

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p -value	direction of effect	RoB
Pilates vs control													
Eyigor 2010	Breast cancer survivors, female	Pilates vs no intervention*	Quality of life, global	end of treatment (8 wks)	EORTC QLQC30 - QoL scale	higher score means better quality of life	27/15	77.02 (21.81)	63.78 (23.80)	NR	0.79	No difference	High
			QoL, functional	end of treatment (8 wks)	EORTC QLQC30 - functioning subscale	higher score means better functioning	27/15	83.26 (14.70)	78.00 (20.54)	NR	0.33	No difference	High
			Fatigue	end of treatment (8 wks)	Brief fatigue inventory	0=no fatigue; 1-3=mild; 4-6=moderate; 7-10=severe	27/15	5.58 (4.67)	6.55 (4.43)	NR	0.66	No difference	High
			Footnotes:	* All participants in the intervention group (27/27) had final assessment data but 10/25 participants in the control group without final assessment data were excluded from the analysis.									
Martin 2013	Breast cancer survivors, female	Pilates vs no intervention	Footnotes:	No data extracted because the study does not include any priority outcome domains or measures.									
Pilates vs 'other'													
Alpozgen 2017	Breast cancer survivors, female (with upper extremity limitations secondary to breast cancer treatment)	Pilates vs combined exercise	Pain (at rest)	end of treatment (8 weeks)	VAS (0-10)	higher score means worse pain	18/18	0.5 (0.99)	0.22 (0.55)	NR	NR*	No difference	Some concerns
			Functional status, upper extremity (at rest)	end of treatment (8 weeks)	DASH (0-100)	higher score means more disability and severity	18/18	23.77 (13.41)	19.36 (12.03)	NR	NR*	No difference	Some concerns
			Footnotes:	*Authors reported within group measures (before and after). Outcome of overall ANOVA did not reach significance, hence individual between group differences were not assessed.									
Alpozgen 2017	Breast cancer survivors, female (with upper extremity limitations secondary to breast cancer treatment)	Pilates vs home exercise	Pain (at rest)	end of treatment (8 weeks)	VAS (0-10)	higher score means worse pain	18/19	0.5 (0.99)	0.21 (0.71)	NR	NR*	No difference	Some concerns
			Functional status, upper extremity (at rest)	end of treatment (8 weeks)	DASH (0-100)	higher score means more disability and severity	18/19	23.77 (13.41)	32.06 (20.15)	NR	NR*	No difference	Some concerns
			Footnotes:	*Authors reported within group measures (before and after). Outcome of overall ANOVA did not reach significance, hence individual between group differences were not assessed.									

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Gajbhiye 2013	Breast cancer (on treatment)	Pilates vs Conventional therapy + Counseling	Quality of life, global	end of treatment (3 weeks)	PCASEE quality of life scale	higher score means better functioning	15/15	77.13 (3.58)	71.87 (3.76)	NR	NR	No difference	High
			Functional status, upper extremity	end of treatment (3 weeks)	Wingate upper extremity functional assessment questionnaire	higher score means better functioning	15/15	26.2 (2.17)	21.3 (2.05)	NR	NR	No difference	High
			Footnotes:										
Martins 2013	Breast cancer survivors, female	Pilates vs resistance training	No data extracted because the study does not include any priority outcome domains or measures.										
Odynets 2018	Breast cancer survivors, female	Pilates vs water physical therapy	Quality of life, global	end of treatment (12 weeks)	EORTC QLQC30 - global score	higher score means better quality of life	34/34	77.93 (2.42)	65.53 (2.22)	NR	> 0.05	Favours intervention	Some concerns
			Physical function	end of treatment (12 weeks)	EORTC QLQC30 - physical function score	higher score means better functioning	34/34	75.80 (2.25)	71.23 (2.40)	NR	NR	No difference	Some concerns
			Functional status, upper extremity (at rest)	end of treatment (12 weeks)	EORTC QLQ-BR23 arm symptoms	higher score means worse symptoms	34/34	31.90 (2.41)	36.30 (3.08)	NR	NR	No difference	Some concerns
			Pain	end of treatment (12 weeks)	EORTC QLQC30 - pain score	higher score means worse symptoms	34/34	27.93 (3.16)	28.80 (4.36)	0.87	> 0.05	No difference	Some concerns
			Fatigue	end of treatment (12 weeks)	EORTC QLQC30 - fatigue score	higher score means worse symptoms	34/34	30.80 (3.69)	41.07 (2.51)	10.27	< 0.05	Favours intervention	Some concerns
			Footnotes:										
Odynets 2019	Breast cancer survivors, female	Pilates vs water physical therapy	Quality of life, global	end of treatment (12 mths)	FACT-B total score	higher score means better quality of life	44/50	116.95 (2.64)	120.91 (1.26)	0.3	> 0.05	No difference	Some concerns

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Odynets 2019	Breast cancer survivors, female	Pilates vs Yoga	Physical function	end of treatment (12 mths)	FACT-B physical wellbeing subscale	higher score means better wellbeing	44/50	22.02 (0.85)	23.75 (0.49)	0.38	> 0.05	No difference	Some concerns
			Functional status, upper extremity	end of treatment (12 mths)	FACT-B arms symptoms subscale	higher score means better symptoms	44/50	15.87 (0.46)	16.48 (0.23)	0.25	> 0.05	No difference	Some concerns
			Footnotes:	Effect size/point estimate=standardised difference between two means (Cohen's d)									
			Quality of life, global	end of treatment (12 mths)	FACT-B total score	higher score means better quality of life	44/30	116.95 (2.64)	119.23 (3.09)	0.13	> 0.05	No difference	Some concerns
			Physical function	end of treatment (12 mths)	FACT-B physical wellbeing subscale	higher score means better wellbeing	44/30	22.02 (0.85)	23.46 (0.48)	0.34	> 0.05	No difference	Some concerns
			Functional status, upper extremity	end of treatment (12 mths)	FACT-B arms symptoms subscale	higher score means better symptoms	44/30	15.87 (0.46)	15.83 (0.53)	0.01	> 0.05	No difference	Some concerns
Sener 2017	Breast cancer survivors, female (with lymphedema secondary to breast cancer treatment)	Pilates vs core stabilisation and home exercise	Quality of life, global	end of treatment (8 wks)	EORTC QLQBR23 - global score	higher score means worse quality of life	30/30	38.51 (8.42)	38.37 (7.48)	NR	0.94	No difference	High
			Functional status, upper extremity	end of treatment (8 wks)	DASH (0-100)	higher score means more disability and severity	30/30	37.99 (15.02)	32.15 (12.11)	NR	0.39	No difference	High
			Lymphedema*	end of treatment (8 wks)	Circumference (20 cm from proximal nail)	lower score better	30/30	19.80 (3.48)	22.63 (3.76)	NR	0.04	Favours intervention	High
			Lymphedema*	end of treatment (8 wks)	Circumference (30 cm from proximal nail)	lower score better	30/30	26.98 (4.14)	29.73 (4.29)	NR	0.01	Favours intervention	High
			Lymphedema*	end of treatment (8 wks)	Circumference (50 cm from proximal nail)	lower score better	30/30	32.43 (4.2)	36.38 (4.82)	NR	0.01	Favours intervention	High
			Footnotes:	Effect size/point estimate=standardised difference between two means (Cohen's d)									

