PCORI Evidence Map and Visualization: Pelvic Floor Muscle Training for Urinary Incontinence in Women

Prepared by:

ECRI Institute-Penn Medicine Evidence-based Practice Center
5200 Butler Pike
Plymouth Meeting, PA 19462

April 17, 2020

Suggested citation:
Contents

Background ........................................................................................................................................... 1

Methods .............................................................................................................................................. 2

Data Sources...................................................................................................................................... 2

Literature Search ............................................................................................................................... 2

Inclusion Criteria ............................................................................................................................... 3

Risk-of-Bias Assessment .................................................................................................................. 4

Data Extraction and Effect Size Computation .................................................................................. 5

Meta-Analysis .................................................................................................................................... 6

Minimal Important Differences ......................................................................................................... 7

Strength of Evidence (SOE) Ratings ............................................................................................... 13

Research Volume Map ...................................................................................................................... 14

Treatment Effects Map .................................................................................................................... 16

Results .............................................................................................................................................. 18

Evidence Base ................................................................................................................................. 18

Selection of Comparisons for the Visualization .............................................................................. 21

Research Volume Map: Observations .............................................................................................. 21

Treatment Effects Map, Effect of PFMT: Observations ................................................................... 22

Treatment Effects Map, Different Methods of PFMT: Observations ............................................. 23

Treatment Effects Map, Add-ons to PFMT: Observations ............................................................... 24

Treatment Effects Map, Alternatives to PFMT: Observations ......................................................... 25

Adverse Effects ............................................................................................................................... 25

Limitations ....................................................................................................................................... 28

Conclusions, Implications, and Future Research ........................................................................... 29

Acknowledgements ......................................................................................................................... 30

References ....................................................................................................................................... 31

Appendices ....................................................................................................................................... 39

Appendix A. Literature Search Methods ........................................................................................... 39

Appendix B: Evidence Tables ............................................................................................................. 41
Figures

Figure 1. Article Flow ............................................................................................................. 20

Tables

Table 1. Use of the MID for Interpreting Results: Examples.................................................. 13
Table 2. Details of Included Studies ......................................................................................... 41
Table 3. Clinicaltrials.gov Records of Studies Not Yet Published ............................................. 47
**Background**

Pelvic floor muscle training (PFMT) is recommended as a first line of treatment for adult women with urinary incontinence (UI) by several professional societies, including the American College of Physicians,\(^1\) the National Institute of Health and Care Excellence,\(^2\) and the European Association of Urology.\(^3\) PFMT’s popularity is likely due to numerous factors such as noninvasiveness, acceptability to women, and effectiveness.

While there is widespread agreement on PFMT’s general efficacy, numerous questions remain about specific PFMT implementations. For many women, it can be challenging to identify the correct pelvic floor muscles used for training, so individualized education may be necessary. The frequency and duration of this education can vary widely, and may also involve the advice and feedback of other patients (e.g., in a group setting). Other physical exercises, such as yoga or exercises that indirectly strengthen the pelvic floor (e.g., the “Paula” method), may be recommended for use in conjunction with PFMT. Investigators have also used supplemental devices to optimize PFMT, such as biofeedback, electrical stimulation, magnetic stimulation, pelvic toner/pelvic trainer devices, and weighted cones/spheres. These devices purportedly give women clearer feedback as to whether they are exercising the correct muscles in the correct manner.

Non-PFMT behavioral strategies are sometimes recommended as first-line treatments, either alone or in conjunction with PFMT. These include anatomy education, bladder training (i.e., increasing void intervals by delaying voiding), dietary modification, smoking cessation, fluid restriction, weight loss, and psychological techniques (e.g., cognitive behavioral therapy). As these strategies are also non-invasive and generally acceptable to women, they represent an important category for comparison to PFMT.

The type of UI (i.e., stress, urgency, or mixed) may influence the recommended first-line treatment. A 2014 American College of Physicians guideline recommended PFMT for women with stress UI, but bladder training for women with urgency UI, and both PFMT and bladder training for women with mixed UI.\(^1\) These recommendations reflect the different underlying pathologies (e.g., weak muscles causing stress UI, vs learned habits and/or bladder overactivity causing urgency UI). Thus, any comparisons of first-line treatments must necessarily consider the type of UI.

Given the complexity of the options surrounding PFMT, women and their caregivers need clear summaries of which options work, how well, and for whom. A 2018 AHRQ report comprehensively evaluated dozens of non-surgical treatments for UI,\(^4\) but did not compare different PFMT implementations or behavioral strategies. At the request of the Patient-Centered Outcomes Research Institute (PCORI), ECRI Institute analyzed data for a series of
Evidence Maps to summarize the evidence on use of PFMT options in patients with UI, and to present our findings in a visually appealing, easy-to-understand format.

Methods

Data Sources

The evidence maps used 6 sources:

- The 2012 AHRQ report\(^5\) entitled *Nonsurgical Treatments for Urinary Incontinence in Adult Women: Diagnosis and Comparative Effectiveness*
- The 2018 AHRQ report\(^4\) entitled *Nonsurgical Treatments for Urinary Incontinence in Women: A Systematic Review Update*
- The 2018 Cochrane report\(^6\) entitled *Pelvic floor muscle training for urinary incontinence in women*
- Our literature search for more recent trials (see more details in the “Literature Search” section below”)
- The PCORI website for pertinent PCORI-funded trials
- Clinicaltrials.gov for other pertinent trials

Literature Search

To identify more recently published trials, an experienced medical information specialist searched PubMed, EMBASE/Medline, and PsycINFO databases, limiting to publications since 12/4/2017, the latest search date for the 2018 AHRQ report on UI. We present the strategies in Appendix A. We translated controlled vocabulary terms and syntax for the PubMed and PsycINFO searches. To identify additional trials in this topic area, we searched the PCORI website and ClinicalTrials.gov. We conducted searches through 7/5/2019.
Inclusion Criteria

Our inclusion criteria were based on the 2012 and 2018 AHRQ reports’ criteria as well as additional data considerations. Specifically, we required that studies:

- Were published in English as a full article
- Were randomized controlled trials (RCTs) reporting at least one of these comparisons:
  - Comparing PFMT (any variant) to inactive treatment (e.g., usual care, waitlist) or adding PFMT to background treatment
  - Comparing two or more variants of PFMT (e.g., group vs individual PFMT)
  - Examining the effect of adding additional intervention(s) (e.g., biofeedback, electrical stimulation, home pelvic floor toner/trainer device) to PFMT
  - Comparing PFMT (any variant) to a non-PFMT behavioral treatment (e.g., bladder training, electrical stimulation, exercise).
- Enrolled a patient population for which at least 90% were adult women with urinary incontinence (any type) in non-acute care settings who were not pregnant, did not have neurogenic lower urinary tract dysfunction or cognitive impairment, and did not have surgically-treated UI. We also excluded studies in which 10% or more of the patients received ancillary treatments for UI during the follow-up period (e.g., medications, surgery). These excluded populations may be less likely to respond to PFMT and related interventions.
- The method of administering PFMT did not vary by type of UI (or if they did, results must have been reported separately for different UI types).
- Reported data on either cure (percentage of patients), reduction in UI events, UI-specific quality of life (QOL), or adverse effects. UI-specific QOL had to be measured with one of four instruments: Incontinence Quality of Life (IQOL), International Consultation on Incontinence Questionnaire-Urinary Incontinence Lower Urinary Tract Symptoms Quality of Life (ICIQ LUTSqol), Overactive Bladder questionnaire on health-related quality-of-life subscale (OAB-q HRQL), or the Urogenital Distress Inventory (UDI). These
were the only four instruments for which we could find published data on the minimal important difference (MID).

- Reported data at least 4 weeks after the start of treatment. This was to avoid cluttering the maps with shorter-term studies whose effects may not persist.
- Reported data on at least 10 patients per group representing at least 50% of those enrolled. With smaller patient numbers or >50% dropout, the reported data are likely unrepresentative, and are also more vulnerable to group imbalance.

**Risk-of-Bias Assessment**

For each study, we rated overall risk of bias as Low, Moderate, or High. After consulting the risk-of-bias methods used in the 2012 and 2018 AHRQ reports, we developed our own process. For each study, we examined five items:

1. Adequate generation of a randomized sequence
2. Allocation concealment
3. Blinding of patients
4. Blinding of outcome assessors
5. Completeness of data

The 2012 and 2018 AHRQ reports each used all five of these items. We rated each individual item as Low, High, or Unclear risk of bias.

We examined items 1 and 2 only once for each study. We tailored the other items (3, 4, and 5) to different outcomes, comparisons and follow-up times:

- For items 3 and 4, we recognized that some outcomes (e.g., pad test) are more objective, and therefore are less influenced by knowledge of treatment group, so we assigned Low risk of bias to outcomes based on a pad test.
- Items 3 and 4 were also sensitive to the specific treatment comparison, as patient expectations differentially influence certain treatment comparisons. Specifically, when a comparison involved *PFMT vs usual care*, or *adding PFMT to background treatment*, or *PFMT alone vs its combination with another treatment*, blinding is important, and so we deemed a lack of blinding as High risk of bias. By contrast, when a comparison involved...
different variants of PFMT, or PFMT vs an active comparator such as general physical therapy, patients likely do not have differential expectations of success, so we did not consider lack of blinding a significant limitation.

- For item 5, we examined incomplete data separately for different outcomes and different follow-up times because the number of patients providing data for each outcome may vary.

To assign a global rating to each data point, we used the following rules:

- **Low risk of bias.** No High risk-of-bias items, AND 3 or more Low risk-of-bias items
- **Moderate risk of bias.** No High risk-of-bias items AND 2 or fewer Low risk-of-bias items
- **Moderate risk of bias.** The only High risk-of-bias item(s) involved blinding (items 3 and 4)
- **High risk of bias.** Any other combination of items.

### Data Extraction and Effect Size Computation

We obtained PDFs of the included articles and extracted the following information:

- Year of publication
- Country in which the patients were enrolled
- Average age of enrolled women, along with range and standard deviation, or median and interquartile range
- Number of patients
- Percentage of women who had different types of UI (stress, urgency, mixed)
- Treatments compared
- Length of follow-up (number of weeks since the start of treatment)
- Percentage of patients in each group who experienced “cure,” and how the authors defined it.
- Frequency of urinary leaks (baseline vs follow-up). This could be based on a patient diary, or on patients’ subjective report of frequency on a Likert scale.
- Impact on quality of life (QOL) as measured by one of four instruments (IQOL, ICIQ LUTSqol, OAB-q HRQOL subscale, or UDI). These were the only instruments for which we could find published data on the minimal important difference (MID).
- Adverse effects (AEs)

We did not extract data on specific PFMT delivery method(s) and/or the extent of clinician involvement unless these metrics were part of the treatment comparisons being made in the study (e.g., comparisons between different methods of PFMT such as group vs. individual PFMT
instruction, PFMT with or without verbal feedback given during a digital exam, fully supervised vs partially supervised PFMT).

