GRADE tables for review question: What is the effectiveness of pelvic floor muscle training (including Kegel exercises, biofeedback, weighted vaginal cones, and electrical stimulation) for improving symptoms of pelvic floor dysfunction?

PFMT versus no treatment/usual care/treatment

Table 7: Clinical evidence profile for comparison: PFMT versus no treatment (or inactive control) for POP

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
Hagen 2011 (SR of RCTs): Self-reported no improvement in prolapse												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	7/19 (36.8%)	16/21 (76.2%)	RR 0.48 (0.26 to 0.91)	396 fewer per 1000 (from 69 fewer to 564 fewer)	MODERATE	CRITICAL
Hagen 2011 (SR of RCTs): Prolapse symptom score (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[2]	none	17	20	-	MD 3.37 lower (6.23 to 0.51 lower)	MODERATE	CRITICAL
Hagen 2011 (SR of RCTs): Prolapse interference with everyday life (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	19	21	-	MD 0.05 lower (0.67 lower to 0.57 higher)	HIGH	CRITICAL
Hagen 2011 (SR of RCTs): increased bother due to bowel emptying difficulty												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6]	none	11/25 (44.0%)	7/15 (46.7%)	RR 0.94 (0.47 to 1.90)	28 fewer per 1000 (from 247 fewer to 420 more)	LOW	CRITICAL
Hagen 2011 (SR of RCTs): increased bother due to flatus leakage												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	16/34 (47.1%)	18/23 (78.3%)	RR 0.68 (0.46 to 0.99)	250 fewer per 1000 (from 423 fewer to 8 fewer)	MODERATE	CRITICAL
Hagen 2011 (SR of RCTs): increased bother due to loose faecal incontinence												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	5/14 (35.7%)	10/10 (100%)	RR 0.38 (0.20 to 0.76)	620 fewer per 1000 (from 800 fewer to 240 fewer)	HIGH	CRITICAL
Hagen 2011 (SR of RCTs): increased bother due to solid faecal incontinence												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious imprecision[6]	none	1/3 (33.3%)	1/2 (50%)	RR 0.67 (0.08 to 5.54)	165 fewer per 1000 (from 460 fewer to 1000 more)	LOW	CRITICAL
Hagen 2011 (SR of RCTs): Ditrovie quality of life score (Better indicated by lower values)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	27	20	-	MD 0.95 lower (1.57 to 0.34 lower)	MODERATE	CRITICAL
Hagen 2011 (SR of RCTs): Satisfaction with treatment (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	27	20	-	MD 3.22 lower (3.79 to 2.65 lower)	MODERATE	IMPORTANT
Hagen 2011 (SR of RCTs): POP-Q stage not improved												
2	randomised trials	very serious[4]	serious[5]	no serious indirectness	serious[1]	none	53/69 (76.8%)	55/59 (93.2%)	RR 0.83 (0.71 to 0.96)	158 fewer per 1000 (from 37 fewer to 270 fewer)	VERY LOW	CRITICAL
Hagen 2011 (SR of RCTs): ICIQ (change score) (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	19	20	-	MD 1.79 lower (3.68 lower to 0.1 higher)	HIGH	CRITICAL
Hagen 2011 (SR of RCTs): Mean bladder symptom score (Better indicated by lower values)												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
1	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	27	20	-	MD 9.22 lower (10.68 to 7.76 lower)	MODERATE	CRITICAL
Ge 2020 (SR of RCTs): Self-reported change in symptoms (better)												
5	randomised trials	serious[3]	very serious[7]	no serious indirectness	no serious imprecision	none	-	-	RR 2.90 (1.72 to 4.89)	-	VERY LOW	CRITICAL
Ge 2020 (SR of RCTs): Self-reported change in symptoms (same)												
4	randomised trials	serious[3]	very serious[7]	no serious indirectness	serious[1]	none	-	-	RR 0.7 (0.45 to 1.09)	-	VERY LOW	CRITICAL
Ge 2020 (SR of RCTs): Self-reported change in symptoms (worse)												
4	randomised trials	serious[3]	very serious[7]	no serious indirectness	very serious[6]	none	-	-	RR 0.67 (0.22 to 2.03)	-	VERY LOW	CRITICAL
Ge 2020 (SR of RCTs): POP-SS (Better indicated by lower values)												
5	randomised trials	serious[3]	very serious[7]	no serious indirectness	no serious imprecision	none	-	-	-	SMD 0.24 lower (0.71 lower to 0.22 higher)	VERY LOW	CRITICAL
Ge 2020 (SR of RCTs): POPDI-6 (Better indicated by lower values)												
4	randomised trials	serious[3]	very serious[7]	no serious indirectness	no serious imprecision	none	-	-	-	SMD 0.14 lower (0.43 lower to 0.15 higher)	VERY LOW	CRITICAL
Ge 2020 (SR of RCTs): CRADI-8 (Better indicated by lower values)												
4	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-	-	-	SMD 0.03 lower (0.16 lower to 0.11 higher)	MODERATE	CRITICAL
Ge 2020 (SR of RCTs): UDI-6 (Better indicated by lower values)												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
4	randomised trials	serious[3]	serious[5]	no serious indirectness	no serious imprecision	none	0	-	-	SMD 0.17 lower (0.43 lower to 0.1 higher)	LOW	CRITICAL
RCT: Recurrence of POP symptoms (final score; 6 months)												
Nyhus 2020	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[6]	none	13/71 (18.3%)	16/73 (21.9%)	RR 0.84 (0.43 to 1.61)	35 fewer per 1000 (from 125 fewer to 134 more)	VERY LOW	CRITICAL
RCT: Sensation of vaginal bulge (final scores; vas 0-100; 6 months) (Better indicated by lower values)												
Nyhus 2020	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	73	75	-	MD 1.4 higher (4.02 lower to 6.82 higher)	LOW	CRITICAL
RCT: Improvement in POP symptoms (final score; 6 months)												
Nyhus 2020	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	62/69 (89.9%)	68/72 (94.4%)	RR 0.95 (0.86 to 1.05)	47 fewer per 1000 (from 132 fewer to 47 more)	LOW	CRITICAL
RCT: POPDI (final score; high score is poor outcome; 60 days post surgery) (Better indicated by lower values)												
Liang 2019	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	47	43	-	MD 1.32 lower (3 lower to 0.36 higher)	LOW	CRITICAL
RCT: CRADI-8 (final score; high score is poor outcome; 60 days post surgery) (Better indicated by lower values)												
Liang 2019	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	47	43	-	MD 0.57 lower (3.14 lower to 2 higher)	LOW	CRITICAL
RCT: UDI-6 (final score; high score is poor outcome; 60 days post surgery) (Better indicated by lower values)												
Liang 2019	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	47	43	-	MD 5.66 lower (9.85 to 1.47 lower)	LOW	CRITICAL
RCT: PFDI-20 (final score; high score is poor outcome; 60 days post surgery) (Better indicated by lower values)												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
Liang 2019	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	47	43	-	MD 7.55 lower (13.9 to 1.2 lower)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 95% CI crosses 1 MID (0.8, 1.25)
2 95% CI crosses 1 MID (0.5 x SD control, 1.45)
3 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
4 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
5 Serious heterogeneity unexplained by subgroup analysis
6 95% CI crosses 2 MIDs (0.8, 1.25)
7 Very serious heterogeneity unexplained by subgroup analysis

Table 8: Clinical evidence profile for comparison: PFMT versus no treatment (or inactive control) for SUI

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
Dumoulin 2018 (SR of RCTs): Patient perceived cure after treatment (treatment duration 3 to 6 months)												
4	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	46/82 (56.1%)	5/83 (6.0%)	RR 8.38 (3.68 to 19.07)	445 more per 1000 (from 161 more to 1000 more)	HIGH	CRITICAL
Dumoulin 2018 (SR of RCTs): Patient perceived cure or improvement after treatment (treatment duration 3 to 6 months)												
3	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	88/119 (73.9%)	14/123 (11.4%)	RR 6.33 (3.88 to 10.33)	607 more per 1000 (from 328 more to 1000 more)	MODERATE	CRITICAL
Dumoulin 2018 (SR of RCTs): Quality of life (King's Health Questionnaire/general health score) (Better indicated by lower values)												
3	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	80	65	-	MD 1.81 higher	MODERATE	CRITICAL

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
										(3.4 lower to 7.03 higher)		
Dumoulin 2018 (SR of RCTs): Participant perceived satisfaction												
2	randomised trials	serious[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36/51 (70.6%)	7/54 (13.0%)	RR 5.32 (2.63 to 10.74)	560 more per 1000 (from 211 more to 1000 more)	MODERATE	IMPORTANT
Imamura 2010 (SR of RCTs): Cure rate												
8	randomised trials	very serious[1]	serious[3]	no serious indirectness	no serious imprecision	none	70/308 (22.7%)	20/297 (6.7%)	OR 5.41 (1.64 to 17.82)	214 more per 1000 (from 39 more to 495 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Improvement rate												
11	randomised trials	very serious[1]	very serious[4]	no serious indirectness	no serious imprecision	none	263/361 (72.9%)	128/337 (38%)	OR 11.75 (3.49 to 39.55)	498 more per 1000 (from 301 more to 581 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Quality of life (Social Activity Index) (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	25	30	-	MD 0.80 higher (0.08 to 1.52 higher)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Quality of life (Norwegian version of the Quality of Life Scale) (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[6]	none	25	30	-	MD 4.9 higher (0.8 lower to 10.60 higher)	VERY LOW	CRITICAL
Moroni 2016 (SR of RCTs): Incontinence specific QoL (Better indicated by lower values)												
2	randomised trials	very serious[1]	no serious inconsistency	Serious[2]	no serious imprecision	none	34	33	-	MD 1.24 lower (1.77 to 0.71 lower)	VERY LOW	CRITICAL
RCT: Improvement in ICIQ sum score (12 weeks)												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
Al-Belushi 2020	randomised trials	serious[7]	no serious inconsistency	serious[8]	no serious imprecision	none	17/36 (47.2%)	2/37 (5.4%)	RR 8.74 (2.17 to 35.13)	418 more per 1000 (from 63 more to 1000 more)	LOW	CRITICAL
RCT: Improved or cured (follow-up 12 weeks)												
Okayama 2019	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	23/31 (74.2%)	7/28 (25%)	RR 2.97 (1.51 to 5.82)	493 more per 1000 (from 127 more to 1000 more)	LOW	CRITICAL
RCT: Cured (follow-up 12 weeks)												
Okayama 2019	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	17/31 (54.8%)	5/28 (17.9%)	RR 3.07 (1.3 to 7.23)	370 more per 1000 (from 54 more to 1000 more)	LOW	CRITICAL
RCT: UI episodes/week (follow-up 12 weeks; Better indicated by lower values)												
Okayama 2019	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[9]	none	30	31	-	Median 1.5 lower Median (IQR): PFMT 0.0(0.0-2.0) Control 1.5(1.0-3.0)	VERY LOW	CRITICAL
ICIQ-SF score (follow-up 12 weeks; Better indicated by lower values)												
Okayama 2019	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[9]	none	30	31	-	Median 1.0 lower Median (IQR): PFMT 5.0(1.0-7.0) Control 6.0(4.3-10.0)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review

