Project

Pilates for preventing and treating health conditions: an evidence evaluation

Prepared for

National Health and Medical Research Council

NHMRC | Natural Therapies Working Committee
Canberra ACT 2601

CONFIDENTIAL

Appendices D to H
prepared by
Health Technology Analysts Pty Ltd

August 2022

Contents

Report information **4**

List of tables **5**

List of figures **8**

List of abbreviations **10**

Appendix D Details of included studies **11**

D1 Neoplasms 11

D1.1 Breast cancer 11

D1.2 Prostate cancer 17

D2 Endocrine, nutritional, or metabolic diseases 21

D2.1 Diabetes mellitus 21

D3 Diseases of the nervous system 26

D3.1 Multiple sclerosis 26

D3.2 Myelopathy (HTLV-1 associated) 34

D3.3 Parkinson's disease 37

D3.4 Rehabilitation after stroke 40

D1 Diseases of the circulatory system 45

D3.5 Hypertension 45

D4 Diseases of the musculoskeletal system or connective tissue 48

D4.1 Osteoarthritis 48

D4.2 Post viral arthropathies 51

D4.3 Spondyloarthritis 54

D4.4 Spinal deformities 58

D4.5 Osteoporosis 63

D4.6 Chronic widespread pain (fibromyalgia) 68

D4.7 Low back pain 71

D4.8 Neck and shoulder pain 84

D1 Disease of the genitourinary system 89

D1.1 Menopausal symptoms or complaint 89

D5 Pregnancy, childbirth, or the puerperium 95

D5.1 Postpartum recovery 95

D6 Injury, poisoning, or other certain other consequences of external causes 98

D6.1 Rehabilitation of the knee after injury 98

D7 External causes of morbidity and mortality 101

D7.1 Rehabilitation of the knee after arthroplasty 101

D8 Prevention of disease, injury, or illness in at risk populations 104

D8.1 Prevention of mental health conditions 104
D8.2 Prevention of metabolic disorders or weight problems associated with sedentary behaviour 107
D8.3 Prevention of age-related physical and mental decline 111
D8.4 Falls prevention 120

Appendix E Risk of bias forms 124

E1 Randomised controlled trials 124
E2 Non-randomised studies of interventions 124

Appendix F Characteristics of included studies 125

F1 Study details 125
F2 Supplementary outcome data 125

Appendix G Differences between protocol & review 126

G1 Methods not implemented 126
G2 Changes from protocol 127

References 130

Report information

Authors

Jorgensen MA[1], Olaya Y[1], Maurin R[1], Iconomou M[1], Miles A[1]

[1] **HT**Analysts, Level 8, 46 Kippax Street, Surry Hills NSW 2010 Australia

Dates

This technical report and accompanying evidence evaluation report received approval from the National Health and Medical Research Council (NHMRC) Natural Therapies Working Committee (NTWC) on 11 August 2021.

The protocol for the evidence evaluation received approval from the NHMRC NTWC on 25 May 2020 and is published on PROSPERO (CRD42020191918).

History

NHMRC has been engaged by the Department of Health and Aged Care (Department) to update the evidence underpinning the *2015 Review of the Australian Government Rebate on Natural Therapies for Private Health Insurance* (2015 Review) (1). The seven natural therapies to be reviewed in the first tranche are naturopathy, Pilates, Rolfing, shiatsu, Tai Chi, Western herbalism and yoga. These therapies are among those excluded from the private health insurance rebate as of 1 April 2019.

To support NHMRC in their evidence review, Health Technology Analysts (**HT**Analysts) has been engaged to conduct a systematic review of the evidence of clinical effectiveness of Pilates. Eligible studies received from the Department's public call for evidence, the Department Natural Therapies Review Expert Advisory Panel (NTREAP) and NTWC will also be included in the evidence evaluation.

This technical report has been developed by **HT**Analysts in conjunction with NHMRC, NTWC, and NTREAP. It provides the appendices and supplementary data related to an evidence valuation of the effect of Pilates for preventing and treating health conditions. The main body of evidence is presented in the evidence evaluation report. All associated materials have been developed in a robust and transparent manner in accordance with relevant best practice standards (2-5).

List of tables

Table D-1 Overview of PICO criteria of included studies: Breast cancer 11

Table D-2 Outcomes considered by the NTWC to be critical or important for decision-making: breast cancer (survivors) 14

Table D-3 Outcomes considered by the NTWC to be critical or important for decision-making: breast cancer (on treatment) 16

Table D-4 Overview of PICO criteria of included studies: Prostate cancer 17

Table D-5 Outcomes considered by the NTWC to be critical or important for decision-making: prostate cancer 19

Table D-6 Overview of PICO criteria of included studies: Diabetes mellitus (type 2) 21

Table D-7 Outcomes considered by the NTWC to be critical or important for decision-making: Diabetes, type 2 23

Table D-8 Overview of PICO criteria of included studies: Multiple sclerosis 26

Table D-9 Outcomes considered by the NTWC to be critical or important for decision-making: Multiple sclerosis 30

Table D-10 Overview of PICO criteria of included studies: Myelopathy (HTLV-1 associated) 34

Table D-11 Outcomes considered by the NTWC to be critical or important for decision-making: Myelopathy (HTLV-1 associated) 35

Table D-12 Overview of PICO criteria of included studies: Parkinson's disease 37

Table D-13 Outcomes considered by the NTWC to be critical or important for decision-making: Parkinson’s disease 38

Table D-14 Overview of PICO criteria of included studies: Rehabilitation after stroke 40

Table D-15 Outcomes considered by the NTWC to be critical or important for decision-making: Rehabilitation after stroke 43

Table D-16 Overview of PICO criteria of included studies: Hypertension 45

Table D-17 Outcomes considered by the NTWC to be critical or important for decision-making: Hypertension 46

Table D-18 Overview of PICO criteria of included studies: Osteoarthritis 48

Table D-19 Outcomes considered by the NTWC to be critical or important for decision-making: Osteoarthritis 49

Table D-20 Overview of PICO criteria of included studies: Post viral arthropathies 51

Table D-21 Outcomes considered by the NTWC to be critical or important for decision-making: Post viral arthropathies 52

Table D-22 Overview of PICO criteria of included studies: Spondyloarthropathies 54

Table D-23 Outcomes considered by the NTWC to be critical or important for decision-making: Spondyloarthropathies 55

Table D-24 Overview of PICO criteria of included studies: Spinal deformities 58

Table D-25 Outcomes considered by the NTWC to be critical or important for decision-making: Spinal Deformities-Scoliosis..... 60

Table D-26 Outcomes considered by the NTWC to be critical or important for decision-making: Spinal Deformities (hyperkyphosis, hyperlordosis, forward head posture)..... 61

Table D-27 Overview of PICO criteria of included studies: Osteoporosis 63

Table D-28 Outcomes considered by the NTWC to be critical or important for decision-making: Osteoporosis 65

Table D-29 Overview of PICO criteria of included studies: Fibromyalgia 68

Table D-30 Outcomes considered by the NTWC to be critical or important for decision-making: Chronic widespread pain (fibromyalgia) 70

Table D-31 Overview of PICO criteria of included studies: Low back pain 71

Table D-32 Outcomes considered by the NTWC to be critical or important for decision-making: Low back pain 78

Table D-33 Overview of PICO criteria of included studies: Neck and shoulder pain 84

Table D-34 Outcomes considered by the NTWC to be critical or important for decision-making: neck pain 86

Table D-35 Outcomes considered by the NTWC to be critical or important for decision-making: shoulder pain 86

Table D-36 Overview of PICO criteria of included studies: Menopausal symptom or complaint..... 89

Table D-37 Outcomes considered by the NTWC to be critical or important for decision-making: Menopausal symptom or complaint..... 91

Table D-38 Overview of PICO criteria of included studies: Postpartum recovery 95

Table D-39 Outcomes considered by the NTWC to be critical or important for decision-making: Postpartum recovery 97

Table D-40 Overview of PICO criteria of included studies: Rehabilitation of the knee after injury 98

Table D-41 Outcomes considered by the NTWC to be critical or important for decision-making: Rehabilitation of the knee after injury..... 99

Table D-42 Overview of PICO criteria of included studies: Rehabilitation of the knee after arthroplasty 101

Table D-43 Outcomes considered by the NTWC to be critical or important for decision-making: Rehabilitation of the knee after arthroplasty 102

Table D-44 Overview of PICO criteria of included studies: Prevention of mental health conditions 104

Table D-45 Outcomes considered by the NTWC to be critical or important for decision-making: Prevention of mental health conditions..... 105

Table D-46 Overview of PICO criteria of included studies: Prevention of metabolic disorders or weight problems associated with sedentary behaviour..... 107

Table D-47 Outcomes considered by the NTWC to be critical or important for decision-making: Prevention of metabolic disorders or weight problems associated with sedentary lifestyles 109

Table D-48 Overview of PICO criteria of included studies: Prevention of age-related physical and mental decline..... 111

Table D-49 Outcomes considered by the NTWC to be critical or important for decision-making: Prevention of age-related physical decline 114

Table D-50 Outcomes considered by the NTWC to be critical or important for decision-making: Prevention of age-related mental decline..... 117

Table D-51 Overview of PICO criteria of included studies: Falls prevention 120

Table D-52 Outcomes considered by the NTWC to be critical or important for decision-making: Falls prevention..... 122

List of figures

Figure D.1 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Breast cancer 13

Figure D.2 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Prostate cancer 18

Figure D.3 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Diabetes, type 2 22

Figure D.4 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Multiple sclerosis 28

Figure D.5 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSI – Multiple sclerosis 30

Figure D.6 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Myelopathy (HTLV-1 associated) 34

Figure D.7 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Parkinson's disease 38

Figure D.8 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Rehabilitation after stroke 41

Figure D.9 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSI – Rehabilitation after stroke 42

Figure D.10 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSIs – Hypertension 46

Figure D.11 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Osteoarthritis 49

Figure D.12 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Post viral arthropathies 51

Figure D.13 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Spondyloarthritis 54

Figure D.14 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Spinal deformities 59

Figure D.15 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Osteoporosis 64

Figure D.16 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Chronic widespread pain (fibromyalgia) 69

Figure D.17 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Low back pain 75

Figure D.18 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSIs – Low back pain 76

Figure D.19 Funnel plot of comparison: Low back pain, outcome: Pain (end of treatment)..... 80

Figure D.20 Funnel plot of comparison: Low back pain, outcome: Disability (end of treatment)..... 82

Figure D.21 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Neck and shoulder pain 85

Figure D.22 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Menopausal symptom or complaint..... 89

Figure D.23 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Postpartum recovery 95

Figure D.24 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Rehabilitation of the knee after injury..... 98

Figure D.25 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Rehabilitation of the knee after arthroplasty 101

Figure D.26 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Prevention of mental health conditions 104

Figure D.27 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Prevention of metabolic conditions or weight problems in sedentary adults..... 107

Figure D.28 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Prevention of age-related decline 112

Figure D.29 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSIs – Prevention of age-related decline 113

Figure D.30 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Falls prevention..... 120

List of abbreviations

BRISA	Regional Base of Health Technology Assessment Reports of the Americas
CINAHL	Cumulative Index to Nursing and Allied Health Literature
COMET	Core Outcome Measures in Effectiveness Trials
GRADE	Grading of Recommendations Assessment, Development and Evaluation
ITT	Intention-to-treat
NHMRC	National Health and Medical Research Council
NRSI	Non-randomised study of an intervention
NTREAP	Natural Therapies Review Expert Advisory Panel
NTWC	Natural Therapies Working Committee
OR	Odds ratios
PAHO	Pan American Health Organization
PICO	Population, Intervention, Comparator, Outcome
PP	Per protocol
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomised controlled trial
RoB	Risk of bias
RR	Risk ratios
SR	Systematic review
SD	Standard deviation
TIDIER	Template for Intervention Description and Replication

Appendix D Details of included studies

This appendix documents the studies that met the prespecified inclusion criteria for a systematic review on the effect of Pilates for preventing and treating any health condition and were prioritised at the population prioritisation phase. It provides an overview of the PICO criteria of these studies, a summary of the risk of bias assessment, and results of the data synthesis for the main comparison.

Additional details concerning the risk of bias judgements for each study are provided in Appendix E1 (RCTs) or Appendix E2 (NRSIs) and characteristics of the included studies are provided in Appendix F1. Outcome date for outcomes considered to be critical or important for this review are provided in Appendix F2.

D1 Neoplasms

D1.1 Breast cancer

D1.1.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-1. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-1 Overview of PICO criteria of included studies: Breast cancer

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Eyigor 2010 (6)	RCT	Breast cancer (survivors)	Pilates exercises (home)	Control (no intervention)	Home exercise	Functional capacity Flexibility Fatigue Depression Quality of life
Martin 2013 (7, 8)	Quasi-RCT	Breast cancer (survivors)	Pilates exercises (chair)	Control (no intervention) OR resistance training^	None	Feasibility Muscular endurance Exercise intensity (RPE)
Pilates versus 'other' intervention*						
Alpozgen 2017 (9)	RCT	Breast cancer (survivors)	Pilates exercises (mat, TheraBand)	Combined exercise (stretching, band, breathing) OR home exercise (individual)	None	Pain Functional status, upper extremity Shoulder function, overall Muscular endurance
Gajbhiye 2013 (10)	RCT	Breast cancer (on treatment)	Pilates exercises (mat)	Control (usual care) plus counselling	None	Functional status, upper extremity Quality of life
Odynets 2018 (11-14)	RCT	Breast cancer (survivors)	Pilates exercises	Yoga	None	Functional status, upper extremity Haemodynamic parameters Pulmonary function Quality of life

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Odynets 2019 (15-17)	RCT	Breast cancer (survivors)	Pilates exercises	Yoga OR water physical rehabilitation programme^	None	Quality of life Anxiety Cardiorespiratory fitness Depression
Sener 2017 (18, 19)	Quasi-RCT	Breast cancer (survivors with lymphedema)	Pilates exercises	Lumbopelvic stability exercises	Home exercise	Body image after cancer Functional status, upper extremity Lymphedema Muscular endurance Pain Quality of life

Abbreviations: RCT, randomised controlled trial; RPE, rate of perceived exertion

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

^ Study included three groups. The inactive control is considered in the evidence synthesis.

D1.1.2 Risk of bias summary

The risk of bias for each item in the included studies for breast cancer is described below and shown graphically in Figure D.1 (details are provided in Appendix E1).

Bias arising from the randomisation process

Two studies (Odynets 2018, Odynets 2019) provided sufficient information on the randomisation process and was at low risk of bias. Both studies randomised participants into groups with the use of sequentially numbered opaque sealed envelopes to ensure concealed randomisation. Odynets 2019 also used an independent person to generate random numbers.

Five studies (Alpozgen 2017, Eyigor 2010, Martin 2013, Gajbhiye 2013, Sener 2017) randomised patients using a simple random number generator or random lottery, but did not provide any information on allocation concealment, causing some concerns. Reported baseline characteristics or baseline outcomes measures appeared matched between treatment groups.

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

Three studies (Alpozgen 2017, Odynets 2018, Odynets 2019) were judged to be at low risk of bias for this domain; any discontinuations from intended interventions were judged to be unrelated to the trial context. Concerns were raised with two studies (Gajbhiye 2013, Martin 2013) due to a lack of information regarding any deviations from intended interventions. Two studies (Eyigor 2010, Sener 2017) were judged to be at high risk of bias, as there were deviations from the trial protocol that appeared unbalanced between treatment groups.

Bias due to missing outcome data

Three studies (Alpozgen 2017, Odynets 2018, Sener 2017) were judged to be at low risk of bias for this domain as outcome data appeared to be available for all (or nearly all) participants. One study (Odynets 2019) had some concerns raised as missingness of the data differ slightly across intervention groups. Two studies (Gajbhiye 2013, Martin 2013) were judged to be at high risk of bias for this domain due to a lack of information regarding missing outcome data and one study (Eyigor 2010) was also at high risk as missing data were unbalanced between treatment groups.

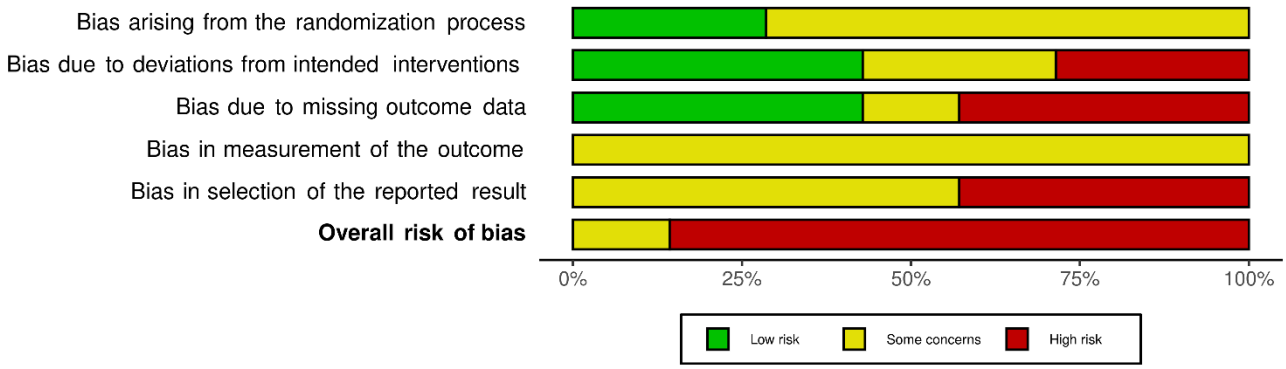
Bias in measurement of the outcome

All included studies were assessed to have some concerns regarding the measurement of outcomes. None of the included studies blinded the participant or outcome assessors and many of the primary or key outcomes were subjective, results of which could be influenced by knowledge of the intervention.

Bias in selection of the reported result

Four studies reported all eligible specified results and, in the absence of an available protocol were judged to be at some concerns for this domain (Alpozgen 2017, Eyigor 2010, Martin 2013, Sener 2017). Three studies 9 Gajbhiye 2013, Odynets 2018, Odynets 2019) were judged to be at high risk of bias because if incomplete reporting, suggesting selective reporting of results. In two studies (Odynets 2018, Odynets 2019), it is possible that the overall direction of bias is against Pilates, as the studies were examining the effectiveness of another intervention, using Pilates as the comparator group.

Figure D.1 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Breast cancer



D1.1.3 Effect of intervention (survivors)

Outcomes considered by the NTWC to be critical or important for decision-making in breast cancer survivors are listed in Table D-2.

Table D-2 Outcomes considered by the NTWC to be critical or important for decision-making: breast cancer (survivors)

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Eyigor 2010*	*Martin 2013*
Quality of life, global	EORTC QLQ-C30 / EORTC QLQ-BR23	Critical	Yes	✓	--
Functional status, upper extremity	DASH	Critical	No	--	--
Pain	Visual analogue scale (VAS)	Critical	No	--	--
Fatigue	FACT – Fatigue or Brief Fatigue Inventory	Critical	Yes	✓	--
Quality of life, functional	EORTC QLQ-C30 – Functional score	Important	Yes	✓	--
Lymphedema	Sum of arm circumference	Important	No	--	--
Physical activity	IPAC- short	Important	No	--	--

Abbreviations: BR23, breast cancer 23-items; C30, general cancer 30-items; DASH, Disabilities of the Arm, Shoulder and Hand scale; EORTC-QLQ, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; FACT, Functional Assessment of Cancer Therapy; IPAC, International Physical Activity Questionnaire

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Two RCTs (Eyigor 2010, Martin 2013) comparing Pilates with no intervention in breast cancer survivors were eligible for this comparison. One RCT (Eyigor 2010) contributed data relevant to three of the seven outcomes. The other RCT (Martin 2013) was a feasibility study and did not measure or assess any outcomes considered critical or important to this review.

There was one study published in a language other than English that compared Pilates with no intervention in breast cancer survivors (total 27 participants) that could have contributed data, but it did not measure or assess any outcomes considered to be critical important for this review (see Appendix C6). There were no ongoing studies eligible for this comparison.

Results for all outcomes were judged to be at high risk of bias as the study that contributed data (Eyigor 2010) had important deviations from the trial protocol that were unbalanced between treatment groups (missing data from 10 control participants). The available evidence comes from one small trial, with no suspicion of non-reporting of results from identified studies (including those awaiting classification or ongoing).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data for all outcomes.

Quality of life (QoL)

One trial (54 participants) reported quality of life measured with EORTC QLQC30 at the end of treatment (8 weeks) (Eyigor 2010). The EORTC QLQC30-global health status is designed to measure cancer patients' physical, psychological and social functions summarised on a scale from 0 (worse) to 100 (best). In people with breast cancer the minimal change in scores to be considered clinically important (MCID) for global health status is 22.4 points (20).

The results showed no change in global quality of life between the Pilates group compared to no intervention (MD –13.24; 95% CI –27.83, 1.35; *p* = 0.08) *(GRADE: Very low)*.

Fatigue

One trial (54 participants) reported fatigue measured with the Brief Fatigue Inventory at the end of treatment (8 weeks) (Eyigor 2010). The Brief Fatigue Inventory is designed to assess the severity and impact of cancer-related fatigue and is summarised on a scale from 0 (no fatigue) to 10 (as bad as you can imagine). In people with cancer, cut points for fatigue level suggested are 1–3 (mild), 4–7 (moderate), and 8–10 (severe), which corelate with functional interference, symptoms, depression, and QOL (21).

The results showed no change in fatigue between the Pilates group compared to no intervention (MD –0.97; 95% CI –3.87, 1.88; *p* = 0.50) *(GRADE: Very low)*. Participants in both groups continue to have 'moderate' fatigue.

Functioning

One trial (54 participants) reported EORTC QLQC30-functional score at the end of treatment (8 weeks) (Eyigor 2010). The EORTC QLQC30-function measures physical, emotional, role, social & cognitive functioning and is summarised on a scale from 0 (worse) to 100 (best). In people with breast cancer the minimal change in scores to be considered clinically important is between 17 and 19.6 points (20, 22).

The results showed no difference in functional score between the Pilates group compared to no intervention (MD –5.26; 95% CI –17.04, 6.52; *p* = 0.38). *(GRADE: Very low)*.

Comparison 2 (vs other)

Four RCTs (Alpozgen 2017, Odynets 2018, Odynets 2019, Sener 2017) comparing Pilates with 'other' interventions in breast cancer survivors were eligible for this comparison and contributed data for six of the seven outcomes. The other RCT (Martin 2013) was a feasibility study and did not measure or assess any outcomes considered critical or important to this review.

Data from these studies are presented in Appendix F2 Supplementary outcome data.

D1.1.4 Effect of intervention (on treatment)

Outcomes considered by the NTWC to be critical or important for decision-making in people with breast cancer who were undergoing radiation or chemotherapy are listed in Table D-3.

Main comparison (vs control)

There were no studies identified comparing Pilates with no intervention in people with breast cancer who were undergoing radiation or chemotherapy.

Table D-3 Outcomes considered by the NTWC to be critical or important for decision-making: breast cancer (on treatment)

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*No studies found*
Quality of life	EORTC QLQ-C30 / EORTC QLQ-BR23	Critical	No	--
Pain	Visual analogue scale	Critical	No	--
Fatigue	Functional Assessment of Cancer Therapy-Fatigue (FACT-F)	Critical	No	--
Functional status, upper extremity	Disabilities of the Arm, Shoulder and Hand scale (DASH)	Critical	No	--
Physical activity	International physical activity questionnaire - short version	Critical	No	--
Anxiety	Hospital Anxiety and Depression Scale - Anxiety	Important	No	--
Emotional wellbeing	Any available	Important	No	--

Abbreviations: BR23, breast cancer 23-items; C30, general cancer 30-items; EORTC-QLQ, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

D1.2 Prostate cancer

D1.2.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-4. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-4 Overview of PICO criteria of included studies: Prostate cancer

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Gomes 2018 (23)	RCT	Prostate cancer (after radical prostatectomy)	Pilates exercises (mat)	Control (no intervention) OR PFM exercise with electrical stimulation^	None	PFM strength Urinary incontinence
Pedriali 2014 (24, 25)	RCT	Prostate cancer (after radical prostatectomy)	Pilates exercises (mat)	Control (no intervention) OR PFM exercise with electrical stimulation^	None	Urinary incontinence
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: PFM, pelvic floor muscle; RCT, randomised control trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

^ Study included three groups. The inactive control is considered in the evidence synthesis.

D1.2.2 Risk of bias summary

The risk of bias for each item in the included studies for prostate cancer is described below and shown graphically in Figure D.2 (details are provided in Appendix E1).

Bias arising from the randomisation process

Gomes 2018 and Pedriali 2014 used sealed envelopes to randomise patients into treatment groups and baseline characteristics appeared balanced between the groups. The studies were therefore judged to be at low risk of bias for this domain

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

Concerns were raised with both studies (Gomes 2018, Pedriali 2014) due possible deviations from the trial protocol. A small number of participants did not complete assigned intervention, some of which were judged to be related to the trial context.

Bias due to missing outcome data

Although outcome data was available for all (or nearly all) participants in both studies (Gomes 2018, Pedriali 2014) there were concerns the missing data could be related to the outcomes, given that participants who were rated continent were excluded from the analysis.

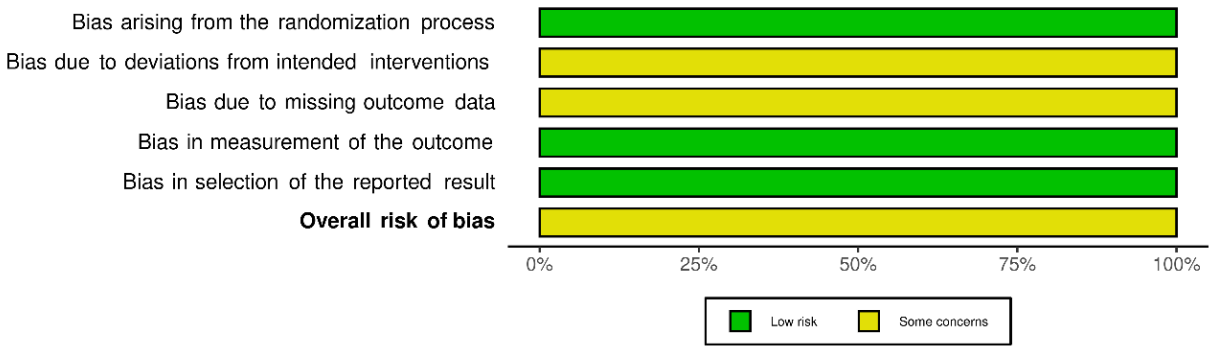
Bias in measurement of the outcome

Both studies (Gomes 2018, Pedriali 2014) were judged to be at low risk of bias for this domain. In both studies the outcome assessors were blinded, and key outcomes were objective and measured using appropriate outcome measurement tools.

Bias in selection of the reported result

Both studies (Gomes 2018, Pedriali 2014) were registered on trial registries, reported all eligible prespecified results and were judged to be at low risk of bias for this domain.

Figure D.2 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Prostate cancer



D1.2.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with prostate cancer are listed in Table D-5.

Table D-5 Outcomes considered by the NTWC to be critical or important for decision-making: prostate cancer

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Gomes 2018*	*Pedriali 2014*
Quality of life, global	No measures reported in eligible studies	Critical	No	--	--
Physical functioning	No measures reported in eligible studies	Critical	No	--	--
Fatigue	No measures reported in eligible studies	Critical	No	--	--
Pain	No measures reported in eligible studies	Critical	No	--	--
Quality of life, disease specific	ICIQ - short Form	Important	Yes	✓	✓
Urinary incontinence	EPIC-26 Urinary incontinence domain / Reduction in daily incontinence (24-hr pad test)	Important	Yes	✓	✓
Urinary frequency/urgency/ irritation	EPIC-26 urinary irritative/obstructive domain	Important	No	--	--
Psychological wellbeing	No measures reported in eligible studies	Important	No	--	--
Sexual function / symptoms	EPIC-26 sexual domain	Important	No	--	--

Abbreviations: EPIC-26, Expanded Prostate Cancer Index Composite; ICIQ, International Consultation on Incontinence Questionnaire

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Two RCTs (Gomes 2018, Pedriali 2014) comparing Pilates with no intervention in participants with post-prostatectomy urinary incontinence were eligible for this comparison and contributed data to two outcomes. There were no addditional studies identified (awaiting classification or ongoing) that compared Pilates with no intervention in men with post-prostatectomy urinary incontinence that could have contributed data to these outcomes (see Appendix C6).

Results for reported outcomes had some concerns of bias (possibly against Pilates) related to deviations attributed to the trial context resulting in missing outcome data that may affect the results.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as none of the studies were judged to be at high risk of bias.

Quality of life – urinary incontinence

Two trials (126 participants) reported urinary incontinence-related quality of life measured with the International Consultation on Incontinence Questionnaire (ICIQ) Short Form questionnaire at the end of treatment (10 weeks) (Gomes 2018, Pedriali 2014). The ICIQ-short form assesses the frequency, severity, and impact on quality of life of urinary incontinence and is summarised as a total score ranging from 0 (no symptoms) to 21 (all the time). The MCID in men for the ICIQ-UI-SF is not known.

The results showed an improvement in QoL scores in the Pilates group compared to no intervention (MD –3.66; 95% CI –5.26, –2.06; $p < 0.0001$) *(GRADE: moderate)*. In the absence of an MCID, a change score of 3.66 (out of a maximum 21) indicated that the magnitude of the effect is moderate (17.4% is between 10% to 20% of the scale).

Urinary incontinence – 24-hr pad test

Two trials (126 participants) reported urinary incontinence measured by the 24-hr pad test (pad weight, grams) at the end-of-treatment (10 weeks) (Gomes 2018, Pedriali 2014). The 24-hr pad test is a gold standard measure of urinary incontinence, with higher urinary loss in the 24-hour period indicating worse incontinence. The degree of incontinence can be classified as mild if urinary loss is less than 100 grams in 24 hours, moderate if urinary loss is between 100 and 400 grams in 24 hours, and high if urinary loss is more than 400 grams in 24 hours (26).

Pooled results suggest an effect favouring control (MD 17.29; 95% CI 6.69, 27.90; $p = 0.001$) however, there was a large variability in pad weight at baseline between treatment groups within and across studies, therefore standardised scores are presented. When standardised, the pooled results showed no effect on urinary incontinence comparing Pilates with no intervention (SMD 0.45; 95% CI –0.28, 1.18, 27.89; $p = 0.23$) *(GRADE: very low)*.

Comparison 2 (vs other)

Two RCTs (Gomes 2018, Pedriali 2014) comparing Pilates with 'other' interventions in people with post-prostatectomy urinary incontinence were eligible for this comparison and contributed data to two outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

D2 Endocrine, nutritional, or metabolic diseases

D2.1 Diabetes mellitus

D2.1.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-6. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-6 Overview of PICO criteria of included studies: Diabetes mellitus (type 2)

Study ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Melo 2020 (27)	QuasiRCT	Diabetes, type 2 (women)	Pilates exercises	Control (no intervention)	Standard medical and dietary care	Glycaemic control Functional capacity
Torabian 2013 (28)	RCT	Diabetes, type 2 (women)	Pilates exercises	Control (no intervention)	Standard medical care	Physical symptoms Anxiety Social dysfunction Depression General health
Yucel 2016 (29)	RCT	Diabetes, type 2 (women)	Pilates exercises (mat)	Control (no intervention)	Standard medical and dietary care	Anxiety Depression Fatigue Pain Physical health Glycaemic control
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: NRSI, non-randomised study of interventions; RCT, randomised control trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

^ Study included three groups. The inactive control is considered in the evidence synthesis.

D2.1.2 Risk of bias summary

The risk of bias for each item in the included studies for diabetes (type 2) is described below and shown graphically in Figure D.3 (details are provided in Appendix E1).

Bias arising from the randomisation process

Concerns of bias were raised in all three studies for this domain. Details relating to method of randomisation or allocation concealment were not provided in Melo 2020; however, baseline characteristics appeared to be matched between treatment groups. Torabian 2013 and Yucel 2016 used simple random sampling methods to all allocate participants but did not provide information about allocation concealment. In Torabian 2013, baseline demographics between the Pilates and control groups were similar but reporting of baseline characteristics was limited in Yucel 2016, raising some concerns.

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

Two studies (Melo 2020, Torabian 2013) had concerns of bias raised for this domain related to missing information. In Melo 2020, two subjects (one from each intervention) were excluded from the analysis

because of the health condition, but no further information was provided. Torabian 2013 did not provide any information regarding deviations from the assigned intervention (no CONSORT provided). Yucel 2016 was at high risk of bias for this domain. A moderate level (~20%) of dropouts and deviations from the intended interventions occurred, with four participants in the control group excluded for reasons considered inconsistent with the trial protocol.

Bias due to missing outcome data

Melo 2020 and Torabian 2013 were assessed to be at low risk of bias for this domain as data was available for all, or nearly all participants. In Yucel 2016, missingness of the outcome data was judged to depend on its true value, with missing data likely to be due to the health status of the patients (assessed as high risk of bias).

Bias in measurement of the outcome

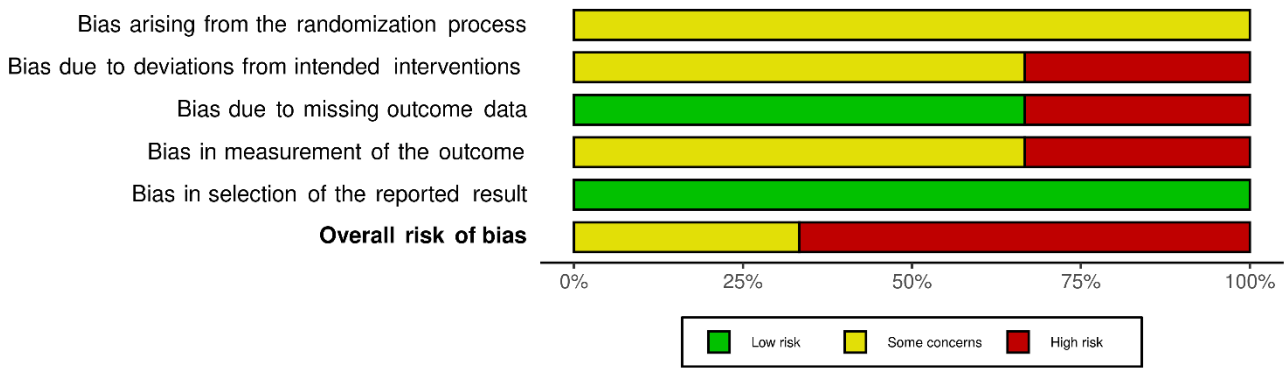
Two studies (Melo 2020, Yucel 2016) were judged to have some concerns for this domain. Outcome measures were patient-reported; however, there was no reason to believe that the subjective outcomes were substantially influenced by knowledge of the intervention received. Measure of glycaemic control (glycated haemoglobin, blood glucose levels) were assessed to be at low risk of bias.

Torabian 2013 was assessed to be at high risk of bias for this domain as the outcome measure used was judged to be inappropriate. The General Health Questionnaire (GHQ-28) is intended for use as a screening instrument (as a measure psychological distress) and is not designed to measure change over time.

Bias in selection of the reported result

All three studies (Melo 2020, Torabian 2013, Yucel 2016) were judged to have low risk for this domain. Specifically, all studies reported pre-specified intentions in sufficient detail and all eligible reported results were available.

Figure D.3 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Diabetes, type 2



D2.1.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with type 2 diabetes are listed in Table D-7.

Table D-7 Outcomes considered by the NTWC to be critical or important for decision-making: Diabetes, type 2

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Melo 2020*	*Torabian 2013*	*Yucel 2016*
Quality of life, global	SF-36-PCS SF-36-MCS	Critical	Yes	--	--	✓
Activities of daily living	GLDAM-composite	Critical	Yes	✓	--	--
Physical functioning	No measures reported in eligible studies	Critical	No	--	--	--
Cardiovascular disease risk	No measures reported in eligible studies	Critical	No	--	--	--
Body composition	Body mass index	Important	Yes	X[1]	X[1]	✓
Fatigue	0-10 visual analogue scale	Important	Yes	--	--	✓
Depression	Hospital Anxiety and Depression Scale - Depression	Important	Yes	--	✓	✓

Abbreviations: 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); MCS, mental component score; PCS, physical component score; SF-36; 36-item short form survey

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Three RCTs (Melo 2020, Torabian 2013, Yucel 2016) comparing Pilates with no intervention in people with type 2 diabetes were eligible for this comparison and contributed data to five of seven outcomes.

There was one additional study published in a language other than English (awaiting classification) that compared Pilates with no intervention in people with type 2 diabetes that could have contributed data to these outcomes but there was no information to make a judgment regarding the extent of missing data (see Appendix C6).

Quality of life

One trial (45 participants) reported quality of life measured with the SF-36 (36-item Short Form Survey) at the end-of-treatment (12 weeks) (Yucel 2016). The other two eligible RCTs did not report QoL, probably because the outcome was not assessed in the studies. Quality of life domains were summarised into two composite scores (physical and mental health) summarised on a scale from 0 (worse) to 100 (best). No MCID has been established in diabetes populations (31) although benchmarks suggest a 1-point change is associated with excess mortality and inability to work (32).

The results show no difference between the Pilates and control groups for mental health (MD 0.00; 95% CI –0.59, 0.59.; *p* = 1.00) *(GRADE: low)* or physical health (MD 0.00; 95% CI –2.34, 2.34; *p* = 1.00) *(GRADE:*

[1] One of the most important aspects of diabetes management is to maintain a healthy weight, with loss of 5% to 10% of body weight likely to reduce the risk of diabetes complications 30. Diabetes Australia. Maintaining a healthy weight [Available from: https://www.diabetesaustralia.com.au/food-activity/maintaining-a-healthy-weight/.. It seems unlikely that the identified studies did not measure BMI or similar.

very low). Noting that data reported by the authors indicated that the presented values are mean (SD) (table 1, (29)), but in the text they state that presented data are median (IQR) (assumed in error).

The study was judged to be at high risk of bias for this outcome related to exclusion of some participants and missing data, but no sensitivity analysis could be conducted as only one study contributed data.

Activities of daily living

One trial (24 participants) reported activities of daily living measured with the Group of Latin American Development to Maturity (GLADM) test battery at the end-of-treatment (12 weeks) (Melo 2020). The other two eligible RCTs did not report activities of daily living, probably because the outcome was not assessed in the studies. The GLADM test battery incorporates a 10 metre walk test, rise from sitting, raise-stand, rise from chair and around, dress- and take- off, which are converted into a summary index score. A higher score suggests worse autonomy. In people aged over 60 years, a score less than 22.28 suggests very good autonomy, whereas a score more than 33.01 indicates insufficient autonomy (33).

The results showed an effect favouring Pilates compared with the control group (MD –8.10, 95% CI –11.55, –4.65; $p < 0.00001$) *(GRADE: low)*. Participants in the control group had a mean score of 35.3 (insufficient autonomy) whereas those in the Pilates group showed a slight improvement, with a score of 27.2 (good autonomy).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as no studies were at high risk of bias for this outcome.

Body composition

One trial (45 participants) reported body composition measured by body mass index (BMI) at the end-of-treatment (12 weeks) (Yucel 2016). It was considered probable that the other two eligible RCTs measured this outcome (or similar), but did not report because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators. A BMI under 18.5 is considered underweight; from 18.5 to 24.9, normal; from 25 to 29.9, overweight; 30 to 39.9, obese; and 40 or higher, extremely obese.

The results show no difference for the outcome of BMI in the Pilates group compared to no intervention (MD 1.67; 95% CI –2.81, 6.15; $p = 0.46$) *(GRADE: very low)*. Participants in both groups remain obese, with the mean BMI being greater than 30 in both groups.

The study was judged to be at high risk of bias for this outcome related to exclusion of some participants and missing data, but no sensitivity analysis could be conducted as only one study contributed data.

Fatigue

One trial (45 participants) reported fatigue measured using a visual analogue scale (VAS) ranging from 0 (no fatigue) to 10 (severe fatigue) at the end-of-treatment (12 weeks) (Yucel 2016). The other two eligible RCTs did not report fatigue, probably because the outcome was not assessed in the studies. An MCID for the fatigue VAS-10 in people with type 2 diabetes has not been established. The MID is reported to range between 0.8 to 1.1 for improvement and 1.1 to 1.3 for worsening in people with rheumatoid arthritis, systemic lupus erythematosus, and cancer (34).

The results show no difference between the Pilates group compared to no intervention for the outcome of fatigue (MD 0.00; 95% CI –0.94, 0.94; $p = 1.00$) *(GRADE: low)*.

The study was judged to be at high risk of bias for this outcome related to exclusion of some participants and missing data, but no sensitivity analysis could be conducted as only one study contributed data.

Depression

One trial (45 participants) reported depression measured with the Hospital Anxiety and Depression Scale (HADS) at the end-of-treatment (12 weeks) (Yucel 2016). One trial (70 participants) reported depression measured with the GHQ-28 subscale at the end-of-treatment (8 weeks) (Torabian 2013). The other eligible RCT did not report depression, probably because the outcome was not assessed in the study.

The 7-item HADS-depression scale measure symptoms of depression and is summarised on a scale from 0 (no depression) to 21 (severe). A score less than 7 indicates no depression, score between 8-10 are considered borderline, and score greater than 11 indicate the presence of increased symptoms of depression (35).

