primary objective of this review was to consider whether one approach to PFMT is better than another. Overall, there were few data, spread over 11 primary Comparisons. The 12th Comparison, that pooled the data from the 11 primary comparisons, was a post hoc subgroup analysis. Therefore, our conclusions are tentative rather than strong. Further, the design of some trials meant that more than one treatment variable differed between the comparison groups; these trials are interpreted cautiously because we are not sure which of the variables has contributed to any of the observed effects.

Summary of main results

Supervision of PFMT

In Comparison 1 (More versus less contact with health professionals, six trials, three subgroups) the trial arms with more contact were consistently more likely to report cure and improvement. Some caution is needed with interpretation of these outcomes because they were self reported, the women were not blind to intervention group, and the group that received more attention reported greater improvement. It was difficult to tell if the group with more contact had better incontinence-specific quality of life. With regard to the secondary outcomes, the pattern was not consistent although the findings were either of no difference between the groups or were in favour of the group with more health professional contact. There were no findings statistically significantly in favour of the group with less health professional contact, although it was interesting to note that just over half the women in the less intensive therapy group reported subjective improvement.

It is possible that more attention has an effect on PFMT and on important outcomes of treatment. For example, women may be prompted to exercise more often and put more effort into their exercises, and a class environment could provide social benefits that are reflected in how women feel about incontinence and PFMT. However, there is also the possibility that the apparent benefits of greater attention are 'experimenter effect' (Field 2003). Potential for experimenter effect is perhaps greatest in those trials where one group receives substantively more health professional contact than the other. Further, none of the trials could feasibly blind women or treatment providers to treatment assignment so those getting more attention knew it. The two outcomes for which there were most data (self reported cure and improvement) could therefore be influenced by attention. For example, women receiving more attention may over-estimate their improvement to please the treatment provider. Thus, 'experimenter effect' is a plausible explanation of the consistent finding of significant difference in the likelihood of self reported cure and improvement. Probably the only way to tease out the effect of attention is for future trials of this type to include an attention control arm in which the content of the attention is not likely to affect the primary outcome of interest, that is incontinence (see Dumoulin 2004 as an example).

Imamura 2010b, in a robust systematic review of all non-surgical treatments for stress urinary incontinence in women, found that more intensive PFMT intervention (either extra sessions with a health professional or the addition of biofeedback) was the most effective non-surgical treatment. Thus, the findings of Imamura 2010b and the present review appear congruent regarding more health professional contact. The more detailed analysis of the trials in the present systematic review has highlighted the difficulty with interpreting this finding, including the plausibility of experimenter effects and potential confounding by differences in PFMT over and above differences in contact.

The data in Comparison 2 (Group versus individual supervision of PFMT, six trials, two subgroups) were more difficult to interpret because all of the six trials in this comparison had more than one treatment variable that differed between the trial arms. The first subgroup in this comparison (individual supervision versus individual and group supervision, with no difference in the PFMT programmes between groups) included three of the trials from Comparison 1. Thus, the findings in this subgroup mirrored those of Comparison 1, and found the that women receiving group supervision were more likely to report improvement. Women receiving the additional group supervision may have benefited from group supervision or contact with other members of the group, the additional health professional contact, or both. We opined that the biggest difference in these three trials was in the amount of health professional contact, rather than in the approach to supervision so we put more weight on the findings of Comparison 1 in our review conclusions. The second subgroup in Comparison 2, was individual supervision only versus group supervision only but all three trials in this subgroup also had differences in the PFMT programme between the groups and so were potentially confounded by another variable. Generally the pattern in the data in this subgroup was one of no difference between the trial arms. Overall, it appeared that more health professional contact was better than less although there was insufficient evidence to be sure that the self reported improvement was echoed in incontinence-specific quality of life or more 'objective' incontinence measures. Based on the limited evidence available it was not clear if the way health professional contact was provided (individually or in a group setting) mattered or not.

Content of PFMT programmes

In Comparison 3 (Direct versus indirect methods PFMT, six trials, four subgroups), subgroup 1 comprised the two trials comparing PFMT and sham PFMT in which the amount of health professional contact was the same in both groups. While the pooled data from the two studies for self reported improvement did not show a statistically significant difference, in the larger trial there was a statistically significant difference in favour of the 'direct' training (Ghoniem 2005); the smaller trial did not find a difference but reported that training adherence was very poor in both groups (Ramsay 1990) and so participants may have done insufficient exercise for this to be a valid comparison. The indirect treatment in both trials comprised forceful abduction of the hips with the legs crossed. Three studies (Bø 1994; Dumoulin 2006; Morin 2004) demonstrated synergistic contraction of hip and pelvic floor muscles (PFM). Physiologically the synergistic contraction in the PFM is insufficient for a clinically important training effect, and a greater training effect in the 'direct' PFM training group in Ghoniem 2005 is congruent with these muscle studies.

Subgroups 2 and 3 comprised the trials comparing PFMT with the 'Paula method' (two trials) and the 'Sapsford' approach (one trial). Outcomes tended to favour the indirect PFMT, but in each of these three trials the women in the indirect training group had more attention than the direct training group. In the absence of control for amount of attention, we considered that the risk of confounding in these trials was high. Further muscle studies have raised doubts about these 'indirect' methods of PFM contraction. Recent 4D ultrasound and EMG study of the 'Paula method' found contraction of one or more 'circular' muscles (such as eye or mouth closure) did not co-activate the PFM muscles or increase PFM activity (Bø 2011; Resende 2011). Further a direct contraction of the PFM produces more 'lift' in the PFM than a transversus abdominis muscle contraction alone or PFM contraction facilitated by a transversus abdominis contraction (Bø 2003); the facilitated approach was that used by Hung 2010. Therefore, the evidence from the physiologic studies of either a lack of PFM co-activation or less effective PFM contraction with these 'indirect' methods adds support to our view that any apparent benefit of the 'indirect' PFM training in these three trials is likely confounded by health professional contact. The remaining trial (PFMT versus

Pilates) in Comparison 3 contributed little because it was a very small pilot study, and every outcome had wide confidence intervals that included no difference.

A single trial contributed data to Comparison 4 (Individualised versus generic PFMT), and this trial also compared group versus individual supervision (i.e. generic training in a group versus individualised training with individual supervision). While there was some difference between the PFMT programmes, both were predominantly strength training programmes, with some elements of co-ordination training; there was probably little difference in the likely physiologic effects of the PFMT programmes being compared. The bigger difference was possibly in the type of health professional contact. While the women with the individualised supervision and training programme appeared to have greater PFM strength post-intervention this was not reflected in differences in self reported improvement or incontinence-specific quality of life.

Comparison 5 (Near maximal versus submaximal contraction) comprised a single trial; no data were collected for any of the primary outcomes of interest in the review and the pattern in the secondary outcomes of interest was one of no difference between the groups. Strength training requires short duration high load (near maximal intensity) contractions, and endurance training light loads (submaximal effort) with high repetitions (ACSM 2009). There was no difference in the training duration (15 minutes, three times per day) in the exercise protocol for the comparison groups. The difference was in contraction intensity (90% versus 60% of maximal voluntary effort). It is possible that these exercise parameters were not sufficiently different to show any difference in training effect between the groups.

in Comparison 10 (PFMT with intravaginal resistance device versus PFMT alone) the consistent pattern in the data from three trials was one of no difference between the resistance and no resistance training groups.The additional effect of strength training, over and above motor relearning (co-ordination type training) in a single trial (Comparison 8, Strength and motor learning versus motor learning PFMT alone) found no differences between training groups. Given near-maximal contractions are need to improve muscle strength, and strength training is facilitated by the addition of resistance (ACSM 2009) the lack of difference between groups in the trials in Comparisons 10 and 8 is potentially counter-intuitive. Other factors (such as training duration, training adherence and so on) may have reduced the training effect. Another possibility is that because all these trials compared two active treatment arms, the difference between the groups was moderate at best and the trials were not sufficiently powered to establish statistically significant differences between the groups.

A single trial contributed data to Comparison 7 (Upright and supine versus supine exercise positions alone) and did not find any differences between the two trials arms that did or did not include upright exercise positions such as sitting and standing.

Finally, a single trial investigated the effect of adding abdominal muscle exercise to PFMT (Comparison 9, PFMT and abdominal muscle exercise versus PFMT alone) and apart from change in vaginal squeeze pressure (which favoured the group doing the additional abdominal muscle exercise) there were no differences between the groups. Vaginal squeeze pressure is influenced by abdominal pressure as well as PFM contraction, so we did not interpret this single outcome as evidence of effect for the group doing additional abdominal muscle exercise.

In summary, from the limited data available direct PFMT was better than a sham method of PFMT (abducting the hips with ankles crossed). It was likely that the trials of some 'indirect' methods of training (such as the 'Paula' or 'Sapsford' approaches) were confounded by the greater amount of attention in the 'indirect' training groups. With few data, and only single trials, it was difficult to reach any conclusion based on any other comparison investigating the content of PFMT programmes.

PFMT frequency

Comparison 6 (Daily versus three times per week PFMT) comprised a single trial that did not find any difference for any measured outcome. Both groups were doing a strength training type of exercise programme. Based on current evidence training three times a week would be sufficient for a strength training and muscle hypertrophy if other exercise parameters (such as effort and repetitions) are optimal (ACSM 2009; ACSM 2011). Thus, it is not surprising that no difference was found in the outcomes of the two exercise frequencies investigated in this trial.

Adherence strategies

Two trials were included in Comparison 11 (PFMT and adherence strategy versus PFMT alone). One of them reported adherence data but no data for incontinence outcomes, so while the trial arm who received an audiotape to support exercise at home did more exercise, there is no way of telling if these women benefited in terms of their continence. Based on the data from the other trial it appeared that women using an electronic device to prompt exercise were more likely to be improved after treatment. However, the secondary outcomes were generally not different between the groups. Further, because exercise adherence was measured only in the trial arm using the device it was not clear if this group did more exercise than the group without the device.

A few trials in other comparisons also measured adherence either based on clinic attendance or training diaries. Adherence was then reported in several ways such as the proportion of prescribed exercise completed, whether women were exercising at home or not, and the proportion or average number of clinic visits attended. Some trials collected adherence data and did not report it, and some collected adherence data in only one trial arm and so there were no data from the other arm for comparison. Overall it was difficult to interpret the few data reported; it seemed adherence rates varied widely between trials and it was not clear why this was so.

PFMT will not work unless the exercise is done. Treatment adherence is likely to have an impact on the size and direction of treatment effect because adherence affects the exercise 'dose'. Maintaining the effect also requires adherence, and medium to long-term exercise adherence is known to be problematic (Sluijs 1998). As PFMT appears to be an effective treatment for urinary incontinence (Dumoulin 2010), the question of how best to facilitate adherence is an important issue that deserves attention. In trials where two active treatments are being compared (such as the trials included in this review) the reasons to measure adherence are to: (a) investigate the effect on adherence from the addition of an adherence strategy in one trial arm; and (b) look at differences in adherence between two therapies as a plausible explanation for differences in the effect of those therapies.

Adherence is certainly difficult to measure (for example, see Stone 2003), and some trialists commented on this (e.g. poor completion of exercise diaries). There is neither consensus on the best ways to measure adherence nor what should be measured. Agreement on how best to collect and report these data, and then a commitment from trialists to doing this, could be valuable additional information for analysing the effects of PFMT.

Socioeconomics, adverse events and follow-up data

Many studies mentioned the socioeconomic impact of incontinence in the introduction to their trials, yet only one trial reported any economic data. Hay-Smith 2002 calculated the costs of treatment and because the same amount of supervision was given in both trial arms the costs of the two interventions was essentially the same. None of the studies reported an economic analysis, or cost-effectiveness analysis. Until PFMT trials routinely report the costs associated with the interventions it will be difficult to do anything other than develop theoretical 'models' of cost-effectiveness or cost-utility. Although it is acknowledged that different currencies and health systems make it difficult to generalise cost information worldwide, reporting costs is the first step toward the cost-effectiveness analyses that are fundamental to discussions about funding health services.

Few trials explicitly reported whether adverse events data were collected. Where such data were collected, no adverse events were reported. Although adverse events are likely to be minor (or uncommon), the only way to be sure is to have a record of adverse events or their absence from as many people as possible.

Two trials had collected longer-term data. It was impossible to interpret these data because one trial combined data from both trial arms, and the other collected data from only one arm of the trial. The data suggested that some women continued to exercise, some maintained the improvement in their symptoms, whereas other women had reduced or stopped exercising, and others had pursued surgical options.

Comment on 'all in one' analysis

We recognise that the heterogeneity of interventions in the 'all in one analysis' was considerable; a more reasonable interpretation might be to consider each of the comparisons separately as we have done above. This analysis was exploratory in nature and its findings need to be interpreted with caution. It might be best to consider them as the basis for further research.

Based on our 'all in one analysis' to compare more versus less intensive interventions (Comparison 12) the very consistent pattern was that the higher the contrast in intervention intensity the more likely it was that there was a statistically significant difference in favour of the more intensive therapy group. The summary statistics from the 'High' contrast subgroups were statistically significantly in favour of the more 'intensive' interventions for all three outcomes. The summary statistics for the 'Low' contrast subgroups were in favour of the more intensive interventions for self reported improvement, but were consistent with no benefit for self reported cure (although this was very close to statistical significance in favour of the less intensive intervention) or for leakage episodes in 24 hours. The trials that were classified as 'High' contrast were those in which there were substantive differences in health professional contact or compared direct with indirect methods of PFMT, therefore it was not surprising this 'all in one analysis' mirrored the findings discussed above. That is, that high levels of supervision are better than low levels (Comparison 1) and direct PFMT is better than indirect PFMT (Comparison 3).

Finding more difference in outcome when there is greater contrast in intensity of intervention seems logical. Interpreting the meaning of this difference is less straightforward. For example, we have already raised the possibility that the lack of control for attention in trials in Comparison 1 means some of the effect could be an 'experimenter' effect in unblinded outcomes such as self reported cure and improvement. However, we feel the advantage of the 'all in one' analysis is that is allows the reader to see as much of the data as possible in three forest plots and the very consistent pattern of effect iterates the findings of the individual comparisons discussed above.

One potentially interesting finding is that in the 'Low' contrast subgroup, there was almost a statistically significant difference in favour of the less intensive intervention for self reported cure. It is possible that those women who concentrated on a more 'basic' PFMT programme benefited from putting their full efforts into this, and were less distracted by additional elements (such as using adjuncts like intravaginal resistance devices or rotating their exercises through multiple body positions). However, the finding for self reported improvement was just in favour of the more intensive intervention, so the inconsistency in the cure and improvement outcomes means we remain uncertain about the importance of the possible difference in the former.

The outcome that matters most to women with urinary incontinence is probably incontinence-specific quality of life (Herbison 2009). Unfortunately the few data for this outcome were not re-

ported in ways that could be used in meta-analysis. If data from psychometrically robust incontinence-specific quality of life measures was also consistent with the findings discussed above (more effect in more intensive contact arm), we would feel more confident that any differences were clinically important. We would also feel more confident if there were more supporting data for more 'objective' outcomes (such as leakage episodes). So far, the pattern appears similar from the few data on leakage episodes but more data are needed.

Finally, the 'all in one' analysis highlighted both that at least half of women receiving a 'less intensive' but active PFMT intervention (that is, the 'control' conditions) were likely to report improvement post-treatment. Interventions with more health professional contact increased the proportion of women reporting improvement, but there might be a trade-off between the resource implications of an intensively supervised programme and the opportunity cost this represents if at least half the women do improve with less intensive supervision. However, we do not have sufficient evidence from follow-up studies to know if there is a difference in medium to longer-term outcomes between more and less intensively supervised groups.

Overall completeness and applicability of evidence

Most of the trials in the review recruited only women with stress urinary incontinence, or women with stress or mixed urinary incontinence. We consider the review findings are most applicable to women with stress incontinence as their only or predominant symptom, and suggest the review findings can not be assumed to apply to women with urgency urinary incontinence as their only or predominant symptom. Trials are need to address questions about the effectiveness of different approaches to PFMT in women with urgency urinary incontinence or mixed symptoms.

In general there were few data in this review spread over many outcomes and comparisons. The fact that many different outcomes were measured, along with the numerous instruments used for every outcome, meant it was hard to gather sufficient data for pooled analysis. This was further complicated by the variety of ways in which trialists reported data (for example, post-treatment data, change from baseline, dichotomising continuous variables and so on). Consequently it was difficult to come to any robust conclusions about the best approaches to PFMT. We have, below, made some comments about improving the design of future PFMT trials so that in time the data in these reviews will be more complete and interpretable.

The use of validated quality of life instruments was more common in recent PFMT trials, although it was rarely chosen as the primary outcome. About half the trials asked women for their opinion about whether their symptoms had improved or not. Both of these outcomes are easily measured using valid and reliable instruments, and all future PFMT trials could do this. The inclusion of these as 'standard' outcomes in future trials would add considerably to the evidence base for PFMT.

Fewer than half the trials reported the number of leakage episodes in a specified time (such as 24 hours, three days or seven days). Urinary or accident diaries are commonly used in clinical and research practice to quantify leakage, and there is more agreement about how to measure this than about pad tests (see Dumoulin 2010 for a full discussion of pad tests). The inclusion of leakage episode outcomes in all PFMT trials would also add considerably to the evidence base. These more 'objective' data would be an interesting supplement to the more 'subjective' quality of life and symptom improvement data.