After data extraction, we calculated effects sizes for cure, reduction in UI events, and QOL. For cure, we calculated the odds ratio, whereas for continuous outcomes (reduction in UI events, and QOL), we calculated Hedges’ g, which is an N-corrected form of the standardized mean difference. In some cases, g was not calculable due to insufficient reporting (e.g., not possible to estimate the standard deviation). In these cases, we still recorded whether the difference between groups was statistically significant. When studies reported multiple follow-up times longer than 4 weeks, we extracted all timepoints that met our inclusion criteria. For the maps, we selected the longest timepoint within a given timeframe (i.e., if a study reported both 4-week data and 11-week data, the maps only display the 11-week data).

Meta-Analysis

We examined all treatment comparisons to determine which comparisons were similar enough to combine in meta-analyses. This was based on approximate similarity of patients, interventions, comparators, outcomes, and timepoints. Final decisions about lumping or splitting were made by the ECRI research team along with the PCORI team. These decisions balanced the need for clinical accuracy with a desire to minimize scrolling and clutter in the treatment effects (TE) map.

When multiple studies reported the same treatment comparison (e.g., PFMT + other exercise vs PFMT alone) and the same outcome category (e.g., cure) at similar follow-up times (e.g., >=12 weeks but <26 weeks), we performed meta-analysis. We analyzed the odds ratio or Hedges’ g using the random-effects method of DerSimonian and Laird. To incorporate the baseline data (where reported), we assumed a pre-post correlation of 0.5. Meta-analysis was performed using STATA (using the command metan). We resolved various meta-analytic dilemmas as follows:

- If a study reported multiple follow-up time points within a given follow-up category, we used the longer follow-up. If the user selected “Any” follow-up, we used the longest follow-up data.
- If a study reported data on multiple types of UI, we used the Stress data when they selected the Stress option, and we used the combined data on other groups when they selected either the Stress/Urge/Mixed option or the Any option (in order to maximize statistical power).
- If a study reported multiple groups receiving PFMT as well as a usual care group, we combined the data for the multiple PFMT groups for the effectiveness of PFMT comparison (in order to maximize statistical power).
For forest plots, we included the I² statistic as a measure of heterogeneity. To determine effect size ranges, we examined all of the cure effect sizes for the effectiveness of PFMT comparison in the TE map and determined that a reasonable range for all of the associated odds ratio outcome forest plots was 0.005 to 200. Similarly, for continuous outcomes (reduction in UI events, and QOL) we used a consistent range of -4 to 4. For the other tabs, we used consistent ranges of 0.02 to 50 for the odds ratio outcome forest plots and -2 to 2 for the continuous outcome forest plots. Each forest plot was exported as a PNG file (for viewability on all platforms) and also an EPS file (for maximal resolution). We created a forest plot even for analyses with only one study, for visual consistency with other evidence bases, but for those we suppressed the “summary” estimate.

We computed meta-analyses for all possible permutations of 3 filters for each tab of the treatment-effects map: the age filter (4 options), the type of UI filter (3 options), and the time-since-start-of-treatment filter (4 options). This meant that we conducted up to 48 meta-analyses (4 x 3 x 4) for each comparison for each outcome. To conduct these meta-analyses efficiently, we used a programming loop in Stata. For more details about the filters, see the sections below on the Research Volume Map and the Treatment Effects Map.

### Minimal Important Differences

To aid interpretation, for each of the effectiveness outcomes (Cure, reduction in UI events, QOL), we defined the minimal important difference (MID). The MID is a numerical threshold below which there is no important impact on patient outcomes. It serves two purposes:

- To help determine whether a **statistically significant difference** is **clinically important**. If the confidence interval (CI) around an effect size (such as an odds ratio) is **fully above the MID**, then clearly the effect is clinically important. If the CI is fully below the MID, then the effect is clearly not clinically important. Or, if the CI straddles the MID, then clinical importance is unclear.

- To help determine whether a **non-statistically significant difference** is small enough to permit a conclusion of **approximate equivalence**. If the CI is fully within one MID of a null effect, then one can conclude approximate equivalence. Otherwise, the data are inconclusive.

The next three sections define our MID’s for the three effectiveness outcomes, and a summary table (}
Table 1) provides examples of how we use the MID’s to draw conclusions.
Cure

Women who have no UI events are considered cured. This does not necessarily mean permanent cure, but many studies report comparative cure rates at an intermediate time interval (e.g., the percentage of women whose 3-day diary indicates 0 incontinence events after 8 weeks of treatment). Given the predefined importance of this outcome, we considered any statistically significant difference to be clinically important.

To interpret non-statistically-significant differences in cure rates, we needed an MID in order to determine whether evidence permits a conclusion of equivalence. For this, we set an equivalence range for the odds ratio from 0.8 to 1.25, which is the same range used by the FDA for judging biologic equivalence. Thus, if the CI for the odds ratio were within the range from 0.8 to 1.25, the evidence could permit a conclusion of approximate equivalence. Otherwise, we judged the evidence for a nonsignificant finding as too imprecise to permit a conclusion. See the top third of
Table 1 for examples.

Reduction in UI events

Women with UI vary greatly in the frequency of UI events prior to treatment. Some randomized studies have reported mean baseline frequencies as low as 1 event per week, and others have reported baselines as high as 55 per week. A typical frequency is about 20 UI events per week (about 3 per day). Some studies define a partial success as a 50% reduction in frequency.10 The 2012 AHRQ report stated that “Importantly, women with daily stress UI perceived important clinical benefit at reductions of approximately 50 percent.”5 A focus group study11 found that 80% of women would consider a 50% reduction to be indicative of a meaningful change. Therefore, we used 50% as the MID for this outcome. See the middle third of
When studies differed in their baseline frequencies, a single MID did not exist. In those cases, we defined the MID as the lowest among the studies of that PICO (i.e., the minimum MID).

QOL

We included four instruments measuring the impact of UI on QOL; these were the only instruments for which we found evidence on the MID. Because we used Hedges’ g to report the impact of UI interventions on QOL, we converted the established MID for each instrument to Hedges’ g by dividing the MID by the average standard deviation (SD) reported in the included studies.

Incontinence Quality of Life (IQOL). The 22-item IQOL instrument has been validated in patients with stress-predominant UI. Scores range from 0 to 100, with higher scores indicating a smaller negative impact of UI on QOL (increased QOL). Yalcin et al. 2006\textsuperscript{12} found that the MID for this instrument is 2.5 points. For our analyses, we used a Hedges’ g of 0.15 (average reported SD for the IQOL was 16).

International Consultation on Incontinence Modular Questionnaire- Lower Urinary Tract Symptoms Quality of Life (ICIQ LUTSqol). The 19-item ICIQ LUTSqol instrument has been validated in patients with stress-predominant UI. Scores range from 19 to 76, with higher scores indicating a larger negative impact of UI on QOL (decreased QOL). Nyström et al. 2015\textsuperscript{13} found that the MID for this instrument is 3.71 points. For our analyses, we used a Hedges’ g of 0.43 (average reported SD for the ICIQ LUTSqol was 43).

Overactive Bladder Questionnaire health-related quality of life subscale (OAB-q HRQL). The 25-item HRQL subscale of the OAB-q instrument has been validated in patients with urge-predominant UI. Scores range from 0 to 100,\textsuperscript{14} with higher scores indicating a smaller negative impact of UI on QOL (increased QOL). Dyer et al. 2011\textsuperscript{15} found that the MID for this instrument is 2.5 points. We did not convert this MID to Hedges’ g, because no included studies used this instrument.

Urogenital Distress Inventory (UDI). The 19-item UDI instrument has been separately validated in patients with stress-predominant UI. Scores range from 0 to 300, with higher scores indicating a larger negative impact of UI on QOL (decreased QOL). For stress-predominant UI, Barber et al. 2009\textsuperscript{16} found that the MID of this instrument is 11 points. For our analyses, we used a Hedges’ g of 0.25 (average reported SD for the UDI was 43).

Two included studies used a 6-item version of the UDI (UDI-6); authors used a 100-point scale range to report UDI-6 scores. Since the 11-point MID of the full UDI instrument is 3.7% of the entire 300-point scale, we imputed that the MID of the UDI-6 should be 3.7 points. We did
not convert this MID to Hedges’ $g$ because of the studies using UDI-6, only one small study\textsuperscript{17} reported SDs.

When studies used different QOL instruments, a single MID did not exist. In those cases, we defined the MID as the lowest among the studies of that PICO (i.e., the minimum MID).
### Table 1. Use of the MID for Interpreting Results: Examples

<table>
<thead>
<tr>
<th>Outcome: Cure</th>
<th>Effect size metric</th>
<th>Point estimate</th>
<th>95% CI</th>
<th>Statistically significant?</th>
<th>MID for this outcome</th>
<th>Possible conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>1.3 to 1.73</td>
<td>Yes</td>
<td>0.8 to 1.25</td>
<td>Clinically important difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>1.1 to 2.04</td>
<td>Yes</td>
<td>0.8 to 1.25</td>
<td>Clinically important difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td>1.01 to 1.20</td>
<td>Yes</td>
<td>0.8 to 1.25</td>
<td>Clinically important difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.9 to 2.5</td>
<td>No</td>
<td>0.8 to 1.25</td>
<td>Evidence is insufficient (too imprecise)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td>0.98 to 1.23</td>
<td>No</td>
<td>0.8 to 1.25</td>
<td>Equivalence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td>0.9 to 1.34</td>
<td>No</td>
<td>0.8 to 1.25</td>
<td>Evidence is insufficient (too imprecise)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome: Reduction in UI events</th>
<th>Effect size metric</th>
<th>Point estimate</th>
<th>95% CI</th>
<th>Statistically significant?</th>
<th>MID for this outcome</th>
<th>Possible conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Between-group difference in change scores</td>
<td>4</td>
<td>+3 to +5</td>
<td>Yes</td>
<td>-2 to +2</td>
<td>Clinically important difference</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>+1 to +7</td>
<td>Yes</td>
<td>-2 to +2</td>
<td>A difference, of unclear clinical importance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>+0.7 to +1.3</td>
<td>Yes</td>
<td>-2 to +2</td>
<td>A difference, but not clinically important</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-1 to +9</td>
<td>No</td>
<td>-2 to +2</td>
<td>Evidence is insufficient (too imprecise)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-0.5 to +1.5</td>
<td>No</td>
<td>-2 to +2</td>
<td>Equivalence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-1 to +3</td>
<td>No</td>
<td>-2 to +2</td>
<td>Evidence is insufficient (too imprecise)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome: QOL, as measured by the IQOL (range 0-100)</th>
<th>Effect size metric</th>
<th>Point estimate</th>
<th>95% CI</th>
<th>Statistically significant?</th>
<th>MID for this outcome</th>
<th>Possible conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Between-group difference in change scores</td>
<td>5</td>
<td>+4 to +6</td>
<td>Yes</td>
<td>-2.5 to +2.5</td>
<td>Clinically important difference</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>+2 to +8</td>
<td>Yes</td>
<td>-2.5 to +2.5</td>
<td>A difference, of unclear clinical importance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>+0.2 to +1.8</td>
<td>Yes</td>
<td>-2.5 to +2.5</td>
<td>A difference, but not clinically important</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-1 to +11</td>
<td>No</td>
<td>-2.5 to +2.5</td>
<td>Evidence is insufficient (too imprecise)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>-1 to +2</td>
<td>No</td>
<td>-2.5 to +2.5</td>
<td>Equivalence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-1 to +3</td>
<td>No</td>
<td>-2.5 to +2.5</td>
<td>Evidence is insufficient (too imprecise)</td>
<td></td>
</tr>
</tbody>
</table>