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Pain	end of treatment (8 wks)	0-10 visual analogue scale	higher score means worse pain	30/30	0.67 (0.84)	0.87 (1.43)	NR	0.51	No difference	High
			Footnotes:	*Authors provide data at 5 cm intervals up to 60 cms. Data reported here correspond to wrist, forearm and upper arm. **Absolute circumference change > 2 cm at 1 or 2 points considered clinically important									
Abbreviations: C, comparison; DASH, Disabilities of the Arm, Shoulder and Hand scale; EORTC QLQC30, European organisation for research and treatment of cancer- quality of life questionnaire; FACT-B, functional assessment of cancer therapy-breast; I, intervention; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Gomes 2018	Prostate cancer (post-prostatectomy urinary incontinence)	Pilates vs Control (waitlist)	QoL, disease specific	end of treatment (10 weeks)	ICIQ-Short Form (0-21)	higher score means worse outcomes	34/35	4.41 (4.96)	8.2 (3.87)	NR	< 0.05	Favours intervention	Some concerns
			Urinary incontinence	end of treatment (10 weeks)	24-hr pad test (pad weight, g)	higher score means worse outcomes	34/35	85.85 (180.6)	72.88 (97.28)	NR	> 0.05	No difference	Some concerns
			Urinary incontinence	end of treatment (10 weeks)	24-hr pad test (% continent, 0 pads)	higher is worse	34/35	NR/34 (58.8)	NR/35 (25.7)	NR	0.005	Favours intervention	Some concerns
			Urinary incontinence	end of treatment (10 weeks)	24-hr pad test (% <8 g)	higher is worse	34/35	NR/34 (47.1)	NR/35 (22.9)	NR	0.034	Favours intervention	Some concerns
			Footnotes:	Type of summary statsitics used for listed outcomes not described, however assumed to be mean +- standard deviation based on other outcomes presented in the study.									
Pedriali 2014	Prostate cancer (post-prostatectomy urinary incontinence)	Pilates vs Control (waitlist)	QoL, disease specific	end of treatment (10 weeks)	ICIQ - Short Form	higher score means worse outcomes	26/31	4.61 (5.3)	8.09 (4)	NR	0.007	Favours intervention	Some concerns
			Urinary incontinence	end of treatment (10 weeks)	24-hr pad test (pad weight, g)	higher score means worse outcomes	26/31	97.65 (20.35)	80.25 (20.86)	NR	0.1	No difference	Some concerns
			Footnotes:										
Pilates vs 'other'													
Gomes 2018	Prostate cancer (post-prostatectomy urinary incontinence)	Pilates vs PFM exercises + AES	QoL, disease specific	end of treatment (10 weeks)	ICIQ - Short Form	higher score means worse outcomes	34/35	4.41 (4.96)	5.77 (4.54)	NR	> 0.05	No difference	Some concerns
			Urinary incontinence	end of treatment (10 weeks)	24-hr pad test (pad weight, g)	higher score means worse outcomes	34/35	85.85 (180.6)	67.4 (131.88)	NR	> 0.05	No difference	Some concerns
			Urinary incontinence	end of treatment (10 weeks)	24-hr pad test (% continent, 0 pads)	higher is worse	34/35	NR/34 (58.8)	NR/35 (54.3)	NR	0.7	No difference	Some concerns
			Urinary incontinence	end of treatment (10 weeks)	24-hr pad test (% <8 g)	higher is worse	34/35	NR/34 (47.1)	NR/35 (34.3)	NR	0.28	No difference	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Footnotes:													
Pedriali 2014	Prostate cancer (post-prostatectomy urinary incontinence)	Pilates vs PFM exercises + AES	QoL, disease specific	end of treatment (10 weeks)	ICIQ - Short Form	higher score means worse outcomes	26/28	4.61 (5.3)	5.6 (4.39)	NR	> 0.05	No difference	Some concerns
			Urinary incontinence	end of treatment (10 weeks)	24-hr pad test (pad weight, g)	higher score means worse outcomes	26/28	97.65 (20.35)	67.14 (12.67)	NR	> 0.05	No difference	Some concerns
Footnotes:													
Abbreviations: AES, anal electrical stimulation; C, Comparator; I, intervention; ICIQ, International Consultation on Incontinence Questionnaire; NR, not reported; PFM, Pelvic floor muscle													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p -value	direction of effect	RoB
Pilates vs control													
Melo 2020	Diabetes, type 2	Pilates vs control (no intervention)	Activities of daily living	end of treatment (12 wks)	GLADM-composite score	Higher score means worse functional autonomy	12/12	27.2 (4.0)	35.3 (4.6)	NR	NR	Favours intervention	Some concerns
			Footnotes:										
Torabian 2013	Diabetes, type 2	Pilates vs control (no intervention)	Depression	end of treatment (8 wks)	GHQ-28 (depression)*	Higher score means more depressed	35/35	6.37 (2.01)	11.29 (2.97)	NR	NR	Favours intervention	High
			Footnotes:	*subscales are not independent of each other and subscores should not be used to indicate specific psychological diagnoses.									
Yucel 2016*	Diabetes, type 2	Pilates vs control (no intervention)	QoL	end of treatment (12 wks)	SF-36-physical (0-100)	Higher score means better QoL	24/21	41 (4)	41 (4)	NR	0.764	No difference	High
			QoL	end of treatment (12 wks)	SF-36-mental (0-100)	Higher score means better QoL	24/21	35 (1)	35 (1)	NR	0.912	No difference	High
			Body composition	end of treatment (12 wks)	BMI (kg/m2)	higher is worse	24/21	32.03 (7.31)	30.36 (7.93)	NR	0.707	No difference	High
			Fatigue	end of treatment (12 wks)	0-10 visual analogue scale	Higher score means more fatigue	24/21	4 (1)	4 (2)	NR	0.645	No difference	High
			Depression	end of treatment (12 wks)	HADS - Depression	Higher score means more depressed	24/21	8 (2)	8 (1)	NR	0.297	No difference	High
			Footnotes:	*authors state in the text that presented data are median (IQR) but the table indicates that the presented values are mean (SD). We have assumed they are mean (SD)									
Pilates vs 'other' - no studies found													
Abbreviations: C, Comparator; GLADM, Group of Latin American Development to Maturity test battery (Includes: 10m walk, rise from sitting, raise-stand, rise from chair and around, dress and take off); GHQ, general health questionnaire; HADS, hospital anxiety and depression scale; I, intervention; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Duff 2018	Multiple sclerosis	Pilates vs control (no intervention)	QoL, disease specific*	end of Treatment (12 wks)	MSQOL-54 - mental health	higher score means better QoL	15/15	68.6 (18.8)	75.5 (13.18)	NR	NR	No difference	Low
			QoL - disease specific*	end of Treatment (12 wks)	MSQOL-54 - physical health	higher score means better QoL	15/15	58.3 (17.6)	61.7 (19.5)	NR	NR	No difference	Low
			Functional mobility*	end of Treatment (12 wks)	Timed Up & Go, left turn (s)	higher score means worse mobility	15/15	8.6 (2.8)	8.9 (5.0)	NR	NR	No difference	Low
			Functional mobility*	end of Treatment (12 wks)	Timed Up & Go, right turn (s)	higher score means worse mobility	15/15	8.8 (3.3)	9.5 (5.5)	NR	NR	No difference	Low
			QoL - disease specific*	baseline, end of treatment (12 wks)	MSQOL-54 - mental health	absolute mean change (95% CI)	15/15	5.9 (-0.5, 12.2)	4.2 (-2.1, 10.6)	NR	0.71	No difference	Low
			QoL - disease specific*	baseline, end of treatment (12 wks)	MSQOL-54 - physical health	absolute mean change (95% CI)	15/15	4.6 (-1.3, 10.5)	2.4 (-3.5, 8.3)	NR	0.60	No difference	Low
			Functional mobility*	baseline, end of treatment (12 wks)	Timed Up & Go, left turn (s)	absolute mean change (95% CI)	15/15	-1.5 (-2.7, -0.4)	0.3 (-0.9, 1.4)	NR	0.03	Favours intervention	Low
			Functional mobility*	baseline, end of treatment (12 wks)	Timed Up & Go, right turn (s)	absolute mean change (95% CI)	15/15	-1.1 (-2.1, -0.1)	0.3 (-0.7, 1.4)	NR	0.06	No difference	Low
			Footnotes:	*Secondary outcomes are exploratory and not sufficiently powered to detect a statistical significant difference. **TUG stratified into right and left turns (once per side) to account for known imbalances that exist in persons with MS.									
Eftekhari 2018*	Multiple sclerosis (women)	Pilates vs control (usual activities)	Balance*	end of treatment (8 wks)	Berg balance scale	higher score means better balance (0-56)	13/12	3.25 (1.42)	3.25 (1.42)	NR	0.003	Favours intervention	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Fatigue	end of treatment (8 wks)	Modified Fatigue Impact Scale (5-items)	higher score means worse fatigue	13/12	6.46 (3.35)	10.5 (4.18)	NR	< 0.000	<i>Favours intervention</i>	High
			Footnotes:	*Individual scores for all 14-items suggest the reported total score for Pilates and control group likely an publishing error (identical and not in the expected range [0-56]).									
Fleming 2019	Multiple sclerosis (women)	Pilates (supervised) vs Control (waitlist)	Fatigue	end of treatment (8 wks)	Modified Fatigue Impact Scale (21-item)	higher score means worse fatigue	3/6	27.7 (6.7)	48.5 (14.2)	NR	< 0.02	<i>Favours intervention</i>	High
			Fatigue	baseline, end of treatment (8 wks)	Modified Fatigue Impact Scale (21-item)	absolute mean change (95% CI)	3/6	-0.26 (-1.65, 1.31)	-0.88 (-2.32,0.57)	NR	NR	<i>No difference</i>	High
			Footnotes:										
Fleming 2019	Multiple sclerosis (women)	Pilates (home-DVD) vs Control (waitlist)	Fatigue	end of treatment (8 wks)	Modified Fatigue Impact Scale (21-item)	higher score means worse outcome	6/6	35.0 (8.6)	48.5 (14.2)	NR	NR	<i>No difference</i>	High
			Fatigue	baseline, end of treatment (8 wks)	Modified Fatigue Impact Scale (21-item)	absolute mean change (95% CI)	6/6	0.38 (-0.76, 1.53)	-0.88 (-2.32,0.57)	NR	NR	<i>No difference</i>	High
			Footnotes:										
Rezvani 2017	Multiple sclerosis (women)	Pilates vs control (no intervention)	Functional mobility*	end of treatment (12 wks)	Timed Up & Go (s)	higher score means worse mobility	10/10	NR	NR	NR	NR	<i>Not reported</i>	High
			Footnotes:	*The results for this outcome measured but not reported or discussed.									
Sisi 2013	Multiple sclerosis (men)	Pilates vs control (no intervention)	Balance*	end of treatment (8 wks)	Berg balance scale (14-items)	higher score means better balance (0-56)	15/15	38.43 (2.87)	31 (2.49)	NR	NR	<i>Not reported</i>	High
			Functional mobility	end of treatment (8 wks)	Timed Up & Go (s)	higher score means worse mobility	15/15	11.72 (3.01)	12.23 (1.81)	NR	0.00	<i>Not reported</i>	High
			Balance	end of treatment (8 wks)	Berg balance scale (14-items)	absolute mean change (95% CI)	15/15	-4.34 (NR)	-0.91 (NR)	NR	0.01	<i>Favours intervention</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Functional mobility	end of treatment (8 wks)	Timed Up & Go (s)	absolute mean change (95% CI)	15/15	0.59 (NR)	0.2 (NR)	NR	0.00	<i>Favours intervention</i>	High
			Footnotes:	*baseline scores between groups are not matched raising serious doubts about the results [mean 34.09 (3.07) in the Pilates group and mean 30.09 (2.51) in the control).									
Pilates vs 'other'													
Abasiyanik 2018	Multiple sclerosis	Pilates vs home exercise	Functional mobility	end of treatment (8 wks)	Timed Up & Go (s)	higher score means worse mobility	16/17	10.27 (6.19)	13.09 (11.91)	NR	0.316	<i>No difference</i>	Some concerns
			Functional mobility	baseline, end of treatment (8 wks)	Timed Up & Go (s)	absolute mean change (95% CI)	16/17	-3.72 (6.56)	-1.92 (3.06)	NR	0.316	No difference	Some concerns
			Physical performance	end of treatment (8 wks)	MS walking scale (12-item)	higher score means worse outcome	16/17	23.81 (9.27)	29.29 (12.70)	NR	0.165	<i>No difference</i>	Some concerns
			Physical performance	baseline, end of treatment (8 wks)	MS walking scale (12-item)	absolute mean change (95% CI)	16/17	-8.06 (6.78)	-5.06 (5.32)	NR	0.165	<i>No difference</i>	Some concerns
			Footnotes:										
Bulguroglu 2015 *	Multiple sclerosis	Pilates (mat) vs home exercise (relaxation and respiration exercises)	Functional mobility	end of treatment (8 wks)	Timed Up & Go (s)	higher score means worse mobility	12/13	5.7 (5.0, 6.5)	4.9 (4.5, 5.3)	NR	NR	<i>Not reported</i>	High
			QoL, disease specific	end of treatment (8 wks)	MSQOL-54 - mental health	higher score means better QoL	12/13	77.23 (70.72, 84.54)	78.52 (64.77, 89.21)	NR	NR	<i>Not reported</i>	High
			QoL, disease specific	end of treatment (8 wks)	MSQOL-54 - physical health	higher score means better QoL	12/13	75.8 (70.83, 86.42)	82.64 (66.77, 91.27)	NR	NR	<i>Not reported</i>	High
			Footnotes:	*Not clear if reported data are (mean, 95% CIs) or (median, IQR) *Comparative stats comparing intervention to control group were not reported. N and n per outcome is assumed									
Bulguroglu 2015 *	Multiple sclerosis	Pilates (reformer) vs home exercise (relaxation and respiration exercises)	Functional mobility	end of treatment (8 wks)	Timed Up & Go (s)	higher score means worse mobility	13/13	5.4 (4.9, 7.1)	4.9 (4.5, 5.3)	NR	NR	<i>Not reported</i>	High
			QoL, disease specific	end of treatment (8 wks)	MSQOL-54 - mental health	higher score means better QoL	N = 25, n = 12, n = 13	74.58 (70.39, 80.58)	78.52 (64.77, 89.21)	NR	NR	<i>Not reported</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Freeman 2012	Multiple sclerosis (women)	Pilates vs Standard physiotherapy	QoL, disease specific	end of treatment (8 wks)	MSQOL-54 - physical health	higher score means better QoL	N = 25, n = 12, n = 13	76.3 (74.39, 83.37)	82.64 (66.77, 91.27)	NR	NR	<i>Not reported</i>	High
			Footnotes:	*Not clear if reported data are (mean, 95% CIs) or (median, IQR) *Comparative statistics comparing intervention to control group were not reported. N and n per outcome is assumed									
			Physical performance	end of treatment (12 wks)	MS walking scale (12-item)	higher score means worse outcome	33/32	63.49 (23.78)	47.84 (24.61)	NR	NR	<i>Favours comparator</i>	Some concerns
Freeman 2012	Multiple sclerosis (women)	Pilates vs Relaxation exercises	Physical performance-adjusted*	end of treatment (12 wks)	MS walking scale (12-item)	adjusted mean difference (95% CI)	33/32	NR	NR	15.65 (29.50, -1.79)	NR	<i>No difference</i>	Low
			Footnotes:	*data were adjusted using LOCF method to account to missing data (relapsers and lost to followup) and fitted to repeated measures mixed models									
			Physical performance	end of treatment (12 wks)	MS walking scale (12-item)	higher score means worse outcome	33/29	63.49 (23.78)	68.39 (23.69)	NR	NR	<i>No difference</i>	Some concerns
Guclu-Gunduz 2014 (NRSI)	Multiple sclerosis	Pilates vs Home exercise	Physical performance-adjusted*	end of treatment (12 wks)	MS walking scale (12-item)	adjusted mean difference (95% CI)	33/29	NR	NR	-4.9 (-19.11, 9.32)	NR	<i>No difference</i>	Low
			Footnotes:	*data were adjusted using LOCF method to account to missing data (relapsers and lost to followup) and fitted to repeated measures mixed models									
			Balance	end of treatment (8 wks)	Berg balance scale	higher score means better balance (0-56)	18/8 median (IQR)	56 (54 to 56)	56 (54.5 to 56)	NR	NR	<i>Not reported</i>	Some concerns
Kalron 2016	Multiple sclerosis	Pilates vs physical therapy	Functional mobility	end of treatment (8 wks)	Timed Up & Go (s)	higher score means worse mobility	18/8 median (IQR)	5.31 (4.56 to 5.99)	6.77 (6.37 to 7.93)	NR	NR	<i>Not reported</i>	Some concerns
			Footnotes:	*Comparative statistics comparing intervention to control group were not reported.									
			Fatigue	end of treatment (12 wks)	Modified Fatigue Impact Scale (21-item)	higher score means worse fatigue	22/23	34.7 (19.5)	28.7 (21.7)	NR	NR	<i>Not reported</i>	Low
Kalron 2016	Multiple sclerosis	Pilates vs physical therapy	Balance	end of treatment (12 wks)	Berg balance scale (14-items)	higher score means better balance	22/23	47.9 (6.4)	44.6 (4.9)	NR	NR	<i>Not reported</i>	Low

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Physical performance	end of treatment (12 wks)	MS walking scale (12-item)	higher score means worse outcome	22/23	36.4 (11.8)	34.8 (11.9)	NR	NR	<i>Not reported</i>	Low
			Functional mobility	end of treatment (12 wks)	Timed Up & Go (s)	higher score means worse mobility	22/23	10.7 (3.3)	9.9 (2.9)	NR	NR	<i>Not reported</i>	Low
			Fatigue	baseline, end of treatment (12 wks)	Modified Fatigue Impact Scale (21-item)	absolute mean change (95% CI)	22/23	-0.6 (13.6)	-1.7 (16.9)	NR	NR	<i>Not reported</i>	Low
			Balance	end of treatment (12 wks)	Berg balance scale (14-items)	absolute mean change (95% CI)	22/23	1.1 (4.2)	1.3 (5.2)	NR	NR	<i>Not reported</i>	Low
			Physical performance	end of treatment (12 wks)	MS walking scale (12-item)	absolute mean change (95% CI)	22/23	2.8 (6.3)	2.4 (5.9)	NR	NR	<i>Not reported</i>	Low
			Functional mobility	end of treatment (12 wks)	Timed Up & Go (s)	absolute mean change (95% CI)	22/23	-1.8 (2.1)	-1.7 (2.1)	NR	NR	<i>Not reported</i>	Low
			Footnotes:	*Comparative statistics comparing intervention to control group were not reported.									
Kara 2017	Multiple sclerosis	Pilates vs aerobic exercise	No data extracted because the study does not include any priority outcome domains or measures.										
Küçük 2015	Multiple sclerosis	Pilates vs traditional exercise programme	Balance	end of treatment (8 wks)	Berg balance scale (14-items)	higher score means better balance (0-56)	11/9	53.55 (6.41)	47.78 (13.89)	NR	NR	<i>Not reported</i>	Some concerns
			Quality of life	end of treatment (8 wks)	MusiQoL (31-items)	Higher score means better QoL	11/9	23.82 (7.53)	40.05 (17.96)	NR	NR	<i>Favours intervention</i>	Some concerns
			Functional mobility	end of treatment (8 wks)	Timed Up & Go, left turn (s)	higher score means worse mobility	11/9	12.2 (8.74)	9.13 (4.25)	NR	NR	<i>Not reported</i>	Some concerns
			Functional mobility	end of treatment (8 wks)	Timed Up & Go, right turn (s)	higher score means worse mobility	11/9	10.74 (7.51)	9.47 (4.3)	NR	NR	<i>Not reported</i>	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Fatigue	end of treatment (12 wks)	Modified Fatigue Impact Scale (21-item) - cognitive	higher score means worse fatigue	11/9	5.82 (5.04)	7.33 (6.60)	NR	NR	Not reported	Some concerns
			Fatigue	end of treatment (12 wks)	Modified Fatigue Impact Scale (21-item) - physical	higher score means worse fatigue	11/9	7.18 (3.63)	7.44 (5.27)	NR	NR	Not reported	Some concerns
			Fatigue	end of treatment (12 wks)	Modified Fatigue Impact Scale (21-item) - social	higher score means worse fatigue	11/9	7.64 (9.60)	13.11 (10.24)	NR	NR	Not reported	Some concerns
			Footnotes:	*Comparative statistics comparing intervention to control group were not reported.									
Rezvani 2017	Multiple sclerosis (women)	Pilates vs exercise	Functional mobility*	end of treatment (12 wks)	Timed Up & Go (s)	higher score means worse mobility	10/10	NR	NR	NR	NR	Not reported	High
			Footnotes:	*The results for this outcome not reported or discussed.									
Sisi 2013	Multiple sclerosis (men)	Pilates vs exercise (rebound therapy)	Balance	Baseline, end of treatment (8 wks)	Berg balance scale (14-items)	higher score means better balance (0-56)	15/15	38.43 (2.87)	36.06 (2.12)	NR	NR	Not reported	High
			Functional mobility	baseline, end of treatment (8 wks)	Timed Up & Go (s)	higher score means worse mobility	15/15	11.72 (3.01)	11.43 (2.37)	NR	NR	Not reported	High
			Balance	Baseline, end of treatment (8 wks)	Berg balance scale (14-items)	absolute mean change (95% CI)	15/15	-4.34 (NR)	-4.35 (NR)	NR	1	No difference	High
			Functional mobility	baseline, end of treatment (8 wks)	Timed Up & Go (s)	absolute mean change (95% CI)	15/15	0.59 (NR)	1.64 (NR)	NR	1	No difference	High
			Footnotes:										
Abbreviations: C, Comparator; I, intervention; LOCF, last-observation carried forward; MSQOL, multiple sclerosis-quality of life; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Borges 2014	Patients infected by the HTLV-1	Pilates vs inactive comparator	Pain	End of treatment (15 wks)	VAS (0-10)	Higher score means more pain	11/11	3.45 (2.54)	7.5 (2.51)	NR	NR*	Favours intervention	Some concerns
			Physical function	End of treatment (15 wks)	SF-36 physical functioning	Higher score means better outcome	11/11	41.82 (20.16)	32.00 (15.31)	NR	NR*	Not reported	Some concerns
			Role - Physical	End of treatment (15 wks)	SF-36 role physical	Higher score means better outcome	11/11	72.73 (32.51)	10.00 (17.48)	NR	NR*	Not reported	Some concerns
			Pain	End of treatment (15 wks)	SF-36 bodily pain	Higher score means better outcome	11/11	60.64 (20.11)	30.50 (14.45)	NR	NR*	Not reported	Some concerns
			General health	End of treatment (15 wks)	SF-36 general health perceptions	Higher score means better outcome	11/11	52.73 (25.73)	32.20 (18.40)	NR	NR*	Not reported	Some concerns
			Vitality	End of treatment (15 wks)	SF-36 vitality	Higher score means better outcome	11/11	56.36 (22.70)	28.00 (23.71)	NR	NR*	Not reported	Some concerns
			Role-social	End of treatment (15 wks)	SF-36 role social	Higher score means better outcome	11/11	69.32 (20.43)	55.00 (25.82)	NR	NR*	Not reported	Some concerns
			Role-emotional	End of treatment (15 wks)	SF-36 role emotional	Higher score means better outcome	11/11	63.65 (40.71)	70.51 (24.59)	NR	NR*	Not reported	Some concerns
			Mental health	End of treatment (15 wks)	SF-36 mental health	Higher score means better outcome	11/11	69.82 (25.45)	54.00 (22.25)	NR	NR*	Not reported	Some concerns
Footnotes:				*Reported inferential statistical analyses were for within group differences (i.e. before-after) but not between group differences (i.e. Pilates vs comparator). The study also reports results at end of cross-over period (i.e. 30 wk follow-up). These outcomes are not reported here.									