In general, more recent trials described the PFMT interventions more fully. There is still considerable scope for improvement. Often it was not clearly stated if the exercise programme was the same in both groups or not. Sometimes the trialists concentrated on explaining what was different, or an additional component in one group, but did not clearly describe the basic content of the PFMT programme. It will continue to be very difficult to apply in practice the findings of individual trials, or systematic reviews of trials, unless the training programmes are reported in detail. Further, it is difficult to determine if there is a reasonable physiologic basis for a training effect if the exercise parameters are not adequately described. At minimum it would be helpful to report:

- whether a correct contraction is confirmed before training (and how);
- the purpose of training (e.g. strength, endurance, behavioural);
- the number of contractions per set and number of sets per day;
- duration of hold of each contraction and duration of rest between;
- the effort with each contraction (maximal, submaximal); and
- the number of weeks of training.

Where the PFMT programmes being compared are expected to act differently (physiologically) on the PFM, it is useful to measure PFM performance to see if the hypothesised differences in strength or endurance or co-ordination are evident. Many trials in this review did measure PFM performance. A wide range of measures were used. Consequently it was difficult to combine the data in any meaningful way; we could only look for any patterns in these data (e.g. tend to be in favour of one group or another, or no difference). It is probably difficult to get any agreement about how best to measure PFM performance; some measures are chosen because pragmatically they cost less, are easier to access, and are already widely used in clinical practice (such as the digital vaginal palpation scores) while some measurement instruments have more robust measurement properties (such as an instrumented speculum). Trialists are encouraged to consider the psychometric properties of PFM performance measures when they choose an instrument for research.

Quality of the evidence

Risk of bias was evaluated from the trial reports so the quality of reporting affected our assessment. Where detail of study methods was incompletely reported we tried to contact the trialists for more information. One of the trials was published only as an abstract (Ramsay 1990), that had few details of study design, methods or data, which made it particularly difficult to assess. One study was published in a PhD thesis (Hay-Smith 2002) with rather more detail of study design and methods than is usual in journal publication.

Twenty-one trials were included in the review and only six of these were judged to be at low risk of selection bias. Given the difficulty of blinding participants and treatment providers to physical therapies, blinding of outcome assessment for as many outcomes as possible is important. Only three trials explicitly stated that blinded assessment of one or more of the outcomes of interest in the review was done, and two trials stated that a lack of blinded outcome assessment was a limitation of their trial. The greatest amount of data in any single analysis in this review was for self reported improvement (see Analysis 12.2). This outcome could not be assessed blind because women knew which treatment group they were in. Consequently, we were cautious about interpreting these data as clear evidence of benefit unless corroborated by at least one other outcome.

With regard to attrition bias, the proportion of losses to follow-up was high in some trials although the differences in the proportion within treatment groups was mostly small. Size-wise, quite a few of the trials were small to moderate so a high proportion of loss to follow-up means they easily become under-powered. Generally, more robust methods of dealing with data for analysis in trials with moderate to large proportions of dropouts are needed.

Overall, the biggest problem with the included trials was incomplete reporting, either of methods (where we often assessed the risk of selection or detection bias as 'unclear') or data. With regard to reporting of methods, trialists, journals and peer reviewers would benefit from closer attention to the CONSORT guidelines for reporting of randomised trials (Boutron 2008; CONSORT 2010; Moher 2001). Data reporting also needs attention. The usefulness of some trials in the review was substantively reduced because of incomplete data reporting. Common problems were:

1. data reported for only one treatment group;
2. reporting of the measure of central tendency such as the mean without a measure of dispersion such as the standard deviation;
3. a lack of raw data in cases where only P values or figures/graphs were reported.

Publishing peer reviewed papers imposes word limits and this makes it difficult to report all data; full reports can be made available via the internet. None of the included trials made their full completed analysis available in this way.

Potential biases in the review process

We were not blinded to the authorship of the papers being screened for eligibility, during risk of bias assessment or data extraction. Between the review authors we already knew many of the included studies and it was not possible to blind them effectively. Two of us (JHS, CD) were the first authors of trials considered for inclusion in the review. We were not involved with the screening, decision about eligibility, data extraction or data entry of these individual trials.

AUTHORS' CONCLUSIONS

Implications for practice

Based on the limited data available it seemed that pelvic floor muscle training (PFMT) with regular (e.g. weekly) supervision was better than PFMT with little or no supervision. It was not clear if the supervision was best provided individually or in a group. Cautiously, because of the few data, we also concluded that PFMT based on direct voluntary contractions of the pelvic floor muscles was better than sham exercises (i.e. crossing the ankles and pulling the legs apart). These two groups of trials (high versus low levels of supervision, direct PFMT versus sham PFMT) made 'high' contrast comparisons of intervention intensity and pooled data from these trials consistently found added benefit from the more intensive interventions.

Most of the trials that compared PFMT with other 'indirect' methods of co-activating the pelvic floor muscles were very likely confounded by differences in the amount of contact time with health professionals. Further, evidence from muscle physiology studies suggests 'indirect' methods are inefficient or ineffective ways to train muscle.

Tentatively, there was no benefit from the addition of intravaginal devices to resist PFMT. There were too few data to drew conclusions for the other comparisons of different approaches to performing PFMT or the addition of adjuncts to training. When considered together, these trials had 'moderate' to 'low' contrast in intervention intensity and their findings were consistent with no benefit for the more intensive intervention.

Substantiating the most effective PFMT programme was identified as a high priority by Buckley 2009 and colleagues, in a process that involved patient groups and clinicians. This review found that the existing evidence is insufficient to make any robust recommendations about the best approach to PFMT, other than women were more likely to report they were improved if they received more attention from a health professional.

Implications for research

Comparisons of approaches to PFMT are comparisons of two ac-

tive treatments and it seems differences between treatments might only be observed if there is a high contrast in intervention intensity. Therefore, it might be difficult to find out which approaches to PFMT are best unless: (a) the differences in outcome are large, (b) the trials are powered to find small to moderate differences in outcome which would probably mean large or very large trials are needed, or (c) the trial is powered to establish equivalence, which again would probably need large trials. Although finding the best approach to PFMT was identified as high priority in recent research involving clinicians and patients, large costly trials may not be the best use of research funds when the difference in outcome between two active treatments is expected to be small, unless there are likely to be significant economic benefits (such as much lower costs for one treatment). Further, because there are so many potential differences in PFMT programmes it would take many trials to investigate each of these using direct comparisons. More attention is needed in choosing (and then adequately describing) interventions that can have a muscle training effect. Approaches other than randomised trials may also need to be considered (see for example, Whiteneck 2009).

With this in mind, along with the findings of the review, we consider that the highest priority for a new randomised trial is to investigate the effect of training supervision in terms of both approach (individual versus group) and amount of contact. In addition to clinical effectiveness, this is an important question because of the resource implications for health service delivery. An important feature of such a proposed trial is the delivery of the same PFMT programme in every trial arm, and this should be based on the best available evidence about training muscle (see for example, ACSM 2009; ACSM 2011). A 2 x 2 trial design with high/low contact and individual/group supervision is suggested. Thus the trial arms would comprise high contact hours with individual supervision versus high contact hours with group supervision versus low contact hours with individual supervision versus low contact hours with group supervision. The choice of the number of hours (for the 'high contact hours' groups) would have to be set against the resource implications of an intensively supervised programme and the opportunity cost this represents. Careful consideration is needed about what number of hours is realistic in everyday practice and in different countries with different health systems, and what could be implemented if the more 'intensive' supervision was shown to be effective.

ACKNOWLEDGEMENTS

The authors of the review would like to thank:

- R Herbison and R George who helped with screening for studies for eligibility and with data extraction;
- M Imamura who helped categorise trial comparisons as 'high', 'moderate' or 'low';
- J Cook for statistical advice on the 'all in one' analysis (Comparison 12);
- K Bø, B Berghmans, R de Bie, E Hendriks and E van Waalwijk van Doorn for their valuable contribution to the previous review, on which the current review was based;
- S Wallace for comprehensive and repeated literature searching; and
- the following trialists who graciously responded to letters or emails: K Bø,T Coombes on behalf of Orelle Corporation Ltd, E Crothers, D Delgado, M Drake, B Fung, M Gallo, S von der Heide, E Hui, C Klingler, M Liebergall-Wischnitzer, I Nygaard, AM Rodrigues and T Wells.

REFERENCES

References to studies included in this review

Borello-France 2006 *{published data only}*

Borello-France DF, Downey PA, Zyczynski HM, Rause CR. Continence and quality-of-life outcomes 6 months following an intensive pelvic-floor muscle exercise program for female stress urinary incontinence: a randomized trial comparing low- and high-frequency maintenance exercise. *Physical Therapy* 2008;**88**(12):1545–53.

* Borello-France DF, Zyczynski HM, Downey PA, Rause CR, Wister JA. Effect of pelvic-floor muscle exercise position on continence and quality-of-life outcomes in women with stress urinary incontinence. *Physical Therapy* 2006;**86**(7):974–86.

Bø 1990 *{published and unpublished data}*

Bø K. Adherence to pelvic floor muscle exercise and long-term effect on stress urinary incontinence. A five-year follow-up study. *Scandinavian Journal of Medicine & Science in Sports* 1995;**5**(1):36–9. [MEDLINE: 3056]

Bø K. Pelvic floor muscle exercise for the treatment of female stress urinary incontinence. Methodological studies and clinical results. *Acta Obstetricia et Gynecologica Scandinavica* 1991;**70**:637–9. [MEDLINE: 2626]

Bø K. Pelvic floor muscle strength and response to pelvic floor muscle training for stress urinary incontinence. *Neurourology and Urodynamics* 2003;**22**(7):654–8. [MEDLINE: 16431]

Bø K, Hagen R, Jorgensen J, Kvarstein B, Larsen S.

The effect of two different pelvic floor muscle exercise programs in treatment of urinary stress incontinence in women. *Neurourology and Urodynamics* 1989;**8**(4):355–6. [MEDLINE: 4514]

Bø K, Hagen R, Kvarstein B, Larsen S. Female stress urinary incontinence and participation in different sports and social activities. *Scandinavian Journal of Sports Science* 1989;**11** (3):117–21. [MEDLINE: 6344]

* Bø K, Hagen RH, Kvarstein B, Jorgensen J, Larsen S. Pelvic floor muscle exercise for the treatment of female stress urinary incontinence: III. Effects of two different degrees of pelvic floor muscle exercises. *Neurourology and Urodynamics* 1990;**9**(5):489–502. [MEDLINE: 2625]

Bø K, Kvarstein B. 15 year follow-up of a randomized controlled trial of pelvic floor muscle training to treat female urodynamic stress incontinence (Abstract). Proceedings of the International Continence Society (34th Annual Meeting) and the International UroGynecological Association, 2004 Aug 23-27, Paris. 2004:Abstract number 658. [MEDLINE: 19081]

Bø K, Kvarstein B, Hagen RH. The effect of two different pelvic floor muscle exercise regimens in treatment of female stress urinary incontinence (Abstract). Proceedings of the American Urogynecology Society, 12th annual meeting, 1191, Oct 23-26, Newport Beach, California, USA. 1991. [MEDLINE: 14574]

Bø K, Kvarstein B, Nygaard I. Lower urinary tract symptoms and pelvic floor muscle exercise adherence after 15 years. *Obstetrics and Gynecology* 2005;**105**(5):999–1005. [MEDLINE: 20358]

Bø K, Larsen S. Pelvic floor muscle exercise for the treatment of female stress urinary incontinence - classification and characterization of responders. *Neurourology and Urodynamics* 1992;**11**(5):497–507. [MEDLINE: 5148]

Bø K, Larsen S, Kvarstein B, Hagen RH. Classification and characterization of responders to pelvic floor muscle exercise for female urinary incontinence. *Neurourology and Urodynamics* 1990;**9**(4):395–7. [MEDLINE: 5113]

Bø K, Talseth T. Long-term effect of pelvic floor muscle exercise 5 years after cessation of organized training. *Obstetrics and Gynecology* 1996;**87**(2):261–5. [MEDLINE: 9901]

de Oliveira 2009 *{published data only}*

* de Oliveira Camargo F, Rodrigues AM, Arruda RM, Ferreira Sartori MG, Girao MJ, Castro RA. Pelvic floor muscle training in female stress urinary incontinence: comparison between group training and individual treatment using PERFECT assessment scheme. *International Urogynecology Journal & Pelvic Floor Dysfunction* 2009;**20**(12, 1433-3023 (Electronic)): 1455–62.

Delgado 2009 *{published data only}*

* Delgado D, Drake M. A randomised study to compare the PelvicToner device against standard pelvic floor exercises in the treatment of stress urinary incontinence in women (Abstract number 486). Proceedings of the 39th Annual Meeting of the International Continence Society (ICS), 2009 Sep 29 - Oct 3, San Francisco, CA 2009.

Diniz Zanetti 2007 *{published data only}*

* Dinez Zanetti MR, de Aquino Castro R, Lyvio Rotta A, Diniz dos Santos P, Sartori M, Castello Giraã o MJB. Impact of supervised physiotherapeutic pelvic floor exercises for treating female stress incontinence. *Sao Paulo Medical Journal* 2007;**125**(5):265–9.

Felicissimo 2010 *{published and unpublished data}*

* Felicissimo MF, Carneiro MM, Saleme CS, Pinto RZ, Da Fonseca AMRM, da Silva Filho AL. Intensive supervised versus unsupervised pelvic floor muscle training for the treatment of stress urinary incontinence: a randomized comparative trial. *International Urogynecology Journal* 2010; **21**:835–40.

Ferguson 1990 *{published data only}*

* Ferguson KL, McKey PL, Bishop KR, Kloen P, Verheul JB, Dougherty MC. Stress urinary incontinence: effect of pelvic muscle exercise. *Obstetrics and Gynecology* 1990;**75** (4):671–5. [MEDLINE: 2632]

Gallo 1997 *{published data only (unpublished sought but not used)}*

* Gallo ML, Staskin DR. Cues to action: pelvic floor muscle exercise compliance in women with stress urinary incontinence. *Neurourology and Urodynamics* 1997;**16**(3): 167–77. [MEDLINE: 4644]

Ghoniem 2005 *{published data only}*

* Ghoniem GM, Van Leeuwen JS, Elser DM, Freeman RM, Zhao YD, Yalcin I, et al.A randomized controlled trial of duloxetine alone, pelvic floor muscle training alone, combined treatment and no active treatment in women with stress urinary incontinence. *Journal of Urology* 2005; **173**:1647–53.

Schagen van Leeuwen JH, Elser D, Freeman R, Ghoniem G, Zhao Y, Yalcin I, et al.Controlled trial of duloxetine alone, pelvic floor muscle training alone, combined treatment, and no treatment in women with stress urinary incontinence (SUI) (Abstract). *European Urology Supplements* 2004;**3**(2): 52.

Hay-Smith 2002 *{published data only}*

* Hay-Smith EJC. Pelvic floor muscle training for female stress urinary incontinence (PhD thesis). University of Otago, Dunedin, New Zealand. Dunedin, New Zealand: University of Otago, 2003.

Hay-Smith EJC, Herbison GP, Wilson PD. Pelvic floor muscle training for women with symptoms of stress urinary incontinence: a randomised trial comparing strengthening and motor relearning approaches (Abstract). *Neurourology and Urodynamics* 2002;**21**(4):371–2. [MEDLINE: 14538]

Hay-Smith EJC, Ryan K, Dean S. The silent, private exercise: experiences of pelvic floor muscle training in a sample of women with stress urinary incontinence. *Physiotherapy* 2007;**93**(1):53–61.

Hung 2010 *{published data only}*

* Hung HC, Hsiao SM, Chih SY, Lin HH, Tsauo JY. An alternative intervention for urinary incontinence: retraining diaphragmatic, deep abdominal and pelvic floor muscle coordinated function. *Manual Therapy* 2010;**15**(3):273–9.

Johnson 2001 *{published data only}*

Johnson VY. *Effects of a submaximal exercise protocol to recondition the circumvaginal musculature in women with genuine stress urinary incontinence (PhD thesis)*. San Antonio, Texas, USA: The University of Texas Health Science Center at San Antonio, 1997.

* Johnson VY. Effects of a submaximal exercise protocol to recondition the pelvic floor musculature. *Nursing Research* 2001;**50**(1):33–41.

Konstantinidou 2007 *{published data only}*

Konstantinidou E, Apostolidis A, Kondelidis N, Tsimtsiou Z, Hatzichristou D, Ioannides E. Short-term efficacy of high-supervisory-intensity group pelvic floor training versus unsupervised, home training in female stress urinary incontinence: a randomised pilot study (Abstract number 678). *European Urology Supplements* 2006;**5**(2):192.

* Konstantinidou E, Apostolidis A, Kondelisis N, Tsimtsiou Z, Hatzchristou D, Ioannides E. Short-term efficacy of group pelvic floor training under intensive supervision versus unsupervised home training for female stress urinary incontinence: a randomized pilot study. *Neurology and Urodynamics* 2007;**26**:486–91.