1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

2 Serious indirectness as comparison includes one study where the intervention is PFMT + BF rather than PFMT alone

3 Serious heterogeneity unexplained by subgroup analysis

4 Very serious heterogeneity unexplained by subgroup analysis

5 95% CI crosses 1 MID (0.5 x SD control, 0.84)
6 95% CI crosses 1 MID (0.5 x SD control, 6.025)
7 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
8 Serious indirectness as comparison group attended a lecture on PFMT rather than receiving no treatment
9 Subjective assessment

Table 9: Clinical evidence profile for comparison PFMT versus no treatment (or inactive control) for UI (SUI or MUI/not reported/UI or OAB)

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
Dumoulin 2018 (SR of RCTs). Patient perceived cure after treatment (treatment duration 3 to 6 months)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	50/144 (34.7%)	9/146 (6.2%)	RR 5.34 (2.78 to 10.26)	268 more per 1000 (from 110 more to 571 more)	MODERATE	CRITICAL
Dumoulin 2018 (SR of RCTs). Patient perceived cure or improvement after treatment (treatment duration 3 to 6 months)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	58/86 (67.4%)	23/80 (28.7%)	RR 2.39 (1.64 to 3.47)	400 more per 1000 (from 184 more to 710 more)	MODERATE	CRITICAL
Dumoulin 2018 (SR of RCTs). Participant-perceived satisfaction												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	45/58 (77.6%)	14/50 (28.0%)	RR 2.77 (1.74 to 4.41)	496 more per 1000 (from 207 more to 955 more)	MODERATE	IMPORTANT
Nie 2017 (SR of RCTs): IIQ7 (Better indicated by lower values)												
2	randomised trials	serious[1]	very serious[2]	serious[4]	no serious imprecision	none	76	80	-	SMD 2.20 lower (4.12 to 0.27 lower)	VERY LOW	CRITICAL
Nie 2017 (SR of RCTs): ICIQ (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	serious[4]	no serious imprecision	none	24	24	-	SMD 1.05 lower (1.65 to 0.44 lower)	LOW	CRITICAL

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	No treatment	Relative (95% CI)	Absolute		
Nie 2017 (SR of RCTs): UDI (Better indicated by lower values)												
2	randomised trials	no serious risk of bias	no serious inconsistency	serious[4]	no serious imprecision	none	76	80	-	MD 7.5 lower (10.41 to 4.58 lower)	MODERATE	CRITICAL
Nie 2017 (SR of RCTs): Quality of life (The General QoL Questionnaire; Incontinence Quality of Life Questionnaire) (Better indicated by higher values												
2	randomised trials	no serious risk of bias	very serious[2]	serious[4]	no serious imprecision[3]	none	51	54	-	SMD 1.67 higher (0.41 to 2.94 higher)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 Very serious heterogeneity unexplained by subgroup analysis
3 Based on 0.5 x control group SD as two different measures were used therefore published MIDs based on a single measure could not be used
4 Serious indirectness due to unclear comparison. Inclusion criteria included PFMT alone or with pamphlet guidance vs no treatment or pamphlet guidance only but no further details given on specific comparison included

Table 10: Clinical evidence profile for comparison: PFMT (antenatal) vs no treatment for faecal/urinary incontinence

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (antenatal)	No treatment	Relative (95% CI)	Absolute		
Woodley 2020 (SR of RCTs): UDI-6 late pregnancy (for treatment or prevention) (range of scores: 0-100; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	150	150	-	MD 1.22 lower (1.96 to 0.48 lower)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): UDI-6 at 0-3 months post-partum (for treatment or prevention) (range of scores: 0-100; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	150	150	-	MD 0.73 lower (1.06 to 0.40 lower)	VERY LOW	CRITICAL

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (antenatal)	No treatment	Relative (95% CI)	Absolute		
Woodley 2020 (SR of RCTs): UDI-6 at >3-6 months post-partum (for treatment or prevention) (range of scores: 0-100; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	150	150	-	MD 0.51 lower (0.74 to 0.28 lower)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): IIQ7 late pregnancy (for treatment or prevention) (range of scores: 0-100; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	150	150	-	MD 1.51 lower (2.78 to 0.24 lower)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): IIQ7 at 0-3 months post-partum (for treatment or prevention) (range of scores: 0-100; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	150	150	-	MD 3.55 lower (4.61 to 2.49 lower)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): IIQ7 at >3-6 months post-partum (for treatment or prevention) (range of scores: 0-100; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	150	150	-	MD 0.79 lower (1.27 to 0.31 lower)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 Serious indirectness due to comparison group ('No PFMT' which included regular antenatal care rather than no treatment)

Table 11: Clinical evidence profile for comparison: PFT (antenatal) versus usual care for faecal/urinary incontinence

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (antenatal)	Usual care	Relative (95% CI)	Absolute		
Woodley 2020 (SR of RCTs): Incontinence-specific QoL (for treatment) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	20	21	-	MD 3.5 lower (6.13 to 0.87 lower)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (antenatal)	Usual care	Relative (95% CI)	Absolute		
Woodley 2020 (SR of RCTs): Incontinence-specific QoL late pregnancy (for treatment or prevention) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	112	112	-	MD 0.2 lower (1.21 lower to 0.81 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Incontinence-specific QoL early postnatal period (for treatment or prevention) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	104	107	-	MD 0.6 lower (1.45 lower to 0.25 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Incontinence-specific QoL late postnatal period (for treatment or prevention) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	93	97	-	MD 0.2 lower (1.2 lower to 0.8 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ bladder score in late pregnancy (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	112	111	-	MD 0.3 lower (0.65 lower to 0.05 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ bladder score at 0-3 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	105	107	-	MD 0.1 lower (0.36 lower to 0.16 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ bladder score at >6-12 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	94	97	-	MD 0.1 lower (0.41 to 0.12 lower)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ bowel score in late pregnancy (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	112	112	-	MD 0.1 lower (0.39 to 0.19 lower)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ bowel score at 0-3 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (antenatal)	Usual care	Relative (95% CI)	Absolute		
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	104	107	-	MD 0.2 lower (0.52 lower to 0.12 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ bowel score at >6-12 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	94	97	-	MD 0.1 lower (0.38 lower to 0.18 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ prolapse score in late pregnancy (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	112	112	-	MD 0 higher (0.34 lower to 0.34 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ prolapse score at 0-3 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	104	107	-	MD 0.2 lower (0.52 lower to 0.12 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): FPFQ prolapse score at >6-12 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	95	97	-	MD 0 higher (0.31 lower to 0.31 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Female Pelvic Floor Questionnaire sex score in late pregnancy (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	79	68	-	MD 0.9 lower (1.54 to 0.26 lower)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): Female Pelvic Floor Questionnaire sex score at 0-3 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	73	77	-	MD 0.4 lower (1.09 lower to 0.29 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Female Pelvic Floor Questionnaire sex score at >6-12 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (antenatal)	Usual care	Relative (95% CI)	Absolute		
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	86	83	-	MD 0.3 lower (0.87 lower to 0.27 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Contilife score in late pregnancy (for treatment or prevention) (range of scores: 0-10; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	108	109	-	MD 0.1 higher (1.54 to 0.26 lower)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): Contilife score at 0-3 months postpartum (for treatment or prevention) (range of scores: 0-10; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	102	101	-	MD 0.1 higher (0.12 lower to 0.32 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Contilife score at >6-12 months (for treatment or prevention) (range of scores: 0-10; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	91	89	-	MD 0 higher (0.32 lower to 0.32 higher)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Sexually active in late pregnancy (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	83/112 (74.1%)	70/112 (62.5%)	RR 1.19 (0.99 to 1.42)	119 more per 1000 (from 6 fewer to 262 more)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): Sexually active at 0-3 months postpartum (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	74/104 (71.2%)	79/106 (74.5%)	RR 0.95 (0.81 to 1.13)	37 fewer per 1000 (from 142 fewer to 97 more)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Sexually active at >6-12 months (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	89/95 (93.7%)	91/97 (93.8%)	RR 1 (0.93 to 1.07)	0 fewer per 1000 (from 66 fewer to 66 more)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): EQ5D in late pregnancy (for treatment or prevention) (range of scores: 0-100; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	111	112	-	MD 1.5 lower (6.35 lower to 3.35 higher)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (antenatal)	Usual care	Relative (95% CI)	Absolute		
Woodley 2020 (SR of RCTs): EQ5D at 0-3 months postpartum (for treatment or prevention) (range of scores: 0-100; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	105	107	-	MD 2.4 higher (2.34 lower to 7.14 higher)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): EQ5D at >6-12 months (for treatment or prevention) (range of scores: 0-100; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	94	97	-	MD 3.9 higher (0.06 lower to 7.86 higher)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): BFLUTS questionnaire: a negative effect on exercise in response to question "does incontinence affect physical activity?"												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	none	47/585 (8%)	41/584 (7%)	RR 1.14 (0.76 to 1.71)	10 more per 1000 (from 17 fewer to 50 more)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): STAI - trait anxiety (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	none	18/85 (21.2%)	20/76 (26.3%)	RR 0.8 (0.46 to 1.40)	53 fewer per 1000 (from 142 fewer to 105 more)	VERY LOW	IMPORTANT
Woodley 2020 (SR of RCTs): STAI - state anxiety (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[7]	none	16/85 (18.8%)	14/76 (18.4%)	RR 1.02 (0.53 to 1.95)	4 more per 1000 (from 87 fewer to 175 more)	VERY LOW	IMPORTANT
Woodley 2020 (SR of RCTs): Sexual satisfaction at 6 years post-delivery (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	34/94 (36.2%)	17/94 (18.1%)	RR 2 (1.2 to 3.32)	181 more per 1000 (from 36 more to 420 more)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): Psychological General Well-being Index (for treatment or prevention) (range of scores: 0-110; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	389	361	-	MD 0.71 higher (0.6 lower to 2.01 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.5 x SD control, 2.8)

3 95% CI crosses 1 MID (0.5 x SD control, 1.05)
4 95% CI crosses 1 MID (0.5 x SD control, 0.65)
5 95% CI crosses 1 MID (0.8, 1.25)
6 95% CI crosses 2 MIDs (EQ5D 0.025)
7 95% CI crosses 2 MIDs (0.8, 1.25)

Table 12: Clinical evidence profile for comparison: PFMT (postnatal) versus usual care for faecal/urinary incontinence