The 7-item GHQ-28 depression subscale identifies the presence of symptoms compared to what is normal for the individual and is summarised on a scale from 0 (better than usual) to 21 (much worse than usual). The GHQ-28 is a screening instrument and is not designed to measure change over time and subscales are not intended to be considered independent of each other. An MCID is therefore not established.

Pooled results show no difference between the Pilates and control groups for the outcome of depression (SMD –0.96; 95% CI –2.84, 0.92; p = 1.00) *(GRADE: very low)*.

Results for this outcome was judged to be at high risk of bias related to exclusion of some participants and missing outcome data, but no sensitivity analysis could be conducted as both studies were at high risk of bias for this outcome.

Comparison 2 (vs other)

There were no studies identified comparing Pilates with 'other' in people with type 2 diabetes.

D3 Diseases of the nervous system

D3.1 Multiple sclerosis

D3.1.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-8. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-8 Overview of PICO criteria of included studies: Multiple sclerosis

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR/s	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Duff 2018 (36)	RCT	Multiple sclerosis (ambulant, PDDS score 0.0 to 6.0)	Pilates exercises (mat + CoreAlign apparatus)	Control (no intervention)	Massage	Functional ability Balance Core stability Physical performance QoL Body composition
Eftekhari 2018 (37, 38)	RCT	Multiple sclerosis (women, EDSS 2.0 to 4.0)	Pilates exercises (mat)	Control (no intervention)	None specified	Functional mobility Fatigue Balance Body composition Serum markers
Fleming 2019 (39)	RCT	Multiple sclerosis (women)	Pilates exercises (mat) OR Pilates exercises (mat, DVD-guided)	Control (waitlist)	None specified	Anxiety Depression Fatigue Mood Physical activity
Marandi 2013 (40-42)	Quasi-RCT	Multiple sclerosis (women, EDSS < 4.5)	Pilates exercises (mat)	Control (no intervention) OR Aqua Fitness^	Usual care	Physical performance Dynamic Balance
Rezvani 2017 (43)	Quasi-RCT	Multiple sclerosis (women, EDSS ≤ 4)	Pilates exercises (mat)	Control (no intervention) OR Physio Ball exercise^	None specified	Disability Functional mobility Dynamic Balance
Sisi 2013 (44)	Quasi-RCT	Multiple sclerosis (men, EDSS ≤ 4)	Pilates exercises (mat)	Control (no intervention) OR Rebound therapy^	None specified	Functional mobility Balance
Pilates versus 'other' intervention*						
Abasiyanik 2020 (45-47)	Quasi-RCT	Multiple sclerosis (able to walk 100m independently)	Pilates exercises (mat)	Home exercise	Home exercise (2 days per week)	Functional mobility Balance Core stability Respiratory muscle strength Cognitive function
Bulguroglu 2015 (48, 49)	Quasi-RCT	Multiple sclerosis (EDSS < 4.5)	Pilates exercises (mat) OR Pilates (Reformer)	Home exercise (relaxation and respiration exercises)	None specified	Core stability Fatigue Balance Functional mobility QoL-disease specific

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR/s	CO-INTERVENTION	OUTCOME DOMAINS
Freeman 2012 (50-53)	RCT	Multiple sclerosis (EDSS 4.0 to 6.5)	Pilates exercises (mat)	Standard physiotherapy OR Relaxation sessions (contract-relax)	None specified	Functional mobility Balance Physical performance Physical activity Functional reach
Guclu-Gunduz 2014 (54)	NRSI	Multiple sclerosis (ambulant)	Pilates exercises	Home exercise	None specified	Balance Functional mobility Physical performance
Kalron 2016 (55, 56)	RCT	Multiple sclerosis (EDSS 3.0 to 6.0)	Pilates exercises (individual)	Physical therapy	Home exercise	Functional mobility Balance Functional reach Physical activity Fatigue Postural control Falls risk
Kara 2017 (57)	NRSI	Multiple sclerosis (EDSS ≤ 6.0)	Pilates exercises (mat)	Aerobic exercise	None specified	Balance Depression Fatigue Cognitive function Functional mobility
Küçük 2015 (58, 59)	Quasi RCT	Multiple sclerosis (EDSS < 6.0)	Pilates exercises (mat)	Standard exercise	None specified	Balance Depression Fatigue Cognitive function Functional mobility Physical performance QoL-disease specific

Abbreviations: EDSS, expanded disability status scale; NRSI, non-randomised study of intervention; PDDS, Patient-Determined Disease Steps; QoL, quality of life; RCT, randomised controlled trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

^ Study included three groups. The inactive control is considered in the evidence synthesis.

D3.1.2 Risk of bias summary

Randomised controlled trials

The risk of bias for each item in the included RCTs for multiple sclerosis is described below and shown graphically in Figure D.4 (details are provided in Appendix E1).

Bias arising from the randomisation process

Three studies (Duff 2018, Freeman 2012, Kalron 2016) were judged to be at low risk of bias for this domain. Details relating to method of randomisation or allocation concealment were not provided in the other studies, raising some concerns (Abasiyanik 2018, Bulguroglu 2015, Fleming 2019, Küçük 2015). Studies that also failed to provide sufficient baseline data (Eftekhari 2018, Marandi 2013, Rezvani 2017, Sisi 2013) were judged to be at high risk of bias.

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

Seven studies were judged to be at low risk of bias for this domain (Abasiyanik 2018, Bulguroglu 2015, Duff 2018, Fleming 2019, Freeman 2012, Kalron 2016, Küçük 2015). Four studies had concerns or serious

concerns raised due to missing information or exclusion of participants who did not adhere to the intended intervention (Eftekhari 2018, Marandi 2013, Rezvani 2017, Sisi 2013).

Bias due to missing outcome data

Data was available for all, or nearly all participants in four studies (Bulguroglu 2015, Duff 2018, Freeman 2012, Kalron 2016) that were judged to be at low risk of bias. In five studies (Abasiyanik 2018, Küçük 2015, Marandi 2013, Rezvani 2017, Sisi 2013) the proportion of missing outcome data across the intervention groups raised some concerns. In Eftekhari 2018 and Fleming 2019, missingness of the outcome data was judged to depend on its true value, with missing data likely to be due to the health status of the patients (assessed as high risk of bias).

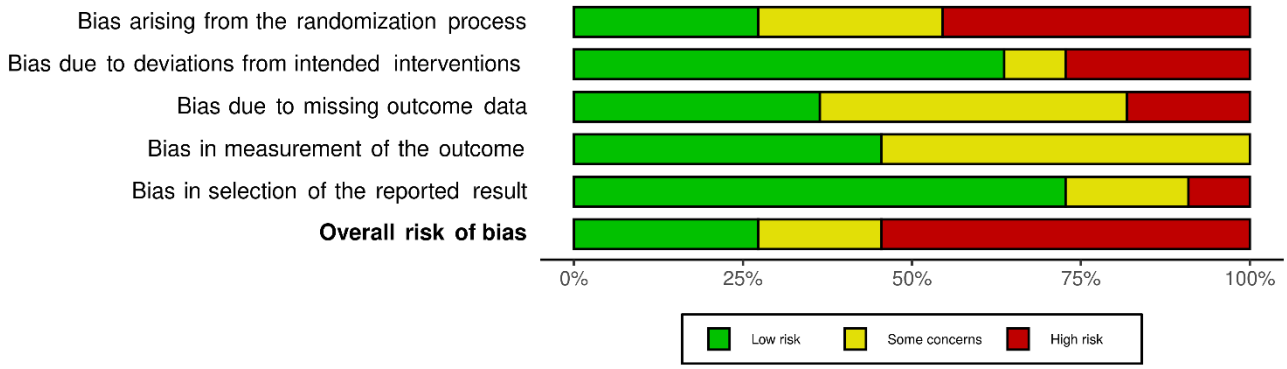
Bias in measurement of the outcome

Five studies were judged to be at low risk of bias for this domain (Bulguroglu 2015, Duff 2018, Eftekhari 2018, Freeman 2012, Kalron 2016), with many reporting blinding of the outcome assessor. Six studies (Abasiyanik 2018, Fleming 2019, Küçük 2015 Marandi 2013, Rezvani 2017, Sisi 2013) had concerns raised relating to subjective outcomes potentially influenced by knowledge of the intervention received.

Bias in selection of the reported result

Eight studies (Abasiyanik 2018, Bulguroglu 2015, Duff 2018, Eftekhari 2018, Fleming 2019, Freeman 2012, Kalron 2016, Küçük 2015) were judged to be at low risk of bias for this domain. In two studies (Marandi 2013, Sisi 2013) there were concerns raised about a lack of information describing the statistical analysis plan, suggesting not all intended outcomes or analyses were reported. One study (Rezvani 2017) was judged to be at high risk of bias, because results of an outcome described in the methods (timed-up and go) are not mentioned or discussed.

Figure D.4 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Multiple sclerosis



Non-randomised studies of interventions

The risk of bias for each item in the included NRSIs for multiple sclerosis is described below and shown graphically in Figure D.5 (details are provided in Appendix E2).

Bias due to confounding

The two studies (Guclu-Gunduz 2014 and Kara 2017) were assessed to be at low risk of bias for this domain. No confounding factors relating to treatment choice were expected.

Bias of selection of participants into the study

The two studies (Guclu-Gunduz 2014 and Kara 2017) were assessed to be at low risk for this domain. All eligible participants were invited to participate in the study and start of interventions coincided.

Bias in classification of interventions

Both studies (Guclu-Gunduz 2014 and Kara 2017) provided clear description and definition of intervention groups and was assessed to be low risk for this domain.

Bias due to deviations from intended interventions

Guclu-Gunduz 2014 did not provide sufficient information on reasons for deviation from intended intervention raising some concerns. Of 24 participants allocated to the intervention or control group, six did not continue the program (25%), this was judged to potentially impact the results. In Kara 2017, 18 of 27 (67%) participants allocated to Pilates did not complete the study, this was unbalanced between the intervention groups and judged to be a substantial deviation from usual practice, placing the study at critical risk of bias.

Bias due to missing data

Guclu-Gunduz 2014 did not provide sufficient information on missing information, therefore it was assumed that the proportions of and reasons for missing data differs slightly across intervention groups, raising some concerns.

Kara 207 was judged to be at critical risk of bias due to deviations from the intended intervention and was therefore not assessed for this domain.

Bias in measurement of outcomes

Guclu-Gunduz 2014 was judged to be moderate risk of bias in measurement of outcomes related to the knowledge of the intervention received by study participants potentially influencing the outcome results.

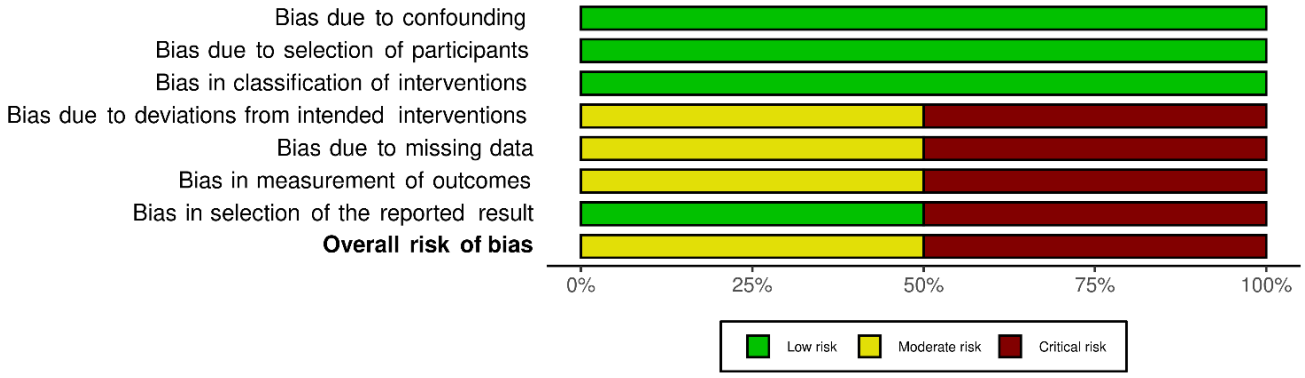
Kara 207 was judged to be at critical risk of bias due to deviations from the intended intervention and was therefore not assessed for this domain.

Bias in selection of the reported result

Guclu-Gunduz 2014 was assessed to be low risk for this domain. There is no indication of selection of the reported outcomes or analysis based on the results.

Kara 207 was judged to be at critical risk of bias due to deviations from the intended intervention and was therefore not assessed for this domain.

Figure D.5 **Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSI – Multiple sclerosis**



D3.1.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with multiple sclerosis are listed in Table D-9.

Table D-9 **Outcomes considered by the NTWC to be critical or important for decision-making: Multiple sclerosis**

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Duff 2018*	*Eftekhari 2018*	*Fleming 2019*	*Marandi 2013*	*Rezvani 2017*	*Sisi 2013*
Quality of life, global	Multiple sclerosis QOL-54	Critical	Yes	✓	--	--	?	--	--
Balance	Berg Balance Scale	Critical	Yes	--	✓	--	--	--	✓
Functional mobility	Timed Up and Go	Critical	Yes	✓	✓	--	?	X	✓
Physical performance	MS walking scale	Critical	No	--	--	--	?	--	--
Disability	No measures reported in eligible studies	Important	No	--	--	--	?	--	--
Fatigue	Fatigue impact scale (40-item) / modified fatigue impact scale (21-item)	Important	Yes	--	✓	✓	?	--	--
Social wellbeing	No measures reported in eligible studies	Important	No	--	--	--	?	--	--

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Six RCTs (Duff 2018, Eftekhari 2018, Fleming 2019, Marandi 2013, Rezvani 2017, Sisi 2013) comparing Pilates with no intervention in people with multiple sclerosis were eligible for this comparison. Four RCTs (Duff 2018, Eftekhari 2018, Fleming 2019, Sisi 2013) contributed data relevant to four of the seven outcomes. One study (Rezvani 2017) did not provide any data for one outcome but is included in the nonquantitative synthesis. One study (Marandi 2013) could have contributed data to these outcomes but there was no information to make a judgment regarding the extent of missing data.

There were eight additional studies awaiting classification or ongoing (total 203+ participants) that compared Pilates with no intervention in people with multiple sclerosis that could have contributed data to the critical or important outcomes but there was no information to make a judgment regarding the extent of missing data (see Appendix C6).

Results for two outcomes (balance and fatigue) were judged to be at high risk of bias. The concern was linked to missing outcome data and issues with the randomisation process.

Quality of life

One trial (30 participants) reported quality of life measured with the MSQoL-54 at the end-of-treatment (12 weeks) (Duff 2018). Four RCTs did not measure or report this outcome, probably because it was not assessed and one RCT there was no information to make a judgement.

The 54-item instrument generates 12 subscales[2] along with two summary scores[3], and two additional single-item measures[4] (60, 61). The study authors reported the two composite scores (physical and mental health), which are summarised on a scale from 0 (worse) to 100 (best). No MCID for the MSQoL-54 has been established but a change of ≥ 5 points in PCS and MCS has been proposed to be clinically meaningful (62, 63).

The results show no between group difference comparing the Pilates and control groups for mental health (MD 6.90; 95% CI –4.72, 18.52; $p = 0.24$) *(GRADE: low)* or physical health (MD 3.40; 95% CI –9.89, 16.69; $p = 0.62$) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data.

Balance

Two trials (55 participants) reported on balance stability during a series of predetermined tasks measured using the Berg Balance test at the end-of-treatment (8 weeks) (Eftekhari 2018, Sisi 2013). Three RCTs measured static and dynamic balance using different measures and one RCT did not measure or report this outcome, probably because it was not assessed.

In most of the 14-items of the Berg Balance test, the subject is asked to maintain a given position for a specific time, with each item consisting of a five-point ordinal scale ranging from 0 to 4. Total scores range from 0 (low balance stability) to 56 (high balance stability). In elderly people, a score of less than 45 indicates

[2] physical function, role limitations-physical, role limitations-emotional, pain, emotional well-being, energy, health perceptions, social function, cognitive function, health distress, overall quality of life, and sexual function

[3] physical health composite summary and the mental health composite summary

[4] satisfaction with sexual function and change in health

individuals may be at greater risk of falling (64). The MCID for improvement in balance in people with multiple sclerosis is 3 points (65).

The results reported by Sisi 2013 (30 participants) suggest an effect in favour of Pilates compared with control (MD –7.43; 95% CI –9.35, –5.51; $p < 0.0001$), however differences in baseline scores between treatment groups suggest the size of the effect may be overstated in favour of the intervention (baseline data were skewed) *(GRADE: very low)*. Scores for participants in both groups remain below 45.

Eftekhari 2018 also reported an effect favouring Pilates ($p = 0.003$), but the available data were not able to be interpreted as the scores do not correlate with expected total values (published total scores were identical in both groups at baseline and end of treatment) (data provided in Appendix F2).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as removal of both studies would leave no result.

Functional mobility

Three trials (80 participants) reported functional mobility measured with the Timed Up and Go test (TUG) at the end-of-treatment (8 or 12 weeks) (Duff 2018, Rezvani 2017, Sisi 2013). The other three RCTs assessed mobility using different measures (focused on aerobic capacity and endurance) and one RCT did not measure or report this outcome, probably because it was not assessed.

Developed for older adults (aged 70 to 84 years), the TUG test has been validated for people with multiple sclerosis (66); with people who take more than 14.4 seconds to complete the test classified as fallers. Pooled results show no between group differences comparing the Pilates and control groups for functional mobility (MD –0.55; 95% CI –2.11, 1.01; $p = 0.49$) *(GRADE: low)*; noting that one study (Rezvani 2017) did not report any data for this outcome (missing data). No MCID has been established in people with MS, with the minimal detectable change reported to be 3.5 seconds in people with Parkinson's disease (67) and 2.9 seconds (68) in people with chronic stroke.

Sensitivity analysis showed no important difference in the observed effect when the two RCTs judged to be at a high risk of bias (Rezvani 2017, Sisi 2013) were not included in the analysis (MD –0.70; 95% CI –3.95, 2.55; $p = 0.67$).

Fatigue

Two trials (34 participants) reported fatigue measured with the modified fatigue impact scale (MFIS) at the end-of-treatment (8 or 12 weeks) (Fleming 2019, Eftekhari 2018). Three RCTs did not measure or report this outcome, probably because it was not assessed and one RCT there was no information to make a judgement.

The MFIS provides an assessment of the perceived impact of fatigue in terms of physical, cognitive, and psychosocial functioning over the previous 4 weeks. The 21-item MFIS (reported by Fleming 2019) is summarised to a total score ranging from 0 (no fatigue) to 84 (severe fatigue); whereas the 5-item MFIS (reported by Eftekhari 2018) is summarised to a total score ranging from 0 (no fatigue) to 20 (severe fatigue). Standardised pooled mean results suggest an effect in favour of Pilates compared with control (SMD –1.13; 95% CI –1.88, –0.37; $p = 0.003$) *(GRADE: very low)*.

Fleming 2019 included two Pilates groups (supervised Pilates classes and home-DVD guided Pilates) compared with control. The supervised Pilates groups is included in the pooled analysis (to avoid double counting of the control group). When the DVD-guided group is considered, the similar results were observed (SMD –1.04; 95% CI –1.74, –0.35; $p = 0.003$).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as removal of both studies would leave no result.

Comparison 2 (vs other)

Seven RCTs (Abasiyanik 2018, Bulguroglu 2015, Freeman 2012, Kalron 2016, Küçük 2015, Rezvani 2017, Sisi 2013) and one NRSI (Guclu-Gunduz 2014) comparing Pilates with 'other' interventions in people with multiple sclerosis were eligible for this comparison and contributed data to six outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

D3.2 Myelopathy (HTLV-1 associated)

D3.2.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-10. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-10 Overview of PICO criteria of included studies: Myelopathy (HTLV-1 associated)

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Borges 2014 (69)	RCT	Myelopathy (HTLV-1 associated)	Pilates exercises (Reformer, Cadillac and mat)	Control (usual activities)	None specified	Pain QoL
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: HTLV, Human T-lymphotropic virus; QoL, quality of life; RCT, randomised controlled trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

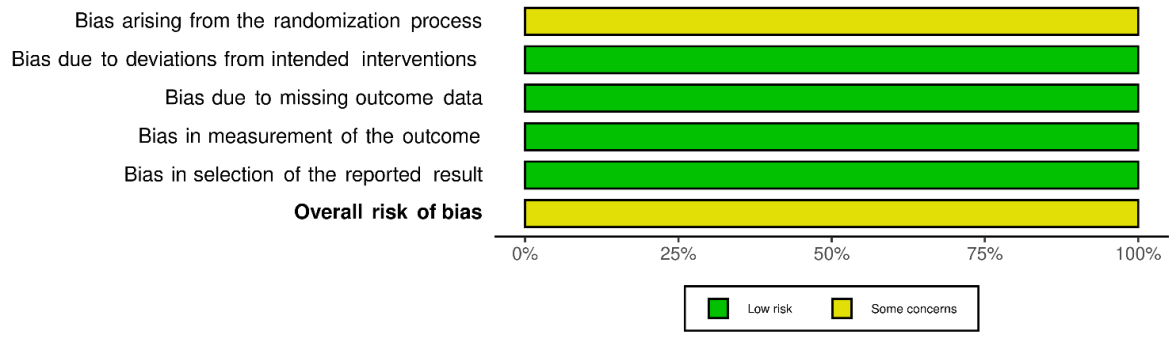
**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D3.2.2 Risk of bias summary

The risk of bias for each item in the included study for myelopathy is described below and shown graphically in Figure D.6 (details are provided in Appendix E1).

The study by Borges 2014 was assessed to be at low risk of bias for all domains except for bias arising from the randomisation process. Patients were randomised by a table of random numbers and were allocated into groups by a blinded team member who did not participate in the assessment, treatment, or statistical analysis phases, however, the absence of baseline characteristics raises some concerns.

Figure D.6 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Myelopathy (HTLV-1 associated)



D3.2.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with HTLV-1 associated myelopathy are listed in Table D-11.

Main comparison (vs control)

One RCT (Borges 2014) comparing Pilates with no intervention in participants with myelopathy (HTLV-1 associated) was eligible for this comparison and contributed data to six of seven outcomes. There were no additional studies identified (awaiting classification or ongoing) in that compared Pilates with no intervention in people with HTLV-1 associated myelopathy that could have contributed data to these outcomes (see Appendix C6).

There were some concerns of bias for all patient-reported outcomes linked to insufficient reporting of baseline characteristics, making it difficult to make a judgement regarding the randomisation process.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data that was not considered to be at high risk of bias.

Table D-11 Outcomes considered by the NTWC to be critical or important for decision-making: Myelopathy (HTLV-1 associated)

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Borges 2014*
Pain	Visual analogue scale	Critical	Yes	✓
Quality of life, global	No measures reported in eligible studies	Critical	No	--
Physical functioning	SF-36 Role - Physical / SF-36 physical function	Critical	Yes	✓
Fatigue	SF-36 vitality	Important	Yes	✓
Mental health	SF-36 mental health	Important	Yes	✓
Mental function	SF-36 Role - Emotional	Important	Yes	✓
Social function	SF-36 Social functioning	Important	Yes	✓

Abbreviations: SF-36, 36-item Short Form Survey

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Pain

One trial (22 participants) reported pain intensity measured with a visual analogue scale (VAS) ranging from 0 (no pain) to 10 (severe pain) at the end-of-treatment (15 weeks) (Borges 2014). The VAS is a subjective assessment of pain, reported by participants and measured on a continuous scale (cm) from 0 (no pain) to 10 (worst imaginable pain). Higher values indicate worse pain. The minimal clinically important difference (MCID) has not been established in people with HTLV-associated myelopathy, with the median absolute MCID reported to be 20 mm (IQR 15–30) in people with chronic pain (70).

The results showed an effect in favour of Pilates compared with the control group (MD –4.05; 95% CI –6.16, –1.94; p = 0.0002) *(GRADE: very low)*.

Quality of life

One trial (22 participants) reported of quality of life measured with the SF-36 (36-item Short Form Survey) at the end-of-treatment (15 weeks) (Borges 2014). An overall global score was not provided. Individual scores for each of the eight domains were reported and are summarised on a scale from 0 (worse) to 100 (best).

The MCID for individual domains of the SF-36 in people with HTLV-associated myelopathy have not been established.

Physical functioning

The SF-36 physical functioning domain is a 10-item measure of physical limitation in a range of activities from vigorous exercise to performing self-care activities. The results showed no difference in scores between the Pilates group compared with the control group (MD – 9.82; 95% CI –24.78, 5.14; $p = 0.20$) *(GRADE: very low)*.

Role physical

The SF-36 role-physical domain contains four items that measures limitations in various roles, including work and daily activities. The results showed an effect in favour of Pilates compared with the control group (MD – 62.73; 95% CI –84.54, –40.92; $p = 0.004$) *(GRADE: very low)*.

Bodily pain

The SF-36 bodily pain domain contains two items that assess pain severity and pain interference. The results showed an effect in favour of Pilates compared with the control group (MD – 30.14; 95% CI –44.77, –15.51; $p < 0.0001$) *(GRADE: very low)*.

General health perceptions

The SF-36 general health perceptions domain contains five items that evaluates a persons' physical health problems and their confidence in progression to better or worse health. The results showed an effect in favour of Pilates compared with the control group (MD –20.53; 95% CI –39.22, –1.84; $p = 0.03$) *(GRADE: very low)*.

Fatigue

The SF-36 vitality domain has four items that measure vitality, energy level, and fatigue and is meant to be a measure of subjective well-being. The results showed an effect in favour of Pilates compared with the control group (MD –28.36; 95% CI –47.76, –8.96; $p = 0.004$) *(GRADE: very low)*.

Mental health

The SF-36 mental health domain has five items that measure anxiety, depression, loss of behavioural/emotional control, and psychological well-being. The results showed no difference in scores between the Pilates group compared with the control group (MD –15.82; 95% CI –35.80, 4.16; $p = 0.12$) *(GRADE: very low)*.

Mental function

The SF-36 role-emotional domain measures role limitations due to mental health difficulties, with three items that focus on amount of time spent on work or other activities, amount of work accomplished, and the care with which work is performed. The results showed no difference in scores between the Pilates group compared with the control group (MD 6.86; 95% CI –21.25, 34.97; $p = 0.63$) *(GRADE: very low)*.

Social function

The SF-36 social function scale includes two items that measure the impact of physical and mental health on social functioning. The results showed no difference in scores between the Pilates group compared with the control group (MD –14.32; 95% CI –33.78, 5.14; $p = 0.15$) *(GRADE: very low)*.

Comparison 2 (vs other)

No studies were identified comparing Pilates with 'other' in people with HTLV-1 associated myelopathy.

D3.3 Parkinson's disease

D3.3.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-12. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-12 Overview of PICO criteria of included studies: Parkinson's disease

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Pandya 2017 (71)	Quasi-RCT	Parkinson's disease	Pilates exercises (mat, Swiss ball, TheraBand)	Control (no intervention)	Conventional balance training	Balance Functional mobility Balance confidence
Pilates versus 'other' intervention*						
Daneshmandi 2017 (72)	Quasi-RCT	Parkinson's disease	Pilates exercises (mat, mini ball, TheraBand)	Walking	None specified	Balance Functional mobility
Mollinedo-Cardalda 2018 (73)	RCT	Parkinson's disease	Pilates exercises (mat, TheraBand)	Aerobic exercises	None specified	Body composition Balance Strength Motor scale

Abbreviations: NRSI, non-randomised study of intervention; QoL, quality of life; RCT, randomised controlled trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D3.3.2 Risk of bias summary

The risk of bias for each item in the included studies for Parkinson's disease is described below and shown graphically in Figure D.7 (details are provided in Appendix E1).

Bias arising from the randomisation process

One study (Mollinedo-Cardalda 2018) was judged to be at low risk of bias for this domain. Details relating to method of randomisation or allocation concealment were not provided in the other two studies (Daneshmandi 2017, Pandya 2017) raising some concerns.

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

One study (Daneshmandi 2017) was judged to be at low risk of bias for this domain, with any deviations judged reflect usual practice. In one study (Pandya 2017) concerns were raised due to missing information regarding the number of analysed participants. One study (Mollinedo-Cardalda 2018) was judged to be at high risk of bias due to deviations being unbalanced between the groups which may have affected the outcome. The direction of bias may be against Pilates, with a higher rate of dropout observed in the comparator group (suggesting those who do not benefit did not complete the study).

Bias due to missing outcome data

Two studies (Daneshmandi 2017, Pandya 2017) were judged to be at low risk of bias for this domain, as the data were reasonably complete, and any missing data was similar across intervention groups. In one study (Mollinedo-Cardalda 2018), concerns were raised due to the missing information being slightly unbalanced between treatment groups.

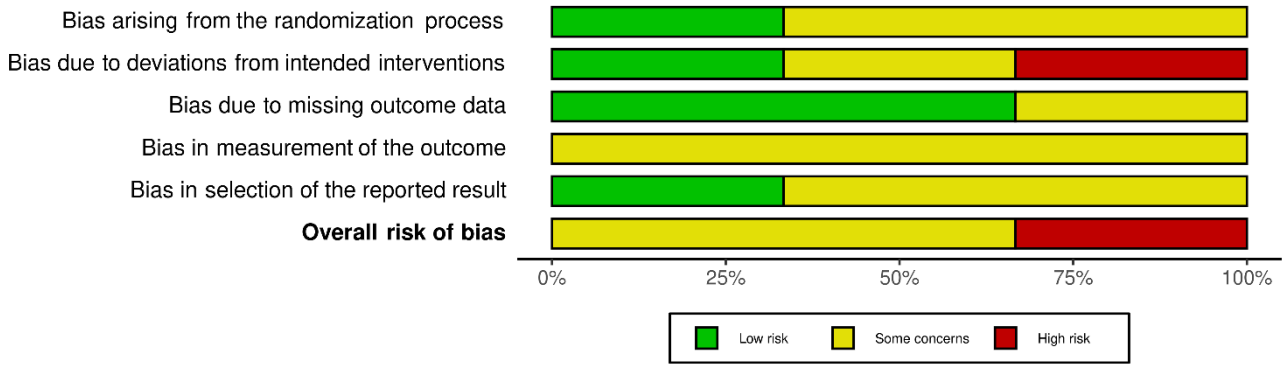
Bias in measurement of the outcome

In all three studies (Daneshmandi 2017, Mollinedo-Cardalda 2018, Pandya 2017) there were concerns of bias related to outcome assessors likely being influenced by knowledge of the intervention received.

Bias in selection of the reported result

One study (Mollinedo-Cardalda 2018) was judged to be at low risk of bias for this domain. The other two studies had concerns raised due to missing information (Pandya 2017) or strong suspicion of selective reporting of results (Daneshmandi 2017).

Figure D.7 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Parkinson's disease



D3.3.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with Parkinson's disease are listed in Table D-13.

Table D-13 Outcomes considered by the NTWC to be critical or important for decision-making: Parkinson's disease

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*
Quality of life, disease specific	Parkinson's Disease Quality of Life Questionnaire (PDQ-8)	Critical	No
Overall motor function	MDS-Unified Parkinson's Disease Rating Scale	Critical	No
Functional mobility	Timed Up and Go (TUG)	Critical	Yes
Gait	No measures reported in eligible studies	Critical	No
Disability	No measures reported in eligible studies	Critical	No
Balance	Berg Balance test	Important	Yes
Falls	Number of falls within previous year	Important	No

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One RCT (Pandya 2017) evaluating the effect of Pilates delivered as an adjunct to conventional balance training in people with Parkinson's disease was eligible for this comparison and contributed data relevant to two of the seven outcomes. There was one additional study published in a language other than in English (106 participants) that compared Pilates with no intervention in Parkinson's disease that could have contributed data to one outcome (balance) (see Appendix C6).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and was not judged to be at high risk of bias.

Functional mobility

One trial (30 participants) reported functional mobility measured with the Timed Up and Go test (TUG) at the end-of-treatment (7 weeks) (Pandya 2017). Developed for older adults (aged 70 to 84 years), the test has been validated for people with Parkinson's disease to identify those at risk for falls (74), with the minimal detectable change reported to be 3.5 seconds (67).

The results reported by Pandya 2017 suggest an effect favouring Pilates compared with the control group (MD –8.53; 95% CI –13.37, –3.69; p = 0.0006) *(GRADE: very low)*.

Balance

One trial (30 participants) reported on the ability to balance during a series of predetermined tasks measured using the Berg Balance test at the end-of-treatment (7 weeks) (Pandya 2017). In most of the 14-items, the subject is asked to maintain a given position for a specific time, with each item consisting of a five-point ordinal scale ranging from 0 to 4. Total scores range from 0 (low balance stability) to 56 (high balance stability). In elderly people, a score of less than 45 indicates individuals may be at greater risk of falling (64), but the test may have poor utility in people with Parkinson's disease due to a ceiling effect (75, 76). The minimal detectable change for people with Parkinson's disease is 5 points (77).

The results reported by Pandya 2017 suggest an effect in favour of Pilates compared with control (MD –5.07; 95% CI –8.90, –1.24; p < 0.010) (*GRADE: very low)*. Participant in both groups remain at greater risk of falling (mean score below 45).

Comparison 2 (vs other)

Two studies (Daneshmandi 2017, Mollinedo-Cardalda 2018) comparing Pilates with 'other' interventions in people with Parkinson's disease were eligible for this comparison and contributed data to three outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

D3.4 Rehabilitation after stroke

D3.4.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-14. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-14 Overview of PICO criteria of included studies: Rehabilitation after stroke

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Lim 2016 (78)	Quasi-RCT	Stroke recovery (chronic)	Pilates exercises (mat)	Control (no intervention)	None	Balance, static Balance, dynamic
Lim 2017 (79)	RCT	Stroke recovery (chronic)	Pilates exercises (mat)	Control (no intervention)	Conventional stroke rehabilitation	Exercise tolerance Functional mobility
Roh 2016 (80)	Quasi-RCT	Stroke recovery (chronic)	Pilates exercises	Control (no exercise)	None	Functional mobility
Sathe 2018 (81)	Quasi-RCT	Stroke recovery (chronic)	Pilates exercises	Control (no intervention)	Conventional balance therapy	Functional mobility Limits of stability
Pilates versus 'other' intervention*						
Yun 2017 (82)	NRSI	Stroke recovery (chronic)	Pilates exercises (mat)	Occupational therapy	None	Motor function Mood Social function QoL

Abbreviations: NRSI, non-randomised study of intervention; QoL, quality of life; RCT, randomised controlled trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D3.4.2 Risk of bias summary

Randomised controlled trials

The risk of bias for each item in the included RCTs for rehabilitation after stroke is described below and shown graphically in Figure D.8 (details are provided in Appendix E1).

Bias arising from the randomisation process

All four RCTs (Lim 2016, Lim 2017, Roh 2016, Sathe 2018) failed to provide sufficient information about the randomisation process or methods for allocation concealment raising some concerns. In addition, a lack of information about baseline characteristics placed two RCTs (Roh 2016, Sathe 2018) at high risk of bias.

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

One study (Lim 2017) was assessed to be at low risk of bias for this domain as there were no deviations from the intended interventions. Concerns were raised with two studies (Roh 2016, Sathe 2018) that did not provide any information on deviations from the trial protocol and one study (Lim 2016) that failed to analyse all randomised participants.

Bias due to missing outcome data

Two studies (Lim 2016, Lim 2017) were assessed to be low risk for this domain as outcome data was available for all participants. Two studies (Roh 2016, Sathe 2018) had concerns raised for this domain as they did not provide any information to make a judgement about the extent of missing outcome data.

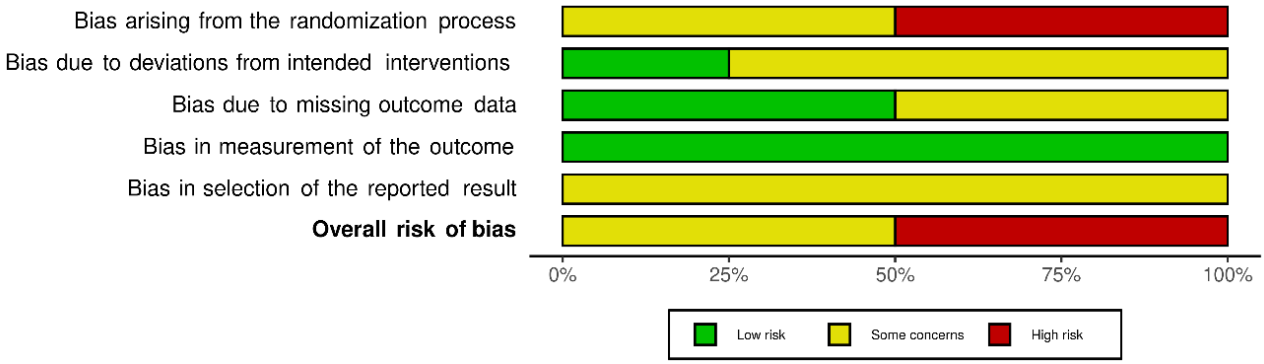
Bias in measurement of the outcome

All four studies (Lim 2016, Lim 2017, Roh 2016, Sathe 2018) were assessed to be low risk for outcome measurements. The methods were comparable across groups and unlikely to be influenced by knowledge of the intervention received.

Bias in selection of the reported result

All four studies (Lim 2016, Lim 2017, Roh 2016, Sathe 2018) had concerns raised due to a lack of information regarding prespecified intentions for outcome measurements and analyses, reducing confidence that all intended outcome measures are reported.

Figure D.8 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Rehabilitation after stroke



Non-randomised studies of interventions

The risk of bias for each item in the included NRSI for rehabilitation after stroke is described below and shown graphically in Figure D.9 (details are provided in Appendix E2).

Bias due to confounding

One study (Yun 2017) was assessed to have some concerns for this domain. Although confounding is expected, there did not appear to be any serious residual confounding and important domains were measured and controlled for.

Bias of selection of participants into the study

One study (Yun 2017) was assessed to be at low risk for this domain. Although not explicitly stated, all eligible participants are likely included in the study, with enrolment and start of intervention coinciding.

Bias in classification of interventions

One study (Yun 2017) was assessed to be at low risk for this domain. The study provided clear description and definition of intervention groups, which were defined at study start.

Bias due to deviations from intended interventions

One study (Yun 2017) was assessed to have some concerns for this domain. The authors did not provide information about deviations from intended interventions. It is assumed that there were no deviations, and if any, the impact on outcomes were minimal.

Bias due to missing data

One study (Yun 2017) was assessed to be at low risk for this domain. The authors did not provide information about any missing data. All available data appear to be included in the analysis.

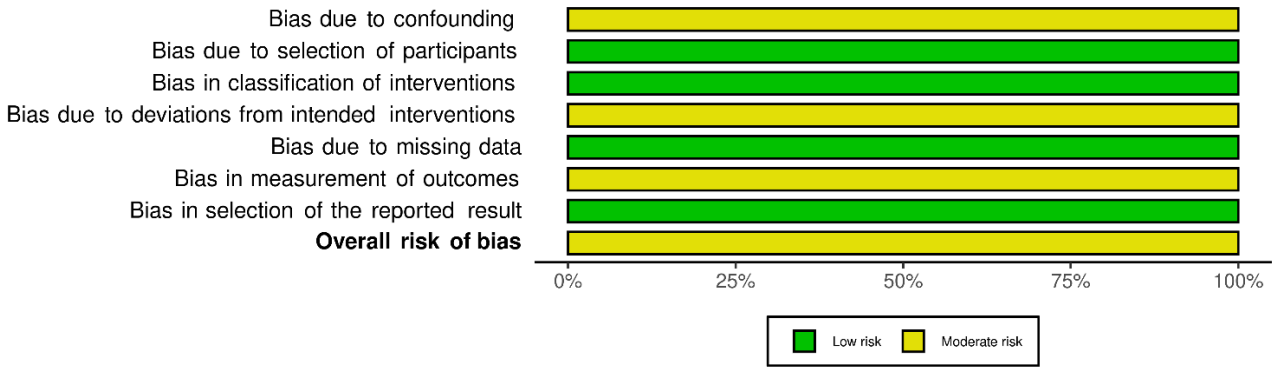
Bias in measurement of outcomes

One study (Yun 2017) was assessed to have some concerns for this domain. Knowledge of the intervention could potentially influence the patient-reported results.

Bias in selection of the reported result

One study (Yun 2017) was assessed to be at low risk for this domain. It is likely that all reported results correspond to all intended outcomes and analyses.

Figure D.9 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSI – Rehabilitation after stroke



D3.4.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with chronic stroke are listed in Table D-15.

Table D-15 Outcomes considered by the NTWC to be critical or important for decision-making: Rehabilitation after stroke

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Lim 2016*	*Lim 2017*	*Roh 2016*	*Sathe 2018*
Disability	PROMIS-10	Critical	No	--	--	--	--
Quality of life	SS-QOL-total PROMIS-10	Critical	No	--	--	--	--
Activities of daily living	PROMIS-10	Critical	No	--	--	--	--
Functional mobility	Gait speed (cm/s) or TUG	Critical	Yes	--	✓	✓	--
Motor function	SS-QOL -physical	Important	No	--	--	--	--
Balance	Berg Balance Test	Important	No	--	--	--	--
Cardiovascular disease risk	No measures reported in eligible studies	Important	No	--	--	--	--

Abbreviations: cm/s, centimetres per second; PROMIS-10; Patient-Reported Outcomes Measurement Information System Global-10; SS-QOL, stroke-specific quality of life

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Four RCTs (Lim 2016, Lim 2017, Roh 2016, Sathe 2018) comparing Pilates with no intervention (or as an adjunct to conventional therapy) in people with chronic stroke were eligible for this comparison. Two RCTs (Lim 2017, Roh 2016) contributed data relevant to one of the seven outcomes. The other two RCTs did not measure or assess any outcomes considered critical or important to this review.

