B.1 Structured exercise therapies or programmes

Overview of the PICO structure

Definition of the intervention	
Exercise is a subcategory of physical activity that is planned, structured, repetitive and purposeful in the sense that improvement or maintenance of one or more components of physical fitness is its objective. Structured exercise therapies or programmes are prescribed or planned by health workers, often delivered with instruction and supervision and may be standardized or individualized. These therapies are broadly defined as "a series of specific movements with the aim of training or developing physical capacity (e.g. muscle and joint strength and function, range of motion or aerobic capacity) by repetition or as physical training to promote good physical health" with the goal of reducing pain and functional limitations *(1)*. They include adopting postures, movements or activities, or a combination (e.g. strengthening, stretching, aerobic exercise) of varying duration, frequency and intensity. Exercise modalities considered for the guideline included: aerobic exercise; muscle strength training; stretching, flexibility or mobilizing exercises; Yoga; core strengthening; motor control exercise; functional restoration exercise; Pilates; Tai Chi; Qigong; aquatic/hydrotherapy; and mixed exercise therapies (i.e. two or more types of exercise in which one did not clearly predominate). Among the trials identified to inform the guideline, this intervention was delivered by health practitioners.	
PICO question	
Population and subgroups	Community-dwelling adults (age 20 years and over) experiencing chronic primary low back pain, with or without leg pain, including older people (aged 60 years and older). Subgroups: • Age (all adults and those aged 60 years and over) • Exercise type • Risk of bias judgement (low vs. not low) • Regional economic development - studies carried out in high-income countries compared with studies in low to middle-income countries
Comparators	a) Placebo/sham b) No or minimal intervention, or where the effect of the intervention can be isolated c) Usual care (described as usual care in the trial)

Outcomes	• Pain • Function • Harms/adverse events

Other Evidence-to-Decision (EtD) considerations

Summary of *values and preferences*	
All adults	**Older people**
No evidence synthesis commissioned for all adults. Judgements made based on experience of GDG members	**# Review findings GRADE-CERQual Assessment of confidence** 12 Participants emphasized the importance of continuity of physical exercises to maintain mobility and to reduce pain. A lack of continuity of physical exercise and instruction could have adverse effects, such as injuries. LOW 12 Participants wanted educational materials for physical interventions which had drawings and descriptions of the exercises. LOW

Summary of *resource considerations*	
All adults	**Older people**
No evidence synthesis commissioned for all adults. Judgements made based on experience of GDG members	No evidence identified

Summary of *equity and human rights considerations*	
All adults	**Older people**

No evidence synthesis commissioned for all adults. Judgements made based on experience of GDG members	**# Review findings GRADE-CERQual Assessment of confidence** 14 Participants saw the need to reduce the stigma associated with doing exercises as treatment for LBP as this was not regarded as legitimate treatment in rural Nigeria. They suggested that changes at the community level such as increasing awareness about the benefits of exercise could change negative community beliefs about exercises to legitimize exercise as treatment for back pain thereby reduce the current stigma associated with it. LOW

Summary of *acceptability considerations*	
All adults	**Older people**
No evidence synthesis commissioned for all adults. Judgements made based on experience of GDG members	**# Review findings GRADE-CERQual Assessment of confidence** 15 Many participants liked a group format for physical exercise classes as these facilitated social support, collaborative learning and social activities, which encouraged increased attendance. MODERATE

Summary of *feasibility considerations*	
All adults	**Older people**

No evidence synthesis commissioned for all adults. Judgements made based on experience of GDG members	**# Review findings GRADE-CERQual Assessment of confidence** 16 Some participants adopted physical exercise or physical supports as a part of their self-management approach to supplement conventional treatments, or when conventional treatments failed or were insufficient. Some viewed this as experimenting to find a solution. MODERATE 17 Participants requested shorter sessions of physical exercises on specific days to fit in with their daily schedule. VERY LOW

Summary of judgements

Domain	All adults	Older people
Benefits	Small; moderate; trivial; uncertain	Small; moderate
Harms	Trivial; uncertain	Uncertain
Balance benefits to harms	Favours exercise; probably favours exercise; uncertain	Probably favours exercise; uncertain
Overall certainty	Low; very low	Very low
Values and preferences	Possibly important uncertainty or variability; no important uncertainty or variability	Possibly important uncertainty or variability; no important uncertainty or variability
Resource considerations	Moderate costs; negligible costs and savings; varies (according to country and health system)	Moderate costs; negligible costs and savings; varies (according to country and health system)
Equity and human rights	Probably increased; probably reduced; no impact; varies	Probably increased; probably reduced; no impact; varies
Acceptability	Yes; probably yes; uncertain; varies	Probably yes; uncertain; varies

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Feasibility	Yes	Yes

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

GRADE Table 1: ***What are the benefits and harms of exercise in the management of community-dwelling adults (including older adults aged 60 years and over) with chronic primary low back pain (with or without leg pain) compared with sham?***

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	sham	Relative (95% CI)	Absolute (95% CI)		
Pain (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
4[1,2,3,4,a]	randomized trials	very serious[b]	serious[c]	not serious[d]	serious[e]	none	192	152	-	MD **1.51 lower** (3.02 lower to 0)	⨁◯◯◯ Very low	CRITICAL
Pain in adults (excluding those aged 60+ years) (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
3[1,2,4,a]	randomized trials	very serious[b]	not serious[f]	not serious[d]	serious[e]	none	152	112	-	MD **0.61 lower** (0.91 lower to 0.31 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in older adults (aged 60+ years) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
1[3]	randomized trials	very serious[b]	not serious[g]	serious[h]	very serious[i]	none	40	40	-	MD **5.54 lower** (6.43 lower to 4.65 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in adults in low- or lower middle-income countries (follow-up: closest to 2 weeks)												
0												
Pain (core strengthening) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	sham	Relative (95% CI)	Absolute (95% CI)		
1[3]	randomized trials	very serious[b]	not serious[g]	serious[h]	very serious[i]	none	40	40	-	MD **5.54 lower** (6.43 lower to 4.65 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (general/muscle strength training) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
1[1,a]	randomized trials	very serious[b]	not serious[g]	serious[h]	very serious[i]	none	22	10	-	MD **0.55 lower** (1.03 lower to 0.07 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (motor control exercise) (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
2[2,4,a]	randomized trials	very serious[b]	serious[j]	not serious[d]	serious[e]	none	106	92	-	MD **0.87 lower** (1.66 lower to 0.09 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (stretching or flexibility/mobilizing exercise) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
1[1,a]	randomized trials	very serious[b]	not serious[g]	serious[h]	very serious[i]	none	24	10	-	MD **0.55 lower** (1.01 lower to 0.09 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (low ROB) (follow-up: closest to 2 weeks; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	sham	Relative (95% CI)	Absolute (95% CI)		
1[2,a]	randomized trials	not serious[k]	not serious[g]	serious[h]	very serious[i]	none	77	77	-	MD **1 lower** (1.85 lower to 0.15 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (high or unclear ROB) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
3[1,3,4,a]	randomized trials	very serious[b]	serious[l]	not serious[d]	serious[e]	none	115	75	-	MD **1.6 lower** (3.44 lower to 0.24 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (motor control exercise, low ROB trial) (follow-up: closest to 12 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
1[2,a]	randomized trials	not serious[k]	not serious[g]	serious[h]	very serious[i]	none	77	77	-	MD **1.3 lower** (2.13 lower to 0.47 lower)	⨁◯◯◯ Very low	CRITICAL
Trials on pain in older adults (aged 60+ years) or in adults in low- or lower middle-income countries not identified												
0												
Function (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI; benefit indicated by lower values; scale: 0 to 24)												
3[1,2,3,a]	randomized trials	very serious[b]	not serious[m]	not serious[d]	serious[e]	none	163	137	-	MD **3.29 lower** (6.22 lower to 0.36 lower)	⨁◯◯◯ Very low	CRITICAL
Function in adults (excluding those aged 60+ years) (follow-up: closest to 2 weeks; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	sham	Relative (95% CI)	Absolute (95% CI)		
2[1,2,a]	randomized trials	very serious[b]	not serious[f]	not serious[d]	serious[e]	none	123	97	-	MD **2.04 lower** (2.86 lower to 1.22 lower)	⨁◯◯◯ Very low	CRITICAL
Function in older adults (aged 60+ years) (follow-up: closest to 2 weeks; assessed with: ODI; benefit indicated by lower values; scale: 0 to 24)												
1[3]	randomized trials	very serious[b]	not serious[g]	serious[h]	very serious[i]	none	40	40	-	MD **6.69 lower** (7.38 lower to 6 lower)	⨁◯◯◯ Very low	CRITICAL
Trial on function in adults in low- or lower middle-income countries not identified												
0												
Function (core strengthening) (follow-up: closest to 2 weeks; assessed with: ODI; benefit indicated by lower values; scale: 0 to 24)												
1[3]	randomized trials	very serious[b]	not serious[g]	serious[h]	very serious[i]	none	40	40	-	MD **6.69 lower** (7.38 lower to 6 lower)	⨁◯◯◯ Very low	CRITICAL
Function (general/muscle strength training) (follow-up: closest to 2 weeks; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												
1[1,a]	randomized trials	very serious[b]	not serious[g]	serious[h]	very serious[i]	none	22	10	-	MD **2.01 lower** (3.32 lower to 0.7 lower)	⨁◯◯◯ Very low	CRITICAL
Function (motor control exercise) (follow-up: closest to 2 weeks; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	sham	Relative (95% CI)	Absolute (95% CI)		
1[2,a]	randomized trials	not serious[k]	not serious[g]	serious[h]	very serious[i]	none	77	77	-	MD **2.3 lower** (4.26 lower to 0.34 lower)	⨁◯◯◯ Very low	CRITICAL
Function (stretching, or flexibility/mobilizing exercise) (follow-up: closest to 2 weeks; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												
1[1,a]	randomized trials	very serious[b]	not serious[g]	serious[h]	very serious[i]	none	24	10	-	MD **1.97 lower** (3.22 lower to 0.72 lower)	⨁◯◯◯ Very low	CRITICAL
Function (low ROB) (follow-up: closest to 2 weeks; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												
1[2]	randomized trials	not serious[k]	not serious[g]	serious[h]	very serious[i]	none	77	77	-	MD **2.3 lower** (4.26 lower to 0.34 lower)	⨁◯◯◯ Very low	CRITICAL
Function (high or unclear ROB) (follow-up: closest to 2 weeks; assessed with: RMDQ; ODI; benefit indicated by lower values; scale: 0 to 24)												
2[1,3,a]	randomized trials	very serious[b]	not serious[n]	serious[h]	very serious[i]	none	86	60	-	MD **3.59 lower** (7.11 lower to 0.07 lower)	⨁◯◯◯ Very low	CRITICAL
Function (motor control exercise, low ROB trial) (follow-up: closest to 12 months; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	sham	Relative (95% CI)	Absolute (95% CI)		
1[2,a]	randomized trials	not serious[k]	not serious[g]	serious[h]	very serious[i]	none	77	77	-	MD **0.9 lower** (3.15 lower to 1.35 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on function in older adults (aged 60+ years) or in adults in low to lower middle-income countries not identified												
0												
Harms												
1[2,o]	randomized trials	not serious[k]	not serious[g]	serious[h]	very serious[i]	none	3/77 (3.9%)	2/77 (2.6%)	**OR 1.52** (0.25 to 9.36)	**13 more per 1,000** (from 19 fewer to 174 more)	⨁◯◯◯ Very low	CRITICAL