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs 'other' - no studies found													
Abbreviations: C, Comparator; I, intervention; NR, not reported; SF-36, 36-item short form survey; VAS, visual analogue scale													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Pandya 2017	Parkinson's disease	Pilates vs control (as adjunct to conventional balance training)	Functional mobility	end of treatment (7 wks)	Timed Up and Go test (s)	Higher score means worse agility	15/15	18 (8.847)	26.53 (3.642)	NR	0.002	Favours intervention	High
			Balance	end of treatment (7 wks)	Berg balance scale (14-items)(0-56)	Higher score means better balance	15/15	42.133 (5.566)	37.066 (5.133)	NR	0.021	Favours intervention	High
			Footnotes:										
Pilates vs 'other'													
Daneshmandi 2017	Parkinson's disease	Pilates vs walking	Functional mobility	end of treatment (8 wks)	Timed Up and Go test (s)	Higher score means worse mobility	15/15	8.67 (2.55)	17.73 (5.47)	MD -9.06 (NR)	0.001	Favours intervention	High
			Balance	end of treatment (8 wks)	Fullerton Advanced Balance Scale (10-items)(0-40)	Higher score means better balance	15/15	32.67 (5.16)	7.27 (2.84)	MD 25.40 (NR)	0.001	Favours intervention	High
			Footnotes:										
Mollinedo-Cardalda 2018	Parkinson's disease	Pilates vs Physical activities (calisthenics)	Functional mobility	end of treatment (12 wk)	Timed Up and Go test (s)	Higher score means worse agility	12/10	7.78 (2.81)	9.22 (2.49)	NR	0.488	No difference	High
			Motor examination	end of treatment (12 wk)	UPDR-Motor scale (Part III)	higher score means worse outcome	12/10	27.92 (12.46)	38.5 (10.81)	NR	0.921	No difference	High
			Footnotes:										
Abbreviations: C, Comparator; I, intervention; NR, not reported; UPDRS; Unified Parkinson's Disease Rating Scale III													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p -value	direction of effect	RoB
Pilates vs control													
Lim 2017	Chronic stroke	Pilates vs control (no intervention) *as an adjunct to conventional stroke rehabilitation	Functional mobility	end of treatment (8 wks)	Timed Up and Go test (s)	Higher score means worse agility	10/10	-19.2 (5.8) mean (SE)	21.7 (6.4) mean (SE)	NR	NR	Not reported	Some concerns
			Functional mobility	baseline, end of treatment (8 wks)	Timed Up and Go test (s)	absolute mean change (SE)	10/10	-3.5 (4.7)	2.5 (1.2)	NR	NR	Not reported	Some concerns
			Footnotes:										
Roh 2016	Chronic stroke	Pilates vs control (no intervention)	Functional mobility	end of treatment (8 wks)	Gait speed (cm/s)	higher score means better outcome	10/10	31.48 (12.81)	21.54 (3.41)	NR	0.05	Favours intervention	High
			Footnotes:										
Sathe 2018	Chronic stroke	Pilates vs control (no intervention) *as an adjunct to conventional balance therapy	Dynamic balance	end of treatment (6 wks)	Tinetti test	higher score means better dynamic balance	5/4	23.4 (NR)	22.5 (NR)	NR	0.05	Favours intervention	High
			Dynamic balance	end of treatment (6 wks)	Tinetti test	absolute mean change (95% CI)	5/4	3.2 (NR)	1.75 (NR)	NR	0.05	Favours intervention	High
			Footnotes:										
Pilates vs 'other'													
Yun 2017	Chronic stroke	Pilates vs usual care (occupational therapy)	QoL - disease specific	baseline, end of treatment (12 wks)	SS-QOL - total score	higher scores mean better quality of life	20/20	3.23 (0.56)	2.8 (0.38)	NR	<0.05	Favours intervention	High
			Physical function	baseline, end of treatment (12 wks)	SS-QOL - physical	higher scores mean better functioning	20/20	3.32 (0.64)	2.92 (0.54)	NR	<0.05	Favours intervention	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Footnotes:													
Abbreviations: C, Comparator; I, intervention; NR, not reported; SS-QOL, Stroke specific quality of life													

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Martins-Meneses 2015	Hypertension (women)	Pilates vs Control (waitlist)	blood pressure, systolic	end of treatment (16 wks)	mmHg	Higher score means worse health	22/22						Critical
			blood pressure, diastolic	end of treatment (16 wks)	mmHg	Higher score means worse health	22/22					Critical	
			blood pressure, mean	end of treatment (16 wks)	mmHg	Higher score means worse health	22/22					Critical	
			heart rate	end of treatment (16 wks)	bpm	Higher score means worse health	22/22					Critical	
			double product (HRxBP)	end of treatment (16 wks)	bpm x mmHg	Higher score means worse health	22/22					Critical	
			Footnotes:	NRSI at critical risk of bias. Outcomes not extracted and not included in evidence synthesis									
Pilates vs 'other' - no studies found													
Abbreviations: bpm, beats per minute; C, Comparator; I, intervention; mmHg, millimetres mercury; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Mazloun 2018a	Osteoarthritis (knee)	Pilates vs Control (usual activities)	Global function/disability	end of treatment (8 weeks)	Lequesne Index	Higher score means more disability	17/16	8.4 (1.9)	10.5 (1.8)	2.1 (1.7, 2.6)	0.00	Favours intervention	High
			Functional performance	end of treatment (8 weeks)	Time to complete activities*	Higher score means worse functional autonomy	17/16	48.9 (5.5)	58.5 (5.8)	9.4 (7.2, 11.6)	0.00	Favours intervention	High
			Footnotes:	* including walking 15 metres, standing up from a chair and walking 15 metres, going up and down 11 stairs (height of 12 centimetres). Authors mention 4 tests, but not clear of additional measure.									
de Oliveria 2019	Chikungunya (chronic, infectious)	Pilates vs Control (no intervention)	Pain	end of treatment (12 weeks)	0-10 visual analogue scale	higher score means worse pain	22/20	4.4 (2.4)	7.8 (2.4)	NR	NR	Not reported	Some concerns
			Physcial wellbeing	end of treatment (12 weeks)	SF-12 Physical component score	higher score means better QoL	22/20	39.9 (9.0)	28.9 (5.0)	NR	NR	Not reported	Some concerns
			Footnotes:										
Pilates vs 'other'													
Mazloun 2018a	Osteoarthritis (knee)	Pilates vs Conventional therapeutic exercises	Global function/disability	end of treatment (8 weeks)	Lequesne Index	Higher score means more disability	17/16	8.4 (1.9)	9.1 (2.0)	0.6 (0.2, 1.1)	0.003	Favours intervention	High
			Functional performance	end of treatment (8 weeks)	Time to complete activities*	Higher score means worse functional autonomy	17/16	48.9 (5.5)	48.0 (4.2)	-3.0 (-2.5, 1.8)	0.938	No difference	High
			Footnotes:	* including walking 15 metres, standing up from a chair and walking 15 metres, going up and down 11 stairs (height of 12 centimetres). Authors mention 4 tests, but not clear of additional measure. **Suspect incorrect reporting of baseline characteristics due to significant difference in baseline funtional status between Pilates and exerise control group.									
Abbreviations: C, Comparator; I, intervention; NR, not reported													

STUDY RESULTS (as reported by the study authors)

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Altan 2012	Ankylosing spondylitis	Pilates vs Control (usual care)	Functional capacity	end treatment (12 weeks)	BAS-Functional Index (BASFI)	Higher score means worse function	30/25	1.7 (1.6)	2.3 (1.7)	NR	NR	Not reported	Some concerns
			Functional capacity	baseline, end treatment (12 weeks)	BAS-Functional Index (BASFI)	Absolute mean change (SD)	30/25	-0.7 (1.5)	0.1 (0.9)	NR	0.059	No difference	Some concerns
			Disease activity	end treatment (12 weeks)	BAS-Disease Activity Index (BASDAI)	Higher score means worse disease activity	30/25	2.1 (2)	3.1 (1.7)	NR	NR	Not reported	Some concerns
			Disease activity	baseline, end treatment (12 wks)	BAS-Disease Activity Index (BASDAI)	Absolute mean change (SD)	30/25	-0.7 (1.8)	0.5 (1.1)	NR	0.003	Favours intervention	Some concerns
			Spinal mobility	end treatment (12 weeks)	BAS-Metrology Index (BASMI)	Higher score means worse spinal mobility	30/25	8.4 (1.9)	8.7 (1.8)	NR	NR	Not reported	Some concerns
			Spinal mobility	baseline, end treatment (12 wks)	BAS-Metrology Index (BASMI)	Absolute mean change (SD)	30/25	-0.4 (0.7)	-0.2 (0.8)	NR	0.304	No difference	Some concerns
			QoL, disease specific	end treatment (12 weeks)	Ankylosing Spondylitis Quality of Life (ASQOL)	higher score means worse quality of life	30/25	4 (4.9)	4 (4.8)	NR	NR	Not reported	Some concerns
			QoL, disease specific	baseline, end treatment (12 wks)	Ankylosing Spondylitis Quality of Life (ASQOL)	Absolute mean change (SD)	30/25	0.3 (2.9)	-0.2 (1.5)	NR	0.343	No difference	Some concerns
Footnotes:			Authors also report data at 24 week follow up which is not included here.										

Pilates vs 'other' - no studies found

Abbreviations: BAS, Bath Ankylosing Spondylitis; C, Comparator; I, intervention; NR, not reported