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (postnatal)	Usual care	Relative (95% CI)	Absolute		
Woodley 2020 (SR of RCTs): Incontinence specific QoL (PFMT for treatment) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	9	9	-	MD 1.66 lower (3.51 lower to 0.19 higher)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): Urinary symptoms (BFLUTS) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	9	9	-	MD 42.83 lower (47.06 to 38.61 lower)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): HADS (for treatment) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	238	219	-	MD 0.79 lower (1.43 to 0.05 lower)	LOW	IMPORTANT
Woodley 2020 (SR of RCTs): Sexual function (attempted sexual intercourse within 3 months of delivery) (PFMT for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	714/819 (87.2%)	681/792 (86%)	RR 1.01 (0.98 to 1.05)	9 more per 1000 (from 17 fewer to 43 more)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): Sexual function (dyspareunia within 3 months post-partum) (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	167/819 (20.4%)	154/792 (19.4%)	RR 1.05 (0.86 to 1.28)	10 more per 1000 (from 27 fewer to 54 more)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): ICIQ-Vag, bulging inside vagina (yes/no) (for treatment or prevention)												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (postnatal)	Usual care	Relative (95% CI)	Absolute		
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	8/87 (9.2%)	22/88 (25%)	RR 0.37 (0.17 to 0.78)	157 fewer per 1000 (from 55 fewer to 207 fewer)	LOW	CRITICAL
Woodley 2020 (SR of RCTs): ICIQ-Vag, bulging outside vagina (yes/no) (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	5/87 (5.7%)	6/88 (6.8%)	RR 0.84 (0.27 to 2.66)	11 fewer per 1000 (from 50 fewer to 113 more)	VERY LOW	CRITICAL
Woodley 2020 (SR of RCTs): POP-Q stage 1 or 2 (for treatment or prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	61/87 (70.1%)	64/88 (72.7%)	RR 0.88 (0.46 to 1.7)	87 fewer per 1000 (from 393 fewer to 509 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.5 x SD control, 1.05)
3 95% CI crosses 1 MID (0.8, 1.25)
4 95% CI crosses 2 MIDs (0.8, 1.25)

Table 13: Clinical evidence profile for comparison: PFMT (postnatal) versus no treatment for faecal/urinary incontinence

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (postnatal)	No treatment	Relative (95% CI)	Absolute		
Woodley 2020 (SR of RCTs): Quality of life - sexual function (reduced vaginal response at 10 months post-partum) (for treatment of prevention)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[3]	serious[2]	none	5/51 (9.8%)	13/56 (23.2%)	RR 0.42 (0.16 to 1.10)	135 fewer per 1000 (from 195 fewer to 23 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference

1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)
3 Serious indirectness due to comparison group ('No PFMT' which included usual postnatal care)

Table 14: Clinical evidence profile for comparison: Magnetic stimulation versus placebo/sham for SUI

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Magnetic stimulation	Placebo	Relative (95% CI)	Absolute		
Peng 2019 (SR of RCTs): Quality of life[2] (follow-up 1 week-14 months; Better indicated by higher values)												
3	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	59	53	-	MD 0.42 higher (0.02 to 0.82 higher)	MODERATE	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 95% CI crosses 1 MID (0.5 x control SD, 0.5)
2 Specific measures used in studies not reported.

Table 15: Clinical evidence profile for comparison: Magnetic stimulation versus placebo/sham for UI

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Magnetic stimulation	Sham	Relative (95% CI)	Absolute		
Lim 2015 (SR of RCTs): Improved incontinence												
3	randomised trials	serious risk of bias[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	65/84 (77.4%)	22/69 (31.9%)	RR 2.29 (1.60 to 3.29)	411 more per 1000 (from 191 more to 730 more)	MODERATE	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

Table 16: Clinical evidence profile for comparison: Vaginal cones versus no treatment for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vaginal cones	No treatment	Relative (95% CI)	Absolute		
Imamura 2010 (SR of RCTs): Improvement rate												
2	randomised trials	very serious[1]	very serious[2]	no serious indirectness	very serious[3]	none	68/106 (64.2%)	54/105 (51.4%)	OR 5.43 (0.07 to 396.77)	338 more per 1000 (from 445 fewer to 483 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Quality of life - Social Activity Index (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	27	30	-	MD 0.3 higher (0.42 lower to 1.02 higher)	VERY LOW	CRITICAL
Herbinson 2013 (SR of RCTs): No subjective improvement or cure												
2	randomised trials	very serious[1]	very serious[2]	serious[5]	serious[6]	none	38/106 (35.8%)	55/109 (50.5%)	RR 0.72 (0.52 to 0.99)	141 fewer per 1000 (from 5 fewer to 242 fewer)	VERY LOW	CRITICAL
Herbinson 2013 (SR of RCTs): No subjective cure												
4	randomised trials	serious[7]	very serious[2]	serious[5]	serious[6]	none	115/151 (76.2%)	190/224 (84.8%)	RR 0.84 (0.76 to 0.94)	136 fewer per 1000 (from 51 fewer to 204 fewer)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR : systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 Very serious heterogeneity unexplained by subgroup analysis
3 95% CI crosses 2 MIDs (0.8, 1.25)
4 95% CI crosses 1 MID (0.5 x control group SD, 0.84)
5 Serious indirectness as control groups included interventions other than no treatment
6 95% CI crosses 1 MID (0.8, 1.25)
7 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

Table 17: Clinical evidence profile for comparison: Vaginal cones versus no treatment for post-natal UI (not specified)

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Vaginal cones	No treatment	Relative (95% CI)	Absolute		
Oblasser 2015 (SR of RCTs): Self-reported urinary incontinence (follow-up 12 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	10/21 (47.6%)	69/91 (75.8%)	RR 0.63 (0.4 to 0.998)	281 fewer per 1000 (from 2 fewer to 455 fewer)	VERY LOW	CRITICAL
Oblasser 2015 (SR of RCTs): Self-reported urinary incontinence (follow-up after 24-44 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	13/19 (68.4%)	20/37 (54.1%)	RR 1.27 (0.83 to 1.94)	146 more per 1000 (from 92 fewer to 508 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)

Table 18: Clinical evidence profile for comparison: Electrical stimulation versus no treatment for SUI

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electrical stimulation	No treatment	Relative (95% CI)	Absolute		
Imamura 2010 (SR of RCTs): Cure rate												
6	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	9/152 (5.9%)	8/136 (5.9%)	OR 1.10 (0.41 to 2.94)	6 more per 1000 (from 34 fewer to 96 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Improvement rate												
7	randomised trials	very serious[1]	serious[3]	no serious indirectness	no serious imprecision	none	71/192 (37%)	23/177 (13%)	OR 3.93 (1.43 to 10.8)	240 more per 1000 (from 46 more to 487 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Incontinence specific QoL (Social Activity Index; IIQ) (change score) (Better indicated by higher values)												
2	randomised trials	very serious[1]	serious[3]	no serious indirectness	no serious imprecision	none	37	42	-	SMD 0.47 higher (0.02 to 0.92 higher)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): UDI (change score) (Better indicated by lower values)												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electrical stimulation	No treatment	Relative (95% CI)	Absolute		
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	12	12	-	MD 8.5 lower (18.65 lower to 1.65 higher)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure (follow-up mean 6 months)												
2	randomised trials	serious[5]	no serious inconsistency	serious[6]	serious[7]	none	18/52 (34.6%)	6/49 (12.2%)	RR 2.31 (1.06 to 5.02)	160 more per 1000 (from 7 more to 492 more)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure or improvement (follow-up 6 weeks to 9 months)												
5	randomised trials	very serious[1]	very serious[10]	serious[6]	no serious imprecision	none	110/174 (63.2%)	66/173 (38.2%)	RR 1.73 (1.41 to 2.11)	278 more per 1000 (from 156 more to 423 more)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Quality of life (KHQ; ICIQ) (follow-up median 6 weeks; Better indicated by lower values)												
6	randomised trials	very serious[1]	very serious[10]	serious[6]	no serious imprecision	none	110113	117	-	SMD 0.72 lower (0.99 to 0.46 lower)	VERY LOW	CRITICAL
Moroni 2016 (SR of RCTs): Incontinence-specific QoL - KHQ; IQoL (intravaginal stimulation) (Better indicated by lower values)												
2	randomised trials	serious[5]	serious[3]	serious[8]	no serious imprecision	none	42	39	-	SMD 1.44 lower (1.94 to 0.95 lower)	LOW	CRITICAL
Moroni 2016 (SR of RCTs): Incontinence-specific QoL - KHQ (superficial stimulation) (Better indicated by lower values)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	22	22	-	MD 50.1 lower (66.77 to 34.25 lower)	LOW	CRITICAL
RCT: UDI-6 (final score; high score is poorer outcome; 8 weeks) (Better indicated by lower values)												
Hwang 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	16	16	-	MD 9 lower (19.11 lower to 1.11 higher)	VERY LOW	CRITICAL
RCT: PISQ - total score (final score; high score is better outcome; 8 weeks) (Better indicated by higher values)												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electrical stimulation	No treatment	Relative (95% CI)	Absolute		
Hwang 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[9]	none	16	16	-	MD 10.88 higher (0.75 to 21.01 higher)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (0.8, 1.25)
3 Serious heterogeneity unexplained by subgroup analysis
4 95% CI crosses 1 MID (UDI, -14)
5 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
6 Serious indirectness due to no treatment groups groups including other interventions
7 95% CI crosses 1 MID (0.8, 1.25)
8 Serious indirectness due to the Castro study control group being 'no active treatment'
9 95% CI crosses 1 MID (PISQ, 6)
10 Very serious heterogeneity unexplained by subgroup analysis

Table 19: Clinical evidence profile for comparison: Electrical stimulation versus no treatment for OAB

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electrical stimulation	No treatment for OAB	Relative (95% CI)	Absolute		
RCT: ICIQ-OAB (final score; high score is poor outcome; 5 weeks) (Better indicated by lower values)												
Teixeira Alve 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	63	25	-	MD 4.92 lower (6.35 to 3.49 lower)	LOW	CRITICAL
RCT: Adherence												
Teixeira Alve 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	63/72 (87.5%)	25/29 (86.2%)	RR 1.01 (0.86 to 1.2)	9 more per 1000 (from 121 fewer to 172 more)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

Table 20: Clinical evidence profile for comparison: Electrical stimulation versus sham for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electrical stimulation	Sham	Relative (95% CI)	Absolute		
Stewart 2017 (SR of RCTs): Subjective cure												
3	randomised trials	serious[1]	serious[2]	no serious indirectness	very serious[3]	none	32/95 (33.7%)	6/63 (9.5%)	RR 2.21 (0.38 to 12.73)	115 more per 1000 (from 59 fewer to 1000 more)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure or improvement												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	71/145 (49%)	18/91 (19.8%)	RR 2.03 (1.02 to 4.07)	204 more per 1000 (from 4 more to 607 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 Serious inconsistency due to significant heterogeneity (I2 = 62%, p=0.07)
3 Confidence intervals cross 2 MIDs (0.8, 1.25)
4 Confidence intervals cross 1 MID (0.8, 1.25)