Liebergall 2005 *{published data only}*

Liebergall M, Hochner-Celnikier D, Manor O, Lavy Y, Paltiel O. Comparison of the effectiveness of circular muscle exercise with pelvic floor training for urinary stress incontinence in women; a randomized controlled trial (Abstract number 264). *International Urogynecology Journal* 2006;**17**(Suppl 2):S206–7.

* Liebergall-Wischnitzer M, Hochner-Celnikier D, Lavy Y, Manor O, Arbel R, Paltiel O. Paula method of circular muscle exercises for urinary stress incontinence--a clinical trial. *International Urogynecology Journal* 2005;**16**(5): 345–51. [MEDLINE: 21306]

Liebergall 2009 *{published data only}*

* Liebergall-Wischnitzer M, Hochner-Celnikier D, Lavy Y, Manor O, Shveiky D, Paltiel O. Randomized trial of circular muscle versus pelvic floor training for stress urinary incontinence in women. *Journal of Women's Health* 2009;**18** (3):377–85.

Ng 2008 *{published data only}*

Lin TL, Chen YC, Hu SW, Chen GD. Nursing intervention to enforce the efficacy of home practice of pelvic floor muscle exercise in mixed incontinence (Abstract 294). Proceedings of the International Continence Society (34th Annual Meeting) and the International UroGynecological Association, 2004 Aug 23-27, Paris. 2004. [MEDLINE: 19053]

* Ng SC, Lin TL, Chang SJ, Tai HL, Hu SW, Chen GD. Nursing intervention to enhance efficacy of home practice of pelvic floor muscle exercises in treating mixed urinary incontinence. *International Urogynecology Journal* 2008;**19** (5):637–42.

Ramsay 1990 *{published data only}*

* Ramsay IN, Thow NB. A randomised, double blind, placebo controlled trial of pelvic floor exercises in the treatment of genuine stress incontinence. *Neurourology and Urodynamics* 1990;**9**(4):398–9.

Savage 2005 *{published data only}*

Savage AM. Is lumbopelvic stability training (using the Pilates model) an effective treatment strategy for women with stress urinary incontinence? A review of the literature and report of a pilot study. *Journal of the Association of Chartered Physiotherapists in Women's Health* 2005;**97**: 33–48.

Sriboonreung 2011a *{published data only}*

* Sriboonreung T, Wongtra-ngan S, Eungpinichpong W, Laopaiboon M. Effectiveness of pelvic floor muscle training in incontinent women at Maharaj Nakorn Chiang Mai Hospital: a randomized controlled trial. *Journal of the Medical Association of Thailand* 2011;**94**(1):1–7.

Sriboonreung 2011b *{published data only}*

* Sriboonreung T, Wongtra-ngan S, Eungpinichpong W, Laopaiboon M. Effectiveness of pelvic floor muscle training in incontinent women at Maharaj Nakorn Chiang Mai Hospital: a randomized controlled trial. *Journal of the Medical Association of Thailand* 2011;**94**(1):1–7.

Sugaya 2003 *{published data only}*

* Sugaya K, Owan T, Hatano T, Nishijima S, Miyazato M, Mukouyama H, et al.Device to promote pelvic floor muscle training for stress incontinence. *International Journal of Urology* 2003;**10**(8):416–22. [MEDLINE: 17358]

Wells 1999 *{published data only}*

* Wells TJ, Major RD, Brink CA, Brown R. Pelvic muscle exercise: a controlled clinical trial. Unpublished manuscript 1999.

References to studies excluded from this review

Crothers 2001 *{published data only}*

Crothers E. A randomised controlled clinical trial of a cueing device to enhance patient compliance to a programme of pelvic floor exercises. The metaRegister of Controlled Trials (mRCT) 2002.

de Jong 2006 *{published data only}*

de Jong JH, Van Kampen MB, Biemans B. The effect of whole body vibration training on women with stress urinary incontinence (Abstract number 416). Proceedings of the International Continence Society (ICS), 36th Annual Meeting, 2006 Nov 27-Dec 1, Christchurch, New Zealand. 2006.

Dumoulin 2003 *{published data only}*

Dumoulin C, Lemieux M, Bourbonnais D, Morin M. Conservative management of stress urinary incontinence: a single-blind, randomized controlled trial of pelvic floor rehabilitation with or without abdominal muscle rehabilitation compared to the absence of treatment (Abstract). *Neurourology and Urodynamics* 2003;**22**(5): 543–4. [MEDLINE: 17101]

Dumoulin C, Morin M, Bourbonnais D, Lemieux M, Gravel D. Effect of adding deep abdominal muscle training to pelvic floor muscle training to treat stress urinary incontinence: a one-year follow up (Abstract).

Proceedings of the International Continence Society (34th Annual Meeting) and the International UroGynecological Association, 23-27 August, 2004, Paris. 2004:Abstract number 662. [MEDLINE: 19083]

Hill 2007 *{published data only}*
Hill LA, Fereday-Smith J, Credgington C, Woodward AF, Knight JC, Williams AJ, et al.Bladders behaving badly: a randomized controlled trial of group versus individual interventions in the management of female urinary incontinence (Abstract). *Journal of the Association of Chartered Physiotherapists in Women's Health* 2007;**101**: 30–6.

Hui 2006 *{published data only}*
Hui E, Lee PSC, Woo J. Management of urinary incontinence in older women using videoconferencing versus conventional management: a randomized controlled trial. *Journal of Telemedicine or Telecare* 2006;**12**(7):343–7.

Klinger 1995 *{published data only}*
Klingler HC, Madersbacher S, Uher EM, Schmidbauer CP. Pelvic floor exercise and endotrainer for treatment of female stress urinary incontinence. Proceedings of the 25th Annual Meeting of the International Continence Society, 1995, Sydney, Australia. 1995.

Nygaard 1996 *{published data only}*
Nygaard IE, Kreder KJ, Lepic MM, Fountain KA, Rhomberg AT. Efficacy of pelvic floor muscle exercises in women with stress, urge, and mixed urinary incontinence. *American Journal of Obstetrics & Gynecology* 1996;**174**(1 (Pt 1)):120–5. [MEDLINE: 2854]

Orelle NatraTone 2008 *{published data only}*
Collier J, McNair P, Williams M. The Orelle Trial. Auckland University of Technology report prepared for Armitage C, Coombes T, Schneiderman R and Orelle Corporation 2008.

Taylor 1986 *{published data only}*
Taylor K, Henderson J. Effects of biofeedback and urinary stress incontinence in older women. *Journal of Gerontological Nursing* 1986;**12**(9):25–30.

Wong 1997 *{published data only}*
Wong KS, Fung BKY, Fung LCW, Ma S. Pelvic floor exercises in the treatment of stress urinary incontinence in Hong Kong Chinese women. Proceedings of the International Continence Society (ICS), 27th Annual Meeting, 23-26 Sept, Yokohama, Japan. 1997.

Yoon 1999 *{published data only}*
Yoon HN, Hong JY, Choi YH, Back SH. The effect of pelvic floor muscle exercises on genuine stress incontinence among Korean women: focusing on its effects on the quality of life. Proceedings of the International Continence Society (ICS), 29th Annual Meeting, 22-26 Aug, Denver, Colorado YR:1999 296. [MEDLINE: 9922]

References to ongoing studies

Kincade 2005 *{published data only}*
Kincade JE, Dougherty MC, Busby-Whitehead J, Carlson JR, Nix WB, Kelsey DT, et al.Self-monitoring and pelvic floor muscle exercises to treat urinary incontinence. *Urologic Nursing* 2005;**25**(5):353–63. [MEDLINE: 21262]
Kincade JE, Dougherty MC, Carlson JR, Hunter GS, Busby-Whitehead J. Randomised clinical trial of efficacy of self-monitoring techniques to treat urinary incontinence in women. *Neurology and Urodynamics* 2007;**26**:507–11.

von der Heide 2003 *{published data only}*
von der Heide S, Emons G, Hilgers R, Viereck V. Effect on muscles of mechanical vibrations produced by the Galileo 2000 in combination with physical therapy in treating female stress urinary incontinence (Abstract). Proceedings of the International Continence Society (ICS), 33rd Annual Meeting, 2003 Oct 5-9, Florence Italy 2003:192–3.

Additional references

ACSM 2009
American College of Sports Medicine. Position stand: progression models in resistance training for healthy adults. *Medicine and Science in Sports and Exercise* 2009;**41**(3):687-708.

ACSM 2011
American College of Sports Medicine. Position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Medicine and Science in Sports and Exercise* 2011;**43**(7):1334–59.

Bartoli 2010
Bartoli S, Aguzzi G, Tarricone R. Impact on quality of life of urinary incontinence and overactive bladder: a systematic literature review. *Urology* 2010;**75**(3):491–500.

Black 1996
Black N, Griffiths J, Pope C. Development of a symptom severity index and a symptom impact index for stress incontinence in women. *Neurourology & Urodynamics* 1996;**15**(6):630–40.

Boutron 2008
Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine* 2008;**148**(4): 295–309.

Brink 1994
Brink CA, Wells TJ, Sampselle CM, Taillie ER, Mayer R. A digital test for pelvic muscle strength in women with urinary incontinence. *Nursing Research* 1994;**43**(6):352–6.

Buckley 2009
Buckley BS, Grant AM, Tincello DG, Wagg AS, Firkins L. Prioritizing research: patients, carers and clinicians working together to identify and prioritize important clinical uncertainties in urinary incontinence. *Neurourology and Urodynamics* 2009;**29**(5):708–14.

Bø 1994
Bø K, Stein R. Needle EMG registration of striated urethral wall and pelvic floor muscle activity patterns during cough,

valsalva, abdominal, hip adductor, and gluteal muscle contraction in nulliparous healthy females. *Neurourology and Urodynamics* 1994;**13**:35–41.

Bø 2003
Bø K, Sherburn M, Allen T. Transabdominal ultrasound measurement of pelvic floor muscle activity when activated directly or via a transversus abdominis muscle contraction. *Neurourology and Urodynamics* 2003;**22**:582–8.

Bø 2011
Bø K, Hilde G, Stær-Jensen J, Brækken IH. Can the Paula method facilitate co-contraction of the pelvic floor muscles? A 4D ultrasound study. *International Urogynecology Journal* 2011;**22**:671-6.

Chen 2003
Chen GD, Lin TL, Hu SW, Chen YC, Lin LY. Prevalence and correlation of urinary incontinence and overactive bladder in Taiwanese women.[see comment]. *Neurourology & Urodynamics* 2003;**22**(2):109–17.

CONSORT 2010
Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Annals of Internal Medicine* 2010;**152**(11):726–32.

Dumoulin 2004
Dumoulin C, Lemieux M-C, Bourbonnais D, Gravel D, Bravo G, Morin M. Physiotherapy for persistent postnatal stress urinary incontinence: a randomized controlled trial. *Obstetrics and Gynecology* 2004;**104**:504-10.

Dumoulin 2006
Dumoulin C, Peng Q, Stodkilde-Jorgensen H, Shishido K, Constantinou C. Changes in levator ani anatomical configuration following physiotherapy in women with stress urinary incontinence. *Journal of Urology* 2007;**178**(3 Pt 1): 970-7; quiz 1129.

Dumoulin 2010
Dumoulin C, Hay-Smith J. Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in women. *Cochrane Database of Systematic Reviews* 2006, Issue 1. [DOI: 10.1002/14651858.CD005654]

Fader 2007
Fader M, Cottenden AM, Getliffe K. Absorbent products for light urinary incontinence in women. *Cochrane Database of Systematic Reviews* 2007, Issue 2. [DOI: 10.1002/14651858.CD001406.pub2]

Fader 2008
Fader M, Cottenden AM, Getliffe K. Absorbent products for moderate-heavy urinary and/or faecal incontinence in women and men. *Cochrane Database of Systematic Reviews* 2008, Issue 4. [DOI: 10.1002/14651858.CD007408]

Field 2003
Field A, Hole G. *How to Design and Report Experiments*. London: Sage Publications, 2003.

Hay-Smith 2008
Hay-Smith J, Mørkved S, Fairbrother KA, Herbison GP. Pelvic floor muscle training for prevention and treatment of urinary and faecal incontinence in antenatal and postnatal women. *Cochrane Database of Systematic Reviews* 2008, Issue 4. [DOI: 10.1002/14651858.CD007471]

Herbison 2002
Herbison GP, Dean N. Weighted vaginal cones for urinary incontinence. *Cochrane Database of Systematic Reviews* 2002, Issue 1. [DOI: 10.1002/14651858.CD002114]

Herbison 2009
Herbison P, Hay-Smith J, Paterson H, Ellis G, Wilson D. Research priorities in urinary incontinence: results from citizens' juries. *BJOG: An International Journal of Obstetrics and Gynaecology* 2009;**116**:713–8.

Herderschee 2011
Herderschee R, Hay-Smith EJC, Herbison GP, Roovers JP, Heineman MJ. Feedback or biofeedback to augment pelvic floor muscle training for urinary incontinence in women. *Cochrane Database of Systematic Reviews* 2011, Issue 7. [DOI: 10.1002/14651858.CD009252]

Higgins 2009
Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley and Sons, 2009.

Hunskaar 2002
Hunskaar S, Burgio K, Diokno AC, Herzog AR, Hjalmas K, Lapitan MC. Epidemiology and natural history of urinary incontinence (UI). In: Abrams P, Cardozo L, Khoury S editor(s). *Incontinence*. 2nd Edition. Health Publication Ltd, 2002.

Imamura 2010a
Imamura M, Williams K, Wells M, McGrother C. Lifestyle interventions for the treatment of urinary incontinence in adults. *Cochrane Database of Systematic Reviews* 2010, Issue 9. [DOI: 10.1002/14651858.CD003505.pub4]

Imamura 2010b
Imamura M, Abrams P, Bain C, Buckley B, Cardozo L, Cody J, et al.Systematic review and economic modelling of the effectiveness and cost-effectiveness of non-surgical treatments for women with stress urinary incontinence. *Health Technology Assessment* 2010;**14**(40):1–506.

Lane 2008
Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharmaceutical Statistics* 2008;**7**:93–106.

Laycock 2001
Laycock J, Jerwood D. Pelvic floor muscle assessment: the PERFECT scheme. *Physiotherapy* 2001;**87**(12):631–42.

Laycock 2008
Laycock J, Whelan MM, Dumoulin C. Patient assessment. In: Haslam J, Laycock J editor(s). *Therapeutic management of incontinence and pelvic pain. Pelvic organ disorders*. 2nd Edition. London: Springer-Verlag, 2008:62.

Lose 1998
Lose G, Fantl JA, Victor A, Walter S, Wells TL, Wyman J, et al.Outcome measures for research in adult women with symptoms of lower urinary tract dysfunction. *Neurourology and Urodynamics* 1998;**17**(3):255–62. [MEDLINE: 98250359]

Miller 1998
Miller JM, Ashton-Miller JA, DeLancey JOL. A pelvic muscle precontraction can reduce cough-related urine loss in selected women with mild SUI. *Journal of the American Geriatric Society* 1998;**46**:870–4.

Milsom 2009
Milsom I, Altman D, Lapitan MC, Nelson R, Sillen U, Thom D. Epidemiology of urinary (UI) and faecal (FI) incontinence and pelvic organ prolapse (POP). *4th International Consultation on Incontinence, July 5-8 2008*. Vol. **1**, Health Publications Ltd, 2009:37–111.

Moher 1998
Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al.Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses?. *Lancet* 1998;**352**(9128):609–13. [MEDLINE: 98417104]

Moher 2001
Moher M, Schulz KF, Altman DG, for the CONSORT group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *JAMA* 2001;**285**: 1987–91.

Morin 2004
Morin M, Bourbonnais D. Pelvic floor muscle function in continent and stress urinary incontinent women using dynamometric measurements. *Neurourology and Urodynamics* 2004;**23**(7):668–74.

Nabi 2006
Nabi G, Cody JD, Ellis G, Hay-Smith J, Herbison GP. Anticholinergic drugs versus placebo for overactive bladder syndrome in adults. *Cochrane Database of Systematic Reviews* 2006, Issue 4. [DOI: 10.1002/14651858.CD003781.pub2]

Ogah 2009
Ogah J, Cody JD, Rogerson L. Minimally invasive synthetic suburethral sling operations for stress urinary incontinence in women. *Cochrane Database of Systematic Reviews* 2009, Issue 4. [DOI: 10.1002/14651858.CD006375.pub2]

Pildal 2007
Pildal J, Hrobjartsson A, Jørgensen KJ, Hilden J, Altman DG, Gøtzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomised trials. *International Journal of Epidemiology* 2007;**36**:847–57.

Resende 2011
Resende APM, Zanetti MRD, Petricelli CD, Castro RA, Alexandre SM, Nakamura MU. Effects of the Paula method in electromyographic activation of the pelvic floor: a comparative study. *International Urogynecology Journal* 2011;**22**:677-80.

Sapsford 2004
Sapsford R. Rehabilitation of pelvic floor muscles utilizing trunk stabilization. *Manual Therapy* 2004;**9**(1):3–12.

Shaikh 2006
Shaikh S, Ong EK, Glavind K, Cook J, N'Dow JMO. Mechanical devices for urinary incontinence in women. *Cochrane Database of Systematic Reviews* 2006, Issue 3. [DOI: 10.1002/14651858.CD001756.pub4]

Sluijs 1998
Sluijs EM, Kerssens JJ, van der Zee J, Myers LB. Adherence to physiotherapy (Chapter 14). In: Myers LB, Midence K editor(s). *Adherence to treatment in medical conditions*. Amsterdam: Harwood Academic Publishers, 1998.