CI: confidence interval; **MD:** mean difference; **NRS:** Numerical Rating Scale; **ODI:** Oswestry Disability Index; **OR:** odds ration; **PSFS:** Patient-Specific Functional Scale; **RMDQ:** Roland Morris Disability Questionnaire; **VAS:** Visual Analog Scale

Explanations

a. Comparison groups were split in half for trials with multiple comparisons.
b. Risk of bias: We downgraded twice. Most or all trials were rated as overall high risk of bias.
c. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 95%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
d. Indirectness: We did not downgrade. Trials conducted in different high-income countries.
e. Imprecision: We downgraded once due to low sample size (OIS would not have been reached).
f. Inconsistency: We did not downgrade. The point estimates are similar with overlapping confidence intervals; statistical heterogeneity is between 0% and 40%, which might not be important (i.e., I2 = 0%).
g. Inconsistency: We did not downgrade; however, there are no additional trials with which to compare findings.
h. Indirectness: We downgraded once. Trial(s) conducted in one country (high income).
i. Imprecision: We downgraded twice due to low sample size (OIS would not have been reached).
j. Inconsistency: We downgraded once. The point estimates are similar with overlapping confidence intervals; statistical heterogeneity is between 0% and 40%, which might not be important (i.e., I2 = 6%).
k. Risk of bias: We did not downgrade. Trial(s) rated as overall low risk of bias.
l. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 96%); this could not be explained due to small subgroups and may represent considerable heterogeneity.

m. Inconsistency: We did not downgrade. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 96%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
n. Inconsistency: We did not downgrade. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 99%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
o. Costa 2009: motor control exercise. Does not include older adults (60+ years). All adverse events were temporary exacerbations of pain.

References

1.Kim. Core Stability and Hip Exercises Improve Physical Function and Activity in Patients with Non-Specific Low Back Pain: A Randomized Controlled Trial. 2020.
2.Costa. Motor control exercise for chronic low back pain: a randomized placebo-controlled trial. 2009.
3.Park. A Randomized Controlled Trial Investigating the Effects of Equine Simulator Riding on Low Back Pain, Morphological Changes, and Trunk Musculature in Elderly Women. 2020.
4.Xu. Effect of Transversus abdominis muscle training on pressure-pain threshold in patients with chronic low Back pain. 2021.

GRADE Table 2: ***What are the benefits and harms of exercise in the management of community-dwelling adults (including older adults aged 60 years and over) with chronic primary low back pain (with or without leg pain) compared with no treatment/no additional treatment?***

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
Pain (follow-up: closest to 2 weeks; assessed with: VAS, NRS, ODI, MPQ; benefit indicated by lower values; scale: 0 to 10)												
41[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,a,b,c,d,e,f,g,h]	randomized trials	very serious[i]	not serious[j]	not serious[k]	not serious[l]	none	1109	959	-	MD **1.32 lower** (1.8 lower to 0.85 lower)	⨁⨁◯◯ Low	CRITICAL
Pain in adults (excluding aged 60+ years) (follow-up: closest to 2 weeks; assessed with: VAS, NRS, ODI; benefit indicated by lower values; scale: 0 to 10)												
35[1,2,3,4,5,6,7,9,10,12,13,14,15,17,18,19,20,21,22,23,24,25,26,28,29,30,31,32,33,34,35,36,37,40,41,a,b,c,d,e,f,g,h]	randomized trials	very serious[i]	not serious[j]	not serious[k]	not serious[l]	none	943	793	-	MD **1.2 lower** (1.7 lower to 0.69 lower)	⨁⨁◯◯ Low	CRITICAL
Pain in older adults (aged 60+ years) (follow-up: closest to 2 weeks; assessed with: VAS, NRS, MPQ; benefit indicated by lower values; scale: 0 to 10)												
6[8,11,16,27,38,39]	randomized trials	very serious[i]	serious[m]	not serious[k]	serious[n]	none	166	166	-	MD **2.31 lower** (3.37 lower to 1.24 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in adults in high or upper-middle income countries (follow-up: closest to 2 weeks; assessed with: VAS, NRS, ODI, MPQ; benefit indicated by lower values; scale: 0 to 10)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect			
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
22[1,8,9,10,11,12,14,15,16,20,22,23,24,25,26,28,29,31,33,36,37,38,a,b,c,d]	randomized trials	very serious[i]	not serious[o]	not serious[p]	not serious[l]	none	708	595	-	MD **1.23 lower** (1.57 lower to 0.89 lower)	⨁⨁◯◯ Low	CRITICAL
Pain in adults in low- or lower middle-income countries (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
19[2,3,4,5,6,7,13,17,18,19,21,27,30,32,34,35,39,40,41,a,e,f,g,h]	randomized trials	very serious[i]	serious[q]	not serious[r]	not serious[l]	none	401	364	-	MD **1.41 lower** (2.23 lower to 0.59 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (aerobic exercise) (follow-up: closest to 2 weeks; assessed with: VAS, NRS, ODI; benefit indicated by lower values; scale: 0 to 10)												
9[1,6,8,9,19,23,29,33,36,a]	randomized trials	very serious[i]	serious[s]	not serious[k]	serious[t]	none	253	214	-	MD **1.61 lower** (3.41 lower to 0.19 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (core strengthening) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
12[4,7,10,16,18,20,21,22,26,30,32,40,a,f,h]	randomized trials	very serious[i]	serious[u]	not serious[k]	serious[n]	none	196	177	-	MD **1.52 lower** (2.02 lower to 1.01 lower)	⨁◯◯◯ Very low	CRITICAL