Table 21: Clinical evidence profile for comparison: PFMT versus electrical stimulation for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	Electrical stimulation	Relative (95% CI)	Absolute		
Imamura 2010 (SR of RCTs): Cure rates												
5	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15/62 (24.2%)	7/62 (11.3%)	OR 2.65 (0.82 to 8.6)	139 more per 1000 (from 18 fewer to 410 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Improvement rates												
6	randomised trials	very serious[1]	serious[3]	no serious indirectness	very serious[7]	none	69/92 (75%)	57/98 (58.2%)	OR 2.18 (0.76 to 6.28)	170 more per 1000 (from 68 fewer to 316 more)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	Electrical stimulation	Relative (95% CI)	Absolute		
Imamura 2010 (SR of RCTs): Social Activity Index (change score) (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	25	25	-	MD 0 higher (0.57 lower to 0.57 higher)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure												
4	randomised trials	very serious[1]	serious[3]	no serious indirectness	serious[2]	none	36/71 (50.7%)	21/72 (29.2%)	RR 1.75 (1.15 to 2.68)	219 more per 1000 (from 44 more to 490 more)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure or improvement												
7	randomised trials	very serious[1]	serious[3]	no serious indirectness	serious[2]	none	79/118 (66.9%)	73/126 (57.9%)	RR 1.18 (0.97 to 1.43)	104 more per 1000 (from 17 fewer to 249 more)	VERY LOW	CRITICAL
Liang 2018 (SR of RCTs): Life quality score (ICI-Q-SF; lower better)												
17[5]	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[6]	none	-	-	-	MD 6.96 lower (from 10.2 lower to 3.72 lower)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review

1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

2 95% CI crosses 1 MID (0.8, 1.25)

3 Serious heterogeneity unexplained by subgroup analysis

4 95% CI crosses 2 MIDs (0.5 x control group SD, 0.51)

5 Number of studies in total NMA

6 95% CI crosses 1 MID (ICIQ-SF, 4)

7 95% CI crosses 2 MIDs (0.8, 1.25)

Table 22: Clinical evidence profile for comparison: PFMT versus vaginal cones for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	Vaginal cones	Relative (95% CI)	Absolute		
Herbison 2013 (SR of RCTs): No subjective improvement or cure												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	73/180 (40.6%)	68/178 (38.2%)	RR 1.03 (0.8 to 1.33)	11 more per 1000 (from 76 fewer to 126 more)	VERY LOW	CRITICAL
Herbison 2013 (SR of RCTs): No subjective cure												
5	randomised trials	serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	128/169 (75.7%)	129/169 (76.3%)	RR 0.99 (0.88 to 1.12)	8 fewer per 1000 (from 92 fewer to 92 more)	LOW	CRITICAL
Imamura 2010 (SR of RCTs): Cure rate												
3	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	6/121 (5%)	11/124 (8.9%)	OR 0.61 (0.09 to 3.95)	33 fewer per 1000 (from 80 fewer to 189 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Improvement rate												
5	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[4]	none	110/167 (65.9%)	108/164 (65.9%)	OR 1.01 (0.52 to 1.95)	2 more per 1000 (from 158 fewer to 131 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Incontinence specific QoL (Social Activity Index; KHQ) (change score)												
2	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	41	57	-	SMD 0.32 higher (0.08 lower to 0.73 higher)	LOW	CRITICAL
Moroni 2016 (SR of RCTs): Incontinence-specific QoL (KHQ; IQoL) (Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	39	39	-	MD 0.56 lower (8.4 lower to 7.28 higher)	MODERATE	CRITICAL
Liang 2018 (SR of RCTs): Life quality score (ICI-Q-SF) (Better indicated by lower values)												
17[5]	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-	-	-	MD 0.01 higher (2.62 lower to 2.64 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review

1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 Serious heterogeneity unexplained by subgroup analysis
3 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
4 95% CI crosses 2 MIDs (0.8, 1.25)
5 This is the total number of studies in the NMA

Table 23: Clinical evidence profile for comparison: PFMT versus vaginal cones for post-natal UI (not specified)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT	Vaginal cones	Relative (95% CI)	Absolute		
Oblasser 2015 (SR of RCTs): Self-reported urinary incontinence (follow-up 12 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	10/21 (47.6%)	9/19 (47.4%)	RR 1.01 (0.52 to 1.93)	5 more per 1000 (from 227 fewer to 441 more)	VERY LOW	CRITICAL
Oblasser 2015 (SR of RCTs): Self-reported urinary incontinence (follow-up after 24-44 months)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	13/19 (68.4%)	10/20 (50%)	RR 1.37 (0.8 to 2.33)	185 more per 1000 (from 100 fewer to 665 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (0.8,1.25)

Table 24: Clinical evidence profile for comparison: PFMT + biofeedback versus electrical stimulation for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + Biofeedback	Electrical stimulation	Relative (95% CI)	Absolute		
Liang 2018 (SR of RCTs): Life quality score (Better indicated by lower values)												
17[1]	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	-	-	-	MD 7.12 lower (3.16 to 11.08 lower)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 This is the number of studies included in the overall NMA
2 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
3 95% CI crosses 1 MID (ICIQ-SF, 4)

Table 25: Clinical evidence profile for comparison: Electrical stimulation versus vaginal cones for SUI

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electrical stimulation	Vaginal cones	Relative (95% CI)	Absolute		
Herbison 2013 (SR of RCTs): No subjective cure or improvement after treatment												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	28/79 (35.4%)	32/72 (44.4%)	RR 0.8 (0.54 to 1.18)	89 fewer per 1000 (from 204 fewer to 80 more)	VERY LOW	CRITICAL
Herbison 2013 (SR of RCTs): No subjective cure or improvement after 6 months												
3	randomised trials	very serious[1]	very serious[3]	no serious indirectness	serious[2]	none	42/81 (51.9%)	49/73 (67.1%)	RR 0.77 (0.59 to 1.01)	154 fewer per 1000 (from 275 fewer to 7 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Cure rates												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	5/55 (9.1%)	4/51 (7.8%)	OR 1 (0.26 to 3.91)	0 fewer per 1000 (from 57 fewer to 171 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Cure rates (long term >1 year)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	12/30 (40%)	10/24 (41.7%)	OR 0.93 (0.31 to 2.78)	18 fewer per 1000 (from 235 fewer to 248 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Improvement rates												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	55/71 (77.5%)	50/70 (71.4%)	OR 1.3 (0.59 to 2.84)	50 more per 1000 (from 118 fewer to 162 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Improvement rates (long term >1 year)												

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electrical stimulation	Vaginal cones	Relative (95% CI)	Absolute		
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	17/30 (56.7%)	17/24 (70.8%)	OR 0.54 (0.17 to 1.68)	141 fewer per 1000 (from 416 fewer to 95 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Social Activity Index (change score) (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	25	27	-	MD 0.5 higher (0.07 lower to 1.07 higher)	VERY LOW	CRITICAL
Moroni 2016 (SR of RCTs): Incontinence specific QoL (Better indicated by lower values)												
2	randomised trials	serious[6]	very serious[3]	no serious indirectness	very serious[7]	none	51	45	-	MD 9.31 higher (2.77 to 15.86 higher)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure												
3	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[4]	none	30/82 (36.6%)	25/75 (33.3%)	RR 1.04 (0.7 to 1.54)	13 more per 1000 (from 100 fewer to 180 more)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure or improvement												
5	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	no serious imprecision	none	140/172 (81.4%)	119/159 (74.8%)	RR 1.09 (0.97 to 1.21)	67 more per 1000 (from 22 fewer to 157 more)	MODERATE	CRITICAL
Stewart 2017 (SR of RCTs): I-QoL (Better indicated by higher values)												
2	randomised trials	serious[6]	no serious inconsistency	no serious indirectness	very serious[7]	none	51	45	-	MD 1.59 higher (3.72 lower to 6.9 higher)	VERY LOW	CRITICAL
Liang 2018 (SR of RCTs): Life quality score (ICI-Q-SF; lower better)												
17[8]	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[9]	none	-	-	-	MID 6.97 higher (3.74 to 10.21 higher)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)
3 Very serious heterogeneity unexplained by subgroup analysis

4 95% CI crosses 2 MIDs (0.8, 1.25)
5 95% CI crosses 1 MID (0.5 x control group SD, 0.53)
6 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
7 95% CI crosses 2 MIDs (I-QoL, 2.5)
8 This is the number of studies included in the overall NMA
9 95% CI crosses 2 MIDs (ICIQ-SF, 4)

Table 26: Clinical evidence profile for comparison: Electrical stimulation versus PTNS for OAB

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Electical stimulation	Transcutaneous posterior tibial nerve stimulation	Relative (95% CI)	Absolute		
RCT: Quality of life (King's Health Questionnaire - symptoms domain; final score; 6 weeks) (Better indicated by lower values)												
Mallmann 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	21	25	-	MD 1.4 higher (1.81 lower to 4.61 higher)	LOW	CRITICAL
RCT: Incontinence Severity Index (6 weeks) - Mild												
Mallmann 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/21 (14.3%)	6/25 (24%)	RR 0.6 (0.17 to 2.1)	96 fewer per 1000 (from 199 fewer to 264 more)	VERY LOW	CRITICAL
RCT: Incontinence Severity Index (6 weeks) - Moderate												
Mallmann 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	14/21 (66.7%)	11/25 (44%)	RR 1.52 (0.89 to 2.59)	229 more per 1000 (from 48 fewer to 700 more)	VERY LOW	CRITICAL
RCT: Incontinence Severity Index (6 weeks) - Severe												
Mallmann 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	4/21 (19%)	8/25 (32%)	RR 0.6 (0.21 to 1.7)	128 fewer per 1000 (from 253 fewer to 224 more)	VERY LOW	CRITICAL
RCT: Incontinence Severity Index (6 weeks) - Very severe												
Mallmann 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/21 (0%)	0/25 (0%)	Not estimable	-	LOW	CRITICAL

RCT: Quality of life (King's Health Questionnaire – total score; final score; 6-8 weeks) (Better indicated by higher values)												
Gungor Urgurlucan 2013	randomised trials	serious[4]	no serious inconsistency	no serious indirectness	very serious[4]	none	35	17	-	MD 66.80 lower (187.61 lower to 54.01 higher)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (0.8, 1.25)
3 95% CI crosses 1 MID (0.8, 1.25)
4 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
5 95% CI crosses 2 MIDs (KHQ, 10-15 for medium effect)

Table 27: Clinical evidence profile for comparison: Vaginal cones versus PFMT + biofeedback for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Vaginal cones**	**PFMT + biofeedback**	**Relative (95% CI)**	**Absolute**		
Liang 2018 (SR of RCTs): Life quality score (Better indicated by lower values)												
17[1]	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	-	-	-	MD 0.14 higher (3.34 lower to 3.62 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 This is the number of studies included in the overall NMA
2 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

Variations of PFMT

Table 28: Clinical evidence profile for comparison: PFMT (more) versus PFMT (less) for UI (SUI/MUI)