Note: This table displays hypothetical data for three effectiveness outcomes, the MID for each outcome, and the possible conclusion resulting from considering three components: 1) the 95% CI, 2) statistical significance, and 3) the MID. On the far right, the reason we say “possible” conclusion, and not just “conclusion”, is that other aspects of the evidence (e.g., risk of bias, consistency) could warrant downgrades to the strength of evidence, resulting in a rating of Insufficient for any of these scenarios, which means “no conclusion.” For reduction in UI events as well as QOL, we converted the MID to Hedges’ g by dividing it by the typical standard deviation for that outcome.

### Strength of Evidence (SOE) Ratings

For cure, reduction in UI events, and QOL outcomes, we rated the SOE for each comparison as High, Moderate, Low, or Insufficient. This categorization corresponds to the EPC system, which is similar to GRADE. The lowest rating (Insufficient) indicates that no conclusion can be drawn from the evidence, whereas the other ratings indicate varying levels of confidence.
For a given comparison, we started the rating at High, as only RCTs were included. We then considered four domains to rate the SOE:

- **Risk of bias** (see earlier section for details). Studies rated at Low risk of bias received no downgrade, moderate risk of bias received a single downgrade, and high risk of bias received a double downgrade.

- **Directness.** Our inclusion criteria required that studies enroll adult women who received interventions and reported patient-oriented outcomes. Therefore, we did not downgrade for indirectness.

- **Consistency.** When multiple studies of the same comparison reported on the same outcome, we determined subjectively whether to downgrade for inconsistent results between studies. For a single-study evidence base, we applied a single downgrade due to the lack of replication of findings.

- **Precision.** We considered precision in two ways. First, we determined whether the evidence was precise enough to potentially permit a conclusion. This was always true for statistically significant differences, and true for nonsignificant findings if the CI was within one MID of a null effect (see the Minimal Important Difference section above). Second, we determined whether the confidence bounds were close to negating the finding (i.e., just barely significant, or just barely within the MID). If so, we downgraded for imprecision. Because consistency and prediction are related (e.g., a random-effects meta-analytic CI could be wide due to either inconsistency or imprecision), we were careful not to penalize evidence bases twice for a single underlying concern.

We rated inconclusive findings as Insufficient. We rated all other findings as High, Moderate, or Low depending on the above domains. All SOE assessments were conducted by one analyst and checked by a second.

**Research Volume Map**

The RV map displays, interactively, the amount of research on different treatment comparisons. There were 4 categories of comparisons (Effect of PFMT, Different Methods of PFMT, Add-ons to PFMT, and Alternatives to PFMT). It uses a bar chart to display the number of women with outcome data (vertical axis) for each of the 4 overall categories (horizontal axis). The top of the chart shows the total number of women, the total number of treatment
comparisons, and the number of RCTs. Five interactive filters along the left side of the map (type of UI, age, number of weeks since the start of treatment, publication year, and region of the world) allow the user to redraw the chart according to their specific interests (the filters can be used in any combination). Hovering over a bar provides additional details such as the top 2 comparisons in that category (by number of women), and clicking on a bar yields counts for each subcategory, per-study Ns, and hyperlinks to trial abstracts. The 5 filters are:

- **Type of UI (3 choices: All, Stress, Stress/Mixed/Urge).** See details below for how we defined these choices.

- **Age (4 choices: All, Age mostly younger than 50, Age mostly 45-65, and Age mostly older than 60).** See details below for how we defined these categories.

- **Number of weeks since start of treatment (4 choices: All, <12 weeks, 12-26 weeks, and >=26 weeks).**

- **Year Published (4 choices: All, 1991 to 2000, 2001 to 2010, and 2011 to mid 2019)**

- **Region (5 choices: All, Asia, Brazil, Europe, or US/Canada).** We separated Brazil due to the relatively large number of Brazilian studies.

For the age filter, we first used reported means and SDs to estimate the percentage of women who were younger than 50, as well as the percentage of women who were older than 60 (assuming normal distributions). For studies reporting the median instead of the mean, we used the median as an estimate of the mean. For studies reporting ranges or interquartile ranges instead of SDs, we estimated SDs using the methods proposed by Wan et al.19 We then examined the bivariate distribution of the 2 percentages (p(age)<50 and p(age)>60). This examination resulted in definitions for the age filter:

- **“Age mostly younger than 50”:** more than 50% of women were under 50 AND fewer than 15% were older than 60

- **“Age mostly older than 60”:** more than 50% of women were over 60 AND fewer than 15% were younger than 50

- **“Age mostly 45-65”:** Study fit neither of the two categories above. We included in this category the studies for which age percentages could not be calculated due to unreported SDs and/or unreported means/medians. We examined a plot of the age data for these studies (means, SDs), and all of the means were above the highest mean for the first category above, and below the lowest mean for the second category above.

For type of UI filter, we examined the percentage of patients in each study that, at baseline, had pure stress UI, pure urge UI, or mixed UI. Some studies had not reported these percentages. After examining the data, we decided that the best way to use type-of-UI
information was to create one category for the pure-stress-Ul studies, and another category for all other studies. Further delineating the latter category, we decided, would be suboptimal.

For 2 of the 73 included trials, we only included their data on adverse effects (because they did not report included data on our effectiveness outcomes). We excluded these 2 studies from the RV map. We also excluded from the RV maps the 3 studies that met inclusion criteria that had measured the effect of PFMT by comparing one group that received a non-PFMT treatment only compared to another group that received the same non-PFMT treatment in combination with PFMT.

**Treatment Effects Map**

The TE map displays, interactively, what the accumulated evidence actually indicates about PFMT. Each of the four comparison categories appears in a different tab, so that the user can focus on a desired portion of the TE map. The tabs were as follows:

- Tab 1: Effectiveness of PFMT
- Tab 2: Different Methods of PFMT
- Tab 3: Add-ons to PFMT
- Tab 4: Alternatives to PFMT

We communicated our findings for each of three outcomes (cure, reduction in UI events, QOL) for each treatment comparison using bubbles:

- Our overall conclusion about superiority, equivalence, or insufficient evidence is indicated by the bubble color, as follows:
  - Conclusions of approximate equivalence as white bubbles
  - Insufficient evidence bases as gray bubbles.
  - No evidence as gray asterisks.
  - Tab 1 (effect of PFMT). Green bubble for when we concluded that there is more benefit with PFMT than usual care
  - Tab 2 (different methods of PFMT). Green bubble for when we concluded that variation 1 is better than variation 2
  - Tab 3 (add-ons to PFMT). Green bubble for when we concluded that the add-on was beneficial, and orange bubble for when we concluded that the add-on yielded less benefit
  - Tab 4 (alternatives to PFMT). The final map, which includes only 1 comparison, had no conclusions, so all cells in the grid are either a gray bubble (insufficient) or a gray asterisk (no evidence).
For conclusions of superiority or inferiority, we use shading to indicate the strength of evidence (SOE), with darker shades denoting higher SOE.

- Regarding filters, we had meta-analyzed data to include all possible permutations of 3 filters (type of UI with 3 options, age with 4 options, and time since the start of treatment with 4 options), and the final tab 1 TE map included only the first two filters (type-of-UI and age), and the other TE map tabs included only the type-of-UI filter.

- For conclusions of superiority, the size of each bubble is linearly related to the estimated effect size (from a meta-analysis if there were multiple studies, or from a single study if there was only one study). We scaled bubble sizing of the three outcomes so that similar effect sizes had similarly sized bubbles. For conclusions of equivalence, or insufficient evidence bases, the bubbles were all the same (relatively small) size.

- Hovering over a bubble reveals our conclusion (“Evidence takeaway”) as well as definitions of the interventions in the comparison.

- Clicking on a bubble reveals a forest plot of the pertinent meta-analysis, the summary effect size, the I^2 as a measure of heterogeneity, hyperlinks to study abstracts. A Summary statement to the left of the plot shows our interpretive statement about the evidence (see next paragraph). A box above the statement is labeled “How to Read a Forest Plot”, which provides explanations of the axes of the plot and how to interpret it. A grid above the forest plot is a small-size version of the entire tab, so that users can quickly switch between the forest plots of that tab.