There was one additional study published in a language other than English (awaiting classification) that compared Pilates with no intervention in people with chronic stroke (6 participants) that could have contributed data to one outcome (balance) (see Appendix C6).

Functional mobility

One trial (20 participants) measure functional mobility by assessing gait speed (cm/s) at the end of treatment (8 weeks) (Roh 2016). The study used 3-D motion analysis using infrared cameras, with participant required to walk on a treadmill for 30 seconds at their preferred walking speed. Five strides in the middle of the recording were used in the analysis. Typically, gait speed is a measure of time taken to walk a specified distance, with a clinically meaningful change estimate in stroke patients reported to be 10 cm/s (83). The results reported by Roh 2016 suggest an effect in favour of Pilates compared with no intervention (MD –9.94; 95% CI –18.16, –1.72; p = 0.02) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as removal of the study would leave no result.

One trial (20 participants) reported functional mobility measured with the Timed Up and Go test (TUG) at the end of treatment (8 weeks) (Lim 2017). The test has been validated for people with chronic stroke to identify those at risk for falls, with the minimal detectable change reported to be 2.9 seconds (68). Results

reported by Lim 2017 indicate no difference between groups comparing Pilates delivered as an adjunct to conventional stroke rehabilitation therapy (MD –2.50; 95% CI –19.43, 14.43; *p* = 0.77) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and was not judged to be at high risk of bias.

Comparison 2 (vs other)

One NRSI (Yun 2017) comparing Pilates with occupational therapy in people with chronic stroke were eligible for this comparison. The study contributed data relevant to two of the seven outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

D1 Diseases of the circulatory system

D3.5 Hypertension

D3.5.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-16. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-16 Overview of PICO criteria of included studies: Hypertension

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Martins-Meneses 2015 (84)	NRSI	Hypertension (women, 30 to 59 years)	Pilates exercises (mat)	Control (waitlist)	None specified	Heart rate Blood pressure BMI
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: BMI, body mass index; NRSI, non-randomised study of intervention

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D3.5.2 Risk of bias summary

The risk of bias for each item in the included studies for hypertension is described below and shown graphically in Figure D.10 (details are provided in Appendix E2).

Bias due to confounding

One study (Martins-Meneses 2015) was assessed at moderate risk of bias in this domain. There were no significant differences between groups at baseline for important potential confounding domains, but baseline characteristics were only presented for participants who remained in the trial after 4-months, potentially obscuring the baseline characteristics for the whole trial population.

Bias of selection of participants into the study

One study (Martins-Meneses 2015) was assessed at low risk of bias in this domain. Participants were followed from the start of the intervention making it unlikely that there was misclassification of outcome status. It is assumed all eligible participants were invited to participate.

Bias in classification of interventions

One study (Martins-Meneses 2015) was assessed at serious risk of bias in this domain. Intervention status was well defined and was determined after enrolment into the study, however participants were allocated based on how readily they could produce a medical certificate to participate. This was considered likely to be affected by knowledge of the outcome.

Bias due to deviations from intended interventions

One study (Martins-Meneses 2015) was assessed at critical risk of bias in this domain due to a higher-than-expected rate of dropouts in both groups. A further 5 were excluded from the analysis because they attended less than 75% of sessions. By the end of the trial, 37% of participants had dropped out or were

excluded from the analysis which is likely to overstate the effect of the intervention. The method of analysis used to account for this missing data (no analysis) was considered inappropriate.

Bias due to missing data

This domain was not assessed.

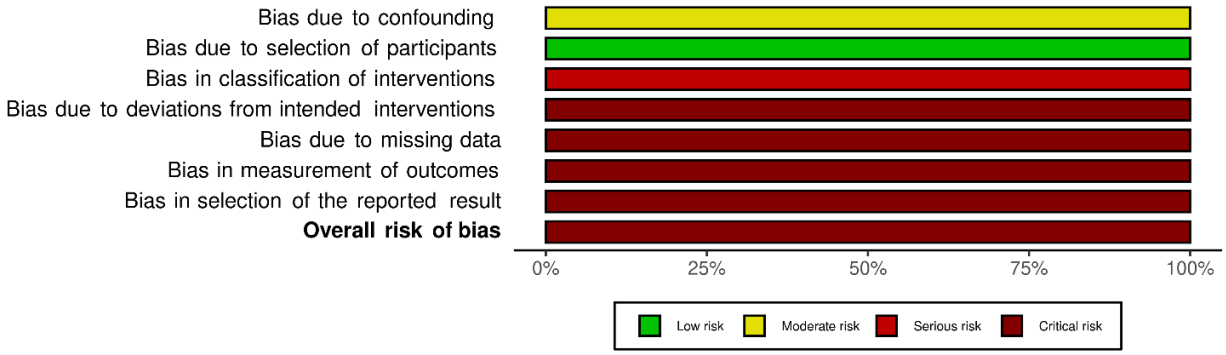
Bias in measurement of outcomes

This domain was not assessed.

Bias in selection of the reported result

This domain was not assessed.

Figure D.10 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSIs – Hypertension



D3.5.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with hypertension are listed in Table D-17.

Table D-17 Outcomes considered by the NTWC to be critical or important for decision-making: Hypertension

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Martins-Meneses 2015*
Quality of life	No measures reported in eligible studies	Critical	No	--
Cardiovascular disease-risk	Blood pressure (systolic, diastolic, mean) Heart rate OR Double product (HR x BP)	Critical	No	✓
Disease progression	No measures reported in eligible studies	Critical	No	--
Fitness/exercise capacity	No measures reported in eligible studies	Critical	No	--
Physical performance	No measures reported in eligible studies	Important	No	--
Body composition	Hip-waist ratio	Important	No	✓

Abbreviations: BP, blood pressure; HR, heart rate

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One NRSI (Martins-Meneses 2015) comparing Pilates with no intervention in women with hypertension was eligible for this comparison. The study reported data relevant to one outcome (cardiovascular disease-risk) but was judged to be at critical risk of bias due to substantial attrition (more than 35% missing data) and was therefore not considered in the reporting of results, evidence synthesis or conclusions.

There was one additional study published in a language other than English (awaiting classification) and one ongoing study that compared Pilates with no intervention in people with hypertension (total 90 participants) that could have contributed data some of the outcomes considered critical or important to this review (see Appendix C6).

Comparison 2 (vs other)

No studies were identified comparing Pilates with 'other' interventions in people with hypertension.

D4 Diseases of the musculoskeletal system or connective tissue

D4.1 Osteoarthritis

D4.1.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-18. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-18 Overview of PICO criteria of included studies: Osteoarthritis

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Mazloum 2018 (85)	Quasi-RCT	Osteoarthritis (knee)	Pilates exercises	Control (usual activities) OR Conventional therapeutic exercises^	None specified	Disability Functional status Physical performance
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: RCT, randomised controlled trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

^ Study included three groups. The inactive control is considered in the evidence synthesis.

D4.1.2 Risk of bias summary

The risk of bias for each item in the included studies for osteoarthritis is described below and shown graphically in Figure D.11 (details are provided in Appendix E1).

Bias arising from the randomisation process

Mazloum 2018 was assessed to have some concerns for this domain due to lack of information provided regarding the randomisation and allocation concealment processes. Minimal baseline characteristics (age, height, weight) were comparable across groups at baseline, and pre-test scores appear comparable between Pilates and control group. Significant difference noted in the pre-test scores for one outcome in the exercise therapy group.

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

Mazloum 2018 was assessed to have some concerns relating to this domain. Both participants and research staff are not blinded to the allocated interventions, and there is a lack of specific information provided regarding deviations from the intended interventions. Eight participants did not receive the intended intervention due to 'personal reasons' and were excluded from the analysis. The number and proportion of patients who did not receive the allocated intervention is balanced between groups, therefore was considered to not have a substantial impact on the result.

Bias due to missing outcome data

Mazloum 2018 was assessed to be at high risk of bias for this domain, due to the missing data from participants who were excluded from the analysis. There were no analyses to test the impact of this

exclusion. Reasons for dropout were not specified, but it is possible the participants health status likely influenced the results. The missing data was therefore considered likely to seriously affect the true value.

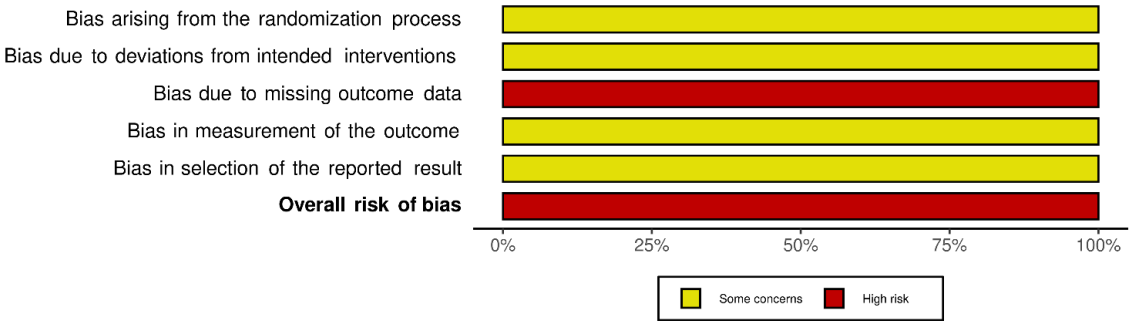
Bias in measurement of the outcome

Mazloum 2018 was assessed to have some concerns in this domain due to the self-reported nature of the outcomes. While the authors specified that the outcome assessor was blinded, the participants were aware of their allocation and could have plausibly biased the reporting of patient-reported outcomes.

Bias in selection of the reported result

Mazloum 2018 was assessed to be at low risk of bias in this domain, as it was considered that there was sufficient information to ensure that all eligible results were reported.

Figure D.11 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Osteoarthritis



D4.1.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with osteoarthritis are listed in Table D-19.

Table D-19 Outcomes considered by the NTWC to be critical or important for decision-making: Osteoarthritis

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Mazloum 2018*
Pain	Visual analogue scale (VAS)	Critical	No	--
Quality of life	EQ-5D-3L	Critical	No	--
Global physical functioning/disability	KOOS-PS or HOOS-PS (or Lequesne Index)	Critical	Yes	✓
Physical performance	time to complete activities	Critical	Yes	✓
Self-efficacy	No measures reported in eligible studies	Important	No	--
Proprioception	Biodex system	Important	No	--
Work status	Ability to work	Important	No	--

Abbreviations: EQ-5D-3L, European Quality of Life Five Dimension Three Level; KOOS-PS; Knee Injury and Osteoarthritis Outcome Score – Physical function short form; HOOS-PS, Hip disability and Osteoarthritis Outcomes Score – Physical function short form

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One study (Mazloum 2018) comparing Pilates with no intervention (or inactive control) in people with osteoarthritis of the knee was eligible for this comparison and contributed data to two of the seven outcomes. There was one additional study published in a language other than English (awaiting classification) that compared Pilates with no intervention in people with knee osteoarthritis that could have contributed data to these outcomes but there was no information to make a judgment regarding the extent of missing data (see Appendix C6).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and was not judged to be at high risk of bias.

Global physical functioning/disability

One trial (33 participants) measured physical functioning/disability using the Lequesne Index at the end of treatment (8 weeks) (Mazloum 2018). The Lequesne Index is used for subjective evaluation of pain and disability in patients with knee osteoarthritis and is considered a moderately reliable instrument for evaluating OA severity in Iran (86). Scores can range between 0 and 24, with higher scores reflecting greater pain and disability. A score lower than 8 is considered to reflect minor to moderate disability.

The results showed an effect in favour of Pilates (MD –2.10; 95% CI –3.36, –0.84, $p = 0.001$) *(GRADE: very low)*, but it is not clear if this effect would be considered clinically significant (87). The MCID of the Lequesne Index in people with knee osteoarthritis is 2.75 (88).

Physical performance

One trial (33 participants) measured physical performance based on the time to complete activities at the end of treatment (8 weeks) (Mazloum 2018). These activities included walking 15 metres (50-foot walk test), standing up from a chair and walking 15 metres (Timed Up and Go), going up and down 11 stairs (height of 12 centimetres) (stair climb test)[5]. Only the composite score was reported. A lower time taken to complete activities is representative of improved physical performance (mobility), but no information was found on the MCID in people with knee OA.

The results showed an effect in favour of Pilates (MD –9.60; 95% CI –13.46, –5.74; $p < 0.00001$) *(GRADE: very low)*.

Comparison 2 (vs other)

One study (Mazloum 2018) comparing Pilates with conventional therapeutic exercises in people with knee osteoarthritis was eligible for this comparison and contributed data relevant to two of the seven outcomes.

Available data from this study are presented in Appendix F2 Supplementary outcome data.

[5] It is not clear if the 30-second Chair Stand Test is also included.

D4.2 Post viral arthropathies

D4.2.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-20. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-20 Overview of PICO criteria of included studies: Post viral arthropathies

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
de Oliveira 2019 (89, 90)	RCT	Post viral arthropathy (Chikungunya, chronic)	Pilates exercises (equipment)	Control (no intervention)	Standard medical care	Functional capacity Joint function Pain Quality of life
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: RCT, randomised controlled trial

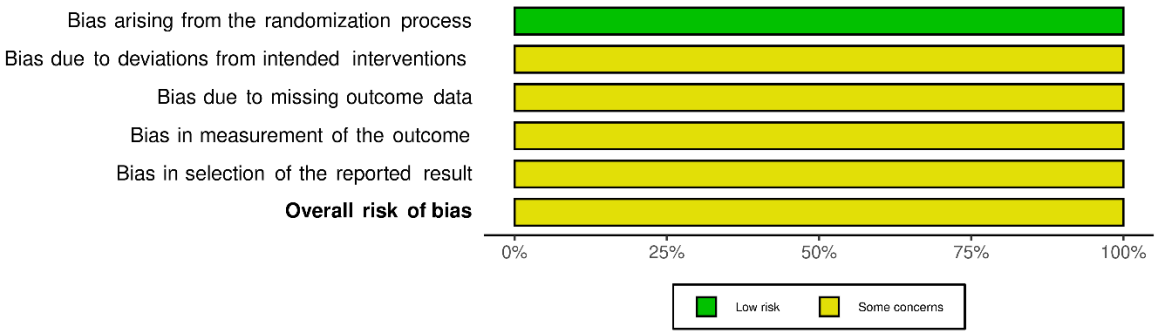
*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D4.2.2 Risk of bias summary

The risk of bias for each item in the included studies for post viral arthropathies is described below and shown graphically in Figure D.12 (details provided in Appendix E1).

Figure D.12 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Post viral arthropathies



Bias arising from the randomisation process

de Oliveira 2019 was assessed to be at low risk of bias in this domain. The method of generating the randomisation sequence and concealing allocation were well described. While there were some differences in baseline characteristics, these are not considered to indicate an issue with the randomisation process (i.e., were considered compatible with chance).

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

de Oliveira 2019 was assessed to have some concerns relating to this domain, due to a lack of blinding of both participants and research staff. A total of nine participants dropouts or failed to complete the assigned intervention, which was judged consistent within the trial context. An appropriate method of analysis (mITT) was used.

Bias due to missing outcome data

de Oliveira 2019 was assessed to have some concerns relating to this domain, due to the high number of participants with missing outcome data (more than 15%) who were excluded from the analysis without evidence that the result was not biased by this missing data.

Bias in measurement of the outcome

de Oliveira 2019 was assessed to have some concerns in this domain due to the self-reported nature of the outcomes and a lack of blinding of participants. While the authors specified that the outcome assessor was blinded, the participants were aware of their allocation and could have plausibly biased the reporting of their answers.

Bias in selection of the reported result

de Oliveira 2019 was assessed to have some concerns in this domain. There was insufficient information provided in the published report and one secondary outcome noted in the trial registry was not reported, suggesting selection and reporting of outcomes on the basis of the results.

D4.2.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with post viral arthropathies are presented in Table D-20.

Table D-21 Outcomes considered by the NTWC to be critical or important for decision-making: Post viral arthropathies

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*de Oliveira 2019*
Pain	Visual analogue scale (VAS) (or SF-36 – Bodily pain; or PROMIS – Pain interference)	Critical	Yes	✓
Quality of life, physical	SF-12 – Physical Component score	Critical	Yes	✓
Fatigue	Visual analogue scale (VAS) (or PROMIS – Fatigue)	Critical	No	?
Global physical functioning	PROMIS – Physical function	Critical	No	?
Global assessment	No measures reported in eligible studies	Critical	No	?
Peripheral joints and entheses	No eligible measures reported in eligible studies	Important	No	?
Acute-phase reactant	No measures reported in eligible studies	Important	No	?

Abbreviations: SF-36; 36-item short form survey; PROMIS, Patient Reported Outcome Measurement Information System; SF-12; 12-item short form survey; VAS, visual analogue scale

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results
? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One study (de Oliveira 2019) comparing Pilates with no intervention (delivered as an adjunct to standard medical care) in people with chronic Chikungunya fever was eligible for this comparison and contributed data relevant to two of the seven critical or important outcomes. There were no studies awaiting classification or ongoing that compared Pilates with no intervention in people with post-viral arthropathies that could have contributed data to these outcomes (see Appendix C6).

Results for all outcomes were judged to have some concerns, related to knowledge of the intervention received and missing data which may (or may not) overestimate the size of the effect.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and it was not judged to be at high risk of bias.

Pain

One trial (42 participants) measured pain intensity using a visual analogue scale (VAS) at the end of treatment (12 weeks) (de Oliveira 2019). The VAS is a subjective assessment of pain, reported by participants and measured on a continuous scale (cm) from 0 (no pain) to 10 (worst imaginable pain). Higher values indicate worse pain, with a 2-point reduction considered by the study authors to indicate the minimal clinically important difference (91).

The results show a benefit in favour of Pilates (MD –3.4; 95% CI –4.85, –1.95; $p < 0.00001$) *(GRADE: very low)*. A significant difference in analgesic use between the two groups at baseline (72.7% in the Pilates group versus 100% in the control group; $p = 0.02$)) creates concerns with this result. It is conceivable that these differences in analgesic use are related to baseline pain (although no differences in baseline VAS were noted), that make it difficult to assess the direction of any potential bias is the observed improvement in pain. The authors do not report analgesic use at the end of treatment.

Quality of life - physical

One trial (42 participants) measured quality of life using the SF-12 (12-item short form survey) at the end-of-treatment (12 weeks) (de Oliveira 2019). The SF-12 is a shortened version of the SF-36 that assesses quality of life across eight domains. Scores were summarised into two composite scores (physical and mental health) reported on a range from 0 to 100, with a population mean of 50 and standard deviation of 10. Higher scores represent improved quality of life. The MCID proposed by the authors was an increase of 3.29 points (92).

The results showed a benefit in favour of Pilates (MD –11.0, 95% CI –15.35, –6.65; $p < 0.00001$) *(GRADE: low)*.

Comparison 2 (vs other)

There were no studies identified comparing Pilates with ‘other’ interventions in people with post viral arthropathies.

D4.3 Spondyloarthritis

D4.3.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-22. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-22 Overview of PICO criteria of included studies: Spondyloarthropathies

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Altan 2012 (93)	RCT	Ankylosing spondylitis	Pilates exercises	Control (usual activities)	Standard medical care	Functional capacity Disease activity Spinal mobility Chest expansion Quality of life
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: RCT, randomised controlled trial

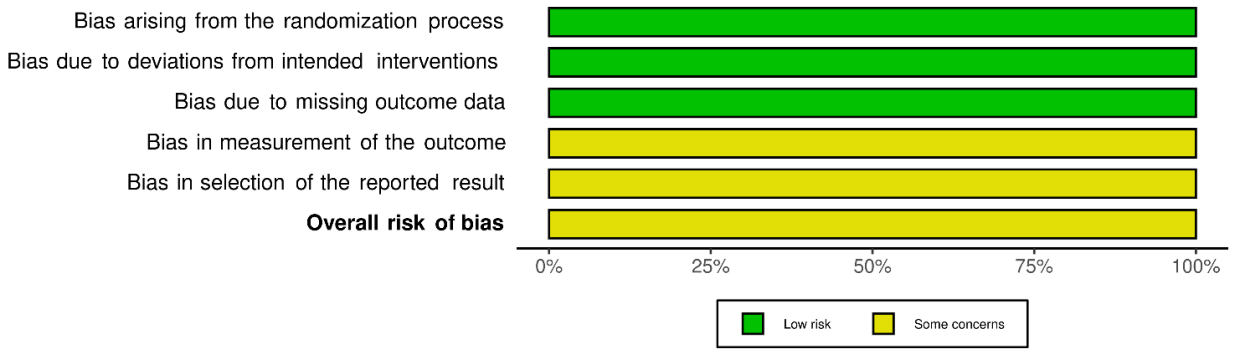
*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D4.3.2 Risk of bias summary

The risk of bias for each item in the included studies for spondyloarthropathies is described below and shown graphically in Figure D.13. Details are provided in Appendix E1.

Figure D.13 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Spondyloarthritis



Bias arising from the randomisation process

One study (Altan 2012) was assessed to be at low risk of bias for this domain. The randomisation process was well described, and baseline characteristics were comparable between the intervention and control groups.

Bias due to deviations from intended interventions (effect of assignment to intervention [ITT])

Altan 2012 was assessed to have low risk of bias in this domain. There was a lack of blinding of participants and research personnel which was considered reasonable given the nature of the intervention. The only reported deviations from the intended intervention were non-completion by some participants, which was in line with what would be expected in routine clinical practice. Two participants were excluded from the final analysis due to discontinuation, which did not raise concern.

Bias due to missing outcome data

Altan 2012 was assessed at low risk of bias for this domain as data was available for nearly all randomised participants.

Bias in measurement of the outcome

Altan 2012 was assessed to have some concerns in this domain due to the self-reported outcome by non-blinded participants. It was considered that, given the nature of the interventions, a lack of blinding could have influenced reporting of the result.

Bias in selection of the reported result

Altan 2012 was assessed to have some concerns for this domain. Three of the four Bath Ankylosing Spondylitis Indices were used to report functional capacity, disease activity and spinal mobility. The fourth index, the Bath Ankylosing Spondylitis Global Score, which assesses well-being was not reported, and it is considered possible that this omission was intentional. There was no pre-specified statistical analysis plan available to confirm whether this measure was collected during the study.

D4.3.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with spondyloarthropathies are presented in Table D-23.

Table D-23 Outcomes considered by the NTWC to be critical or important for decision-making: Spondyloarthropathies

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Altan 2012*
Global disease assessment	Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)	Critical	Yes	✓
Physical function	Bath Ankylosing Spondylitis Functional Index (BASFI)	Critical	Yes	✓
Quality of life	Ankylosing Spondylitis Quality of Life	Critical	Yes	✓
Pain	Visual analogue scale	Critical	No	?
Fatigue	BASDAI - Fatigue	Important	No	?
Spinal mobility	Bath Ankylosing Spondylitis Metrology Index (BASMI)	Important	Yes	✓
Symptoms of peripheral joints and entheses	Number of swollen joints	Important	No	?

Abbreviations: BASDI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; BASMI, Bath Ankylosing Spondylitis Metrology Index

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One study (Altan 2012) comparing Pilates to control (no intervention) in people with ankylosing spondylitis was eligible for this comparison and contributed data to four of the seven critical or important outcomes. There were three additional studies awaiting classification (two published in a language other than English) that compared Pilates with no intervention in people with ankylosing spondylitis (total 82 participants) that could have contributed data to all four outcomes (see Appendix C6).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and was not judged to be at high risk of bias.

Global assessment

One trial (55 participants) measured global disease assessment using the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) at the end of treatment (12 weeks) (Altan 2012). The results show an effect in favour of Pilates (MD –1.0; 95% CI –1.98, –0.02; $p = 0.05$) *(GRADE: low)*..

The BASDAI is considered the gold standard for measuring and following disease activity in patients with ankylosing spondylitis. It consists of six 10-cm visual analogue scales to measure severity of fatigue, spinal and peripheral joint pain, localised tenderness, and morning stiffness (94). Final scores range from 0 to 10, with a higher score indicating worse function (94). It is estimated that the minimum clinically important improvement in BASDAI score is 0.7 points, increasing to 1.1 in patients with active disease (i.e., a baseline BASDAI score of 4 or more) (95).

Global physical functioning

One trial (55 participants) assessed physical functioning using the Bath Ankylosing Spondylitis Functional Index (BASFI) at the end of treatment (12 weeks) (Altan 2012). The results showed no difference (MD –0.60; 95% CI –1.48, 0.28; $p = 0.18$) *(GRADE: low)*..

The BASFI is a set of 10 questions designed to measure the degree of functional limitation for patients with ankylosing spondylitis. The first 8 questions relate to ability to complete everyday tasks, and the last 2 questions assess the patient's ability to cope with everyday life. The final score ranges from 0 to 10, with higher scores indicating a higher degree of functional limitation (94). It is estimated that the minimum clinically important improvement in BASFI score is 0.4 points, increasing to 0.6 in patients with active disease (i.e., a baseline BASDAI score of 4 or more) (95).

Quality of life

One trial (55 participants) assessed quality of life using the Ankylosing Spondylitis Quality of Life (ASQoL) tool at the end of treatment (12 weeks) (Altan 2012). The results show no difference in quality of life between the intervention groups (MD 0.0; 95% CI –2.57, 2.57; $p = 1.0$) *(GRADE: low)*..

The ASQoL consists of 18 yes or no questions, with final scores ranging from 0 to 19. A higher score indicates a poorer quality of life (96). It is estimated that the minimum clinically important difference in ASQoL score is 3 points (97).

Spinal mobility

One trial (55 participants) reported spinal mobility using the Bath Ankylosing Spondylitis Metrology Index (BASMI) at the end of treatment (12 weeks) (Altan 2012). The results show no significant difference between treatment groups (MD –0.30; 95% CI –1.28, 0.68; *p* = 0.55) *(GRADE: low)*..

The BASMI is composed of five mobility tests that includes: cervical rotation, tragus to wall distance, lumbar side flexion, modifies Schober's and intermalleolar distance. The scale ranges from 0 to 10, where a higher score indicates more severe mobility limitations (94). Estimates for a minimally clinically important difference for BASMI is yet to be defined (98).

Comparison 2 (vs other)

The were no studies identified that compared Pilates with 'other' interventions in people with spondyloarthropathies.

D4.4 Spinal deformities

D4.4.1 List of studies

An overview of the PICO criteria of included studies is provided in Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-24 Overview of PICO criteria of included studies: Spinal deformities

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Alves de Araujo 2010 (99, 100)	RCT	Scoliosis (18 to 25 years)	Pilates exercises	Control (no intervention)	None specified	Flexibility Pain Deformity progression
Junges 2012 (101-104)	Quasi-RCT	Hyperkyphosis (women, > 45 years)	Pilates exercises	Control (usual activities)	None specified	Aesthetics Body composition Deformity progression Range of Motion
Pilates versus 'other' intervention*						
Lee 2016b (105)	Quasi-RCT	Forward head (women, 20 to 39 years)	Pilates exercises (equipment)	Combined exercises	None	Range of Motion Function/disability Pain
Navega 2016 (106, 107)	Quasi-RCT	Hyperkyphosis	Pilates (mat)	Active control (education)	None specified	Balance Thoracic kyphosis
Kudchadhar 2019 (108, 109)	RCT	Hyperlordosis (18 to 40 years)	Pilates exercises (mat)	Egoscue OR Lumbar stabilisation exercises	Stretching	Lumbar lordosis Pelvic tilt Exercise tolerance
Kim 2016 (110)	Quasi-RCT	Scoliosis	Pilates exercises	Schroth exercises	None specified	Deformity progression

Abbreviations: RCT, randomised controlled trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D4.4.2 Risk of bias summary

The risk of bias for each item in the included studies for spinal deformities is described below and shown graphically in Figure D.14. Details are provided in Appendix E1.

Bias arising from the randomisation process

All six studies (Alves de Araujo 2010, Junges 2012, Kim 2016, Kuchadkar 2019, Lee 2016b, Navega 2016) were judged to have some concerns for this domain as they did not provide information regarding allocation concealment. Four studies (Junges 2012, Kim 2016, Lee 2016b, Navega 2016) also did not provide specific information about generation of the randomisation sequence, however there were not notable differences between the treatment groups in terms of their baseline characteristics. There were some differences in

baseline characteristics in one study (Kuchadkar 2019), but these were considered likely compatible with chance and not related to issues with the randomisation process.

Bias due to deviations from the intended intervention

One study was judged to be at low risk (Kuchadkar 2019) and three studies (Junges 2012, Lee 2016b, Navega 2016) had some concerns raised for this domain. As all participants received the allocated intervention, and any deviations were considered to occur outside the trial context. Two studies were judged to be at high risk of bias related to missing information regarding the number of participants randomised to the intervention (Alves de Araujo 2010, Kim 2016).

Bias due to missing outcome data

One study was assessed to be at low risk and one study had some concerns raised as outcome data were available for nearly all participants (Kuchadkar 2019) or any missingness of the data was considered to not likely to affect the result (Lee 2016b). In fours studies there was no information to suggest that outcome data was available for all, or nearly all participants and there was high suspicion that missing outcome data could affect the result (Alves de Araujo 2010, Junges 2012, Kim 2016, Navega 2016,).

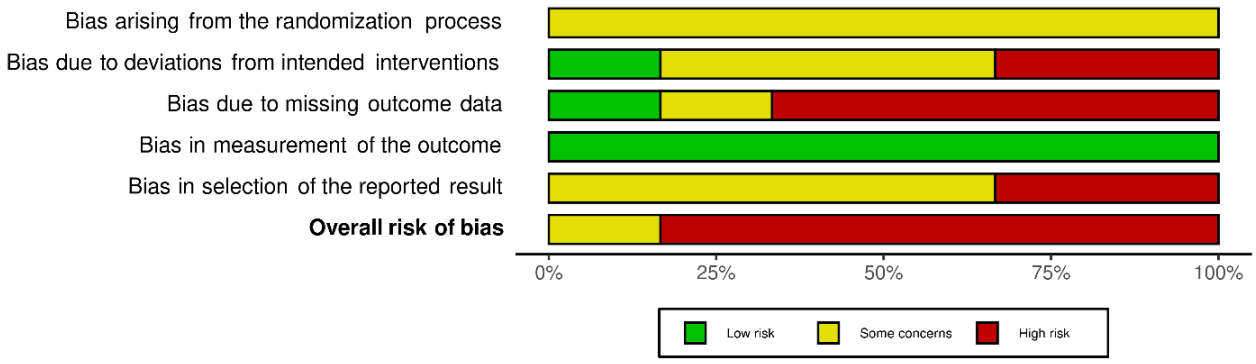
Bias in the measurement of the outcome

All six studies (Alves de Araujo 2010, Junges 2012, Kim 2016, Kuchadkar 2019, Lee 2016b, Navega 2016) were assessed to be at low risk of bias for this domain as the method of measuring the outcome was appropriate and not likely that the observer-reported outcomes were influenced by knowledge of the intervention received (Kim 2016, Lee 2016b, Navega 2016, Junges 2012, Kuchadkar 2019, Alves de Araujo 2010).

Bias in selection of the reported result

Four studies (Alves de Araujo 2010, Junges 2012, Kuchadkar 2019, Navega 2016) had some concerns raised for this domain as the researcher's pre-specified intentions were not available but were sufficiently described and there was no indication of selection of the reported measures on the basis of the results. Two studies were at high risk of bias, as insufficient information available suggesting selective reporting of results (Kim 2016, Lee 2016b).

Figure D.14 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Spinal deformities



D4.4.3 Effect of intervention (scoliosis)

Outcomes considered by the NTWC to be critical or important for decision-making in people with scoliosis are listed in Table D-25.

Table D-25 Outcomes considered by the NTWC to be critical or important for decision-making: Spinal Deformities-Scoliosis

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Alves de Araujo 2010*
Pain	0-10 numeric rating scale	Critical	Yes	✓
Disability	Oswestry Disability Index	Critical	No	?
Quality of life	SF-36 (or SRS-22/24/30)	Critical	No	?
Psychological wellbeing	No measures reported in eligible studies	Important	No	?
Flexibility/ROM	No eligible measures reported	Important	No	--
Deformity progression	Degree of curvature (Cobb angle)	Important	Yes	✓
Balance	No measures reported in eligible studies	Important	No	?

Abbreviations: CR, category ratio; ROM, range of motion; SF-36, 36-Item Short Form Survey; SRS, Scoliosis Research Society

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One RCT (Alves de Araujo 2010) comparing Pilates with no intervention (attention control) in people with non-structural scoliosis was eligible for this comparison and contributed data relevant to two of the five outcomes. There were no additional studies identified (awaiting classification or ongoing) that compared Pilates with no intervention people with non-structural scoliosis that could have contributed data to these outcomes (see Appendix C6).

Results for all outcomes were judged to be at high risk of bias related to missing information regarding trial withdrawals or discontinuations and potential missing outcome data.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data its removal would leave no results.

Pain

One trial (31 participants) measured pain using a numeric rating scale (Borg CR10) at the end of treatment (12 weeks) (Alves de Araujo 2010). The Borg CR10 is a general intensity scale based on a category-ratio scale anchored from 0 (no pain) to 10 (extreme pain). The MCID for pain in people with non-structural scoliosis is unknown but a reduction of 2 points is reported to be clinically important in people with diabetic neuropathy, postherpetic neuralgia, chronic low back pain, fibromyalgia and osteoarthritis (91).

The results show an effect favouring Pilates when compared to the control group (MD –2.00; 95% CI –3.80, –0.20; $p = 0.03$) *(GRADE: very low)*.

Deformity progression

One trial (31 participants) measured deformity progression using degree of curvature (Cobb angle) at the end of treatment (12 weeks) (Alves de Araujo 2010). The Cobb angle is used to assess the status of scoliosis with success of conservative treatment defined by a curve progression of 5 degrees or less in the Cobb angle at the end of treatment.

The results show an effect in favour of Pilates when compared to the control group (MD –2.10; 95% CI –4.13, –0.07; $p = 0.04$) *(GRADE: very low)*. A change of 2.10 degrees was not considered clinically important.

Comparison 2 (vs other) – Scoliosis

One quasi-RCT (Kim 2016) comparing Pilates with one 'other' intervention in people with idiopathic scoliosis was eligible for this comparison and contributed data to one outcome.

Available data from this study are presented in Appendix F2 Supplementary outcome data.

D4.4.4 Effect of intervention (hyperkyphosis, hyperlordosis, forward head)

Outcomes considered by the NTWC to be critical or important for decision-making in people with spinal deformities including hyperkyphosis, hyperlordosis and forward head are listed in Table D-26.

Table D-26 Outcomes considered by the NTWC to be critical or important for decision-making: Spinal Deformities (hyperkyphosis, hyperlordosis, forward head posture)

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Junges 2012*
All conditions				
Pain	0-10 numerical pain scale	Critical	No	--
Disability	Neck disability index, Oswestry Disability Index (or other)	Critical	No	--
Quality of life	SF-36	Critical	No	--
Global perceived effect	No measures reported in eligible studies	Important	No	--
Work status	No eligible measures reported	Important	No	--
Condition specific				
Hyperkyphosis				
Deformity progression	Degree of curvature (Cobb angle) (or craniovertebral angle)	Important	Yes	✓
Flexibility/ROM	No eligible measures reported	Important	No	--
Hyperlordosis				
Degree of lumbar lordosis	Index of lordosis	Important	No	NA
Anterior pelvic tilt	Pelvic inclinometer	Important	No	NA
Forward head posture				
Global physical functioning	SF-36 - physical function	Critical	No	NA
Deformity progression	Craniovertebral angle	Important	No	NA

Abbreviations: NA, not applicable; ROM, Range of motion; SF-36, 36-Item Short Form Survey

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One quasi-RCT (Junges 2012) comparing Pilates with no intervention (usual activities) in people with hyperkyphosis was eligible for this comparison and contributed data relevant to one outcome.

There were two studies identified (awaiting classification or ongoing) that compared Pilates with no intervention people with hyperlordosis that could have contributed data to the critical or important outcomes specific to people with hyperlordosis (see Appendix C6).

Results were judged to be at high risk of bias due to missingness of outcome data, which may overestimate the size of the effect. No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no results.

Deformity progression

One study (41 participants) reported deformity progression based on the degree of curvature (Cobb angle) at the end of treatment (30 weeks) (Junges 2012). The Cobb angle is used to assess the status of scoliosis with success of conservative treatment defined by a curve progression of 5 degrees or less in the Cobb angle at the end of treatment.

The results show no difference between the Pilates and control groups (MD –2.72; 95% CI –9.04, 3.60; $p = 0.40$) *(GRADE: very low)*. A change of 2.72 degrees was not considered clinically important.

Comparison 2 (vs other)

Three studies (Kudchadkar 2019, Lee 2016b, Navega 2016) comparing Pilates with 'other' interventions in people with spinal deformities (including forward head posture, hyperkyphosis, hyperlordosis) were eligible for this comparison and contributed data to three outcomes.

Available data from these studies are presented in Appendix F2 Supplementary outcome data.

D4.5 Osteoporosis

D4.5.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-27. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-27 Overview of PICO criteria of included studies: Osteoporosis

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Angin 2015 (111)	Quasi-RCT	Osteoporosis (without fracture, post menopause)	Pilates exercises (TheraBand, balls)	Control (no intervention)	None specified	Bone mineral density Pain Physical performance Activities of daily living Jobs around the house Mobility Social function General health status Mental function
Oksuz 2014 (112-114)	RCT	Osteoporosis (without fracture, post menopause)	Pilates exercises (not specified)	Control (usual activities)	None specified	Kinesiophobia Falls risk Pain Disability Balance Strength and endurance Flexibility Disease specific-QoL Emotional wellbeing Functional impairment Life Satisfaction
Pilates versus 'other' intervention*						
Kucukcakir 2013 (115)	Quasi-RCT	Osteoporosis (without fracture, post menopause)	Pilates exercises (mat, ball)	Home exercise (Unbalanced support surface exercise)	None specified	Mobility Pain Physical performance Disease specific-QoL QoL-global

Abbreviations: RCT, randomised controlled trial; QoL, quality of life

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D4.5.2 Risk of bias summary

The risk of bias for each item in the included studies for osteoporosis is described below and shown graphically in Figure D.15. Details are provided in Appendix E.

Bias arising from the randomisation process

Two studies (Kucukcakir 2015 and Oksuz 2014) were assessed to have some concerns in this domain due to lack of information provided regarding the allocation concealment process. Only one study (Oksuz 2014) reported on the method of generating the randomisation sequence. Demographic and clinical characteristics

appeared comparable between groups at baseline, however these were only presented for participants who finished the studies leading to some concerns.

One study (Angin 2015) was assessed at high risk of bias due to likely skewed data and significant differences across multiple domains that measure quality of life, suggesting the control group had better quality of life at baseline. This study did not describe the method of random sequence generation or allocation concealment.

Bias due to deviations from the intended intervention

One study (Angin 2015) was assessed to have some concerns for this domain due to the participants who discontinued the study, all in the control group. As this was an inactive control this was considered unlikely related to the trial context. Two studies (Kucukcakir 2015, Oksuz 2014) were at high risk of bias due to discontinuations that were likely or possibly related to the trial context. Three participants in Kucukcakir 2015 discontinued due to unwillingness to participate in their allocated intervention compared to none in the control group. Seven discontinuations occurred in Oksuz 2014 which were only reported in the abstract of the study and not in the main trial publication. No information on these participants was available.

Bias due to missing outcome data

One study (Angin 2015) was assessed at low risk of bias for this domain as outcome data was available for most participants. Three discontinuations in the control group were not considered likely to be related to the true outcome value given the duration of the trial and the inactive control. Two studies (Kucukcakir 2015, Oksuz 2014) had concerns raised due to the proportion of participants with missing outcome data that could potentially affected the true value of the outcome.

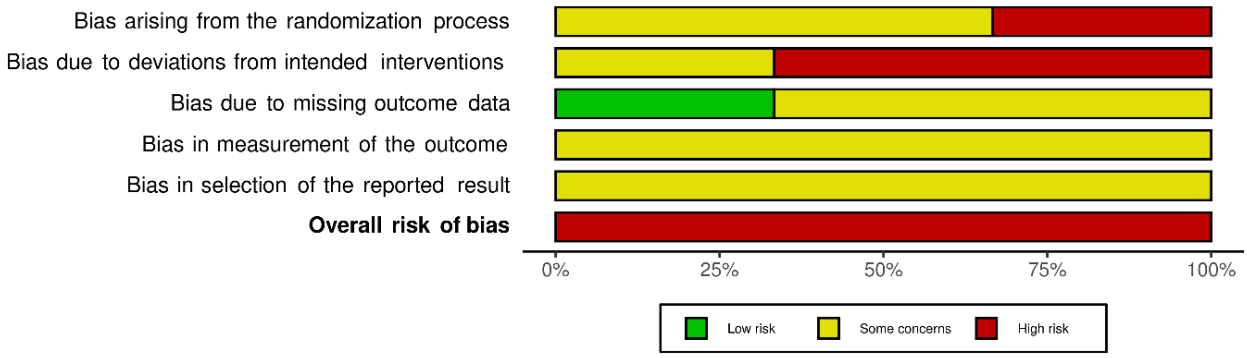
Bias in the measurement of the outcome

All studies were assessed to have some concerns in this domain due to a lack of blinding of participants and the self-reported nature of the outcome. It was considered plausible that participants who were aware of their intervention status would differentially report their outcomes.