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
Pain (general/muscle strength training) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
3[3,14,34,a]	randomized trials	serious[v]	serious[w]	not serious[k]	very serious[x]	none	92	84	-	MD **0.61 higher** (1.62 lower to 2.84 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (mixed exercise) (follow-up: closest to 2 weeks; assessed with: VAS, MPQ; benefit indicated by lower values; scale: 0 to 10)												
7[11,12,27,36,37,38,39,a,b,c,d]	randomized trials	very serious[i]	serious[y]	not serious[k]	not serious[l]	none	250	203	-	MD **1.52 lower** (2.58 lower to 0.47 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (motor control exercise) (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
5[2,13,25,35,41,a]	randomized trials	very serious[i]	serious[z]	not serious[k]	very serious[x]	none	104	92	-	MD **0.78 lower** (1.79 lower to 0.23 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (Pilates) (follow-up: closest to 2 weeks; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
1[28,e]	randomized trials	serious[v]	not serious[aa]	serious[ab]	very serious[x]	none	43	43	-	MD **2.1 lower** (3.07 lower to 1.13 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (Qigong) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
2[15,24]	randomized trials	very serious[i]	not serious[ac]	serious[ab]	very serious[x]	none	60	60	-	MD **0.93 lower** (1.45 lower to 0.4 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (stretching or flexibility/mobilizing exercise) (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
5[5,17,31,34,40,a,g]	randomized trials	very serious[i]	not serious[ad]	not serious[k]	very serious[x]	none	96	79	-	MD **1.52 lower** (2.08 lower to 0.95 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (Tai Chi) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
1[26]	randomized trials	very serious[i]	not serious[aa]	serious[ab]	very serious[x]	none	15	7	-	MD **2.38 lower** (3.16 lower to 1.6 lower)	⨁◯◯◯ Very low	CRITICAL

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
Pain (low ROB trials) (follow-up: closest to 2 weeks; assessed with: VAS 0 to 100; benefit indicated by lower values)												
1[42]	randomized trials	not serious[ae]	not serious[aa]	serious[ab]	very serious[x]	none	Smeets 2008: 119 participants total. Mixed exercise vs no/no additional treatment. Participants performed combination treatment (active physical treatment [aerobic and core strengthening exercises] + graded activity with problem-solving training) vs graded activity with problem solving training alone. Between-group MD (VAS 0-100) graded activity with problem-solving training alone vs combination treatment = 5.35, 95% CI -3.73 to 14.42.				⨁◯◯◯ Very low	CRITICAL
Pain (follow-up: closest to 3 months; assessed with: VAS, ODI; benefit indicated by lower values; scale: 0 to 10)												
5[23,33,36,37,43,a]	randomized trials	very serious[i]	not serious[af]	not serious[p]	serious[n]	none	191	156	-	MD **0.54 lower** (0.88 lower to 0.2 lower)	⨁◯◯◯ Very low	CRITICAL
Trials on pain in older adults or in adults in low- or lower middle-income countries not identified												
0												
Pain (aerobic exercise) (follow-up: closest to 3 months; assessed with: VAS, ODI; benefit indicated by lower values; scale: 0 to 10)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
3[23,33,36,a]	randomized trials	very serious[i]	not serious[af]	not serious[p]	serious[n]	none	111	70	-	MD **0.73 lower** (1.35 lower to 0.11 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (core strengthening) (follow-up: closest to 3 months; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
1[43]	randomized trials	very serious[i]	not serious[aa]	serious[ab]	very serious[x]	none	47	47	-	MD **0.53 lower** (0.97 lower to 0.09 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (mixed exercise) (follow-up: closest to 3 months; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
2[36,37,a]	randomized trials	serious[v]	not serious[af]	not serious[p]	very serious[x]	none	33	39	-	MD **0.05 lower** (1.13 lower to 1.02 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (low ROB trials) (follow-up: closest to 3 months)												
0												
Pain (follow-up: closest to 12 months; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect			
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
1[14,ag]	randomized trials	serious[v]	not serious[aa]	serious[ab]	very serious[x]	none	35	35	-	MD **0.1 lower** (1.32 lower to 1.12 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on pain in older adults or in adults in low- or lower middle-income countries not identified												
0												
Pain (general/muscle strength training) (follow-up: closest to 12 months; assessed with: benefit indicated by lower values; scale: 0 to 10)												
1[14]	randomized trials	serious[v]	not serious[aa]	serious[ab]	very serious[x]	none	35	35	-	MD **0.1 lower** (1.32 lower to 1.12 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (mixed exercise, low ROB trial) (follow-up: closest to 12 months; assessed with: VAS 0-100; benefit indicated by lower values)												
1[42]	randomized trials	not serious[ae]	not serious[aa]	serious[ab]	very serious[x]	none	Smeets 2008 (119 participants). Participants performed combination treatment (active physical treatment [aerobic and core strengthening exercises] + graded activity with problem solving training) vs graded activity with problem solving training alone. Between-group MD (VAS 0-100) graded activity with problem solving training alone vs combination treatment = 6.25, 95% CI -2.94 to 15.44.			⨁◯◯◯ Very low	CRITICAL
Function (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, modified ODI, Quebec Back Pain Disability Scale, Hannover, PROMIS, WI; benefit indicated by lower values)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
39[1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18,19,21,23,24,25,27,28,29,30,31,32,33,34,35,36,37,38,40,41,44,45,46,a,ah,ai,aj,ak,al,am,an,ao,ap,aq]	randomized trials	very serious[i]	serious[ar]	not serious[k]	not serious[l]	none	1077	956	-	SMD **0.8 lower** (1.07 lower to 0.53 lower)	⨁◯◯◯ Very low	CRITICAL
Function in adults (excluding aged 60+ years) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, modified ODI, Quebec Back Pain Disability Scale, Hannover, PROMIS, WI; benefit indicated by lower values)												
35[1,2,3,4,5,6,7,9,10,12,13,14,15,17,18,19,21,23,24,25,28,29,30,31,32,33,34,35,36,37,40,41,44,45,46,a,ah,ai,aj,ak,al,am,an,ao,ap,aq]	randomized trials	very serious[i]	serious[ar]	not serious[k]	not serious[l]	none	933	811	-	SMD **0.8 lower** (1.1 lower to 0.5 lower)	⨁◯◯◯ Very low	CRITICAL
Function in older adults (aged 60+ years) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI; benefit indicated by lower values)												
4[8,16,27,38,a]	randomized trials	very serious[i]	serious[as]	not serious[k]	serious[n]	none	144	145	-	SMD **0.85 lower** (1.66 lower to 0.04 lower)	⨁◯◯◯ Very low	CRITICAL
Function in adults in high or upper-middle income countries (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, Hannover, PROMIS, WI; benefit indicated by lower values)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
18[1,8,9,10,12,14,15,16,23,24,25,28,29,31,33,36,37,38,a,ah,ai,aj,am,ap]	randomized trials	very serious[i]	not serious[o]	not serious[p]	not serious[l]	none	637	544	-	SMD **0.48 lower** (0.7 lower to 0.27 lower)	⨁⨁◯◯ Low	CRITICAL
Function in adults in low- or lower middle-income countries (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, modified ODI, Quebec Back Pain Disability Scale; benefit indicated by lower values)												
21[2,3,4,5,6,7,13,17,18,19,21,27,30,32,34,35,40,41,44,45,46,a,ak,al,an,ao,aq]	randomized trials	very serious[i]	not serious[at]	not serious[r]	not serious[l]	none	440	412	-	SMD **1.19 lower** (1.74 lower to 0.64 lower)	⨁⨁◯◯ Low	CRITICAL
Function (aerobic exercise) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, Quebec Back Pain Disability Scale, Hannover, PROMIS; benefit indicated by lower values)												
10[1,6,8,9,19,23,29,33,36,44,a]	randomized trials	very serious[i]	not serious[au]	not serious[k]	not serious[l]	none	263	224	-	SMD **0.98 lower** (1.51 lower to 0.45 lower)	⨁⨁◯◯ Low	CRITICAL
Function (core strengthening) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI; benefit indicated by lower values)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
10[4,7,10,16,18,21,30,32,40,45,a,ak,ap,aq]	randomized trials	very serious[i]	not serious[av]	not serious[k]	serious[n]	none	186	178	-	SMD **1.08 lower** (1.47 lower to 0.69 lower)	⨁◯◯◯ Very low	CRITICAL
Function (general/muscle strength training) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI; benefit indicated by lower values)												
3[3,14,34,a]	randomized trials	serious[v]	serious[aw]	not serious[k]	very serious[x]	none	92	84	-	SMD **1.09 higher** (0.99 lower to 3.17 higher)	⨁◯◯◯ Very low	CRITICAL
Function (mixed exercise) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, WI; benefit indicated by lower values)												
6[12,27,36,37,38,46,a,ah,ai,aj,am,an,ao]	randomized trials	very serious[i]	serious[ax]	not serious[k]	not serious[l]	none	233	196	-	SMD **0.83 lower** (1.38 lower to 0.29 lower)	⨁◯◯◯ Very low	CRITICAL
Function (motor control exercise) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, modified ODI; benefit indicated by lower values)												
5[2,13,25,35,41,a]	randomized trials	very serious[i]	serious[ay]	not serious[k]	very serious[x]	none	104	92	-	SMD **0.82 lower** (1.65 lower to 0.02 higher)	⨁◯◯◯ Very low	CRITICAL