As interpretational aids, for each conclusion of superiority, we translated each odds ratio or Hedges’ g into more easily understood numerical sentences. For example, for tab 1 (effect of PFMT), 1 evidence base for cure found an estimated odds ratio of 5.9 in favor of PFMT with add-ons over usual care. In addition to providing the 5.9 effect size, the planned hover also stated “This odds ratio means that if the cure rate for Usual Care is 10%, then the estimated cure rate for PFMT Alone is 40%.” We used 10% as the hypothetical cure rate for usual care for all such statements, because 10% was a typical cure rate for usual care. For reduction in UI events and QOL, for usual care we hypothesized an improvement of 1/week (e.g., a Hedges’ g of 0.6 was interpreted to mean “This effect means that if Usual Care results in 1 fewer episodes per week, then PFMT Alone results in 6.8 fewer episodes per week”). For QOL, we used the most commonly-used instrument (the IQOL), and converted the Hedges’ g to the IQOL’s 0-100 scale (e.g., a Hedges’ g of 1.1 was interpreted to mean “This effect means that, compared to Usual Care, PFMT with Add-ons results in 18.4 points better on the IQOL (which ranges from 0 to 100).”). For other tabs, for these interpretive statements we used a lower cure rate of 40% and a lower reduction in UI events of 7/week.
Results

Evidence Base

We included a total of 73 RCTs (see article flow in Figure 1 below):

- 33 had been included by the 2012 AHRQ report
- 22 had been included by the 2018 AHRQ report
- 3 had been excluded by one or both AHRQ reports
- 4 had been included by the 2018 Cochrane report
- 11 were identified in our search update

Our searches of clinicaltrials.gov identified 47 potentially relevant records; 13 of these were randomized trials that made pertinent treatment comparisons (and were not already included in our evidence base). These trials are all listed in TE – Treatment effects; QOL – Quality of Life.
Table 3 in Appendix B. Our searches of the PCORI website identified no PCORI-funded studies on PFMT for women with UI.

Adverse effects were rarely reported, so we decided to exclude them from the maps (to reduce clutter). Instead, a separate section (below) describes the adverse effect data that we found in our included studies. Two of the 73 trials were excluded from both the RV map and the TE map, as the only included outcomes were adverse events.

Appendix B contains a table with study details (Table 2). The rightmost column identifies the TE map tabs where the study is included. Multi-arm trials could contribute to multiple tabs depending on the comparisons they made.
Figure 1. Article Flow

154 publications identified in our recent search (Dec 2017 or later)

Abstracts screened

125 excluded:
57: Not the right patient population
30: Review, pooled analysis, commentary, guideline, protocol, not an RCT
17: Not a comparison of interest
13: No PFMT group
6: No outcomes of interest
2: Duplicate

29 Full text review

18 excluded:
7: No outcomes of interest
3: Already included by 2018 AHRQ report
3: Not a comparison of interest
2: No the right patient population
1: Ancillary drug treatment
1: No data on separate groups
1: Only outcome of interest, which was reported nonsensically

33 RCTs included by the 2012 AHRQ report
22 more RCTs included by the 2018 AHRQ report
3 pre-2018 RCTs excluded by both AHRQ reports
4 more RCTs included by the 2018 Cochrane report
11 more RCTs found by our recent search

73 RCTs included
Selection of Comparisons for the Visualization

Sixty-eight of the included studies are summarized in the web-based interactive format, either as part of the visual or in a table (accessible using a link in the TE tabs). Specifically, the RV map includes 68 of the 73 included trials, since it does not depict the 2 trials whose only included data were on adverse events, and it also does not depict the 3 trials that measured the effect of PFMT when it is being added to another treatment. These 5 trials also do not appear in any of the TE maps. For the other 68 trials, any comparisons studied by only a single study do not appear as bubbles, but instead are accessible using a linked box in the bottom left of each TE tab (“View treatment comparisons by only one trial”). The Effect of PFMT tab does not have this box because all 3 of its comparisons had been investigated by at least 2 trials. Overall, the treatment comparisons represented as bubbles include any comparisons among the 68 trials that were investigated by at least 2 trials.

In the next 5 sections below, we provide our observations about the visualized data in the RV map as well as each tab of the TE map.

Research Volume Map: Observations

- The initial screen of the RV map shows a grand total of 68 trials reporting 89 treatment comparisons with data on 7,307 women. Among the 4 comparison categories, the Add-on and Different methods bars had the largest numbers of women (2,310 and 2,240 women, respectively).
- Selecting stress UI only (using the filter) reduces the total number by more than half (3,500 women with stress UI), and the four comparison categories all have similar numbers of women (between 825 and 924 women per comparison category).
- Using the age filter to focus on younger women (mostly women under 50) reduces the grand total to only 2,384 women, with the largest enrollment in comparisons of different methods of PFMT (868 women). Focusing on older women (mostly women older than 60) reduce the total even more (total 1,804 women) and the most-represented category is Effect of PFMT (930 women).
- Using the filter “# of weeks since the start of treatment” shows that very few trials followed women for 26 weeks or longer (only 3 of 68 trials). Far more trials (43 of 68) reported data between 12-26 weeks after the start of treatment.
- For publication year, most of the recent literature (2011 to mid 2019, 32 trials) has addressed either the effect of PFMT (12 trials) or add-ons to PFMT (18 trials). Only 4 recent trials have compared different methods of PFMT.
- For region of the world, only 15 of 68 trials (22%) have been conducted in the USA or Canada. Other prominent regions include Europe (22 trials, or 32%) and Asia (22 trials,
or 32%). Selecting both Europe and Brazil reveals that they provide the majority of the data comparing different methods of PFMT (14 of 22 trials).

**Treatment Effects Map, Effect of PFMT: Observations**

- The initial screen of the Effect of PFMT tab shows a 3x3 grid of unanimously green bubbles, which is an immediate indication that PFMT has been shown to be effective. For example, our largest meta-analysis (15 trials) is represented by the top center green bubble. Clicking on that bubble reveals the corresponding forest plot, which shows each individual study along the left side, and the combined summary statistic at the bottom right (summary Hedges’ g of 0.8 with a 95% CI from 0.6 to 1.0). This g=0.8 is difficult to understand on its own, so we have provided an interpretive statement to the left of the plot: “This effect means that if Usual Care results in 1 fewer episode per week, then PFMT alone or with Add-ons results in 9.3 fewer episodes per week.” This statement helps women understand that the typical benefit of PFMT is relatively large.

- Many women are interested in cure of UI, not merely a reduction in UI events. Therefore, cure data is shown in the left column of green bubbles. Clicking on the top left bubble shows the corresponding forest plot for our meta-analysis of 10 trials, with a summary odds ratio of 7.5 and a 95% CI from 3.6 to 15.2. Odds ratios are also difficult to understand, so we provided an interpretive statement: “This odds ratio means that if the cure rate for Usual Care is 10%, then the cure rate for PFMT alone or with Add-ons is 45%.”

- Some portion of the PFMT benefit is likely due to the additional interventions (such as bladder training) that 12 of the 20 trials gave to all of the women who tried PFMT. The other 8 trials examined the isolated effect of PFMT, and the middle row of the 3x3 grid summarized the corresponding data. Clicking on the green bubble in the middle of the grid reveals its meta-analysis (7 trials) and forest plot. As expected, the benefit is attenuated (summary g=0.6 with a 95% CI from 0.4 to 0.8), with an interpretive statement “This effect means that if Usual Care results in 1 fewer episode per week, then PFMT alone results in 6.8 fewer episodes per week.” The cure data, however, do not show this attenuation (summary odds ratio 12.3, cure rates 10% usual care vs 58% PFMT alone). Note that only 3 of the 8 PFMT-alone trials reported cure data, and so this latter estimate may be too high due to selective outcome reporting (those 3 authors may have chosen to report cure data only because they showed a large benefit of PFMT alone).

- Quality of life data (right column of bubbles) also show a benefit of PFMT. However, these data were reported less commonly (6 of 20 trials for the top right bubble, 2 of 8 trials for the middle right bubble, and 4 of 13 trials for the bottom right bubble).
• PFMT is often a first-line treatment for Stress UI, but not necessarily for Mixed or Urge UI. We conducted meta-analyses specifically for Stress UI, and these are visible by using the Type of UI filter to the left of the 3x3 grid. Selecting “Stress” shows that 8 of the 9 bubbles are still green, indicating that the findings show PFMT benefit even when the evidence only includes Stress UI. By contrast, if one selects “Stress/Mixed/Urge” (thereby excluding studies that only enrolled women with Stress UI), then only 6 of the 9 bubbles are still green. Particular notable is the middle row that contains a gray asterisk for Cure (indicating that none of the PFMT alone vs Usual Care trials that enrolled some women with Mixed or Urge UI reported cure data), a green bubble for Reduction in UI events (indicating PFMT benefit), and a gray bubble (indicating inconclusive data) for quality of life. This overall pattern of data confirms the clinical opinion that PFMT best targets Stress UI.

• The map also suggests that the evidence for PFMT appears to be stronger for younger women than for older women. Using the age filter and selecting “Mostly younger than 50” shows that the middle row (PFMT alone vs usual care) is green for all 3 bubbles. By contrast, selecting “Mostly older than 60” shows a complete lack of evidence on the isolated effect of PFMT on older women (shown as gray asterisks).

**Treatment Effects Map, Different Methods of PFMT: Observations**

• This visual shows 4 comparisons: 1) fully supervised vs partially supervised or unsupervised; 2) Group training vs individual training; 3) verbal feedback vs biofeedback; 4) verbal feedback vs. insertable training device. For these 4 comparisons, the map shows our 6 conclusions:
  o Across type of UI, the reduction in UI events is greater with fully supervised PFMT than with partially supervised or unsupervised PFMT. This is shown by the green bubble in the top center, when no filter is selected.
  o Across type of UI, the reduction in UI events is similar for verbal feedback and biofeedback. This is shown by the white bubble in the third row in the center, when no filter is selected.
  o When stress-Ul-only data are excluded (by selecting Stress/Mixed/Urge in the Type of UI filter), the reduction in UI events is greater with fully supervised or unsupervised PFMT.
  o For Stress/Mixed/Urge, the reduction in UI events is similar for group training and individual training.
  o For Stress/Mixed/Urge, the reduction in UI events is similar for verbal feedback and biofeedback.
  o For Stress/Mixed/Urge, the reduction in UI events is greater for verbal feedback than for insertable training device.
• For cure and quality of life for these 4 comparisons, all data were inconclusive, either due to no evidence (gray asterisks) or inconclusive evidence (gray bubbles). The typical reason for inconclusivity was a wide confidence interval around a summary effect size. Wide intervals are due to small numbers of trials, low patient enrollment, and/or high variability between patients.