Bias in the selection of the reported result

All studies were assessed to have some concerns for this domain as it was considered possible that the outcomes reported could have plausibly been selected from a range of eligible outcome domains.

Figure D.15 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Osteoporosis



D4.5.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with osteoporosis are presented in Table D-28.

Table D-28 Outcomes considered by the NTWC to be critical or important for decision-making: Osteoporosis

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Oksuz 2014,*	*Angin 2015*
Functional mobility	6-minute walk test	Critical	Yes	--	✓
Pain	Short-Form McGill Pain Questionnaire (or Visual Analogue Scale)	Critical	Yes	✓	✓
Quality of life	SF-36 Global Score (or QUALEFFO-41)	Critical	Yes	✓	✓
Bone mineral density	Bone mineral density (g/cm2) (or T-score)	Important	Yes	--	✓
Balance	Berg Balance Test	Important	Yes	✓	✓
Falls	Incidence/rate	Important	No	--	--
Global assessment	Outcomes and Assessment Information Set (OASIS)	Important	No	--	--

Abbreviations: SF-36; 36-item short form survey; QUALEFFO, Quality of Life Questionnaire of the European Foundation for Osteoporosis

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Two studies (Oksuz 2014, Angin 2015) comparing Pilates with no intervention (or usual activities) in people with osteoporosis were eligible for this comparison and contributed data to five of the seven outcomes. There were no studies awaiting classification or ongoing that compared Pilates with no intervention in people with osteoporosis that could have contributed data to these outcomes (see Appendix C6).

Results for all outcomes were judged to be at high risk of bias related to randomisation (with skewed baseline data) and missing information (high dropout and missing data) that may over (or under) state the effect.

Functional mobility

One study (41 participants) reported functional mobility using the 6-minute walk test at the end of treatment (24 weeks) (Angin 2015). The 6-minute walk test is used to assess aerobic capacity and endurance, with the distance covered over the 6-minute period used to assess changes in performance capacity. An MCID for the 6-minute walk test in women with osteoporosis had not been reported, with the expected walking distance in women aged between 60 to 64 years to around 500 to 600 metres (116). Among a range of conditions, a change of 14.0 to 30.5 meters may be clinically important across multiple patient groups (117).

The results showed an effect favouring Pilates that did not reach statistical significance (MD –53.40; 95% CI –110.61, 3.81, p = 0.07) (likely due to small sample size and wide confidence intervals) *(GRADE: very low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Pain

Two studies (81 participants) reported pain measured using a visual analogue scale (VAS) at the end of treatment (6- and 24 weeks, respectively) (Oksuz 2014, Angin 2015). The VAS scale ranged from 0 to 10, with higher values indicating more severe pain. Both studies reported the mean change from baseline for pain (at rest) and pain (while moving). A reduction of 2 points (or 30%) on the Numeric Pain Scale is reported to be clinically important in people with diabetic neuropathy, postherpetic neuralgia, chronic low back pain, fibromyalgia and osteoarthritis (91).

Pooled results showed an effect in favour of Pilates for pain (at rest) (MD –2.29; 95% CI –2.88, –1.70; $p < 0.001$; $I^2 = 0\%$) *(GRADE: not assessed)* and for pain (while moving) (MD –3.25; 95% CI –6.26, –0.23 $p = 0.04$; $I^2 = 97\%$) *(GRADE: very low)*.

The observed heterogeneity for pain (while moving) may be related to duration of treatment, with the effect observed at 24 weeks (MD –4.78; 95% CI –5.42, –4.14; $p < 0.001$) better than that observed after 6 weeks (MD –1.70; 95% CI –2.43, –0.97; $p < 0.001$).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as removal of both studies would leave no result.

Quality of life

Two studies (81 participants) reported quality of life measured with the Quality-of-Life Questionnaire of the European Foundation for Osteoporosis QUALEFFO-41 at the end of treatment (6- and 24- weeks respectively) (Oksuz 2014, Angin 2015). QUALEFFO-41 measures changes in everyday functioning, well-being and health related quality of life in patients with vertebral fractures. The questionnaire covers five domains: pain, physical functioning (activities of daily living, jobs around the house, mobility), leisure/social activities, general health perceptions and mental function that are summarised into a total score ranging from 41 (no problem) to 205 (severe problems) (or less when some questions are not answered). Scores are transformed to a scale of 0 (worse) to 100 (best). No MCID has been established for the QUALEFFO-41 (118).

One study (Oksuz 2014) reported change from baseline scores for the total QUALEFFO-41 score, which showed an effect in favour of Pilates (MD –6.21; 95% CI –7.97, –4.45; $p < 0.00001$) *(GRADE: low)*. Change from baseline scores for individual domains were also provided (see Appendix F2 Supplementary outcome data).

One study (Angin 2015) presented individual scores for each question, but a total score was not provided. The authors reported an effect in favour of Pilates compared to control for each of the seven questions when comparing change from baseline score (see Appendix F2 Supplementary outcome data). The effect size is uncertain as baseline data were skewed (119).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as removal of both studies would leave no result.

Bone mineral density

One study (41 participants) assessed bone mineral density and reported a T-score at the end of treatment (24 weeks) (Angin 2015). BMD was measured via dual energy X-ray absorptiometry (DXA). The standard deviation between the participant's BMD and that of healthy young adults is the T-score. A negative T-score

indicates the bone is weaker than normal. According to the WHO, a T-score less than –1 indicates low BMD (osteopenia) and a T-score less than –2.5 is diagnostic of osteoporosis (120).

The results suggest an effect in favour of Pilates (MD 0.34; 95% CI 0.19, 0.49; *p* = 0.003) *(GRADE: very low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Balance

One study (40 participants) reported balance measured the Berg Balance test using change from baseline at the end of treatment (6 weeks) (Oksuz 2014). In most of the 14-items, the subject is asked to maintain a given position for a specific time, with each item consisting of a five-point ordinal scale ranging from 0 to 4. Total scores range from 0 (low balance stability) to 56 (high balance stability). In elderly people, a score of less than 45 indicates individuals may be at greater risk of falling (64), however the test may be subject to ceiling effects in community dwelling older adults (121). A minimum detectable clinical difference in older adults is proposed to be 6.5 points (122).

The results showed a minimal to no effect in favour of Pilates (MD –1.70; 95% CI –2.26, –1.14; *p* = 0.00001) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Comparison 2 (vs other)

One study (Kucukcakir 2015) comparing Pilates to one 'other' comparator in people with osteoporosis was eligible for this comparison and contributed data relevant to four of the critical or important outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

D4.6 Chronic widespread pain (fibromyalgia)

D4.6.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-29. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-29 Overview of PICO criteria of included studies: Fibromyalgia

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
No studies found.						
Pilates versus 'other' intervention*						
Altan 2009 (123)	RCT	Fibromyalgia (women)	Pilates exercises (band, ball)	Home exercise (relaxation-stretching)	None specified	Functional capacity Pain Tenderness QoL-global
de Medeiros 2020 (124, 125)	RCT	Fibromyalgia (women)	Pilates exercises (mat, Swiss ball)	Aqua aerobics	None specified	Pain Psychosocial Sleep quality QoL-global QoL-disease specific
Ekici 2014 (126, 127)	Quasi-RCT	Fibromyalgia (women)	Pilates exercises	Connective tissue massage	None specified	Anxiety Depression Fatigue Functional capacity Pain Stiffness Sleep quality

Abbreviations: RCT, randomised controlled trial; QoL, quality of life

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D4.6.2 Risk of bias summary

The risk of bias for each item in the included studies for chronic widespread pain (fibromyalgia) is described below and shown graphically in Figure D.16. Details are provided in Appendix E1.

Bias arising from the randomisation process

One study (de Medeiros 2020) was considered to be at low risk of bias for this domain as participants were randomised via an independent researcher with allocations concealed in sequentially numbered sealed opaque envelopes (de Medeiros 2020). One study (Altan 2009) randomised participants using a random number table to generate the sequence but did not provide sufficient information on the allocation concealment process, raising some concerns. In both these studies baseline characteristics were sufficiently matched, indicating no serious issues with randomisation. One study (Ekici 2017) randomised participants using alternate allocation, with two letters enclosed in sealed envelopes that guided the assigned treatment and instructor. Significant baseline differences were observed in several outcome measures that raised serious doubt about the randomisation process.

Bias due to deviations from the intended intervention

One study (Altan 2009) was judged to be at low risk of bias for this domain, as all randomised participants received the intervention, and any deviations were considered to occur outside the trial context. One study (de Medeiros 2020) had some concerns raised related to potential deviations (e.g., worsening symptoms) related to the trial context. One study (Ekici 2017) was judged to be at high risk of bias for this domain as changes from the assigned intervention were inconsistent with the trial protocol and were not balanced between treatment groups.

Bias due to missing outcome data

One study was assessed to be at low risk of bias for this domain as outcome data were available for nearly all participants (Altan 2009). One study (de Medeiros 2020) has some concerns raised due to the missingness of the data potentially influence the true value of the outcome. In one study (Ekici 2017), differences in the proportions of missing outcome data between treatment groups increased the threat of bias to high.

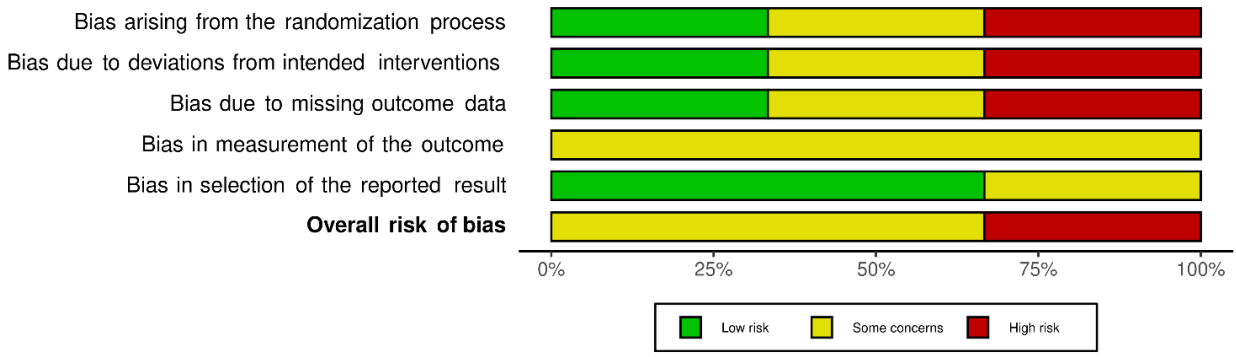
Bias in the measurement of the outcome

All studies were assessed to have some concerns for this domain as participant-reported outcomes could be influenced by knowledge of the intervention received, potentially overstating the effect in favour of the intervention received (Altan 2009, de Medeiros 2020, Ekici 2017).

Bias in selection of the reported result

Two studies reported all eligible pre-specified results for the outcome of interest and were judged to be at low risk of bias for this domain (Altan 2009, de Medeiros 2020). One study (Ekici 2017) had some concerns raised for this domain as the researchers pre-specified intention are not available but are sufficiently described and data analysis was performed accordingly (Altan 2009, Ekici 2017).

Figure D.16 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Chronic widespread pain (fibromyalgia)



D4.6.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with chronic widespread pain (fibromyalgia) are listed in Table D-30.

Table D-30 Outcomes considered by the NTWC to be critical or important for decision-making: Chronic widespread pain (fibromyalgia)

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*
Tenderness	Number of tender points	Important	No
Functional capacity	FIQ function	Critical	No
Pain	FIQ pain	Critical	No
Fatigue	Multidimensional Fatigue Inventory	Critical	No
Quality of life, disease specific	FIQ	Critical	No
Sleep quality	FIQ-morning rest	Important	No
Stiffness	FIQ-stiffness	Important	No

Abbreviations: FIQ, Fibromyalgia Impact Questionnaire

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

There were no studies identified comparing Pilates with no intervention in people with chronic widespread pain (fibromyalgia). There was one ongoing study (completed) that compared Pilates with no intervention in people with fibromyalgia (target 50 participants) that could have contributed data to the outcomes considered critical or important to this review, but the study results had not been published (see Appendix C6).

Comparison 2 (vs other)

Three RCTs (Altan 2009, de Medeiros 2020, Ekici 2014) comparing Pilates with 'other' interventions in participants with chronic widespread pain (fibromyalgia) were eligible for this comparison and contributed data relevant to six of the seven outcomes.

Available data from this study are presented in Appendix F2 Supplementary outcome data.

D4.7 Low back pain

D4.7.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-31. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-31 Overview of PICO criteria of included studies: Low back pain

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Cruz-Diaz 2015 (128)	RCT	Low back pain (> 3 months, women > 65 years)	Pilates exercises	Control (no intervention)	Conventional physiotherapy (TENS, massage, stretching)	Pain intensity Functional mobility Fear of falling
Cruz-Diaz 2016 (129)	RCT	Low back pain (> 3 months, women, 45 to 75 years)	Pilates exercises (equipment)	Control (no intervention)	Conventional physiotherapy (electrotherapy, joint mobilisation)	Pain intensity Functional impairment
Cruz-Diaz 2017 (130)	RCT	Low back pain (> 3 months, women)	Pilates exercises (mat) OR Pilates exercises (reformer)	Control (no intervention)	None specified	Pain intensity Functional disability Kinesiophobia Muscle activation
Cruz-Diaz 2018 (131)	RCT	Low back pain (> 3 months, women)	Pilates exercises (mat)	Control (education booklet)	None specified	Pain intensity Functional disability Kinesiophobia Muscle activation
da Fonseca 2009 (132)	Quasi-RCT	Low back pain (> 6 months)	Pilates exercises	Control (usual activities)	Standard medical care	Pain intensity Gait analysis
Gladwell 2006 (133)	Quasi-RCT	Low back pain (> 3 months)	Pilates exercises	Control (usual activities)	Standard medical care (analgesics)	Pain intensity Functional disability Global perceived effect Sports Functioning QoL-global Static balance Flexibility
Hasanpour-Dehkordi 2017 (134)	Quasi-RCT	Low back pain (> 3 months, men 40 to 55 years)	Pilates exercises	Control (no intervention) OR McKenzie exercises^	None specified	Pain Psychological wellbeing
Kliziene 2017 (135)	NRSI	Low back pain (> 3 months, women)	Pilates exercises	Control (no details)	None specified	Muscle strength Muscle endurance Pain intensity
Kofotolis 2016 (136)	RCT	Low back pain (> 3 months, women)	Pilates exercises (mat)	Control (usual activities) OR Standard strengthening exercises^	None specified	QoL-global Functional disability
Lopes 2014 (137, 138)	RCT	Low back pain (> 3 months)	Pilates exercises	Control (no intervention)	None specified	Pain intensity Dynamic balance

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Mazloum 2016 (139, 140)	Quasi-RCT	Low back pain (> 3 months)	Pilates exercises	Control (usual activities) OR Extension-based exercises^	None specified	Pain intensity Functional disability Lumbar ROM Lumbar curvature
Miyamoto 2011 (141, 142)	RCT	Low back pain (> 3 months)	Pilates exercises (modified)	Control (no intervention)	Education booklet	Pain intensity Functional disability Global perceived effect Kinesiophobia Function (patient specific)
Miyamoto 2016 (143-145)	RCT	Low back pain (> 3 months)	Pilates exercises (equipment) (three groups with either 1, 2 or 3 sessions per week)	Control (no intervention)	Educational booklet	Pain Functional disability Global perceived effect Kinesiophobia QoL-global Function (patient specific)
Natour 2011 (146, 147)	RCT	Low back pain (> 3 months)	Pilates exercises	Control (no intervention)	Standard medical care (NSAIDs)	Pain intensity Functional disability QoL-global Flexibility Analgesic use
Notarnicola 2014 (148)	NRSI	Low back pain (> 6 months, 30 years or older)	Pilates exercises (mat, equipment)	Control (usual activities)	None specified	Functional disability QoL-global Functional capacity
Pappas 2013 (149)	NRSI	Low back pain (> 6 weeks)	Pilates exercises (mat, Fitball)	Control (no intervention)	None specified	Functional disability Balance Flexibility Mood
Patti 2016 (150)	RCT	Low back pain (> 3 months)	Pilates exercises (mat)	Control (usual activities)	None specified	Functional disability Dynamic balance
Quinn 2011 (151)	RCT	Low back pain (> 3 months)	Pilates exercises (mat, modified)	Control (no intervention)	None specified	Pain intensity Functional disability Lumbopelvic control
Rydeard 2006 (152)	Quasi-RCT	Low back pain (> 6 weeks)	Pilates exercises (mat, Reformer)	Control (usual care)	None specified	Pain Functional disability
Valenza 2017 (153)	RCT	Low back pain (> 3 months)	Pilates exercises (mat, ball)	Control (educational advice)	None specified	Pain intensity Functional disability Lumbar ROM Flexibility Balance
Zeada 2012 (154)	Quasi-RCT	Low back pain (> 3 months, 20 to 25 years)	Pilates exercises (mat)	Control (usual care)	None specified	Muscle endurance Lumbar ROM Functional disability Urine catecholamine
Pilates versus 'other' intervention*						
Albert Anand 2014 (155)	RCT	Low back pain (> 3 months)	Modified Pilates exercises	Therapeutic exercises	Educational advice	Pain Functional disability
Avila Ribeiro 2015 (156)	Quasi-RCT	Low back pain (> 3 months)	Pilates exercises	Classical kinesiotherapy exercises	None specified	Pain Functional Disability

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Bhadauria 2017 (157)	RCT	Low back pain (> 3 months)	Pilates exercises	Lumbar stabilisation exercises OR Dynamic strengthening exercises	Conventional physiotherapy (HMP & IFC)	Pain Functional disability Core muscle strength Lumbar ROM
Brooks 2012 (158-160)	RCT	Low back pain (> 3 months)	Pilates exercises	Aerobic exercise (stationary cycling)	None specified	Pain Functional disability Pain catastrophising Kinesiophobia
Devasahayam 2016 (161)	RCT	Low back pain, with lower limb musculoskeletal injury (unilateral)	Pilates exercises	Conventional physical therapy	None specified	Pain intensity Global perceived effect Function (patient specific)
Donzelli 2006 (162)	Quasi-RCT	Low back pain (> 3 months)	Pilates exercises	Back school method	None specified	Pain intensity Functional disability
Dsa 2014 (163)	RCT	Low back pain (> 3 months)	Pilates exercises (mat, modified)	Conventional core exercises	HMP before intervention	Pain intensity Functional disability
Gonzalez-Galvez 2019 (164)	Quasi-RCT	Low back pain (> 12 months, adolescents 14 to 16 years)	Pilates exercises	Physical education sessions	None specified	Flexibility Muscle endurance Body Mass Index
Mostagi 2015 (165)	RCT	Low back pain (> 3 months)	Pilates exercises	General exercise	None specified	Pain intensity Functional disability Flexibility Muscle endurance
Rajpal 2008 (166)	Quasi-RCT	Low back pain (> 3 months, women 20 to 30 years)	Pilates exercises	McKenzie method	None specified	Pain intensity Functional disability Postural tilt Core muscle strength
Silva 2018 (167)	Quasi-RCT	Low back pain (> 3 months, 30 to 60 years)	Pilates exercises	Conventional physiotherapy	None specified	Pain Functional disability
Wajswelner 2011 (168, 169)	RCT	Low back pain (> 3 months)	Pilates exercises (individualised) (Reformer, trapeze)	Conventional physiotherapy (individualised)	None specified	Pain intensity Functional disability Pain confidence Function (patient specific) Global perceived effect QoL-global

Abbreviations: HMP, hot moist pack; IFC, inferential current; QoL, quality of life; RCT, randomised controlled trial; ROM, range of motion; TENS, transcutaneous electrical nerve simulation

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

^ Study included three groups. The inactive control is considered in the evidence synthesis.

D4.7.2 Risk of bias summary

Randomised controlled trials

The risk of bias for each item in the included RCTs for low back pain are described below and shown graphically Figure D.17. Details are provided in Appendix E1.

Bias arising from the randomisation process

Sixteen studies (Anand 2014, Brooks 2012, Cruz Diaz 2016, Cruz Diaz 2017, Cruz Diaz 2018, Devasahayam 2016, Kofotolis 2016, Lopes 2017, Miyamoto 2016, Mostagi 2015, Natour 2011, Patti 2016, Quinn 2011, Rydeard 2006, Valenza 2017, Wajswelner 2011) provided sufficient information on the randomisation process and were at low risk of bias for this domain.

Seven studies (Cruz Diaz 2015, Da Fonesca 2009, Dsa 2014, Gladwell 2006, Mazloum 2016, Rajpal 2008, Silva 2018) had some concerns raised due to missing information about methods to generate the randomisation sequence or concealing treatment allocation; but there were no differences in demographics or outcome measures observed at baseline. Concerns in one study (Miyamoto 2011) were due to differences between groups with regards to the duration of symptoms at baseline, suggesting issues with the randomisation process.

Six studies were judged to be at high risk for this domain, because the authors did not provide information on the method of randomisation or allocation concealment and either did not provide useful baseline information (Avila Riberio 2015, Gonzalez-Galvez 2019, Zaeda 2012) or differences between groups with regards to pre-treatment outcome measures suggested an issue with the randomisation process (Bhadauria 2017, Donzelli 2006, Hasanpour-Dehkordi 2017).

Bias due to deviations from the intended intervention

Sixteen studies were judged to be at low risk of bias for this domain; any discontinuations from intended interventions were judged to be unrelated to the trial context (Avila Riberio 2015, Bhadauria 2017, Cruz-Diaz 2015, Cruz-Diaz 2018, Gonzalez-Galvez 2019, Kofotolis 2016, Lopes 2017, Mazloum 2016, Miyamoto 2011, Miyamoto 2016, Natour 2011, Patti 2016, Quinn 2011, Rajpal 2008, Valenza 2017, Wajswelner 2011).

Seven studies had some concerns raised for this domain as there were deviations from the treatment allocation possible related to the trial context, but their impact on the outcome was expected to be slight (Brooks 2012, Cruz Diaz 2016, Cruz-Diaz 2007, Devasahayam 2016, Dsa 2014, Rydeard 2006, Zaeda 2012).

Seven studies were judged to be at high risk of bias for this domain because the authors did not provide any information to allow an assessment about the number or potential impact of participants deviating from the assigned intervention (Anand 2014, Brooks 2012, Cruz-Diaz 2017, Da Fonesca 2009, Donzelli 2006, Gladwell 2006, Hasanpour-Dehkordi 2017, Mostagi 2015, Silva 2018).

Bias due to missing outcome data

Sixteen studies were assessed to be at low risk of bias for this domain as outcome data were available for all (or nearly all) participants (Avila Riberio 2015, Cruz Diaz 2015, Cruz-Diaz 2018, Donzelli 2006, Gonzalez-Galvez 2019, Lopes 2017, Miyamoto 2011, Miyamoto 2016, Natour 2011, Patti 2016, Quinn 2011, Rajpal 2008, Rydeard 2006, Valenza 2017, Wajswelner 2011, Zaeda 2012).

Five studies had some concerns raised as outcome data was not available for all (or nearly all) participants, but missingness of the data was balanced across intervention groups and considered not likely to substantially impact the results (Cruz-Diaz 2016, Cruz-Diaz 2017, Mazloum 2016, Silva 2018).

Seven studies (Anand 2014, Bhadauria 2017, Da Fonesca 2009, Devasahayam 2016, Dsa 2014, Hasanpour-Dehkordi 2017, Kofotolis 2016) provided no information regarding the extent of missing outcome data, therefore were judged to be at high risk of bias for this domain. A further three studies (Brooks 2012, Gladwell 2006, Mostagi 2015) had missing data that was not balanced between treatment groups, and the analysis was unlikely to have adjusted for any impact the missingness of the outcome could have on the results.

Bias in the measurement of the outcome

Four studies were assessed to be at low risk for this domain as the testing measures used to assess the outcomes were considered appropriate, the same methods were used between the two groups and outcome assessors were also blinded to the intervention allocation (Gonzalez-Galvez 2019, Lopes 2017, Rajpal 2008, Zaeda 2012).

The remaining twenty-six studies were assessed to have some concerns for this domain because the participant-reported outcomes could be influenced by knowledge of the intervention received. There were no reasons to suspect that patient-reported outcomes were substantially influenced by their treatment experience (Anand 2014, Avila Riberio 2015, Bhadauria 2017, Brooks 2012, Cruz-Diaz 2015, Cruz-Diaz 2016, Cruz-Diaz 2017, Cruz-Diaz 2018, Da Fonesca 2009, Devasahayam 2016, Donzelli 2006, Dsa 2014, Gladwell 2006, Hasanpour-Dehkordi 2017, Kofotolis 2016, Mazloum 2016, Miyamoto 2011, Miyamoto 2016, Mostagi 2015, Natour 2011, Patti 2016, Quinn 2011, Rydeard 2006, Silva 2018, Valenza 2017, Wajswelner 2012).

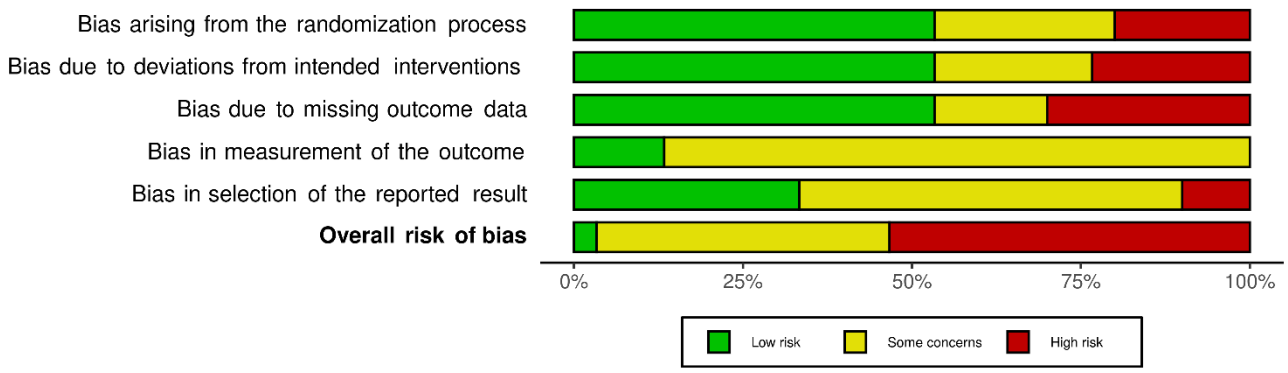
Bias in selection of the reported result

Ten studies reported all eligible pre-specified results and were judged to be of low risk for this domain (Brooks 2012, Cruz-Diaz 2015, Gonzalez-Galvez 2019, Kofotolis 2016, Lopes 2017, Miyamoto 2011, Miyamoto 2016, Mostagi 2015, Rydeard 2006, Wajswelner 2011).

Seventeen studies did not report analysis intentions in sufficient detail to enable an assessment and were judged to have some concerns (AvilaRiberio 2015, Bhadauria 2017, Cruz-Diaz 2016, Da Fonesca 2009, Devasahayam 2016, Donzelli 2006, Dsa 2017, Gladwell 2006, Hasanpour-Dehkordi 2017, Mazloum 2016, Natour 2011, Patti 2016, Quinn 2011, Rajpal 2008, Silva 2018, Valenza 2017, Zaeda 2012).

Three studies were judged to be at high risk of bias because of incomplete reporting suggesting selective reporting of outcome results (Anand 2014, Cruz-Diaz 2017, Cruz-Diaz 2018).

Figure D.17 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Low back pain



Non-randomised studies of interventions

The risk of bias for each item in the included NRSIs for low back pain are described below and shown graphically in Figure D.18 (details are provided in Appendix E2).

Bias due to confounding

One study (Notarnicola 2014) was judged to be at low risk of bias for this domain. The authors used multiple logistic regression to evaluate the influence of gender, age, smoking habits, carrying heavy loads, sporting

activity, bone diseases, number of days of back pain and the baseline values for testing. No influence is said to have been found by the trial authors. In two studies (Kliziene 2017, Pappas 2013) there was no attempt to adjust for potential confounding factors raising some or serious concerns.

Bias of selection of participants into the study

Two studies (Kliziene 2017, Notarnicola 2014) were at low risk of bias as all eligible participants were invited to participate in the study and start of interventions coincided. In one study (Pappas 2013) the selection of participants raised serious concerns, relating to the subjective inclusion of participant in the study, based on clinical discretion.

Bias in classification of interventions

All three studies (Kliziene 2017, Notarnicola 2014, Pappas 2013) provided a clear description and definition of intervention groups prior to enrolment and were assessed to be at low risk for this domain.

Bias due to deviations from intended interventions

All three studies (Kliziene 2017, Notarnicola 2014, Pappas 2013) failed to provide any information relating to the number of participants allocated to the intervention groups, or if there were any deviations or dropouts that occurred during the prospective study.

Bias due to missing data

All three studies (Kliziene 2017, Notarnicola 2014, Pappas 2013) failed to provide any information relating to the amount of missing data.

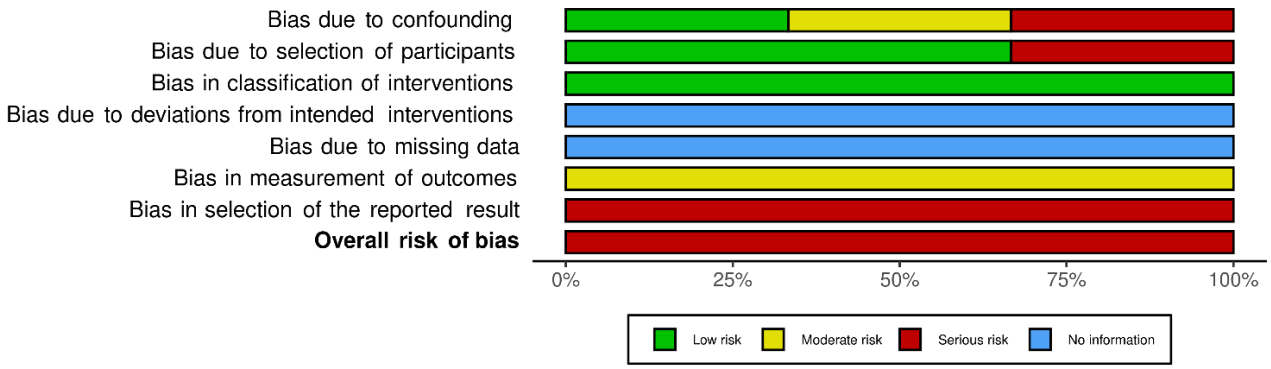
Bias in measurement of outcomes

All three studies (Kliziene 2017, Notarnicola 2014, Pappas 2013) were judged to be at moderate risk of bias for measurement of the outcomes related to the knowledge of the intervention received by study participants potentially influencing the outcome results.

Bias in selection of the reported result

All three studies (Kliziene 2017, Notarnicola 2014, Pappas 2013) were judged to be at serious risk of bias as there was high suspicion of selection of the selection of reported outcomes or analysis based on the results.

Figure D.18 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSIs – Low back pain



D4.7.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with chronic low back pain are listed in Table D-32.

Table D-32 Outcomes considered by the NTWC to be critical or important for decision-making: Low back pain

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Cruz-Diaz 2015*	*Cruz-Diaz 2016*	*Cruz-Diaz 2017*	*Cruz-Diaz 2018*	*da Fonseca 2009*	*Gladwell 2006*	*Hasanpour-Dehkordi 2017*	*Kliziene 2017*	*Kofotolis 2016*	*Lopes 2014*	*Mazloum 2016*	*Miyamoto 2011*	*Miyamoto 2016*	*Natour 2011*	*Notarnicola 2014*	*Pappas 2013*	*Patti 2016*	*Quinn 2011*	*Rydeard 2006*	*Valenza 2017*	*Zeada 2012*
Pain	McGill Pain Questionnaire (or other)	Critical	Yes	✓	✓	✓	✓	✓	✓	✓	✓^	✓	✓	✓	✓	✓	✓	✓^	✓^	--	✓	✓	✓	--
Disability	Oswestry Disability Index (or RMDQ)	Critical	Yes	?	✓	✓	✓	?	✓	?	?	✓	?	✓	✓	✓	✓	✓^	✓^	✓	✓	✓	✓	✓
Functional capacity	Patient Specific Functional Scale	Critical	Yes	--	--	--	--	--	--	--	--	--	--	--	✓	✓	--	✓^	--	--	--	--	--	--
Quality of Life	EQ5D-3L (or other)	Critical	No	--	--	--	--	--	--	--	--	✓	--	--	--	✓	✓	✓^	--	--	--	--	--	--
Physical performance	No eligible measures reported	Important	No	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Analgesic use	Narcotic or non-narcotic use	Important	Yes	--	--	--	--	--	--	--	--	--	--	--	--	--	✓	--	--	--	--	--	--	--
Work Status	Time to return to work	Important	No	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Abbreviations: EQ5D-3L, EuroQol- 5 Dimension three level; NPRS, Numeric Pain Rating Scale; RMDQ, Roland Morris Disability Questionnaire

^ NRSI with serious risk of bias not included in the evidence synthesis.

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Eighteen RCTs comparing Pilates with control (no intervention, usual care, educational advice) in people with low back pain were eligible for this comparison and contributed data relevant to five of the six outcomes (Cruz-Diaz 2015, Cruz Diaz 2016, Cruz-Diaz 2017, Cruz-Diaz 2018, da Fonseca 2009, Gladwell 2006, Hasanpour-Dehkordi 2017, Kofotolis 2016, Lopes 2014, Mazloum 2016, Miyamoto 2011, Miyamoto 2016, Natour 2011, Patti 2016, Quinn 2011, Rydeard 2006, Valenza 2017, Zeada 2012).

The three NRSIs (Kliziene 2017, Notarnicola 2014, Pappas 2013) were judged to be at serious risk of bias and were therefore not considered in the reporting of results, evidence synthesis or conclusions.

Pain

Fourteen RCTs (1072 participants) reported pain measured using the Numeric Pain Rating Scale (0-10) at the end of treatment (less than 12 weeks) (Cruz Diaz 2015, Cruz-Diaz 2016, Cruz-Diaz 2017, Cruz-Diaz 2018, da Fonseca 2009, Gladwell 2006, Lopes 2014, Mazloum 2016, Miyamoto 2011, Miyamoto 2016, Natour 2011, Quinn 2011, Rydeard 2006, Valenza 2017). Three RCTs (194 participants) had data not included in the meta-analysis (see Appendix F2).

The Numeric Pain Scale (0-10) is a segmented numeric version of the VAS. The 11-point numeric scale ranges from 0 (representing no pain) to 10 (representing pain as bad as you can imagine). The participant selects the whole number (between 0 and 10) that best represents the intensity of their pain. The Numeric Pain Scale is administered verbally or graphically for self-completion. A reduction of 2 points (or 30%) on the Numeric Pain Scale is estimated to be clinically important in people with diabetic neuropathy, postherpetic neuralgia, chronic low back pain, fibromyalgia and osteoarthritis (91). The pooled results suggest an effect favouring Pilates when compared to control (MD –1.69; 95% CI –2.39, –0.98; $p < 0.00001$; $I^2 = 88\%$) but the clinical importance was not reached *(GRADE: low)*.

In a sensitivity analysis that examined the impact of RCTs judged to be at a high risk of bias (3 studies) no important change in the result was observed (MD –1.69; 95% CI –2.36, –1.03; $p < 0.00001$; $I^2 = 80\%$)

One RCT (24 participants) reported pain measured by the McGill Pain Questionnaire at end of treatment (six weeks) (Hasanpour-Dehkordi 2017). The McGill Pain Questionnaire (0-78) is a self-reported measure of pain, which assesses both the quality and the intensity of subjective pain. It consists of 78 words, of which people with low back pain choose those that best describe their experience of pain. Scores are tabulated by summing values associated with each word. Scores range from zero, indicating no pain, to 78, indicating severe pain. The pooled results suggest an effect favouring Pilates when compared to control (MD –22.75; 95% CI –31.37, –14.13; $p < 0.00001$) *(GRADE: low)*.

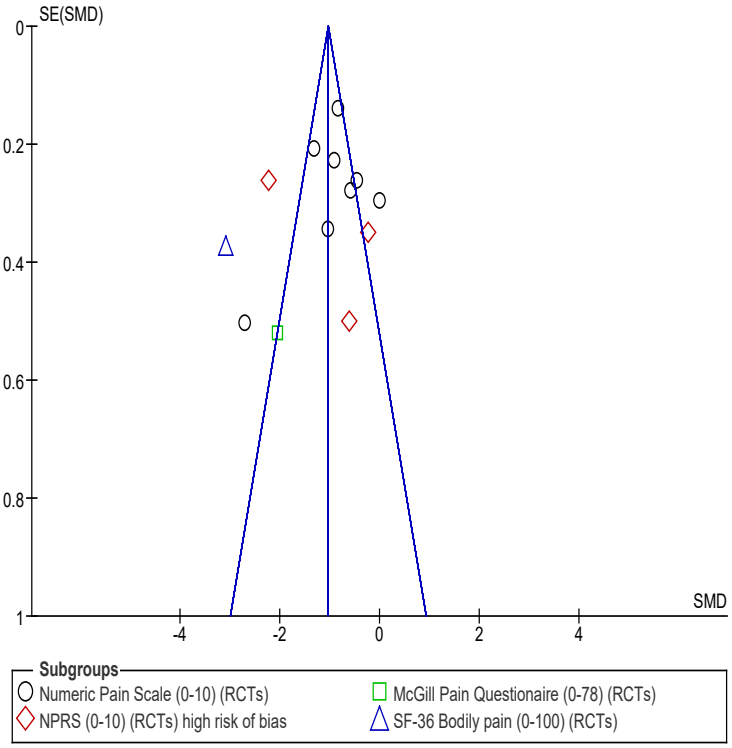
One RCT (65 participants) reported pain measured with the SF-36 bodily pain subscale at end of treatment (eight weeks) (Kotofolis 2016). The SF-36 bodily pain domain contains two items that assess pain severity and pain interference. Scores are summarised on a scale from 0 to 100, with a population mean of 50 and standard deviation of 10. The results showed an effect in favour of Pilates compared with the control group (MD – 37.53; 95% CI –44.00, –31.06; $p < 0.0001$) *(GRADE: low)*.

In combining the different pain measures, pooled results (13 studies with usable data) show an effect favouring Pilates when compared to control, however there was substantial heterogeneity (SMD –1.18; 95% CI –1.67, –0.75; $p < 0.00001$; $I^2 = 87\%$). Visual inspection of the funnel plot (see Figure D.19) suggests a possibility that results are missing from the meta-analysis, but it is not clear if it is due to non-reporting biases or other factors (e.g. inflated effects in smaller studies, artefactual correlations).

In a sensitivity analysis that examined the impact of small studies, a slight shift towards the larger study results was observed when using a fixed effects (SMD –1.04; 95% CI –1.18, –0.89; $p < 0.00001$; $I^2 = 87\%$).

In a sensitivity analysis that examined the impact of RCTs judged to be at a high risk of bias (5 studies) the size (but not overall direction) of the effect was reduced (MD –0.89; 95% CI –1.25, –0.53; $p < 0.00001$; $I^2 = 77\%$).

Figure D.19 Funnel plot of comparison: Low back pain, outcome: Pain (end of treatment)



Disability

Five RCTs (270 participants) reported disability measured with the Oswestry Disability Index (0-100) at the end of treatment (range 6 to 24 weeks) (Cruz Diaz 2016, Gladwell 2006, Mazloum 2016, Patti 2016, Valenza 2017). The Oswestry Disability Index is used to quantify disability related to lower back pain. The questionnaire is comprised of 10 questions that assess the ability of people with low back pain to manage everyday life. Answers are scored on a 0 (no disability) to 5 (great deal of disability) scale. The final score ranges from 0-100, with a score of 0-20 indicating minimal disability, 21-40 indicates moderate disability, 41-60 indicates severe disability, 61-80 indicates crippled (back pain impinges on all aspects of life), and 81-100 indicating complete disability (bed-bound). In people with chronic low back pain the minimal important change is calculated to be 12.88 (sensitivity 88%, specificity 85%) (170).

The pooled results show an effect favouring Pilates compared to control (no intervention or usual care) (MD –3.00; 95% CI –4.43, –1.58; $p < 0.0001$; $I^2 = 16\%$) *(GRADE: low)*, which was not considered clinically important. In a sensitivity analysis that examined the impact of the one RCT judged to be at a high risk of bias (Gladwell 2006) no important change in the result was observed (MD – 3.18; 95% CI –4.86, –1.49; $p = 0.0002$; $I^2 = 30\%$).

Eight RCTs[6] (Cruz-Diaz 2017, Kofotolis 2016, Miyamoto 2011, Miyamoto 2016, Natour 2011, Rydeard 2006, Valenza 2017, Zeada 2012) reported disability measured with the Roland Morris Disability Questionnaire (RMDQ) at the end of treatment (range 6 to 24 weeks) (total 721 participants). Data were from two RCTs (total 91 participants) not able to be included in the analysis (see Appendix F2).