Stone 2003
Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Controlled Clinical Trials* 2003;**24**(2):182–99.

Wallace 2004
Wallace SA, Roe B, Williams K, Palmer M. Bladder training for urinary incontinence in adults. *Cochrane Database of Systematic Reviews* 2004, Issue 1. [DOI: 10.1002/14651858.CD001308.pub2]

Whiteneck 2009
Whiteneck G, GassawayJ, Dijkers M, Jha A. New approach to study the contents and outcomes of spinal cord injury rehabilitation: The SCIRehab project. *Journal of Spinal Cord Medicine* 2009;**32**(3):251-9.

Wilson 1987
Wilson PD, Al Samarrai T, Deakin M, Kolbe E, Brown AD. An objective assessment of physiotherapy for female genuine stress incontinence. British Journal of Obstetrics & Gynaecology 1987; Vol. 94, issue 6:575–82.

References to other published versions of this review

Hay-Smith 2001
Hay-Smith EJC, Bø K, Berghmans LCM, Hendriks HJM, de Bie RA, van Waalwijk van Doorn ESC. Pelvic floor muscle training for urinary incontinence in women. *Cochrane Database of Systematic Reviews* 2001, Issue 1. [DOI: 10.1002/14651858.CD001407]

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Borello-France 2006

Methods	2-arm RCT, parallel design Comparison: supine PFMT versus supine and upright PFMT A priori power calculation: no
Participants	44 women with symptoms of SUI from a single centre in the USA Inclusion criteria: aged 38 to 70 years, not pregnant, ambulatory, symptoms of SUI at least once per week Exclusion criteria: UUI symptoms, prior treatment for SUI, previously taught PFM contractions and prescribed PFMT, pacemaker, IUD, vaginal wall prolapse, inability to demonstrate palpable PFM contraction, sensory loss below the L4 dermatome, atrophic vaginitis, lumbo-sacral/pelvic pain, inability to tolerate supine position, detrusor instability, abdominal leak point pressure < 60 cmH_2O, and history of pelvic cancer, severe endometriosis or neurologic/metabolic conditions likely to impair sphincter function Mean (SD) age in years: 51.7 (8.9) versus 53.6 (8.1) Mean (SD) incontinence episodes per week: PFMT 6.9 (7.0) versus 7.2 (5.5) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 7 1. PFMT supine (n = 22): instructed to perform exercises while supine 2. PFMT supine and upright (n = 22): instructed to alternate exercise between supine, sitting and standing positions
Outcomes	Primary endpoint: 9 to 12 weeks Primary outcome measure(s): not stated Other outcome measures: IIQ, urinary diary, modified 1-hour pad test, digital PFM assessment (Brink scale)
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Randomly assigned; adjustments to allocation to balance for age and severity
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes	High risk	Not feasible to blind participants or treatment provider. Outcome assessor blinded for urodynamic outcome assessment. Not blinded for other outcomes

Borello-France 2006 (*Continued*)

Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 8/44 Dropouts by group: supine PFMT 5/22 versus supine + upright PFMT 3/22 ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? Yes, which dealt with missing data using last urinary diary outcome carried forward
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: yes, National Institutes of Health/National Institute on Aging Ethical approval: unclear from trial report if informed consent was given to examination or to participation in trial or both Conflict of interest: not stated

Bø 1990

Methods	2-arm RCT, parallel design Comparison: home PFMT versus home and group supervised PFMT Stratification: by previous surgery, menopause and incontinence severity A priori power calculation: no
Participants	52 women with USI, from a single centre in Norway Inclusion criteria: USI Exclusion criteria: detrusor instability, UTI Mean (range) age in years: 45.9 (35 to 63) versus 44.9 (24 to 64) Mean (range) duration of symptoms in years: 10.8 (1 to 30) versus 8.5 (2 to 27) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 1 1. Home PFMT (n = 29) 2. Home + group supervised PFMT (n = 23 + 5?): addition of weekly 45-minute group exercise session
Outcomes	Primary endpoint: 6 months Primary outcome measure(s): not stated Other outcome measures: Leakage Index, 90-second pad test, urodynamics, vaginal squeeze pressure (perineometer), subjective measure of improvement (5-point Likert scale, worse to continent), Social Activity Index, subjective report of degree of leakage with particular "sporting" activities

Bø 1990 (*Continued*)

Notes	Follow-up data only for the home + class PFMT group	
Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Randomly assigned", random number generation
Allocation concealment (selection bias)	Low risk	Adequate concealment confirmed by first author. Sealed opaque envelopes used
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 5/52 Dropouts by group: all from home + class group ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: Foundation for Education and Research in Physical Therapy, and The Research Council for Science and Humanities Ethical approval: not stated Conflict of interest: not stated

de Oliveira 2009

Methods	2-arm RCT, parallel design Comparison: individual supervision of individualised PFMT versus group supervision of standard PFMT Stratification: no Power calculation: post hoc
Participants	61 women with USI, from a single centre in Brazil Inclusion criteria: 30 to 75 years, no detrusor overactivity, positive stress test, > 3 g leakage on pad test with 200 ml standardised bladder volume, predominant SUI symptoms with at least 3 leakage episodes per week

	Exclusion criteria: chronic neurological or muscular diseases, abdominal genital bleeding, uterine prolapse, active genitor-urinary tract infection, pregnancy, vaginal atrophy, urodynamically confirmed intrinsic sphincter deficiency Mean (SD) age in years: 50.3 (8.7) versus 51.6 (9.6) Mean (SD) duration of symptoms (not stated if months or years): 5.0 (3.9) versus 4.9 (3.0) Mean (SD) leakage episodes per week: 3.1 (1.5) versus 3.3 (1.4) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 2 and Table 4 1. Individual supervision and individualised PFMT (n = 30): PFMT according to PERFECT scheme 2. Group supervision and standard PFMT (n = 30)
Outcomes	Primary endpoint: 12 weeks Primary outcome(s): negative 1-hour pad test (< 2 g weight gain) with 200 ml standardised bladder volume (as per Lose et al 1988) Other outcomes: digital PFM assessment (Oxford scale), 7-day voiding diary, KHQ, subjective cure (dichotomous - satisfied or not)
Notes	Three months of topical hormone therapy was provided before treatment if postmenopausal

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Allocation undertaken using a computer-generated random number generator
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Main investigator was blinded to intervention group allocation, but unclear if main investigator was the same person as outcome assessor
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 1/61 Dropouts by group: not stated ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data

Other bias	Low risk	Funding or financial assistance: none stated Ethical approval: yes, approved by Institutional Review Board Committee Conflict of interest: stated "none"

Delgado 2009

Methods	2-arm RCT, parallel design Comparison: home PFMT versus home PFMT with intravaginal resistance device A priori power calculation: no
Participants	52 women with symptoms of pure SUI or stress-predominant MUI, from a single centre in England Inclusion criteria: > 18 years, minimum of 3 stress leaks per week based on bladder diary, no surgery for incontinence. Exclusion criteria: pregnancy, < 12 weeks post partum, taking duloxetine hydrochloride, recent or recurrent UTI, neurological disease, post void residual > 100 ml and significant pelvic organ prolapse Mean (range) age in years for both groups: 49.6 (36 to 68) Mean (range) duration of symptoms in years for both groups: 5 (0.5 to 30) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 10 1. PFMT (n = 26?) 2. PFMT + resistance (n = 26?)
Outcomes	Primary endpoint:16 weeks, interim at 8 weeks Primary outcome measure(s): self reported improvement (based on response to question 11 of the ICIQ-FLUTS) Other outcome measures: ICIQ-UI short form, ICIQ-LUTSquol, patient satisfaction questionnaire, PGI, estimated per cent improved
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Randomisation sequence was generated independently of the investigator
Allocation concealment (selection bias)	Low risk	Randomisation slips placed into opaque, sequentially-numbered envelopes, which were sealed until interventions were assigned

Delgado 2009 (*Continued*)

Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts 12/52 Dropouts by group: not stated. Review authors estimate 7/26 PFMT versus 5/26 PFMT+ device. ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: stated "study undertaking independently of company input" Ethical approval: yes, approved by Southmead Research Ethics Committee Conflict of interest: not stated

Diniz Zanetti 2007

Methods	2-arm RCT, parallel design Comparison: unsupervised home PFMT versus supervised PFMT Stratified: states that stratification used but no information about the stratification variable(s) A priori power calculation: no
Participants	44 women with USI from a single centre in Brazil Inclusion criteria: urinary leakage observed during a physical examination, at least 3 months of HRT if post-menopausal. Exclusion criteria: any kind of disorder affecting muscle or nerve tissues, genital bleeding, pregnancy, UTI, vulvovaginitis, genital prolapse beyond the hymen, atrophic vaginitis, cardiac pacemaker Median age in years: 54 versus 56 Median symptom duration in years: 5 versus 5 Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 1 1. Unsupervised home PFMT (n = 21): monthly clinic visit for PFM evaluation only 2. Supervised PFMT (n = 23): twice-weekly clinic visits

Diniz Zanetti 2007 (*Continued*)

Outcomes	Primary endpoint: 12 weeks Primary outcome measure (s): not stated Other outcome measures: 7-day urinary diary, 1-hour pad test, I-QoL, satisfaction with progress	
Notes		

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Computer-generated random number table used
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Dropouts: not stated, possibly none ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Unclear risk	Funding or financial assistance: stated "none" Ethical approval: yes, approved by Unifesp-EPM Research Ethics Committee Conflict of interest: stated "none"

Felicissimo 2010

Methods	2-arm RCT, parallel design Comparison: unsupervised home PFMT versus group supervised and home PFMT A priori power calculation: yes
Participants	62 women with USI from a single centre in Brazil Inclusion criteria: predominant SUI with an average of at least 3 stress incontinence episodes per week Exclusion criteria: chronic neurological or muscular diseases, abnormal genital bleeding, genital prolapse (stage 2 or higher of POP-Q), UTI, pregnancy and women who preferred

	surgery, urodynamic confirmed intrinsic sphincter deficiency Mean (SD) age in years: 48.1 (7.7) versus 51.2 (9.4) Mean (SD) duration of symptoms in months: 60.0 (30.1) versus 60.0 (12.1) Trialists stated that the groups were comparable at baseline, except for educational level (which was higher in the supervised PFMT group)
Interventions	Details of PFMT in Table 1 1. Unsupervised home PFMT (n = 31) 2. Group supervised and home PFMT (n = 31): addition of 16 supervised group sessions
Outcomes	Primary endpoint: 8 weeks Primary outcome: 24-hour pad test (as per Wilson 1987 and Laycock 2008) Other outcomes: digital PFM assessment (Oxford scale), ICIQ-UI short form, subjective cure (4-item Likert scale; cured, better, unchanged, worse), satisfaction with treatment
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Allocation of the groups was undertaken using a computer-generated random number generator. Participants were "randomly assigned to two distinct groups"
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 3/62 Dropouts by group: 1/31 unsupervised PFMT versus 2/31 supervised PFMT ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: not stated Ethical approval: yes, approved by Ethics Committee, Federal University of Minas Gerais

Felicissimo 2010 (*Continued*)

		Conflict of interest: stated "none"

Ferguson 1990

Methods	2-arm RCT, parallel design Comparison: PFMT versus PFMT with intravaginal resistance device A priori power calculation: no
Participants	20 women with USI from a single centre in the USA Inclusion criteria: not described Exclusion criteria: postmenopausal, previous urologic surgery, taking medication affecting bladder or skeletal muscle, urinary urgency/frequency or nocturia Mean (SD) age in years: 35.8 (4.6) versus 37.1 (6.4) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 10 1. PFMT (n = 10): home training 2. PFMT+ resistance (n = 10): as for PFMT group, with addition of intravaginal balloon device
Outcomes	Primary endpoint: 6 weeks Longer-term follow-up: 12 to 24 months Primary outcome measure(s): not stated Other outcome measures: 24-hour home pad test, 30-minute office pad test (n = 14), vaginal squeeze pressure (perineometer), urodynamics
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: none ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No

Ferguson 1990 (*Continued*)

Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: National Center for Nursing Research, the National Institutes of Health, and the Department of Obstetrics and Gynecology, University of Florida Ethical approval: not described Conflict of interest: not stated

Gallo 1997

Methods	2 arm quasi-RCT Comparison: PFMT versus PFMT + audiotape A priori power calculation: yes
Participants	86 women with USI from a single centre in the USA Inclusion: aged 20 to 80, USI, desire for conservative treatment, ability to complete questionnaire, willingness to participate Exclusion: pregnancy, psychological disorders making it difficult to follow PFMT instructions Mean (range) age in years for both groups: 60 (29 to 80) Trialists did not state whether the groups were comparable at baseline
Interventions	Details of PFMT in Table 11 1. PFMT (n = 43) 2. PFMT+ audiotape (n = 43): received audiotape of exercise instructions
Outcomes	Primary endpoint: 4 to 6 weeks Primary outcome measure(s): not stated Other outcome measures: frequency of exercise per day, number of minutes of exercise per day, number of seconds of hold per contraction, what prompted woman to perform exercise
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	High risk	"every other patient was randomly assigned by the nurse to the experimental group"
Allocation concealment (selection bias)	High risk	Not further described, but no way to conceal allocation if alternation was used

Gallo 1997 (*Continued*)

Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 11/86 Dropouts by group: 9/43 PFMT, 2/43 PFMT + audiotape ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Unclear risk	Funding or financial assistance: Incare Medical Products Ethical approval: yes, approved by Review Board approval at Beth Israel Hospital Pelvic Floor Unit Conflict of interest: not stated

Ghoniem 2005

Methods	4-arm RCT, 2 x 2 factorial design Comparison: placebo drug + indirect PFMT versus placebo + direct PFMT (and 2 arms not considered in this review: drug + indirect PFMT versus drug + direct PFMT) A priori power calculation: yes
Participants	201 women with USI or SUI from multiple centres in the USA, UK and the Netherlands Inclusion criteria: aged 18 to 75, USI or positive cough stress test and normal micturition frequency Exclusion criteria: USI with DO within 6 months before trial entry, pelvic organ prolapse, active or recurrent UTI, continence surgery within year, current device or pharmaceutical incontinence treatment, prior hip fracture or replacement, prior formal PFMT with continence nurse or physical therapist Mean (range) age in years: 51 (29 to 68) versus 54 (36 to 75) (and 53 (34 to 70) versus 54 (31 to 75)). Median (range) incontinence episodes per week: 18.9 (10.3 to 299.4) versus 22.0 (13.0 to 140.9) (and 18.3 (6.4 to 78.5) versus 19.4 (10.0, 70.5)) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 3 1. Placebo drug and indirect PFMT (n = 47): hip adductor exercise 2. Placebo drug and direct PFMT (n = 50): taught PFMT by "qualified instructor" 3. Duloxetine and indirect PFMT (n = 52): this arm not considered in this review 4. Duloxetine and direct PFMT (n = 54): this arm not considered in this review

Outcomes	Primary endpoint: 12 weeks Primary outcome measure(s): incontinence episode frequency Other outcome measures: number of incontinence pads used, I-QOL score, PGI	
Notes		

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"random"
Allocation concealment (selection bias)	Low risk	Allocation of concealment was independent of trial sites - computer voice response system
Blinding (performance bias and detection bias) All outcomes	Unclear risk	It was not feasible to blind treatment providers to the PFMT intervention. An attempt was made to blind participants by using an "imitation" PFMT programme (i.e. indirect PFMT). Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Low risk	Total dropouts 21/201 Dropouts by group: 7/97 receiving placebo and 17/107 receiving duloxetine (but not further broken down by PFMT) ITTA: 1. Participants analysed in group to which assigned? Yes 2. Authors stated analysis by intention-to-treat? Yes, which dealt with missing data by using last outcome carried forward
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	High risk	Funding or financial assistance: yes, Elli Lilly and Company, and Boehringer Ingelheim Ethical approval: states that "study received ethical approval" Conflict of interest: all but the first author declared a financial interest and/or other relationship with Eli Lilly and Company

Hay-Smith 2002

Methods	2-arm RCT, parallel design Comparison: motor learning PFMT versus motor learning and strength training PFMT A priori power calculation: yes
Participants	128 women with symptoms of SUI from a single centre in New Zealand Inclusion criteria: reported SUI symptoms, 2 or more leakage episodes/week, toileted independently, lived in the community Exclusion criteria: reversible causes of incontinence, uncontrolled metabolic conditions, clinical history/uroflowmetry indicated voiding difficulty, UTI, pelvic organ prolapse below hymenal ring, unable to perform correct VPFMC after instruction, use of concomitant therapies for incontinence, age < 16, inability to read, write or speak English Mean (SD) age in years: 48.9 (13.1) versus 48.7 (13.2) Mean (SD) duration SUI symptoms in years: 9.1 (9.1) versus 8.7 (9.4) Mean (SD) leakage episodes per day: 1.7 (1.7) versus 1.9 (2.2) Trialist stated that the groups were comparable at baseline, except for digital palpation of PFM grade (there was a greater proportion of women with grade 4 or 5 contraction in the motor learning PFMT group)
Interventions	Details of PFMT in Table 8 1. Motor learning PFMT (n = 64) 2. Motor learning and strength training PFMT (n = 64)
Outcomes	Primary endpoint: 20 weeks Primary outcome measures: Paper Towel Test Other outcome measures: women's observation of change in leakage (6-point Likert scale, much worse to cured), leakage episodes (taken from diary), desire for further treatment, 24-hour pad test, other urinary diary measures, vaginal squeeze pressure, KHQ
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Block randomisation, using randomly selected block length (2, 4, 6 or 8)
Allocation concealment (selection bias)	Low risk	Random number sequence decoded by research assistant, transferred to plain white card and sealed in sequentially numbered sealed opaque envelopes. Released one at a time to researcher once participant agreed to participate
Blinding (performance bias and detection bias) All outcomes	Low risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor

Hay-Smith 2002 (*Continued*)

Incomplete outcome data (attrition bias) All outcomes	Low risk	Total dropouts 5/128 Dropouts by group: 3/64 motor learning PFMT versus 2/64 motor learning + strengthening PFMT ITTA: 1. Participants analysed in group to which assigned? Yes 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Unclear risk	Funding or financial assistance: Otago Medical Research Foundation, and NZ Physiotherapy Trust Fund Ethical approval: yes, approved by Otago Ethics Committee Conflict of interest: not stated

Hung 2010

Methods	2-arm RCT, parallel design Comparison: unsupervised home PFMT versus supervised "Sapsford" approach to PFMT A priori power calculation: yes
Participants	70 women with symptoms of SUI or MUI, from a single centre in Taiwan Inclusion criteria: aged 18 to 65, > 1 episode of SUI in previous month Exclusion criteria: UUI only, pregnant, < 3 months postpartum, systemic neuromuscular disease, previous surgery, previous intensive PFMT, severe low back or pelvic pain, concurrent treatment for incontinence or low back pain, radical hysterectomy, ongoing UTI Mean (SD) age in years: 48.9 (6.4) versus 48.6 (6.4) Mean (SD) duration of symptoms in months: 98.9 (71.2) versus 104.5 (89.7) Median (IQR) leakage episodes per day: 0 (0 to 0.9) versus 0 (0 to 0.7) Trialists stated that the groups were comparable at baseline except for the proportion of women with MUI (there were more women with urgency and UUI in the "Sapsford" PFMT group)
Interventions	Details of intervention in Table 1 and Table 3 1. Unsupervised home PFMT (n = 35) 2. Supervised "Sapsford" PFMT (n = 35): 16-week highly structured programme. Started with correct diaphragmatic breathing. Progressed to combination of transversus abdominus muscle contraction with PFM contraction, under increasingly demanding conditions

Outcomes	Primary endpoint: 4 months Primary outcome measures: self reported improvement Other outcome measure: 20-minute pad test with standardised bladder volume, vaginal squeeze pressure (perineometer), Chinese version (Chen 2003) of the Symptom Impact Index (Black 1996), adverse events, attendance
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Block randomisation with maximum of 6
Allocation concealment (selection bias)	Low risk	Each participant chose and opened one opaque sealed envelope by herself
Blinding (performance bias and detection bias) All outcomes	Low risk	Not feasible to blind participants or treatment provider. All measures, except pad test, were administered by the same blinded evaluator. A female technician blinded to participants' allocation administered the pad test
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 6/70 Dropouts by group: 2/35 PFMT versus 4/35 Sapsford PFMT ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? Yes, which dealt with missing data by using baseline or last value carried forward
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Unclear risk	Funding or financial assistance: National Science Council of the Republic of China Ethical approval: yes, approved by Ethics Committee, Federal University of Minas Gerais Conflict of interest: not stated

Johnson 2001

Methods	2-arm RCT, parallel design Comparison: strengthening PFMT versus endurance PFMT A priori power calculation: yes
Participants	37 women with USI, from a single centre in the USA Inclusion criteria: self report of > 2 SUI leak episodes/day, aged 35 to 65, confirmed USI, English speaking, not pregnant, free of bladder or vaginal infection, not currently taking medications for SUI, adequate oestrogenisation of vaginal mucosa. Exclusion criteria: history of urethral collagen injection, neuromuscular disease or radical pelvic/perineal surgery, or other serious physical or psychological problems Mean (SD) age in years: 49.5 (11.09) versus 51 (10.21) Mean (SD) symptom duration in months: 149.38 (160.38) versus 98.25 (131.38) Mean (SD) leakage episodes per day: 3.16 (1.85) versus 4.04 (3.32) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Additional Table 5 1. Strengthening PFMT (n = 16 after dropouts): near-maximal voluntary contractions performed at 90% maximal force 2. Endurance PFMT (n = 16 after dropouts): sub-maximal voluntary contractions performed at 60 % maximal force
Outcomes	Primary endpoint: 7 weeks Primary outcome measure: not stated Other outcome measures: PFM EMG (endurance, duration of individual sustained contractions, mean maximal contractions, muscle activity recruitment), leakage episodes from daily diary, self reported leakage severity (7-point Likert scale), 10-hour pad test
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Randomly assigned using table of random numbers
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes	High risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 5/37 Dropouts by group: not stated ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-

Johnson 2001 (*Continued*)

		treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: TTUHSC School of Nursing, EA Franklin Foundation, Iota Mu Chapter of Sigma Theta Tau International, and support-in-kind from Incare Medical Products Ethical approval: yes, approved by local research ethics committee Conflict of interest: not stated

Konstantinidou 2007

Methods	2-arm quasi-RCT, parallel design Comparison: home PFMT versus home and group PFMT A priori power calculation: yes
Participants	30 women with USI, from a single centre in Greece Inclusion criteria: aged > 18 years, clinical diagnosis SUI for > 3 months, > 7 episodes incontinence/week, daytime frequency < 8 voids, < 3 night-time voids, positive stress test (urine leakage with cough at 400 ml bladder capacity), positive 24-hour pad test (> 4 g in 24 hours), PFM Oxford grade 3 or 4 Exclusion criteria: symptoms of urgency or UUI, any degree of pelvic organ prolapse, pregnancy, co-morbidities from or affecting the urinary tract such as diabetes mellitus, neurological disease, psychiatric illness, use of medication affecting micturition, prior pharmacological or surgical treatment for SUI, chronic debilitating diseases such as renal failure Mean (SD) age in years for both groups: 47.8 years (7.5) Mean (SD) duration of SUI symptoms in years: 6.4 (3.9) versus 5.7 (2.8) Mean (SD) incontinence episodes per week: 14.8 (6.1) versus 12.2 (4.8) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 2 1. Home PFMT (n = 15) 2. Home and group supervised PFMT (n = 15): addition of small-group (5 participants) training session weekly
Outcomes	Primary endpoint: 12 weeks Primary outcome measure: PGI Other outcome measures: number of incontinence episodes from voiding diaries, digital PFM assessment (Oxford scale), 24-hour pad test (negative if < 2 g), Quality of Life index
Notes	
Risk of bias	

Konstantinidou 2007 (*Continued*)

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Randomised by the recruiting physician in "consecutive alternate fashion according to their hospital administration sequence"
Allocation concealment (selection bias)	High risk	See above
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 8/30 Dropouts by group: 3/15 home PFMT versus 5/15 home and group PFMT ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: not stated Ethical approval: yes, approved by local research committee board Conflict of interest: stated "none"

Liebergall 2005

Methods	2-arm RCT, parallel design Comparison: group PFMT versus individual "Paula Method" PFMT Stratification: by age A priori power calculation: yes
Participants	63 women with signs of SUI or MUI, from multiple centres in Israel Inclusion criteria: aged 20 to 65, pad test leak of > 1 g urine at initial assessment Exclusion criteria: pregnancy, severe cardiac/respiratory disease, pelvic surgery within previous 6 months, grade 3 or 4 cystocele, previous pelvic radiation, active mucosal lesion in the perineum/vagina Per cent aged 51 and 65 years: 51.7 versus 50.0 Per cent leaking once or more per day: 44.8 versus 43.3 Trialists did not state whether the groups were comparable at baseline
Interventions	Details of PFMT in Table 2 and Table 3 1. Group PFMT (n = 29 after dropouts) 2. Paula method PFMT (n = 30 after dropouts): "Paula method" of circular muscle

	(sphincter) contraction
Outcomes	Primary endpoint: 12 weeks Primary outcome measure: change in 1-hour pad test Other outcome measures: continence questionnaire, change in headache, constipation, backache, I-QOL, digital PFM assessment (normal or not normal tone), vaginal squeeze pressure (perineometer)
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Randomisation done according to random number table in blocks of 4
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Study gynaecologists were blinded to allocation but it is not clear if the gynaecologists were the outcome assessors or not
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 4/59 Dropouts by group: 2/29 PFMT versus 2/30 Paula PFMT ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: Internal grant for Paramedical Personnel at Hadassah, the Lillian Silverstein Fund Ethical approval: yes, approved by Institutional Review Board at Hadassah Medical Organisation. Conflict of interest: not stated

Liebergall 2009

Methods	2-arm RCT, parallel design Comparison: individually supervised PFMT versus individual "Paula Method" PFMT Stratification: by age and place of residence A priori power calculation: yes
Participants	245 women with symptoms and signs of SUI from a (multiple?) centres in Israel Inclusion criteria: aged 20 to 65, history of SUI, > 1 g leakage on 1-hour clinic pad test, able to understand Hebrew or English Exclusion criteria: pregnancy or breastfeeding, within 12 weeks of delivery or 6 weeks of abortion or 6 months of pelvic surgery, cardiac, respiratory, psychiatric, neurologic illness that limit physical activity, SUI symptoms but < 1 g leakage on pad test, grade III or higher uterovaginal prolapse, previous UI surgery, previous pelvic radiation therapy Mean (SD) age in years: 47.9 (8.4) versus 47.3 (8.4) Trialists stated that the groups were comparable at baseline, except for uterine prolapse (a higher proportion of women with prolapse were in the "Paula" PFMT group)
Interventions	Details of PFMT in Table 2 and Table 3 1. Individually supervised PFMT (n = 117) 2. Individually supervised Paula method PFMT (n = 123): "Paula method" of circular muscle (sphincter) contraction
Outcomes	Primary endpoint: 12 weeks Primary outcome measures: 1-hour pad test Other outcome measures: subjective assessment of symptoms, I-QoL (Hebrew version)
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Randomisation by using a table of random numbers prepared by a biostatistician; block size of 4
Allocation concealment (selection bias)	Low risk	Study numbers were assigned to the women when they completed the questionnaires. Phone number of research co-ordinator was given from whom they could obtain their assignment after informing her of their study number. The randomisation list was kept by research co-ordinator and not broken until the last participant completed her postintervention test
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Primary outcome (pad test) measurement was blinded. Otherwise, blinding of outcome assessors not described

Liebergall 2009 (*Continued*)

Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Dropouts: 5/245 dropped out immediately after randomisation and before intervention. A further 57/245 dropouts in the intervention phase Dropouts by group: 36/123 PFMT versus 21/117 Paula PFMT ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? A best-worst case scenario in order to approximate an intention-to-treat analysis was used
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Unclear risk	Funding or financial assistance: The Hadassah Women's Health Research Fund, the Berman Family Foundation Ethical approval: yes, approved by Institutional Review Board of Hadassah Conflict of interest: stated "none"

Ng 2008

Methods	2-arm RCT, parallel design Comparison: PFMT versus PFMT with phone call follow-up A priori power calculation: yes
Participants	88 women with symptoms of MUI, from a single centre in Taiwan Inclusion criteria: interest in behavioral training, ambulatory, available for phone contact Exclusion criteria: no educational background, not independent in daily activities Mean (SD) age in years: 52.3 (14) versus 54.0 (13.6) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 1 1. PFMT (n = 34 after dropouts) 2. PFMT with phone call follow-up (n = 34 after dropouts): fortnightly telephone contact with nurse
Outcomes	Primary endpoint: probably 6 months Primary outcome measure: not stated Other outcome measures: B-FLUFTS, Symptom Impact Index (as per Black et al 1996)
Notes	
Risk of bias	

Ng 2008 (*Continued*)

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 20/88 Dropouts by group: 10/44 PFMT versus 10/44 PFMT with phone calls ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? Yes, which dealt with missing data using baseline values carried forward and self reported improvement was given as unchanged
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: National Science Council in Taiwan Ethical approval: yes, approved by the Chung Shan Institutional Review Board Conflict of interest: not stated

Ramsay 1990

Methods	2-arm RCT, parallel design Comparison: indirect PFMT versus direct PFMT A priori power calculation: no
Participants	44 women with SUI from a single centre in the UK Inclusion criteria: not described Exclusion criteria: not described Trialists did not state whether the groups were comparable at baseline
Interventions	Details of PFMT in Table 3 1. Indirect PFMT (n = 22): maximal hip adductor contractions with feet crossed at ankles 2. Direct PFMT (n = 22)

Ramsay 1990 (*Continued*)

Outcomes	Primary endpoint: 3 months Primary outcome measures: not stated Other outcome measures: subjective assessment of severity of problem, amount and frequency of urine lost, vaginal squeeze pressure (perineometry), pad test
Notes	Conference abstract only

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"random"
Allocation concealment (selection bias)	Unclear risk	See above. Not further described.
Blinding (performance bias and detection bias) All outcomes	Unclear risk	It was not feasible to blind treatment providers to the PFMT intervention. An attempt was made to blind participants by using a "sham" PFMT programme (i.e. indirect PFMT). Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: not stated ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Unclear risk	Abstract only. Incomplete data.
Other bias	Unclear risk	Funding or financial assistance: not stated Ethical approval: not stated Conflict of interest: not stated

Savage 2005

Methods	2-arm RCT, parallel design Comparison: PFMT versus "Pilates" approach to PFMT A priori power calculation: no, pilot study
Participants	11 women with SUI, from multiple centres in the UK Inclusion criteria: clinical history of SUI (as per Laycock 2001), leaks on cough/sneeze, jump or movement Exclusion criteria: incontinence symptoms other than SUI, prolapse, positive urinalysis, pregnancy, concomitant treatments, pathology affecting ability to exercise, neurological or psychiatric conditions, birth or gynaecological surgery in previous 6 months, physio-

	therapy for SUI in previous 2 years, already practising Pilates, unable to attend training sessions Mean (range) age in years: 54.6 (37 to 79) versus 48.1 (not stated) Mean (range) duration of symptoms in years: 7.5 (1 to 14) versus 6.0 (1.2 to 10) Trialists stated that the groups were comparable at baseline, although with such small numbers in each group it was difficult to be sure
Interventions	Details of PFMT in Table 3 1. PFMT (n = 5) 2. Pilates approach to PFMT (n = 6): 12 week structured lumbopelvic stability (Pilates) training programme
Outcomes	Primary endpoint: 13 weeks Primary outcome measures: KHQ Other outcome measures: digital PFM assessment (Oxford scale), satisfaction with treatment
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"Block randomisation"
Allocation concealment (selection bias)	Unclear risk	See above. Not further described.
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinded outcome assessment of PFM function but blinding of outcome assessment not further described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 1/11 Dropouts by group: 1/5 PFMT ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Unclear risk	Funding or financial assistance: not stated (note: masters thesis) Ethical approval: yes, approved by local research ethics committee Conflict of interest: not stated

Sriboonreung 2011a

Methods	3-arm RCT, parallel design Comparison: daily home PFMT versus thrice weekly home PFMT (and one arm considered in Sriboonreung 2011b: versus thrice-weekly home PFMT and abdominal muscle training) A priori power calculation: yes
Participants	68 women with symptoms of SUI from a single centre in Thailand Inclusion criteria: aged 35 to 65 years, SUI symptoms and signs (based on ICIQ-SF, frequency volume chart for 3 days, physical examination and positive cough stress test), pad test weight more than 2 grams, and signed consent form Exclusion criteria: prolapsed uterus, reversible cause of urinary incontinence (e.g. fecal impaction, drug effect), uncontrolled metabolic condition (e.g. diabetes mellitus), serious chronic condition that may result in neurogenic bladder dysfunction, residual urine > 100 ml, urinary tract infection, genito-urinary fistula, previous surgery for SUI, inability to correctly perform a pelvic muscle contraction on digital examination, neurological disease that resulted in combination of bladder and sphincter dysfunction Mean (SD) age in years: 51.4 (6.1) versus 51.5 (6.6) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 6 1. Daily home PFMT (n = 23) 2. Thrice-weekly home PFMT (n = 22)
Outcomes	Primary endpoint: 12 weeks Primary outcome measures: 1-hour pad test Other outcome measures: vaginal squeeze pressure (perineometry) and treatment satisfaction
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"the subjects were randomly allocated into each of three treatment arms using block randomized allocation with block sizes of 3, 6 and 9 enclosed in envelopes"
Allocation concealment (selection bias)	Unclear risk	See above. Not further described.
Blinding (performance bias and detection bias) All outcomes	High risk	Not feasible to blind participants or treatment provider. Outcome assessor was not blinded
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 8/68 Dropouts by group: 3/23 daily PFMT versus 3/22 thrice weekly PFMT ITTA:

Sriboonreung 2011a (*Continued*)

		1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Low risk	Funding or financial assistance: not stated Ethical approval: yes, approved by ethics committee of the Faculty of Medicine, Chiang Mai University Conflict of interest: stated "none"