Quality assessment						No of patients			Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (more)	PFMT (less)	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not cured (more vs less contact with health professionals: additional group supervision)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	43/52 (82.7%)	55/59 (93.2%)	RR 0.89 (0.78 to 1.03)	103 fewer per 1000 (from 205 fewer to 28 more)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not cured (more vs less contact with health professionals: individual supervision vs no supervision)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	26/31 (83.9%)	32/33 (97%)	RR 0.86 (0.73 to 1.02)	136 fewer per 1000 (from 262 fewer to 19 more)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not improved (more vs less contact with health professionals: additional group supervision)												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	9/87 (10.3%)	39/90 (43.3%)	RR 0.29 (0.15 to 0.55)	308 fewer per 1000 (from 195 fewer to 368 fewer)	MODERATE	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not improved (more vs less contact with health professionals: individual supervision vs no supervision)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	1/31 (3.2%)	11/33 (33.3%)	RR 0.1 (0.01 to 0.71)	300 fewer per 1000 (from 97 fewer to 330 fewer)	MODERATE	CRITICAL
HaySmith 2011 (SR of RCTs): Quality of Life Index ("How would you feel if you had to spend the rest of your life with the same urinary problem") (more vs less contact with health professionals: additional group supervision) (Better indicated by lower values)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	12	10	-	MD 1.9 lower (2.93 to 0.87 lower)	LOW	CRITICAL
HaySmith 2011 (SR of RCTs): Symptom impact index (Chinese version) - avoiding activities due to worry about leaking (more vs less contact with health professionals: individual supervision vs no supervision)												

Quality assessment						No of patients			Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (more)	PFMT (less)	Relative (95% CI)	Absolute		
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	8/31 (25.8%)	15/31 (48.4%)	RR 0.53 (0.27 to 1.07)	227 fewer per 1000 (from 353 fewer to 34 more)	LOW	CRITICAL
HaySmith 2011 (SR of RCTs): Symptom impact index (Chinese version) - avoiding activities due needing a toilet (more vs less contact with health professionals: individual supervision vs no supervision)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	7/31 (22.6%)	16/31 (51.6%)	RR 0.44 (0.21 to 0.91)	289 fewer per 1000 (from 46 fewer to 408 fewer)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)
3 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

Table 29: Clinical evidence profile for comparison: PFMT (more) versus PFMT (less) for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (more)	PFMT (less)	Relative (95% CI)	Absolute		
Imamura 2010 (SR of RCTs): Cure rate (PFMT with additional sessions vs PFMT)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	25/58 (43.1%)	9/60 (15%)	OR 8.81 (2.33 to 33.27)	459 more per 1000 (from 141 more to 704 more)	LOW	CRITICAL
Imamura 2010 (SR of RCTs): Improvement rate (PFMT with additional sessions vs PFMT)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	34/35 (97.1%)	21/39 (53.8%)	OR 20.74 (3.58 to 120.25)	422 more per 1000 (from 268 more to 454 more)	LOW	CRITICAL
Imamura 2010 (SR of RCTs): Cure rate (long term >1 year) (PFMT with additional sessions vs PFMT)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	6/20 (30%)	4/25 (16%)	OR 2.25 (0.54 to 9.44)	140 more per 1000 (from 67 fewer to 483 more)	VERY LOW	CRITICAL
Imamura 2010 (SR of RCTs): Incontinence specific quality of life (Social Activity Index; quality of life index) (PFMT with additional sessions vs PFMT)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	35	39	-	SMD 0.12 higher (0.37 lower to 0.61 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (0.8, 1.25)

Table 30: Clinical evidence profile for comparison: PFMT (group) versus PFMT (individual) for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**PFMT (group)**	**PFMT (individual)**	**Relative (95% CI)**	**Absolute**		
Moroni 2016 (SR of RCTs): Incontinence-specific QoL (KHQ) (Better indicated by lower values)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	45	45	-	MD 7.96 higher (2.69 lower to 18.60 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (KHQ, 10-15 for medium effect)

Table 31: Clinical evidence profile for comparison: PFMT (group) vs PFMT (individual) for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**PFMT (group)**	**PFMT (individual)**	**Relative (95% CI)**	**Absolute**		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not cured (individual supervision only vs individual and group supervision)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (group)	PFMT (individual)	Relative (95% CI)	Absolute		
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	43/52 (82.7%)	55/59 (93.2%)	RR 0.89 (0.78 to 1.03)	103 fewer per 1000 (from 205 fewer to 28 more)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved (individual and group supervision vs individual supervision)												
3	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	3/64 (4.7%)	23/69 (33.3%)	RR 0.16 (0.05 to 0.46)	280 fewer per 1000 (from 180 fewer to 317 fewer)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved (group supervision vs individual supervision)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	12/30 (40%)	10/30 (33.3%)	RR 1.2 (0.61 to 2.34)	67 more per 1000 (from 130 fewer to 447 more)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Quality of Life Index ("How would you feel if you had to spend the rest of your life with the same urinary problem") (Better indicated by lower values)												
1	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	12	10	-	MD 1.9 lower (2.93 to 0.87 lower)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): KHQ (incontinence impact) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	30	30	-	MD 6.7 higher (5.91 lower to 19.31 higher)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): KHQ (severity) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	30	30	-	MD 0.9 higher (9.37 lower to 11.17 higher)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): IQoL (change in total score) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	29	30	-	MD 13.2 lower (39.2 lower to 12.8 higher)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): IQoL (total score) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	none	123	117	-	MD 5 lower (9.14 to 0.86 lower)	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (group)	PFMT (individual)	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Adherence (participated in >50% of supervised sessions)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	16/84 (19%)	6/92 (6.5%)	RR 2.92 (1.20 to 7.12)	125 more per 1000 (from 13 more to 399 more)	LOW	IMPORTANT
Hay-Smith 2011 (SR of RCTs): Adherence (did not attend any sessions)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	11/84 (13.1%)	12/92 (13%)	RR 1 (0.47 to 2.15)	0 fewer per 1000 (from 69 fewer to 150 more)	VERY LOW	IMPORTANT
Hay-Smith 2011 (SR of RCTs): Adherence (no exercise at home)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	100/123 (81.3%)	86/117 (73.5%)	RR 1.11 (0.96 to 1.27)	81 more per 1000 (from 29 fewer to 198 more)	LOW	IMPORTANT
RCT: PGI-I - perceived benefit (1 year)												
Dumoulin 2020	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	144/166 (86.7%)	146/171 (85.4%)	RR 1.02 (0.93 to 1.11)	17 more per 1000 (from 60 fewer to 94 more)	LOW	CRITICAL
RCT: Satisfaction (1 year)												
Dumoulin 2020	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	150/165 (90.9%)	154/171 (90.1%)	RR 1.01 (0.94 to 1.08)	9 more per 1000 (from 54 fewer to 72 more)	LOW	CRITICAL
RCT: KHQ - severity (final score; high score is poor outcome; 6 months) (Better indicated by lower values)												
Figueiredo 2020	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	very serious[8]	none	30	30	-	MD 1.4 lower (11.52 lower to 8.72 higher)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)
3 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
4 95% CI crosses 2 MIDs (0.8, 1.25)
5 95% CI crosses 1 MID (KHQ, 10-15 for medium effect)
6 95% CI crosses 2 MIDs (I-QoL, 2.5)

7 95% CI crosses 1 MID (I-QoL, 2.5)
8 95% CI crosses 1 MID (KHQ, 5-6 for small effect)

Table 32: Clinical evidence profile for comparison: PFMT (direct) versus PFMT (indirect) for UI (SUI or MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (direct)	PFMT (indirect)	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not cured (PFMT vs Sapsford approach)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32/33 (97%)	26/31 (83.9%)	RR 1.16 (0.98 to 1.36)	134 more per 1000 (from 17 fewer to 302 more)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved (PFMT vs sham/imitation PFMT)												
2	randomised trials	very serious[3]	no serious inconsistency	no serious indirectness	serious[2]	none	25/71 (35.2%)	34/67 (50.7%)	RR 0.69 (0.47 to 1.02)	157 fewer per 1000 (from 269 fewer to 10 more)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved (PFMT vs Sapsford approach)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	11/33 (33.3%)	1/31 (3.2%)	RR 0 (1.42 to 75.41)	32 fewer per 1000 (from 14 more to 1000 more)	MODERATE	CRITICAL
Hay-Smith 2011 (SR of RCTs): I-QoL (change in total score) (PFMT vs Paula method) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	29	30	-	MD 13.2 lower (39.2 lower to 12.8 higher)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): I-QoL (total score) (PFMT vs Paula method) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	123	117	-	MD 5 lower (9.14 to 0.86 lower)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Adherence (participated in <50% of supervised sessions) (PFMT vs Paula method)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	16/84 (19%)	6/92 (6.5%)	RR 2.92 (1.2 to 7.12)	125 more per 1000 (from 13 more to 399 more)	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (direct)	PFMT (indirect)	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Adherence (did not attend any supervision sessions) (PFMT vs Paula method)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	11/84 (13.1%)	12/92 (13%)	RR 1 (0.47 to 2.15)	0 fewer per 1000 (from 69 fewer to 150 more)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Adherence (documented no exercise at home) (PFMT vs Paula method)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	100/123 (81.3%)	86/117 (73.5%)	RR 1.11 (0.96 to 1.27)	81 more per 1000 (from 29 fewer to 198 more)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Symptom impact index (Chinese version) - avoiding activities due to worry about leaking (PFMT vs Sapsford approach)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15/31 (48.4%)	8/31 (25.8%)	RR 1.88 (0.93 to 3.77)	227 more per 1000 (from 18 fewer to 715 more)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Symptom impact index (Chinese version) - avoiding activities due to needing a toilet (PFMT vs Sapsford approach)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	16/31 (51.6%)	7/31 (22.6%)	RR 1.43 (0.62 to 3.27)	97 more per 1000 (from 86 fewer to 513 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference: SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)
3 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
4 95% CI crosses 2 MIDs (I-QoL, 2.5)
5 95% CI crosses 1 MID (I-QoL, 2.5)
6 95% CI crosses 2 MIDs (0.8, 1.25)

Table 33: Clinical evidence profile for comparison: PFMT (individualised) versus PFMT (generic) for UI (SUI/MUI)

Quality assessment							Number of participants		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Individualised PFMT	Generic PFMT	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	10/30 (33.3%)	12/30 (40%)	RR 0.83 (0.43 to 1.63)	68 fewer per 1000 (from 228 fewer to 252 more)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): KHQ (incontinence impact) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	30	30	-	MD 6.7 lower (19.31 lower to 5.91 higher)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): KHQ (severity) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	30	30	-	MD 0.90 lower (11.17 lower to 9.37 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (0.8, 1.25)
3 95% CI crosses 1 MID (KHQ, 10-15 for medium effect)

Table 34: Clinical evidence profile for comparison: PFMT (daily) vs PFMT (3x per week) for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (daily)	PFMT (3x per week)	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not cured												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	16/19 (84.2%)	15/21 (71.4%)	RR 1.18 (0.84 to 1.65)	129 more per 1000 (from 114 fewer to 464 more)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (daily)	PFMT (3x per week)	Relative (95% CI)	Absolute		
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/19 (0%)	0/21 (0%)	-	-	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)

Table 35: Clinical evidence profile for comparison: PFMT (upright and supine) vs PFMT (supine) for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (upright and supine)	PFMT (supine)	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Incontinence-specific quality of life (IIQ) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	19	17	-	MD 2.9 lower (23.78 lower to 17.98 higher)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Treatment adherence (number of clinic visits) (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	22	22	-	MD 0.5 higher (1.21 lower to 2.21 higher)	VERY LOW	IMPORTANT