The RMDQ is a measure of how low-back pain affects functional activities in people with mild to moderate acute or chronic low back pain. Answers are scored on a range from 0 (no disability) to 24 (severe disability). There are also 18-item or 21-item versions. In people with chronic low back pain the minimal important difference is reported to be 5 points (171, 172), with an RMDQ threshold value of 4 (out of 24) suggested to identify those who met their goals compared with those who did not (173).

The pooled results show an effect favouring Pilates compared to control (no intervention or usual care) but it does not reach the minimal important difference (MD –3.64; 95% CI –5.17, –2.10; $p < 0.0001$; $I^2 = 82\%$) *(GRADE: low)*. In a sensitivity analysis that examined the impact of RCTs judged to be at a high risk of bias (2 RCTs) a change in the size (but not direction) of effect was observed (MD –2.85; 95% CI –4.20, –1.49; $p = 0.006$; $I^2 = 69\%$).

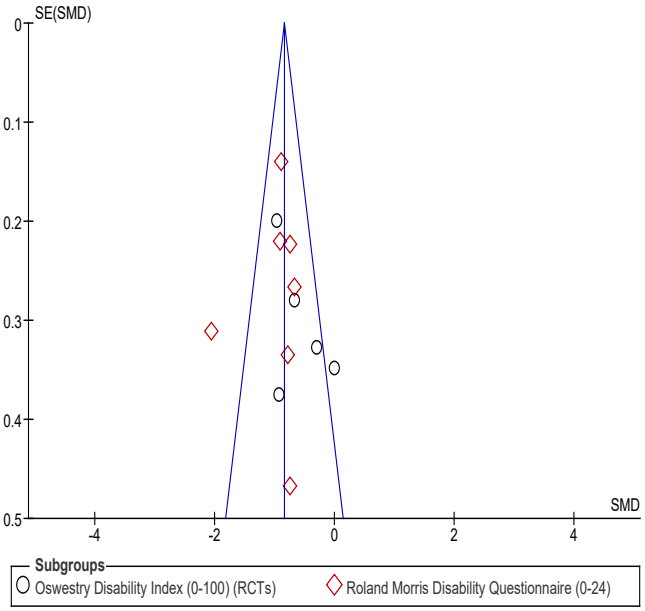
Taken together, the standardised mean result shows an effect favouring Pilates compared to control (no intervention or usual care)[7] (SMD –0.82; 95% CI –1.05, –0.59, $p = 0.008$; $I^2 = 56\%$) *(GRADE: low)*. In a sensitivity analysis that examined the impact of RCTs judged to be at a high risk of bias (3 RCTs) the size of effect was reduced (SMD –0.80; 95% CI –0.95, –0.64; $p < 0.0001$; $I^2 = 0\%$).

Visual inspection of the funnel plot (see Figure D.20) suggests no evidence of assymetry. In a sensitivity analysis that examined the impact of small studies, no material change in the result was observed when using a fixed effects model (SMD –0.84; 95% CI –0.98, –0.70; $p < 0.0001$; $I^2 = 56\%$).

[6] Includes one RCT (Valenza 2017) that measured disability using both the RMDQ and ODI (54 participants).

[7] RMDQ data from one study (Valenza 2017) not included; ODI is the preferred measure.

Figure D.20 Funnel plot of comparison: Low back pain, outcome: Disability (end of treatment).



Functional capacity

Two RCTs (381 participants) reported functional capacity measured with the patient specific functional scale (3-items) at the end of treatment (six weeks) (Miyamoto 2011, Miyamoto 2016).

The patient specific functional scale is a self-reported outcome measure of function for people with back, neck, knee, and upper extremity problems. People with low back pain are asked to identify up to five important activities they are unable to perform or have difficulty with because of their low back pain and to rate the difficulty associated with each activity on an 11-point scale. A score of zero indicates an inability to perform the activity and a score of 11 indicates the ability to perform the activity at a level prior to having low back pain. The final score is the sum of the activity scores divided by the number of activities. A minimum detectable change (90% confidence interval) for the average score is two points and for a single activity score it is three points (174).

The pooled results show an effect favouring Pilates when compared to the control group (education booklet), but it does not reach the minimal importance difference (MD –1.47; 95% CI –2.04, –0.90; $p <$ 0.00001; $I^2 = 14\%$) *(GRADE: LOW)*. No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as no studies were judged to be at high risk of bias.

Quality of life

Three RCTs (participants) reported quality of life measured with the SF-36 (Natour 2011), the SF-12 (Gladwell 2006) or the SF-6D (Miyamoto 2016) at the end of treatment (six or 12 weeks). The other studies did not report quality of life, probably because they did not measure the outcome.

Miyamoto 2016 reported the SF-6D, which uses information from the SF-36 (reduced to six dimensions), with an algorithm used to generate a continuous index for health. The minimal important difference of the SF-6D for the general popualiton is reported to range from 0.010 to 0.05 (20, 175). In people with chronic low back pain, the minimal important change score value is 0.031 (170)

The results showed an effect favouring Pilates when compared with the control group (education booklet) (MD –0.04; 95% CI –0.06, –0.02; *p* = 0.0002) *(GRADE: moderate)*. No sensitivity analysis was conducted as the study was not judged to be at high risk of bias.

Natour 2011 reported individual scores for each of the eight domains of the SF-36, which are summarised on a scale from 0 (worse) to 100 (best). A higher score means better quality of life. The MCID for individual domains of the SF-36 in people with chronic low back pain have not been established.

No important difference was observed between groups for any domain:

- physical functioning (MD –8.54; 95% CI –20.50, 3.42; *p* = 0.16),
- role-physical (MD –6.34; 95% CI –24.53, 11.85; *p* = 0.49),
- bodily pain (MD –8.04; 95% CI –20.51, 4.43; *p* = 0.21),
- general health perceptions (MD –10.88; 95% CI –21.23, –0.53; *p* = 0.04),
- vitality (MD –10.58; 95% CI –21.00, –0.16; *p* = 0.05),
- role-social (MD –4.23; 95% CI –16.88, 8.42; *p* = 0.05)
- role-emotional (MD –6.68; 95% CI –21.68, –8.32; *p* = 0.38) and
- mental health (MD –8.67; 95% CI –19.91, 2.57; *p* = 0.13).

Gladwell 2006 reported individual scores for each of the eight domains of the SF-12 (see Appendix F2), however, the SF-12 has only one or two items from each of the eight health concepts of the SF-36 and scores are intended to be yield scores for Physical and Mental Components Score (not individual domains).

The SF-36 measures the amount of limitation a patient is experiencing; but it does not enable trade-offs between different dimensions of health (176). As such, results from Miyamoto 2016 were used (cannot be combined with SF-36) in the GRADE summary findings.

Analgesic use

One trial (60 participants) reported non-narcotic (sodium diclofenac) intake measured with a chart at the end of treatment (90 days) (Natour 2011). It is not clear what the measure is (e.g. pills or mg per day, total). Non-narcotic intake is a self-reported measure of analgesic use. The results show no difference in analgesic use between the Pilates and no intervention groups (MD –5.66; 95% CI –13.73, 2.41; *p* = 0.17).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and was not judged to be at high risk of bias.

Comparison 2 (vs other)

There were fifteen studies comparing Pilates with 'other' interventions in people with low back pain that were eligible for this comparison. Nine RCTs (Albert Anand 2014, Avila Ribeiro 2015, Bhaduria 2017, Brooks 2012, Devasahayam 2016, Donzelli 2006, Gladwell 2006, Silva 2018, Wajswelner 2011) contributed data relevant to two of the seven outcomes.

Data from these studies are presented in Appendix F2 Supplementary outcome data.

D4.8 Neck and shoulder pain

D4.8.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-33. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-33 Overview of PICO criteria of included studies: Neck and shoulder pain

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Cazotti 2013 (177-179)	RCT	Neck pain (chronic, mechanical)	Pilates exercises (mat & equipment)	Control (usual care)	None specified	Pain Function QoL Use of analgesics
Dunleavy 2016 (180)	Quasi-RCT	Neck pain (chronic, mechanical)	Pilates exercises (mat)	Control (no intervention) OR yoga^	None specified	Disability Pain Range of movement Posture
Pilates versus 'other' intervention*						
Atilgan 2017 (181)	RCT	Shoulder pain (chronic, nonmechanical)	Pilates exercises	Conventional exercise	Hot pack, TENS	Pain Physical functioning
Ulug 2018 (182)	RCT	Neck pain (chronic, NOS)	Pilates exercises (mat)	Yoga (Iyengar) OR isometric exercise	Hot pack, ultrasound, TENS	Pain QoL Disability Range of motion Depression Muscle size

Abbreviations: NOS, not otherwise specified; NRSI, non-randomised study of interventions; RCT, randomised controlled trial; QoL, quality of life; TENS, transcutaneous electrical nerve stimulation

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

^ Study included three groups. The inactive control is considered in the evidence synthesis.

D4.8.2 Risk of bias summary

The risk of bias for each item in the included studies for neck and shoulder pain is described below and shown graphically in Figure D.21. Details are provided in Appendix E1.

Bias arising from the randomisation process

Two studies (Cazotti 2015, Ulug 2018) provided sufficient information on the randomisation process and were at low risk of bias for this domain. Some concerns were raised in one study (Dunleavy 2016) due to randomisation of participants based on geographic and time convenience with reason to suspect that the enrolling investigator or the participant had knowledge of the forthcoming allocation. One study had concerns raised (Atilgan 2017) due to a baseline imbalance reported in the duration of pain between groups before treatment initiated, suggesting a possible issue with the randomisation process.

Bias due to deviations from the intended intervention

One study (Atilgan 2017) did not report any deviations relating to the trial context and analysed results appropriately and thus was judged to be at a low risk of bias. Two studies had some concerns raised: Cazotti 2015 reported participant dropouts due to health reasons and Ulug 2018 failed to provide information on the reasons for dropouts, which raised concerns of bias but unlikely to substantially impact results. One study (Dunleavy 2016) excluded participants from the analysis due to noncompliance, which may overstate the effect of the intervention in favour, as participants who do not comply cannot benefit from the intervention.

Bias due to missing outcome data

Two studies (Atilgan 2017, Cazotti 2015) were assessed to be at low risk of bias for this domain as outcome data were available for all (or nearly all) participants. Dunleavy 2016 reported a high proportion of missing data (36.4%) that was balanced between groups. There was no analysis to assess the impact of missing data, raising some concerns. One study did not provide reasons for discontinuations and was assessed to be of high risk of bias with missingness of the data considered to affect the true value of the outcome (Ulug 2018).

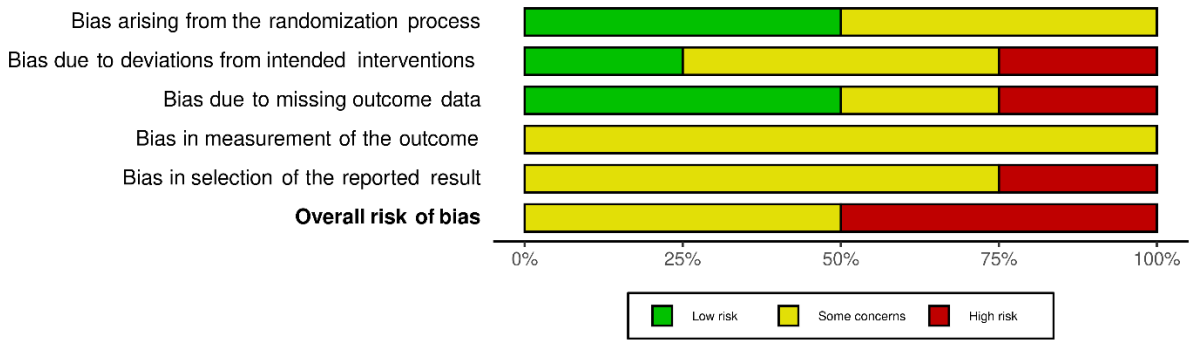
Bias in the measurement of the outcome

All studies (Atilgan 2017, Cazotti 2015, Dunleavy 2016, Ulug 2018) were judged to have some concerns of bias for this domain due to participant-reported outcomes, which could be influenced by knowledge of the intervention received. It was considered unlikely that the participant treatment experience substantially influenced the results.

Bias in selection of the reported result

Three studies (Atilgan 2017, Cazotti 2015, Ulug 2018) were judged to have some concerns of bias for this domain. The researcher's pre-specified intentions are not reported but are sufficiently described and data analysis was performed accordingly. The remaining study (Dunleavy 2015) did not report all outcomes listed in the methods, suggesting selective bias in reporting of results and was assessed to have high risk of bias.

Figure D.21 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Neck and shoulder pain



D4.8.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people with neck pain are listed in Table D-34.

Outcomes considered by the NTWC to be critical or important for decision-making in people with shoulder pain are listed in Table D-35.

Table D-34 Outcomes considered by the NTWC to be critical or important for decision-making: neck pain

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Cazotti 2013*	*Dunleavy 2016*
Pain	McGill Pain Questionnaire (Or Numeric Pain Scale)	Critical	Yes	✓	✓
Function/ Disability	Neck Disability Index	Critical	Yes	✓	✓
Quality of life	SF-36 – total score (or individual domain scores)	Critical	Yes	✓	--
Flexibility/ ROM	No eligible measures reported	Important	No	--	X
Psychosocial wellbeing	Beck Depression Inventory	Important	No	--	--
Fatigue	No measures reported in eligible studies	Important	No	--	--
Work status	No measures reported in eligible studies	Important	No	--	--

Abbreviations: ROM, range of motion; SF-36, 36-item short form survey

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Table D-35 Outcomes considered by the NTWC to be critical or important for decision-making: shoulder pain

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*
Pain	Shoulder Pain and Disability Index - Pain	Critical	No
Function/Disability	Shoulder Pain and Disability Index - Disability	Critical	No
Quality of life	SF-36 – total score (or individual domain scores)	Critical	No
Flexibility/ range of motion	No eligible measures reported	Important	No
Global perceived effect	No eligible measures reported	Important	No
Work status	No measures reported in eligible studies	Important	No

Abbreviations: SF-36, short form 36 questions

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One RCT (Cazotti 2015) and one quasi-RCT (Dunleavy 2016) comparing Pilates with control (no intervention or usual care) in people with neck pain were eligible for this comparison and contributed data relevant to two of the seven outcomes. There were two additional studies awaiting classification (available as abstracts only) and no ongoing studies that compared Pilates with no intervention in people with neck pain (total 97

participants) that could have contributed data to the outcomes considered critical or important to this review (see Appendix C6).

Outcome results were judged to be at high risk of bias (disability and pain) or had concerns raised (quality of life), which may over (or under) estimate the size of the effect. The concern is linked to missing outcome data and possible issues with the randomisation process.

Pain

Two studies (127 participants) reported pain measured with the Numeric Pain Scale (0-10) at the end of treatment (12 weeks) (Cazotti 2015, Dunleavy 2016). The Numeric Pain Scale (0-10) is a is a segmented numeric version of the VAS that is administered verbally or graphically for self-completion. The 11-point numeric scale ranges from 0 (representing no pain) to 10 (representing pain as bad as you can imagine). The participant selects the whole number (between 0 and 10) that best represents the intensity of their pain. A reduction of 2 points (or 30%) on the Numeric Pain Scale is reported to be clinically important in people with diabetic neuropathy, postherpetic neuralgia, chronic low back pain, fibromyalgia and osteoarthritis (91).

Pooled results show an effect favouring Pilates when compared with control (no intervention or usual care) (MD –3.10; 95% CI –5.22, –0.97; p = 0.004; I^2 = 89%) *(GRADE: very low)*.

Sensitivity analysis showed no change in the direction of the effect when the RCT judged to be at a high risk of bias (Dunleavy 2016) was removed from the analysis (MD –4.17; 95% CI –5.09, –3.25; p = 0.0002). The overall result and conclusions are not changed.

Function/Disability

Two trials (127 participants) reported disability measured with the Neck Disability Index at the end of treatment (12 weeks) (Cazotti 2013, Dunleavy 2016). The Neck Disability Index is designed to measure neck-specific disability using a questionnaire with 10 items relating to pain and activities of daily living. Each item is scored out of five giving a total score out of 50, with higher scores indicating greater disability. The minimal detectable change is estimated to be between 4.7 and 5.0 points (183).

Pooled results show an effect in favour of Pilates when compared with control (no intervention or usual care) (MD –6.55; 95% CI –8.80, –4.30; p < 0.00001; I^2 = 0%) *(GRADE: low)*.

Sensitivity analysis showed no change in the direction of the effect when the RCT judged to be at a high risk of bias (Dunleavy 2016) was removed from the analysis (MD –7.03; 95% CI –9.84, –4.22; p < 0.0001). The overall result and conclusions are not changed.

Quality of life

One trial (64 participants) reported quality of life measured with the SF-36 at the end of treatment (12 weeks) (Cazotti 2013). An overall global score was not provided, but individual scores for each of the eight domains were provided and are summarised on a scale from 0 (worse) to 100 (best) with a population mean of 50 and standard deviation of 10. Higher scores indicate a better health status. A change of 2.6 points in the SF-36 physical component summary score and a change of 15.5 points for SF-36-bodily pain has been reported to be clinically meaningful in people with chronic non-specific neck pain (184). No MCID for the SF-36 mental component scores has been established.

The results show no difference between the Pilates and control (usual care) groups for general health perceptions (MD –4.40; 95% CI –16.16, 7.36; p = 0.46) *(GRADE: very low)* or role-emotional (MD –7.30; 95% CI –25.93, 11.33; p = 0.44) *(GRADE: very low)*.

An effect favouring Pilates is reported for the other six domains: physical functioning (MD –11.40; 95% CI –19.50, –3.30; *p* = 0.006) *(GRADE: very low)*; role-physical (MD –27.00; 95% CI –45.28, –8.72; *p* = 0.004) *(GRADE: very low)*; bodily pain (MD –13.90; 95% CI –23.08, –4.72; *p* = 0.003) *(GRADE: very low)*; vitality (MD –12.60; 95% CI –23.50, –1.70; *p* = 0.02) *(GRADE: very low)*; role-social (MD –17.60; 95% CI –28.24, –6.96; *p* = 0.001) *(GRADE: very low)*; and mental health (MD –14.10; 95% CI –22.90, –5.30; *p* = 0.002) *(GRADE: very low)*. Across each domain, the clinical importance of the effect varies[8].

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and it was not judged to be at high risk of bias.

Comparison 2 (vs other)

Two RCTs (Ulug 2018, Atilgan 2017) comparing Pilates with 'other' interventions in people with neck or shoulder pain were eligible for this comparison and contributed data for two of the seven outcomes.

Data from these studies are presented in Appendix F2 Supplementary outcome data.

[8] not different (general health perceptions, role-emotional); not important (bodily pain); moderate (physical function, vitality, role-social, mental health); or large (role-physical).

D1 Disease of the genitourinary system

D1.1 Menopausal symptoms or complaint

D4.8.4 List of studies

An overview of the PICO criteria of included studies is provided in Table D-36. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-36 Overview of PICO criteria of included studies: Menopausal symptom or complaint

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Ahmadinezhad 2017 (185, 186)	RCT	Postmenopausal women (aged 40 to 60 years)	Pilates exercises	Control (no intervention) OR Acupressure^	None specified	Sleep quality Anxiety
Campos de Oliveira 2018 (187-189)	RCT	Postmenopausal women (aged 40 to 70 years)	Pilates exercises (equipment)	Control (no intervention) OR Whole-body vibration^	None specified	Muscle strength Bone mineral density Quality of life Balance
Lee 2016a (190)	Quasi-RCT	Postmenopausal women (aged 45 to 60 years)	Pilates exercises	Control (no intervention)	None specified	Flexibility Quality of life
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: RCT, randomised controlled trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

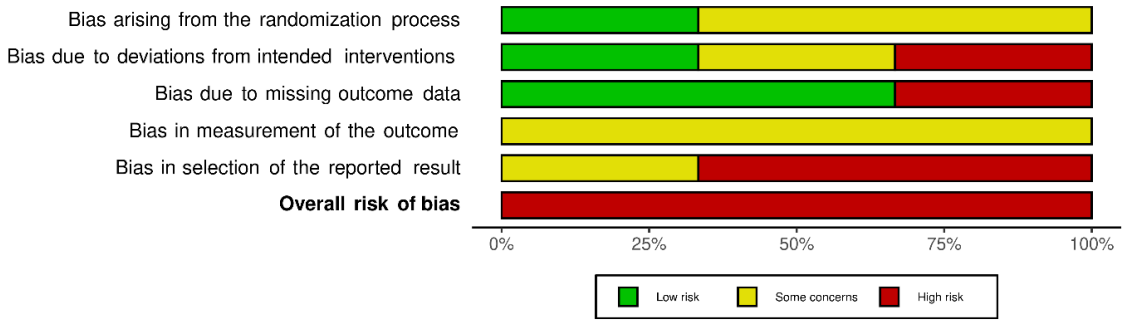
**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

^ Study included three groups. The inactive control is considered in the evidence synthesis.

D4.8.5 Risk of bias summary

The risk of bias for each item in the included studies for menopausal symptoms or complaints is described below and shown graphically in Figure D.22. Details are provided in Appendix E1.

Figure D.22 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Menopausal symptom or complaint



Bias arising from the randomisation process

One study (Campos de Oliveira 2018) was judged to be at low risk of bias for this domain. Details relating to method of randomisation or allocation concealment were not provided in the other two studies (Ahmadinezhad 2017, Lee 2016a) raising some concerns.

Bias due to deviations from the intended intervention

One study (Campos de Oliveira 2018) was judged to be at low risk of bias for this domain. Concerns were raised in one study (Ahmadinezhad 2017) because it was not clear if all randomised participants were included in the analysis. There was some indication that participants who missed more than two sessions were excluded from the analysis, but details were lacking. One study (Lee 2016a) was judged to be at high risk of bias, as the number of participants was not balanced between groups and there was no information provided to make any further assessment.

Bias due to missing outcome data

Two studies (Ahmadinezhad 2017, Campos de Oliveira 2018) were judged to be at low risk of bias for this domain as information appeared to be available for all, or nearly all, participants. One study (Lee 2016a) was judged to be at high risk of bias for this domain, as there were concerns regarding missing outcome data that may affect the true value of the outcome.

Bias in the measurement of the outcome

All three studies (Ahmadinezhad 2017, Campos de Oliveira 2018, Lee 2016a) had concerns raised relating to subjective outcomes potentially influenced by knowledge of the intervention received.

Bias in selection of the reported result

One study (Lee 2016a) had concerns raised due a minimal information describing the statistical analysis plan, suggesting not all intended outcomes or analyses were reported. In two studies (Ahmadinezhad 2017, Campos de Oliveira 2018) there were multiple post hoc analyses suggestive of selective reporting and one or more outcome domains (primary or secondary) listed in the prespecified trial registry were not reported or discussed, increasing the risk of bias to high.

D4.8.6 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in otherwise healthy menopausal or postmenopausal women are listed in Table D-37.

Table D-37 Outcomes considered by the NTWC to be critical or important for decision-making: Menopausal symptom or complaint

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Ahmadinezhad 2017*	*Campos de Oliveira 2018*	*Lee 2016a*
Quality of life, global	SF-36 – total score	Critical	Yes [a]	--	X	--
Sleep quality	PSQI – global score	Important	Yes	✓	--	--
Vasomotor symptoms	MSQ-Vasomotor	Important	Yes	--	--	✓
Global physical functioning	MSQ - Physical (or SF-36 - Physical Component Score)	Important	Yes	--	--	✓
Physical performance	30 second Chair-Stand test (or isokinetic muscular strength)	Important	Yes	--	✓	--
Bone mineral density	T-score (or other BMD measure)	Important	Yes	--	✓	--
Depression	Hospital Anxiety and Depression Scale	Important	No	?	--	--

Abbreviations: BMD, bone mineral density; MSQ, Menopausal symptoms questionnaire; PSQI, Pittsburgh Sleep Quality Index; SF-36, 36-item Short Form Survey

a. total score not reported by study authors, individual domain scores provided here

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Three RCTs (Ahmadinezhad 2017, Campos de Oliveira 2018, Lee 2016a) comparing Pilates with no intervention in otherwise healthy menopausal or postmenopausal women were eligible for this comparison. All three RCTs contributed data relevant to six of the seven outcomes. There were no additional studies awaiting classification or ongoing that compared Pilates with no intervention in otherwise healthy menopausal or postmenopausal women that could have contributed data to the outcomes considered critical or important to this review (see Appendix C6).

Results for all outcomes were judged to be at high risk of bias, which may over (or under) estimate the size of the effect. The concern is linked to missing information, skewed data, and selective reporting of results.

Quality of life, global

One trial (34 participants) reported of quality of life measured with the SF-36 at the end-of-treatment (26 weeks) (Campos de Oliveira 2018). An overall global score was not provided. Individual scores for each of the eight domains were provided and are summarised on a scale from 0 (worse) to 100 (best). The data were considered nonparametric (i.e., skewed) and presented as median (IQR), with post hoc Bonferroni test used to test multiple comparisons between the pairs *(GRADE: very low)*.

Available data are summarised below and provided in Appendix F2 Supplementary outcome data.

Physical functioning

The SF-36 physical functioning domain is a 10-item measure of physical limitation in a range of activities from vigorous exercise to performing self-care activities. The post-hoc analysis showed no difference between the Pilates group compared with the control group (p = NR).

Role-physical

The SF-36 role-physical domain contains four items that measures limitations in various roles, including work and daily activities. The post-hoc analysis showed a difference between the Pilates group compared with the control group (p = 0.033), with the Pilates group reporting better scores than the control group at the end of treatment.

Bodily pain

The SF-36 bodily pain domain contains two items that assess pain severity and pain interference. The post-hoc analysis showed a difference between the Pilates group compared with the control group (p = 0.035), with the Pilates group reporting higher quality of life scores than the control group at the end of treatment.

General health perceptions

The SF-36 general health perceptions domain contains five items that evaluates a persons' physical health problems and their confidence in progression to better or worse health. The post-hoc analysis showed no difference between the Pilates group compared with the control group (p = NR).

Fatigue

The SF-36 vitality domain has four items that measure vitality, energy level, and fatigue and is intended to be a measure of subjective well-being. The post-hoc analysis showed no difference between the Pilates group compared with the control group (p = NR).

Social function

The SF-36 role-social domain includes two items that measure the impact of physical and mental health on social functioning. The post-hoc analysis showed a difference between the Pilates group compared with the control group (p = 0.025), with the Pilates group reporting higher scores than the control group at the end of treatment.

Mental function

The SF-36 role-emotional domain measures role limitations due to mental health difficulties, with three items that focus on amount of time spent on work or other activities, amount of work accomplished, and the care with which work is performed. The post-hoc analysis showed a difference between the Pilates group compared with the control group (p = 0.011), with the Pilates group reporting higher scores than the control group at the end of treatment.

Mental health

The SF-36 mental health domain has five items that measure anxiety, depression, loss of behavioural/emotional control, and psychological well-being. The post-hoc analysis showed no difference between the Pilates group compared with the control group (p = NR).

Sleep quality

One trial (72 participants) reported sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI) at the end-of-treatment (6 weeks) (Ahmadinezhad 2017). The PSQI is a nine-item[9] questionnaire that assesses the sleep quality of an individual in the previous month. It assesses seven sleep components including

[9] The PSQI can include five additional partner-rated questions (Item 10) that do not contribute to the total PSQI score (not used in the included study).

subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disorder (sleep fragmentation), use of sleeping medication, and daytime dysfunction (191). Each item is scored (range from 0 to 3) with the total global score ranging from 0 (no problems) to 21 (severe problems). A score of five or more is associated with poor sleep quality. A validated MCID for PSQI in postmenopausal women was not found.

The results showed an effect favouring Pilates compared with no intervention (MD –9.83; 95% CI –11.11, –8.55; *p* = 0.0006) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Vasomotor symptoms

One trial (74 participants) reported vasomotor symptoms measured with the Menopausal Symptoms Questionnaire (MSQ) at the end-of-treatment (8 weeks) (Lee 2016a). The vasomotor symptoms subscale of the MSQ includes seven items that are answered on a scale of 0 (no symptoms) to 6 (symptoms are very severe). A higher score indicates worse symptoms, but the clinical importance of this is not known (192).

The results showed an effect favouring Pilates compared with no intervention (MD –8.88; 95% CI –13.40, –4.36; *p* = 0.0001) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Global physical functioning

One trial (74 participants) reported physical symptoms measured with the Menopausal Symptoms Questionnaire (MSQ) at the end-of-treatment (8 weeks) (Lee 2016a). The physical symptoms subscale of the MSQ includes 11 items that are answered on a scale of 0 (no symptoms) to 6 (symptoms are very severe). A higher score indicates worse symptoms, but the clinical importance of this is not known (192).

The results showed an effect favouring Pilates compared with no intervention (MD –14.44; 95% CI –20.19, –8.69; *p* < 0.00001) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Physical performance

One study (34 participants) reported physical performance measured using an isokinetic dynamometer at the end of treatment (26 weeks) (Campos de Oliveira 2018). The equipment isolates the joint of interest, which allows targeted testing of selected muscle groups, and was used to measure peak isokinetic torque (Newtons per metre) of the knee extensor and knee flexors at the angular velocity of 60 and 180 degrees per second. A higher score indicates higher peak strength (torque), which is suggested to offset age-related muscle decline in menopausal and postmenopausal women (193). Baseline scores were not balanced between groups, with a post hoc Bonferroni test used to test multiple comparisons between the pairs.

End of treatment results showed improvement in peak strength of the knee extensors in the Pilate group compared to the control at a velocity of 60° per second (MD –14.60; 95% CI –28.66, –0.54; *p* = 0.04) *(GRADE: very low)* and 180° per second (MD –9.30; 95% CI –16.37, –2.23; *p* = 0.01) *(GRADE: very low)*.

Results showed no improvements in peak strength of the knee flexors at a velocity of 60° per second (MD –6.80; 95% CI –14.81, 1.21; *p* = 0.10) *(GRADE: very low)* or 180° per second (MD –3.70; 95% CI –10.15, 2.75;

p = 0.26) *(GRADE: very low)*. The change from baseline results suggested an effect in favour of Pilates for all measures (see Appendix F2 Supplementary outcome data).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Bone mineral density

One study (34 participants) reported BMD measured via DXA at the end of treatment (26 weeks) (Campos de Oliveira 2018). The study authors assessed and reported six different bone regions including lumbar spine (L1-L4), femoral neck, total hip, trochanter, intertrochanter, and Ward's area but did not provide an overall T-score[10] (or Z-score[11]) at the end of treatment. BMD is a measure how much calcium and other types of minerals are in an area of your bone and provide an assessment of bone density and fracture risk related to osteopenia or osteoporosis. Absolute BMD values (g/cm^2), calculated by comparing the participant measurements against the manufacturer's calibration standard, allows comparison between measurements at different times with changes between 4–7% of the baseline BMD value (i.e. about 0.050 g/cm^2) likely to be associated with clinically significant BMD change (194, 195).

The study found no difference between the Pilates and control groups for any BMD values at the end of treatment: Lumbar spine (MD 0.000; 95% CI –0.09, 0.09; *p* = 1.0), femoral neck (MD –0.02; 95% CI –0.09, 0.05; *p* = 0.57), total hip (MD –0.01; 95% CI –0.07, 0.05; *p* = 0.76) *(GRADE: very low)*.

Baseline data between treatment groups were not balanced and a post hoc Bonferroni test was used to test multiple comparisons between the pairs. The authors reported a difference between change scores comparing Pilates with control for lumbar spine (MD 0.016; 95% CI 0.007, 0.025; *p* = 0.008) and trochanter (MD 0.020; 95% CI 0.010, 0.031; *p* = 0.005) (see Appendix F2 Supplementary outcome data).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Comparison 2 (vs other)

Two RCTs (Ahmadinezhad 2017, Campos de Oliveira 2018) comparing Pilates with 'other' interventions in menopausal or postmenopausal women were eligible for this comparison and contributed data to four outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

[10] compared with what is normally expected in healthy young women.

[11] compared with what is normally expected in women of similar age, sex, weight, and ethnic or racial origin

D5 Pregnancy, childbirth, or the puerperium

D5.1 Postpartum recovery

D5.1.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-38. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-38 Overview of PICO criteria of included studies: Postpartum recovery

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Mirmohammadali 2012 (196-199)	RCT (cluster)	Postpartum women (72 hours to 1 week after vaginal delivery)	Pilates exercises (mat)	Control (no intervention)	None	Sleep quality Fatigue
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: RCT, randomised controlled trial

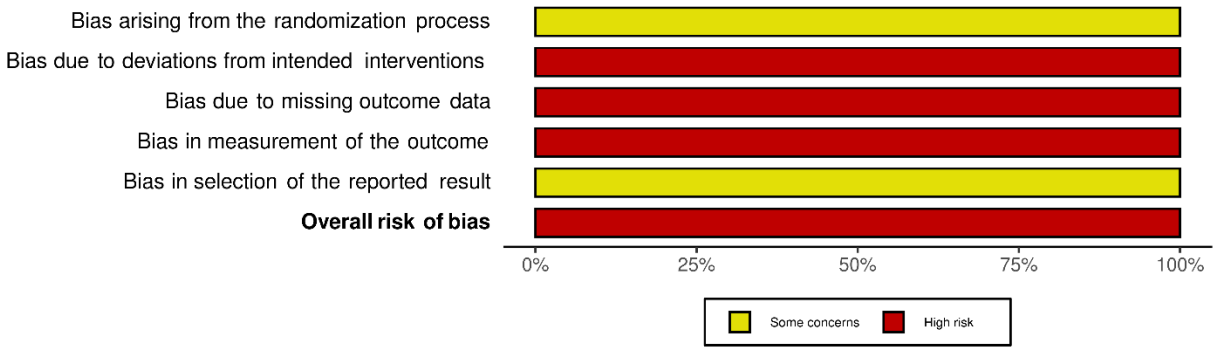
*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D5.1.2 Risk of bias summary

The risk of bias for each item in the included studies for Postnatal care is described below and shown graphically in Figure D.23. Details are provided in Appendix E1.

Figure D.23 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Postpartum recovery



Bias arising from the randomisation process

One study (Mirmohammadali 2012) was a cluster randomised trial. Participants at four of seven health centres (selected using randomised block-level sampling) were allocated to the Pilates group, with participants at the remaining three health centres allocated to control. The authors did not report on

allocation concealment, raising some concerns, but baseline characteristics did not suggest any substantial problems with the randomisation process.

Bias arising from the timing of identification and recruitment of individual participants

Mirmohammadali 2012 was assessed to be at low risk of bias for this domain because the women had been referred to the centres for prenatal care prior to randomisation.

Bias due to deviations from the intended intervention

Mirmohammadali 2012 was assessed to be at high risk of bias for this domain because there was no information provided to assess the potential impact of deviations from the intended intervention. It was noted that participants who did not perform the exercises for three consecutive sessions or had more than five interrupted sessions had been excluded from the analysis, the impact of this on the outcome result is not known.

Bias due to missing outcome data

Mirmohammadali 2012 was assessed be at high risk of bias for this domain. The study provides no information regarding dropouts, or the extent of missing outcome data and the analysis is unlikely to have removed the risk of bias arising from the missing data.

Bias in the measurement of the outcome

Mirmohammadali 2012 was assessed to be at high risk of bias for this domain. The testing measures and timing used to assess the outcomes were considered appropriate, but because outcomes were self-reported and the participants were clearly motivated (e.g., daily exercise diary) it is considered highly likely that participants would be biased in their reporting of the outcome in favour of the intervention.

Bias in selection of the reported result

Mirmohammadali 2012 was assessed to have some concerns in this domain. Reporting of the outcome results was not based on multiple eligible measures, analyses or time points but there was no pre-specified analysis plan, and the study was added to the trial registry retrospectively.

D5.1.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in postpartum women are listed in Table D-39.

Table D-39 Outcomes considered by the NTWC to be critical or important for decision-making: Postpartum recovery

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Mirmohammadali 2012*
Pelvic pain and dysfunction	ICIQ – Urinary Incontinence Short Form	Critical	No	--
Quality of life	PROMIS Global 10	Critical	No	--
Pelvic floor muscle function	No measures reported in eligible studies	Important	No	--
Body composition	No measures reported in eligible studies	Important	No	--
Fatigue	Multidimensional Fatigue Inventory	Important	Yes	✓
Mental health	Edinburgh Postnatal Depression Scale	Important	No	--
Exercise capacity	No measures reported in eligible studies	Important	No	--

Abbreviations: ICIQ, International Consultation on Incontinence Questionnaire; PROMIS, Patient Reported Outcomes Measurement Information System

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One study (Mirmohammadali 2012) comparing Pilates to control in postpartum mothers was eligible for this comparison and contributed data to one of the seven critical or important outcomes. There were no additional studies awaiting classification or ongoing that compared Pilates with no intervention in postpartum mothers that could have contributed data to the outcomes considered critical or important to this review (see Appendix C6).

Fatigue

One study (80 participants) measured general fatigue using the Multidimensional Fatigue Inventory at the end of an 8-week intervention period (Mirmohammadali 2012). The Multidimensional Fatigue Inventory is a 20-item scale that measures fatigue in the previous days across five domains: general fatigue, physical fatigue, reduced activity, reduced motivation, and mental fatigue. Each domain has a possible score ranging from 4 to 20. To our knowledge, the MFI-20 has not been validated in postpartum women (200). In people with cancer undergoing radiotherapy the MCID change scores range from 1.36 to 2.39 units in each domain. The applicability of this in the postpartum population is not known (201).

Results for each domain show a treatment effect in favour of Pilates (see Appendix F2). Results for the general fatigue domain showed an effect favouring Pilates (MD –4.92; 95% CI –5.77, –4.07; $p < 0.001$) *(GRADE: very low)*. No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Comparison 2 (vs other)

There were no studies identified comparing Pilates with 'other' in postpartum women.

D6 Injury, poisoning, or other certain other consequences of external causes

D6.1 Rehabilitation of the knee after injury

D6.1.1 List of studies

An overview of the PICO criteria of included studies is provided in Table D-40. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-40 Overview of PICO criteria of included studies: Rehabilitation of the knee after injury

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Celik 2017 (202)	RCT	Isolated ACL injury	Pilates exercises	Control (waitlist)	None reported	Knee function Muscle strength Global improvement
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: RCT, randomised controlled trial; ACL, anterior cruciate ligament

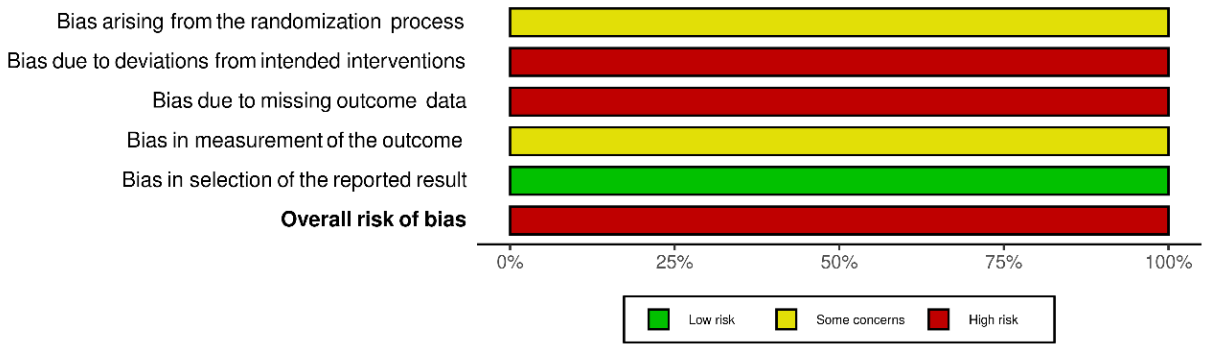
*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D6.1.2 Risk of bias summary

The risk of bias for each item in the included studies for rehabilitation of the knee after injury is described below and shown graphically in Figure D.24. Details are provided in Appendix E1.

Figure D.24 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Rehabilitation of the knee after injury



Bias arising from the randomisation process

Celik 2017 was assessed to have some concerns for this domain due to lack of information provided regarding the allocation concealment process. Baseline characteristics did not suggest any substantial problems with the randomisation process.

Bias due to deviations from the intended intervention

Celik 2017 was assessed at high risk of bias in this domain due to the unbalanced rate of loss to follow up between the two groups. In the Pilates group, 25% of participants discontinued for reasons not provided and likely to be related to the trial context. There was 10% dropout in the control group. There was insufficient information to assess the potential impact any deviation from the assigned intervention had on the outcome. The method of analysis was considered appropriate (modified ITT).

Bias due to missing outcome data

Celik 2017 was assessed to be at high risk of bias for this domain due to the large amounts of missing data and lack of analysis presented to assess the impact of this. Without reasons for drop out, it is difficult to assess whether this was likely related to the true value of the outcome.

Bias in the measurement of the outcome

Celik 2017 was assessed to have some concerns for this domain due to the self-reported nature of the outcomes and a lack of blinding of participants.

Bias in selection of the reported result

Celik 2017 was considered to be at low risk of bias for this domain. There was no suspicion of reported results being selected on the basis of multiple eligible domains, time points or analyses.

D6.1.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people undergoing knee rehabilitation after injury are listed in Table D-41.