Sriboonreung 2011b

Methods	3-arm RCT, parallel design Comparison: thrice-weekly home PFMT versus thrice-weekly home PFMT and abdominal muscle training (and one arm considered in Sriboonreung 2011a: versus daily home PFMT) A priori power calculation: yes
Participants	68 women with symptoms of SUI from a single centre in Thailand Inclusion criteria: aged 35 to 65 years, SUI symptoms and signs (based on ICIQ-SF, frequency volume chart for 3 days, physical examination and positive cough stress test), pad test weight more than 2 grams, and signed consent form Exclusion criteria: prolapsed uterus, reversible cause of urinary incontinence (e.g. fecal impaction, drug effect), uncontrolled metabolic condition (e.g. diabetes mellitus), serious chronic condition that may result in neurogenic bladder dysfunction, residual urine > 100 ml, urinary tract infection, genito-urinary fistula, previous surgery for SUI, inability to correctly perform a pelvic muscle contraction on digital examination, neurological disease that resulted in combination of bladder and sphincter dysfunction Mean age in years (SD): 53.8 (5.6) versus 51.5 (6.6) Mean pad weight in grams (SD): 4.7 (1.6) versus 4.0 (1.5) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 9 1. Thrice weekly home PFMT (n = 22) 2. Thrice weekly home PFMT and abdominal muscle training (n = 23): abdominal muscle training was also 3 times a week focused on the transversus abdominus and internal oblique muscles
Outcomes	Primary endpoint: 12 weeks Primary outcome measures: 1-hour pad test Other outcome measures: vaginal squeeze pressure (perineometry) and treatment satisfaction
Notes	

Sriboonreung 2011b (*Continued*)

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	See Sriboonreung 2011a
Allocation concealment (selection bias)	Unclear risk	See Sriboonreung 2011a
Blinding (performance bias and detection bias) All outcomes	High risk	See Sriboonreung 2011a
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 8/68 Dropouts by group: 3/22 thrice-weekly PFMT versus 2/23 thrice-weekly PFMT and abdominal training ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No 3. Full intention-to-treat used? No
Selective reporting (reporting bias)	Low risk	See Sriboonreung 2011a
Other bias	Low risk	See Sriboonreung 2011a

Sugaya 2003

Methods	2-arm quasi-RCT, parallel design Comparison: PFMT versus PFMT with chime device A priori power calculation: no
Participants	46 women with symptoms of SUI, from multiple centres in Japan Inclusion criteria: SUI Exclusion criteria: UTI, prior experience of PFMT, organic bladder disease Mean (SD) age in years: 58.4 (10.8) versus 58.1 (13.7) Mean (SD) duration of symptoms in years: 5.1 (3.1) versus 7.3 (4.7) Mean (SD) incontinence episodes per day: 3.3 (2.2) versus 3.4 (2.1) Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 11 1. PFMT (n = 23) 2. PFMT + chime device (n = 23): alarm device that also included recorded PFMT programme instructions

Sugaya 2003 (*Continued*)

Outcomes	Primary endpoint: 8 weeks Primary outcome measure(s): not stated Other outcome measures: incontinence episodes, number of pads used, Quality of Life index for urination, 1-hour pad test, compliance (self reported and recorded by device for group 2)
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	High risk	"divided into two groups" "in order of presentation"
Allocation concealment (selection bias)	High risk	See above. No further description.
Blinding (performance bias and detection bias) All outcomes	Unclear risk	Not feasible to blind participants or treatment provider. Blinding of outcome assessor not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts 5/46 Dropouts by group: 3/23 PFMT versus 2/23 PFMT with device ITTA: 1. Participants analysed in group to which assigned? Not stated 2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Unclear risk	Funding or financial assistance: Kyusyu Micturition Disorder Study Group Ethical approval: not stated Conflict of interest: not stated

Wells 1999

Methods	4-arm RCT, parallel design Comparison: PFMT versus PFMT with intravaginal resistance device (and 2 arms not considered in this review: "self-insight" versus "health promotion") A priori power calculation: no

Participants	286 women with stress or mixed urinary incontinence from a single centre in the USA Inclusion criteria: stress or mixed urinary incontinence, aged 21 years or more, independent in self care, able to speak and hear conversation in English adequately over the telephone, negative urinalysis, able to contract the PFM as demonstrated on physical examination, and able to read, understand and agree to the diagnostic consent form Exclusion criteria: degenerative neurological disorder, pregnancy, high risk of infection following urologic instrumentation Mean (SD) age in years in all groups: 56 (12.76) Symptom duration > 1 year in all groups: 68% Trialists stated that the groups were comparable at baseline
Interventions	Details of PFMT in Table 10 1. PFMT (n = 71) 2. PFMT with intravaginal resistance device (n = 71) 3. "Self-insight" (n = 72): observation of voiding behaviour and lifestyle. No clinic visits. This arm not considered in this review. 4. "Health promotion" (n = 72): good bladder hygiene, individualised lifestyles intervention and monthly clinic visits. This arm not considered in this review
Outcomes	Primary endpoint: 5 months Primary outcome measure(s): not stated Other outcome measures: PFM EMG (PerryMeter anal or vaginal sensor attached to the Dantec urodyn 5000 system), digital PFM assessment (Brink Scale, Brink 1994) , urethral pressure and cough test (urodynamics), 10-point VAS for leakage (0 = no leakage, 10 = a lot of leakage), number of leak episodes per day, pad test with standardised bladder volume
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"randomly assigned"
Allocation concealment (selection bias)	Unclear risk	See above. Not further described.
Blinding (performance bias and detection bias) All outcomes	High risk	Not feasible to blind participants or treatment provider. Outcome assessors not blinded
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total dropouts: 127/286 Dropouts by group: 30/71 PFMT, 32/71 PFMT + device (35/72 self insight, 30/72 health promotion) ITTA: 1. Participants analysed in group to which assigned? Not stated

		2. Authors stated analysis by intention-to-treat? No
Selective reporting (reporting bias)	Low risk	Nothing to suggest selective reporting of data
Other bias	Unclear risk	Funding or financial assistance: National Institute for Nursing Research Ethical approval: yes, Institutional Review Board of the University of Rochester Conflict of interest: not stated

BFLUTS = Bristol Female Lower Urinary Tract Symptoms questionnaire, DO = detrusor overactivity, EMG = electromyography, HRT = hormone replacement therapy, ICIQ-FLUTS = International Consultation on Incontinence-Female Lower Urinary Tract symptoms, ICIQ-LUTSqol = International Consultation on Incontinence-Lower Urinary Tract symptoms quality of life, ICIQ-UI short form = International Consultation on Incontinence-Urinary Incontinence short form, IIQ = Incontinence Impact Questionnaire, I-QoL = Incontinence Quality of Life, IQR = interquartile range, ITTA = intention-to-treat analysis, IUD = intrauterine device, KHQ = Kings Health Questionnaire,kn = kneeling, LUT = lower urinary tract, ly = lying, min(s) = minutes(s), MMSE = mini mental state examination, MUI = mixed urinary incontinence, OAB = overactive bladder, PERFECT = power or pressure, endurance, repetitions, fast contractions, every contraction timed, PFM = pelvic floor muscle(s), PFMT = pelvic floor muscle training, PGI = Patient Global Impression of improvement, POP-Q = pelvic organ prolapse quantified, RCT = randomised controlled trial, SD = standard deviation, sec(s) = second(s), sitt = sitting, st = standing, SUI = stress urinary incontinence, USI = urodynamic stress incontinence, UTI = urinary tract infection, UUI = urgency urinary incontinence, VAS = visual analogue scale, VPFMC = voluntary pelvic floor muscle contraction.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Crothers 2001	2-arm RCT comparing PFMT with adherence device versus PFMT alone in women with urinary incontinence. The trial record was found in a register of randomised trials. The first author confirmed that she no longer has the data. A copy of the trial report was not available from the funder; it appeared the report was discarded. The trial was excluded because no data were available
de Jong 2006	2-arm RCT comparing PFMT with whole body vibration versus PFMT alone. Reported only as a brief conference abstract that contained no data. Excluded because no were data available
Dumoulin 2003	3-arm RCT comparing pelvic floor rehabilitation versus pelvic floor rehabilitation with abdominal muscle training versus control (shoulder and back massage). Reported in 2 conference abstracts that were considered for inclusion. However, in the full publication (Dumoulin C et al, Obstet Gynecol 2004; 104:504-10) it was clear the participants were postnatal women with persistent stress urinary incontinence. As the participants were postnatal women the trial was excluded
Hill 2007	2-arm RCT comparing group versus individual management of urinary incontinence in women. Reported in a conference abstract. The abstract suggested that both arms received bladder training in addition to

(*Continued*)

	PFMT and the full trial report (Lamb SE et al, BMC Women's Health 2009, 9:26 doi:10.1186/1472-6874-9-26) confirmed this. The trial was excluded because of the addition of bladder training to PFMT
Hui 2006	2-arm RCT comparing a telemedicine continence programme versus conventional outpatient continence service in women with urgency or stress incontinence. All women had an initial PFMT session with a health professional. The first author confirmed that both programmes included bladder training. The trial was excluded because of the addition of bladder training to PFMT
Klinger 1995	2-arn RCT comparing PFMT versus PFMT with "Endotrainer" in women with stress urinary incontinence. Reported only as a brief conference abstract. The abstract stated that the Endotrainer was an "intermittent gas filled balloon placed in the vagina which has to be compressed by the patient under audio-visual biofeedback control". Excluded because, even if this were a trial of an intravaginal resistance device, the comparison was confounded by the use of biofeedback in only one arm of the trial (see Types of interventions).
Nygaard 1996	2-arm RCT comparing PFMT + audiotape versus PFMT alone in women with stress, urge or mixed urinary incontinence. The paper reported data by diagnostic group rather than by the group to which participants were assigned. The first author confirmed that these data were no longer retrievable. The trial was excluded because no data were available
Orelle NatraTone 2008	2-arm RCT comparing PFMT with Orelle NatraTone device versus PFMT alone. A pilot randomised trial completed on behalf of Orelle Corporation by the Health and Rehabilitation Research Centre, Auckland University of Technology, New Zealand. Orelle Corporation kindly provided a copy of the report; the report contained no raw data for any of the outcomes of interest. An enquiry to the report authors, via Orelle Corporation, did not result in release of the data. The trial was excluded because no data were available
Taylor 1986	4-arm RCT of approximately 12 participants, comparing PFMT versus PFMT with home biofeedback versus PFMT with intravaginal resistance device and home biofeedback versus PFMT with clinic biofeedback. While one comparison in this trial met the inclusion criteria for the review (i.e. PFMT with home biofeedback versus PFMT with intravaginal resistance device and home biofeedback) it was excluded for the following reasons (as we had no response from the authors); 1) Participants were randomised to one of 4 groups but neither the number of participants for the whole trial, nor per group was stated; 2) No endpoint was stated; 3) No useable data were reported
Wong 1997	2-arm RCT comparing clinic PFMT and home PFMT in women with urodynamic stress urinary incontinence. Reported only as a brief conference abstract that contained no data. No data were provided by the authors. Excluded because no were data available
Yoon 1999	3-arm, probably quasi-randomised, trial. Compared functional electrical stimulation and biofeedback assisted PFMT versus intensive PFMT versus verbal instruction in PFMT in women with stress urinary incontinence. Reported only as a brief conference abstract that contained no data. No response from authors. Excluded because no data were available

PFMT = pelvic floor muscle training, RCT = randomised controlled trial.

Characteristics of ongoing studies *[ordered by study ID]*

Kincade 2005

Trial name or title	
Methods	RCT. In phase I women randomised to either self monitoring (lifestyle advice and The Knack) or wait list control. After 3 weeks all women have urodynamic investigation and enter phase II, where randomised to one of 3 arms - PFMT, PFMT with biofeedback, or attentional control (clinic visits and information on maintenance of a healthy lifestyle)
Participants	184 women who completed phase 1 of study Age > 18 with SUI or MUI
Interventions	Phase I: self monitoring, wait list control Phase II: PFMT, PFMT + BF, attentional control
Outcomes	3-day bladder diary, 48-hour pad test, pelvic floor muscle EMG, Incontinence Impact Questionnaire, patient self report of improvement (Likert scale), satisfaction with outcome (Likert scale), rating of improvement (VAS)
Starting date	May 2002
Contact information	Molly C. Dougherty, School of Nursing, CB#7460 The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7460, USA m_dougherty@unc.edu
Notes	To date 2 papers have been published - methods, phase I data. Data of interest for this review are those from phase II

von der Heide 2003

Trial name or title	
Methods	3- arm RCT, parallel design
Participants	29 women with USI, from a single centre in Germany
Interventions	1. PFMT for 12 weeks, addition of vibration for last 12 weeks 2. PFMT + vibration for 24 weeks 3. PFMT + vibration for 12 weeks, continued without vibration for last 12 weeks
Outcomes	Primary endpoint: 24 weeks Primary outcome measures: cough test Other outcome measures: subjective frequency of urine loss, PFM strength, objective cure
Starting date	
Contact information	Silke von der Heide. s.vdh@t-online.de

Notes	This study was reported in a Doctoral Dissertation submitted in 2007. The first author has confirmed a full publication of the data is planned. Until the paper is published the first author prefers not to release the data for inclusion in the review

BF = biofeedback, EMG = electromyography, ICIQ-FLUTS = International Consultation on Incontinence-Female Lower Urinary Tract Symptom score, MRC = Medical Research Council (UK), MUI = mixed urinary incontinence, PFM = pelvic floor muscle(s), PFMT = pelvic floor muscle training, RCT = randomised controlled trial, SUI = stress urinary incontinence, USI = urodynamic stress incontinence, VAS = visual analogue scale, VPFMC = voluntary pelvic floor muscle contraction

DATA AND ANALYSES

Comparison 1. More versus less contact with health professionals

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not cured	3		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
1.1 Additional group supervision (with no difference in PFMT)	2	111	Risk Ratio (M-H, Fixed, 95% CI)	0.89 [0.78, 1.03]
1.2 Additional phone calls (with no difference in PFMT)	0	0	Risk Ratio (M-H, Fixed, 95% CI)	0.0 [0.0, 0.0]
1.3 Individual supervision versus no supervision (with differences in PFMT)	1	64	Risk Ratio (M-H, Fixed, 95% CI)	0.86 [0.73, 1.02]
2 Patients' perception of change in incontinence - not improved	5		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
2.1 Additional group supervision (with no difference in PFMT)	4	177	Risk Ratio (M-H, Random, 95% CI)	0.29 [0.15, 0.55]
2.2 Additional phone calls (with no difference in PFMT)	0	0	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
2.3 Individual supervision versus no supervision (with difference in PFMT)	1	64	Risk Ratio (M-H, Random, 95% CI)	0.10 [0.01, 0.71]
3 Incontinence specific quality of life			Other data	No numeric data
3.1 Additional group supervision (with no difference in PFMT)			Other data	No numeric data
3.2 Additional phone calls (with no difference in PFMT)			Other data	No numeric data
3.3 Individual supervision versus no supervision (difference in PFMT)			Other data	No numeric data
4 Symptom impact			Other data	No numeric data
4.1 Additional group supervision (with no difference in PFMT)			Other data	No numeric data
4.2 Additional phone calls (with no difference in PFMT)			Other data	No numeric data
4.3 Individual supervision versus no supervision (difference in PFMT)			Other data	No numeric data
5 Frequency of leakage - leakage episodes in 24 hours	1	22	Mean Difference (IV, Fixed, 95% CI)	-1.38 [-2.04, -0.72]

5.1 Additional group supervision (with no difference in PFMT)	1	22	Mean Difference (IV, Fixed, 95% CI)	-1.38 [-2.04, -0.72]
5.2 Additional phone calls (with no difference in PFMT)	0	0	Mean Difference (IV, Fixed, 95% CI)	0.0 [0.0, 0.0]
5.3 Individual supervision versus no supervision (with difference in PFMT)	0	0	Mean Difference (IV, Fixed, 95% CI)	0.0 [0.0, 0.0]
6 Frequency of leakage - other measures			Other data	No numeric data
6.1 Additional group supervision (with no difference in PFMT)			Other data	No numeric data
6.2 Additional phone calls (with no difference in PFMT)			Other data	No numeric data
6.3 Individual supervision versus no supervision (with difference in PFMT)			Other data	No numeric data
7 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
7.1 Additional group supervision (with no difference in PFMT)			Other data	No numeric data
7.2 Additional phone calls (with no difference in PFMT)			Other data	No numeric data
7.3 Individual versus no supervision (with difference in PFMT)			Other data	No numeric data
8 Amount of leakage - other measures			Other data	No numeric data
8.1 Additional group supervision (with no difference in PFMT)			Other data	No numeric data
8.2 Additional phone calls (with no difference in PFMT)			Other data	No numeric data
8.3 Individualised versus no supervision (with difference in PFMT)			Other data	No numeric data
9 Voiding frequency			Other data	No numeric data
9.1 Additional group supervision (with no difference in PFMT)			Other data	No numeric data
9.2 Additional phone calls (with no difference in PFMT)			Other data	No numeric data
9.3 Individual versus no supervision (with difference in PFMT)			Other data	No numeric data
10 PFM performance			Other data	No numeric data
10.1 Additional group supervision (with no difference in PFMT)			Other data	No numeric data