**Treatment Effects Map, Add-ons to PFMT: Observations**

• This visual shows 6 comparisons: 1) Adding exercise to PFMT; 2) Adding behavioral therapy to PFMT; 3) Adding biofeedback to PFMT; 4) Adding electrical stimulation to PFMT; 5) Adding insertable training device to PFMT; 6) Adding verbal feedback to PFMT. For these 6 comparisons, the map shows our 7 conclusions:
  o Across type of UI, the reduction in UI events is similar for adding vs not adding behavioral training to PFMT. This is shown by the white bubble in the second row in the center, when no filter is selected.
  o Specifically for Stress UI, the reduction in UI events is greater when any exercise is added to PFMT. This is shown by the green bubble in the top center, when Stress is selected.
  o Specifically for Stress UI, the reduction in UI events is similar for adding vs not adding behavioral training to PFMT. This is shown by the white bubble in the second row in the center, when Stress is selected.
  o Specifically for Stress UI, the improvement in quality of life is greater when behavioral training is added to PFMT. This is shown by the green bubble in the second row on the right, when Stress is selected.
  o Specifically for Stress/Mixed/Urge UI (i.e., excluding Stress-only studies), cure rates are higher when any exercise is added to PFMT. This is shown by the green bubble in the top row on the left, when Stress/Mixed/Urge is selected.
  o Specifically for Stress/Mixed/Urge UI, the reduction in UI events is similar for adding vs not adding behavioral training to PFMT. This is shown by the white bubble in the second row in the center, when Stress/Mixed/Urge is selected.
  o Specifically for Stress/Mixed/Urge UI, cure rates are higher when an insertable training device is *not* added to PFMT. This is shown by the orange bubble in the 5th row on the left, when Stress/Mixed/Urge is selected.
• Other analyses for these 6 comparisons were all inconclusive, either due to no evidence, or inconclusivity.
Treatment Effects Map, Alternatives to PFMT: Observations

- This tab shows 1 comparison: PFMT with verbal feedback vs electrical stimulation without PFMT. All of these data were inconclusive. Both trials had enrolled only patients with Stress UI.

Adverse Effects

Four studies evaluating PFMT effectiveness (TE map tab 1) reported on adverse effects.

Adverse effects reported in these studies included:

- Any adverse effect (0% in both the PFMT group and the non-PFMT group)
- Dry mouth (22/63 in the PFMT group vs 34/62 in the non-PFMT group)
- Constipation (14/63 in the PFMT group vs 23/62 in the non-PFMT group)
- Blurred vision (6/63 in the PFMT group vs 6/62 in the non-PFMT group)
- Confusion (4/63 in the PFMT group vs 7/62 in the non-PFMT group)
- Inability to void (4/63 in the PFMT group vs 2/62 in the non-PFMT group)
- Vaginal bleeding (0% in both groups)
- Urinary infection (0% in both groups)
- Vulvovaginitis (0% in both groups)
- Tenderness and bleeding (0% in both groups)
- Discomfort (0% in both groups)
- Abdominal pain (0% in both groups)
- Bleeding (0% in both groups)

Four studies evaluating different methods of PFMT (TE map tab 2) reported on adverse effects.

Adverse effects reported in these studies included:

- Any adverse effect (0% in both PFMT variation groups)
- Vaginal irritation (0% in both groups)
- Lower abdominal pain when conducting PFMT (1/124 in the internet-based PFMT with email support group vs 0/126 in the mail-based PFMT without email support group)
- Vaginal bleeding (0% in both groups)
- Urinary infection (0% in both groups)
- Vulvovaginitis (0% in both groups)
Seven studies\textsuperscript{10,20,21,31,34-36} combining other treatments with PFMT (TE map tab 3) reported on adverse effects. Adverse effects reported in these studies included:

- Any adverse effect (4/35 in the combination with PFMT group vs 1/30 in the PFMT alone group in Porta Roda et al. (2014)\textsuperscript{21}; 17/28 in the PFMT + yoga group vs 14/28 in the PFMT alone group in Huang et al. (2019)\textsuperscript{35}; 0\% in both the combination with PFMT group and the PFMT alone group in the other studies)\textsuperscript{10,20,21,34}
- Gastrointestinal adverse events (0/28 in the PFMT + yoga group vs 2/28 in the PFMT alone group)\textsuperscript{35}
- Genitourinary adverse events (1/28 in the PFMT + yoga group vs 0/28 in the PFMT alone group)\textsuperscript{35}
- Musculoskeletal adverse events (8/28 in the PFMT + yoga group vs 7/28 in the PFMT alone group)\textsuperscript{35}
- Neurological or psychological adverse events (4/28 in the PFMT + yoga group vs 4/28 in the PFMT alone group)\textsuperscript{35}
- Ophthalmologic adverse events (1/28 in the PFMT + yoga group vs 3/28 in the PFMT alone group)\textsuperscript{35}
- Respiratory or sinus adverse events (8/28 in the PFMT + yoga group vs 9/28 in the PFMT alone group)\textsuperscript{35}
- Serious adverse events (0\% in both groups)\textsuperscript{35}
- Hypersensitivity (1/35 in the PFMT + weights group vs 0/30 in the PFMT alone group)\textsuperscript{21}
- Irritation (1/35 in the PFMT + weights group vs 0/30 in the PFMT alone group)\textsuperscript{21}
- Itching (1/35 in the PFMT + weights group vs 0/30 in the PFMT alone group)\textsuperscript{21}
- Local discomfort (1/35 in the PFMT + weights group vs 0/30 in the PFMT alone group)\textsuperscript{21}
- Low tolerance to physical therapy (0/35 in the PFMT + weights group vs 1/30 in the PFMT alone group)\textsuperscript{21}
- Vaginal irritation (4/67 in the PFMT + electrical stimulation group vs 0/67 in the PFMT alone group)\textsuperscript{31}
- Vaginal discharge (23/39 in the PFMT + assistive device group vs 0/46 in the PFMT alone group)\textsuperscript{36}
- Pain (3/39 in the PFMT + assistive device group vs 0/46 in the PFMT alone group)\textsuperscript{36}
- Spotting (6/39 in the PFMT + assistive device group vs 0/46 in the PFMT alone group)\textsuperscript{36}

Three studies\textsuperscript{28,30,37} evaluating alternatives to PFMT (TE map tab 4) reported on adverse effects. Adverse effects reported in these studies included:

- Any adverse effect (0\% in both the PFMT group and the alternative to PFMT group)\textsuperscript{37}
- Tenderness and bleeding (0/25 in the PFMT group vs 1/25 in the electrical stimulation group and 0/27 in the vaginal cones group)\textsuperscript{30}
- Discomfort (0/25 in the PFMT group vs 1/25 in the electrical stimulation group and 0/27 in the vaginal cones group)\textsuperscript{30}
- Vulvovaginitis (0/25 in the PFMT group vs 0/25 in the electrical stimulation group and 1/27 in the vaginal cones group)\textsuperscript{30}
- Abdominal pain (0/25 in the PFMT group vs 0/25 in the electrical stimulation group and 2/27 in the vaginal cones group)\textsuperscript{30}
- Bleeding (0/25 in the PFMT group vs 0/25 in the electrical stimulation group and 1/27 in the vaginal cones group)\textsuperscript{30}
- Vaginal bleeding (0% in both groups)\textsuperscript{28}
- Urinary infection (0% in both groups)\textsuperscript{28}
- Vulvovaginitis (0% in both groups)\textsuperscript{28}
Limitations

We acknowledge several important limitations to this work:

First, some studies compared treatment groups that differed in the amount of contact time between patients and healthcare providers. We believe that contact time with healthcare providers, regardless of the activity, can influence patient outcomes (perhaps via implicit reminders to perform the prescribed PFMT). Consider a 3-group study where two groups had similar contact times (e.g., 2 variants of PFMT involving a similar number of clinic visits), but the third group had less contact time (e.g., usual care without clinic visits). If we factored in possible healthcare interaction confounding into our risk-of-bias assessment, the comparison of the first two groups would receive Low risk of bias for confounding, whereas comparisons involving the third group would receive a High risk of bias for confounding. While we feel that these differences may contribute to patient outcomes, in order to be consistent with the risk-of-bias assessment methods used in the 2012 and 2018 AHRQ reports, we did not consider healthcare interaction confounding when performing our own risk-of-bias assessment.

Second, for each included trial, only a single person performed a risk-of-bias assessment. Guidelines generally recommend dual assessment, due to the subjective nature of the task. However, at the start, the 3-person team did perform risk-of-bias assessments on the same set of 10 articles, to resolve any interpretational differences regarding the risk-of-bias items. Despite this effort, we recognize that our risk-of-bias assessments may have been inconsistent.

Third, for tab 2 of the TE map, there was a dizzying array of different methods of PFMT, so for brevity’s sake, we sometimes grouped and performed meta-analyses on studies that had varying levels of clinician involvement (e.g., studies using partially supervised or unsupervised PFMT), or studies that used different PFMT add-ons (e.g., PFMT with or without bladder training; with or without behavioral training; with or without biofeedback). It is possible that our conclusions could change if these studies had been analyzed separately. Additionally, for the studies represented in tabs 1, 3, and 4, we did not extract data regarding the extent of clinician involvement and/or the specific PFMT delivery method (e.g., PFMT information provided in a brochure vs in-person PFMT instruction), because these did not differ between treatment groups. Therefore, we may have meta-analyzed studies that varied in clinician involvement and/or PFMT delivery method. If this were the case, the conclusions we reached in tabs 1, 3, and 4 could change if we had considered these variables and determined that they differed enough to warrant separate analyses.

Finally, in the field of first-line treatments of UI, new trials are often being published, and updating the maps accordingly would be quite resource-intensive. For each new trial, one would have to determine trial eligibility, categorize treatment comparisons/outcomes, compute
new effect sizes and standard errors, perform new meta-analyses, interpret the results, rate SOE (dual ratings) for the newly combined evidence, resolve discrepancies, update the source files accordingly for all filter combinations, add links to the trial abstract, and re-draw the maps. None of these steps are fully automatable.

**Conclusions, Implications, and Future Research**

Traditionally, evidence-based medicine has not taken advantage of interactive tools to promote informed, shared decision-making for patients, clinicians, and policymakers. The use of Evidence Maps to distill the results of a comprehensive synthesis, using an interactive Web-based format, is a promising new development.

Our visual maps summarize data from 68 trials reporting 89 treatment comparisons with data on 7,307 women. Trials were often conducted in Europe or Asia (64%), and about half (53%) were conducted in 2010 or earlier. Only 4 trials had measured outcomes at least 6 months after the start of treatment. Many trials enrolled a mix of women at different ages and with different types of UI (Stress or Mixed or Urge).

Our maps generally confirm the efficacy of PFMT, at least in the short term. Meta-analyses isolating the effect of PFMT alone suggests about 7 fewer UI episodes per week, compared to only 1 fewer episode under usual care. The evidence appears to be stronger for younger women as well as women with stress UI, confirming the general clinical opinion that PFMT should be a first-line treatment for stress UI.

Most other evidentiary aspects of PFMT are unclear, such as comparing different methods of PFMT, whether various add-ons are beneficial, and whether non-PFMT treatment are better or worse than PFMT. The inconclusiveness is mostly due to limited evidence for specific comparison/outcome combinations, such as low numbers of trials, low patient enrollment, and high variability among patients.