Table D-41 Outcomes considered by the NTWC to be critical or important for decision-making: Rehabilitation of the knee after injury

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Celik 2017*
Knee function	Cincinnati Knee Rating System	Critical	Yes	✓
Quality of life	No measures reported in eligible studies	Critical	No	--
Return to activity/sports	No measures reported in eligible studies	Critical	No	--
Knee stability	Global Rating of Change Scale	Important	Yes	✓
Physical functioning	No measures reported in eligible studies	Important	No	--
Isokinetic muscle strength	Flexion peak torque, Extension peak torque	Important	Yes	✓
Requirement for surgery	No measures reported in eligible studies	Important	No	--

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One study (Celik 2017) comparing Pilates with control (waitlist) in people undergoing knee rehabilitation after injury was eligible for this comparison and contributed data to three of the seven critical or important

outcomes. There were no additional studies awaiting classification or ongoing that compared Pilates with no intervention in people undergoing knee rehabilitation after injury that could have contributed data to the outcomes considered critical or important to this review (see Appendix C6).

All outcomes were judged to be at high risk of bias, related to concerns about deviations from the intervention that were not balanced between groups and missing outcome data that could potentially over (or under) state the effect.

Knee function

One trial (50 participants) reported knee function measured with the Cincinnati Knee Rating System at the end of treatment (12 weeks) (Celik 2017). The Cincinnati Knee Rating System has multiple versions which assess knee function based on symptoms of pain, swelling, and giving way. Scores range from 120 (worse) to 420 (best). No MCID has been established.

The results showed no difference in knee function for the Pilates group compared to no intervention (MD –4.10; 95% CI –10.10, 1.90; $p = 0.18$) *(GRADE: very low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Knee Stability

One trial (50 participants) reported improvement in stability with the Global Rating of Change scale at the end of treatment (12 weeks) (Celik 2017). Participants were asked to rate their knee condition compared to the beginning of the exercise program by stating if they had improved, stayed the same, or deteriorated. All subjects in the Pilates arm stated they felt much better (22 out of 24 participants) or slightly better (2 out of 24 participants), compared to the control arm where only 6 out of 26 participants stated they felt slightly better.

The results for improvement showed an effect in favour of Pilates (RR 4.07; 95% CI 2.08, 7.97; $p < 0.001$) *(GRADE: low)*. No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Strength

One trial (50 participants) reported isokinetic muscle strength of the injured leg, measured with a dynamometer at the end of treatment (12 weeks) (Celik 2017). The equipment isolates the joint of interest, which allows targeted testing of selected muscle groups, and was used to measure peak isokinetic torque (Newtons per metre) of the quadriceps and hamstrings (using both flexion and extension) at the angular velocity of 180 degrees per second.

The results showed no difference in peak strength for flexion comparing the Pilates and control groups (MD –9.10; 95% CI –23.16, 4.96; $p = 0.20$) *(GRADE: very low)*. A treatment effect favouring Pilates was observed for peak extension (MD –23.90; 95% CI –39.59, –8.21; $p = 0.003$) *(GRADE: very low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Comparison 2 (vs other)

There were no studies identified comparing Pilates with 'other' comparator.

D7 External causes of morbidity and mortality

D7.1 Rehabilitation of the knee after arthroplasty

An overview of the PICO criteria of included studies is provided in Table D-42. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-42 Overview of PICO criteria of included studies: Rehabilitation of the knee after arthroplasty

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Karaman 2017 (203)	Quasi-RCT	Rehabilitation after total knee arthroplasty	Pilates exercises	Control (no intervention)	Standard rehabilitation exercises	Postural stability Physical functioning Quality of life
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: RCT, randomised controlled trial

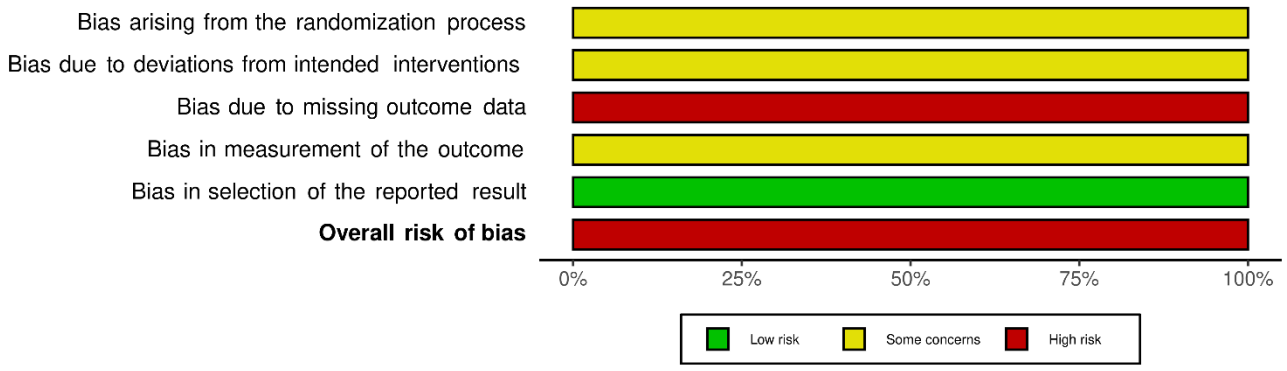
*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D7.1.1 Risk of bias summary

The risk of bias for each item in the included studies for rehabilitation of the knee after arthroplasty is described below and shown graphically in Figure D.25. Details are provided in Appendix E1.

Figure D.25 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Rehabilitation of the knee after arthroplasty



Bias arising from the randomisation process

Karaman 2017 was assessed to have some concerns for this domain due to a lack of information regarding allocation concealment. Randomisation was conducted using a computer-generated random number table and there were no baseline differences between groups to suggest a problem with randomisation.

Bias due to deviations from the intended intervention

Karaman 2017 was assessed to have some concerns for this domain due to the high level of drop out between the two treatment groups. Reasons for discontinuation were possibly related to the trial context but were balanced between treatment groups.

Bias due to missing outcome data

Karaman 2017 was assessed to be at high risk of bias for this domain due to the large proportion of missing data (26% of participants), which is likely related to the health status of the participants (e.g., exacerbation of pain, need for additional surgery). No analysis was presented to adjust for this missingness or to demonstrate that the result was not biased.

Bias in the measurement of the outcome

Karaman 2017 was assessed to have some concerns for this domain due to the self-reported nature of the outcomes which could potentially be influenced by knowledge of the intervention received .

Bias in selection of the reported result

Karaman 2017 was assessed at low risk of bias for this domain as it was not considered likely that the reported result was selected on the basis of multiple eligible domains, time points or analyses.

D7.1.2 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people requiring rehabilitation of the knee after arthroplasty are listed in Table D-43.

Table D-43 Outcomes considered by the NTWC to be critical or important for decision-making: Rehabilitation of the knee after arthroplasty

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Karaman 2017*
Quality of life	SF-36 – total score (OR Physical/Mental Component Scores OR individual domains)	Critical	Yes	✓
Balance	Berg Balance Test	Critical	Yes	✓

Abbreviations: SF-36; 36-item short form survey

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One study (Karaman 2017) comparing Pilates with no intervention (delivered as an adjunct to standard rehabilitation exercises) in people after total knee arthroplasty (TKA) was eligible for this comparison and contributed data to all outcomes. There were no additional studies awaiting classification or ongoing that compared Pilates with no intervention in people after TKA that could have contributed data to the outcomes considered critical or important to this review (see Appendix C6).

All outcomes were judged to be at high risk of bias, related to concerns about skewed baseline scores and missing outcome data that could potentially over (or under) state the effect.

Available data are presented in Appendix F2 Supplementary outcome data.

Quality of life

One trial (36 participants) reported quality of life measured with the SF-36 (36-item Short Form Survey) at the end-of-treatment (6 weeks) (Karaman 2017). Results were summarised into two composite scores (physical and mental health) that are reported on a scale from 0 (worse) to 100 (best). Individual domain scores were also provided (see Appendix F2).

MCID values for SF-36 summary scores in people after primary total knee replacement at 6 weeks are not available (204). One study (205) provides estimates for domain scores in people 6 months after surgery (ranging from 0.11 for general health to 12.8 for bodily pain), however given likely time variance the usefulness of these data is limited.

The results show an effect favouring Pilates for health-related physical (MD –6.70; 95% CI –11.24, –2.16.; p = 0.004) *(GRADE: very low)* and mental health (MD –12.50; 95% CI –20.30, –4.70; p = 0.002) *(GRADE: very low)* when delivered as an adjunct to standard post-operative exercises, however the data were not normally distributed at baseline making interpretation of the results uncertain.

Balance

One trial (36 participants) reported on the ability to balance during a series of predetermined tasks measured using the Berg Balance test at the end-of-treatment (6 weeks) (Karaman 2017). In most of the 14-items, the subject is asked to maintain a given position for a specific time, with each item consisting of a five-point ordinal scale ranging from 0 to 4. Total scores range from 0 (low balance stability) to 56 (high balance stability). In elderly people, a score of less than 45 indicates individuals may be at greater risk of falling (64), noting the test may have poor utility in people after TKA due to a ceiling effect observed at 12 and 24 weeks after TKA (206).

The results show an effect favouring Pilates when delivered as an adjunct to standard post-operative exercises (MD –12.55, 95% CI –20.30, –4.70; p = 0.002) *(GRADE: very low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its removal would leave no result.

Comparison 2 (vs other)

There were no studies identified that compared Pilates with an 'active' comparator in rehabilitation of the knee after arthroplasty.

D8 Prevention of disease, injury, or illness in at risk populations

D8.1 Prevention of mental health conditions

D8.1.1 List of studies

An overview of the PICO criteria of included studies for prevention of mental health conditions is provided in Table D-44. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-44 Overview of PICO criteria of included studies: Prevention of mental health conditions

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Abavisani 2019 (207, 208)	Quasi-RCT	Employment conditions, at-risk adults (emergency department, 18 to 40 years)	Pilates exercises	Control (usual activities)	None specified	Anxiety
Pilates versus 'other' intervention*						
No studies found.						

Abbreviations: RCT, randomised controlled trial

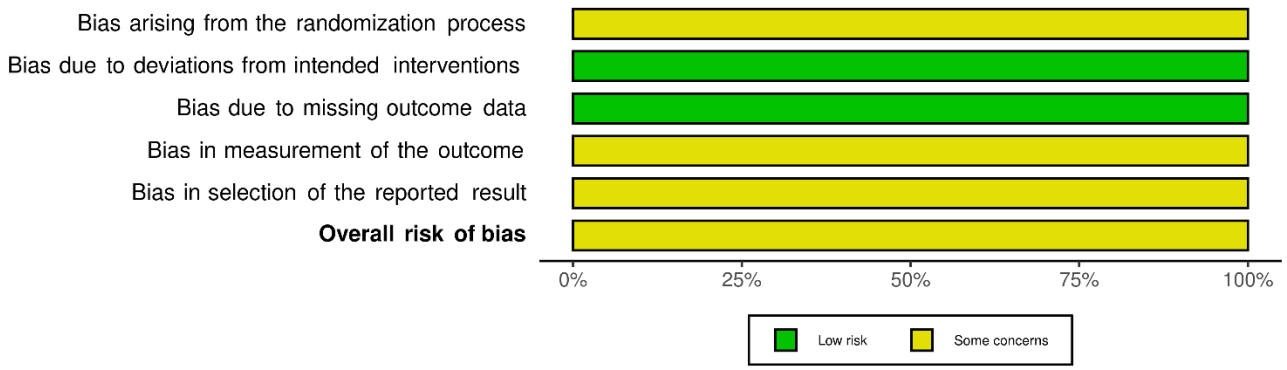
*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D8.1.2 Risk of bias summary

The risk of bias for each item in the included studies for prevention of mental health conditions is described below and shown graphically in Figure D.26.

Figure D.26 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Prevention of mental health conditions



Bias arising from the randomisation process

One study (Abavisani 2019) provided basic information regarding randomisation method supported by comparable baseline characteristics between the two groups. No information regarding concealment allocation was described raising some concerns.

Bias due to deviations from the intended intervention

One study (Abavisani 2019) was judged to be at low risk of bias for this domain due as there were no deviations or discontinuations from intended interventions.

Bias due to missing outcome data

One study (Abavisani 2019) was judged to be at low risk of bias for this domain as outcome data were available for all enrolled participants.

Bias in the measurement of the outcome

One study (Abavisani 2019) had some concerns raised relating to the measurement of outcomes. Participants or outcome assessors were not blinded, and primary outcomes were subjective, results of which could be influenced by knowledge of the intervention.

Bias in selection of the reported result

One study (Abavisani 2019) reported all eligible specified results for one outcome but data for a second outcome listed in the trial registry was not provided. In the absence of an available protocol, the study was judged to be at some concerns for this domain.

D8.1.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people at risk of mental health disorders are listed in Table D-45.

Table D-45 Outcomes considered by the NTWC to be critical or important for decision-making: Prevention of mental health conditions

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Abavisani 2019*
Quality of life	No measures reported in eligible studies	Critical	No	?
Active coping	No measures reported in eligible studies	Important	No	?
Anxiety	Spielberger anxiety questionnaire	Important	Yes	✓
Depression	No measures reported in eligible studies	Important	No	?
Physical stress symptoms	No measures reported in eligible studies	Important	No	?
Fatigue	No measures reported in eligible studies	Important	No	?
Stress/ stress perception/ burnout	No measures reported in eligible studies	Important	No	?

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One quasi-RCT (Abavisani 2019) comparing Pilates with no intervention in adults aged between 18 and 40 years at risk of anxiety (associated with employment conditions) was eligible for this comparison and contributed data relevant to one of the seven outcomes. There were no additional studies awaiting classification or ongoing that compared Pilates with no intervention in people at risk of mental health conditions that could have contributed data to the outcomes considered critical or important to this review (see Appendix C6).

Anxiety

One trial (62 participants) reported anxiety measured with the Spielberger's anxiety questionnaire at the end of treatment (8 weeks) (Abavisani 2019). The study had each participant complete the self-assessment which consisted of 20 questions evaluating obvious (state) anxiety and 20 questions evaluating hidden (trait) anxiety. State anxiety, evaluates the individuals feeling in the moment and trait anxiety, measures the individuals usual and general feelings. Determining meaningful difference can be difficult for the trait anxiety subscale as it is intended to identify susceptibility and is less responsive to change compared to state anxiety. For the state anxiety subscale, a cut point of 39-40 is suggested to detect clinically significant symptoms (209).

The results suggest an effect favouring Pilates for state anxiety (MD –5.46, 95% CI –9.08, –1.84; $p = 0.003$) *(GRADE: low)* and trait anxiety (MD –10.52, 95% CI –14.29, –6.75; $p < 0.00001$) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and its was not judged to be at high risk of bias.

Comparison 2 (vs other)

No studies were identified comparing Pilates with 'other' interventions in people at risk of mental health conditions.

D8.2 Prevention of metabolic disorders or weight problems associated with sedentary behaviour

D8.2.1 List of studies

An overview of the PICO criteria of included studies for prevention of metabolic disorders or weight problems associated with sedentary behaviour is provided in Table D-40. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-46 Overview of PICO criteria of included studies: Prevention of metabolic disorders or weight problems associated with sedentary behaviour

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Garcia-Soidan 2014 (210)	Quasi-RCT	At-risk adults (sedentary, 40 to 60 years)	Pilates exercises	Control (no intervention)	None	Accelerometry QoL Sleep quality
Pilates versus 'other' intervention*						
Sahinci Gokgul 2017 (211)	Quasi-RCT	At-risk adults (sedentary, women 25 to 55 years)	Pilates exercises	Cyclic exercise	None	Anthropometrics BMI Cholesterol biomarkers Flexibility Balance Fitness

Abbreviations: BMI, body mass index; QoL, quality of life

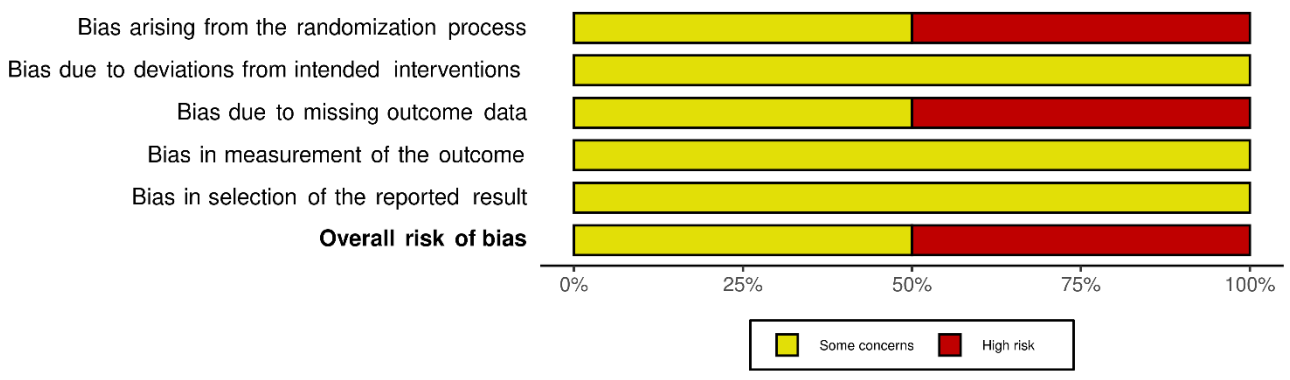
*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D8.2.2 Risk of bias summary

The risk of bias for each item in the included studies for prevention of metabolic disorders or weight problems associated with sedentary behaviour is described below and shown graphically in Figure D.27. Details are provided in Appendix E1.

Figure D.27 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Prevention of metabolic conditions or weight problems in sedentary adults



Bias arising from the randomisation process

One study (Garcia-Soidan 2014) reported random allocation supported by no significant differences in baseline characteristics, but the absence of concealment allocation raised some concerns. One study (Sahinci Gokgul 2017) reported random allocation but did not provide sufficient details and, in the absence of concealment allocation and potential group differences based on pre-test measurements, was judged to be at high risk of bias for this domain.

Bias due to deviations from the intended intervention

Garcia-Soidan 2014 reported ~20% dropout and deviation. Deviations were balanced between the two groups and were due to participants not adhering to tracking machines. Dropouts were slightly greater among the Pilates group than the control group and reasons were not provided, raising some concerns. Sahinci Gokgul 2017 reported two participants did not complete the assigned intervention, however, was judged to have some concerns due to not reporting the number of participants allocated to the intervention and control groups and insufficient information regarding discontinuation or deviation from trial.

Bias due to missing outcome data

One study (Garcia-Soidan 2014) was judged to have some concerns due to the missing outcome data being unbalanced between the two groups. Sahinci Gokgul 2017 was judged to be at high risk of bias for this domain as authors did not report on missing outcome data.

Bias in the measurement of the outcome

Two studies (Garcia-Soidan 2014, Sahinci Gokgul 2017) were assessed to have some concerns regarding the measurement of outcomes. Participants or outcome assessors were not blinded, and primary outcomes were subjective, results of which could be influenced by knowledge of the intervention.

Bias in selection of the reported result

Two studies (Garcia-Soidan 2014, Sahinci Gokgul 2017) reported all eligible specified results and, in the absence of an available protocol, were judged to be at some concerns for this domain.

D8.2.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people at risk of metabolic disorders or weight problems associated with sedentary behaviour are listed in Table D-47.

Table D-47 Outcomes considered by the NTWC to be critical or important for decision-making: Prevention of metabolic disorders or weight problems associated with sedentary lifestyles

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Garcia-Soidan 2014*
Functional/ physical performance	Accelerometry	Critical	Yes	✓
Quality of life	SF-36, global	Critical	Yes	✓
Sedentary behaviour	No measures reported in eligible studies	Important	No	?
Physical functioning	No measures reported in eligible studies	Important	No	?
Glycaemic control	No measures reported in eligible studies	Important	No	?
Cardiovascular disease risk	High density lipoprotein levels AND low-density lipoprotein levels	Important	No	?
Anthropometrics / body composition	Circumference of hip AND Circumference of waist	Important	No	?

Abbreviations: SF-36; 36-item short form survey

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

One quasi-RCT (Garcia-Soidan 2014) comparing Pilates with no intervention in people at risk of metabolic disorders or weight problems associated with sedentary behaviour was eligible for this comparison and contributed data relevant to two of the seven outcomes.

There were no additional studies awaiting classification or ongoing that compared Pilates with no intervention in people at risk of at risk of metabolic disorders or weight problems that could have contributed data to the outcomes considered critical or important to this review (see Appendix C6).

Functional/ physical performance

One study (99 participants) reported general activity using an accelerometer, which participants were required to wear for the seven days before and after the study (Garcia-Soidan 2014). The reliable and valid device measures activity by count/minute, accounting for individual differences in posture and gait (212). Different cut points have been proposed to determine physical activity intensity. Furthermore, different population groups (e.g. overweight, cardiovascular disease) may have different patterns of engaging in physical activity which can lead to substantial heterogeneity (212). The authors considered 2020 counts/minute or more as indicative of moderate-to-vigorous physical activity (assumed to be per day), presenting an average count for the full 7 days.

The results showed no difference in activity count in the Pilates group when compared with the control group (MD –422.0, 95% CI – 3770.23, 2926.23; p = 0.80) *(GRADE: very low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data. The study was not judged to be at high risk of bias.

Quality of life

One study (99 participants) evaluated quality of life using the SF-36 questionnaire at the end of treatment (12 weeks) (Garcia-Soidan 2014). The multidimensional questionnaire comprises 36 items measuring eight subscales that assess two distinct components, the physical component and the mental component. There is no single measure of health-related quality of life provided by the SF-36 questionnaire and is reported as inappropriate by the developers of the questionnaire, although some researchers may extrapolate measures to provide a single score (213). Given this, determining a clinically importance difference can be difficult. Individual scores for each of the eight domains were provided and are summarised on a scale from 0 (worse) to 100 (best). Garcia-Soidan 2014 did not report a global SF-36 score.

The results showed significant improvement in the Pilates group compared with the control across three of the four domains associated with physical wellbeing: physical functioning (MD –9.80; 95% CI –12.60, –7.00; $p < 0.00001$) *(GRADE: low)*, role-physical (MD –9.20; 95% CI –11.76, –6.64; $p < 0.00001$) *(GRADE: low)* and general health perceptions (MD –18.30; 95% CI –21.43, –15.17; $p < 0.00001$) *(GRADE: low)* as well as significant improvement in three of four domains associated with mental wellbeing: vitality (MD –17.40; 95% CI –18.58, –16.22; $p < 0.00001$) *(GRADE: low)*, role-emotional (MD –31.20; 95% CI –48.65, –13.75; $p = 0.0005$) *(GRADE: low)*, and social functioning (MD –12.40; 95% CI –16.29, –8.51; $p < 0.00001$) *(GRADE: low)*.

The effect on the mental health domain did not reach statistical significance (MD –16.30; 95% CI –33.74, 1.14; $p = 0.07$) *(GRADE: low)*, and an effect favouring the control was reported for bodily pain (MD 12.00; 95% CI 9.52, 14.48; $p < 0.0001$) *(GRADE: low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data. The study was not judged to be at high risk of bias.

Comparison 2 (vs other)

One study (Sahinci Gokgul 2017) was identified comparing Pilates with 'other' interventions in people at risk of metabolic disorders or weight problems associated with sedentary behaviour and contributed data relevant to two of the seven outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

D8.3 Prevention of age-related physical and mental decline

D8.3.1 List of studies

An overview of the PICO criteria of included studies for the prevention of age-related physical and mental decline in at-risk people is provided in Table D-48. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-48 Overview of PICO criteria of included studies: Prevention of age-related physical and mental decline

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Curi 2018 (214, 215)	RCT	Adults > 60 years (sedentary women, BMI > 24)	Pilates exercises (mat & equipment)	Control (no intervention)	None specified	Life satisfaction Sleep quality General health (mental) Senior fitness test
de Andrade Mesquita 2015 (216-219)	RCT	Adults > 60 years (sedentary women)	Pilates exercises (mat & equipment)	Control (usual activities) OR Proprioceptive Neuromuscular Facilitation^	None specified	Postural control Functional mobility Balance Flexibility Isokinetic muscle strength
Gandolfi 2020 (220)	NRSI	Adults > 60 years (sedentary women)	Pilates exercises (mat & equipment)	Control (no intervention)	None specified	Quality of life Biological markers (bone remodelling, physical health)
Irez 2011 (221)	Quasi-RCT	Adults > 60 years (sedentary women)	Pilates exercises (mat & equipment)	Control (no intervention)	None specified	Dynamic balance Flexibility Muscle strength (hips) Reaction Time
Liposcki 2019 (222)	Quasi-RCT	Adults > 60 years (sedentary women)	Pilates exercises (mat & equipment)	Control (usual activities)	None specified	Quality of life
Pilates versus 'other' intervention*						
Aibar-Almazan 2019 (223-225)	RCT	Adults > 60 years (sedentary women, BMI > 24)	Pilates exercises (mat & equipment)	Education and usual activities	None specified	Balance confidence Fear of falling Sleep quality Anxiety Depression Fatigue

Abbreviations: BMI, body mass index; NRSI, non-randomised study of interventions; RCT, randomised controlled trial

*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

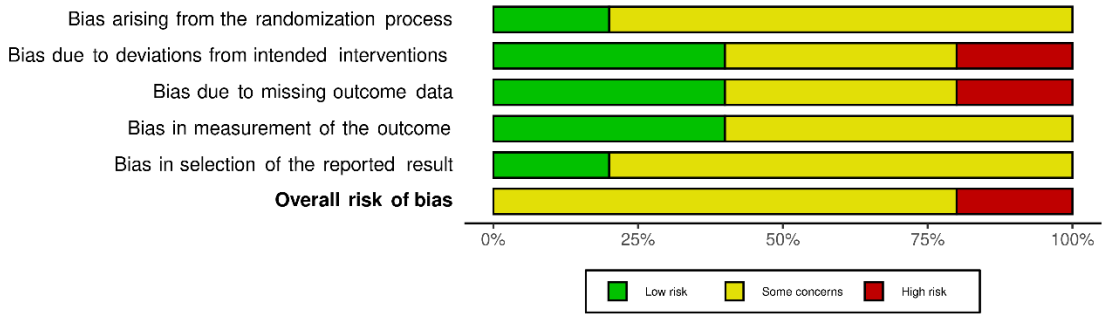
^ Study included three groups. The inactive control is considered in the evidence synthesis.

D8.3.2 Risk of bias summary

Randomised controlled trials

The risk of bias for each item in the included RCTs for prevention of age-related physical and mental decline in at-risk people is described below and shown graphically in Figure D.28. Details are provided in Appendix E1.

Figure D.28 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Prevention of age-related decline



Bias arising from the randomisation process

One study (Aibar-Almazan 2019) provided sufficient information on the randomisation and allocation concealment process and was at low risk of bias for this domain. Four studies (Curi 2018, de Andrade Mesquita 2015, Irez 2011, Liposcki 2019) had some concerns raised due to missing information about the method of allocation concealment but reported baseline characteristics appeared balanced between groups.

Bias due to deviations from the intended intervention

Two studies were at low risk of bias for this domain, with any discontinuation or deviations judged to be unrelated to the trial context (Aibar-Almazan 2019, Curi 2018). Two studies had some concerns raised relating to dropouts or withdrawals being unbalanced between the groups (de Andrade Mesquita 2015, Irez 2011). One study was assessed to be at high risk of bias as the deviations was unbalanced between groups and participants who did not attend more than 90% sessions were excluded from the analysis (Liposcki 2019).

Bias due to missing outcome data

Two studies were at low risk of bias for this domain, with outcome data available for all (or nearly all) randomised participants (Aibar-Almazan 2019, Curi 2018). Two studies had some concerns raised for this domain, with missingness of the data considered not likely to substantially impact the results (de Andrade Mesquita 2015, Irez 2011). One study was assessed to be at high risk of bias as missingness of the data was considered likely to substantially impact the true value (Liposcki 2019).

Bias in the measurement of the outcome

Two studies were at low risk of bias for this domain, with measurement of the outcome data considered appropriate and not likely to be influenced by knowledge of the intervention received (de Andrade Mesquita 2015, Irez 2011). Three studies had some concerns raised because the participant-reported outcomes could be influenced by knowledge of the intervention received (Aibar-Almazan 2019, Curi 2018, Liposcki 2019).

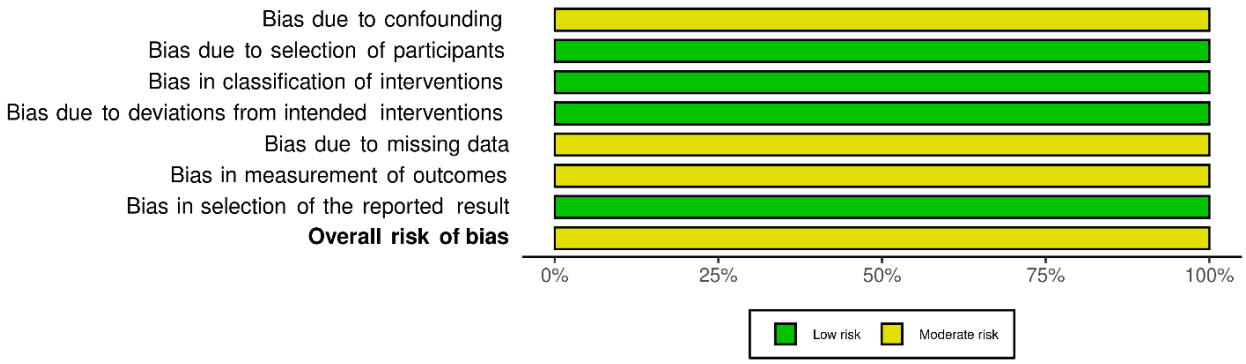
Bias in selection of the reported result

One study reported all eligible pre-specified results and were judged to be of low risk for this domain (de Andrade Mesquita 2015). Four studies had some concerns raised because analysis intentions were not available or not described in sufficient detail to enable an assessment (Aibar-Almazan 2019, Curi 2018, Irez 2011, Liposcki 2019).

Non-randomised studies of interventions

The risk of bias for each item in the included NRSIs for prevention of age-related physical and mental decline in at-risk people is described below and shown graphically in Figure D.29 (details are provided in Appendix E2).

Figure D.29 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included NRSIs – Prevention of age-related decline



Bias due to confounding

Gandolfi 2020 enrolled of a group of elderly women based on prespecified criteria (age, BMI), which reduced the potential for confounding. The study had some concerns raised about potential confounding factors which were not adjusted for in the analysis.

Bias of selection of participants into the study

Gandolfi 2020 was judged to be at low risk of bias for this domain, with all eligible participants invited to participate in the study and the start of interventions coincided.

Bias in classification of interventions

Gandolfi 2020 provided a clear description and definition of intervention groups prior to enrolment and were assessed to be at low risk for this domain.

Bias due to deviations from intended interventions

Gandolfi 2020 was judged to be at low risk of bias for this domain with any discontinuation or deviations judged to be unrelated to the trial context.

Bias due to missing data

One study (Gandolfi 2020) had some concerns raised about the missingness of the data that could influence the true value of the outcome.

Bias in measurement of outcomes

One study (Gandolfi 2020) was judged to be at moderate risk of bias for this domain related to the knowledge of the intervention received that could potentially influence the participant reported outcomes.

Bias in selection of the reported result

Gandolfi 2020 was judged to be at low risk of bias for this domain as all reported results appeared to correspond to all intended outcomes.

D8.3.3 Effect of intervention (physical decline)

Outcomes considered by the NTWC to be critical or important for decision-making the prevention of age-related physical decline are listed in Table D-49.

Table D-49 Outcomes considered by the NTWC to be critical or important for decision-making: Prevention of age-related physical decline

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Curi 2018*	*de Andrade Mesquita 2015*	*Gandolfi 2020*	*Irez 2011*	*Liposcki 2019*
Functional mobility	6-minute walk test (or TUG)	Critical	Yes	✓	✓	--	?	--
Physical functioning/ ADL	GLADM – composite (or SF-36 – physical functioning)	Critical	No	--	--	✓	--	✓
Quality of life	WHOQOL – total score (or SF-36 – General health perceptions)	Critical	Yes	--	--	--	--	--
Balance	Berg balance scale (or CTSIB)	Critical	Yes	--	✓	--	--	--
Pain	Geriatric Pain Measure (or SF-36 – bodily pain)	Important	No	--	--	✓	--	✓
Aerobic capacity/ fitness	Fullerton Functional Fitness Test (or aerobic endurance)	Important	No	?	--	--	--	--
Strength	Isokinetic muscular strength	Important	No	--	--	--	--	--

Abbreviations: ADL, Activities of daily living; CTSIB, clinical test of sensory interaction on balance; 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); SF-36; 36-item short form survey; TUG, Timed Up and Go; WHOQOL, World Health Organization quality of life questionnaire

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Five studies comparing Pilates with control (no intervention or usual activities) in sedentary older women at-risk of age-related physical or mental decline were eligible for this comparison. Three studies (Curi 2018, de Andrade Mesquita 2015, Liposcki 2019) and one NRSI (Gandolfi 2020) contributed data relevant to four of the seven outcomes. One study (Irez 2011) did not measure or assess any outcomes considered critical or important to this review and it is unclear if there is any missing (non-reporting) data.

There were five additional studies published in a language other than English (awaiting classification) and two ongoing studies that compared Pilates with no intervention in people at risk of age-related physical or mental decline (total 193 participants), that could have contributed data to some of the outcomes considered critical or important to this review (see Appendix C6).

Functional mobility

Two studies (total 102 participants) reported functional mobility measured using the Timed Up and Go (TUG) test at end of treatment (4 or 16 weeks) (Curi 2018, de Andrade Mesquita 2015). Data were missing from one study published in a language other than English (30 participants) (see Appendix C6).

Developed for older adults (aged 70 to 84 years), the TUG test has been found to be a sensitive and specific measure for predicting the risk of falls (rather than ruling out), with a TUG time over 13.5 seconds indicating a high risk of falling (226). No MCID has been established in older adults, with the minimal detectable change reported to be 3.5 seconds in people with Parkinson's disease (67) and 2.9 seconds (68) in people with chronic stroke.

The pooled results showed an effect favouring Pilates when compared with the control group (MD –3.75; 95% CI –8.33, 0.84; $p < 0.0001$; $I^2 = 94\%$) *(GRADE: very low)*. Many of the participants were not at high risk of falling at baseline (TUG time is less than 13.5 seconds), therefore the clinical important of the results are difficult to interpret. Due to a possible ceiling effect, the tests ability to detect improvements in balance in people not at risk of falling maybe limited (226).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as there were no studies judged to be at high risk of bias included in the analysis.

Quality of life – Physical component score

One RCT and one NRSI (total 60 participants) reported physical wellbeing measured with the SF-36 at the end of treatment (20 or 26 weeks) (Gandolfi 2020, Liposcki 2019). Data were missing from one study published in a language other than English (21 participants) (see Appendix C6). Individual scores for each of the four domains relating to physical wellbeing were provided and are summarised on a scale from 0 (worse) to 100 (best).

The pooled results showed an effect favouring Pilates when compared with the control group for three of the four domains: physical functioning[12] (MD –30.30; 95% CI –38.98, –21.63; $p < 0.00001$; $I^2 = 0\%$) *(GRADE: low)*, role-physical[13] (MD –52.09; 95% CI –69.31, –34.87; $p < 0.00001$; $I^2 = 7\%$) *(GRADE: low)* and general health perceptions[14] (MD –12.65; 95% CI –23.34, –1.97; $p = 0.02$; $I^2 = 98\%$) *(GRADE: very low)*. The effect

[12] SF-36 physical functioning subscale is a 10-item measure of physical limitation in a range of activities from vigorous exercise to performing self-care activities

[13] SF-36 role-physical subscale contains four items that measures limitations in various roles, including work and daily activities

[14] SF-36 general health perceptions subscale contains five items that evaluates a persons' physical health problems and their confidence in progression to better or worse health.

of Pilates on bodily pain[15] was variable (MD –23.36; 95% CI –62.06, 15.34; $p = 0.24$; $I^2 = 66\%$) *(GRADE: very low)*.

When the RCT and NRSI results were considered separately (sensitivity analysis to examine the impact of the RCT at high risk of bias and/or the inclusion of the NRSI), there were no important differences in the results for physical functioning, role-physical or general health perceptions, but both studies showing an effect favouring Pilates for bodily pain.

The RCT results suggest an effect favouring Pilates when compared with the control group for all four domains: physical functioning (MD –29.00; 95% CI –46.18, –11.82; $p < 0.0009$), role-physical (MD –40.50; 95% CI –68.19, –12.81; $p < 0.004$), general health perceptions (MD –19.40; 95% CI –30.85, –7.95; $p = 0.0009$) and bodily pain (MD –43.50; 95% CI –54.78, –32.22; $p < 0.00001$).

The NRSI results also suggest an effect favouring Pilates when compared with the control group for all four domains: physical functioning (MD –30.75; 95% CI –40.80, –20.70; $p < 0.0001$), role-physical (MD –58.75; 95% CI –79.26, –38.24; $p < 0.00001$), general health perceptions (MD –8.25; 95% CI –13.80, –2.70; $p = 0.004$) and bodily pain (MD –4.00; 95% CI –7.10, –0.90; $p = 0.01$).

Balance

One study (38 participants) reported balance measured using the 14-item Berg Balance Scale at end of treatment (4 weeks) (de Andrade Mesquita 2015). Data were missing from one study published in a language other than English (45 participants) (see Appendix C6). In most of the 14-items, the subject is asked to maintain a given position for a specific time, with each item consisting of a five-point ordinal scale ranging from 0 to 4. Total scores range from 0 (low balance stability) to 56 (high balance stability). In elderly people, a score of less than 45 indicates individuals may be at greater risk of falling (64) however the test may be subject to ceiling effects in community dwelling older adults (121). A minimum detectable clinical difference in older adults is proposed to be 6.5 points (122).

The results showed an effect favouring Pilates when compared with the control group (MD – 5.00; 95% CI – 6.62, –3.38; $p < 0.00001$) *(GRADE: very low)*. It was not possible to measure a large improvement because scores were so close to the maximum (i.e. ceiling effect).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and it was not judged to be at high risk of bias.

Aerobic capacity/ fitness

One study (64 participants) reported aerobic capacity measured as part of the senior fitness test at the end of treatment (16 weeks) (Curi 2018). Data were missing from one study published in a language other than English (21 participants) (see Appendix C6). The senior fitness test includes measures of common activities such as getting up from a chair (Chair-Stand Test), bending and stretching (Sit and Reach test, Back Scratch test), walking (6-minute walk test), as well as testing strength (arm-curl test) and agility (Timed Up and Go) (227). The 6-minute walk test is used to assess aerobic capacity and endurance, with the distance covered over the 6-minute period used to assess changes in performance capacity.

The results showed little or no effect (MD –1.50; 95% CI –3.03, 0.003, $p = 0.06$), however we are unable to interpret the results. The study mentions the 6-minute walk test, which should be measured in distance (higher is better), but the reported results are listed as minutes (higher scores means worse aerobic

[15] SF-36 bodily pain subscale contains two items that assess pain severity and pain interference.

endurance) and do not correlate with the expected walking distance of older adults (116) (around 500 to 600 metres in women between 60 to 64 years). It is not clear if this is an error in reporting or if a different test was used (e.g. 10M walk test). Given the discrepancy, we have reported the result below but cannot interpret the direction or size of the effect (results are not included in the GRADE summary of findings).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias, as only one study contributed data and it was not judged to be at high risk of bias.

Comparison 2 (vs other)

Two studies (Aibar-Almazan 2019, de Andrade Mesquita 2015) comparing Pilates with another intervention in sedentary older women at-risk of age-related physical decline were eligible for this comparison. One study (de Andrade Mesquita 2015) contributed data relevant to two of the seven outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

D8.3.4 Effect of intervention (mental decline)

Outcomes considered by the NTWC to be critical or important for decision-making the prevention of age-related mental decline are listed in Table D-50.

Table D-50 Outcomes considered by the NTWC to be critical or important for decision-making: Prevention of age-related mental decline

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Curi 2018*	*de Andrade Mesquita 2015*	*Gandolfi 2020*	*Irez 2011*	*Liposcki 2019*
Quality of life	WHOQOL – total score (or SF-36 – MCS)	Critical	Yes	--	--	✓	--	✓
General (mental) health	GHQ-12	Important	Yes	✓	--	--	--	--
Emotional wellbeing	Geriatric Depression Scale-Short form (or HADS)	Important	No	--	--	--	--	--
Sleep	PSQI – Global score	Important	Yes	✓	--	--	--	--
Carer burden	Zarit Burden Interview	Important	No	--	--	--	--	--
Loneliness / isolation	UCLA Loneliness scale	Important	No	--	--	--	--	--
Cognitive function	Global – MMSE Verbal fluency (Isaacs test) or Executive function (trail making test)	Important	No	--	--	--	--	--

Abbreviations: GHQ-12, 12-item General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; MCS, mental component score; MMSE, Mini Mental State Examination; PSQI, Pittsburgh Sleep Quality Index; SF-36; 36-item short form survey; UCLA, WHOQOL, World Health Organization quality of life questionnaire

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Five studies comparing Pilates with control (no intervention or usual activities) in sedentary older women at-risk of age-related mental decline were eligible for this comparison. Two RCTs (Curi 2018, Liposcki 2019) and one NRSI (Gandolfi 2020) contributed data relevant to three of the seven outcomes.