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
Function (Pilates) (follow-up: closest to 2 weeks; assessed with: RMDQ; benefit indicated by lower values)												
1[28]	randomized trials	serious[v]	not serious[aa]	serious[ab]	very serious[x]	none	43	43	-	SMD **0.74 lower** (1.18 lower to 0.3 lower)	⨁◯◯◯ Very low	CRITICAL
Function (Qigong) (follow-up: closest to 2 weeks; assessed with: ODI; benefit indicated by lower values)												
2[15,24]	randomized trials	very serious[i]	not serious[az]	serious[ab]	very serious[x]	none	60	60	-	SMD **1.16 lower** (1.87 lower to 0.45 lower)	⨁◯◯◯ Very low	CRITICAL
Function (stretching or flexibility/mobilizing exercise) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI; benefit indicated by lower values)												
5[5,17,31,34,40,a,al,ao]	randomized trials	very serious[i]	serious[ba]	not serious[k]	very serious[x]	none	96	79	-	SMD **0.62 lower** (1.36 lower to 0.13 higher)	⨁◯◯◯ Very low	CRITICAL
Function (Tai Chi) (follow-up: closest to 2 weeks; assessed with: ODI 0-50; benefit indicated by lower values)												
1[47]	randomized trials	very serious[i]	serious[aa]	serious[ab]	very serious[x]	none	Liu 2018: 43 participants total. Authors reported the average ODI score in each domain of Tai Chi group decreased significantly compared to comparison group (overall scores not reported).				⨁◯◯◯ Very low	

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
Function (low ROB trials) (follow-up: closest to 2 weeks)												
1[42]	randomized trials	not serious[ae]	not serious[aa]	serious[ab]	very serious[x]	none	Smeets 2008 (119 participants). Participants performed combination treatment (active physical treatment [aerobic and core strengthening exercises] + graded activity with problem solving training) vs graded activity with problem solving training alone. Between-group MD (RMDQ 0-24) graded activity with problem solving training alone vs combination treatment = 0.58, 95% CI -1.08 to 2.24.				⨁◯◯◯ Very low	CRITICAL
Function (follow-up: closest to 3 months; assessed with: ODI, Hannover, Functional Rating Test, WI; benefit indicated by lower values)												
5[23,33,37,43,48,a]	randomized trials	very serious[i]	serious[as]	not serious[k]	serious[n]	none	211	163	-	SMD **0.99 lower** (1.69 lower to 0.3 lower)	⨁◯◯◯ Very low	CRITICAL
Function in older adults (aged 60+ years) (follow-up: closest to 3 months)												
0												
Function in adults in high or upper-middle income countries (follow-up: closest to 3 months; assessed with: ODI, Hannover, WI; benefit indicated by lower values)												
4[23,33,37,43]	randomized trials	very serious[i]	not serious[af]	not serious[p]	serious[n]	none	173	129	-	SMD **0.43 lower** (0.66 lower to 0.19 lower)	⨁◯◯◯ Very low	CRITICAL

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
Function in adults in low- or lower middle-income countries (follow-up: closest to 3 months; assessed with: Functional Rating Test; benefit indicated by lower values)												
1[48,a]	randomized trials	very serious[i]	not serious[aa]	serious[bb]	very serious[x]	none	38	34	-	SMD **2.87 lower** (6.68 lower to 0.93 higher)	⨁◯◯◯ Very low	CRITICAL
Function (aerobic exercise) (follow-up: closest to 3 months; assessed with: ODI, Hannover; benefit indicated by lower values)												
2[23,33]	randomized trials	very serious[i]	not serious[af]	not serious[p]	very serious[x]	none	102	56	-	SMD **0.27 lower** (0.6 lower to 0.07 higher)	⨁◯◯◯ Very low	CRITICAL
Function (core strengthening) (follow-up: closest to 3 months; assessed with: ODI; benefit indicated by lower values)												
1[43]	randomized trials	very serious[i]	not serious[aa]	serious[ab]	very serious[x]	none	47	47	-	SMD **0.66 lower** (1.07 lower to 0.24 lower)	⨁◯◯◯ Very low	CRITICAL
Function (mixed exercise) (follow-up: closest to 3 months; assessed with: WI; benefit indicated by lower values)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
1[37]	randomized trials	serious[v]	not serious[aa]	serious[ab]	very serious[x]	none	24	26	-	SMD **0.44 lower** (1.01 lower to 0.12 higher)	⨁◯◯◯ Very low	CRITICAL
Function (stretching or flexibility/mobilizing exercise) (follow-up: closest to 3 months; assessed with: Functional Rating Scale (unspecified scale range); benefit indicated by lower values)												
1[48,a]	randomized trials	very serious[i]	not serious[aa]	serious[bb]	very serious[x]	none	38	34	-	SMD **2.87 lower** (6.68 lower to 0.93 higher)	⨁◯◯◯ Very low	CRITICAL
Function (low ROB trials) (follow-up: closest to 3 months)												
0												
Function (follow-up: closest to 12 months; assessed with: RMDQ; benefit indicated by lower vales; scale: 0 to 24)												
1[14,bc]	randomized trials	serious[v]	not serious[aa]	serious[ab]	very serious[x]	none	35	35	-	MD **0.2 lower** (2.73 lower to 2.33 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on function in older adults or in adults in low- or lower middle-income countries not identified												
0												
Function (general strength training) (follow-up: closest to 12 months; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
1[14]	randomized trials	serious[v]	not serious[aa]	serious[ab]	very serious[x]	none	35	35	-	MD **0.2 lower** (2.73 lower to 2.33 higher)	⨁◯◯◯ Very low	CRITICAL
Function (mixed exercise, low ROB trial) (follow-up: closest to 12 months; assessed with: RMDQ 0-24; benefit indicated by lower values)												
1[42]	randomized trials	not serious[ae]	not serious[aa]	serious[ab]	very serious[x]	none	Smeets 2008: 119 participants. Participants performed combination treatment (active physical treatment [aerobic and core strengthening exercises] + graded activity with problem solving training) vs graded activity with problem solving training alone. Between-group MD (RMDQ 0-24) graded activity with problem solving training alone vs combination treatment = 1.11, 95% CI -0.56 to 2.79.				⨁◯◯◯ Very low	CRITICAL
Harms												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	no treatment	Relative (95% CI)	Absolute (95% CI)		
6[23,28,32,33,38,42]	randomized trials	serious[v]	not serious	not serious	not serious	none	Lang 2021 (aerobic exercise; 174 participants total): no harms reported. Miyamoto 2013 (Pilates; 86 participants total): no harms reported. Rahbar 2018 (core strengthening; 80 participants total): no harms reported. Rotter 2022 (aerobic exercise; 55 participants total): no harms reported. Smeets 2008 (mixed exercise; 119 participants total): 3 (5%) of participants in exercise group had increased back pain. Weiner 2008 (older adults) (mixed exercise; 200 participants total): no significant intervention-associated adverse events reported. One participant (2%) had increased back pain. One participant (2%) had decreased functional status.				⨁⨁⨁◯ Moderate	CRITICAL

CI: confidence interval; **Hannover:** Hannover Functional Ability Questionnaire; **MD:** mean difference; **MPQ:** McGill Pain Questionnaire; **NRS:** Numerical Rating Scale; **ODI:** Oswestry Disability Index; **PROMIS:** Patient-Reported Outcomes Measurement Information System; **PSFS:** Patient-Specific Functional Scale; **RMDQ:** Roland Morris Disability Questionnaire; **SMD:** standardized mean difference; **VAS:** Visual Analog Scale; **WI:** Waddell Disability Index