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (IIQ, 16)
3 95% CI crosses 1 MID (0.5 x control SD, 1.4)

Table 36: Clinical evidence profile for comparison: PFMT (more intensive) vs PFMT (less intensive) for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (more intensive)	PFMT (less intensive)	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not cured (high contrast)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	69/83 (83.1%)	87/92 (94.6%)	RR 0.89 (0.8 to 0.98)	104 fewer per 1000 (from 19 fewer to 189 fewer)	MODERATE	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not cured (low contrast)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	148/161 (91.9%)	126/143 (88.1%)	RR 1.06 (1 to 1.13)	53 more per 1000 (from 0 more to 115 more)	MODERATE	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved (high contrast)												
6	randomised trials	serious[1]	serious[2]	no serious indirectness	serious[3]	none	29/166 (17.5%)	68/169 (40.2%)	RR 0.37 (0.17 to 0.84)	253 fewer per 1000 (from 64 fewer to 334 fewer)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved (moderate contrast)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	6/23 (26.1%)	16/21 (76.2%)	RR 0.34 (0.17 to 0.71)	503 fewer per 1000 (from 221 fewer to 632 fewer)	MODERATE	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change in incontinence - not improved (low contrast)												
7	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	serious[3]	none	50/212 (23.6%)	78/193 (40.4%)	RR 0.75 (0.59 to 0.95)	101 fewer per 1000 (from 20 fewer to 166 fewer)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review

1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

2 Serious heterogeneity unexplained by subgroup analysis

3 95% CI crosses 1 MID (0.8, 1.25)

4 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

Table 37: Clinical evidence profile for comparison: PFMT (app based) vs PFMT (written) for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (app based)	PFMT (written) for UI	Relative (95% CI)	Absolute		
RCT: Adherence (Number of protocol repetitions; final score; 3 months) (Better indicated by higher values)												
Araujo 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	12	9	-	MD 26.1 higher (19.64 to 32.56 higher)	LOW	CRITICAL
RCT: Adherence (Self-reported adherence; final score; 3 months) (Better indicated by higher values)												
Araujo 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	12	9	-	MD 1.23 higher (0.37 to 2.09 higher)	VERY LOW	CRITICAL
RCT: QUID (final score; 3 months) (Better indicated by lower values)												
Araujo 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	12	9	-	MD 3.6 higher (2.01 lower to 9.21 higher)	VERY LOW	CRITICAL
RCT: ICIQ-UI SF (final score; 3 months) (Better indicated by lower values)												
Araujo 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	12	9	-	MD 0.6 lower (6.3 lower to 5.1 higher)	VERY LOW	CRITICAL
RCT: ICIQ-Vaginal Symptoms (final score; 3 months) (Better indicated by lower values)												
Araujo 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	12	9	-	MD 0.8 higher (4.84 lower to 6.44 higher)	VERY LOW	CRITICAL
RCT: ICIQ - Sexual function (final score; 3 months) (Better indicated by lower values)												
Araujo 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	12	9	-	MD 5.5 higher (6.53 lower to 17.53 higher)	VERY LOW	CRITICAL
RCT: ICIQ - QoL (final score; 3 months) (Better indicated by lower values)												
Araujo 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	12	9	-	MD 4.3 higher (1.22 to 7.38 higher)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference

1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.5 x control group SD, 0.65)
3 95% CI crosses 1 MID (0.5 x control group SD, 3.7)
4 95% CI crosses 2 MIDs (ICIQ-SF, 4)
5 95% CI crosses 1 MID (ICIQ-SF, 4)

Table 38: Clinical evidence profile for comparison: PFMT (outpatient) vs PFMT (home) for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**PFMT (outpatient)**	**PFMT (home) for SUI**	**Relative (95% CI)**	**Absolute**		
RCT: I-QoL - avoidance and limiting behaviour (final score; high score is good outcome; 3 months) (Better indicated by higher values)												
Fitz 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	28	-	MD 1.1 higher (15.48 lower to 17.68 higher)	VERY LOW	CRITICAL
RCT: I-QoL - psychosocial impacts (final score; high score is good outcome; 3 months) (Better indicated by higher values)												
Fitz 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	28	-	MD 7.8 lower (26.5 lower to 10.9 higher)	VERY LOW	CRITICAL
RCT: I-QoL - social embarrassment (final score; high score is good outcome; 3 months) (Better indicated by higher values)												
Fitz 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	28	28	-	MD 10 lower (24.19 lower to 4.19 higher)	VERY LOW	CRITICAL
RCT: Adherence (3 months) (Better indicated by higher values)												
Fitz 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	28	28	-	MD 6.9 higher (1.22 lower to 15.02 higher)	VERY LOW	CRITICAL
RCT: Patient satisfaction (3 months)												
Fitz 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	24/34 (70.6%)	18/35 (51.4%)	RR 1.37 (0.93 to 2.02)	190 more per 1000 (from 36 fewer to 525 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (I-QoL, 2.5)

3 95% CI crosses 1 MID (0.5 x control group SD, 9.9)
4 95% CI crosses 1 MID (0.8, 1.25)

Table 39: Clinical evidence profile for comparison: PFMT + BF vs PFMT for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
Liang 2018 (SR of RCTs): Life quality score (ICIQ-SF) (follow-up 4-24 weeks; Better indicated by lower values)												
17[1]	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0	-	-	MD 0.15 lower (2.43 lower to 2.12 higher)	LOW	CRITICAL
Imanura 2010 (SR of RCTs): cure rates												
8	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	87/179 (48.6%)	64/191 (33.5%)	OR 1.88 (1.23 to 2.86)	151 more per 1000 (from 48 more to 255 more)	LOW	CRITICAL
Imanura 2010 (SR of RCTs): improvement rates												
7	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[3]	none	119/139 (85.6%)	120/157 (76.4%)	OR 1.83 (1.01 to 3.34)	91 more per 1000 (from 2 more to 151 more)	VERY LOW	CRITICAL
Imanura 2010 (SR of RCTs): Quality of life (Social Activity Index) (follow-up 6 months; Better indicated by higher values)												
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[4]	none	36	34	-	MD 0.1 higher (0.22 lower to 0.42 higher)	VERY LOW	CRITICAL
Imanura 2010 (SR of RCTs): Quality of life (Modified PRAFAB) (Better indicated by lower values)												
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[5]	none	20	20	-	MD 2.00 lower (6.57 lower to 2.57 higher)	VERY LOW	CRITICAL
Imanura 2010 (SR of RCTs): Quality of life (Kings Health Questionnaire; change score) (Better indicated by lower values)												
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[6]	none	22	16	-	MD 1.99 lower (7.13 lower to 3.15 higher)	VERY LOW	CRITICAL
Imanura 2010 (SR of RCTs): Quality of life (Incontinence Impact Questionnaire; change score) (Better indicated by lower values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
1	randomised trials	very serious[2]	no serious inconsistency	no serious indirectness	serious[7]	none	10	7	-	MD 16 lower (30.7 to 1.3 lower)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Number of studies in total NMA
2 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
3 95% CI crosses 1 MID (0.8, 1.25)
4 95% CI crosses 1 MID (0.5x control group SD, 0.37)
5 95% CI crosses 1 MID (0.5x control group SD, 4.3)
6 95% CI crosses 1 MID (KHQ, 5-6 for small effect)
7 95% CI crosses 1 MID (IIQ, 16)