There were five additional studies published in a language other than English (awaiting classification) and two ongoing studies that compared Pilates with no intervention in people at risk of age-related physical or mental decline (total 193 participants), that could have contributed data to some of the outcomes considered critical or important to this review (see Appendix C6).

Quality of life – Mental component score

One RCT and one NRSI (total 60 participants) reported mental wellbeing measured with the SF-36 at the end of treatment (20 or 26 weeks) (Gandolfi 2020, Liposcki 2019). Individual scores for each of the four domains relating to mental wellbeing were reported and are summarised on a scale from 0 (worse) to 100 (best).

The pooled results showed an effect favouring Pilates when compared with the control group for three of the four domains: vitality[16] (MD –19.21; 95% CI –27.57, –10.84; $p < 0.00001$; $I^2 = 0\%$) *(GRADE: very low)*, role-emotional[17] (MD –46.51; 95% CI –64.73, –28.28; $p < 0.00001$; $I^2 = 0\%$) *(GRADE: very low)* and mental health[18] (MD –14.62; 95% CI –23.51, –5.74; $p = 0.001$; $I^2 = 0\%$) *(GRADE: very low)*. The effect of Pilates on role-social[19] was variable (MD –5.19; 95% CI –31.42, 21.03; $p = 0.70$; $I^2 = 89\%$) *(GRADE: very low)*.

When the RCT and NRSI results were considered separately (sensitivity analysis to examine the impact of the RCT at high risk of bias and/or the inclusion of the NRSI), there were no important differences in the results for vitality, role-emotional or mental health, but different results were observed for role-social.

The RCT results showed an effect favouring Pilates when compared with the control group for all four domains: vitality (MD –15.50; 95% CI –27.96, –3.04; $p = 0.01$), role-emotional (MD –35.10; 95% CI –67.75, –2.45; $p = 0.04$), mental health (MD –13.68; 95% CI –26.14, –1.22; $p = 0.03$) and role-social (MD –19.30; 95% CI –34.43, –4.17; $p = 0.01$).

The NRSI results showed an effect favouring Pilates when compared with the control group for three of the four domains: vitality (MD –22.25; 95% CI –33.54, –10.96; $p = 0.0001$), role-emotional (MD –51.67; 95% CI –73.64, –29.70; $p < 0.00001$) and mental health (MD –15.60; 95% CI –28.28, –2.92; $p = 0.02$). No effect on role-social was observed (MD 7.50; 95% CI –1.63, 16.63; $p = 0.11$).

[16] The SF-36 vitality subscale has four items that measure vitality, energy level, and fatigue and is intended to be a measure of subjective well-being.

[17] The SF-36 role-emotional subscale measures role limitations due to mental health difficulties, with three items that focus on amount of time spent on work or other activities, amount of work accomplished, and the care with which work is performed.

[18] The SF-36 mental health subscale has five items that measure anxiety, depression, loss of behavioural/emotional control, and psychological well-being.

[19] The SF-36 role-social subscale includes two items that measure the impact of physical and mental health on social functioning.

General health perceptions

One study (64 participants) reported general health perceptions measured with the GHQ-12 at the end of treatment (16 weeks) (Curi 2018). The GHQ-12 is intended to screen for general (non-psychotic) mental health problems among primary care patients (228). It consists of 12-items that measure concerns related to mental health, with responses measured on a four-point scale using a timeframe of "in the last two weeks". Positive phrased outcomes are scored higher, with the optimal threshold in the general population varying from 1/2 to 11/12 (i.e. scores below this are considered typical), depending on whether a bimodal scoring method (0-0-1-1) (total maximum score of 12) or Likert scoring system (0-1-2-3) is used (maximum score of 36) (228). It is assumed a Likert scoring system was used (not detailed in the study report), with scores greater than 15 suggesting evidence of distress, and scores greater than 20 considered severe problems with psychological distress.

The results showing an effect favouring Pilates when compared with the control (no intervention) group (MD –5.08; 95% CI –7.73, –2.43; *p* = 0.0002) *(GRADE: low)*, however given the control group are within the range of typical, the clinical relevance of the observed improvement is not important.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias as only one study contributed data and it was not judged to be at high risk of bias.

Sleep quality

One trial (64 participants) reported sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI) at the end of treatment (16 weeks) (Curi 2018). The PSQI is a nine-item[20] questionnaire that assesses the sleep quality of an individual in the previous month. It assesses seven sleep components including subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disorder (sleep fragmentation), use of sleeping medication, and daytime dysfunction (191). Each item is scored (range from 0 to 3) with the total global score ranging from 0 (no problems) to 21 (severe problems). A score of five or more is associated with poor sleep quality.

The results showed difference between treatment groups when comparing Pilates with control (no intervention) (MD –1.99; 95% CI –4.25, 0.27; *p* = 0.08) *(GRADE: very low)*.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias as only one study contributed data and it was not judged to be at high risk of bias.

Comparison 2 (vs other)

Two studies (Aibar-Almazan 2019, de Andrade Mesquita 2015) comparing Pilates with another intervention in sedentary older women at-risk of age-related mental decline were eligible for this comparison. One study (Aibar-Almazan 2019) contributed data relevant to two of the seven outcomes.

Available data are presented in Appendix F2 Supplementary outcome data.

[20] The PSQI can include five additional partner-rated questions (Item 10) that do not contribute to the total PSQI score (not used in the included study).

D8.4 Falls prevention

D8.4.1 List of studies

An overview of the PICO criteria of included studies for falls prevention is provided in Table D-51. Study details, including all outcome domains and measures reported by the included studies are provided in Appendix F1. Outcome data for critical or important outcomes are provided in Appendix F2.

Table D-51 Overview of PICO criteria of included studies: Falls prevention

STUDY ID	Study design	POPULATION	INTERVENTION	COMPARATOR	CO-INTERVENTION	OUTCOME DOMAINS
Pilates versus control (no intervention, waitlist, inactive usual care)*						
Barker 2016 (229, 230)	RCT	Healthy adults (> 60 years) at risk of falls (prior history)	Pilates (equipment)	Control (usual care)	Home exercise	Falls Balance Postural control Muscle strength Flexibility
Roller 2018 (231)	RCT	Healthy adults (> 65 years) at risk of falls (prior history or impaired balance)	Pilates (reformer)	Control (no intervention)	None specified	Balance Postural control Flexibility
Pilates versus 'other' intervention*						
Josephs 2016 (232)	RCT	Healthy adults (> 65 years) at risk of falls (prior history or impaired balance)	Pilates (equipment)	Conventional balance exercises	None specified	Balance Balance confidence

Abbreviations: RCT, randomised controlled trial

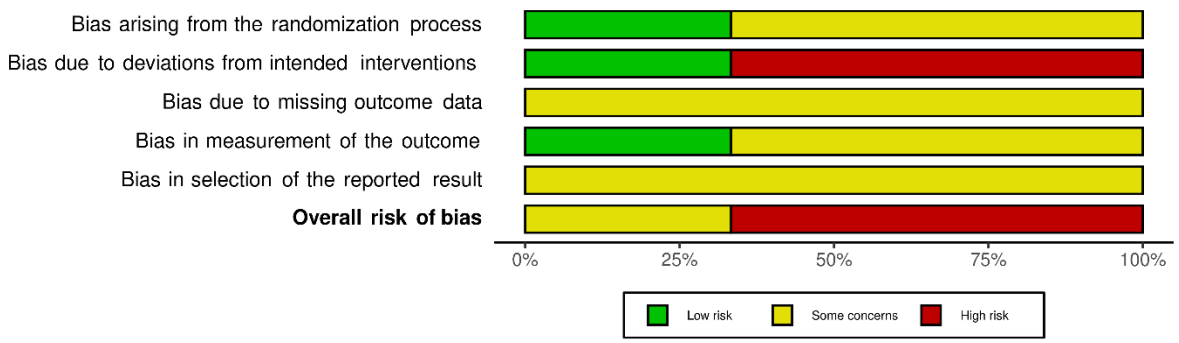
*Studies that compared Pilates with an inactive control were eligible for inclusion in the evidence synthesis and are included in the Summary of findings tables if they reported outcomes considered critical or important to this review.

**Studies that compared Pilates with an active intervention are included in the supplementary outcome tables (Appendix F2) if they reported data for outcomes considered critical or important to this review.

D8.4.2 Risk of bias summary

The risk of bias for each item in the included studies for falls prevention is described below and shown graphically in Figure D.30. Details are provided in Appendix E1.

Figure D.30 Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included RCTs – Falls prevention



Bias arising from the randomisation process

One study (Barker 2016) was judged to be at low risk of bias as the random sequence generation and allocation concealment processes were well described. Two studies (Josephs 2016, Roller 2018) were assessed to have some concerns due to a lack of information regarding the allocation concealment process. Any differences in baseline characteristics were not deemed to indicate a problem with the randomisation process.

Bias due to deviations from the intended intervention

Two studies (Barker 2016, Josephs 2016) were assessed at high risk of bias due to high and uneven rates of drop out between the intervention groups. One study (Roller 2018) was assessed at low risk of bias for this domain, any deviations from the intended intervention were considered in line with what would occur in clinical practice.

Bias due to missing outcome data

Three studies (Barker 2016, Josephs 2016, Roller 2018) were assessed to have some concerns for this domain due to large amounts of missing data (over 20% missing) that could affect the true values of the outcome. In one study (Barker 2016) the authors reporting correcting for the missing data (using the last observation carried forward method) but details about the influence of the missingness of the data were not provided.

Bias in the measurement of the outcome

Two studies (Barker 2016, Josephs 2016) were assessed to have some concerns for this domain due to the self-reported nature of some outcomes and the lack of blinding of the outcome assessors, which could be subject to performance bias. Roller 2018 was assessed to be at low risk of bias as the outcome assessor was blinded for most outcomes. A second (non-blinded) outcome assessor measured objective outcomes.

Bias in selection of the reported result

All studies were assessed to have some concerns for this domain due to the lack of information regarding the pre-specified analysis plan. It was not considered likely that results were not reported based on multiple eligible domains or analyses.

D8.4.3 Effect of intervention

Outcomes considered by the NTWC to be critical or important for decision-making in people at risk of falls are listed in Table D-52.

Table D-52 Outcomes considered by the NTWC to be critical or important for decision-making: Falls prevention

Outcome domain	*Measured with*	*consensus rating*	*Data available for main comparison?*	*Barker 2016*	*Roller 2018*
Falls	Number of falls (count) AND Number of falls (rate per 1,000 person days)	Critical	Yes	✓	?
Falls injury	Falls injury rate (per 1,000 person days)	Critical	Yes	✓	?
Balance	Berg Balance Scale	Critical	Yes		✓
Physical functioning/ functional capacity	Latin American Development Group for Elderly	Critical	No	--	--
Quality of life	SF-36 Global Score	Critical	No	--	--
Functional mobility	6 Minute Walk Test (or Timed Up and Go)	Important	Yes	✓	✓
Psychological consequences	Falls Efficacy Scale – International	Important	No	--	--

Abbreviations: SF-36, 36-item short form survey

✓ A study result is available for inclusion in the synthesis

X No study result is available for inclusion, (probably) because the P value, magnitude or direction of the results generated were considered unfavourable by the study investigators

--No study result is available for inclusion, (probably) because the outcome was not assessed, or for a reason unrelated to the P value, magnitude or direction of the results

? No study result is available for inclusion, and it is unclear if the outcome was assessed in the study

Main comparison (vs control)

Two studies (Barker 2016, Roller 2018) comparing Pilates with no intervention (or inactive control) in people at risk of falls were eligible for this comparison and contributed data to four of the seven outcomes. There was one additional study published in a language other than English (awaiting classification) and three ongoing studies (complete but results not available) that compared Pilates with no intervention in people at risk of falls (total 241 participants) that could have contributed data to some of the outcomes considered critical or important to this review (see Appendix C6).

Falls

One study (Barker 2016) reported rate of falls per 1,000 person days at the end of follow-up (24 weeks). Results were presented by the study authors as incidence ratios (95% CI) for the individual intervention groups, with no standard deviation provided. It was considered likely that the results presented were calculated on transformed values, making them impossible to interpret.

The results as reported by the study authors showed no difference in rate of falls between Pilates and control at end of follow-up (incidence rate ratio 1.17; 95% CI 0.43, 3.16; $p = 0.754$) *(GRADE: very low)*. An IRR > 1 indicates that the incident rate is greater in the Pilates group compared to control.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias as only one study contributed data and its removal would leave no result.

Fall injury

One study (Barker 2016) reported rate of fall injury per 1,000 person days at the end of follow-up (24 weeks). Results were presented by the study authors as incidence ratios (95% CI) for the individual

intervention groups, with no standard deviation provided. It was considered likely that the results presented were calculated on transformed values, making them impossible to interpret.

The results as reported by the study authors showed no difference in rate of falls injury between Pilates and control at end of follow-up (incidence rate ratio 0.36; 95% CI 0.09, 1.38; *p* = 0.136) *(GRADE: very low)*. An IRR < 1 indicates that the incident rate is reduced in the Pilates group compared to control.

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias as only one study contributed data and its removal would leave no result.

Balance

One study (56 participants) reported balance using the Berg Balance Scale (BBS) at end of treatment (10 weeks) (Roller 2018). One additional study published in a language other than English (awaiting classification) (30 participants) reported this outcome (see Appendix C6).

In most of the 14-items, the subject is asked to maintain a given position for a specific time, with each item consisting of a five-point ordinal scale ranging from 0 to 4. Total scores range from 0 (low balance stability) to 56 (high balance stability). In elderly people, a score of less than 45 indicates individuals may be at greater risk of falling (64) however the test may be subject to ceiling effects in community dwelling older adults (121). The minimum detectable clinical difference in older adults proposed by the study was 6.5 points (122). Given the mean baseline BBS score (50.63 in Pilates, 52.11 in control), this proposed minimum detectable change would result in scores above the maximum (56 points). It is considered that this measure is unlikely to be sufficiently sensitive to detect true change in balance stability for this population.

The results showed no difference in balance between Pilates and control at end of treatment (MD –0.52; 95% CI –2.03, 0.99), but it was not possible to measure a bigger improvement because scores were so close to the maximum (i.e. ceiling effect).

No sensitivity analysis was conducted that examined the impact of RCTs judged to be at a high risk of bias as only one study contributed data and it was not judged to be at high risk of bias.

Functional mobility

Two studies (Barker 2016, Roller 2018; 104 participants) reported functional mobility using the Timed Up and Go (TUG) test at end of treatment (10-12 weeks). One ongoing study (60 participants) with results (not published) for this outcome (see Appendix C6). Developed for older adults (aged 70 to 84 yeas), the test has been found to be a sensitive and specific measure for predicting the risk of falls, with TUG time over 13.5 seconds associated with fall risk (233). The MCID in older adults in not established. A minimal detectable change is reported to be 3.5 seconds in people with Parkinson's disease (67) and 2.9 seconds (68) in people with chronic stroke.

Results show no difference between Pilates and control for functional mobility assessed with TUG test (MD –0.65; 95% CI –1.94, 0.64; *p* = 0.26; I^2 = 0%).

Sensitivity analysis showed no important difference in the observed effect when the RCT judged to be at a high risk of bias (Barker 2016) was not included in the analysis (MD –0.56; 95% CI –2.57, 1.45; *p* = 0.59).

Appendix E Risk of bias forms

This appendix documents the risk of bias judgements made on studies that met the prespecified inclusion criteria for a systematic review on the effect of Pilates for preventing and treating any health condition and were conducted in populations prioritised for inclusion in the evidence synthesis.

Risk of bias was assessed using the most appropriate risk of bias assessment tool (see www.riskofbias.info) according to the type of study as follows:

- RCTs: Revised Cochrane Risk of Bias tool v2.0 (234, 235).
- NRSIs: ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions (236)

Where possible, the assessment was based on the primary outcome for that study (or that for which the study was powered). In some circumstance, two assessments were made to account for risk of bias associated with different (e.g., subjective and objective) outcome measures.

E1 Randomised controlled trials

Appendix E1 (see attachment E1) lists the included RCTs and quasi-RCTs (for priority populations) in order of ICD-11 category. Studies within the ICD-11 category are then ordered by the prioritised condition and listed alphabetically. For each study there are two columns: column one is the judgement applied to each signalling question[21] associated with each risk of bias domain (answered as yes, partial yes, no, partial no, no information or not applicable); column two is a comment that briefly explains the reasoning that underpins the judgement.

E2 Non-randomised studies of interventions

Appendix E2 (see attachment E2) lists the included NRSIs (for priority populations) in order of ICD-11 category. Studies within the ICD-11 category are then ordered by the prioritised condition and listed alphabetically. For each study there are two columns: column one is the judgement applied to each signalling question[22] associated with each risk of bias domain (answered as yes, partial yes, no, partial no, no information or not applicable); column two is a comment that briefly explains the reasoning that underpins the judgement.

[21] see https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2

[22] https://www.riskofbias.info/welcome/home/current-version-of-robins-i/robins-i-detailed-guidance-2016

Appendix F Characteristics of included studies

This appendix documents the data extracted from studies that met the prespecified inclusion criteria for a systematic review on the effect of Pilates for preventing and treating any health condition and were conducted in populations prioritised for inclusion in the evidence synthesis.

All extracted data is presented, including that which was not synthesised in the main report.

F1 Study details

Appendix F1 (see attachment F1) lists the characteristics of each included study (for priority populations) in order of ICD-11 category. Studies within the ICD-11 category are then ordered by the prioritised condition and listed alphabetically.

For each study, the data extraction has included (but was not limited to) the following characteristics: study design, year conducted, setting and location, participant inclusion criteria, intervention and comparator characteristics (including number of treatment sessions, program duration, co-interventions), outcomes (including measurement method and timing), and funding sources.

Outcome domains and measures considered critical or important for inclusion in the review are highlighted with a blue box. Conversely, outcome domains and measures that were of limited importance are not highlighted.

F2 Supplementary outcome data

Appendix F2 (see attachment F2) lists the data extracted for critical or important outcomes identified in each included study (for priority populations) in order of ICD-11 category. Studies within the ICD-11 category are then ordered by the prioritised condition. Within each sheet, studies are listed by comparison (Pilates vs control or Pilates vs 'other') with the study results per outcome reported (critical or important outcome measures) that includes (but is not limited to) the following: outcome domain, timing, outcome measure, measure details, number of included participants, point estimates, *p*-value, direction of effect.

Data extracted is that reported by the study authors at the end of treatment (where possible) with footnotes included if further explanation was required (e.g., authors do not provide end-of treatment results therefore the mean change from baseline data are reported). The final column lists the risk of bias assessment for that outcome as made by the review authors (see Appendix E1 – RCTs or Appendix E2 – NRSIs).

Appendix G Differences between protocol & review

G1 Methods not implemented

There were some methods that were not implemented in the review relating to the following sections:

Studies identified in the literature search

It was intended that, if a study did not contain the required PICO information for a decision to be made regarding its eligibility, the information would be sought from the study's authors through an open-ended request. Given time and resource constraints, we did not contact authors for additional information regarding eligibility criteria.

Requests for data

Eligible primary studies not published in English, ongoing trials and studies published as conference abstracts with incomplete results were identified for inclusion and listed as either '*Ongoing*' or within the '*Studies Awaiting Classification'*. It was intended that study authors would be contacted through an open-ended request for further information, and, if available, the study would be included in the evidence appraisal. Given time and resource constraints, we did not contact study authors for additional information regarding missing data.

Risk of reporting bias across studies

To assess potential bias due to 'non-reporting', it was intended that funnel plots (of effect estimates against their standard errors) would be generated in RevMan 5.4 (if there were more than ten RCTs included for a PICO); with visual inspection of the funnel plot being used to look for evidence of asymmetry (suggesting small-study effects or missing results). Other possible reasons for funnel plot asymmetry were to be considered at this time (e.g. poor methodological quality, true heterogeneity, chance) (237). There were less than ten RCTs included for most PICOs (except low back pain), therefore funnel plot asymmetry was not able to be assessed. In the absence of funnel plots, non-reporting bias was suspected when the evidence was limited to a small number of small trials reporting favourable results; supplemented through inspection of outcomes reported in the 'Ongoing Studies' and 'Studies Awaiting Classification' (if available) (see Appendix B3.3).

Quantitative synthesis

The NTWC could request that data comparing Pilates with 'other' (active) intervention be synthesised (prior to provision of the first draft evaluation report), where:

i. at least two studies compare the effect of Pilates with the same active comparator, and the comparator is sufficiently homogenous across studies to support synthesis, and
ii. at least two of these studies are at low or moderate risk of bias, and
iii. the comparator represents an accepted, evidence-based 'gold standard' of care for the population in question.

No such cases were identified or requested.

Subgroup analyses and investigations of heterogeneity

We did not plan to undertake any subgroup analyses of subsets of participants within or across studies, unless there was substantial inconsistency between effect estimates. Any subgroup analysis was intended to explore possible sources of heterogeneity relating to delivery of the intervention. Studies were to be

grouped according to intervention characteristics (i.e. intensity, duration, mode of delivery) and a standard test for heterogeneity across the subgroups was to be reported. Due to time and resource constraints, we did not undertake a subgroup analysis of intervention characteristics (intensity, duration, mode of delivery) to explain statistical heterogeneity. Note that Cochrane (10.11.5.1) recommends that at least 10 studies are needed for subgroup analysis and most conditions did not meet this.

G2 Changes from protocol

There were some differences between the protocol and review relating to the following sections:

Types of participants

Additional clarification on what constitutes an 'at-risk' healthy population was made prior to data extraction and evidence synthesis, to avoid ambiguity regarding eligibility and to establish a minimum threshold requirement for inclusion. The NTWC agreed that, where a study could provide sufficient evidence of the *individual* participant being 'at-risk' then it was eligible for inclusion. This meant that studies that enrolled participants at a *population* level were not eligible unless there was some form of prespecified enrolment criteria for the otherwise healthy participants or there were baseline data that indicated all participants met a certain criterion.

For example, a study that enrolled healthy nurses from the local hospital and examined the effect on Pilates on preventing stress, anxiety or burnout was excluded, unless the study participants had been enrolled based on help-seeking behaviour (e.g., referral after visit to medical practitioner), the participants had been screened for elevated stress prior to study entry (e.g., enrolment based on a certain perceived stress score [PSS]), or baseline data suggested all participants met a preclinical condition (e.g. all participants had elevated stress validated by an established PSS cut-off at baseline). A similar example would be a study that examined falls risk in otherwise healthy older adults (aged over 65 years), with eligible studies being those in which the participants had a prior history of falls or had been judged by a clinician prior to study entry to have a balance impairment (or met frailty criteria). Where there was ambiguity, information on participants and the aim of the study was provided to the NTWC for a decision about eligibility.

Comparators

Additional clarification on what constitutes an 'inactive' or 'active' control was made prior to data extraction and evidence synthesis. Comparators that provided minimal intervention or change to the participants day-to-day activities were judged to be 'inactive', whereas comparators that required the participants involvement over the course of the study were judged to be 'active'.

For example, an education booklet or handout providing health advice that given to participants at the study start was judged 'inactive', whereas education in the form of weekly group sessions was judged 'active'. Similarly, a control group that received a weekly phone call was judged 'inactive', but if the control group received a weekly health check up with clinical advice (either at home or requiring site visits), then this was judged to be 'active'.

Outcome measures and timepoints of interest

It was intended that outcomes reported at different timepoints were to be grouped and considered as either: short term, intermediate term, long-term (or not specified); with the NTWC to decide during outcome prioritisation as to whether evidence reported at multiple timepoints would be considered critical or important for decision-making (to be considered and reported separately). During the preliminary data extraction (and prior to outcome prioritisation), it became apparent that very few studies reported anything

beyond baseline and end of treatment scores (i.e., there was minimal reporting of mid-treatment or follow-up results after completion of the Pilates programme).

A pragmatic decision was therefore made to maximise the available data eligible for inclusion, with 'end-of-treatment' outcomes being the sole timepoint of interest to be considered in the evidence synthesis (unless there was good rationale for selecting an alternative timeframe).

Studies identified in the literature search

It was intended that the lead reviewer would reinspect a random 20% sample of articles marked as excluded to ensure adherence to the *a priori* exclusion criteria. In fact, the lead reviewer screened approximately 40% of articles marked for exclusion, but the selection of articles screened in duplicate was not random, rather it was targeted towards studies excluded for the following reasons: population out of scope, comparator out of scope, outcome out of scope, or study design out of scope.

Risk of bias

It was intended that, for any included NRSI, any potential confounders or cointerventions would be identified and agreed through discussion with the NTWC prior to assessment of the risk of bias. Given the small number of NRSIs identified for inclusion, these studies were judged by the evidence review team alone.

We had stated in the protocol that NRSIs judged to be at critical risk of bias in any domain would be excluded from the reporting of results, synthesis and conclusion. We also stated that only NRSIs that are judged to be at a low to moderate risk of bias will be meta-analysed. Both these statements are true and have been adhered to in this report, however our intended approach did not specify how NRSIs judged to be at serious risk of bias in any domain were to be handled. There were three NRSIs judged to be at serious risk of bias that were identified in people with low back pain. The studies reported data for the outcomes of pain or disability. Because the intent of including NRSIs was to ensure the evidence review adequately covered the breadth of health conditions and outcomes, these studies were not included in the reporting of results, synthesis and conclusion for low back pain as RCT evidence was available for both outcomes. Data from these studies are included in Appendix F2 Supplementary outcome data.

Subgroup analyses and investigations of heterogeneity

We had specified that studies were to be stratified based on whether the participants receive instructor-led Pilates, to allow for potential subgroup analysis (and to investigate heterogeneity). However, given the small number of studies for each comparison, and a lack of studies in the main analysis specifying they were not instructor-lead, we did not stratify studies on this basis.

Summary of findings and certainty of the evidence

We had specified that the evidence from RCTs and NRSIs would be evaluated separately in the summary of findings table, but there was only one NRSI included in the evidence synthesis. A pragmatic decision was made to report of this study alongside the RCT, as it was considered to not seriously alter the results or evidence statement made.

Appendix H How comments from methodological review were addressed

Methodological review (or peer review) was conducted to appraise the methodological quality and assess the appropriateness of reporting for this systematic review (including appendices).

For reporting, the methodological review assessed the systematic review against the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) Checklist (2020) and where applicable, the MECIR (Methodological Expectations of Cochrane Intervention Reviews) manual.

The ROBIS (Risk of Bias in Systematic Reviews) tool was used to assess the methodological quality of the systematic review, to ensure it was designed and conducted in accordance with:

- NHMRC's Developing your Guideline module in NHMRC's Guidelines for Guidelines Handbook
- Cochrane Handbook for Systematic Reviews of Interventions (updated 2022)
- GRADE guidance and GRADE working group criteria for determining whether the GRADE approach was used (GRADE handbook).

The ROBIS assessment included specification and application of criteria for considering studies for the review and synthesis, search methods, data extraction and analysis, assessment of risk of bias of studies, assessment of the certainty of evidence using GRADE, and the interpretation and summary of findings.

The systematic review (including appendices) has been updated to reflect the amendments suggested by methodological review and NHMRC's Natural Therapies Working Committee, where appropriate. In summary, updates included additional information and/ or clarification of the Plain Language Summary, Executive Summary, Results sections, and Appendices, including:

- Information on minimally clinically important difference (MCID) was added to the Summary of Findings tables, where appropriate and checked. Information on MCID was also clarified in Appendix B3.1.2.
- The basis for GRADE judgements were elaborated on in both Summary of Findings Tables and Appendix B.4.1 for transparency, with GRADE judgements clarified and confirmed where appropriate. This included clarifying where sensitivity analyses were conducted for judgements about risk of bias and the inclusion of funnel plots (for low back pain) to inform judgments for publication bias.
- Included data from the Roland Morris Disability index for the disability outcome for low back pain, which changed the certainty from low to moderate and effect from none to moderate. This makes the result consistent with the previous systematic review on effects of Pilates in Low Back Pain. Additional data was also added for the SF-36 for pain in the Low Back Pain condition, which did not change the certainty or effect size; and included the SF-6D for quality of life in Low Back Pain. Other conditions were checked, and no additional changes were required. The omission of these measures in the earlier version of the report was an artefact of piloting the outcome prioritisation process, the process has now been rectified across reviews.
- The process for assigning studies to conditions has been clarified in the main report and appendices.
- It has been made clearer throughout the main report and Appendices which studies were included in analysis and why studies were not included in the analysis.

A detailed record of responses to all comments indicating changes that were made, was provided to the NHMRC together with the amended Report and Appendices documents for transparency.

References

1. Department of Health. The 2015 Review of the Australian Government Rebate on Private Health Insurance for Natural Therapies Canberra: Australian Government Department of Health; 2019 [Available from: https://www.health.gov.au/internet/main/publishing.nsf/Content/phi-natural-therapies.
2. Higgins JPT, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, et al. Methodological Expectations of Cochrane Intervention Reviews. London: Cochrane; 2019. Available from: https://community.cochrane.org/mecir-manual.
3. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
4. Churchill R, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, et al. Standards for the reporting of new Cochrane Intervention Reviews. 2019 [cited 4 June 2020]. In: Methodological Expectations of Cochrane Intervention Reviews [Internet]. London: Cochrane, [cited 4 June 2020]. Available from: https://community.cochrane.org/mecir-manual/standards-reporting-new-cochrane-intervention-reviews-r1-r109.
5. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach [Internet]. 2013; (Updated October 2013). Available from: https://gdt.gradepro.org/app/handbook/handbook.html.
6. Eyigor S, Karapolat H, Yesil H, Uslu R, Durmaz B. Effects of pilates exercises on functional capacity, flexibility, fatigue, depression and quality of life in female breast cancer patients: a randomized controlled study. European journal of physical & rehabilitation medicine. 2010;46(4):481-7.
7. Martin E, Battaglini C, Groff D, Naumann F. Improving muscular endurance with the MVe Fitness Chair™ in breast cancer survivors: a feasibility and efficacy study. J Sci Med Sport. 2013;16(4):372-6.
8. Martin E, Battaglini C, Groff D, Naumann F. Improving muscular endurance with the MVe Fitness Chair in breast cancer survivors: Afeasibility and efficacy study. Journal of Science and Medicine in Sport. 2011;14:e116.
9. Zengin Alpozgen A, Razak Ozdincler A, Karanlik H, Yaman Agaoglu F, Narin AN. Effectiveness of Pilates-based exercises on upper extremity disorders related with breast cancer treatment. European Journal of Cancer Care. 2017;26(6):e12532.
10. Gajbhiye PP, Deshpande L. To compare the effects of Pilates exercises and Conventional therapy on Upper Extremity Function and Quality of Life in women with breast cancer. Indian Journal of Occupational Therapy. 2013;45(1):3-9.
11. Odynets T, Briskin Y, Perederiy A, Pityn M, Svistelnyk I. Effect of water physical therapy on quality of life in breast cancer survivors. Physiotherapy Quarterly. 2018;26(4):11-6.
12. Odynets T, Briskin Y, Yefremova A, Goncharenko I. The effectiveness of two individualized physical interventions on the upper limb condition after radical mastectomy. Physiotherapy Quarterly. 2019;27(1):12-7.
13. Odynets T, Briskin Y, Zakharina I, Yefremova A. Influence of a water physical rehabilitation program on the hemodynamic parameters in breast cancer survivors. Physiotherapy Quarterly. 2019;27(2):6-10.
14. Odynets T, Briskin Y, Zakharina I, Yefremova A. Impact of a 12-week water program on the respiratory function in breast cancer survivors. Postepy Rehabilitacji [Advances in Rehabilitation]. 2019;33(2):5-11.
15. Odynets T, Briskin Y, Todorova V. Effects of Different Exercise Interventions on Quality of Life in Breast Cancer Patients: A Randomized Controlled Trial. Integrative Cancer Therapies. 2019;18:1534735419880598.
16. Odynets T, Briskin Y, Todorova V, Bondarenko O. Impact of different exercise interventions on anxiety and depression in breast cancer patients. Physiotherapy Quarterly. 2019;27(4):31-6.
17. Odynets T, Briskin Y, Todorova V, Tyshchenko V, Bondarenko O. Effect of yoga in the modulation of heart rate variability in patients with breast cancer. Advances in Rehabilitation [Postepy Rehabilitacji]. 2019;33(4):5-11.
18. Sener HO, Malkoc M, Karadibak D, Ergin G, Yavuzsen T. The effect of clinical pilates exercises on lymphedema secondary to breast cancer treatments. Fizyoterapi Rehabilitasyon. 2015;26 (2):S161-S2.
19. Şener H, Malkoç M, Ergin G, Karadibak D, Yavuzşen T. Effects of Clinical Pilates Exercises on Patients Developing Lymphedema after Breast Cancer Treatment: A Randomized Clinical Trial. J Breast Health. 2017;13(1):16-22.
20. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life-a systematic review. J Clin Epidemiol. 2017;89:188-98.
21. Chang YJ, Lee JS, Lee CG, Lee WS, Lee KS, Bang SM, et al. Assessment of clinical relevant fatigue level in cancer. Support Care Cancer. 2007;15(7):891-6.

22. Giesinger JM, Kuijpers W, Young T, Tomaszewski KA, Friend E, Zabernigg A, et al. Thresholds for clinical importance for four key domains of the EORTC QLQ-C30: physical functioning, emotional functioning, fatigue and pain. Health Qual Life Outcomes. 2016;14(1):87.
23. Gomes CS, Pedriali FR, Urbano MR, Moreira EH, Averbeck MA, Almeida SHM. The effects of Pilates method on pelvic floor muscle strength in patients with post-prostatectomy urinary incontinence: A randomized clinical trial. Neurourology & Urodynamics. 2018;37(1):346-53.
24. Pedriali F, Gomes C, Soares L, Urbano M, Moreira E, De Almeida S. The efficacy of pilates compared to pelvic floor muscle training associated with electrical stimulation in the recovery of post-prostatectomy urinary incontinence: A randomized controlled trial. Neurourology and Urodynamics. 2014;33 (6)(6):742-3.
25. Pedriali FR, Gomes CS, Soares L, Urbano MR, Moreira EC, Averbeck MA, et al. Is pilates as effective as conventional pelvic floor muscle exercises in the conservative treatment of post-prostatectomy urinary incontinence? A randomised controlled trial. Neurourology & Urodynamics. 2016;35(5):615-21.
26. Malik RD, Cohn JA, Fedunok PA, Chung DE, Bales GT. Assessing variability of the 24-hour pad weight test in men with post-prostatectomy incontinence. Int Braz J Urol. 2016;42(2):327-33.
27. Melo KCB, Araújo FS, Cordeiro Júnior CCM, de Andrade KTP, Moreira SR. Pilates Method Training: Functional and Blood Glucose Responses of Older Women With Type 2 Diabetes. J Strength Cond Res. 2020;34(4):1001-7.
28. Torabian M, Taghadosi M, Ajorpaz N, Khorasanifar L. The effect of Pilates exercises on general health in women with type 2 diabetes. Life Science Journal. 2013;10:283-8.
29. Yucel H, Uysal O. Pilates-based mat exercises and parameters of quality of life in women with type 2 diabetes. Iranian Red Crescent Medical Journal. 2018;18 (3) (no pagination)(e21919).
30. Diabetes Australia. Maintaining a healthy weight [Available from: https://www.diabetesaustralia.com.au/food-activity/maintaining-a-healthy-weight/.
31. CADTH. Clinical Review Report: Insulin Degludec (Tresiba): (Novo Nordisk Canada Inc): Indication: For once-daily treatment of adults with diabetes mellitus to improve glycemic control [Internet]. 2017 Dec. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health. Available from: https://www.ncbi.nlm.nih.gov/books/NBK533972/.
32. Bjorner JB, Lyng Wolden M, Gundgaard J, Miller KA. Benchmarks for interpretation of score differences on the SF-36 health survey for patients with diabetes. Value Health. 2013;16(6):993-1000.
33. Dantas E, Figueira H, Emygdio R, Vale R. Functional Autonomy Gdlam Protocol Classification Pattern in Elderly Women. INDIAN JOURNAL OF APPLIED RESEARCH. 2014;4.
34. Nordin A, Taft C, Lundgren-Nilsson A, Dencker A. Minimal important differences for fatigue patient reported outcome measures-a systematic review. https://wwwmjacomau/sites/default/files/NHMRClevelsofevidence2008-09pdf. 2016;16:62.
35. Stern AF. The Hospital Anxiety and Depression Scale. Occupational Medicine. 2014;64(5):393-4.
36. Duff WR, Andrushko JW, Renshaw DW, Chilibeck PD, Farthing JP, Danielson J, et al. Impact of pilates exercise in multiple sclerosis: A randomized controlled trial: International Journal of MS Care. 20 (2) (pp 92-100), 2018. Date of Publication: 01 Mar 2018.; 2018.
37. Eftekhari E, Etemadifar M. Impact of clinical mat pilates on body composition and functional indices in female patients with multiple sclerosis. Crescent Journal of Medical and Biological Sciences. 2018;5(4):297-305.
38. Eftekhari E, Etemadifar M. Interleukin-10 and brain-derived neurotrophic factor responses to the Mat Pilates training in women with multiple sclerosis. Scientia Medica. 2018;28 (4) (no pagination)(31668).
39. Fleming KM, Coote SB, Herring MP. The feasibility of Pilates to improve symptoms of anxiety, depression, and fatigue among people with Multiple Sclerosis: An eight-week randomized controlled pilot trial: Psychology of Sport and Exercise. 45 (no pagination), 2019. Article Number: 101573. Date of Publication: November 2019.; 2019.
40. Marandi SM, Shahnazari Z, Minacian V, Zahed A. A comparison between Pilates exercise and aquatic training effects on mascular strength in women with Mulitple sclorosis. Pakistan Journal of Medical Sciences. 2013;29(1 SUPPL.):285-9.
41. Marandi S, Nejad V, Shanazari Z, Zolaktaf V. A comparison of 12 weeks of Pilates and aquatic training on the dynamic balance of women with Multiple Sclerosis. Int J Prev Med. 2013;4(Supp 1):S110-7.
42. Shanazari Z, Marandi SM, Minasian V. Effect of 12-week pilates and aquatic training on fatigue in women with multiple sclerosis. [Persian]. Journal of Mazandaran University of Medical Sciences. 2013;23(98):257-64.
43. Rezvani 1* MH. Comparing Pilates and Physioball Exercise Regimens on Balance and Motor Control in Women with Multiple Sclerosis. International Journal of Health Studies; Volume 3, Issue 2, 2017DO - 1022100/ijhsv3i2202. 2017.
44. Sisi SZH, Sadeghi H, Nabavi SM. The effects of 8 weeks of rebound therapy and Pilates practices on static and dynamic balances in males with multiple sclerosis. Advances in Environmental Biology. 2013:4290-3.

45. Abasiyanik Z, Ertekin O, Kahraman T, Yigit P, Ozakbas S. The effects of clinical Pilates training on walking, balance, fall risk, respiratory and cognitive functions in persons with multiple sclerosis: A randomized controlled trial. Multiple Sclerosis Journal. 2018;24 (6):862.
46. Abasiyanik Z, Yigit P, Ozdogar AT, Ertekin O, Ozakbas S. A comparison of the effects of yoga and clinical pilates exercise on mobility, respiratory muscle strength and cognition in persons with multiple sclerosis. Multiple Sclerosis Journal. 2018;24 (2 Supplement):976-7.
47. Abasiyanik Z, Ertekin O, Kahraman T, Ozakbas S. Effects of clinical Pilates training on walking, balance, and fall risk, respiratory and cognitive function in patients with multiple sclerosis: A randomized controlled trial. Fizyoterapi Rehabilitasyon. 2018;29 (2):S20-S1.
48. Bulguroglu I, Guclu-Gunduz A, Gokhan Y, Ozkul C, Irkec C, Batur-Caglayan HZ, et al. Comparison of the effects of mat Pilates and reformer Pilates on balance, strength, mobility, fatique and quality of life in patients with multiple sclerosis. European Journal of Neurology. 2015;1):672.
49. Bulguroglu I, Guclu-Gunduz A, Yazici G, Ozkul C, Irkec C, Nazliel B, et al. The effects of Mat Pilates and Reformer Pilates in patients with Multiple Sclerosis: A randomized controlled study. Neurorehabilitation. 2017;41(2):413-22.
50. Fox E, Freeman J, Hough A. The effects of Pilates upon deep abdominal muscle activity in people with multiple sclerosis: An exploratory ultrasound study. Multiple Sclerosis. 2014;20 (7)(7):981.
51. Fox E, Hough A, Creanor S, Gear M, Freeman J. Effects of Pilates-based core stability training in ambulant people with, multiple sclerosis: Multicenter, assessor-blinded, randomized, controlled trial. Physical Therapy. 2016;98(8):1170-8.
52. Fox EE, Hough AD, Creanor S, Gear M, Freeman JA. The effects of Pilates-based core stability training in ambulant people with multiple sclerosis: A multicentre, block randomised, double blinded placebo controlled trial. Multiple Sclerosis. 2013;1(11_suppl):40.
53. Freeman J, Fox E, Gear M, Hough A. Pilates based core stability training in ambulant individuals with multiple sclerosis: protocol for a multi-centre randomised controlled trial. BMC Neurology. 2012;12:19.
54. Guclu-Gunduz A, Citaker S, Irkec C, Nazliel B, Batur-Caglayan HZ. The effects of pilates on balance, mobility and strength in patients with multiple sclerosis. Neurorehabilitation. 2014;34(2):337-42.
55. Kalron A, Frid L, Berkowitz S, Achiron A. Changes in gait and balance in people with multiple sclerosis attending a 12-week Pilates exercise program. Multiple Sclerosis. 2016;22 (Supplement 3)(3_suppl):396.
56. Kalron A, Rosenblum U, Frid L, Achiron A. Pilates exercise training vs. physical therapy for improving walking and balance in people with multiple sclerosis: a randomized controlled trial. Clinical Rehabilitation. 2017;31(3):319-28.
57. Kara B, Kucuk F, Poyraz EC, Tomruk MS, Idiman E. Different types of exercise in Multiple Sclerosis: Aerobic exercise or Pilates, a single-blind clinical study. Journal of Back & Musculoskeletal Rehabilitation. 2017;30(3):565-73.
58. Kucuk F, Kara B, Coskuner Poyraz E, Idiman E, Senol H. Comparison of the effects of clinical pilates and exercise treatment in multiple sclerosis patients: A single blind randomised study. Fizyoterapi Rehabilitasyon. 2015;26 (2):S62.
59. Küçük F, Kara B, Poyraz E, İdiman E. Improvements in cognition, quality of life, and physical performance with clinical Pilates in multiple sclerosis: a randomized controlled trial. J Phys Ther Sci. 2016;28(3):761-8.
60. Vickrey B. Multiple Sclerosis Quality of Life (MSQOL)-54 Instrument 1995. Available from: https://www.nationalmssociety.org/For-Professionals/Researchers/Resources-for-MS-Researchers/Research-Tools/Clinical-Study-Measures/Multiple-Sclerosis-Quality-of-Life-54-(MSQOL-54).
61. Fiest KM, Greenfield J, Metz LM, Patten SB, Jetté N, Marrie RA. Discriminative ability of quality of life measures in multiple sclerosis. Health Qual Life Outcomes. 2017;15(1):246-.
62. Jongen PJ. Health-Related Quality of Life in Patients with Multiple Sclerosis: Impact of Disease-Modifying Drugs. CNS Drugs. 2017;31(7):585-602.
63. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003;41(5):582-92.
64. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health. 1992;83 Suppl 2:S7-11.
65. Gervasoni E, Jonsdottir J, Montesano A, Cattaneo D. Minimal Clinically Important Difference of Berg Balance Scale in People With Multiple Sclerosis. Arch Phys Med Rehabil. 2017;98(2):337-40.e2.
66. Sebastião E, Sandroff B, Learmonth Y, Motl R. Validity of The Timed Up and Go as A Measure of Functional Mobility in Persons with Multiple Sclerosis. Archives of Physical Medicine and Rehabilitation. 2016;97(7):1072-7.