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Alves de Araújo 2010	Scoliosis (nonstructural)	Pilates vs Attention control (weekly meetings)	Deformity progression	End of treatment (12 weeks)	Degree of curvature (Cobb angle)	Higher score means worse posture	20/11	4.8 (2.0)	6.9 (3.1)	NR	0.028	<i>Favours intervention</i>	High
			Pain	End of treatment (12 weeks)	0-10 numeric rating scale (Borg CR10)	Higher score means worse pain	20/11	1.8 (1.9)	3.8 (2.7)	NR	0.004	<i>Favours intervention</i>	High
			Footnotes:										
Junges 2012	hyperkyphosis (thoracic spine)	Pilates vs Control (usual activities)	Deformity progression	End of treatment (30 weeks)	Degree of curvature (Cobb angle)	Higher score means worse posture	22/19	55.50 (11.97)	58.22 (8.59)	NR	< 0.001	<i>Favours intervention</i>	High
Pilates vs 'other'													
Kim 2016	Scoliosis (idiopathic)	Pilates vs Schroth exercise	Deformity progression	End of treatment (12 weeks)	Degree of curvature (Cobb angle)	Higher score means worse posture	12/12	16.0 (6.9)	12.0 (4.7)	NR	< 0.05	<i>Favours comparator</i>	High
Footnotes:													
Kudchadkar 2019	hyperlordosis (lumbar spine)	Pilates vs Egoscue exercises	Degree of lumbar lordosis	End of treatment (4 weeks)	Index of lordosis (61 cm Surveyors flexi curve)	Maximum width and the total length of the curve	17/17	39.16 (4.35)	38.09 (6.08)	NR	0.68	<i>Favours intervention</i>	Some concerns
Footnotes:													
Kudchadkar 2019	hyperlordosis (lumbar spine)	Pilates vs Lumbar stabilization exercises	Degree of lumbar lordosis	End of treatment (4 weeks)	Index of lordosis (61 cm Surveyors flexi curve)	Maximum width and the total length of the curve	17/17	39.16 (4.35)	43.43 (6.04)	NR	< 0.001*	<i>Favours intervention</i>	Some concerns
Footnotes:				*Significant difference refers to Pilates and Egoscue groups were significantly different from the lumbar stabilisation exercise group.									
Lee 2016b	forward head posture (cervical spine)	Pilates vs Combined exercise	Function/disability	End of treatment (10 weeks)	Neck disability index	Higher is worse	14/14	5.1 (2.1)	5.9 (3.4)	NR	NR	<i>Not reported</i>	High
Footnotes:													
Navega 2016	hyperkyphosis (thoracic spine)	Pilates vs Education	Thoracic kyphosis	End of treatment (8 weeks)	Computed biophotogrammetry	Angle of the spine	14/17	53.43	54.88	NR	0.723	<i>No difference</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Footnotes:				No priority outcome measures reported but included a potentially relevant outcome. SD not reported.									
Abbreviations: C, Comparator; I, intervention; NR, not reported;													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Angin 2015	Osteoporosis (without fracture, postmenopause)	Pilates vs Control (no intervention)	Pain (active)	End of treatment (24 weeks)	Visual Analogue Scale (0-10)	Higher score means more pain	22/19	1.95 (1.31)	6.89 (1.56)	NR	NR	<i>Not reported</i>	High
			Pain (at rest)	End of treatment (24 weeks)	Visual Analogue Scale (0-10)	Higher score means more pain	22/19	0.39 (0.69)	3.37 (2.45)	NR	NR	<i>Not reported</i>	High
			Quality of life - Pain	End of treatment (24 weeks)	QUALEFFO-41 - Pain	Higher score means better QoL	22/19	63.18 (12.30)	56.84 (15.20)	NR	NR	<i>Not reported</i>	High
			Quality of life - Physical function	End of treatment (24 weeks)	QUALEFFO-41 - Daily activities	Higher score means better QoL	22/19	81.60 (13.79)	81.24 (10.82)	NR	NR	<i>Not reported</i>	High
			Quality of life - Physical function	End of treatment (24 weeks)	QUALEFFO-41 - House work	Higher score means better QoL	22/19	62.27 (15.47)	60.79 (10.84)	NR	NR	<i>Not reported</i>	High
			Quality of life - Physical function	End of treatment (24 weeks)	QUALEFFO-41 - Mobility	Higher score means better QoL	22/19	73.12 (13.22)	85.11 (12.83)	NR	NR	<i>Not reported</i>	High
			Quality of life - Social	End of treatment (24 weeks)	QUALEFFO-41 - Social activities	Higher score means better QoL	22/19	39.98 (14.81)	30.94 (18.42)	NR	NR	<i>Not reported</i>	High
			Quality of life - General	End of treatment (24 weeks)	QUALEFFO-41 - General health	Higher score means better QoL	22/19	35.07 (12.83)	32.38 (15.32)	NR	NR	<i>Not reported</i>	High
			Quality of life - Mental	End of treatment (24 weeks)	QUALEFFO-41 - Mental functions	Higher score means better QoL	22/19	51.95 (10.09)	54.94 (10.53)	NR	NR	<i>Not reported</i>	High
			Bone mineral density	End of treatment (24 weeks)	Bone mineral density T-score (L2-L4)	Higher is better	22/19	-2.49 (0.37)	-2.81 (0.33)	NR	NR	<i>Not reported</i>	High
			Bone mineral density	End of treatment (24 weeks)	Bone mineral density (L2-L4) (g/cm2)	Higher is better	22/19	0.714 (0.009)	0.653 (0.007)	NR	NR	<i>Not reported</i>	High
			Functional mobility	End of treatment (24 weeks)	6-minute walk test (m)	Higher is better	22/19	453.82 (93.07)	400.42 (93.31)	NR	NR	<i>Not reported</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Pain (active)	Baseline, end of treatment (24 wks)	Visual Analogue Scale (0-10)	absolute mean change (SD)	22/19	4.23 (1.39)	-0.55 (0.60)	NR	0.00	Favours intervention	High
			Pain (at rest)	Baseline, end of treatment (24 wks)	Visual Analogue Scale (0-10)	absolute mean change (SD)	22/19	1.87 (1.51)	-0.29 (0.75)	NR	0.00	Favours intervention	High
			Quality of life - Pain	Baseline, end of treatment (24 wks)	QUALEFFO-41 - Pain	absolute mean change (SD)	22/19	13.91 (7.81)	-3.16 (6.71)	NR	0.00	Favours intervention	High
			Quality of life - Physical function	Baseline, end of treatment (24 wks)	QUALEFFO-41 - Daily activities	absolute mean change (SD)	22/19	7.68 (14.08)	-2.30 (3.72)	NR	0.00	Favours intervention	High
			Quality of life - Physical function	Baseline, end of treatment (24 wks)	QUALEFFO-41 - House work	absolute mean change (SD)	22/19	10.86 (6.86)	-2.37 (4.21)	NR	0.00	Favours intervention	High
			Quality of life - Physical function	Baseline, end of treatment (24 wks)	QUALEFFO-41 - Mobility	absolute mean change (SD)	22/19	11.32 (12.07)	-2.32 (6.06)	NR	0.00	Favours intervention	High
			Quality of life - Social	Baseline, end of treatment (24 wks)	QUALEFFO-41 - Social activities	absolute mean change (SD)	22/19	22.03 (11.87)	-1.65 (3.75)	NR	0.00	Favours intervention	High
			Quality of life - General	Baseline, end of treatment (24 wks)	QUALEFFO-41 - General health	absolute mean change (SD)	22/19	16.97 (10.77)	-3.07 (4.98)	NR	0.00	Favours intervention	High
			Quality of life - Mental	Baseline, end of treatment (24 wks)	QUALEFFO-41 - Mental functions	absolute mean change (SD)	22/19	8.24 (3.97)	-1.02 (2.81)	NR	0.00	Favours intervention	High
			Footnotes:										
Oksuz 2014	Osteoporosis (without fracture, postmenopause)	Pilates vs Control (usual activities)	Pain (active)	Baseline, end of treatment (6 wks)	Visual Analogue Scale (0-10)	absolute mean change (SD)*	20/20	0.28 (0.53)	-1.42 (1.58)	NR	0.00	Favours intervention	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
)		Pain (at rest)	Baseline, end of treatment (6 wks)	Visual Analogue Scale (0-10)	absolute mean change (SD)*	20/20	0.21 (1.03)	-2.36 (2.15)	NR	0.00	Favours intervention	High
			Quality of life	Baseline, end of treatment (6 wks)	QUALEFFO-41 - Total	absolute mean change (SD)*	20/20	6.90 (3.82)	0.69 (1.25)	NR	0.00	Favours intervention	High
			Quality of life - Pain	Baseline, end of treatment (6 wks)	QUALEFFO-41 - Pain	absolute mean change (SD)*	20/20	11.00 (14.10)	0.50 (2.76)	NR	0.00	Favours intervention	High
			Quality of life - Physical function	Baseline, end of treatment (6 wks)	QUALEFFO-41 - Physical function	absolute mean change (SD)*	20/20	6.5 (4.23)	0.81 (1.94)	NR	0.00	Favours intervention	High
			Quality of life - Social	Baseline, end of treatment (6 wks)	QUALEFFO-41 - Social activities	absolute mean change (SD)*	20/20	9.36 (9.30)	0.00 (0.00)	NR	0.00	Favours intervention	High
			Quality of life - General health	Baseline, end of treatment (6 wks)	QUALEFFO-41 - General health	absolute mean change (SD)*	20/20	5.83 (5.47)	1.25 (3.05)	NR	0.003	Favours intervention	High
			Quality of life - Mental	Baseline, end of treatment (6 wks)	QUALEFFO-41 - Mental functions	absolute mean change (SD)*	20/20	6.81 (6.40)	0.83 (2.40)	NR	0.00	Favours intervention	High
			Balance stability	baseline, end of treatment (6 wks)	Berg Balance Test	absolute mean change (SD)*	20/20	1.75 (1.25)	0.05 (0.22)	NR	0.00	Favours intervention	High
			Footnotes:	*baseline, end of treatment scores not provided.									
Pilates vs 'other'													
Kucukcakir 2013	Osteoporosis (without fracture, postmenopause	Pilates vs Home exercise	Quality of life - Total	End of treatment (1 year)	QUALEFFO-41 - Total	Higher score means worse	30/30	14.8 (9.1)	33.6 (11.4)	NR	<0.001	Favours intervention	High

Abbreviations: C, Comparator; I, intervention; NR, not reported; QUALEFFO, Questionnaire of the European Foundation for Osteoporosis

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Mirmohammad ali 2012	Postpartum women	Control (attention control)*	General fatigue	End of treatment (8 weeks postpartum)	MFI - general fatigue	Higher score means more fatigue	40/40	7.80 (2.07)	12.72 (1.79)	NR	<0.001	Favours intervention	High
			Physical fatigue	End of treatment (8 weeks postpartum)	MFI - physical fatigue	Higher scores mean more fatigue	40/40	7.12 (1.41)	10.42 (2.02)	NR	<0.001	Favours intervention	High
			Reduced activity	End of treatment (8 weeks postpartum)	MFI - reduced activity	Higher scores mean more fatigue	40/40	6.95 (1.35)	11.27 (1.70)	NR	<0.001	Favours intervention	High
			Reduced motivation	End of treatment (8 weeks postpartum)	MFI - reduced motivation	Higher scores mean more fatigue	40/40	6.20 (1.01)	9.80 (2.04)	NR	<0.001	Favours intervention	High
			Mental fatigue	End of treatment (8 weeks postpartum)	MFI - mental fatigue	Higher scores mean more fatigue	40/40	6.85 (1.45)	10.72 (1.98)	NR	<0.001	Favours intervention	High
			Footnotes:	*Participants attended one educational session on postnatal care and were followed up with weekly phone calls									
Pilates vs 'other' - no studies found													
Abbreviations: C, Comparator; I, intervention; MFI, Multidimensional Fatigue Inventory; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Ahmadinezhad 2017	Menopausal symptom or complaint	Pilates vs Control (no intervention)	Sleep quality	End of treatment (6 weeks)	Pittsburgh Sleep Quality Index - total score	Higher score means worse sleep quality	36/36	4.75 (2.62)	14.58 (2.9)	NR	NR	<i>Favours intervention</i>	High
Footnotes:													
Campos de Oliveira 2018	Menopausal symptom or complaint	Pilates vs Control (no intervention)	Physical function	End of treatment (26 weeks)	SF-36 physical functioning	Higher score means better outcome	17/17	95 (78-95)*	85 (65-90)*	NR	NR	<i>No difference</i>	High
			Role - Physical	End of treatment (26 weeks)	SF-36 role physical	Higher score means better outcome	17/17	100 (100-100)*	75 (75-100)*	F=2.546	0.033**	<i>Favours intervention</i>	High
			Bodily Pain	End of treatment (26 weeks)	SF-36 bodily pain	Higher score means better outcome	17/17	82 (62-100)*	61 (22-74)*	F=2.525	0.035**	<i>Favours intervention</i>	High
			General health	End of treatment (26 weeks)	SF-36 general health perceptions	Higher score means better outcome	17/17	82 (77-92)*	82 (47-92)*	NR	NR	<i>No difference</i>	High
			Vitality	End of treatment (26 weeks)	SF-36 vitality	Higher score means better outcome	17/17	85 (75-90)*	70 (55-80)*	NR	NR	<i>No difference</i>	High
			Role-social	End of treatment (26 weeks)	SF-36 role social	Higher score means better outcome	17/17	100 (87-100)*	63 (50-100)*	F=2.632	0.025**	<i>Favours intervention</i>	High
			Role-emotional	End of treatment (26 weeks)	SF-36 role emotional	Higher score means better outcome	17/17	100 (100-100)*	63 (33-67)*	2.905	0.011**	<i>Favours intervention</i>	High
			Mental health	End of treatment (26 weeks)	SF-36 mental health	Higher score means better outcome	17/17	84 (78-96)*	72 (52-84)*	NR	NR	<i>No difference</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Lumbar spine (L1-L4)	higher score means better bone density	17/17	0.94 (0.19)	0.94 (0.09)	NR	NR***	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Femoral neck	higher score means better bone density	17/17	0.77 (0.12)	0.75 (0.08)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Total hip	higher score means better bone density	17/17	0.9 (0.11)	0.89 (0.08)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Trochanter	higher score means better bone density	17/17	0.62 (0.1)	0.61 (0.06)	NR	NR***	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Interchanter	higher score means better bone density	17/17	1.04 (0.13)	1.04 (0.09)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Ward's area	higher score means better bone density	17/17	0.52 (0.13)	0.53 (0.12)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Lumbar spine (L1-L4)	absolute mean change (SD)	17/17	0.017 (0.017)	0.001 (0.010)	Cohen's d 1.15	0.008	Favours intervention	High
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Femoral neck	absolute mean change (SD)	17/17	0.008 (0.050)	0.003 (0.028)	Cohen's d 0.12	1	No difference	High
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Total hip	absolute mean change (SD)	17/17	0.016 (0.031)	0.002 (0.012)	Cohen's d 0.60	0.304	No difference	High
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Trochanter	absolute mean change (SD)	17/17	0.021 (0.014)	0.001 (0.017)	Cohen's d 1.28	0.005	Favours intervention	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Interchanter	absolute mean change (SD)	17/17	0.011 (0.041)	-0.001 (0.024)	Cohen's d 0.36	0.764	No difference	High
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Ward's area	absolute mean change (SD)	17/17	0.017 (0.057)	-0.001 (0.040)	Cohen's d 0.37	0.915	No difference	High
			Isokinetic muscle strength	End of treatment (26 weeks)	Knee extensors 60 degree/s	Peak isokinetic torque (Nm)	17/17	110 (20)	95.4 (21.8)	NR	NR	Favours intervention	High
			Isokinetic muscle strength	End of treatment (26 weeks)	Knee flexors 60 degree/s	Peak isokinetic torque (Nm)	17/17	57.5 (12.4)	50.7 (11.4)	NR	NR	Favours intervention	High
			Isokinetic muscle strength	End of treatment (26 weeks)	Knee extensors 180 degree/s	Peak isokinetic torque (Nm)	17/17	65.2 (11.0)	55.9 (10.0)	NR	NR	Favours intervention	High
			Isokinetic muscle strength	End of treatment (26 weeks)	Knee flexors 180 degree/s	Peak isokinetic torque (Nm)	17/17	40.6 (11.3)	35.5 (7.8)	NR	NR	Favours intervention	High
			Isokinetic muscle strength	Baseline, end of treatment (26 weeks)	Knee extensors 60 degree/s	absolute mean change (SD)	17/17	6.5 (9.5)	-1.1 (10.1)	Cohen's d 0.78	0.015	Favours intervention	High
			Isokinetic muscle strength	Baseline, end of treatment (26 weeks)	Knee flexors 60 degree/s	absolute mean change (SD)	17/17	6.4 (6.8)	1.0 (5.8)	Cohen's d 0.83	0.024	Favours intervention	High
			Isokinetic muscle strength	Baseline, end of treatment (26 weeks)	Knee extensors 180 degree/s	absolute mean change (SD)	17/17	4.0 (5.5)	-0.6 (5.7)	Cohen's d 0.82	0.028	Favours intervention	High
			Isokinetic muscle strength	Baseline, end of treatment (26 weeks)	Knee flexors 180 degree/s	absolute mean change (SD)	17/17	4.1 (2.6)	0.0 (5.4)	Cohen's d 0.96	0.013	Favours intervention	High
			Footnotes:	<p>*Nonparametric data presented as median (IQR), and as a post hoc, the Bonferroni test was used in multiple comparisons between pairs</p> <p>** Between group comparison (Post Hoc Dunn test)</p> <p>*** Authors report significant difference between change scores comparing pilates with control for lumbar spine (0.016; 95% CI 0.007, 0.025; p=0.008) and trochanter (0.020; 95% CI 0.010, 0.031; p=0.005)</p>									
Lee 2016a	Menopausal symptom or complaint	Pilates vs Control (no intervention)	Vasomotor symptoms	End of treatment (8 weeks)	MSQ- Vasomotor subscale	Higher score means worse symptoms	45/29	8.86 (5.97)	17.74 (11.47)	NR	0.001	Favours intervention	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Physical health	End of treatment (8 weeks)	MSQ-Physical subscale	Higher score means worse symptoms	45/29	18.77 (8.54)	33.21 (14.24)	NR	0.001	<i>Favours intervention</i>	High
Footnotes:													
Pilates vs 'other'													
Ahmadinezhad 2017	Menopausal symptom or complaint	Pilates vs Acupressure	Sleep quality	End of treatment (6 weeks)	Pittsburgh Sleep Quality Index - total score	Higher score means worse sleep quality	36/36	7.48 (3.25)	14.58 (2.9)	NR	>0.999	<i>No difference</i>	High
Footnotes:													
Campos de Oliveira 2018	Menopausal symptom or complaint	Pilates vs Whole body vibration	Physical function	End of treatment (26 weeks)	SF-36 physical functioning	Higher score means better outcome	17/17	95 (78-95)*	85 (68-90)*	NR	NR	<i>No difference</i>	High
			Role - Physical	End of treatment (26 weeks)	SF-36 role physical	Higher score means better outcome	17/17	100 (100-100)*	100 (38-100)*	F=1.689	0.274**	<i>No difference</i>	High
			Bodily Pain	End of treatment (26 weeks)	SF-36 bodily pain	Higher score means better outcome	17/17	82 (62-100)*	74 (62-85)*	F=0.484	1.000**	<i>No difference</i>	High
			General health	End of treatment (26 weeks)	SF-36 general health perceptions	Higher score means better outcome	17/17	82 (77-92)*	92 (85-96)*	NR	NR	<i>No difference</i>	High
			Vitality	End of treatment (26 weeks)	SF-36 vitality	Higher score means better outcome	17/17	85 (75-90)*	70 (55-80)*	NR	NR	<i>No difference</i>	High
			Role-social	End of treatment (26 weeks)	SF-36 role social	Higher score means better outcome	17/17	100 (87-100)*	88 (63-100)*	F=1.446	0.445**	<i>No difference</i>	High
			Role-emotional	End of treatment (26 weeks)	SF-36 role emotional	Higher score means better outcome	17/17	100 (100-100)*	100 (33-100)*	F=1.006	0.943**	<i>No difference</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Mental health	End of treatment (26 weeks)	SF-36 mental health	Higher score means better outcome	17/17	84 (78-96)*	80 (68-90)*	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Lumbar spine (L1-L4)	higher score means better bone density	17/17	0.94 (0.19)	0.98 (0.11)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Femoral neck	higher score means better bone density	17/17	0.77 (0.12)	0.78 (0.1)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Total hip	higher score means better bone density	17/17	0.9 (0.11)	0.91 (0.10)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Trochanter	higher score means better bone density	17/17	0.62 (0.1)	0.63 (0.08)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Interchanter	higher score means better bone density	17/17	1.04 (0.13)	1.06 (0.11)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	End of treatment (26 weeks)	Ward's area	higher score means better bone density	17/17	0.52 (0.13)	0.55 (0.14)	NR	NR	No difference	High
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Lumbar spine (L1-L4)	absolute mean change (SD)	17/17	0.017 (0.017)	0.015 (0.013)	Cohen's d 0.13	1	No difference	High
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Femoral neck	absolute mean change (SD)	17/17	0.008 (0.050)	0.013 (0.036)	Cohen's d 0.11	1	No difference	High
			Bone mineral density (g/cm2)	Baseline, end of treatment (26 weeks)	Total hip	absolute mean change (SD)	17/17	0.016 (0.031)	0.007 (0.024)	Cohen's d 0.32	1	No difference	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Bone mineral density (g/cm ²)	Baseline, end of treatment (26 weeks)	Trochanter	absolute mean change (SD)	17/17	0.021 (0.014)	0.019 (0.018)	Cohen's d 0.12	1	No difference	High
			Bone mineral density (g/cm ²)	Baseline, end of treatment (26 weeks)	Interchanter	absolute mean change (SD)	17/17	0.011 (0.041)	0.005 (0.021)	Cohen's d 0.18	1	No difference	High
			Bone mineral density (g/cm ²)	Baseline, end of treatment (26 weeks)	Ward's area	absolute mean change (SD)	17/17	0.017 (0.057)	0.006 (0.039)	Cohen's d 0.23	1	No difference	High
			Isokinetic muscle strength	End of treatment (26 weeks)	Knee extensors 60 degree/s	Peak isokinetic torque (Nm)	17/17	110 (20)	89.2 (15.7)	NR	NR	No difference	High
			Isokinetic muscle strength	End of treatment (26 weeks)	Knee flexors 60 degree/s	Peak isokinetic torque (Nm)	17/17	57.5 (12.4)	47.7 (10.7)	NR	NR	Favours intervention	High
			Isokinetic muscle strength	End of treatment (26 weeks)	Knee extensors 180 degree/s ***	Peak isokinetic torque (Nm)	17/17	65.2 (11.0)	54.5 (10.2)	NR	NR	No difference	High
			Isokinetic muscle strength	End of treatment (26 weeks)	Knee flexors 180 degree/s	Peak isokinetic torque (Nm)	17/17	40.6 (11.3)	35.5 (7.8)	NR	NR	No difference	High
			Isokinetic muscle strength	Baseline, end of treatment (26 weeks)	Knee extensors 60 degree/s	absolute mean change (SD)	17/17	6.5 (9.5)	3.4 (7.6)	Cohen's d 0.37	0.188	No difference	High
			Isokinetic muscle strength	Baseline, end of treatment (26 weeks)	Knee flexors 60 degree/s	absolute mean change (SD)	17/17	6.4 (6.8)	1.2 (8.0)	Cohen's d 0.70	0.009	Favours intervention	High
			Isokinetic muscle strength	Baseline, end of treatment (26 weeks)	Knee extensors 180 degree/s	absolute mean change (SD)	17/17	4.0 (5.5)	3.1 (7.2)	Cohen's d 0.13	0.584	No difference	High
			Isokinetic muscle strength	Baseline, end of treatment (26 weeks)	Knee flexors 180 degree/s	absolute mean change (SD)	17/17	4.1 (2.6)	1.1 (3.2)	Cohen's d 1.00	0.113	No difference	High
			Footnotes:	*Nonparametric data presented as median (IQR), and as a post hoc, the Bonferroni test was used in multiple comparisons between pairs ** Between group comparison (Post Hoc Dunn test) *** significant difference between groups at baseline for this measure									