Explanations

a. Comparison groups were split for trials with multiple comparisons.

b. Dalichau 2003: not included in meta-analysis due to missing data. Rated as high overall risk of bias; 90 participants total. Mixed exercise vs no/no additional treatment: authors reported greater pain reduction in exercise group (unclear effect estimates).

c. McIlveen 1998: not included in meta-analysis due to missing data. Rated as high overall risk of bias; 95 participants total. Mixed exercise vs no/no additional treatment: no significant difference in the number of participants who improved more than 1 point between exercise and comparison; p=0.13 (McGill Pain Questionnaire 1-5, benefit indicated by lower values).

d. Smeets 2008: not included in meta-analysis due to missing data. Rated as low overall risk of bias; 119 participants total. Mixed exercise vs no/no additional treatment. Participants performed combination treatment (active physical treatment [aerobic and core strengthening exercises] + graded activity with problem solving training) vs graded activity with problem solving training alone. Between-group MD (VAS 0-100, benefit indicated by lower values) graded activity with problem solving training alone vs combination treatment = 5.35, 95% CI -3.73 to 14.42.

e. Sokhanguei 2017: not included in meta-analysis due to missing data. Rated as high overall risk of bias; 34 participants total. Pilates exercise vs no/no additional treatment. Authors reported greater pain reduction in Pilates group; mean difference (SEM): -2.3 (0.72); p=0.003.

f. Kanwal 2021: not included in meta-analysis due to missing data. Rated as high overall risk of bias; 24 participants total. Core strengthening vs no/no additional treatment. Authors reported no significant difference in pain between groups; p=0.317.

g. Raza 2020: not included in meta-analysis due to missing data. 40 participants, rated as overall high risk of bias, stretching, or flexibility/mobilizing exercise. Authors reported no significant difference in median pain between groups; p=0.112.
h. Rathi 2013: not included in meta-analysis due to missing data. 30 participants, rated as overall high risk of bias, core strengthening. Authors reported significantly greater mean pain reduction in exercise group (3.8, SD 1.0) than in no treatment group (2.9, SD 0.8); $p < 0.05$ (VAS 0-10, benefit indicated by lower values).
i. Risk of bias: We downgraded twice. Most or all trials were rated as overall high risk of bias.
j. Inconsistency: We did not downgrade. There is similarity in the majority of point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 97%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
k. Indirectness: We did not downgrade. Trials conducted in different countries both high and low income.
l. Imprecision: We did not downgrade. OIS would have been reached. The point estimate reached the pre-specified threshold for what may be considered appreciable benefit (MD = -1 or SMD = -0.2); the confidence interval does not cross the null.
m. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 97%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
n. Imprecision: We downgraded once due to low sample size (OIS would not have been reached).
o. Inconsistency: We did not downgrade. There is similarity in the majority of point estimates with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 65%); this could not be explained due to small subgroups and may represent substantial heterogeneity.
p. Indirectness: We did not downgrade. Trials conducted in different high-income countries.
q. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 98%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
r. Indirectness: We did not downgrade. Trials conducted in different low-income countries.
s. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 99%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
t. Imprecision: We downgraded once. OIS would have been reached. The point estimate reached the pre-specified threshold for what may be considered appreciable benefit (MD = -1 or SMD = -0.2); the confidence interval crosses the null.
u. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 73%); this could not be explained due to small subgroups and may represent substantial heterogeneity.
v. Risk of bias: We downgraded once. Some of the weight (>50%) comes from trials with unclear risk of bias.
w. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 95%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
x. Imprecision: We downgraded twice due to low sample size (OIS would not have been reached).
y. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 82%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
z. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 90%); this could not be explained due to small subgroups and may represent substantial heterogeneity.
aa. Inconsistency: We did not downgrade; however, there are no additional trials with which to compare findings.
ab. Indirectness: We downgraded once. Trial(s) conducted in one country (high income).
ac. Inconsistency: We did not downgrade. There is similarity in the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 43%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
ad. Inconsistency: We did not downgrade. There is similarity in the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 32%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
ae. Risk of bias: We did not downgrade. Trial(s) rated as overall low risk of bias.
af. Inconsistency: We did not downgrade. The point estimates are similar with overlapping confidence intervals; statistical heterogeneity is between 0% and 40%, which might not be important (i.e., I2 = 0%).
ag. Smeets 2008 was not included in the meta-analysis (provided within-group mean changes; no follow-up scores). 119 participants, rated as overall low risk of bias. Participants performed combination treatment (active physical treatment [aerobic and core strengthening exercises] + graded activity with problem solving training) vs graded activity with problem solving training alone. Between-group MD (VAS 0-100, benefit indicated by lower values) graded activity with problem solving training alone vs combination treatment = 6.25, 95% CI -2.94 to 15.44.

ah. Smeets 2008 was not included in the meta-analysis (provided within-group mean changes; no follow-up scores). 119 participants, rated as overall low risk of bias. Participants performed combination treatment (active physical treatment [aerobic and core strengthening exercises] + graded activity with problem solving training) vs graded activity with problem solving training alone. Between-group MD (RMDQ 0-24, benefit indicated by lower values) graded activity with problem solving training alone vs combination treatment = 0.58, 95% CI -1.08 to 2.24.
ai. Dalichau 2003: not included in meta-analysis due to missing data. Rated as high overall risk of bias; 90 participants total. Mixed exercise vs no/no additional treatment: authors reported greater disability improvement in exercise group (unclear effect estimates).
aj. McIlveen 1998: not included in meta-analysis due to missing data. Rated as high overall risk of bias; 95 participants total. Mixed exercise vs no/no additional treatment: authors reported significantly greater number of participants improved more than 10 points in the exercise group (27%) than in the no treatment group (8%); p=0.04 (ODI 0-100).
ak. Kanwal 2021: not included in meta-analysis due to missing data. Rated as high overall risk of bias; 24 participants total. Core strengthening vs no/no additional treatment. Authors reported no significant difference in disability between groups; p=0.692.
al. Raza 2020: not included in meta-analysis due to missing data. 40 participants, rated as overall high risk of bias, stretching, or flexibility/mobilizing exercise. Authors reported significantly lower median item scores in the exercise group for personal care (exercise: median 1, IQR 0; no treatment: median 1, IQR 1; p=0.041) and travelling (exercise: median 1, IQR 0; no treatment: median 1, IQR 0; p=0.027); no significant difference for other items (ODI individual items; 0-5).
am. Da Silva 2014: not included in meta-analysis due to missing data. 18 participants total, rated as overall high risk of bias, mixed exercise. Authors reported significantly greater mean % improvement from baseline in exercise group (45% improvement) vs no exercise (2% worsening); p=0.008 (RMDQ 0-24, benefit indicated by lower values).
an. Wattamwar 2012: not included in meta-analysis due to missing data. 24 participants total, rated as overall high risk of bias, yoga exercise. Authors reported no significant difference in change scores between groups; p=0.146.
ao. Sedaghati 2017: not included in meta-analysis due to missing data. 34 participants total, rated as overall high risk of bias, mixed exercise (in and out of water) and stretching or flexibility/mobilizing exercise. Authors reported a significant difference in follow-up scores between mixed exercise (mean 23.0, SD 3.0) and no treatment (mean 27.5, SD 3.0) (Quebec Back Pain Disability Scale 0-100, benefit indicated by lower values). No significant difference in follow-up scores between stretching or flexibility/mobilizing group and no treatment.
ap. Liu 2018: not included in meta-analysis due to missing data. 43 participants total, rated as overall high risk of bias, Tai Chi and core strengthening. Authors reported the average ODI score in each domain of both exercise groups decreased significantly compared to comparison group (overall scores not reported) (ODI 0-50, benefit indicated by lower values).
aq. Rathi 2013: not included in meta-analysis due to missing data. 30 participants total, rated as overall high risk of bias, core strengthening. Authors reported significantly greater mean disability improvement in exercise group (24.1, SD 3.2) than in no treatment group (19.73, SD 3.58); $p < 0.05$ (ODI 0-100; benefit indicated by lower values).
ar. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 87%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
as. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 89%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
at. Inconsistency: We did not downgrade. There is similarity in the majority of point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 92%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
au. Inconsistency: We did not downgrade. There is similarity in the majority of point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 84%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
av. Inconsistency: We did not downgrade. There is similarity in the majority of point estimates with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 63%); this could not be explained due to small subgroups and may represent substantial heterogeneity.
aw. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 97%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
ax. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 83%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
ay. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 88%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
az. Inconsistency: We did not downgrade. There is similarity in the majority of point estimates with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 70%); this could not be explained due to small subgroups and may represent substantial heterogeneity.
ba. Inconsistency: We downgraded once. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 82%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
bb. Indirectness: We downgraded once. Trial(s) conducted in one country (low income).

bc. Smeets 2008 was not included in the meta-analysis (provided within-group mean changes; no follow-up scores). 119 participants, rated as overall low risk of bias. Participants performed combination treatment (active physical treatment [aerobic and core strengthening exercises] + graded activity with problem solving training) vs graded activity with problem solving training alone. Between-group MD (RMDQ 0-24, benefit indicated by lower values) graded activity with problem solving training alone vs combination treatment = 1.11, 95% CI -0.56 to 2.79.