Table 40: Clinical evidence profile for comparison: PFMT + BF vs PFMT for UI (UUI/MUI/SUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
Herderschee 2011 (SR of RCTs): Quality of life (Protection, Amount, Frequency, Adjustment, Body Image; PRAFAB, short version) (no difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	20	20	-	MD 0.27 lower (0.89 lower to 0.36 higher)	MODERATE	CRITICAL
Herderschee 2011 (SR of RCTs): Quality of life (KHQ total score, change score) (no difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	22	16	-	MD 1.99 lower (4.42 lower to 0.44 higher)	MODERATE	CRITICAL
Herderschee 2011 (SR of RCTs): Quality of life (IIQ, final score) (no difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	10	10	-	MD 41.60 lower (78.62 to 4.58 lower)	LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
Herderschee 2011 (SR of RCTs): Quality of life (KHQ total score, final score) (no difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	11	11	-	MD 4.45 lower (18.64 lower to 9.74 higher)	VERY LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Quality of life (PRAFAB, change score) (no difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	18	15	-	MD 0.36 lower (1.05 lower to 0.33 higher)	MODERATE	CRITICAL
Herderschee 2011 (SR of RCTs): Quality of life (KHQ - incontinence impact) (difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[13]	none	34	34	-	MD 31.39 higher (11.09 lower to 73.89 higher)	VERY LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Quality of life (KHQ - severity measures) (difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	34	34	-	MD 5.94 higher (6.56 lower to 18.44 higher)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Perception of change - not cured or improved (No difference in PFMT)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	58/88 (65.9%)	68/89 (76.4%)	RR 0.87 (0.72 to 1.05)	99 fewer per 1000 (from 214 fewer to 38 more)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Perception of change - not cured or improved (difference in PFMT)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	80/162 (49.4%)	131/181 (72.4%)	RR 0.69 (0.58 to 0.83)	224 fewer per 1000 (from 123 fewer to 304 fewer)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Perception of change - not cured (combined no difference in PFMT and difference in PFMT)												
5	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	108/155 (69.7%)	126/166 (75.9%)	RR 0.92 (0.81 to 1.05)	61 fewer per 1000 (from 144 fewer to 38 more)	MODERATE	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
Herderschee 2011 (SR of RCTs): Women's satisfaction with progress - not satisfied (combined no difference in PFMT and difference in PFMT)												
7	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	39/147 (26.5%)	60/101 (59.4%)	RR 0.65 (0.49 to 0.9)	208 fewer per 1000 (from 59 fewer to 303 fewer)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Symptom distress/Quality of life (UDI - total score) (No difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[6]	none	10	10	-	MD 31.7 lower (80.36 lower to 16.96 higher)	VERY LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Symptom distress/Quality of life (Social activity index) (No difference in PFMT) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[7]	none	48	46	-	MD 0.10 higher (0.18 lower to 0.38 higher)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Anxiety (Hopkins Symptom Checklist - anxiety) (Difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	47	40	-	MD 1.40 lower (6.74 lower to 3.94 higher)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Depression (Hopkins Symptom Checklist - depression) (Difference in PFMT) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[8]	none	47	40	-	MD 2.40 lower (7.59 lower to 2.79 higher)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (adherence to clinical sessions) (no difference in PFMT)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	20/20 (100%)	20/20 (100%)	RR 1.00 (0.91 to 1.1)[9]	0 fewer per 1000 (from 90 fewer to 100 more)	MODERATE	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (adherence to home treatment) (no difference in PFMT)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[10]	none	17/22 (77.3%)	13/16 (81.3%)	RR 0.95 (0.69 to 1.32)[9]	41 fewer per 1000 (from 252 fewer to 260 more)	VERY LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (exercised > 3x per week) (no difference in PFMT)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43/48 (89.6%)	39/46 (84.8%)	RR 1.06 (0.9 to 1.23)[9]	51 more per 1000 (from 85 fewer to 195 more)	MODERATE	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (adherence to exercises - rarely) (no difference in PFMT)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[10]	none	0/15 (0%)	1/22 (4.5%)	RR 0.48 (0.02 to 11.03)	24 fewer per 1000 (from 45 fewer to 456 more)	VERY LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (adherence to exercises - occasionally) (no difference in PFMT)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	5/15 (33.3%)	15/22 (68.2%)	RR 0.49 (0.23 to 1.06)	348 fewer per 1000 (from 525 fewer to 41 more)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (adherence to exercises - frequently)(no difference in PFMT)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	9/15 (60%)	6/22 (27.3%)	RR 2.20 (0.99 to 4.89)	327 more per 1000 (from 3 fewer to 1000 more)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (adherence to exercises - all the time)(no difference in PFMT)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[10]	none	1/15 (6.7%)	0/22 (0%)	RR 4.31 (0.19 to 99.27)	-	VERY LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (participants exercising regularly) (difference in PFMT)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	17/19 (89.5%)	7/14 (50%)	RR 1.79 (1.04 to 3.09)	395 more per 1000 (from 20 more to 1000 more)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Adherence (compliance) (difference in PFMT)												
1	randomised trials	very serious[12]	no serious inconsistency	no serious indirectness	serious[11]	none	19/16 (118.8%)	16/18 (88.9%)	RR 1.12 (0.92 to 1.36)	107 more per 1000 (from 71 fewer to 320 more)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
Herderschee 2011 (SR of RCTs): Follow up data: Symptom distress/Quality of life (UDI - total score at follow up) (No difference in PFMT) (follow-up 24 weeks; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[6]	none	10	9	-	MD 61.70 lower (109.85 to 13.55 lower)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Follow up data: Quality of life (IIQ - total score at follow up) (No difference in PFMT) (follow-up 24 weeks; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	10	9	-	MD 39.10 lower (79.81 lower to 1.61 higher)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Follow up data: Quality of life (KHQ - total score at follow up) (No difference in PFMT) (follow-up 3 months; Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[4]	none	11	11	-	MD 8.18 lower (25.52 lower to 9.16 higher)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Follow up data: Adherence (women still doing PFMT exercise regularly) (difference in PFMT) (follow-up 2-3 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	17/19 (89.5%)	7/14 (50%)	RR 1.79 (1.04 to 3.09)	395 more per 1000 (from 20 more to 1000 more)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Follow up data: Women still subjective cured (difference in PFMT) (follow-up 2-3 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[10]	none	5/19 (26.3%)	0/14 (0%)	RR 8.25 (0.49 to 137.94)	-	VERY LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Follow up data: Women still subjective improved (difference in PFMT) (follow-up 2-3 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	8/19 (42.1%)	4/14 (28.6%)	RR 2.39 (0.99 to 5.79)	397 more per 1000 (from 3 fewer to 1000 more)	LOW	CRITICAL
Herderschee 2011 (SR of RCTs): Follow up data: Subjective cure (difference in PFMT) (follow-up 3 months)												
1	randomised trials	very serious[12]	no serious inconsistency	no serious indirectness	very serious[10]	none	8/13 (61.5%)	19/27 (70.4%)	RR 0.87 (0.53 to 1.43)	91 fewer per 1000 (from 331 fewer to 303 more)	VERY LOW	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
Herderschee 2011 (SR of RCTs): Follow up data: Symptomatic improvement - much better (difference in PFMT) (follow-up 3 months)												
1	randomised trials	very serious[12]	no serious inconsistency	no serious indirectness	very serious[10]	none	3/14 (21.4%)	2/15 (13.3%)	RR 1.61 (0.31 to 8.24)	81 more per 1000 (from 92 fewer to 965 more)	VERY LOW	CRITICAL
RCT: Adherence (number of appointments attended, 0-6) (Better indicated by higher values)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	295	298	-	MD 0.2 higher (0.12 lower to 0.52 higher)	MODERATE	CRITICAL
RCT: ICIQ-UI SF (final score; high is poor outcome; 24 months) (Better indicated by lower values)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	225	235	-	MD 0.3 lower (1.21 lower to 0.61 higher)	MODERATE	CRITICAL
RCT: Cure (Negative response to both "how often do you leak urine?" and "how much urine do you usually leak?"; 24 months)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[10]	none	18/229 (7.9%)	20/238 (8.4%)	RR 0.94 (0.51 to 1.72)	5 fewer per 1000 (from 41 fewer to 61 more)	VERY LOW	CRITICAL
RCT: Improvement (Reduction ICIQ of ≥3 points from baseline; 24 months)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	None	135/225 (60%)	147/235 (62.6%)	RR 0.96 (0.83 to 1.11)	25 fewer per 1000 (from 106 fewer to 69 more)	MODERATE	CRITICAL
RCT: PGI-I (Very much better or much better; 24 months)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[11]	none	93/227 (41%)	90/236 (38.1%)	RR 1.07 (0.86 to 1.35)	27 more per 1000 (from 53 fewer to 133 more)	LOW	CRITICAL
RCT: ICIQ-FLUTS incontinence (final score; high is poor outcome; 24 months) (Better indicated by lower values)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	164	169	-	MD 0.5 higher (0.39 lower to 1.39 higher)	MODERATE	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + BF	PFMT	Relative (95% CI)	Absolute		
RCT: ICIQ-LUTSqol (final score; high is poor outcome; 24 months) (Better indicated by lower values)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	164	169	-	MD 0 higher (2.67 lower to 2.67 higher)	MODERATE	CRITICAL
RCT: Adherence (adherence during clinic appointment - any adherence in clinic)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	231/290 (79.7%)	231/292 (79.1%)	RR 1.01 (0.93 to 1.09)	8 more per 1000 (from 55 fewer to 71 more)	MODERATE	CRITICAL
RCT: ICIQ-FLUTS filling score (final score; high is poor outcome; 24 months) (Better indicated by lower values)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	167	168	-	MD 0.1 lower (0.63 lower to 0.43 higher)	MODERATE	CRITICAL
RCT: ICIQ-FLUTS voiding score (final score; high is poor outcome; 24 months) (Better indicated by lower values)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	165	169	-	MD 0 higher (0.39 lower to 0.39 higher)	MODERATE	CRITICAL
RCT: ICIQ-LUTSqol bother (final score; high is poor outcome; 24 months) (Better indicated by lower values)												
Hagen 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	163	169	-	MD 0.1 higher (0.55 lower to 0.75 higher)	MODERATE	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (IIQ, 16)
3 95% CI crosses 1 MID (KHQ, 5-6 for small effect)
4 95% CI crosses 1 MID (KHQ, 10-15 for medium effect)
5 95% CI crosses 1 MID (0.5 x control group SD, 6.1)
6 95% CI crosses 1 MID (UDI, -14)
7 95% CI crosses 1 MID (0.5 x control group SD, 0.35)
8 95% CI crosses 1 MID (0.5 x control group SD, 6.25)
9 Herdesrschee 2011 did not report RR (only reported % and not effect estimate)
10 95% CI crosses 2 MIDs (0.8, 1.25)
11 95% CI crosses 1 MID (0.8, 1.25)

12 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
13 95% CI crosses 1 MID (KHQ, 10-15 for medium effect)

Table 41: Clinical evidence profile for comparison: PFMT + BF vs PFMT for FI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + biofeedback	PFMT for FI	Relative (95% CI)	Absolute		
RCT: Cleveland score (clinical severity; high score is poorer outcome; 3 months) (Better indicated by lower values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	36	36	-	MD 0.38 lower (2.66 lower to 1.90 higher)	VERY LOW	CRITICAL
RCT: FIQL - lifestyle (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	36	-	MD 0.08 higher (0.22 lower to 0.38 higher)	LOW	CRITICAL
RCT: FIQL - depression (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	36	-	MD 0.02 higher (0.32 lower to 0.36 higher)	LOW	CRITICAL
RCT: FIQL - coping (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	36	36	-	MD 0.13 higher (0.18 lower to 0.44 higher)	VERY LOW	CRITICAL
RCT: FIQL - embarrassment (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	36	36	-	MD 0.07 lower (0.44 lower to 0.3 higher)	VERY LOW	CRITICAL
RCT: EQ5D (high score is good outcome; 3 months) (Better indicated by higher values)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + biofeedback	PFMT for FI	Relative (95% CI)	Absolute		
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	36	36	-	MD 0.07 higher (0.06 lower to 0.2 higher)	VERY LOW	CRITICAL
RCT: ICIQ-UI (low score is good outcome; 3 months) (Better indicated by lower values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	17	13	-	MD 4.32 higher (0.28 lower to 8.92 higher)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 Confidence intervals crossed 1 MID (0.5 x control SD, 2.07)
3 Confidence intervals crossed 1 MID (FIQL, 0.4)
4 Confidence interval crosses 2 MIDs (EQ5D 0.025)
5 Confidence intervals crossed 1 MID (ICIQ-SF, 4)

Table 42: Clinical evidence profile for comparison: PFMT + Feedback vs PFMT for UI (UUI/MUI/SUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + Feedback	PFMT	Relative (95% CI)	Absolute		
Herderschee 2011 (SR of RCTs): Perception of change - not cured or improved												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	21/57 (36.8%)	45/65 (69.2%)	RR 0.53 (0.37 to 0.78)	325 fewer per 1000 (from 152 fewer to 436 fewer)	MODERATE	CRITICAL
Herderschee 2011 (SR of RCTs): Satisfaction with progress - not satisfied												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	8/55 (14.5%)	27/61 (44.3%)	RR 0.33 (0.16 to 0.66)	297 fewer per 1000 (from 150 fewer to 372 fewer)	MODERATE	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

PFMT + treatment versus PFMT alone

Table 43: Clinical evidence profile for comparison: PFMT + VC vs PFMT for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + VC	PFMT	Relative (95% CI)	Absolute		
Imanura 2010 (SR of RCTs): Cure rates												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/21 (23.8%)	3/25 (12%)	OR 2.29 (0.48 to 11.01)	118 more per 1000 (from 59 fewer to 480 more)	VERY LOW	CRITICAL
Imanura 2010 (SR of RCTs): Improvement rates												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	11/21 (52.4%)	12/25 (48%)	OR 1.19 (0.37 to 3.81)	43 more per 1000 (from 225 fewer to 299 more)	VERY LOW	CRITICAL
Herbinson 2013 (SR of RCTs): No subjective improvement or cure (follow-up 6 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	13/21 (61.9%)	11/25 (44%)	RR 1.41 (0.81 to 2.45)	180 more per 1000 (from 84 fewer to 638 more)	LOW	CRITICAL
Herbinson 2013 (SR of RCTs): No subjective improvement or cure (follow-up 12 weeks)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	10/21 (47.6%)	13/25 (52%)	RR 0.92 (0.51 to 1.64)	42 fewer per 1000 (from 255 fewer to 333 more)	VERY LOW	CRITICAL
Herbinson 2013 (SR of RCTs): No subjective cure												
1	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[2]	none	8/14 (57.1%)	9/19 (47.4%)	RR 1.21 (0.63 to 2.32)	99 more per 1000 (from 175 fewer to 625 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (0.8, 1.25)
3 95% CI crosses 1 MID (0.8, 1.25)
4 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