We note 4 specific areas where conclusions were possible. First, fully supervised PFMT yields greater reductions in UI events than partially supervised or unsupervised PFMT. Second, we found similar outcomes with verbal-feedback PFMT and biofeedback-PFMT. Third, for stress UI, adding exercise to PFMT reduces the frequency of UI events, but this result is based on only 2 studies, therefore adding exercise to PFMT is a ripe target for future research. Fourth, there were similar outcomes with adding vs not adding behavioral training to PFMT.

Future research should report longer-term outcomes (e.g., 26 weeks or longer), and report quality-of-life data using a validated instrument with a defined minimal important difference. Most studies report reductions in UI events, but fewer reported cure rates, and even fewer reported quality of life data. These latter outcomes, in the longer term, are likely more important to most women who suffer from UI.
Acknowledgements

We acknowledge Dr. Diane Newman of the Perelman School of Medicine at the University of Pennsylvania for her contributions to the project, including her input on how treatments might be categorized and which comparisons, outcomes, and timeframes are clinically important.
References


Appendices

Appendix A. Literature Search Methods

Literature Search:

In July 2019, we conducted a literature review to identify research focused on pelvic floor muscle training for female urinary incontinence. Our search protocol included PubMed and EMBASE/Medline databases via the Embase.com platform. We present the strategies in Embase.com syntax (using EMTREE) in the tables below. We translated the controlled vocabulary terms and syntax for the PubMed searches.

Clinical Trials and PCORI website:

To identify additional trials in this topic area, we searched ClinicalTrials.gov through December 6, 2018 and the PCORI website through July 5, 2019.

Search terms used for ClinicalTrials.gov include:

Condition or disease: urinary incontinence OR enuresis OR stress urinary incontinence OR incontinence, urge OR incontinence, urinary OR incontinence, urinary stress OR incontinence, stress OR urinary incontinence OR urine; incontinence, stress

Intervention/treatment: exercise OR exercise* OR kegel OR kegel* OR "pelvic floor" OR "pelvic floor muscle training" OR "physical therapy" OR physiotherapy OR physio* OR pilates OR pfm OR pfmt

The following checktags were also applied: recruiting, not yet recruiting, active not recruiting, enrolling by invitation, female, adult (18-64) older adult (65+)

Search terms used for PCORI website include:

urinary; urine; bladder; incontinent; incontinence
### Bibliographic search strategy:

**Embase.com [2017-2018]**

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Concept</th>
<th>Search Statement</th>
<th># Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Urinary Incontinence</td>
<td>enuresis/exp OR 'mixed incontinence'/exp OR 'stress incontinence'/exp OR 'urge incontinence'/exp OR 'urine incontinence'/exp OR ((enuresis OR incontinen*) NEAR/2 (bladder OR mixed OR stress* OR urge OR urin*)) OR ui OR sui</td>
<td>109 309</td>
</tr>
<tr>
<td>S2</td>
<td>Female Population</td>
<td>female/exp OR (female* OR women* OR woman*) OR [female]/lim</td>
<td>9 042 473</td>
</tr>
<tr>
<td>S3</td>
<td>Pelvic Floor Muscle Training</td>
<td>'pelvic floor muscle training'/exp OR (pelvic NEAR/3 (exercis* OR train*))) OR exercis* OR kegel* OR pfm OR pfmt OR 'physical therapy' OR physiotherapy/exp OR physiotherap* OR pilates</td>
<td>673 536</td>
</tr>
<tr>
<td>S4</td>
<td>Limits for Study Designs/Publication Types/Date and Language</td>
<td>'randomized controlled trial'/exp OR 'randomized controlled trial'' OR random*:ti NOT (abstract:nc OR annual:nc OR book/exp OR 'case report'/exp OR 'case study'/exp OR conference:nc OR 'conference abstract'.it OR 'conference paper'/exp OR 'conference paper':it OR 'conference proceeding':pt OR 'conference review':it OR congress:nc OR editorial/exp OR editorial:it OR erratum/exp OR letter:it OR note/exp OR note:it OR meeting:nc OR sessions:nc OR 'short survey'/exp OR symposium:nc) AND [english]/lim AND [2017-2018]/py</td>
<td>81 388</td>
</tr>
<tr>
<td>S5</td>
<td>Combine Sets</td>
<td>S1 AND S2 AND S3 AND S4</td>
<td>72</td>
</tr>
</tbody>
</table>