References

1.Abadi. The effect of aquatic exercise program on low-back pain disability in obese women. 2019.
2.Abdel-Aziem. The Effects of Stabilization Exercises Combined With Pelvic Floor Exercise in Women With Nonspecific Low Back Pain: A Randomized Clinical Study. 2021.
3.Afzal. Effects of virtual reality exercises and routine physical therapy on pain intensity and functional disability in patients with chronic low back pain. 2022.
4.Ahmadizadeh. The Effect of Stabilization Exercises Along With Self-care Training on Transverse Abdominal Activity, Pain, and Disability in Mothers With Low Back Pain Having Children With Cerebral Palsy: A Randomized Controlled Trial. 2020.
5.Ain. Effects of Sustained Natural Apophyseal Glides with and without thoracic posture correction techniques on mechanical back pain: a randomized control trial. 2019.
6.Alikhajeh. A Comparison of 6 Weeks of Aquatic Exercise and Kinesio Taping in Patients With Chronic Nonspecific Low Back Pain. 2021.
7.Alvani. Neuromuscular exercises on pain intensity, functional disability, proprioception, and balance of military personnel with chronic low back pain. 2021.
8.Barni. Does an Aerobic Exercise Improve Outcomes in Older Sedentary Nonspecific Low Back Pain Subjects? A Randomized Controlled Study. 2018.
9.Bruehl. Are endogenous opioid mechanisms involved in the effects of aerobic exercise training on chronic low back pain? A randomized controlled trial. 2020.
10.da Luz. Neuromuscular electrical stimulation associated with core stability exercises in nonspecific postural low back pain: a randomized clinical trial. 2019.
11.da Silva. Back school program for back pain: education or physical exercise? 2014.
12.Dalichau. Quantification of spinal configuration and postural capacity by ultrasound topometry for evaluation of different muscle strengthening programs in the therapy of back pain. 2005.
13.Dineshkumar. Effect of abdominal drawing-in maneuver along with resisted ankle dorsi-flexion to activate transverse abdominis muscle in chronic non-specific low back pain. 2015.
14.Fukuda. Does adding hip strengthening exercises to manual therapy and segmental stabilization improve outcomes in patients with nonspecific low back pain? A randomized controlled trial. 2021.
15.Gao. Effect observation of Baduanjin training on chronic nonspecific low back pain. 2018.
16.Ge. Effects of core stability training on older women with low back pain: a randomized controlled trial. 2022.
17.Hatefi. The effect of static stretching exercises on hip range of motion, pain, and disability in patients with non-specific low back pain. 2021.
18.Hemmati. Effects of consecutive supervised core stability training on pain and disability in women with nonspecific chronic low back pain. 2011.
19.Idowu. Efficacy of Graded Activity with and without Daily-Monitored-Walking on Pain and Back Endurance among Patients with Concomitant Low-Back Pain and Type-2 Diabetes: A Randomized Trial. 2020.
20.Joseph. Effects of Massage as a Combination Therapy with Lumbopelvic Stability Exercises as Compared to Standard Massage Therapy in Low Back Pain: a Randomized Cross-Over Study. 2018.
21.Karimzadeh. The effect of 8 weeks core stabilization exercises on pain and functional disability induced by low back pain in the mothers of children with cerebral palsy. 2016.
22.Kim. Effects of 3D Moving Platform Exercise on Physiological Parameters and Pain in Patients with Chronic Low Back Pain. 2020.
23.Lang. A randomized controlled trial investigating effects of an individualized pedometer driven walking program on chronic low back pain. 2021.
24.Li. Clinical observation of treating chronic nonspecific low back pain with Baduanjin combining sling exercise therapy. 2015.
25.Li. The Effect of Virtual Reality Training on Anticipatory Postural Adjustments in Patients with Chronic Nonspecific Low Back Pain: A Preliminary Study. 2021.
26.Liu. Chen-Style Tai Chi for Individuals (Aged 50 Years Old or Above) with Chronic Non-Specific Low Back Pain: A Randomized Controlled Trial. 2019.
27.Madadi-Shad. Effect of a corrective exercise program on gait kinetics and muscle activities in older adults with both low back pain and pronated feet: A double-blind, randomized controlled trial. 2020.
28.Miyamoto. Efficacy of the addition of modified Pilates exercises to a minimal intervention in patients with chronic low back pain: a randomized controlled trial. 2013.
29.Nardin. Effects of photobiomodulation and deep water running in patients with chronic non-specific low back pain: a randomized controlled trial. 2022.
30.Noormohammadpour. The Role of a Multi-Step Core Stability Exercise Program in the Treatment of Nurses with Chronic Low Back Pain: A Single-Blinded Randomized Controlled Trial. 2018.
31.Prado. Influence of isostretching on patients with chronic low back pain. A randomized controlled trial. 2019.
32.Rahbar. Effect of hippotherapy simulator on pain, disability and range of motion of the spinal column in subjects with mechanical low back pain: A randomized single-blind clinical trial. 2018.
33.Rotter. Mindful Walking in Patients with Chronic Low Back Pain: A Randomized Controlled Trial. 2022.
34.Shamsi. Effects of static stretching and strengthening exercises on flexion relaxation ratio in patients with LBP: A randomized clinical trial. 2022.
35.Srivastav. Comparison between Effectiveness of Lumbar Stabilization Exercises and Conventional Physical Therapy in the Management of Mechanical Low Back Pain . 2018.
36.Suh. The effect of lumbar stabilization and walking exercises on chronic low back pain: A randomized controlled trial. 2019.
37.Takinaci. Clinical efficacy of spa therapy (balneotherapy) for chronic low back pain: A randomized single-blind trial. 2019.
38.Weiner. Efficacy of percutaneous electrical nerve stimulation and therapeutic exercise for older adults with chronic low back pain: a randomized controlled trial. 2008.
39.Yalfani. Effects of an 8-Week Virtual Reality Training Program on Pain, Fall Risk, and Quality of Life in Elderly Women with Chronic Low Back Pain: Double-Blind Randomized Clinical Trial. 2022.

40.Zakari. Comparison of the Effects of Positional Release Therapy and Lumbar Stabilization Exercises in the Management of Chronic Mechanical Low Back Pain—Randomized Controlled Trial. 2019.
41.Karimi. The effects of consecutive supervised functional lumbar stabilizing exercises on the postural balance and functional disability in low back pain. 2014.
42.Smeets. Chronic low back pain: physical training, graded activity with problem solving training, or both? The one-year post-treatment results of a randomized controlled trial. 2008.
43.Shu. Observation on the Curative Effect of Massage Manipulation Combined with Core Strength Training in Patients with Chronic Nonspecific Low Back Pain. 2021.
44.Ansari. The effect of aquatic exercise on functional disability, flexibility and function of trunk muscles in postmenopausal women with chronic non-specific low back pain: Randomized controlled trial. 2020.
45.Fouda. Effects of proprioceptive neuromuscular facilitation techniques in treating chronic nonspecific low back pain patients. 2021.
46.Heidari. Comparison of the Effects of 8 Weeks of Core Stability Exercise on Ball and Sling Exercise on The Quality of Life and Pain in The Female With Non-Specific Chronic Low Back Pain (NSLBP). 2018.
47.Liu. Effects of Tai Chi on the event-related potential of patients with chronic non-specific low back pain. 2018.
48.Shariat. The impact of modified exercise and relaxation therapy on chronic lower back pain in office workers: a randomized clinical trial. 2019.