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Abbreviations: C, Comparator; I, intervention; MSQ, menopausal symptoms questionnaire; Nm, Newtons per metre; NR, not reported;													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control - No studies found													
Pilates vs 'other'													
Altan 2009	Fibromyalgia (women)	Pilates vs Home exercise (relaxation/stretching)	QoL - disease specific	End of treatment (12 weeks)	FIQ-total score	Higher score means worse health	25/25	63.5 (19.6)	77.5 (21.4)	NR	0.01	Favours intervention	Some concerns
			Tenderness	End of treatment (12 weeks)	number of tender points	Higher score means more tenderness	25/25	13.2 (3.6)	14.1 (4.5)	NR	0.481	No difference	Some concerns
			Footnotes:										
de Medeiros 2020	Fibromyalgia (women)	Pilates vs Aqua Aerobics	QoL - disease specific	End of treatment (12 weeks)	FIQ-total score	Higher score means worse health	21/21	51 (17)	58 (16)	0.67 (-0.37, 1.72)*	0.2*	No difference	Some concerns
			Sleep quality	End of treatment (12 weeks)	Pittsburgh Sleep quality index	Higher means worse sleep quality	21/21	9.9 (3.7)	9.5 (3.7)	-0.33 (-2.7, 2.0)*	0.77*	No difference	Some concerns
			Footnotes: *Difference between change from baseline scores										
Ekici 2014	Fibromyalgia (women)	Pilates vs Connective tissue massage	QoL - disease specific	End of treatment (4 weeks)	FIQ-total score	Higher score means worse health	15/21	22.12 (4.58)	28.68 (14.22)	NR	0.205	No difference	High
			Functional capacity	End of treatment (4 weeks)	FIQ - function	Higher score means worse health	15/21	1.06 (0.99)	2.05 (1.59)	NR	0.054	No difference	High
			Pain	End of treatment (4 weeks)	FIQ - pain	Higher score means worse health	15/21	2.43 (1.21)	3.38 (2.22)	NR	0.258	No difference	High
			Sleep quality	End of treatment (4 weeks)	FIQ - morning rest	Higher score means worse health	15/21	2.83 (1.23)	4.67 (2.73)	NR	0.064	No difference	High
			Stiffness	End of treatment (4 weeks)	FIQ - stiffness (0–10)	Higher score means worse health	15/21	4.17 (1.77)	3.69 (2.85)	NR	0.401	No difference	High
			Footnotes: *data not adjusted for differences in baseline values. Adjustment for differences in baseline values did not change significance of results.										
Abbreviations: C, Comparator; FIQ, Fibromyalgia Impact Questionnaire; I, intervention; NR, not reported;													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Cruz-Diaz 2015	Low back pain for more than 12 weeks (women, >65 yrs)	Pilates vs Control (no intervention) *adjunct to physiotherapy	Pain	End of treatment (6 weeks)	Numeric rating scale (0-10)	Higher score means more pain intensity	51/ 52	NR*	NR*	t(95) = 6.91	<0.000	<i>Favours intervention</i>	Some concerns
			Footnotes:	*Data presented in graphs and not extracted here. Authors note a significant between group difference.									
Cruz-Diaz 2016	Low back pain for more than 12 weeks (women, 45-75 yrs)	Pilates vs Control (no intervention) *adjunct to physiotherapy	Pain	End of treatment (6 weeks)	Numeric rating scale (0-10)	Higher score means more pain intensity	57/ 55	3.81 (1.21)	5.69 (1.63)	NR	NR	<i>Favours intervention</i>	Some concerns
			Functional disability	End of treatment (6 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	57/ 55	16.55 (2.24)	19.29 (3.34)	NR	NR	<i>Favours intervention</i>	Some concerns
			Footnotes:	*Only 53 participants in the Pilates group and 48 participants in the control group were analysed (because they had completed the protocol).									
Cruz-Diaz 2017	Low back pain for more than 12 weeks (adults, 18-50 yrs)	Pilates vs Control (no intervention)	Pain	End of treatment (12 weeks)	VAS (0-10)	Higher score means more pain intensity	68^/34	1.9 (1.39)	4.96 (1.31)	NR	NR		
			Pain	End of treatment (12 weeks)	VAS (0-10)	Higher score means more pain intensity	34/34 *	2.1 (1.36)	4.96 (1.31)	NR	NR	<i>Favours intervention</i>	High
			Pain	End of treatment (12 weeks)	VAS (0-10)	Higher score means more pain intensity	34/34 **	1.70 (1.41)	4.96 (1.31)	NR	NR	<i>Favours intervention</i>	High
			Functional disability	End of treatment (12 weeks)	RMDQ (0-24)	Higher score means more disability	68^/34	5.57 (5.13)	10.41 (5.6)	NR	NR		
			Functional disability	End of treatment (12 weeks)	RMDQ (0-24)	Higher score means more disability	34/34*	6.35 (5.3)	10.41 (5.6)	NR	NR	<i>Favours intervention</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Functional disability	End of treatment (12 weeks)	RMDQ (0-24)	Higher score means more disability	34/30**	4.79 (4.9)	10.41 (5.6)	NR	NR	<i>Favours intervention</i>	High
			Footnotes:	* Grp1: Pilates mat vs control ** Grp2: Equipment based Pilates vs control. ^Groups combined as per Cochrane Chapter 6 (values in bold)									
Cruz-Diaz 2018	Low back pain for more than 12 weeks (adults, 18-50 yrs)	Pilates vs Control (no intervention)	Pain	End of treatment (12 weeks)	VAS (0-10)	Higher score means more pain intensity	32/30*	1.95 (95% CI: 1.81-2.37)	4.35 (95% CI: 4.31-5.21)	NR	<0.001	<i>Favours intervention</i>	High
			Functional disability	End of treatment (12 weeks)	RMDQ (0-24)	Higher score means more disability	32/30*	5 (95% CI: 3.51-4.87)	9 (95% CI: 8.8-10.13)	NR	<0.001	<i>Favours intervention</i>	High
			Footnotes:	*Non-normal distributed data, values are expressed as median and 95% CI.									
da Fonesca 2019	Low back pain for more than 6 months	Pilates vs Control (usual activities)	Pain	End of treatment (7 weeks)	VAS (0-10)	Higher score means more pain intensity	8/9	3.0 (3.4)	4.9 (2.5)	NR	NR	<i>Not reported</i>	High
			Footnotes:										
Gladwell 2006	Low back pain for more than 12 weeks (adults)	Pilates vs Control (usual activities)	Pain	End of treatment (6 weeks)	Roland Morris pain rating scale	Higher score means more pain intensity	20/14	2.2 (0.9)	2.4 (0.9)	NR	0.05	<i>Favours comparator</i>	High
			Functional disability	End of treatment (6 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	20/14	18.1 (11.2)	18.1 (13.0)	NR	0.05	<i>Favours comparator</i>	High
			Quality of Life	End of treatment (6 weeks)	SF-12-Symptom report	Higher score means better QoL *	20/14	2.2 (0.6)	2.3 (0.6)	NR	NR	<i>No difference</i>	High
			Quality of Life	End of treatment (6 weeks)	SF-12 - General health	Higher score means better QoL *	20/14	3.7 (0.7)	3.6 (1.0)	NR	NR	<i>No difference</i>	High
			Quality of Life	End of treatment (6 weeks)	SF-12 - physical functioning	Higher score means better QoL *	20/14	3.2 (0.3)	3.0 (0.7)	NR	NR	<i>No difference</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Quality of Life	End of treatment (6 weeks)	SF-12 - role-functioning	Higher score means better QoL *	20/14	3.2 (0.8)	3.4 (0.6)	NR	NR	<i>No difference</i>	High
			Quality of Life	End of treatment (6 weeks)	SF-12 Social functioning	Higher score means better QoL *	20/14	3.5 (0.6)	3.9 (0.8)	NR	NR	<i>No difference</i>	High
			Quality of Life	End of treatment (6 weeks)	SF-12 - Bodily pain	Higher score means better QoL *	20/14	3.4 (1.0)	2.8 (0.7)	NR	NR	<i>No difference</i>	High
			Quality of Life	End of treatment (6 weeks)	SF-12 - Health perception	Higher score means better QoL *	20/14	2.5 (0.9)	2.9 (0.7)	NR	NR	<i>No difference</i>	High
			Quality of Life	End of treatment (6 weeks)	SF-12 - Sports functioning	Higher score means better QoL *	20/14	3.0 (0.8)	2.4 (0.9)	NR	NR	<i>No difference</i>	High
			Footnotes:	* Data not able to be interpreted. SF-12 scores should be on a scale of 0-100. The SF-12 items should provided a PCS and MCS summary, not individual subscales.									
Hasanpour-Dehkordi 2017	Low back pain for more than 12 weeks (men)	Pilates vs Control (no intervention)	Pain	End of treatment (6 weeks)	McGill Pain Questionnaire (0-78)	Higher score means more pain intensity	12/12	13.25 (6.38)	36.00 (13.84)	NR	NR*	<i>Not reported</i>	High
			Pain	baseline, end of treatment (6 weeks)	McGill Pain Questionnaire (0-78)	Absolute mean change (SD)	12/12	-8.17 (5.7)	1.83 (1.99)	NR	NR*	<i>Favours intervention</i>	High
			Footnotes:	*significant difference between groups at baseline (skewed data)									
Kofotolis 2016	Low back pain for more than 3 months	Pilates vs Control (no intervention)	Pain	End of treatment (4 weeks)*	SF-36 Bodily pain	Higher score means better pain-related QoL	37/28*	79.14 (7.93)	41.61 (16.05)	NR	NR	<i>Favours intervention</i>	Some concerns
			Functional disability	End of treatment (4 weeks)*	RMDQ (0-24)	Higher score means more disability	37/28*	3.32 (1.78)	10.09 (4.55)	NR	NR	<i>Favours intervention</i>	Some concerns
			Footnotes:	*3-month followup data not reported here.									