10.2 More individual contact versus less individual contact (with no difference in PFMT)		Other data	No numeric data
10.3 Individual versus no supervision (with difference in PFMT)		Other data	No numeric data
11 Treatment adherence		Other data	No numeric data
11.1 Additional group supervision (with no difference in PFMT)		Other data	No numeric data
11.2 Additional phone calls (with no difference in PFMT)		Other data	No numeric data
11.3 Individual versus no supervision (with difference in PFMT)		Other data	No numeric data
12 Follow-up data		Other data	No numeric data
12.1 Additional group supervision (with no difference in PFMT)		Other data	No numeric data
12.2 Additional phone calls (with no difference in PFMT)		Other data	No numeric data
12.3 Individual versus no supervision (with difference in PFMT)		Other data	No numeric data

Comparison 2. Group versus individual supervision of PFMT

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not cured	2		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
1.1 Individual supervision only versus individual and group supervision (no differences in PFMT)	2	111	Risk Ratio (M-H, Fixed, 95% CI)	0.89 [0.78, 1.03]
1.2 Individual supervision only versus group supervision only (with difference in PFMT)	0	0	Risk Ratio (M-H, Fixed, 95% CI)	0.0 [0.0, 0.0]
2 Patients' perception of change in incontinence - not improved	4		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
2.1 Individual and group supervision only versus individual supervision (no difference in PFMT)	3	133	Risk Ratio (M-H, Fixed, 95% CI)	0.16 [0.05, 0.46]
2.2 Group supervision only versus individual supervision only (with difference in PFMT)	1	60	Risk Ratio (M-H, Fixed, 95% CI)	1.2 [0.61, 2.34]

3 Incontinence-specific quality of life			Other data	No numeric data
3.1 Individual supervision only versus individual and group supervision (no difference in PFMT)			Other data	No numeric data
3.2 Individual supervision only versus group supervision only (with difference in PFMT)			Other data	No numeric data
4 Symptom impact			Other data	No numeric data
4.1 Individual supervision only versus individual and group supervision (no difference in PFMT)			Other data	No numeric data
4.2 Individual supervision only versus group supervision only (with difference in PFMT)			Other data	No numeric data
5 Frequency of leakage - leakage episodes in 24 hours	2		Mean Difference (IV, Fixed, 95% CI)	Subtotals only
5.1 Individual supervision only versus individual and group supervision (no difference in PFMT)	1	22	Mean Difference (IV, Fixed, 95% CI)	-1.38 [-2.04, -0.72]
5.2 Individual supervision only versus group supervision only (with difference in PFMT)	1	60	Mean Difference (IV, Fixed, 95% CI)	0.1 [-0.16, 0.36]
6 Frequency of leakage - other measures			Other data	No numeric data
6.1 Individual supervision only versus individual and group supervision (no difference in PFMT)			Other data	No numeric data
6.2 Individual supervision only versus group supervision only (with difference in PFMT)			Other data	No numeric data
7 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
7.1 Individual supervision only versus individual and group supervision (no difference in PFMT)			Other data	No numeric data
7.2 Individual supervision only versus group supervision only (with difference in PFMT)			Other data	No numeric data
8 Amount of leakage - other measures			Other data	No numeric data
8.1 Individual supervision only versus individual and group supervision (no difference in PFMT)			Other data	No numeric data

8.2 Individual supervision only versus group supervision only (with difference in PFMT)	Other data	No numeric data
9 Voiding frequency	Other data	No numeric data
9.1 Individual supervision only versus individual and group supervision (no difference in PFMT)	Other data	No numeric data
9.2 Individual supervision only versus group supervision only (with difference in PFMT)	Other data	No numeric data
10 PFM performance	Other data	No numeric data
10.1 Individual supervision only versus individual and group supervision (no difference in PFMT)	Other data	No numeric data
10.2 Individual supervision only versus group supervision only (with difference in PFMT)	Other data	No numeric data
11 Treatment adherence	Other data	No numeric data
11.1 Individual supervision only versus individual and group supervision (no difference in PFMT)	Other data	No numeric data
11.2 Individual supervision only versus group supervision only (with difference in PFMT)	Other data	No numeric data
12 Follow-up data	Other data	No numeric data
12.1 More group contact versus less individual contact	Other data	No numeric data
12.2 More individual contact versus less individual contact	Other data	No numeric data

Comparison 3. Direct versus indirect methods of PFMT

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not cured	1		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
1.1 PFMT versus sham or imitation PFMT	0	0	Risk Ratio (M-H, Fixed, 95% CI)	0.0 [0.0, 0.0]
1.2 PFMT versus 'Paula method'	0	0	Risk Ratio (M-H, Fixed, 95% CI)	0.0 [0.0, 0.0]
1.3 PFMT versus 'Sapsford' approach	1	64	Risk Ratio (M-H, Fixed, 95% CI)	1.16 [0.98, 1.36]
1.4 PFMT versus Pilates	0	0	Risk Ratio (M-H, Fixed, 95% CI)	0.0 [0.0, 0.0]
2 Patients' perception of change in incontinence - not improved	3		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only

2.1 PFMT versus sham or imitation PFMT	2	138	Risk Ratio (M-H, Fixed, 95% CI)	0.69 [0.47, 1.02]
2.2 PFMT versus 'Paula method'	0	0	Risk Ratio (M-H, Fixed, 95% CI)	0.0 [0.0, 0.0]
2.3 PFMT versus 'Sapsford' approach	1	64	Risk Ratio (M-H, Fixed, 95% CI)	10.33 [1.42, 75.41]
2.4 PFMT versus Pilates	0	0	Risk Ratio (M-H, Fixed, 95% CI)	0.0 [0.0, 0.0]
3 Incontinence specific quality of life			Other data	No numeric data
3.1 PFMT versus sham or imitation PFMT			Other data	No numeric data
3.2 PFMT versus 'Paula method'			Other data	No numeric data
3.3 PFMT versus 'Sapsford' approach			Other data	No numeric data
3.4 PFMT versus Pilates			Other data	No numeric data
4 Symptom impact			Other data	No numeric data
4.1 PFMT versus sham or imitation PFMT			Other data	No numeric data
4.2 PFMT versus 'Paula' method			Other data	No numeric data
4.3 PFMT versus 'Sapsford' approach			Other data	No numeric data
4.4 PFMT versus Pilates			Other data	No numeric data
5 Frequency of leakage - other measures			Other data	No numeric data
5.1 PFMT versus sham or imitation PFMT			Other data	No numeric data
5.2 PFMT versus 'Paula method'			Other data	No numeric data
5.3 PFMT versus 'Sapsford' approach			Other data	No numeric data
5.4 PFMT versus Pilates			Other data	No numeric data
6 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
6.1 PFMT versus sham or imitation PFMT			Other data	No numeric data
6.2 PFMT versus 'Paula method'			Other data	No numeric data
6.3 PFMT versus 'Sapsford' approach			Other data	No numeric data
6.4 PFMT versus Pilates			Other data	No numeric data
7 Amount of leakage - other measures			Other data	No numeric data
7.1 PFMT versus sham or imitation PFMT			Other data	No numeric data
7.2 PFMT versus 'Paula method'			Other data	No numeric data
7.3 PFMT versus 'Sapsford' approach			Other data	No numeric data
7.4 PFMT versus Pilates			Other data	No numeric data

8 Voiding frequency	Other data	No numeric data
8.1 PFMT versus sham or imitation PFMT	Other data	No numeric data
8.2 PFMT versus 'Paula method'	Other data	No numeric data
8.3 PFMT versus 'Sapsford' approach	Other data	No numeric data
8.4 PFMT versus Pilates	Other data	No numeric data
9 PFM performance	Other data	No numeric data
9.1 PFMT versus sham or imitation PFMT	Other data	No numeric data
9.2 PFMT versus 'Paula method'	Other data	No numeric data
9.3 PFMT versus 'Sapsford' approach	Other data	No numeric data
9.4 PFMT versus Pilates	Other data	No numeric data
10 Treatment adherence	Other data	No numeric data
10.1 PFMT versus sham or imitation PFMT	Other data	No numeric data
10.2 PFMT versus 'Paula method'	Other data	No numeric data
10.3 PFMT versus 'Sapsford' approach	Other data	No numeric data
10.4 PFMT versus Pilates	Other data	No numeric data

Comparison 4. Individualised versus generic PFMT

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not improved	1	60	Risk Ratio (M-H, Fixed, 95% CI)	0.83 [0.43, 1.63]
2 Incontinence specific quality of life			Other data	No numeric data
3 Frequency of leakage - leakage episodes in 24 hours	1	60	Mean Difference (IV, Fixed, 95% CI)	-0.1 [-0.36, 0.16]
4 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
5 PFM performance			Other data	No numeric data
6 Treatment adherence			Other data	No numeric data

Comparison 5. Near maximal versus submaximal contractions

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Frequency of leakage - leakage episodes in 24 hours	1		Mean Difference (IV, Fixed, 95% CI)	Subtotals only
2 Frequency of leakage - other measures			Other data	No numeric data
3 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
4 PFM performance			Other data	No numeric data

Comparison 6. Daily versus three times per week PFMT

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not cured	1		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
2 Patients' perception of change in incontinence - not improved	1		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
3 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
4 PFM performance			Other data	No numeric data

Comparison 7. Upright and supine versus supine exercise positions alone

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Incontinence-specific quality of life			Other data	No numeric data
2 Frequency of leakage - leakage episodes in 24 hours	1		Mean Difference (IV, Fixed, 95% CI)	Subtotals only
3 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
4 PFM performance			Other data	No numeric data
5 Treatment adherence			Other data	No numeric data

Comparison 8. Strength and motor learning versus motor learning PFMT alone

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not cured	1	123	Risk Ratio (M-H, Fixed, 95% CI)	1.05 [0.98, 1.13]
2 Patients' perception of change in incontinence - not improved	1	123	Risk Ratio (M-H, Fixed, 95% CI)	0.65 [0.31, 1.40]
3 Incontinence-specific quality of life			Other data	No numeric data
4 Frequency of leakage - leakage episodes in 24 hours	1	97	Mean Difference (IV, Fixed, 95% CI)	-0.20 [-0.55, 0.15]
5 Amount of leakage - pad, paper towel test and cough tests			Other data	No numeric data
6 Amount of leakage - other measures			Other data	No numeric data
7 Voiding frequency			Other data	No numeric data

Comparison 9. PFMT and abdominal muscle exercise versus PFMT alone

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not cured	1		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
2 Patients' perception of change in incontinence - not improved	1		Risk Ratio (M-H, Fixed, 95% CI)	Subtotals only
3 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
4 PFM performance			Other data	No numeric data

Comparison 10. PFMT with intravaginal resistance device versus PFMT alone

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not cured	2	120	Risk Ratio (M-H, Fixed, 95% CI)	1.07 [0.96, 1.20]
2 Patients' perception of change in incontinence - not improved	2	120	Risk Ratio (M-H, Fixed, 95% CI)	0.86 [0.62, 1.20]
3 Symptom impact			Other data	No numeric data
4 Frequency of leakage - leakage episodes in 24 hours	1	80	Mean Difference (IV, Fixed, 95% CI)	0.20 [-0.13, 0.53]
5 Frequency of leakage - other measures			Other data	No numeric data

6 Amount of leakage - pad, paper towel and cough tests		Other data	No numeric data
7 Amount of leakage - other measures		Other data	No numeric data
8 PFM performance		Other data	No numeric data

Comparison 11. PFMT and adherence strategy versus PFMT alone

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not improved	1	41	Risk Ratio (M-H, Fixed, 95% CI)	0.56 [0.34, 0.91]
2 Frequency of leakage - leakage episodes in 24 hours	1	41	Mean Difference (IV, Fixed, 95% CI)	-0.50 [-1.55, 0.55]
3 Amount of leakage - pad, paper towel and cough tests			Other data	No numeric data
4 Amount of leakage - other measures			Other data	No numeric data
5 Treatment adherence			Other data	No numeric data

Comparison 12. 'More intensive' versus 'less intensive' PFMT programmes

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Patients' perception of change in incontinence - not cured	8		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
1.1 'High' contrast	3	175	Risk Ratio (M-H, Random, 95% CI)	0.89 [0.80, 0.98]
1.2 'Moderate' contrast	0	0	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
1.3 'Low' contrast	5	304	Risk Ratio (M-H, Random, 95% CI)	1.06 [1.00, 1.13]
2 Patients' perception of change in incontinence - not improved	14		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
2.1 'High' contrast	6	335	Risk Ratio (M-H, Random, 95% CI)	0.37 [0.17, 0.84]
2.2 'Moderate' contrast	1	44	Risk Ratio (M-H, Random, 95% CI)	0.34 [0.17, 0.71]
2.3 'Low' contrast	7	405	Risk Ratio (M-H, Random, 95% CI)	0.75 [0.59, 0.95]
3 Frequency of leakage - leakage episodes in 24 hours	7		Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1 'High' contrast	1	22	Mean Difference (IV, Random, 95% CI)	-1.38 [-2.04, -0.72]
3.2 'Moderate' contrast	0	0	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3.3 'Low' contrast	6	346	Mean Difference (IV, Random, 95% CI)	-0.03 [-0.19, 0.14]

ADDITIONAL TABLES

Table 1. More versus less contact with health professionals

Study	Group	Intervention	Duration	Supervision	Category
Bø 1990	Individual supervision: home PFMT	Correct VPFMC confirmed. PFMT: 8 to 12 near maximal contractions (with 6 to 8 sec hold and rests) 3 times daily. Monthly clinic visits for perineometer biofeedback of PFM strength	6 months	Physiotherapist monthly	Strength
	Group supervision: weekly exercise class	As above, with addition of weekly 45-min group exercise session which included PFMT, abdominal, gluteal and thigh exercises. The PFMT comprised near maximal contractions for 6 to 8 sec each and 3 to 4 fast contractions, repeated 8 to 12 times, in standing, sitting, lying and kneeling positions	6 months	As above, plus weekly in a group	Strength
Diniz Zanetti 2007	Individual supervision: home PFMT	PFMT: 10 contractions with 5-sec hold and 5-sec rest, 20 contractions of 1-sec hold and 1-sec rest, 5 contractions of 10-sec hold and 10-sec rest, 5 strong contractions with cough, and 1-minute intervals between sets. Monthly clinic visits for assessment only	12 weeks	Physiotherapist monthly	Strength and co-ordination

Table 1. More versus less contact with health professionals (*Continued*)

	Group supervision: twice- weekly supervision	PFMT as above, with 45-min twice-weekly supervision (no clear if individual or group)	12 weeks	Physiotherapist monthly, plus fortnightly in a group?	Strength and co-ordination
Felicissimo 2010	Individual supervision: home PFMT	Correct VPFMC confirmed. PFMT: 10 contractions with 6-sec hold and 12-sec rest in different positions 9 (?) times per day. Start with 90 contractions in first week, then 180 a day for remaining 7 weeks	8 weeks	Physiotherapist at initial session	Endurance
	Group supervision: twice- weekly exercise group	As above, with addition of twice-weekly 50-min group exercise session	8 weeks	As above, plus twice weekly in a group	Endurance
Hung 2010	Less supervision: 'direct' PFMT at home and no supervision	Correct VPFMC confirmed. Oral instruction in PFMT. No other detail given	16 weeks	None	Uncertain
	More supervision: 'indirect' PFMT at home and fortnightly supervision	Correct VPFMC confirmed. 'Indirect' PFMT: weeks 1 to 4 diaphragmatic breathing, weeks 2 to 5 tonic transversus abdominus and PFM activation, weeks 4 to 7 tonic activation with activities of daily living and walking, weeks 6 to 16 muscle strengthening, weeks 8 to 16 functional expiratory patterns, and weeks 10 to 16	16 weeks	Fortnightly with physiotherapist	Indirect

Table 1. More versus less contact with health professionals (*Continued*)

		impact activities. A very full description of the programme is given in the paper by Hung 2010. Participants in this group were "asked not to perform isolated voluntary pelvic floor muscle contraction exercise during the intervention period"			
Konstantinidou 2007	Individual supervision: home PFMT	Correct VPFMC confirmed. PFMT: Individualised programme of 3 sets of fast contractions, 3 to 4 sets of slow contractions daily in lying, standing and sitting positions. Individual follow-up in hospital every 4 weeks	12 weeks	Physiotherapist monthly	Strength
	Group supervision: weekly exercise group	As above, with addition of weekly exercise group	12 weeks	As above, plus weekly in a group	Strength
Ng 2008	Less supervision: PFMT at home and clinic visits	Not clear if correct VPFMC confirmed. Home PFMT progressing to "50 to 75 contractions three times a day". Taught urgency strategies. One-hour clinic visits twice a week for 4 weeks with nurse	6 months?	Nurse twice a week for 4 weeks	Uncertain
	More supervision: PFMT at home, clinic visits and phone calls	As above, then phone calls twice a week from the nurse after cessation of clinic visits to encourage	6 months?	Nurse twice a week for 4 weeks, then twice weekly phone calls	Uncertain

Table 1. More versus less contact with health professionals (*Continued*)

		exercise			

PFM = pelvic floor muscle(s), PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 2. Group versus individual supervision of PFMT