GRADE Table 3: ***What are the benefits and harms of exercise in the management of community-dwelling adults (including older adults aged 60 years and over) with chronic primary low back pain (with or without leg pain) compared with usual care?***

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	usual care	Relative (95% CI)	Absolute (95% CI)		
Pain (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
5[1,2,3,4,5,a,b,c]	randomized trials	very serious[d]	not serious[e]	not serious[f]	serious[g]	none	288	166	-	MD **0.89 lower** (1.27 lower to 0.5 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in adults (excluding those aged 60+ years) (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
3[1,3,4,a,b,c]	randomized trials	very serious[d]	not serious[h]	not serious[f]	serious[g]	none	232	115	-	MD **0.93 lower** (1.4 lower to 0.45 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in older adults (aged 60+ years) (follow-up: closest to 2 weeks; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
2[2,5]	randomized trials	very serious[d]	not serious[i]	not serious[j]	very serious[k]	none	56	51	-	MD **0.65 lower** (1.5 lower to 0.19 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (high or upper-middle income countries) (follow-up: closest to 2 weeks; assessed with: VAS, NRS; benefit indicated by lower values; scale: 0 to 10)												
4[2,3,4,5,a,b,c]	randomized trials	very serious[d]	not serious[l]	not serious[j]	serious[g]	none	243	118	-	MD **1.01 lower** (1.32 lower to 0.7 lower)	⨁◯◯◯ Very low	CRITICAL

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	usual care	Relative (95% CI)	Absolute (95% CI)		
Pain (low- or lower middle-income countries) (follow-up: closest to 2 weeks; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
1[1]	randomized trials	serious[m]	not serious[n]	serious[o]	very serious[k]	none	45	48	-	MD **0.1 higher** (0.81 lower to 1.01 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (core strengthening) (follow-up: closest to 2 weeks; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
1[4,c]	randomized trials	very serious[d]	not serious[n]	serious[p]	very serious[k]	none	7	7	-	MD **2.3 lower** (3.96 lower to 0.64 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (general/muscle strength training) (follow-up: closest to 2 weeks; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
1[3,a]	randomized trials	very serious[d]	not serious[n]	serious[p]	serious[g]	none	180	60	-	MD **1.01 lower** (1.36 lower to 0.65 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (mixed exercise) (follow-up: closest to 2 weeks; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
3[1,2,5]	randomized trials	very serious[d]	not serious[i]	not serious[f]	serious[g]	none	101	99	-	MD **0.31 lower** (0.93 lower to 0.31 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (yoga) (follow-up: closest to 2 weeks; assessed with: Aberdeen Back Pain Scale, 0-100; benefit indicated by lower values)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	usual care	Relative (95% CI)	Absolute (95% CI)		
16,q	randomized trials	seriousm	not seriousn	seriousp	seriousg	none	Yoga vs usual care: difference in mean change -2.42, 95% CI -4.97 to 0.12 (313 participants total).				⨁◯◯◯ Very low	CRITICAL
Pain (low ROB trials) (follow-up: closest to 2 weeks)												
0									-		-	
Pain (older adults aged 60+ years, mixed exercise, unclear ROB trial) (follow-up: closest to 3 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
12	randomized trials	seriousm	not seritableousn	seriousp	very seriousk	none	26	22	-	MD **0.3 lower** (1.66 lower to 1.06 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (low- or lower middle-income countries) (follow-up: closest to 3 months)												
0												
Pain (yoga exercise) (follow-up: closest to 12 months; assessed with: Aberdeen Back Pain Scale, 0-100; benefit indicated by lower values)												
16,q	randomized trials	seriousm	not seriousn	seriousp	seriousg	none	Yoga vs usual care: difference in mean change -0.73, 95% CI -3.30 to 1.84 (313 participants total).				⨁◯◯◯ Very low	CRITICAL
Low ROB trial on pain or trials on pain in older adults or adults in low or lower middle-income countries not identified												
0												
Function (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, modified ODI; benefit indicated by lower values; scale: 0 to 100)												
61,2,3,4,5,7,a,r	randomized trials	very seriousd	not serious s	not seriousf	not serious t	none	303	181	-	MD **9.72 lower** (13.72 lower to 5.72 lower)	⨁⨁◯◯ Low	CRITICAL

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	usual care	Relative (95% CI)	Absolute (95% CI)		
Function in adults (excluding those aged 60+ years) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI, modified ODI; benefit indicated by lower values; scale: 0 to 100)												
4[1,3,4,7,a,r]	randomized trials	very serious[d]	not serious[u]	not serious[f]	serious[g]	none	247	130	-	MD **9.72 lower** (14.37 lower to 5.07 lower)	⨁◯◯◯ Very low	CRITICAL
Function in older adults (aged 60+ years) (follow-up: closest to 2 weeks; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 100)												
2[2,5]	randomized trials	very serious[d]	not serious[i]	not serious[j]	very serious[k]	none	56	51	-	MD **9.81 lower** (16.11 lower to 3.52 lower)	⨁◯◯◯ Very low	CRITICAL
Function (high or upper-middle income countries) (follow-up: closest to 2 weeks; assessed with: RMDQ, ODI; benefit indicated by lower values; scale: 0 to 100)												
4[2,3,4,5,a,r]	randomized trials	very serious[d]	not serious[v]	not serious[j]	serious[g]	none	243	118	-	MD **8.13 lower** (10.69 lower to 5.58 lower)	⨁◯◯◯ Very low	CRITICAL
Function (low or lower middle-income countries) (follow-up: closest to 2 weeks; assessed with: ODI, modified ODI; benefit indicated by lower values; scale: 0 to 100)												
2[1,7]	randomized trials	very serious[d]	not serious[w]	serious[o]	very serious[k]	none	60	63	-	MD **14.02 lower** (19.75 lower to 8.3 lower)	⨁◯◯◯ Very low	CRITICAL
Function (aerobic exercise) (follow-up: closest to 2 weeks; assessed with: ODI; benefit indicated by lower values; scale: 0 to 100)												

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	usual care	Relative (95% CI)	Absolute (95% CI)		
1[7]	randomized trials	very serious[d]	not serious[n]	serious[o]	very serious[k]	none	15	15	-	MD **16 lower** (17.59 lower to 14.41 lower)	⨁◯◯◯ Very low	CRITICAL
Function (core strengthening) (follow-up: closest to 2 weeks; assessed with: ODI; benefit indicated by lower values; scale: 0 to 100)												
1[4]	randomized trials	very serious[d]	not serious[n]	serious[p]	very serious[k]	none	7	7	-	MD **4.3 lower** (9.64 lower to 1.04 higher)	⨁◯◯◯ Very low	CRITICAL
Function (general/muscle strength training) (follow-up: closest to 2 weeks; assessed with: ODI; benefit indicated by lower values; scale: 0 to 100)												
1[3,a]	randomized trials	very serious[d]	not serious[n]	serious[p]	serious[g]	none	180	60	-	MD **8.95 lower** (11.96 lower to 5.93 lower)	⨁◯◯◯ Very low	CRITICAL
Function (mixed exercise) (follow-up: closest to 2 weeks; assessed with: RMDQ, modified ODI; benefit indicated by lower values; scale: 0 to 100)												
3[1,2,5]	randomized trials	very serious[d]	not serious[i]	not serious[f]	serious[g]	none	101	99	-	MD **9.77 lower** (14.64 lower to 4.89 lower)	⨁◯◯◯ Very low	CRITICAL
Function (yoga) (follow-up: closest to 2 weeks; assessed with: RMDQ, 0-24; benefit indicated by lower values)												
1[6]	randomized trials	serious[m]	not serious[n]	serious[p]	serious[g]	none	Yoga vs usual care: difference in mean change -2.17, 95% CI -3.31 to -1.03 (313 participants total).				⨁◯◯◯ Very low	CRITICAL

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	exercise	usual care	Relative (95% CI)	Absolute (95% CI)		
Function (low ROB trials) (follow-up: closest to 2 weeks)												
0									-		-	0
Function (older adults aged 60+ years, mixed exercise, unclear ROB trial) (follow-up: closest to 3 months; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												
1[2]	randomized trials	serious[m]	not serious[n]	serious[p]	very serious[k]	none	26	22	-	MD **2.3 lower** (4.92 lower to 0.32 higher)	⨁◯◯◯ Very low	CRITICAL
Function (low or lower middle-income countries) (follow-up: closest to 3 months)												
Function (yoga) (follow-up: closest to 12 months; assessed with: RMDQ 0 to 24; benefit indicated by lower values)												
1[6]	randomized trials	serious[m]	not serious[n]	serious[p]	serious[g]	none	Yoga vs usual care: difference in mean change -1.57, 95% CI -2.71 to -0.42 (313 participants total).				⨁◯◯◯ Very low	CRITICAL
Low ROB trial on function or trials of function in older adults or in adults in low or lower middle countries not identified												
0												CRITICAL
Harms												
2[5,6]	randomized trials	serious[m]	not serious	not serious[j]	serious[g]	none	Tilbrook 2011: yoga vs usual care; 313 participants total: Minor adverse events: 11 of 156 (7.1%) yoga participants events were classified as nonserious and mostly related to increased pain. Major adverse events. 1 yoga participant experienced severe pain (possibly associated with yoga). In usual care group, 1 participant died; 1 had severe accident/injury. Zadro 2019: mixed exercise vs usual care; 60 older participants total: no adverse events reported.				⨁⨁◯◯ Low	CRITICAL