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Lopes 2014	Low back pain for more than 12 weeks (students)	Pilates vs Control (waitlist)	Pain	End of treatment (immediately after one session)	VAS (0-10)	Higher score means more pain intensity	23/23	1.7 (1.4)	1.7 (1.3)	NR	0.228	<i>No difference</i>	Low
			Footnotes:										
Mazloun 2018b	Low back pain for more than 12 weeks (adults)	Pilates vs Control (usual activities)	Functional disability	End of treatment (6 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	16/16	22.7 (3.1)	26.6 (4.9)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (6 weeks)	VAS (0-10)	Higher score means more pain intensity	16/16	3.4 (1.0)	6.6 (1.3)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Footnotes:										
Miyamoto 2011	Low back pain for more than 12 weeks (adults)	Pilates vs Control (education booklet)	Pain	End of treatment (6 weeks)	VAS (0-10)	Higher score means worse pain	43/43	3.1 (2.3)	5.2 (2.3)	2.2 (1.1-3.2)*	<0.01	<i>Favours intervention</i>	Some concerns
			Functional disability	End of treatment (6 weeks)	RMDQ (0-24)	Higher score means more disability	43/43	3.6 (3.4)	7.1 (5.7)	2.7 (1.0-4.4)*	<0.01	<i>Favours intervention</i>	Some concerns
			Function (patient specific)	End of treatment (6 weeks)	Patient-Specific Functional Scale (0-10)	Higher score means greater ability to perform activity at pre-injury level	43/43	7.5 (2.1)	6.4 (2.0)	-0.4 (-1.3 to 0.4)*	0.35	<i>No difference</i>	Some concerns
			Footnotes: *Adjusted mean difference (95% CI) reported that incorporate time, group interactions										
Miyamoto 2016	Low back pain for more than 12 weeks (adults)	Pilates vs Control (education booklet)*	Function (patient specific)	End of treatment (6 weeks)	Patient-specific functional scale (0-10)	Higher score means greater ability to perform activity at pre-injury level	222/73	6.7 (2.1)	5.0 (2.6)	NR	NR	<i>Favours intervention</i>	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Function (patient specific)	End of treatment (6 weeks)	Patient-specific functional scale (0-10)	Higher is better	74/73	6.3 (2.0)	5.0 (2.6)	1.2 (0.4 to 2.0)	<0.01	<i>Favours intervention</i>	Some concerns
			Function (patient specific)	End of treatment (6 weeks)	Patient-specific functional scale (0-10)	Higher is better	74/73	6.9 (2.1)	5.0 (2.6)	1.8 (1.0 to 2.5)	<0.01	<i>Favours intervention</i>	Some concerns
			Function (patient specific)	End of treatment (6 weeks)	Patient-specific functional scale (0-10)	Higher is better	74/73	6.8 (2.1)	5.0 (2.6)	1.4 (0.6 to 2.3)	<0.01	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (6 weeks)	VAS (0-10)	Higher score means more pain intensity	222/73	3.5 (2.5)	5.6 (2.6)	NR	NR	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (6 weeks)	VAS (0-10)	Higher score means more pain intensity	74/73	4.0 (2.7)	5.6 (2.6)	-1.2 (-2.2 to -0.3)	<0.01	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (6 weeks)	VAS (0-10)	Higher score means more pain intensity	74/73	3.3 (2.3)	5.6 (2.6)	-2.3 (-3.2 to -1.4)	<0.001	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (6 weeks)	VAS (0-10)	Higher score means more pain intensity	74/73	3.2 (2.4)	5.6 (2.6)	-2.1 (-3.0 to -1.1)	<0.001	<i>Favours intervention</i>	Some concerns
			Functional disability	End of treatment (6 weeks)	RMDQ (0-24)	Higher score means more disability	222/73	6.7 (4.82)	11.3 (6.1)	NR	NR	<i>Favours intervention</i>	Some concerns
			Functional disability	End of treatment (6 weeks)	RMDQ (0-24)	Higher score means more disability	74/73	7.8 (5.2)	11.3 (6.1)	-1.9(-3.6 to -0.1)	<0.05	<i>Favours intervention</i>	Some concerns
			Functional disability	End of treatment (6 weeks)	RMDQ (0-24)	Higher score means more disability	74/73	6.8 (2.1)	11.3 (6.1)	-4.7(-6.4 to -3.0)	<0.001	<i>Favours intervention</i>	Some concerns
			Functional disability	End of treatment (6 weeks)	RMDQ (0-24)	Higher score means more disability	74/73	6.1 (5.5)	11.3 (6.1)	-3.3(-5.0 to 1.6)	<0.001	<i>Favours intervention</i>	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Quality of Life	End of treatment (6 weeks)	SF-6D	Higher score means better QoL	222/73	0.82 (0.08)	0.78 (0.08)	NR	NR	<i>Favours intervention</i>	Some concerns
			Quality of Life	End of treatment (6 weeks)	SF-6D	Higher score means better QoL	74/73	0.81 (0.08)	0.78 (0.08)	0.03 (0.01 to 0.06)	<0.01	<i>Favours intervention</i>	Some concerns
			Quality of Life	End of treatment (6 weeks)	SF-6D	Higher score means better QoL	74/73	0.81 (0.07)	0.78 (0.08)	0.03 (0.01 to 0.06)	<0.01	<i>Favours intervention</i>	Some concerns
			Quality of Life	End of treatment (6 weeks)	SF-6D	Higher score means better QoL	74/73	0.83 (0.09)	0.78 (0.08)	0.04 (0.01 to 0.06)	<0.001	<i>Favours intervention</i>	Some concerns
			Footnotes:	*4 study arms: one control and three pilates arms (difference being number of sessions per week). Intervention arms were combined according to Chapter 6 Cochrane handbook (values in bold).									
Natour 2011	Low back pain for more than 12 months (adults)	Pilates vs Control (no intervention) as adjunct to usual medical care (NSAIDS)	Pain	End of treatment (12 weeks)	VAS (0-10)	Higher score means more pain intensity	30/30	4.04 (2.42)	5.16 (2.53)	ES -0.57 (-1.08 to 0.05)	<0.001	<i>Favours intervention</i>	Some concerns
			Functional disability	End of treatment (12 weeks)	RMDQ (0-24)	Higher score means more disability	30/30	6.79 (5.34)	10.59 (5.88)	ES -0.67 (-1.19 to 0.15)	<0.001	<i>Favours intervention</i>	Some concerns
			Quality of Life	End of treatment (12 weeks)	SF-36 physical functioning	Higher score means better QoL	30/30	65.83 (27.96)	57.29 (18.29)	MD: 8.54	ES 0.24 (-0.27 to 0.75)	<i>Favours intervention</i>	Some concerns
			Quality of Life	End of treatment (12 weeks)	SF-36 role physical	Higher score means better QoL	30/30	49 (37.27)	42.66 (34.57)	MD: 6.34	NR	<i>No difference</i>	Some concerns
			Quality of Life	End of treatment (12 weeks)	SF-36 bodily pain	Higher score means better QoL	30/30	54.45 (23.41)	46.41 (25.83)	MD: 8.04	EF 0.30 (-0.21 to 0.81)	<i>Favours intervention</i>	Some concerns
			Quality of Life	End of treatment (12 weeks)	SF-36 general health perceptions	Higher score means better QoL	30/30	68.58 (21.92)	57.70 (18.86)	MD: 10.88	NR	<i>No difference</i>	Some concerns
			Quality of Life	End of treatment (12 weeks)	SF-36 vitality	Higher score means better QoL	30/30	64.58 (21.15)	54.00 (20.02)	MD 10.58	ES 0.23 (-0.28 to 0.74)	<i>Favours intervention</i>	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Quality of Life	End of treatment (12 weeks)	SF-36 role social	Higher score means better QoL	30/30	83.75 (24.51)	79.52 (25.49)	MD: 4.23	NR	No difference	Some concerns
			Quality of Life	End of treatment (12 weeks)	SF-36 role emotional	Higher score means better QoL	30/30	80.43 (29.72)	73.75 (29.56)	MD: 6.68	NR	No difference	Some concerns
			Quality of Life	End of treatment (12 weeks)	SF-36 mental health	Higher score means better QoL	30/30	69.30 (21.14)	60.63 (23.23)	MD 8.67	NR	No difference	Some concerns
			Analgaesic use (NSAID)	End of treatment (12 weeks)	Patient record (sodium diclofenac)	Lower is better	30/30	6.7 (12.77)	12.36 (18.59)	ES -0.48 (-1.00 to 0.03)	<0.01	Favours intervention	Some concerns
			Footnotes:	NSAIDS=non-steroidal antiinflammatory drugs. ES=effect size.									
Patti 2016	Low back pain for more than 12 months (adults)	Pilates vs Control (no intervention)	Functional disability	End of treatment (14 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	19/19	6.5 (4)	8.4 (7.8)	NR	NR	Not reported	Some concerns
			Footnotes:										
Quinn 2011	Low back pain for more than 12 weeks (adults)	Pilates vs Control (no intervention)	Pain	End of treatment (8 weeks)	VAS (0-100)	absolute mean change (range)*	15/14	-9.5 (-16 to 45)	4.7 (-35 to 24)	NR	0.047	Favours intervention	Some concerns
			Functional disability	End of treatment (8 weeks)	RMDQ (0-24)	absolute mean change (range)**	15/14	-5.4	-0.21	NR	0.301	Favours intervention	Some concerns
			Footnotes:	*Authors report mean change (range) which is not considered reliable to estimate the SD. (See Cochrane Chapter 6) ** Authors report mean change only.									
Rydeard 2006	Low back pain for more than 6 weeks (adults)	Pilates vs Control (usual care)	Pain	End of treatment (8 weeks)	VAS (0-100)	Higher score means more pain intensity	21/18*	18.3 (3.2)	33.9 (3.5)	NR	0.002	Favours intervention	Some concerns
			Functional disability	End of treatment (8 weeks)	RMDQ (0-24)	Higher score means more disability	21/18*	2.0 (0.3)	3.2 (0.4)	NR	0.023	Favours intervention	Some concerns
			Footnotes:	*Data are mean (SEM).									