**DOWNLOADED RECORDS** 72
### Table 2. Details of Included Studies

<table>
<thead>
<tr>
<th>Author/year</th>
<th>Mean Age</th>
<th>Type of UI</th>
<th>Follow-ups (weeks)</th>
<th>Treatments</th>
<th>Outcomes</th>
<th>TE map tab 1</th>
<th>TE map tab 2</th>
<th>TE map tab 3</th>
<th>TE map tab 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdulaziz and Hasan (2012)(^38)</td>
<td>44</td>
<td>Stress</td>
<td>13.1</td>
<td>Fully supervised PFMT&lt;br&gt;Usual care</td>
<td>Cure</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aksac et al. (2003)(^39)</td>
<td>53</td>
<td>Stress</td>
<td>8</td>
<td>Fully supervised PFMT + verbal feedback&lt;br&gt;Fully supervised PFMT + biofeedback&lt;br&gt;Usual care</td>
<td>Cure</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arvonen et al. (2001)(^40)</td>
<td>48</td>
<td>Stress</td>
<td>17.4</td>
<td>Fully supervised PFMT + verbal feedback&lt;br&gt;Fully supervised PFMT + weights (vaginal balls)</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asklund I., et al. (2017)(^41)</td>
<td>45</td>
<td>Stress</td>
<td>12</td>
<td>Unsupervised, app-based PFMT&lt;br&gt;Usual care</td>
<td>Reduction, QOL</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berghmans et al. (1996)(^42)</td>
<td>48</td>
<td>Stress</td>
<td>4</td>
<td>Fully supervised PFMT&lt;br&gt;Fully supervised PFMT + biofeedback</td>
<td>Cure, Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bo et al. (1999)(^30)</td>
<td>49.5</td>
<td>Stress</td>
<td>26</td>
<td>Fully supervised PFMT&lt;br&gt;Electrical stimulation&lt;br&gt;Weights (vaginal cones)&lt;br&gt;Usual care</td>
<td>Cure, Reduction, AEs</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Borello-France et al. (2008)(^43)</td>
<td>53</td>
<td>Stress</td>
<td>26.1</td>
<td>Unsupervised PFMT maintenance program (4x/week)&lt;br&gt;Unsupervised PFMT maintenance program (1x/week)</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burgio et al. (1998)(^29)</td>
<td>67.5</td>
<td>Urge, mixed</td>
<td>8</td>
<td>Fully supervised PFMT + biofeedback + bladder training&lt;br&gt;Placebo</td>
<td>Cure, Reduction, AEs</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burgio et al. (2002)(^46)</td>
<td>65</td>
<td>Urge/Mixed</td>
<td>8</td>
<td>Fully supervised behavioral training (PFMT, home exercises, bladder training, and bladder diary self-monitoring) + verbal feedback&lt;br&gt;Fully supervised behavioral training (PFMT, home exercises, bladder training, and bladder diary self-monitoring) + biofeedback&lt;br&gt;Unsupervised behavioral training (self-help booklet)</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cammu and Van Nylen (1998)(^47)</td>
<td>56</td>
<td>Stress</td>
<td>12</td>
<td>Fully supervised PFMT + verbal and biofeedback&lt;br&gt;Fully supervised PFMT + weights (vaginal cones)</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author/year</td>
<td>Mean age</td>
<td>Type of UI</td>
<td>Follow-ups (weeks)</td>
<td>Treatments</td>
<td>Outcomes</td>
<td>TE map tab 1</td>
<td>TE map tab 2</td>
<td>TE map tab 3</td>
<td>TE map tab 4</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>----------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Castro et al. (2008)²⁸</td>
<td>54</td>
<td>Stress</td>
<td>26.1</td>
<td>Fully supervised PFMT + verbal feedback&lt;br&gt;Fully supervised electrical stimulation&lt;br&gt;Fully supervised PFMT + weights (vaginal cones)&lt;br&gt;Usual care</td>
<td>Cure, Reduction, QOL, AEs</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Celiker et al. (2015)⁴⁸</td>
<td>52</td>
<td>Stress/Urge/Mixed</td>
<td>12</td>
<td>Fully supervised PFMT&lt;br&gt;Usual care</td>
<td>Reduction</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Oliveira Camargo et al. (2009)⁴⁹</td>
<td>51</td>
<td>Stress</td>
<td>12</td>
<td>Fully supervised PFMT (group)&lt;br&gt;Fully supervised PFMT (individual)</td>
<td>Cure, Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Souza Abreu et al. (2017)⁵⁰</td>
<td>54</td>
<td>Stress</td>
<td>5, 17.9</td>
<td>Fully supervised PFMT&lt;br&gt;Fully supervised PFMT + dynamic lumbopelvic stabilization exercise</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delgado et al. (2013)³¹</td>
<td>50</td>
<td>Stress</td>
<td>16, 42</td>
<td>Fully supervised PFMT + pelvic toner device&lt;br&gt;Fully supervised PFMT</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demain et al. (2001)⁵²</td>
<td>50</td>
<td>Stress/Urge/Mixed</td>
<td>12</td>
<td>Fully supervised PFMT + bladder training (group)&lt;br&gt;Fully supervised PFMT + bladder training (individual)</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diokno et al. (2018)⁵³</td>
<td>65</td>
<td>Stress/Urge/Mixed</td>
<td>13.1, 52</td>
<td>Partially supervised behavioral training (PFMT, bladder training), administered in a group setting&lt;br&gt;Usual care</td>
<td>Reduction, QOL</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Due et al. (2018)⁵⁴</td>
<td>45</td>
<td>Stress</td>
<td>12</td>
<td>Partially supervised PFMT + biofeedback&lt;br&gt;Partially supervised PFMT + biofeedback + KAATSU (thigh occlusion training)</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dumoulin et al. (2004)⁵⁵</td>
<td>36</td>
<td>Stress</td>
<td>8</td>
<td>Fully supervised PFMT + electrical stimulation + biofeedback&lt;br&gt;Fully supervised PFMT + electrical stimulation + biofeedback + abdominal exercises</td>
<td>Cure, QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felicissimo et al. (2010)⁵⁶</td>
<td>50</td>
<td>Stress</td>
<td>8</td>
<td>Fully supervised PFMT&lt;br&gt;Unsupervised PFMT</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferreira et al. (2014)⁵⁷</td>
<td>19.3</td>
<td>Stress</td>
<td>12</td>
<td>Unsupervised PFMT + education&lt;br&gt;Usual care</td>
<td>Reduction</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firra et al. (2013)⁵⁸</td>
<td>57.5</td>
<td>Stress, Urge</td>
<td>8</td>
<td>Fully supervised PFMT+ES&lt;br&gt;Fully supervised PFMT&lt;br&gt;Usual care</td>
<td>Reduction</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitz et al. (2017)⁵⁹</td>
<td>56</td>
<td>Stress</td>
<td>12, 36</td>
<td>Partially supervised PFMT&lt;br&gt;Partially supervised PFMT + biofeedback</td>
<td>Cure, Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fürst et al. (2014)²⁴</td>
<td>52</td>
<td>Stress</td>
<td>13.1</td>
<td>Fully supervised PFMT + electrical stimulation&lt;br&gt;Fully supervised electrical stimulation</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author/year</td>
<td>Mean age</td>
<td>Type of UI</td>
<td>Follow-ups (weeks)</td>
<td>Treatments</td>
<td>Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glavind et al. (1996)</td>
<td>45</td>
<td>Stress</td>
<td>12</td>
<td>Fully supervised PFMT&lt;br&gt; Fully supervised PFMT + biofeedback</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Golmakani et al. (2014)</td>
<td>46</td>
<td>Stress</td>
<td>12</td>
<td>Fully supervised PFMT + bladder training&lt;br&gt; Fully supervised PFMT + weights (vaginal cones)</td>
<td>Reduction, QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goode et al. (2003)</td>
<td>56</td>
<td>Stress/Mixed (&quot;Mixed&quot;)</td>
<td>8</td>
<td>Fully supervised behavioral training (PFMT + biofeedback, home exercises, bladder training, and bladder diary self-monitoring)&lt;br&gt; Fully supervised behavioral training (PFMT + biofeedback, home exercises, bladder training, and bladder diary self-monitoring) + electrical stimulation&lt;br&gt; Unsupervised behavioral training (self-help booklet)</td>
<td>Reduction, AEs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hahn et al. (1991)</td>
<td>47</td>
<td>Stress</td>
<td>13.1</td>
<td>Partially supervised PFMT + verbal feedback&lt;br&gt; Partially supervised electrical stimulation</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hirakawa et al. (2013)</td>
<td>57</td>
<td>Stress</td>
<td>12</td>
<td>Fully supervised PFMT + biofeedback&lt;br&gt; Fully supervised PFMT</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huang et al. (2019)</td>
<td>66.1</td>
<td>Stress, Urge, Mixed</td>
<td>12</td>
<td>Unsupervised behavioral training (PFMT + bladder training) + yoga&lt;br&gt; Unsupervised behavioral training (PFMT + bladder training)</td>
<td>Reduction, QOL, AEs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hui et al. (2006)</td>
<td>74</td>
<td>Stress/Urge (&quot;Mixed&quot;)</td>
<td>8</td>
<td>Fully supervised behavioral training (PFMT, bladder training, diet and fluid management) via videoconferencing&lt;br&gt; Fully supervised behavioral training (PFMT, bladder training, diet and fluid management) + digital and biofeedback via in-person care</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hung et al. (2010)</td>
<td>49</td>
<td>Stress/Mixed (&quot;Mixed&quot;)</td>
<td>17.4</td>
<td>Fully supervised PFMT + verbal feedback + abdominal exercises&lt;br&gt; Partially supervised PFMT + verbal feedback</td>
<td>Cure, Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Janssen et al. (2001)</td>
<td>48</td>
<td>Stress/Urge/Mixed (&quot;Mixed&quot;)</td>
<td>13.1, 52</td>
<td>Fully supervised PFMT + bladder training (group)&lt;br&gt; Fully supervised PFMT + bladder training (individual)</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kashanian et al. (2011)</td>
<td>40</td>
<td>Stress, Mixed</td>
<td>17.4, 26.1</td>
<td>Fully supervised PFMT + assistive device (Kegelmaster)&lt;br&gt; Fully supervised PFMT</td>
<td>Cure, QOL, AEs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaya et al. (2015)</td>
<td>50</td>
<td>Stress, Urge, Mixed (reported separately)</td>
<td>6</td>
<td>Fully supervised PFMT + bladder training&lt;br&gt; Fully supervised bladder training</td>
<td>Cure, Reduction, QOL, AEs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenton et al. (2012)</td>
<td>50</td>
<td>Stress</td>
<td>13.1</td>
<td>Fully supervised behavioral training (PFMT, bladder training)&lt;br&gt; Partially supervised pessary treatment</td>
<td>QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim (2001)</td>
<td>54</td>
<td>Stress/Urge/Mixed (&quot;Mixed&quot;)</td>
<td>12</td>
<td>Fully supervised behavioral training (PFMT, bladder training, audiovisual tape, calendar, pamphlet, followup phone calls)&lt;br&gt; Partially supervised PFMT</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author/year</td>
<td>Mean age</td>
<td>Type of UI</td>
<td>Follow-ups (weeks)</td>
<td>Treatments</td>
<td>Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>------------</td>
<td>--------------------</td>
<td>------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et al. (2007)</td>
<td>77</td>
<td>Stress</td>
<td>13.1, 52</td>
<td>Fully supervised PFMT + exercise program&lt;br&gt;Usual care</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et al. (2011a)</td>
<td>76</td>
<td>Stress, Urge, Mixed</td>
<td>13.1</td>
<td>Fully supervised PFMT/strength training + heat &amp; steam generating sheet&lt;br&gt;Fully supervised PFMT/strength training&lt;br&gt;Heat &amp; steam generating sheet&lt;br&gt;General education</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et al. (2011b)</td>
<td>76</td>
<td>Stress, Urge, Mixed</td>
<td>13.1, 43.5</td>
<td>Fully supervised PFMT + strength training&lt;br&gt;General education</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Konstantinidou et al. (2007)</td>
<td>48</td>
<td>Stress</td>
<td>12</td>
<td>Fully supervised PFMT (group)&lt;br&gt;Fully supervised PFMT (individual)</td>
<td>Improv</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kumari et al. (2008)</td>
<td>45</td>
<td>Unclassified (&quot;Mixed&quot;)</td>
<td>8, 21.1, 34.8</td>
<td>Fully supervised behavioral training (PFMT, bladder training, and bladder and exercise diaries)&lt;br&gt;Usual care</td>
<td>Cure, Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamb et al. (2009)</td>
<td>51</td>
<td>Stress/Urage/Mixed (&quot;Mixed&quot;)</td>
<td>6, 21.8</td>
<td>Fully supervised group behavioral training (PFMT, bladder training)&lt;br&gt;Fully supervised individual behavioral training (PFMT, bladder training)</td>
<td>QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lausen et al. (2018)</td>
<td>50</td>
<td>Unclassified (&quot;Mixed&quot;)</td>
<td>20</td>
<td>Fully supervised PFMT + Group Pilates&lt;br&gt;Fully supervised PFMT</td>
<td>QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leong et al. (2015)</td>
<td>74</td>
<td>Stress/Urage/Mixed (&quot;Mixed&quot;)</td>
<td>12</td>
<td>Fully supervised behavioral training (PFMT + verbal feedback, bladder training, education)&lt;br&gt;Unsupervised PFMT self-education</td>
<td>Reduction, AEs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liebergall-Wischnitzer et al. (2009)</td>
<td>48</td>
<td>Stress</td>
<td>12</td>
<td>Fully supervised PFMT&lt;br&gt;Fully supervised Paula method exercises</td>
<td>Cure, QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manonai et al. (2015)</td>
<td>48</td>
<td>Stress</td>
<td>16</td>
<td>Fully supervised PFMT&lt;br&gt;Fully supervised PFMT + single biofeedback session</td>
<td>QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mateus-Vasconcelos et al. (2018)</td>
<td>53</td>
<td>Unclassified (&quot;Mixed&quot;)</td>
<td>8</td>
<td>Partially supervised PFMT&lt;br&gt;Fully supervised PFMT + verbal feedback with pelvic tilt&lt;br&gt;Fully supervised PFMT + verbal feedback&lt;br&gt;Fully supervised PFMT + electrical stimulation</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McLean et al. (2013)</td>
<td>51.8</td>
<td>Stress, Mixed</td>
<td>12</td>
<td>Fully supervised PFMT&lt;br&gt;Usual care</td>
<td>Reduction, QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morkved et al. (2002)</td>
<td>47</td>
<td>Stress</td>
<td>26.1</td>
<td>Fully supervised PFMT + verbal feedback&lt;br&gt;Fully supervised PFMT + biofeedback</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author/year</td>
<td>Mean age</td>
<td>Type of UI</td>
<td>Follow-ups (weeks)</td>
<td>Treatments</td>
<td>Outcomes</td>
<td>TE map tab 1</td>
<td>TE map tab 2</td>
<td>TE map tab 3</td>
<td>TE map tab 4</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
</tbody>
</table>
| Okayama et al. (2019)<sup>79</sup> | 44       | Stress              | 6, 12              | Unsupervised PFMT  
Wearing supportive underwear  
Usual care                      | Cure, Reduction            | ✓                         |              |              |              |
| Oldham et al. (2013)<sup>34</sup> | 48       | Stress/Urge/ Mixed  | 12                 | Unsupervised PFMT + electrical stimulation (Pelviva)  
Unsupervised PFMT               | Reduction, AE             |              |              |              |              |
| Ong et al. (2015)<sup>80</sup>   | 52       | Stress              | 4, 16              | Fully supervised PFMT  
Fully supervised PFMT + biofeedback (Vibranc Kegel) | Cure                      |              |              |              |              |
| Orhan et al. (2018)<sup>81</sup> | 49       | Stress              | 4, 8, 12           | Fully supervised PFMT + vaginal tampon training  
Fully supervised PFMT           | Cure, AEs                  |              |              |              |              |
| Pages et al. (2001)<sup>10</sup> | 51       | Stress              | 4, 12.7            | Partially supervised PFMT (group)  
Partially supervised PFMT (group) + biofeedback | Cure, Reduction           |              |              |              |              |
| Pereira et al. (2013)<sup>20,82</sup> | 63      | Stress              | 58                 | Fully supervised PFMT + weights (vaginal cones)  
Fully supervised PFMT           | AEs                       |              |              |              |              |
| Porta Roda et al. (2014)<sup>21</sup> | NR      | Stress/ Mixed       | Any                | Partially supervised PFMT + weights (vaginal spheres)  
Partially supervised PFMT       | AEs                       |              |              |              |              |
| Ptak et al. (2019)<sup>63</sup>   | 53.1     | Stress              | 12                 | Fully supervised PFMT + transverse abdominal muscle training  
Fully supervised PFMT           | QOL                       |              |              |              |              |
| Sacomori et al. (2015)<sup>22</sup> | 50      | Stress/Urge/ Mixed  | 12.9               | Partially supervised PFMT + motivational videotape  
Partially supervised PFMT       | Cure                      |              |              |              |              |
| Sari et al. (2009)<sup>84</sup>   | 43.2     | Stress, Mixed       | 8                  | Partially supervised PFMT + The Knack  
Usual care                       | Reduction, QOL             | ✓                         |              |              |              |
| Sherman et al. (1997)<sup>85</sup> | 29       | Stress/ Mixed       | 8.7                | Fully supervised PFMT + bladder training  
Fully supervised PFMT + biofeedback + bladder training | Reduction                  |              |              |              |              |
| Sjöström et al. (2013)<sup>33,86,87</sup> | 49      | Stress              | 17.4               | Partially supervised behavioral training (PFMT, education, bladder diary) via internet with email support  
Unsupervised behavioral training (PFMT, education, bladder diary) via regular mail without email support | Redution, QOL, AEs          |              |              |              |              |
| Sran et al. (2016)<sup>27</sup>   | 67       | Stress/Urge/ Mixed  | 13.1, 52           | Fully supervised behavioral training (PFMT + biofeedback, bladder training, motor control exercises, dietary changes)  
Usual care                       | Reduction, AEs             | ✓                         |              |              |              |
| Sung et al. (2000)<sup>88</sup>   | NR       | Stress              | 6                  | Fully supervised intensive PFMT  
Fully supervised PFMT + electrical stimulation + biofeedback | Reduction                  |              |              |              |              |
<table>
<thead>
<tr>
<th>Author/year</th>
<th>Mean/Type of UI</th>
<th>Follow-ups (weeks)</th>
<th>Treatments</th>
<th>Outcomes</th>
<th>TE map tab 1</th>
<th>TE map tab 2</th>
<th>TE map tab 3</th>
<th>TE map tab 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talley et al. (2017)</td>
<td>85 Stress/Urge/Mixed (&quot;Mixed&quot;)</td>
<td>12</td>
<td>Partially supervised behavioral training (PFMT, bladder training, exercise)</td>
<td>Partially supervised PFMT, bladder training, exercise</td>
<td>Reduction, QOL</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtuoso et al. (2019)</td>
<td>64.8 Stress, Mixed</td>
<td>4, 12, 16</td>
<td>Fully supervised PFMT + weight training</td>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wagg et al. (2019)</td>
<td>64.6 NR</td>
<td>24</td>
<td>Fully supervised PFMT + bladder training</td>
<td>Cure, Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams et al. (2006)</td>
<td>57 Stress/Mixed (&quot;Mixed&quot;)</td>
<td>13.1</td>
<td>Fully supervised PFMT + verbal feedback</td>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wong et al. (2001)</td>
<td>46 Stress</td>
<td>4</td>
<td>Fully supervised PFMT + vaginal feedback</td>
<td>Reduction, QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyman et al. (1998)</td>
<td>61 Stress, Urge/Mixed</td>
<td>12, 25.1</td>
<td>Partially supervised PFMT + biofeedback</td>
<td>Cure, Reduction, QOL</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Zanetti et al. (2007)</td>
<td>55 Stress</td>
<td>13.1</td>
<td>Fully supervised PFMT</td>
<td>Cure, Reduced, QOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For Type of UI, a study that enrolled women with multiple types of UI is referred to as Mixed in the table. For outcomes, “Cure” was defined differently by different studies, “Improv.” is an abbreviation for Improvement and the study reported data on the change in frequency of UI events, and “QOL” means the study reported data on one of the four quality-of-life instruments that we included.