67. Huang SL, Hsieh CL, Wu RM, Tai CH, Lin CH, Lu WS. Minimal detectable change of the timed "up & go" test and the dynamic gait index in people with Parkinson disease. Phys Ther. 2011;91(1):114-21.
68. Flansbjer UB, Holmbäck AM, Downham D, Patten C, Lexell J. Reliability of gait performance tests in men and women with hemiparesis after stroke. J Rehabil Med. 2005;37(2):75-82.
69. Borges J, Baptista AF, Santana N, Souza I, Kruschewsky RA, Galvao-Castro B, et al. Pilates exercises improve low back pain and quality of life in patients with HTLV-1 virus: a randomized crossover clinical trial. Journal of Bodywork & Movement Therapies. 2014;18(1):68-74.
70. Olsen MF, Bjerre E, Hansen MD, Tendal B, Hilden J, Hróbjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. Journal of Clinical Epidemiology. 2018;101:87-106.e2.
71. Pandya S, Nagendran T, Shah A, Chandrabharu V. Effect of Pilates Training Program on Balance in Participants with Idiopathic Parkinson's Disease - an Interventional Study. International Journal of Health Sciences and Research. 2017;7(6):186-96.
72. Daneshmandi H, Sayyar S, Bakhshayesh B. The Effect of a Selective Pilates Program on Functional Balance and Falling Risk in Patients with Parkinson's Disease. Zahedan J Res Med Sci. 2017;19(4):e7886.
73. Mollinedo-Cardalda I, Cancela-Carral JM, Vila-Suarez MH. Effect of a Mat Pilates Program with TheraBand on Dynamic Balance in Patients with Parkinson's Disease: Feasibility Study and Randomized Controlled Trial. Rejuvenation Research. 2018;21(5):423-30.
74. Nocera JR, Stegemöller EL, Malaty IA, Okun MS, Marsiske M, Hass CJ, et al. Using the Timed Up & Go test in a clinical setting to predict falling in Parkinson's disease. Archives of physical medicine and rehabilitation. 2013;94(7):1300-5.
75. Winser SJ, Kannan P, Bello UM, Whitney SL. Measures of balance and falls risk prediction in people with Parkinson's disease: a systematic review of psychometric properties. Clinical Rehabilitation. 2019;33(12):1949-62.
76. Krzysztoń K, Stolarski J, Kochanowski J. Evaluation of Balance Disorders in Parkinson's Disease Using Simple Diagnostic Tests-Not So Simple to Choose. Front Neurol. 2018;9:932-.
77. Donoghue D, Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. J Rehabil Med. 2009;41(5):343-6.
78. Lim HS, Kim YL, Lee SM. The effects of Pilates exercise training on static and dynamic balance in chronic stroke patients: a randomized controlled trial. J Phys Ther Sci. 2016;28(6):1819-24.
79. Lim HS, Yoon S. The effects of Pilates exercise on cardiopulmonary function in the chronic stroke patients: a randomized controlled trials. J Phys Ther Sci. 2017;29(5):959-63.
80. Roh S, Gil HJ, Yoon S. Effects of 8 weeks of mat-based Pilates exercise on gait in chronic stroke patients. J Phys Ther Sci. 2016;28(9):2615-9.
81. Sathe P, Chitre P, Ghodey S. Added effect of pilates mat exercises on balance and limits of stability in chronic stroke patients - a pilot study. International Journal of Physiotherapy and Research. 2018;6(3):2732-9.
82. Yun S-M, Park S-K, Lim HS. Influence of pilates training on the quality of life of chronic stroke patients. J Phys Ther Sci. 2017;29(10):1830-5.
83. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc. 2006;54(5):743-9.
84. Martins-Meneses DT, Antunes HK, de Oliveira NR, Medeiros A. Mat Pilates training reduced clinical and ambulatory blood pressure in hypertensive women using antihypertensive medications. Int J Cardiol. 2015;179:262-8.
85. Mazloum V, Rabiei P, Rahnama N, Sabzehparvar E. The comparison of the effectiveness of conventional therapeutic exercises and Pilates on pain and function in patients with knee osteoarthritis. Complementary Therapies in Clinical Practice. 2018;31:343-8.
86. Nadrian H, Moghimi N, Nadrian E, Moradzadeh R, Bahmanpour K, Iranpour A, et al. Validity and reliability of the Persian versions of WOMAC Osteoarthritis Index and Lequesne Algofunctional Index. Clin Rheumatol. 2012;31(7):1097-102.
87. Concoff A, Rosen J, Fu F, Bhandari M, Boyer K, Karlsson J, et al. A Comparison of Treatment Effects for Nonsurgical Therapies and the Minimum Clinically Important Difference in Knee Osteoarthritis: A Systematic Review. JBJS Rev. 2019;7(8):e5-e.
88. Salehi R, Valizadeh L, Negahban H, Karimi M, Goharpey S, Shahali S. The Western Ontario and McMaster Universities Osteoarthritis, Lequesne Algofunctional index, Arthritis Impact Measurement Scale-short form, and Visual Analogue Scale in patients with knee osteoarthritis: responsiveness and minimal clinically important differences. Disability and Rehabilitation. 2022:1-7.

89. tdpn RBR. The Effects of Pilates Exercise on Joint Pain and Mobility of Chikungunya Fever Patients. http://wwwwhoint/trialsearch/Trial2aspx?TrialID=RBR-99tdpn. 2017.
90. de Oliveira BFA, Carvalho PRC, de Souza Holanda AS, Dos Santos R, da Silva FAX, Barros GWP, et al. Pilates method in the treatment of patients with Chikungunya fever: a randomized controlled trial. Clinical Rehabilitation. 2019;33(10):1614-24.
91. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001;94(2):149-58.
92. Díaz-Arribas MJ F-SM, Royuela A, Kovacs FM, Gallego-Izquierdo T, Ramos-Sánchez M, et al. . Minimal Clinically Important Difference in Quality of Life for Patients with Low Back Pain. . Spine (Philadelphia, Pa 1976) 2017;42(24):1908-16.
93. Altan L, Korkmaz N, Dizdar M, Yurtkuran M. Effect of Pilates training on people with ankylosing spondylitis. Rheumatology International. 2012;32(7):2093-9.
94. Sieper J RM, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. . The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. . Annals of the rheumatic diseases 2009;68(Suppl 2):ii1-7.
95. Kviatkovsky MJ, Ramiro S, Landewé R, Dougados M, Tubach F, Bellamy N, et al. The Minimum Clinically Important Improvement and Patient-acceptable Symptom State in the BASDAI and BASFI for Patients with Ankylosing Spondylitis. J Rheumatol. 2016;43(9):1680-6.
96. Doward LC SA, Cook SA, Whalley D, Helliwell PS, Kay LJ, et al. Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. . Annals of the Rheumatic Diseases 2003;62(1):20.
97. Richard N, Haroon N, Tomlinson GA, Sari I, Touma Z, Inman RD. FRI0208 Ankylosing spondylitis quality of life: defining minimal clinically important change. Annals of the Rheumatic Diseases. 2018;77(Suppl 2):645.
98. Barnett R MA, Young M, Cavill C, Freeth M, Sengupta R. . Long-term improvement in axial spondyloarthritis clinical outcomes following 2-weeks of intensive education and rehabilitation: results from the Bath residential rehabilitation programme. . Rheumatology 2021;60(Supplement_1).
99. Alves dAM, Bezerra dSE, Bragade MD. The effectiveness of the Pilates method: Reducing the degree of non-structural scoliosis, and improving flexibility and pain in female college students. Journal of Bodywork and Movement Therapies. 2012;16(2):191-8.
100. Araújo MEAd, Silva EBd, Vieira PC, Cader SA, Mello DBd, Dantas EHM. Redução da dor crônica associada à escoliose não estrutural, em universitárias submetidas ao método Pilates. Motriz rev educ fís (Impr). 2010;16(4):958-66.
101. Junges S. A randomized clinical trial comparing posture and respiratory functions of women with kyphosis. European Geriatric Medicine. 2012;1):S68.
102. Junges S, Dias Molina R, Sartori J, Ferreira L, da Silva Filho IG. Mudança da pressão expiratória com a aplicação do método Pilates em mulheres adultas com hipercifose. Fisioterapia Brasil. 2017;18(5):643-9.
103. Junges S, Dias Molina R, Valle Gottlieb MG, da Silva Filho IG. Efeito de 30 semanas do Método Pilates na composição corporal de mulheres adultas com cifose. Fisioterapia Brasil. 2016;17(1):59-65.
104. Junges S, Gottlieb MG, Baptista RR, Quadros CBd, Resende TdL, Gomes I. Effectiveness of pilates method for the posture and flexibility of women with hyperkyphosis. Rev bras ciênc mov. 2012;20(1):21-3.
105. Lee SM, Lee CH, O'Sullivan D, Jung JH, Park JJ. Clinical effectiveness of a Pilates treatment for forward head posture. J Phys Ther Sci. 2016;28(7):2009-13.
106. jd8nq RBR. Effect of the Pilates in elderly. http://wwwwhoint/trialsearch/Trial2aspx?TrialID=RBR-6jd8nq. 2015.
107. Navega MT, Furlanetto MG, Lorenzo DM, Morcelli MH, Tozim BM. Effect of the Mat Pilates method on postural balance and thoracic hyperkyphosis among elderly women: a randomized controlled trial. Rev bras geriatr gerontol. 2016;19(3):465-72.
108. Kudchadkar G, Gurudut P, Welling A. Comparative effect of mat pilates and egoscue exercises in asymptomatic individuals with lumbar hyperlordosis: A randomized controlled trial. Indian Journal of Physical Therapy and Research. 2019;1(2):79-88.
109. Ctri. Effect Of Two Types Of Back Exercises On increased Curve Of Lower Back Spine On Normal Healthy Subjects. http://wwwwhoint/trialsearch/Trial2aspx?TrialID=CTRI/2018/07/015086. 2018.
110. Kim G, Hwangbo PN. Effects of Schroth and Pilates exercises on the Cobb angle and weight distribution of patients with scoliosis. Journal of Physical Therapy Science 2016 Mar;28(3):1012-1015. 2016;28(3):1012-5.
111. Angin E, Erden Z, Can F. The effects of clinical pilates exercises on bone mineral density, physical performance and quality of life of women with postmenopausal osteoporosis. Journal of Back & Musculoskeletal Rehabilitation. 2015;28(4):849-58.

112. Oksuz S, Unal E, Dizmek P, Aydin Ozcan D. The effects of clinical pilates exercises on kinesophobia in women with osteoporosis. Annals of the Rheumatic Diseases Conference: Annual European Congress of Rheumatology of the European League Against Rheumatism, EULAR. 2014;73(SUPPL. 2):1225.4-6.
113. Oksuz S, Ual E. The effect of the clinical pilates exercises on kinesiophobia and other symptoms related to osteoporosis: Randomised controlled trial. Complementary Therapies in Clinical Practice. 2017;26(3):68-72.
114. Nct. Effect Of Exercises On Kinesiophobia, And Other Symptoms Related to Osteoporosis. https://clinicaltrialsgov/show/NCT02716844. 2016.
115. Kucukcakir N, Altan L, Korkmaz N. Effects of Pilates exercises on pain, functional status and quality of life in women with postmenopausal osteoporosis. Journal of Bodywork & Movement Therapies. 2013;17(2):204-11.
116. Jones CJ, Rikli RE. Measuring Functional Fitness in Older Adults. The Journal of Active Ageing. 2002:25-30.
117. Bohannon RW, Crouch R. Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. J Eval Clin Pract. 2017;23(2):377-81.
118. Gibbs JC MN, Ponzano M, Templeton JA, Thabane L, Papaioannou A, et al. . Exercise for improving outcomes after osteoporotic vertebral fracture. . Cochrane Database of Systematic Reviews 2019(7).
119. Deeks J HJ, Altman D. . Chapter 10: Analysing data and undertaking meta-analyses. 2021. In: . Cochrane Handbook for Systematic Reviews of Interventions version 62 [Internet].
120. World Health Organization. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: report of a WHO study group [meeting held in Rome from 22 to 25 June 1992]. Geneva: World Health Organization; 1994.
121. Pardasaney PK, Latham NK, Jette AM, Wagenaar RC, Ni P, Slavin MD, et al. Sensitivity to Change and Responsiveness of Four Balance Measures for Community-Dwelling Older Adults. Physical Therapy. 2012;92(3):388-97.
122. Romero S BM, Velozo CA, Light K. Minimum detectable change of the Berg Balance Scale and Dynamic Gait Index in older persons at risk for falling. . J Geriatr Phys Ther 2011;34(3):131-7.
123. Altan L, Korkmaz N, Bingol U, Gunay B. Effect of pilates training on people with fibromyalgia syndrome: a pilot study. Archives of Physical Medicine & Rehabilitation. 2009;90(12):1983-8.
124. de Medeiros SA, de Almeida Silva HJ, do Nascimento RM, da Silva Maia JB, de Almeida Lins CA, de Souza MC. Mat Pilates is as effective as aquatic aerobic exercise in treating women with fibromyalgia: a clinical, randomized and blind trial. Advances in Rheumatology. 2020;60(1):21.
125. de Almeida Silva HJ, de Almeida Lins CA, Nobre TTX, de Sousa VPS, Caldas RTJ, de Souza MC. Mat Pilates and aquatic aerobic exercises for women with fibromyalgia: A protocol for a randomised controlled blind study. BMJ Open. 2019;9(2):(no pagination): e022306.
126. Ekici G, Unal E, Akbayrak T, Vardar-Yagli N, Yakut Y, Karabulut E. Effects of active/passive interventions on pain, anxiety, and quality of life in women with fibromyalgia: Randomized controlled pilot trial. Women & Health. 2017;57(1):88-107.
127. Ekici G, Ünal E, Akbayrak T, Yaɬli NV, Yakut Y. Active versus passive therapy in females with fibromyalgia; “pilates exercises and connective tissue massage”: a randomized controlled pilot trial. Fizyoterapi Rehabilitasyon. 2014;25(1):S14-S5.
128. Cruz-Diaz D, Martinez-Amat A, De la Torre-Cruz MJ, Casuso RA, de Guevara NM, Hita-Contreras F. Effects of a six-week Pilates intervention on balance and fear of falling in women aged over 65 with chronic low-back pain: A randomized controlled trial. Maturitas. 2015;82(4):371-6.
129. Cruz-Diaz D, Martinez-Amat A, Osuna-Perez MC, De la Torre-Cruz MJ, Hita-Contreras F. Short- and long-term effects of a six-week clinical Pilates program in addition to physical therapy on postmenopausal women with chronic low back pain: a randomized controlled trial. Disability & Rehabilitation. 2016;38(13):1300-8.
130. Cruz-Diaz D, Bergamin M, Gobbo S, Martinez-Amat A, Hita-Contreras F. Comparative effects of 12 weeks of equipment based and mat Pilates in patients with Chronic Low Back Pain on pain, function and transversus abdominis activation. A randomized controlled trial. Complementary Therapies in Medicine. 2017;33:72-7.
131. Cruz-Diaz D, Romeu M, Velasco-Gonzalez C, Martinez-Amat A, Hita-Contreras F. The effectiveness of 12 weeks of Pilates intervention on disability, pain and kinesiophobia in patients with chronic low back pain: a randomized controlled trial. Clinical Rehabilitation. 2018;32(9):1249-57.
132. da Fonseca JL, Magini M, de Freitas TH. Laboratory gait analysis in patients with low back pain before and after a pilates intervention. Journal of Sport Rehabilitation. 2009;18(2):269-82.
133. Gladwell V, Head S, Haggar M, Beneke R. Does a program of Pilates improve chronic non-specific low back pain? . Journal of Sport Rehabilitation. 2006;15(4):338-50.
134. Hasanpour-Dehkordi A, Dehghani A, Solati K. A Comparison of the Effects of Pilates and McKenzie Training on Pain and General Health in Men with Chronic Low Back Pain: A Randomized Trial. Indian Journal of Palliative Care. 2017;23(1):36-40.

135. Kliziene I, Sipaviciene S, Vilkiene J, Astrauskiene A, Cibulskas G, Klizas S, et al. Effects of a 16-week Pilates exercises training program for isometric trunk extension and flexion strength. Journal of Bodywork and Movement Therapies. 2017;21(1):124-32.
136. Kofotolis N, Kellis E, Vlachopoulos SP, Gouitas I, Theodorakis Y. Effects of Pilates and trunk strengthening exercises on health-related quality of life in women with chronic low back pain. Journal of Back & Musculoskeletal Rehabilitation. 2016;29(4):649-59.
137. Lopes S, Correia C, Felix G, Lopes M, Cruz A, Ribeiro F. Immediate effects of Pilates based therapeutic exercise on postural control of young individuals with non-specific low back pain: A randomized controlled trial. Complementary Therapies in Medicine. 2017;34:104-10.
138. Lopes S, Costa R, Rodriguesa M, Correia C, Lopes M, Goncalves P, et al. Immediate effects of four mat pilates exercises on postural sway of young adults with low back pain. Atencion Primaria. 2014;46 (Supplement 5):55-6.
139. Mazloum V, Sahebozamani M, Barati A, Nakhaee N. Comparing the effects of pilates training and McKenzie exercises on core muscles cross-sectional area and strength in patients with chronic non-specific low back pain: A clinical trial. [Persian]. Journal of Mazandaran University of Medical Sciences. 2016;26(143):48-61.
140. Mazloum V, Sahebozamani M, Barati A, Nakhaee N, Rabiei P. The effects of selective Pilates versus extension-based exercises on rehabilitation of low back pain. Journal of Bodywork & Movement Therapies. 2018;22(4):999-1003.
141. Miyamoto GC, Costa LO, Galvanin T, Cabral CM. The efficacy of the addition of the Pilates method over a minimal intervention in the treatment of chronic nonspecific low back pain: a study protocol of a randomized controlled trial. Journal of Chiropractic Medicine. 2011;10(4):248-54.
142. Miyamoto G, Costa L, Galvanin T, Cabral C. Efficacy of the addition of modified pilates exercises to a minimal intervention in patients with chronic low back pain: A randomized controlled trial. Physical Therapy. 2013;93(3):310-20.
143. Amaral DDV, Miyamoto GC, Franco KFM, Dos Santos Franco YR, Bastos De Oliveira NT, Hancock MJ, et al. Examination of a Subgroup of Patients With Chronic Low Back Pain Likely to Benefit More From Pilates-Based Exercises Compared to an Educational Booklet. Journal of Orthopaedic & Sports Physical Therapy. 2020;50(4):189-97.
144. Miyamoto G, Ferro MK, dos SFY, Teixeira BdON, Diulgeroglo VAD, Nery CBA, et al. Effectiveness and cost-effectiveness of different weekly frequencies of, pilates for chronic low back pain: Randomized controlled trial. Physical Therapy. 2016;96(3):382-9.
145. Miyamoto GC, Franco KFM, van Dongen JM, Franco Y, de Oliveira NTB, Amaral DDV, et al. Different doses of Pilates-based exercise therapy for chronic low back pain: a randomised controlled trial with economic evaluation. British Journal of Sports Medicine. 2018;52(13):859-68.
146. Natour J, Baptista AS, Cazotti LA, Ribeiro LHC, Jones A. Pilates to treat chronic non-specific low back pain. Arthritis and Rheumatism Conference: Annual Scientific Meeting of the American College of Rheumatology and Association of Rheumatology Health Professionals. 2011;63(10 SUPPL. 1).
147. Natour J, Cazotti Ld.e A, Ribeiro LH, Baptista AS, Jones A. Pilates improves pain, function and quality of life in patients with chronic low back pain: a randomized controlled trial. Clinical Rehabilitation. 2015;29(1):59-68.
148. Notarnicola A, Fischetti F, Maccagnano G, Comes R, Tafuri S, Moretti B. Daily pilates exercise or inactivity for patients with low back pain: a clinical prospective observational study. European journal of physical & rehabilitation medicine. 2014;50(1):59-66.
149. Pappas E, Panou H, Souglis A. The effect of a pilates exercise programme using fitball on people suffering from chronic low-back pain in terms of pain reduction and function imrovement. Journal of Physical Education & Sport. 2013;13(4):606-11.
150. Patti A, Bianco A, Paoli A, Messina G, Montalto MA, Bellafiore M, et al. Pain Perception and Stabilometric Parameters in People With Chronic Low Back Pain After a Pilates Exercise Program: A Randomized Controlled Trial. Medicine. 2016;95(2):e2414.
151. Quinn K, Barry S, Barry L. Do patients with chronic low back pain benefit from attending Pilates classes after completing conventional physiotherapy treatment? Physiotherapy Ireland. 2011;32(1):5-12.
152. Rydeard R, Leger A, Smith D. Pilates-based therapeutic exercise: effect on subjects with nonspecific chronic low back pain and functional disability: a randomized controlled trial. Journal of Orthopaedic & Sports Physical Therapy. 2006;36(7):472-84.
153. Valenza MC, Rodriguez-Torres J, Cabrera-Martos I, Diaz-Pelegrina A, Aguilar-Ferrandiz ME, Castellote-Caballero Y. Results of a Pilates exercise program in patients with chronic non-specific low back pain: a randomized controlled trial. Clinical Rehabilitation. 2017;31(6):753-60.

154. Zeada MA. Effects of Pilates on low back pain and urine catecholamine. Ovidius University Annals, Series Physical Education and Sport/Science, Movement and Health 2012;12(1):41-47. 2012.
155. Albert Anand U, Mariet Caroline P, Arun B, Lakshmi Gomathi G. A study to analyse the efficacy of modified Pilates based exercises and therapeutic exercises in individuals with chronic non specific low back pain: a randomized controlled trial. International Journal of Physiotherapy and Research 2014 May-Jun;2(3):525-529. 2014.
156. Avila Ribeiro I, Damé de Oliveira T, Redin Blois C. Effects of Pilates and Classical Kinesiotherapy on chronic low back pain: a case study. Fisioterapia em Movimento. 2015;28(4):759-65.
157. Bhadauria EA, Gurudut P. Comparative effectiveness of lumbar stabilization, dynamic strengthening, and Pilates on chronic low back pain: randomized clinical trial. Journal of Exercise Rehabilitation. 2017;13(4):477-85.
158. Brooks C, Kennedy S, Marshall PW. Specific trunk and general exercise elicit similar changes in anticipatory postural adjustments in patients with chronic low back pain: a randomized controlled trial. Spine. 2012;37(25):E1543-50.
159. Kennedy S. Exercise rehabilitation programs for chronic non-specific low back pain: A comparison of Pilates exercise and general aerobic exercise. Journal of Science and Medicine in Sport. 2012;15 (SUPPL.1):S80.
160. Marshall PW, Kennedy S, Brooks C, Lonsdale C. Pilates exercise or stationary cycling for chronic nonspecific low back pain: does it matter? a randomized controlled trial with 6-month follow-up. Spine. 2013;38(15):E952-9.
161. Devasahayam AJ, Ho DRY, Leung EYS, Goh MR, Koh P. The effects of a novel pilates exercise prescription method on people with non-specific unilateral musculoskeletal pain: A randomised pilot trial. Proceedings of Singapore Healthcare. 2016;25(4):201-6.
162. Donzelli S, Di DE, Cova A, Galletti R, Giunta N. Two different techniques in the rehabilitation treatment of low back pain: a randomized controlled trial. Eura Medicophys. 2006;42(3):205-10.
163. Dsa CF, Rengaramanujam K, Kudchadkar MS. To assess the effect of modified pilates compared to conventional core stabilization exercises on pain and disability in chronic non-specific low back pain-randomized controlled trial: Indian Journal of Physiotherapy and Occupational Therapy. 8 (3) (pp 202-207), 2014. Date of Publication: 2014.; 2014.
164. Gonzalez-Galvez N, Marcos-Pardo P, Carrasco-Poyatos M. Functional improvements after a pilates program in adolescents with a history of back pain: A randomised controlled trial. Complementary Therapies in Clinical Practice. 2019;35(1):1-7.
165. Mostagi F, Dias JM, Pereira LM, Obara K, Mazuquin BF, Silva MF, et al. Pilates versus general exercise effectiveness on pain and functionality in non-specific chronic low back pain subjects. Journal of Bodywork and Movement Therapies. 2015;19(4):636-45.
166. Rajpal N, Arora M, Chauhan V. A study on efficacy of Pilates & Pilates & Mckenzie exercise in postural low back pain - a rehabilitative protocol. Physiotherapy and occupational therapy journal. 2008;1(1):33-56.
167. Silva PHBd, Silva DFd, Oliveira JKdS, Oliveira FBd. The effect of the Pilates method on the treatment of chronic low back pain: a clinical, randomized, controlled study. BrJP. 2018;1(1):21-8.
168. Wajswelner H, Bennell K, Metcalf B. Clinical pilates for the management of chronic low back pain: A randomized controlled trial. Physiotherapy (United Kingdom). 2011;1):eS1318.
169. Wajswelner H, Metcalf B, Bennell K. Clinical pilates versus general exercise for chronic low back pain: randomized trial. Medicine & Science in Sports & Exercise. 2012;44(7):1197-205.
170. Johnsen LG, Hellum C, Nygaard OP, Storheim K, Brox JI, Rossvoll I, et al. Comparison of the SF6D, the EQ5D, and the oswestry disability index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskelet Disord. 2013;14:148.
171. Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. J Clin Epidemiol. 2006;59(1):45-52.
172. Ostelo RWJG, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting Change Scores for Pain and Functional Status in Low Back Pain: Towards International Consensus Regarding Minimal Important Change. Spine. 2008;33(1):90-4.
173. Stratford PW, Riddle DL. A Roland Morris Disability Questionnaire Target Value to Distinguish between Functional and Dysfunctional States in People with Low Back Pain. Physiother Can. 2016;68(1):29-35.
174. Stratford P. Assessing Disability and Change on Individual Patients: A Report of a Patient Specific Measure. Physiotherapy Canada 47(4):258-63.
175. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes. 2003;1:4.
176. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. 2002;21(2):271-92.

177. Cazotti LA, Jones A, Ribeiro LHC, Silva DR, Natour J. Effectiveness of the Pilates method in the treatment of chronic mechanical-postural neck pain. Annals of the Rheumatic Diseases. 2015;2)(Suppl 2):1345-6.
178. Cazotti LA, Jones A, Silva DR, Ribeiro LHC, Natour J. Effectiveness of pilates method for the treatment of chronic mechanical neck pain. Arthritis and Rheumatism. 2013;10)(10 Suppl):S776-S7.
179. de ACL, Jones A, Roger-Silva D, Ribeiro L, Natour J. Effectiveness of the Pilates Method in the Treatment of Chronic Mechanical Neck Pain: A Randomized Controlled Trial. Archives of Physical Medicine and Rehabilitation. 2018;99(9):1740-6.
180. Dunleavy K, Kava K, Goldberg A, Malek MH, Talley SA, Tutag-Lehr V, et al. Comparative effectiveness of Pilates and yoga group exercise interventions for chronic mechanical neck pain: quasi-randomised parallel controlled study. Physiotherapy. 2016;102(3):236-42.
181. Atilgan E, Aytar A, Caglar A, Tigli A, Arin G, Yapali G, et al. The effects of Clinical Pilates exercises on patients with shoulder pain: A randomised clinical trial. Journal of Bodywork and Movement Therapies. 2017;21(4):847-51.
182. Ulug N, Tunca Yilmaz O, Kara M, Ozcakar L. Effects of Pilates and yoga in patients with chronic neck pain: a sonographic study. Journal of Rehabilitation Medicine 2018 Jan;50(1):80-85. 2018;50(1):80-5.
183. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther. 2009;39(5):400-17.
184. Lauche R, Langhorst J, Dobos GJ, Cramer H. Clinically meaningful differences in pain, disability and quality of life for chronic nonspecific neck pain - a reanalysis of 4 randomized controlled trials of cupping therapy. Complement Ther Med. 2013;21(4):342-7.
185. Ahmadinezhad M, Kargar M, Vizeshfar F, Hadianfard MJ. Comparison of the Effect of Acupressure and Pilates-Based Exercises on Sleep Quality of Postmenopausal Women: A Randomized Controlled Trial. IrIran J Nurs Midwifery Res. 2017;22(2):140-6.
186. Irct2014042017344N. Pilates exercise and acupressure effect on anxiety and sleep quality in postmenopausal women. http://wwwwhoint/trialsearch/Trial2aspx?TrialID=IRCT2014042017344N1. 2014.
187. Campos de Oliveira L, Gonçalves de Oliveira R, de Almeida Pires-Oliveira DA. Effects of Whole-Body Vibration Versus Pilates Exercise on Bone Mineral Density in Postmenopausal Women: A Randomized and Controlled Clinical Trial. Journal of Geriatric Physical Therapy. 2019;42(2):E23-E31.
188. Nct. Effects of Whole Body Vibration and Pilates on Bone Mineral Density in Postmenopausal Women. https://clinicaltrialsgov/show/NCT02769143. 2016.
189. Campos de Oliveira L, Gonçalves de Oliveira R, de Almeida Pires-Oliveira DA. Effects of the Pilates exercise compared to whole body vibration and no treatment controls on muscular strength and quality of life in postmenopausal women: A randomized controlled trial. Isokinetics and Exercise Science. 2018;26(2):149-61.
190. Lee H, Caguicla JM, Park S, Kwak DJ, Won DY, Park Y, et al. Effects of 8-week Pilates exercise program on menopausal symptoms and lumbar strength and flexibility in postmenopausal women. Journal of Exercise Rehabilitation. 2016;12(3):247-51.
191. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213.
192. Kwak EK, Park HS, Kang NM. Menopause Knowledge, Attitude, Symptom and Management among Midlife Employed Women. J Menopausal Med. 2014;20(3):118-25.
193. Greeves JP, Cable NT, Reilly T, Kingsland C. Changes in muscle strength in women following the menopause: a longitudinal assessment of the efficacy of hormone replacement therapy. Clin Sci (Lond). 1999;97(1):79-84.
194. Phillips PJ, Phillipov G. Bone mineral density - frequently asked questions. Aust Fam Physician. 2006;35(5):341-4.
195. Phillipov G, Seaborn CJ, Phillips PJ. Reproducibility of DXA: potential impact on serial measurements and misclassification of osteoporosis. Osteoporos Int. 2001;12(1):49-54.
196. Mirmohammadali M, Ashrafinia F, Rajabi H, Amelvalizadeh M, Haghighi KS, Kazemnejad A. Effect of exercise on quality of sleep in post-partum women: HAYAT. 18 (1) (no pagination), 2012. Date of Publication: 2012.; 2012.
197. Ashrafinia F, Mirmohammadali M, Rajabi H, Kazemnejad A, SadeghniiatHaghighi K, Amelvalizadeh M, et al. The effects of Pilates exercise on sleep quality in postpartum women. Journal of Bodywork and Movement Therapies. 2014;18(2):190-9.
198. Ashrafinia F, Mirmohammadali M, Rajabi H, Kazemnejad A, Sadeghniiat Haghighi K, Amelvalizadeh M. Effect of Pilates exercises on postpartum maternal fatigue. Singapore Medical Journal. 2015;56(3):169-73.
199. Irct201102275912N. Effect of exercise on fatigue and sleep quality mothers in postpartum. http://wwwwhoint/trialsearch/Trial2aspx?TrialID=IRCT201102275912N2. 2011.
200. Nordin Å, Taft C, Lundgren-Nilsson Å, Dencker A. Minimal important differences for fatigue patient reported outcome measures-a systematic review. BMC medical research methodology. 2016;16:62-.

201. Purcell A FJ, Bennett S, Burmeister B, Haines T. . Determining the minimal clinically important difference criteria for the Multidimensional Fatigue Inventory in a radiotherapy population. . Supportive care in cancer 2010;18(3):307-15.
202. Celik D, Turkel N. The effectiveness of Pilates for partial anterior cruciate ligament injury. Knee Surgery, Sports Traumatology, Arthroscopy. 2017;25(8):2357-64.
203. Karaman A, Yuksel I, Kinikli GI, Caglar O. Do Pilates-based exercises following total knee arthroplasty improve postural control and quality of life? Physiotherapy Theory & Practice. 2017;33(4):289-95.
204. Maredupaka S, Meshram P, Chatte M, Kim WH, Kim TK. Minimal clinically important difference of commonly used patient-reported outcome measures in total knee arthroplasty: review of terminologies, methods and proposed values. Knee Surg Relat Res. 2020;32(1):19.
205. Quintana JM, Escobar A, Arostegui I, Bilbao A, Azkarate J, Goenaga JI, et al. Health-related quality of life and appropriateness of knee or hip joint replacement. Arch Intern Med. 2006;166(2):220-6.
206. Chan ACM, Pang MYC. Assessing Balance Function in Patients With Total Knee Arthroplasty. Physical Therapy. 2015;95(10):1397-407.
207. Irct20181118041692N. The effect of Pilates exercise on anxiety and self-esteem. http://wwwwhoint/trialsearch/Trial2aspx?TrialID=IRCT20181118041692N1. 2019.
208. Abavisani M, Lakzian R, Sarchahi Z, Sarchahi Z, Pourjafar H. The effect of Pilates exercise on anxiety in students of department of medical emergency - a clinical trial. Journal of Evolution of Medical and Dental Sciences. 2019;8(21):1755-8.
209. Julian LJ. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care Res (Hoboken). 2011;63 Suppl 11:S467-72.
210. Garcia-Soidan JL, Arufe Giraldez V, Cachon Zagalaz J, Lara-Sanchez AJ. Does pilates exercise increase physical activity, quality of life, latency, and sleep quantity in middle-aged people?: Perceptual and Motor Skills. 119 (3) (pp 838-850), 2014. Date of Publication: 2014.; 2014.
211. Sahinci Gokgul B, Hazar S. The effect of eight-week cyclic exercises and pilates exercises in women to some physical parameters and blood lipids. Turkish Journal of Sport and Exercise. 2017;19(1):60-4.
212. Arigo D MJ, Brown MM, Roberts SR, Pasko K, Butryn ML, et al. . Differences between accelerometer cut point methods among midlife women with cardiovascular risk markers. . Menopause 2020;27(5):559-67.
213. Lins L CF. SF-36 total score as a single measure of health-related quality of life: Scoping review. . SAGE Open Medicine 2016;4(0):205031211667172.
214. Curi VS, Haas AN, Alves-Vilaça J, Fernandes HM. Effects of 16-weeks of Pilates on functional autonomy and life satisfaction among elderly women. Journal of Bodywork and Movement Therapies. 2018;22(2):424-9.
215. Curi VS, Vilaca J, Haas AN, Fernandes HM. Effects of 16-weeks of Pilates on health perception and sleep quality among elderly women. Archives of Gerontology & Geriatrics. 2018;74:118-22.
216. de Andrade Mesquita LS, de Carvalho FT, de Andrade Freire LS, Neto OP, Zangaro RA. Effects of two exercise protocols on postural balance of elderly women: a randomized controlled trial. BMC Geriatrics 2015 Jun 2;15(61):Epub. 2015.
217. Nct. Exercise Protocols in Postural Balance Of Elderly Women. https://clinicaltrialsgov/show/NCT02278731. 2014.
218. Nct. Pilates and PNF Methods Induces Similar Strength Gains. https://clinicaltrialsgov/show/NCT02274909. 2014.
219. de Carvalho FT, de Andrade Mesquita LS, Pereira R, Neto OP, Zangaro RA. Pilates and proprioceptive neuromuscular facilitation methods induce similar strength gains but different neuromuscular adaptations in elderly women. Experimental Aging Research 2017 Oct-Dec;43(5):440-452. 2017.
220. Gandolfi NRS, Corrente JE, De Vitta A, Gollino L, Mazeto G. The influence of the Pilates method on quality of life and bone remodelling in older women: a controlled study. Quality of Life Research. 2020;29(2):381-9.
221. Irez GB, Ozdemir RA, Evin R, Irez SG, Korkusuz F. Integrating Pilates exercise into an exercise program for 65+ year-old women to reduce falls [with consumer summary]. Journal of Sports Science and Medicine 2011 Mar;10(1):105-111. 2011.
222. Liposcki D, da SNI, Silvano G, Zanella K, Schneider R. Influence of a Pilates exercise program on the quality of life of sedentary elderly people: A randomized clinical trial. Journal of Bodywork and Movement Therapies. 2019;23(2):390-3.
223. Aibar-Almazan A, Martinez-Amat A, Cruz-Diaz D, De la Torre-Cruz MJ, Jimenez-Garcia JD, Zagalaz-Anula N, et al. Effects of Pilates on fall risk factors in community-dwelling elderly women: A randomized, controlled trial. European Journal of Sport Science EJSS : Official Journal of the European College of Sport Science. 2019;19(10):1386-94.

224. Aibar-Almazan A, Hita-Contreras F, Cruz-Diaz D, de la Torre-Cruz M, Jimenez-Garcia JD, Martinez-Amat A. Effects of Pilates training on sleep quality, anxiety, depression and fatigue in postmenopausal women: a randomized controlled trial [with consumer summary]. Maturitas 2019 Jun;124:62-67. 2019.
225. Nct. Obesity, Sarcopenic, Risk of Falls in Spanish Postmenopausal Women. https://clinicaltrialsgov/show/NCT03201107. 2017.
226. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta- analysis. BMC Geriatrics. 2014;14(1):14.
227. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist. 2013;53(2):255-67.
228. Anjara SG, Bonetto C, Van Bortel T, Brayne C. Using the GHQ-12 to screen for mental health problems among primary care patients: psychometrics and practical considerations. International Journal of Mental Health Systems. 2020;14(1):62.
229. Barker AL, Talevski J, Bohensky MA, Brand CA, Cameron PA, Morello RT. Feasibility of Pilates exercise to decrease falls risk: a pilot randomized controlled trial in community-dwelling older people. Clinical Rehabilitation. 2016;30(10):984-96.
230. Actrn. Pilates: fitness fad or next frontier in falls prevention and bone health? A pilot randomised controlled trial. http://wwwwhoint/trialsearch/Trial2aspx?TrialID=ACTRN12612000224820. 2012.
231. Roller M, Kachingwe A, Beling J, Ickes DM, Cabot A, Shrier G. Pilates Reformer exercises for fall risk reduction in older adults: A randomized controlled trial. Journal of Bodywork & Movement Therapies. 2018;22(4):983-98.
232. Josephs S, Pratt ML, Calk Meadows E, Thurmond S, Wagner A. The effectiveness of Pilates on balance and falls in community dwelling older adults. Journal of Bodywork & Movement Therapies. 2016;20(4):815-23.
233. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. Phys Ther. 2000;80(9):896-903.
234. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
235. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 6,0. (updated July 2019): Cochrane 2019.
236. Sterne J, Hernán M, McAleenan A, Reeves B, Higgins J. Chapter 25: Assessing risk of bias in a non-randomized study. 2019. In: Cochrane Handbook for Systematic Reviews of Interventions version 6,0 [Internet]. Cochrane. Available from: www.training.cochrane.org/handbook.
237. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.