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Valenza 2017	Low back pain for more than 12 weeks (adults)	Pilates vs Control (usual activities and educational leaflet)	Functional disability	End of treatment (8 weeks)	Oswestry Disability Index (0-100)	absolute mean change (SD)*	27/27	16.35 (14.07)	4.5 (20.52)	MD 12.32 (18.5) (95% CI 6.47 to 19.41)	<0.001	Favours intervention	Some concerns
			Functional disability	End of treatment (8 weeks)	RMDQ (0-24)	absolute mean change (SD)*	27/27	5.31 (3.37)	2.4 (6.78)	MD 3.2 (4.12) (95% CI 1.02 to 5.46)	0.003	Favours intervention	Some concerns
			Pain (current)	End of treatment (8 weeks)	VAS (0-10)	absolute mean change (SD)*	27/27	2.3 (1.9)	0.9 (2.8)	MD 1.4 (2.3) (95% CI 0.7 to 2.9)	0.002	Favours intervention	Some concerns
			Footnotes:	*Mean change from baseline reported (only pre-intervention mean SD reported and not post intervention mean SD).									
Zeada 2012	Low back pain for more than 3 months (adults)	Pilates vs control (usual care)	Functional disability	End of treatment (8 weeks)	RMDQ (0-24)	Higher score means more disability	10/10	4.65 (2.8)	6.35 (1.3)	NR	<0.05*	Favours intervention	Some concerns
			Footnotes:	* Figures under <0.05 were reported as significant, actual figures not reported									
Pilates vs 'other'													
Anand 2014	Low back pain for more than 12 weeks (adults)	Pilates vs therapeutic and general flexibility exercises	Functional disability	End of treatment (8 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	15/15	41.36 (2.10)	64.66 (3.72)	NR	NR	Favours intervention	High
			Pain	End of treatment (8 weeks)	VAS (0-10)	Higher score means worse pain	15/15	3.93 (0.92)	6.53 (0.56)	NR	NR	Favours intervention	High
			Footnotes:	Unpaired 'T' test value was 13.44.									
Avila Ribeiro 2015	Low back pain for more than 12 weeks (adults)	Pilates vs classical kinesiotherapy	Functional disability	End of treatment (10 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	3/2	3 (2.7)	2 (2.8)	NR	0.767	No difference	High
			Pain	End of treatment (10 weeks)	VAS (0-10)	Higher score means worse pain	3/2	0 (0)	0 (0)	NR	0.999	No difference	High
			Footnotes:										
Bhaduria 2017	Low back pain for more than 12 weeks (adults)	Pilates vs lumbar stabilisation exercise	Functional disability	End of treatment (3 weeks)	Modified Oswestry Disability Index	Higher score means more disability in ADL	15/15	8.42 (5.14)	6.92 (2.47)	NR	NR	Not reported	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
		*as adjunct to conventional treatment	Pain	End of treatment (3 weeks)	VAS (0-10)	Higher score means more pain	15/15	1.33 (0.98)	1.17 (0.72)	NR	NR	<i>Not reported</i>	High
			Footnotes:										
Bhaduria 2017	Low back pain for more than 12 weeks (adults)	Pilates vs dynamic strengthening exercise	Functional disability	End of treatment (3 weeks)	Modified Oswestry Disability Index	Higher score means more disability in ADL	15/14	8.42 (5.14)	23.42 (11.1)	NR	NR	<i>Not reported</i>	High
		*as adjunct to conventional treatment	Pain	End of treatment (3 weeks)	VAS (0-10)	Higher score means more pain	15/14	1.33 (0.98)	2.00 (1.35)	NR	NR	<i>Not reported</i>	High
			Footnotes:										
Brooks 2012	Low back pain for more than 12 weeks (adults)	Pilates vs stationary cycling exercise	Functional disability	End of treatment (8 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	32/32	(-)10.4 (95% CI - 14.2 to -6.7)	(-)3.9(95% CI - 7.8 to 0)	MD -6.8 (95% CI 0.9 to 11.9)	<0.05	<i>Favours intervention</i>	High
			Pain	End of treatment (8 weeks)	VAS (0-10)	Higher score means more pain	32/32	(-)1.9 (95% CI - 2.6 to -1.2)	(-) 0.8 (95% CI - 1.5 to -0.1)	MD -1.1 (95% CI 0.1 to 2.1)	<0.05	<i>Favours comparator</i>	High
			Footnotes: Results reported are adjusted for baseline values. 6 month follow-up reported in Marshall 2013.										
Devasahayam 2016	Low back pain and longterm unilateral musculoskeletal injury to lower limb (adults)	Pilates vs gym-based exercises determined as per participant needs	Function (patient specific)	End of treatment (6 weeks)	Patient-specific functional scale	Higher score means greater ability to perform activity at pre-injury level	9/6	8.00 (2.00)	8.28 (1.03)	0.28 (change score)	0.277	<i>No difference</i>	High
			Pain	End of treatment (6 weeks)	Numeric rating scale (0-10)	Higher score means more pain	9/6	3.00 (2.10)	2.33 (2.00)	0.67 (change score)	0.837	<i>No difference</i>	High
			Footnotes: Pilates N = 14 and control N = 10 participants assigned to each group.										
Donzelli 2006	Low back pain for more than 12 weeks (adults)	Pilates vs Back School	Functional disability	End of treatment (10 days)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	21/22	NR	NR	NR	NR	<i>Not reported</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Dsa 2014	Low back pain for more than 12 weeks (adults)	Pilates vs Core stabilisation exercises	Pain	End of treatment (10 days)	VAS (0-10)	Higher score means more pain intensity	21/22	NR	NR	NR	NR	Not reported	High
			Footnotes:	Results presented as graph so no values could be extracted.									
			Functional disability	End of treatment (2 weeks)	RMDQ (0-24)*	Higher score means more disability	17/21	39.32 (14.670)	9.32 (8.781)	NR	<0.001	Favours intervention	High
			Pain	End of treatment (2 weeks)	VAS (0-10)	Higher score means more pain intensity	16/17	6.08 (1.882)	2.58 (1.352)	NR	<0.001	Favours intervention	High
Footnotes:	*Not clear what RMDQ is used. Typically max score of 24, but intervention group has mean sore of 39.32 (max score 67) so may have been transformed?												
Gonzalez-Galvez 2019	Low back pain for more than 12 months (adolescents)	Pilates vs control (education session)	No data extracted because the study does not include any priority outcome domains or measures.										
Hasanpour-Dehkordi 2017	Low back pain for more than 12 weeks (men)	Pilates vs Mckenzie training	Pain	End of treatment (6 wks)	McGill Pain Questionnaire (0-78)	Higher score means more pain intensity	12/12	13.25 (6.38)	19.25 (7.46)	NR	NR	Not reported	High
			Pain	baseline, end of treatment (6 weeks)	McGill Pain Questionnaire (0-78)	Absolute mean change (SD)	12/12	-8.17 (5.7)	-6.25 (1.84)	NR	NR*	Not reported	High
			Footnotes:										
Kofotolis 2016	Low back pain for more than 3 months	Pilates vs strengthening exercises	Pain	End of treatment (4 weeks)*	SF-36 Bodily pain	Higher score means better pain-related QoL	37/36	79.14 (7.93)	71.28 (11.27)	NR	NR	Not reported	Some concerns
			Functional disability	End of treatment (4 weeks)*	RMDQ (0-24)	Higher score means more disability	37/36	3.32 (1.78)	4.88 (1.60)	NR	NR	Not reported	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Footnotes:	*3-month followup data not reported here.									
Mazloum 2018b	Low back pain for more than 12 weeks (adults)	Pilates vs Extension based exercises	Functional disability	End of treatment (6 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	16/15	22.7 (3.1)	23.2 (7.6)	NR	0.05	Not reported	Some concerns
			Pain	End of treatment (6 weeks)	VAS (0-10)	Higher score means more pain intensity	16/15	3.4 (1.0)	5.3 (1.3)	NR	0.001	Not reported	Some concerns
			Footnotes:										
Mostagi 2015	Low back pain for more than 12 weeks (adults)	Pilates vs General exercise	Functional disability	End of treatment (8 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	11/11	md 0.4 (IQR 0; 3.5)	md 0.50 (IQR 0; 2.1)	NR	NR	Not reported	Some concerns
			Pain	End of treatment (8 weeks)	VAS (0-10)	Higher score means more pain intensity	11/11	15.5 (10.3)	11.7 (9)	MD 17.7 (95% CI 7.5 to 27.9)	0.02	No difference	Some concerns
			Footnotes:	IQR, interquartile range; md: median; MD = mean difference									
Rajpal 2008	Low back pain for more than 12 weeks (women)	Pilates vs. McKenzie Method	Functional disability	End of treatment (4 weeks)	Back performance scale (Sock test, Pick-up test, Roll-up test, Fingertip to floor test and lift test)	Higher score means more disability in daily living activities	17/15	1.941 (0.555)	1.533 (0.639)	NR	0.063	No difference	Some concerns
			Pain	End of treatment (4 weeks)		VAS (0-10)	Higher score means more pain intensity	17/15	3.352 (0.931)	2.000 (1.000)	NR	0.00	Favours intervention
			Footnotes:	Back performance score t value (1.930) and VAS t value (3.692)									
Silva 2018	Low back pain for more than 12 weeks (adults)	Pilates vs conventional stretching and strengthening	Functional disability	End of treatment (8 weeks)	Oswestry Disability Index (0-100)	Higher score means more disability in ADL	8/8	24.75 (11.56)	37.75 (19.25)	NR	0.1238	No difference	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
		exercises	Pain	End of treatment (8 weeks)	VAS (0-10)	Higher score means more pain intensity	8/8	2.00 (2.56)	3.25 (3.27)	NR	0.4177	No difference	Some concerns
			Footnotes:										
Wajswelner 2011	Low back pain for more than 12 weeks (adults)	Pilates vs individualised exercise program based on conventional physiotherapy	Function (patient specific)	End of treatment (6 weeks)	Patient-Specific Functional Scale (3-items)	Higher score means better function (max score=30)	44/43	19.0 (6.2)	18.9 (5.9)	MD 1.0 (95% CI -1.5 to 3.5)	NR	No difference	High
			Functional disability	End of treatment (6 weeks)	Quebec Back Pain Disability Scale (20-item)	Higher score means more disability in daily living activities	44/43	15.3 (9.1)	17.1 (13.4)	MD -3.5 (95% CI -7.3 to 0.3)	NR	No difference	High
			Pain	End of treatment (6 weeks)	Numeric rating scale (0-10)	Higher score means more pain intensity	44/43	2.8 (1.6)	3.2 (2.1)	MD -0.5 (95% CI -1.3 to 0.3)	NR	No difference	High
			Pain confidence	End of treatment (6 weeks)	Pain Self-efficacy Questionnaire	Higher score means more confidence in managing pain	44/43	51.2 (10.4)	50.7 (8.0)	MD 2.1 (95% CI -0.8 to 5.1)	NR	No difference	High
			Participant perceived global change in pain	End of treatment (6 weeks)	5-point Likert scale	higher is better	44/43	68.5 (22.3)	71.9 (18.9)	MD -0.8 (95% CI -9.6 to 8.1)	NR	No difference	High
			Footnotes: MD = difference between groups, adjusted mean (means adjusted for baseline scores)										
Abbreviations: ADL, activites of daly living; C, Comparator; I, intervention; MD, mean difference; NR, not reported; RMDQ, Roland Morris Disability Questionnaire													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Cazotti 2015	Chronic mechanical neck pain	Pilates vs Control (usual care)	Function	End of treatment (12 weeks)	Neck Disability Index (0-50)	Higher score means more disability	32/32	3.56 (3.74)	10.59 (7.19)	NR	NR	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (12 weeks)	Numeric Pain Scale (0-10)	Higher score means more pain	32/32	1.3 (1.66)	5.47 (2.09)	NR	NR	<i>Favours intervention</i>	Some concerns
			Physical function	End of treatment (12 weeks)	SF-36 physical functioning	Higher score means better outcome	32/32	84.2 (11.4)	72.8 (20.4)	NR	0.019	<i>Favours intervention</i>	Some concerns
			Role - Physical	End of treatment (12 weeks)	SF-36 role physical	Higher score means better outcome	32/32	87.5 (26.7)	60.5 (45.5)	NR	0.072	<i>No difference</i>	Some concerns
			Pain	End of treatment (12 weeks)	SF-36 bodily pain	Higher score means better outcome	32/32	66.3 (20.5)	52.4 (16.8)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			General health	End of treatment (12 weeks)	SF-36 general health perceptions	Higher score means better outcome	32/32	80.0 (22.4)	75.6 (25.5)	NR	0.022	<i>Favours intervention</i>	Some concerns
			Vitality	End of treatment (12 weeks)	SF-36 vitality	Higher score means better outcome	32/32	69.8 (22.5)	57.2 (22.0)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Role-social	End of treatment (12 weeks)	SF-36 role social	Higher score means better outcome	32/32	93.8 (11.9)	76.2 (28.3)	NR	0.382	<i>No difference</i>	Some concerns
			Role-emotional	End of treatment (12 weeks)	SF-36 role emotional	Higher score means better outcome	32/32	80.2 (36.8)	72.9 (39.2)	NR	0.172	<i>No difference</i>	Some concerns
			Mental health	End of treatment (12 weeks)	SF-36 mental health	Higher score means better outcome	32/32	79.3 (15.9)	65.2 (19.8)	NR	0.012	<i>Favours intervention</i>	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Footnotes:	p-value generated from ANOVA comparison over time between groups. (T0, T45, T90 and T180 days).									
Dunleavy 2016	Chronic neck pain (NOS)	Pilates vs Control (no intervention)	Pain	End of treatment (12 weeks)	Numeric Pain Scale (0-10)	Higher score means more pain	20/17	1.9 (1.6)	3.9 (1.6)	NR	NR	<i>Not reported</i>	High
			Disability	End of treatment (12 weeks)	Neck Disability Index (0-50)	Higher score means greater diasbility	20/17	6.8 (4.3)	12.5 (6.8)	NR	NR	<i>Not reported</i>	High
			Footnotes:	Numeric pain scale is average pain compared using analysis of covariance with mean baseline average pain as a covariate.									
Pilates vs 'other'													
Atilgan 2017	Persistent shoulder pain for at least four weeks	Pilates vs Conventional exercises	Pain	End of treatement (10 days)	Visual analogue scale (rest)	Higher score means worse pain	15/15	1.06 (1.48)	2.24 (2.39)	NR	0.002	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (10 days)	Visual analogue scale (during activity)	Higher score means worse pain	15/15	0.98 (1.73)	2.40 (2.06)	NR	0.002	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (10 days)	Shoulder Pain and Disability Index - Pain	Higher score means worse pain	15/15	33.90 (20.12)	44.26 (19.82)	NR	0.012	<i>Favours intervention</i>	Some concerns
			Physical function/activity	End of treatment (10 days)	Shoulder Pain and Disability Index - Disability	Higher score means more disability	15/15	26.83 (18.95)	33.50 (22.42)	NR	0.085	<i>No difference</i>	Some concerns
			Footnotes:										
Dunleavy 2016	Chronic neck pain (NOS)	Pilates vs Yoga	Pain	End of treatment (12 weeks)	Numeric Pain Scale (0-10)	Higher score means more pain	20/17	1.9 (1.6)	2.3 (1.6)	NR	NR	<i>Not reported</i>	High
			Disability	End of treatment (12 weeks)	Neck Disability Index (0-50)	Higher score means more disability	20/17	6.8 (4.3)	8.1 (5.6)	NR	NR	<i>Not reported</i>	High
			Footnotes:	Numeric pain scale is average pain compared using analysis of covariance with mean baseline average pain as a covariate.									
Ulug 2018	Chronic neck pain (NOS)	Pilates vs Isometric exercise	Pain	End of treatment (6 weeks)	Short-Form McGill Pain Questionnaire	Higher score related to higher pain	20/20	2.4 (3.5)	2.2 (3.7)	NR	>0.05	<i>No difference</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Ulug 2018	Chronic neck pain (NOS)	Pilates vs Yoga	Disability	End of treatment (6 weeks)	Neck Disability Index	Higher score means more disability	20/20	10.0 (4.8)	11.3 (6.3)	NR	>0.05	No difference	High
			Footnotes:	Baseline values varied between groups.									
			Pain	End of treatment (6 wks)	Short-Form McGill Pain Questionnaire	Higher score related to higher pain	20/20	2.4 (3.5)	1.3 (2.2)	NR	>0.05	No difference	High
			Disability	End of treatment (6 wks)	Neck Disability Index	Higher score means more disability	20/20	10.0 (4.8)	8.2 (4.8)	NR	>0.05	No difference	High
			Footnotes:	Baseline values varied between groups.									
Abbreviations: C, Comparator; I, intervention; NOS, not otherwise specified; NR, not reported;													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Celik 2017	Rehabilitation after anterior cruciate ligament injury	Control (no intervention)	Knee function	End of treatment (12 weeks)	Cincinnati Knee Rating System	Higher score means better function	24/26	87.1 (9.0)	83.0 (12.5)	NR	Not reported	No difference	High
			Isokinetic muscle strength	End of treatment (12 weeks)	Knee flexion peak torque	Higher score means better knee strength	24/26	132.0 (17.2)	122.9 (31.9)	NR	Not reported	No difference	High
			Isokinetic muscle strength	End of treatment (12 weeks)	Knee extension peak torque	Higher score means better knee strength	24/26	182.6 (3.0)	158.7 (40.7)	NR	0.03	Favours intervention	High
			Improvement in stability	End of treatment (12 weeks)	Global rating of change	Proportion reporting improvement	24/26	24/24 (100%)	6/26 (23%)	NR	Not reported	Not reported	High
Footnotes:													
Pilates vs 'other' - no studies found													
Abbreviations: C, Comparator; I, intervention; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Karaman 2017*	Rehabilitation after total knee arthroplasty	Control (no intervention) *as adjunct to standard post-operative exercise program	Balance	End of treatment (6 wks)	Berg Balance Test (0-56)	Higher score means better balance	17/17	50.6 (3.9)	41.1 (5.1)	NR	NR**	<i>Favours intervention</i>	High
			Physical wellbeing	End of treatment (6 wks)	SF-36 Physical Component Score	Higher score means better wellbeing	17/17	44.2 (7.1)	37.5 (6.4)	NR	NR**	<i>Favours intervention</i>	High
			Mental wellbeing	End of treatment (6 wks)	SF-36 Mental Component Score	Higher score means better wellbeing	17/17	53.6 (10.4)	41.1 (12.7)	NR	NR**	<i>No difference</i>	High
			Physical function	End of treatment (6 wks)	SF-36 physical functioning	Higher score means better outcome	17/17	67.7 (18.9)	49.8 (19.6)	NR	NR**	<i>Favours intervention</i>	High
			Role - Physical	End of treatment (6 wks)	SF-36 role physical	Higher score means better outcome	17/17	64.2 (39.5)	27.9 (45)	NR	NR**	<i>Favours intervention</i>	High
			Pain	End of treatment (6 wks)	SF-36 bodily pain	Higher score means better outcome	17/17	59.1 (25.2)	45.1 (32.1)	NR	NR**	<i>No difference</i>	High
			General health	End of treatment (6 wks)	SF-36 general health perceptions	Higher score means better outcome	17/17	81 (16.8)	64.4 (18.8)	NR	NR**	<i>No difference</i>	High
			Vitality	End of treatment (6 wks)	SF-36 vitality	Higher score means better outcome	17/17	67.1 (20.4)	49.2 (22.7)	NR	NR**	<i>No difference</i>	High
			Role-social	End of treatment (6 wks)	SF-36 role social	Higher score means better outcome	17/17	81.6 (27.6)	50 (35.5)	NR	NR**	<i>No difference</i>	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Role-emotional	End of treatment (6 wks)	SF-36 role emotional	Higher score means better outcome	17/17	76.5 (40.4)	31.4 (46.4)	NR	NR**	No difference	High
			Mental health	End of treatment (6 wks)	SF-36 mental health	Higher score means better outcome	17/17	76.2 (15.8)	60.9 (22.1)	NR	NR**	No difference	High
			Footnotes:	*authors reported direction of effect is in relation to change scores ** data were not normally distributed at baseline, with nonparametric statistical tests used (details not provided)									
Pilates vs 'other' - no studies identified													
Abbreviations: C, Comparator; I, intervention; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Abavisani 2019	Employment conditions (at risk of anxiety)	Control (usual activities)	State anxiety	End of treatment (8 weeks)	Spielberger anxiety questionnaire	Higher score means more anxiety	31/31	52.74 (7.41)	49.63 (7.39)*	NR	0.001	Favours intervention	High
			Trait anxiety	End of treatment (8 weeks)	Spielberger anxiety questionnaire	Higher score means more anxiety	31/31	52.26 (7.30)	42.77 (7.71)	NR	0.001	Favours intervention	High
			Footnotes:	*authors reported 49.63 in body text and 46.35 in Table 2									
Pilates vs 'other' - no studies identified													
Abbreviations: C, Comparator; I, intervention; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Garcia-Soidan 2014	Sedentary behaviour	Control (no intervention)	Accelerometry	7 days before and 7 days after study	ActiGraph model GT3X® (count)	Higher score means more activity	51/48	13,095.8 (SE 1,291.2)	12,673.8 (SE 1,118.7)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Physical function	End of treatment (12 weeks)	SF-36 (functional capacity)	Higher scores indicate better state of health	51/48	87.6 (SE 1.16)	77.8 (SE 0.83)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Role - Physical	End of treatment (12 weeks)	SF-36 (physical aspects)	Higher scores indicate better state of health	51/48	86.6 (SE 0.75)	77.4 (SE 1.07)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (12 weeks)	SF-36 (bodily pain)	Higher scores indicate better state of health	51/48	66.1 (SE 0.90)	78.1 (SE 0.89)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			General health	End of treatment (12 weeks)	SF-36 (general health)	Higher scores indicate better state of health	51/48	81.0 (SE 1.51)	62.7 (SE 0.52)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Vitality	End of treatment (12 weeks)	SF-36 (vitality)	Higher scores indicate better state of health	51/48	71.0 (SE 0.12)	53.6 (SE 0.59*)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Role-social	End of treatment (12 weeks)	SF-36 (social aspects)	Higher scores indicate better state of health	51/48	85.2 (SE 1.3)	72.8 (SE 0.15)	NR	<0.001	<i>Favours intervention</i>	Some concerns
			Role-emotional	End of treatment (12 weeks)	SF-36 (emotional aspects)	Higher scores indicate better state of health	51/48	75.3 (SE 8.90)**	44.1 (SE 0.19)	NR	<0.001	<i>Favours intervention</i>	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p -value	direction of effect	RoB
			Mental health	End of treatment (12 weeks)	SF-36 (metal health)	Higher scores indicate better state of health	51/48	73.4 (SE 8.90)**	57.1 (SE 0.08)	NR	<0.001	Favours intervention	Some concerns
			Footnotes:	*error in reporting by authors. SE for vitality subscale was 53.6. SE for baseline used as a proxy due to minimal change. ** SE appears to be an error (repeated number in table may not be correct).									
Pilates vs 'other'													
Sahinci Gokgul 2017	Sedentary behaviour	Cyclic exercises	Anthropometric	End of treatment (8 weeks)	Hip circumference	cm	NR	97.5 (7.326)	100.9 (7.324)	NR	NR*	Favours intervention	High
			Anthropometric	End of treatment (8 weeks)	Waist circumference	cm	NR	74.8 (8.23)	80.5 (11.394)	NR	NR*	Favours intervention	High
			Anthropometric	End of treatment (8 weeks)	Chest	cm	NR	87.5 (6.587)	93.3 (11.605)	NR	NR*	Favours intervention	High
			Anthropometric	End of treatment (8 weeks)	Abdominal	cm	NR	89.1 (7.233)	91.5 (9.834)	NR	NR*	Favours intervention	High
			CVD risk	End of treatment (8 weeks)	HDL	mg/dl	NR	33.86 (0.889)	55.10 (12.433)	NR	NR*	No difference	High
			CVD risk	End of treatment (8 weeks)	LDL	mg/dl	NR	90.87 (30.062)	91.08 (55.998)	NR	NR*	No difference	High
			Footnotes:	*authors reported p value for pre-post comparison within treatment group.									
Abbreviations: C, Comparator; I, intervention; NR, not reported; SE, standard error													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Curi 2018	Women > 60 yrs at risk of physical or mental decline	Pilates vs Control (no intervention)	Sleep quality	End of treatment (16 weeks)	Pittsburgh sleep quality index	higher score means worse sleep quality	33/31	5.61 (2.93)	7.60 (5.77)	NR	NR	No difference	Some concerns
			General (mental) health	End of treatment (16 weeks)	GHQ-12	higher means better mental health	33/31	7.32 (4.36)	12.40 (6.23)	n2=0.19	<0.001*	Favours intervention	Some concerns
			Dynamic balance/mobility	End of treatment (16 weeks)	timed up and go (s)	higher means worse agility	33/31	6.34 (0.99)	7.86 (1.12)	F=9.388	0.003*	Favours intervention	Some concerns
			Aerobic endurance	End of treatment (16 weeks)	NR**	higher is worse	33/31	18.84 (2.99)	20.34 (3.25)	NR	NR	No difference	Some concerns
			Footnotes:	*data are between group X time interaction calculated using repeated measure ANOVA **authors mention the 6-minute walk test, which should be measured in distance (higher is better), but the reported results are measured in minutes (higher scores indicting worse aerobic endurance). It is assumed it is a distance-based walk test.									
de Andrade Mesquita 2015	Women > 60 yrs at risk of physical or mental decline	Pilates vs Control (usual activites)	Dynamic balance/mobility	baseline, end of treatment (4 weeks)	timed up and go (s)	higher means worse agility	20/18	7.7 (1.5)	13.9 (4.3)	NR	NR	No difference	Some concerns
			Balance	baseline, end of treatment (4 weeks)	Berg Balance test (0-56)	higher is better	20/18	56.0 (0.1)	51.0 (3.5)	NR	<0.05	Favours intervention	Some concerns
			Isokinetic muscle strength	baseline, end of treatment (4 weeks)	knee extensors (n/M)	higher is better	20/18	data presented in figures and not included here.				Favours intervention	Some concerns
			Isokinetic muscle strength	baseline, end of treatment (4 weeks)	knee flexors (n/M)	higher is better	20/18	data presented in figures and not included here.				Favours intervention	Some concerns
			Footnotes:										