Table 44: Clinical evidence profile for comparison: PFMT + ES vs PFMT for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + ES	PFMT	Relative (95% CI)	Absolute		
Imanura 2010 (SR of RCTs): Cure rates												
4	randomised trials	serious[1]	serious[2]	no serious indirectness	very serious[3]	none	22/108 (20.4%)	22/104 (21.2%)	OR 0.95 (0.49 to 1.85)	8 fewer per 1000 (from 95 fewer to 120 more)	VERY LOW	CRITICAL
Imanura 2010 (SR of RCTs): Improvement rate												
3	randomised trials	very serious[4]	no serious inconsistency	no serious indirectness	very serious[3]	none	68/81 (84%)	65/79 (82.3%)	OR 1.13 (0.49 to 2.58)	17 more per 1000 (from 128 fewer to 100 more)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	9/49 (18.4%)	12/50 (24%)	RR 0.76 (0.38 to 1.52)	58 fewer per 1000 (from 149 fewer to 125 more)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective cure or improvement												
8	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	117/175 (66.9%)	85/133 (63.9%)	RR 1.10 (0.95 to 1.28)	64 more per 1000 (from 32 fewer to 179 more)	LOW	CRITICAL
Stewart 2017 (SR of RCTs): Quality of life (Better indicated by lower values)												
4	randomised trials	serious[1]	very serious[6]	no serious indirectness	no serious imprecision	none	99	94	-	SMD 0.35 lower (0.64 to 0.05 lower)	VERY LOW	CRITICAL
Stewart 2017 (SR of RCTs): Subjective assessment (VAS) (Better indicated by lower values)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	77	73	-	SMD 0.57 lower (0.9 to 0.24 lower)	MODERATE	CRITICAL
RCT: Quality of Life (Wagner's QoL scale; final score; 4 weeks) (Better indicated by lower values)												
Karaman 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	20	28	-	MD 11.1 lower (14.74 to 7.46 lower)	MODERATE	CRITICAL
RCT: UI recurrence (final score; 4 weeks)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + ES	PFMT	Relative (95% CI)	Absolute		
Karaman 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	2/20 (10%)	5/28 (17.9%)	RR 0.56 (0.12 to 2.6)	79 fewer per 1000 (from 157 fewer to 286 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 Serious heterogeneity unexplained by subgroup analysis
3 95% CI crosses 2 MIDs (0.8, 1.25)
4 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
5 95% CI crosses 1 MID (0.8, 1.25)
6 Very serious heterogeneity unexplained by subgroup analysis

Table 45: Clinical evidence profile for comparison: PFMT + ES vs PFMT for UI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + Electrical stimulation	PFMT for FI	Relative (95% CI)	Absolute		
RCT: PISQ (6 months) (Better indicated by lower values)												
Jha 2018	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	30	34	-	MD 5 lower (12.04 lower to 2.04 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (PISQ, 6)

Table 46: Clinical evidence profile for comparison: PFMT + ES vs PFMT for FI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + Electrical stimulation	PFMT for FI	Relative (95% CI)	Absolute		
RCT: Cleveland score (clinical severity; high score is poorer outcome; 3 months) (Better indicated by lower values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	39	36	-	MD 1.61 lower (3.68 lower to 0.46 higher)	VERY LOW	CRITICAL
RCT: FIQL - lifestyle (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	39	36	-	MD 0.15 higher (0.14 lower to 0.44 higher)	VERY LOW	CRITICAL
RCT: FIQL - depression (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	39	36	-	MD 0.18 higher (0.11 lower to 0.47 higher)	VERY LOW	CRITICAL
RCT: FIQL - coping (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	39	36	-	MD 0.21 higher (0.15 lower to 0.57 higher)	VERY LOW	CRITICAL
RCT: FIQL - embarrassment (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	39	36	-	MD 0.08 higher (0.29 lower to 0.45 higher)	VERY LOW	CRITICAL
RCT: EQ5D (high score is good outcome; 3 months) (Better indicated by higher values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[4]	none	39	36	-	MD 0.19 higher (0.08 lower to 0.30 higher)	VERY LOW	CRITICAL
RCT: ICIQ-UI (low score is good outcome; 3 months) (Better indicated by lower values)												
Mundet 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[5]	none	15	17	-	MD 1.89 lower (6.13 lower to 2.35 higher)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference

1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.5 x control group SD, 2.07)
3 95% CI crosses 1 MID (FIQL, 0,4)
4 95% CI crosses 2 MIDs (EQ5D 0.025)
5 95% CI crosses 1 MID (ICIQ-SF, 4)

Table 47: Clinical evidence profile for comparison: PFMT (strength and motor learning) vs PFMT (motor learning alone) for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT (strength and motor learning)	PFMT (motor learning alone)	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not cured												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60/61 (98.4%)	58/62 (93.5%)	RR 1.05 (0.98 to 1.13)	47 more per 1000 (from 19 fewer to 122 more)	MODERATE	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not improved												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	9/61 (14.8%)	14/62 (22.6%)	RR 0.65 (0.31 to 1.4)	79 fewer per 1000 (from 156 fewer to 90 more)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Quality of life (KHQ - incontinence impact) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	60	55	-	MD 10.6 higher (0.9 to 20.4 higher)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Quality of life (KHQ - severity measures) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	57	50	-	MD 6.9 higher (1.6 lower to 15.3 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (0.8, 1.25)
3 95% CI crosses 1 MID (KHQ, 10-15 for medium effect)

Table 48: Clinical evidence profile for comparison: PFMT + abdominal exercise vs PFMT for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + abdominal exercise	PFMT	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not cured												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	15/21 (71.4%)	15/19 (78.9%)	RR 0.9 (0.63 to 1.29)	79 fewer per 1000 (from 292 fewer to 229 more)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not improved												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/21 (0%)	0/19 (0%)	Not estimable[3]	Risk difference 0 higher (9 lower to 9 higher)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 2 MIDs (0.8, 1.25)
3 Hay-Smith 2011 used RR rather than RD and so estimate was 'not estimable'

Table 49: Clinical evidence profile for comparison: PFMT + abdominal exercise vs PFMT for SUI

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT+abdominal exercise	PFMT for SUI	Relative (95% CI)	Absolute		
RCT: ICIQ LUTS QOL (final score; 3 months) (Better indicated by lower values)												
Ptak 2020	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	70	70	-	MD 102.6 lower (131.9 to 73.3 lower)	LOW	CRITICAL
RCT: IIQ (final score; 8 weeks) (Better indicated by lower values)												
Kucukkaya 2020	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 4.5 lower (7.13 to 1.87 lower)	MODERATE	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT+abdominal exercise	PFMT for SUI	Relative (95% CI)	Absolute		
RCT: UDI (final score; 8 weeks) (Better indicated by lower values)												
Kucukkaya 2020	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 7.3 lower (11.36 to 3.24 lower)	MODERATE	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment

Table 50: Clinical evidence profile for comparison: PFMT + abdominal exercise vs PFMT for PFD (UI/POP/FI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT+abdominal exercise	PFMT for PFD (UI/POP/AI)	Relative (95% CI)	Absolute		
RCT: PFDI-20 (Change score; 12 months) (Better indicated by lower values)												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 15.93 higher (2.35 to 29.51 higher)	MODERATE	CRITICAL
RCT: POPDI (Change score; 12 months) (Better indicated by lower values)												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 7.01 higher (1.74 to 12.28 higher)	MODERATE	CRITICAL
RCT: CRADI (Change score; 12 months) (Better indicated by lower values)												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 3.96 higher (0.89 lower to 8.81 higher)	MODERATE	CRITICAL

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT+abdominal exercise	PFMT for PFD (UI/POP/AI)	Relative (95% CI)	Absolute		
RCT: UDI (Change score; 12 months) (Better indicated by lower values)												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 4.8 higher (1.65 lower to 11.25 higher)	MODERATE	CRITICAL
RCT: PFIQ-7 (Change score; 12 months) (Better indicated by lower values)												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	32	-	MD 12.28 higher (2.6 to 21.96 higher)	LOW	CRITICAL
RCT: POPIQ (Change score; 12 months) (Better indicated by lower values)												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 4.86 higher (1.04 to 8.68 higher)	MODERATE	CRITICAL
RCT: CRAIQ (Change score; 12 months) (Better indicated by lower values)												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 4.97 higher (2.18 to 7.76 higher)	MODERATE	CRITICAL
RCT: UIQ (Change score; 12 months) (Better indicated by lower values)												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 2.85 higher (2.91 lower to 8.61 higher)	MODERATE	CRITICAL
RCT: Adherence												
Navarro-Brazalez 2020	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	23/32 (71.9%)	21/32 (65.6%)	RR 1.1 (0.79 to 1.53)	66 more per 1000 (from 138 fewer to 348 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference
1 Serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.5 x control group SD, 21.86)

3 95% CI crosses 2 MIDs (0.8, 1.25)

Table 51: Clinical evidence profile for comparison: PFMT + intravaginal device vs PFMT for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + intravaginal device	PFMT	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not cured												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	57/60 (95%)	53/60 (88.3%)	RR 1.07 (0.96 to 1.2)	62 more per 1000 (from 35 fewer to 177 more)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not improved												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	30/60 (50%)	35/60 (58.3%)	RR 0.86 (0.62 to 1.2)	82 fewer per 1000 (from 222 fewer to 117 more)	VERY LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference: SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)

Table 52: Clinical evidence profile for comparison: PFMT + adherence strategy vs PFMT for UI (SUI/MUI)

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + adherence strategy	PFMT	Relative (95% CI)	Absolute		
Hay-Smith 2011 (SR of RCTs): Patients' perception of change - not improved												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	10/21 (47.6%)	17/20 (85%)	RR 0.56 (0.34 to 0.91)	374 fewer per 1000 (from 76 fewer to 561 fewer)	VERY LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Adherence (did not do routine PFMT)												

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PFMT + adherence strategy	PFMT	Relative (95% CI)	Absolute		
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/41 (0%)	12/34 (35.3%)	RR 0.03 (0 to 0.54)	342 fewer per 1000 (from 162 fewer to 353 fewer)	LOW	CRITICAL
Hay-Smith 2011 (SR of RCTs): Adherence (did not do twice daily PFMT as recommended)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	7/41 (17.1%)	30/34 (88.2%)	RR 0.19 (0.1 to 0.38)	715 fewer per 1000 (from 547 fewer to 794 fewer)	LOW	CRITICAL

CI: confidence interval; MID: minimal important difference; MD: mean difference; OR: odds ratio; RCT: randomised controlled trial; RR: relative risk; SMD: standardised mean difference; SR: systematic review
1 Very serious risk of bias in the evidence contributing to the outcomes as per RoB assessment
2 95% CI crosses 1 MID (0.8, 1.25)

Click here to enter text.