AEs – Adverse effects; Improv. – Improvement; NR – Not reported; PFMT – Pelvic floor muscle training; TE – Treatment effects; QOL – Quality of Life
### Table 3. Clinicaltrials.gov Records of Studies Not Yet Published

<table>
<thead>
<tr>
<th>NCT Number</th>
<th>Title</th>
<th>Start Date and Projected End Date</th>
<th>Study Design</th>
<th>Number of Patients</th>
<th>Intervention(s)</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCT03194789</td>
<td>The Effect of Footwear Generated Biomechanical Manipulation on Symptoms of Stress Urinary Incontinence</td>
<td>November 2017 to December 2020</td>
<td>RCT</td>
<td>64</td>
<td>PFMT; Footwear + PFMT</td>
<td>QOL (UDI-6)</td>
</tr>
<tr>
<td>NCT03238716</td>
<td>Neuromuscular Re-education, Exercise and Electric Dry Needling versus Neuromuscular Re-education and Exercise for Stress Urinary Incontinence</td>
<td>July 2017 to May 2019</td>
<td>RCT</td>
<td>60</td>
<td>Neuromuscular reeducation (PFMT + BF); dry needling; exercise</td>
<td>QOL (UDI full)</td>
</tr>
<tr>
<td>NCT02039830</td>
<td>Group Versus Individual Physiotherapy for Urinary Incontinence in Aging Women</td>
<td>July 2012 to June 2018 (record last updated February 2018)</td>
<td>RCT</td>
<td>364</td>
<td>Group PFMT; individual PFMT</td>
<td>Cure, Improvement (complete or partial reduction in number of UI episodes, as recorded in bladder diary); QOL (ICIQ LUTSsql)</td>
</tr>
<tr>
<td>NCT03632447</td>
<td>Comparing Use of a Digital Health System of Pelvic Floor Exercise Program to Kegel Exercises in Stress Urinary Incontinence</td>
<td>October 2018 to March 2019</td>
<td>RCT</td>
<td>225</td>
<td>Home-based PFMT alone; Home-based PFMT + probe-based biofeedback, delivered via smartphone</td>
<td>Cure, Improvement (complete or partial reduction in number of UI episodes, as recorded in bladder diary); QOL (UDI-6)</td>
</tr>
<tr>
<td>NCT02318251</td>
<td>Stress Urinary Incontinence Physiotherapy</td>
<td>January 2015 to July 2019</td>
<td>RCT</td>
<td>96</td>
<td>PFMT; involuntary PFM contractions from exercise</td>
<td>Cure (negative pad test); QOL (ICIQ LUTSsql)</td>
</tr>
<tr>
<td>NCT03722719</td>
<td>The Knack on Female Stress Urinary Incontinence</td>
<td>March 2018 to April 2020</td>
<td>RCT</td>
<td>210</td>
<td>PFMT; The Knack; PFMT + the Knack</td>
<td>Cure (negative pad test, subjective report of cure); Improvement (partial reduction in number of UI episodes, as recorded in bladder diary); QOL (IQOL)</td>
</tr>
<tr>
<td>NCT03203798</td>
<td>Effects of Training of Pelvic Floor Muscles (MAP) on Stress Urinary Incontinence</td>
<td>September 2017 to August 2020</td>
<td>RCT</td>
<td>42</td>
<td>PFMT; Hipopressive abdominal exercises</td>
<td>Cure, Improvement (complete or partial reduction in number of UI episodes, as recorded in bladder diary)</td>
</tr>
<tr>
<td>NCT03296462</td>
<td>Hip External Rotation Physical Therapy Trial</td>
<td>November 2016 to March 2019</td>
<td>RCT</td>
<td>30</td>
<td>PFMT; Hip external rotation</td>
<td>Cure, Improvement (complete or partial reduction in number of UI episodes, as recorded in bladder diary); QOL (UDI-6)</td>
</tr>
<tr>
<td>NCT03166150</td>
<td>Muscle and Functional Assessment in Leakage Study</td>
<td>July 2017 to April 2019</td>
<td>RCT</td>
<td>200</td>
<td>PFMT; PFMT + non-PFMT exercise</td>
<td>Cure, Improvement (complete or partial reduction in number of UI episodes, as recorded in bladder diary)</td>
</tr>
<tr>
<td>NCT03911362</td>
<td>A Comparison of Lumbopelvic Stabilisation and Pelvic Floor Exercises on the Stress Incontinence</td>
<td>12/1/2018 to 8/31/2019</td>
<td>RCT</td>
<td>NR</td>
<td>Lumbopelvic Stabilisation Exercises; Pelvic Floor Exercises</td>
<td>Improvement (number of UI episodes, as recorded in bladder diary); QOL (UDI, IIQ)</td>
</tr>
</tbody>
</table>

PCORI Evidence Map: Pelvic Floor Muscle Training for Urinary Incontinence on Women
<table>
<thead>
<tr>
<th>NCT Number</th>
<th>Title</th>
<th>Start Date and Projected End Date</th>
<th>Study Design</th>
<th>Number of Patients</th>
<th>Intervention(s)</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCT03862326</td>
<td>Pilot Study - Pelvic Strength</td>
<td>8/1/2016 to 5/1/2019</td>
<td>RCT</td>
<td>35</td>
<td>Group PFMT; Individual PFMT; Individual PFMT with biofeedback</td>
<td>QOL (UDI-6, IIQ-7)</td>
</tr>
<tr>
<td>NCT03969368</td>
<td>rPMS Compared With Pelvic Floor Exercises for Treatment of Urinary Incontinence</td>
<td>6/1/2019 to 1/30/2021</td>
<td>RCT</td>
<td>150</td>
<td>rPMS Pelvic Floor Stimulation Device; Pelvic floor exercises</td>
<td>Improvement (number of UI episodes, as recorded in bladder diary), QOL (UDI-6, ICIQ-lutsQOL)</td>
</tr>
<tr>
<td>NCT04084340</td>
<td>tDCS and Female Urinary Incontinence</td>
<td>11/14/2019 to 11/30/2021</td>
<td>RCT</td>
<td>60</td>
<td>Real transcranial direct current stimulation plus pelvic exercises; Sham transcranial direct current stimulation plus pelvic exercises;</td>
<td>Improvement (number of UI episodes, as recorded in bladder diary), QOL (ISI, ICIQ UI short form, KHQ)</td>
</tr>
</tbody>
</table>

BF – Biofeedback; CIQ LUTS qol – International Consultation on Incontinence Questionnaire-Urinary Incontinence Lower Urinary Tract Symptoms Quality of Life; NR – Not reported; PFMT – Pelvic floor muscle training; QOL – Quality of life; RCT – Randomized controlled trial; UDI – Urogenital Distress Inventory; UI – Urinary incontinence