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Gandolfi 2020	Women > 60 yrs at risk of physical or mental decline	Pilates vs Control (no intervention)	Quality of life	End of treatment (20 weeks)	SF-36 total score (0-100)	Higher score means better outcome	20/20	77.60 (4.86)	54.63 (15.86)	NR	<0.0001*	<i>Favours intervention</i>	Some concerns
			Physical wellbeing	End of treatment (20 weeks)	SF-36 physical component score	Higher score means better outcome	20/20	79.70 (3.83)	54.90 (15.05)	NR	<0.0001*	<i>Favours intervention</i>	Some concerns
			Emotional wellbeing	End of treatment (20 weeks)	SF-36 mental component score	Higher score means better outcome	20/20	74.10 (8.37)	54.18 (22.55)	NR	<0.0001*	<i>Favours intervention</i>	Some concerns
			Physical function	End of treatment (20 weeks)	SF-36 physical functioning	Higher score means better outcome	20/20	86.25 (9.58)	55.5 (20.83)	NR	0.0003*	<i>Favours intervention</i>	Some concerns
			Role - Physical	End of treatment (20 weeks)	SF-36 role physical	Higher score means better outcome	20/20	100 (0)	41.25 (46.79)	NR	0.0009*	<i>Favours intervention</i>	Some concerns
			Pain	End of treatment (20 weeks)	SF-36 bodily pain	Higher score means better outcome	20/20	50.50 (5.10)	46.50 (4.89)	NR	0.1672	<i>No difference</i>	Some concerns
			General health	End of treatment (20 weeks)	SF-36 general health perceptions	Higher score means better outcome	20/20	79.25 (6.34)	71 (10.95)	NR	0.1342*	<i>No difference</i>	Some concerns
			Vitality	End of treatment (20 weeks)	SF-36 vitality	Higher score means better outcome	20/20	82.50 (14.28)	60.25 (21.43)	NR	0.0110*	<i>Favours intervention</i>	Some concerns
			Role-social	End of treatment (20 weeks)	SF-36 role social	Higher score means better outcome	20/20	42.50 (13.69)	50 (15.71)	NR	0.3871*	<i>No difference</i>	Some concerns
			Role-emotional	End of treatment (20 weeks)	SF-36 role emotional	Higher score means better outcome	20/20	100 (0)	48.33 (50.12)	NR	0.0046*	<i>Favours intervention</i>	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Mental health	End of treatment (20 weeks)	SF-36 mental health	Higher score means better outcome	20/20	79.80 (19.31)	64.20 (21.54)	NR	0.1414*	No difference	Some concerns
			Footnotes:	*analysis in time-repeated measures (ANOVA), followed by Tukey's multiple comparison test.									
Irez 2011	Sedentary behaviour	Control (no intervention)		No data extracted because the study does not include any priority outcome domains or measures.									
Liposcki 2019	Women > 60 yrs at risk of physical or mental decline	Pilates vs Control (usual activities)	Physical function	End of treatment (26 weeks)	SF-36 physical functioning	Higher score means better outcome	9/11	91.6 (14.3)	62.6 (24.4)	NR	0.00	Favours intervention	High
			Role - Physical	End of treatment (26 weeks)	SF-36 role physical	Higher score means better outcome	9/12	92.7 (14.8)	52.2 (43.9)	NR	0.03	Favours intervention	High
			Pain	End of treatment (26 weeks)	SF-36 bodily pain	Higher score means better outcome	9/13	95.7 (6.9)	52.2 (17.5)	NR	0.00	Favours intervention	High
			General health	End of treatment (26 weeks)	SF-36 general health perceptions	Higher score means better outcome	9/14	89.4 (11.2)	76.7 (16.3)	NR	0.04	Favours intervention	High
			Vitality	End of treatment (26 weeks)	SF-36 vitality	Higher score means better outcome	9/15	85.5 (13.5)	70.0 (14.9)	NR	0.02	Favours intervention	High
			Role-social	End of treatment (26 weeks)	SF-36 role social	Higher score means better outcome	9/16	97.2 (8.3)	77.9 (23.9)	NR	0.03	Favours intervention	High
			Role-emotional	End of treatment (26 weeks)	SF-36 role emotional	Higher score means better outcome	9/17	92.6 (22.0)	57.5 (49.6)	NR	0.06	No difference	High
			Mental health	End of treatment (26 weeks)	SF-36 mental health	Higher score means better outcome	9/18	88.88 (10.5)	75.2 (17.6)	NR	0.05	Favours intervention	High

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Footnotes:										
Pilates vs 'other'													
Aibar-Almazan 2019	Women > 60 yrs at risk of physical or mental decline	Pilates vs non-exercise intervention (education)	Sleep quality	End of treatment (12 weeks)	Pittsburgh sleep quality index	higher score means worse sleep quality	55/52	7.16 (4.9)	8.38 (4.28)	F=0.17	0.677*	No difference	Some concerns
			Anxiety	End of treatment (12 weeks)	HADS (0-21)	<7 means no anxiety	55/52	4.76 (3.73)	9.37 (3.52)	F=11.74	<0.01	Favours intervention	Some concerns
			Depression	End of treatment (12 weeks)	HADS (0-21)	<7 means no depression	55/52	3.98 (2.93)	6.81 (3.6)	F=4.31	>0.04	Favours intervention	Some concerns
			Footnotes: *authors noted an association between BMI and PSQI total score at baseline, as well as an association between age and baseline anxiety on PSQI subscales. When included in an analysis of covariables (ANCOVA), group X time interactions for PSQI total score were statistically significant (p<0.001).										
de Andrade Mesquita 2015	Women > 60 yrs at risk of physical or mental decline	Pilates vs Proprioceptive Neuromuscular Facilitation	Dynamic balance/mobility	baseline, end of treatment (4 weeks)	timed up and go (s)	absolute mean change (SD)	20/20	7.7 (1.5)	8.1 (1.9)	NR	NR	Not reported	Some concerns
			Balance	baseline, end of treatment (4 weeks)	Berg Balance test	absolute mean change (SD)	20/20	56.0 (0.1)	55.8 (0.4)	NR	NR	Not reported	Some concerns
			Footnotes:										
Abbreviations: C, Comparator; HADS, hospital anxiety and depression scale; I, intervention; NR, not reported													

STUDY RESULTS (as reported by the study authors)													
Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
Pilates vs control													
Barker 2016	Healthy adults at risk of falls	Control (usual care)	Falls	Follow up (24 weeks)	Number of falls (count)	higher score means worse outcome	20/24*	13	11	NR	NR	<i>Not reported</i>	High
			Falls	Follow up (24 weeks)	Falls injury (rate per 1000 person days) **	higher score means worse outcome	20/24*	1.23 (0.40 to 2.88)	2.02 (0.81 to 4.16)	0.58 (0.18 to 1.84)	0.347	<i>No difference</i>	High
			Falls	Follow up (24 weeks)	Number of Falls (rate per 1000 person days)**	higher score means worse outcome	20/24*	3.21 (1.71 to 5.48)	3.17 (1.58 to 5.68)	1.17 (0.43 to 3.16)	0.754	<i>No difference</i>	High
			Falls	Follow up (24 weeks)	Injurious fall rates (rate per 1000 person days)**	higher score means worse outcome	20/24*	0.74 (0.15 to 2.16)	2.02 (0.81 to 4.16)	0.36 (0.09 to 1.38)	0.136	<i>No difference</i>	High
			Functional mobility	end of treatment (12 weeks)	Timed up and go (s) (comfortable pace)	higher score mean worse agility	20/29	9.33 (2.09)	9.98 (2.50)	NR	NR	<i>Not reported</i>	High
			Functional mobility	end of treatment (12 weeks)	Timed up and go (s) (fast pace)	higher score mean worse agility	20/29	7.12 (1.36)	7.80 (1.59)	NR	NR	<i>Not reported</i>	High
			Functional mobility	end of treatment (12 weeks)	Timed up and go test (s) (manual task)	higher score mean worse agility	20/29	8.23 (1.51)	9.28 (1.87)	NR	NR	<i>Not reported</i>	High
			Functional mobility	end of treatment (12 weeks)	Timed up and go test (s) (cognitive task)	higher score mean worse agility	20/29	8.01 (1.76)	8.55 (2.65)	NR	NR	<i>Not reported</i>	High
			Footnotes:	*Only followup data reported (24 weeks), with data available for N=24 in the Control group. **Rate of falls and falls injury are reported at rate (confidence interval). No standard deviation or standard error provided.									
Roller 2018	Healthy adults at risk of falls	Control (no intervention)	Functional mobility	end of treatment (10 weeks)	Timed Up and Go (s)	Higher score means worse agility	27/28	10.98 (4.67)	11.54 (2.62)	NR	>0.05	<i>No difference</i>	Some concerns

Study ID	Condition	Comparison	Outcome	Timing	Outcome measure	measure details	# participants (I/C)	(intervention) n/N (%) or mean (SD)	(comparator) n/N (%) or mean (SD)	Point estimate (95% CI)	p-value	direction of effect	RoB
			Balance	end of treatment (10 weeks)	Berg Balance Scale (0-56)	higher score means better balance (static)	27/28	53.22 (2.59)	52.70 (3.11)	NR	>0.05	No difference	Some concerns
			Footnotes:										
			Pilates vs 'other'										
Josephs 2016	Healthy adults at risk of falls	Conventional balance sessions	Functional mobility	end of treatment (12 weeks)	Timed up and go (s)	higher score means worse balance	13/11	13.07 (4.59)	11.26 (2.89)	NR	NR	No difference	High
			Balance	end of treatment (12 weeks)	Fullerton Advanced Balance Scale	higher score means better balance confidence	13/11	24.85 (12.54)	27.27 (6.41)	NR	NR	No difference	High
			Footnotes:										
Abbreviations: C, Comparator; I, intervention; NR, not reported; ROM , range of motion; s, seconds													