CI: confidence interval; **MD:** mean difference; **NRS:** Numerical Rating Scale; **ODI:** Oswestry Disability Index; **RMDQ:** Roland Morris Disability Questionnaire; **VAS:** Visual Analog Scale

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Explanations

a. Comparison groups were split for trials with multiple comparisons.
b. Tilbrook 2011: not included in meta-analysis (only reported within-group changes; follow-up scores not provided). Rated as unclear overall risk of bias. Yoga vs usual care: difference in mean change -2.42, 95% CI -4.97 to 0.12 (313 participants total; Aberdeen Back Pain Scale 0-100, benefit indicated by lower values).
c. Raoul 2019: not included in meta-analysis due to missing data. Rated as high overall risk of bias. Core strengthening vs usual care: greater mean pain reduction in exercise group (3.91, SD 2.88) than in comparison group (1.83, SD 2.80), p<0.01(67 participants total; NRS 0-10, benefit indicated by lower values).
d. Risk of bias: We downgraded twice. Most or all trials were rated as overall high risk of bias.
e. Inconsistency: We did not downgrade. There is similarity in the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 50%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
f. Indirectness: We did not downgrade. Trials conducted in different countries both high and low income.
g. Imprecision: We downgraded once due to low sample size (OIS would not have been reached).
h. Inconsistency: We did not downgrade. There is similarity in the majority of point estimates with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 65%); this could not be explained due to small subgroups and may represent substantial heterogeneity.
i. Inconsistency: We did not downgrade. The point estimates are similar with overlapping confidence intervals; statistical heterogeneity is between 0% and 40%, which might not be important (i.e., I2 = 0%).
j. Indirectness: We did not downgrade. Trials conducted in different high-income countries.
k. Imprecision: We downgraded twice due to low sample size (OIS would not have been reached).
l. Inconsistency: We did not downgrade. The point estimates are similar with overlapping confidence intervals; statistical heterogeneity is between 0% and 40%, which might not be important (i.e., I2 = 26%).
m. Risk of bias: We downgraded once. Some (>50%) or all weight comes from trials with unclear risk of bias.
n. Inconsistency: We did not downgrade; however, there are no additional trials with which to compare findings.
o. Indirectness: We downgraded once. Trial(s) conducted in one country (low income).
p. Indirectness: We downgraded once. Trial(s) conducted in one country (high income).
q. Tillbrook 2011: not included in meta-analysis (only reported within-group changes; follow-up scores not provided).
r. Tillbrook 2011: not included in meta-analysis (only reported within-group changes; follow-up scores not provided). Rated as unclear overall risk of bias. Yoga vs usual care: difference in mean change -2.17, 95% CI -3.31 to -1.03 (313 participants total; RMDQ 0-24, benefit indicated by lower values).
s. Inconsistency: We did not downgrade. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 80%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
t. Imprecision: We did not downgrade. OIS would have been reached. The point estimate did not reach the pre-specified threshold for what may be considered appreciable benefit (MD = -10 or SMD = -0.2); the confidence interval does not cross the null.
u. Inconsistency: We did not downgrade. There is similarity in some of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 85%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
v. Inconsistency: We did not downgrade. The point estimates are similar with overlapping confidence intervals; statistical heterogeneity is between 0% and 40%, which might not be important (i.e., I2 = 9%).
w. Inconsistency: We did not downgrade. There is similarity in the majority of point estimates with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 59%); this could not be explained due to small subgroups and may represent substantial heterogeneity.

References

1.Chhabra. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. 2018.
2.Jinnouchi. Effects of brief self-exercise education on the management of chronic low back pain: A community-based, randomized, parallel-group pragmatic trial. 2020.
3.Kell. The response of persons with chronic nonspecific low back pain to three different volumes of periodized musculoskeletal rehabilitation. 2011.
4.Mendes. Core stabilisation exercises reduce chronic low back pain in Air Force fighter pilots: a randomized controlled trial. 2022.
5.Zadro. Video-Game-Based Exercises for Older People With Chronic Low Back Pain: A Randomized Controlledtable Trial (GAMEBACK). 2019.
6.Tillbrook. Yoga for chronic low back pain: a randomized trial. 2011.
7.Gupta. The Effectiveness of Aerobic Exercise Program for Improving Functional Performance and Quality of Life in Chronic Low Back Pain. 2019.

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

GRADE Table 4: ***What are the benefits and harms of exercise compared with a combined comparator of placebo, no intervention or usual care for adults with chronic primary low back pain?***

This GRADE Evidence Profile Table presents data from the Cochrane review by Hayden et al. (2021) with certainty assessments conducted by an independent methodologist. The certainty assessments highlighted in green illustrate where changes have been proposed compared with the original review.

Setting: Community and health facility-based

Bibliography: Hayden JA, Ellis J, Ogilvie R, Malmivaara A, van Tulder MW. Exercise therapy for chronic low back pain. *Cochrane Database of Systematic Reviews* 2021, Issue 9. Art. No.: CD009790. DOI: https://doi.org/10.1002/14651858.CD009790.pub2. Independent ROBIS evaluation on Hayden 2021 review and re-created GRADE table below.

Certainty assessment							№ of patients		Effect		Certainty assessment for GDG by independent methodologist	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Exercise	Placebo, No intervention or Usual care	Relative (95% CI)	Absolute (95% CI)		
Pain intensity (0 - 100; 0 = no pain): Earliest follow-up (time point closest to 3 months) (scale: 0 to 100)												
35[a]	randomized trials	not serious[b]	serious[c]	serious[d]	not serious	none	1531	1215	-	MD **15.22 / 100 lower** (18.26 lower to 12.18 lower)	⨁⨁◯◯ Low	CRITICAL
Functional limitations ((0 - 100; 0 = no functional limitations): Earliest follow-up (time point closest to 3 months) (scale: 0 to 100)												
38[e]	randomized trials	not serious[f]	not serious[g]	serious[d]	not serious	publication bias strongly suspected[h]	1664	1278	-	MD **6.82 /100 lower** (8.32 lower to 5.32 lower)	⨁⨁◯◯ Low	CRITICAL

Web Annex D.B1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

CI: confidence interval; **MD:** mean difference

Explanations

a. 35 trials with 47 study groups

b. Risk of bias: *From Hayden review*: Seven studies (10 groups; 526 participants) were judged to have high risk of bias (19% of participant data). Exclusion of these studies in sensitivity analysis did not change conclusions.

c. Inconsistency: *From Hayden review*: Serious unexplained inconsistency (substantial heterogeneity I^2 = 75%, point estimates and confidence intervals varied considerably).

d. Indirectness: *From Independent ROBIS evaluation:* No trials were conducted in low-income countries and no trials were conducted on the African continent, potentially limiting the applicability to all global regions. The comparator combined usual care, placebo/sham and no intervention unlike the WHO PICO which separated these comparators; however, this was not considered a reason to further downgrade. Most trials were conducted in health facilities and few in the community, limiting generalizability to settings outside health facilities. However, this was not considered sufficient to further downgrade.

e. 38 studies with 50 study groups

f. Risk of Bias: *From Hayden review*: Nine studies (13 groups; 495 participants) were judged to have high risk of bias (17% of participant data). Exclusion of these studies in sensitivity analysis did not change conclusions.

g. Inconsistency: *From Hayden review*: Some unexplained inconsistency (moderate heterogeneity I^2 = 38%, point estimates and confidence intervals varied).

h. Other considerations: *From Hayden review:* Some evidence of publication bias (Egger's test, P = 0.005).

Reference

1. Abenhaim L, Rossignol M, Valat JP, Nordin M, Avouac B, Blotman F et al. The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. Spine (Phila Pa 1976). 2000;25:1s-33s. doi: 10.1097/00007632-200002151-00001.