Study	**Group**	**Intervention**	**Duration**	**Supervision**	Category
Bø 1990	Individual supervision: home PFMT	Correct VPFMC confirmed. PFMT: 8 to 12 near maximal contractions (with 6 to 8-sec hold and rests) 3 times daily. Monthly clinic visits for perineometer biofeedback of PFM strength	6 months	Physiotherapist monthly	Strength
	Group supervision: weekly exercise class	As above, with addition of weekly 45-min group exercise session which included PFMT, abdominal, gluteal and thigh exercises. The PFMT comprised near maximal contractions for 6 to 8 sec each and 3 to 4 fast contractions, repeated 8 to 12 times, in standing, sitting, lying and kneeling positions	6 months	As above, plus weekly in a group	Strength
De Oliveira Camargo 2009	Individual supervision	Individualised programme based on PERFECT scheme. 10 slow and 10 fast contractions with 10-sec rest; 10 alternating fast and slow contractions and 5 slow contractions with a cough	12 weeks	Urogynaecology physiotherapist	Strength and co-ordination

Table 2. Group versus individual supervision of PFMT (*Continued*)

	Group supervision	VPFMC confirmed by DVP, in 'orthostatic' position. 10 contractions/5-sec hold/5-sec rest; and 20 contractions/1-sec hold/1-sec rest; and 5x strong contractions with stimulated cough. 1-minute interval between sets	12 weeks	Urogynaecology physiotherapist	Strength and co-ordination
Felicissimo 2010	Individual supervision: home PFMT	Correct VPFMC confirmed. PFMT: 10 contractions with 6-sec hold and 12-sec rest in different positions 9 (?) times per day. Start with 90 contractions in first week, then 180 a day for remaining 7 weeks	8 weeks	Physiotherapist at initial session	Endurance
	Group supervision: twice-weekly exercise group	As above, with addition of twice-weekly 50-min group exercise session	8 weeks	As above, plus twice-weekly in a group	Endurance
Konstantinidou 2007	Individual supervision: home PFMT	Correct VPFMC confirmed. PFMT: individualised programme of 3 sets of fast contractions, 3 to 4 sets of slow contractions daily in lying, standing and sitting positions. Individual follow-up in hospital every 4 weeks	12 weeks	Physiotherapist monthly	Strength
	Group supervision: weekly exercise group	As above, with addition of weekly exercise group	12 weeks	As above, plus weekly in a group	Strength

Table 2. Group versus individual supervision of PFMT (*Continued*)

Liebergall-Wischnitzer 2007	Control: Direct PFMT	Weekly 30-minute lesson for 4 weeks in groups of 5. Encouraged to practise at home 15 min/day. Fortnightly phone call from physiotherapist	12 weeks	Physiotherapist	Uncertain
	Experimental: Indirect PFMT - 'Paula' method	Taught 'Paula' method of sphincter contraction: programme of exercises performed in lying, sitting or standing positions. Exercises were contraction of circular muscles including the pubococcygeal muscle, the anal sphincter, eye and eyelid, mouth, and grip. Programme focusses on strengthening the circular muscles of the body, based on the theory that all sphincters in the body work simultaneously and thus exercising circular muscles in one area of the body will result in strengthening of other sphincters. Weekly individual 45-min training for 12 weeks. Encouraged to practise daily 15 to 45 min	12 weeks	Therapist certified in the 'Paula' method	Indirect
Liebergall-Wirschnitzer 2009	Control: PFMT individual	Unclear if VPFMC was confirmed. Weekly individual sessions of 45 minutes and daily home exer-	12 weeks	3 registered instructors	Indirect

Table 2. Group versus individual supervision of PFMT (*Continued*)

		cises for 45 minutes for 12 weeks. Paula method was taught; first 2 steps were 'rhythmically' contracting the PFM 'with gradual intensity'. Last 3 steps involved contraction and relaxation of eyelids, movement of the mouth and fingers. Rationale based on the idea that all sphincters in the body work together and can affect one another			
	Experimental: PFMT group	VPFMC confirmed, based on observation. 6 group classes (1 to 10 women) of 30 minutes each. Once weekly for first 4 weeks, 2 more classes in last 2 months. Exercises in different positions. Separate contractions of levator ani and anal sphincter. Prolonged, rapid and gradual contractions. 1 to 2 minutes between exercises	12 weeks	10 physiotherapists	Uncertain

DVP = Digital vaginal palpitation, PERFECT = power or pressure, endurance, repetitions, fast contractions, every contraction timed, PFM = pelvic floor muscle(s), PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 3. Direct versus indirect methods of PFMT

Study	Group	Intervention	Duration	Supervision	Category
Ghoniem 2005	Direct PFMT	Correct VPFMC confirmed. PFMT: 3 sets of 10 long (6 to 8-sec hold) and 2 sets of 10 rapid (1 to 2-sec hold) contractions 4 days a week, for a total of 200 contractions weekly. Also 'The Knack'. Four clinic visits	12 weeks	Physiotherapist.	Endurance and co-ordination
	Indirect PFMT: "imitation" PFMT	Correct VPFMC not confirmed. Programme of exercises above with abduction of the hips with legs crossed at the ankles (and knees and hips flexed while sitting or supine)	12 weeks	Physiotherapist	Indirect
Hung 2010	Direct PFMT	Correct VPFMC confirmed. Oral instruction in PFMT. No other detail given	16 weeks	None	Uncertain
	Indirect PFMT: 'Sapsford' approach	Correct VPFMC confirmed. 'Indirect' PFMT: weeks 1 to 4 diaphragmatic breathing, weeks 2 to 5 tonic transversus abdominus and PFM activation, weeks 4 to 7 tonic activation with activities of daily living and walking, weeks 6 to 16 muscle strengthening, weeks 8 to 16 functional expiratory patterns, and	16 weeks	Fortnightly with physiotherapist.	Indirect

Table 3. Direct versus indirect methods of PFMT (*Continued*)

		weeks 10 to 16 impact activities. A very full description of the programme is given in the paper by Hung 2010. Participants in this group were "asked not to perform isolated voluntary pelvic floor muscle contraction exercise during the intervention period"			
Liebergall-Wischnitzer 2007	Direct PFMT	Weekly 30-minute lesson for 4 weeks in groups of 5. Encouraged to practise at home 15 min/day. Fortnightly phone call from physiotherapist	12 weeks	Physiotherapist	Uncertain
	Indirect PFMT: 'Paula method'	Taught 'Paula' method of sphincter contraction: programme of exercises performed in lying, sitting or standing positions. Exercises were contraction of circular muscles including the pubococcygeal muscle, the anal sphincter, eye and eyelid, mouth, and grip. Programme focusses on strengthening the circular muscles of the body, based on the theory that all sphincters in the body work simultaneously and thus exercising circular muscles in one	12 weeks	Therapist certified in the 'Paula' method	Indirect

Table 3. Direct versus indirect methods of PFMT (*Continued*)

		area of the body will result in strengthening of other sphincters. Weekly individual 45-min training for 12 weeks. Encouraged to practice daily 15 to 45 min			
Liebergall-Wirschnitzer 2009	Direct PFMT	Unclear if VPFMC was confirmed. Weekly individual sessions of 45 minutes and daily home exercises for 45 minutes for 12 weeks. Paula method was taught; first 2 steps were 'rhythmically' contracting the PFM 'with gradual intensity'. Last 3 steps involved contraction and relaxation of eyelids, movement of the mouth and fingers. Rationale based on the idea that all sphincters in the body work together and can affect one another	12 weeks	3 registered instructors	Indirect
	Indirect PFMT: 'Paula method'	VPFMC confirmed, based on observation. 6 group classes (1 to 10 women) of 30 minutes each. Once weekly for first 4 weeks, 2 more classes in last 2 months. Exercises in different positions. Separate contractions of levator	12 weeks	10 physiotherapists	Uncertain

Table 3. Direct versus indirect methods of PFMT (*Continued*)

		ani and anal sphincter. Prolonged, rapid and gradual contractions. 1 to 2 minutes between exercises			
Ramsay 1990	Direct PFMT	4 maximum isometric contractions, 4-sec hold, 10-sec rest. Repeated once every waking hour, daily	12 weeks	Not clear	Endurance
	Indirect PFMT: "placebo" PFMT	4 maximum hip abductor contractions (with feet crossed at ankles), otherwise as above	12 weeks	Not clear	Indirect
Savage 2005	Direct PFMT	VPFMC confirmed. Individualised PFMT according to PERFECT scheme including maximal contractions with 1 to 2-sec hold, submaximal contractions, and "staged' contractions (slowly tighten to maximum and slow release). Also 'The Knack. Encouraged to practise several times a day. 6 clinic visits	12 weeks	Physiotherapist	Strength and co-ordination
	Indirect PFMT: Pilates	Not clear if VPFMC confirmed in this group. Stage 1: activate and control deep abdominal muscle and PFM co-contraction. Stage 2: as stage 1 in anti-gravity positions. Stage 3: added limb movement for	12 weeks	Physiotherapist	Indirect

Table 3. Direct versus indirect methods of PFMT (*Continued*)

		low muscle loading consistent with "Pilates principles of concentration, centring, breathing, isolation, routine, precision, control and flowing movement". Encouraged to do 10 to 15 min at home every other day. Not encouraged to do isolated PFM contractions. 6 clinic visits			

PFM = pelvic floor muscle(s), PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 4. Individualised versus generic PFMT

Study	Group	Intervention	Duration	Supervision	Category
De Oliveira Camargo 2009	Individual supervision	Individualised programme based on PERFECT scheme. 10 slow and 10 fast contractions with 10-sec rest; 10 alternating fast and slow contractions and 5 slow contractions with a cough	12 weeks	Urogynaecology physiotherapist	Strength and co-ordination
	Group supervision	VPFMC confirmed by DVP, in 'orthostatic' position. 10 contractions/5-sec hold/5-sec rest; and 20 contractions/1-sec hold/1-sec rest; and 5x strong contractions with stimulated cough. 1 minute interval between sets	12 weeks	Urogynaecology physiotherapist	Strength and co-ordination

DVP= Digital vaginal palpitation,PERFECT = power or pressure, endurance, repetitions, fast contractions, every contraction timed, VPFMC = voluntary pelvic floor muscle contraction

Table 5. Near maximal versus submaximal contractions

Study	Group	Intervention	Duration	Supervision	Category
Johnson 2001	Near maximal PFMT	Correct VPFMC confirmed. VPFMC at 90% maximal force, 15 min, 3 times daily. Given home biofeedback device and exercise diary. Two clinic visits	6 weeks	"The Investigator"	Strength
	Submaximal PFMT	As above except VPFMC at 60% maximal strength	6 weeks	"The Investigator"	Endurance

PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 6. Daily versus three times per week PFMT

Study	Group	Intervention	Duration	Supervision	Category
Sriboonreung (b)	Daily PFMT	Correct VPFMC confirmed. Eight to 12 maximal contractions with 6 to 8-sec hold with 6 to 8 fast contractions, with 6 to 8-sec rest, daily, 3 times a day	12 weeks	Physiotherapist	Strength
	3 times weekly PFMT	As above except 3 sets a day, 3 days a week.	12 weeks	Physiotherapist	Strength

PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 7. Upright and supine versus supine exercise positions alone

Study	Group	Intervention	Duration	Supervision	Category
Borello-France 2006	Upright and supine PFMT	Correct VPFMC confirmed. Began training with maximum of 2 sets of 10 repetitions of both 3-sec maximal contraction and 12-sec contraction twice-daily, progressing at discretion of therapist. Exercise alternated between supine, sitting and standing. Also 'The Knack'. 12 clinic visits	9 to 12 weeks	Physiotherapist	Strength and co-ordination
	Supine PFMT	As above, but exercise only in supine	9 to 12 weeks	Physiotherapist	Strength and co-ordination

PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 8. Strength and motor learning versus motor learning PFMT alone

Study	Group	Intervention	Duration	Supervision	Category
Hay-Smith 2002	Strengthening and motor learning PFMT	Correct VPFMC confirmed. Motor learning PFMT: VPFMC in variety of body positions (squat, sit, stand, kneel and so on) and movements (reach, sit to stand, walk and so on). Bracing and hold (transversus abdominus and PFM) with effort (such as cough, sneeze, lift). Progressed from discrete, stable, close tasks to continuous,	20 weeks	Physiotherapist	Strength and co-ordination

Table 8. Strength and motor learning versus motor learning PFMT alone (*Continued*)

		open, mobile tasks. Strengthening PFMT: individualised progressing to 12 maximal effort contractions with 8 sec hold and 8 sec rest, 3 times a day, daily. Three clinic visits and 3 phone calls			
	Motor learning PFMT	As above, except motor learning programme only	20 weeks	Physiotherapist	Co-ordination

PFM = pelvic floor muscle(s), PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 9. PFMT and abdominal muscle exercise versus PFMT alone

Study	Group	Intervention	Duration	Supervision	Category
Sriboonreung (a)	3 times weekly PFMT and abdominal muscle training	Presumed as below except addition of abdominal (specifically transversus abdominus and internal oblique muscle) training. No further detail given	12 weeks	Physiotherapist	Strength
	3 times weekly PFMT	Correct VPFMC confirmed. Eight to 12 maximal contractions with 6 to 8-sec hold with 6 to 8 fast contractions, with 6 to 8-sec rest, 3 times a day, 3 days a week	12 weeks	Physiotherapist	Strength

PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 10. PFMT with intravaginal resistance device versus PFMT alone

Study	Group	Intervention	Duration	Supervision	Category
Delgado 2010	PFMT with intravaginal resistance device	Correct VPFMC confirmed. PFMT: 5 quick and 5 slow (sustained), high-intensity contractions daily. Advised to hold contractions as long as possible, relaxing their PFM for an equivalent time before repeating the process. Intravaginal resistance: instructions to use the Pelvic-Toner Device concurrently whilst exercising. Two clinic visits and one phone call	16 weeks	Urology research nurse	Strength
	PFMT	As above without device	16 weeks	Urology research nurse	Strength
Ferguson 1990	PFMT with intravaginal resistance device	PFMT: exercises at home for strength and endurance, using audio-tape to guide exercises at home. Intravaginal resistance: use of intravaginal balloon. Weekly phone call	6 weeks	?None	Strength and endurance
	PFMT	As above without device	6 weeks	?None	Strength and endurance
Wells	PFMT with intravaginal resistance	Correct VPFMC confirmed. PFMT: minimum of 80 VPFMC with 10-sec hold and 10-sec rest per day distributed in individual pattern throughout the day. Intravaginal resistance: Fit-	5 months	Nurse practitioner	Endurance

Table 10. PFMT with intravaginal resistance device versus PFMT alone (*Continued*)

		ted with vaginal dilator to use as resistive device. Monthly clinic visits			
	PFMT	As above without device			Endurance

PFM = pelvic floor muscle(s), PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

Table 11. PFMT and adherence strategy versus PFMT alone

Study	Group	Intervention	Duration	Supervision	Category
Gallo 1997	PFMT	Correct VPFMC confirmed. Adherence strategy: audiocassette tape for use twice a day (contained verbal instruction and counted aloud 25 consecutive PFM contractions, with 10-sec hold and 10-sec relaxation)	4 to 6 weeks	Nurse	Strength
	PFMT with adherence strategy	Encouraged to exercise 10 minutes twice a day, potential times suggested depending on lifestyle	4 to 6 weeks	Nurse	Uncertain
Sugaya 2003	PFMT	PFMT: 1 minute of rapid contractions every 2 sec, followed by 1 minute of slow 10-sec contractions with 10-sec rests, performed 3 times a day. Adherence strategy: pocket size device with chime or blinking light to indicate time to exercise (3 times a day) and when activated leads person through PFMT programme,	8 weeks	"Clinician"	Strength

Table 11. PFMT and adherence strategy versus PFMT alone (*Continued*)

		and kept record of exercise. Fortnightly follow-up			
	PFMT with adherence strategy	As above, without device	8 weeks	"Clinician"	Strength

PFMT = pelvic floor muscle training, VPFMC = voluntary pelvic floor muscle contraction

HISTORY

Review first published: Issue 12, 2011

CONTRIBUTIONS OF AUTHORS

JHS led the review process, and wrote the first drafts of the protocol and full review. RH and JHS (except where a third person was needed due to JHS's involvement in a study) screened the studies and checked eligibility. RH took the lead role in data extraction and data entry (which JHS cross-checked). CD and PH commented on all drafts of the review, and PH gave statistical and methodological advice as needed.

DECLARATIONS OF INTEREST

Jean Hay-Smith and Chantale Dumoulin were first authors of trials considered for inclusion in the review (Dumoulin 2003; Hay-Smith 2002).

SOURCES OF SUPPORT

Internal sources

- University of Otago, New Zealand.

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

The protocol included PFMT with or without adjunctive biofeedback. There were so many studies of biofeedback or feedback these became a separate review (see Herderschee 2011). We used a different subgroup analysis than planned (see Subgroup analysis and investigation of heterogeneity and Effects of interventions).

INDEX TERMS

Medical Subject Headings (MeSH)

Directly Observed Therapy [methods]; Exercise Therapy [*methods]; Pelvic Floor [*physiology]; Randomized Controlled Trials as Topic; Resistance Training [methods]; Time Factors; Urinary Incontinence [*rehabilitation]

MeSH check words

Female; Humans