Web Annex D.A1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

A.1 Structured and standardized education/advice

Overview of the PICO structure

Definition of the intervention	
"Education and/or advice" aims to improve the understanding of the pain experience for a person with CPLBP and guide their self-management and well-being. Evidence reviewed for the guideline included "structured and standardized education and/or advice", defined as the provision of structured/standardized information delivered by health workers(s) to a person with CPLBP. This is distinct and separate from education/advice provided by a health worker to a person with CPLBP as part of a clinical encounter. Structured/standardized advice may not be tailored or personalized. Among the trials identified to inform the guideline, this intervention was delivered by health practitioners.	
PICO question	
Population and subgroups	Community-dwelling adults (aged 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) • Gender and/or sex • Presence of leg pain (radicular, non-radicular, mixed) • Race/ethnicity - studies of populations who were historically marginalized compared with studies of those who were not • 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)

Web Annex D.A1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

Outcomes	Critical outcomes constructs (all adults) • Pain • Back-specific function/disability • General function/disability • Health-related quality of life • Psychosocial function • Social participation • Change in the use of medications • Health literacy • Adverse events (as reported in trials) • Back-specific function/disability • General function/disability • Health-related quality of life • Psychosocial function • Change in the use of medications • Adverse events (as reported in trials) • Falls	Critical outcomes constructs (older adults, aged ≥ 60 years) Pain

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	No evidence identified

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	No evidence identified

Summary of *acceptability considerations*	
All adults	**Older people**
No evidence synthesis commissioned for all adults. Judgements made based on experience of GDG members	Peer support interventions appeared to be acceptable and sought after by some participants. They were seen as an acceptable way of gaining support and sharing information or advice. **# Review findings GRADE-CERQual Assessment of confidence** 21 Participants broadly had positive views of peer support although they found it was difficult to access and did not know of support groups in their area. Empathy and "being believed" through common experience were the most important attributes in a peer supporter. Participants believed it would be helpful to share information and receive or exchange support and advice. LOW

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

Web Annex D.A1: ETD summary for WHO Guideline on non-surgical management of chronic primary low back pain in adults

No evidence synthesis commissioned for all adults. Judgements made based on experience of GDG members	No evidence identified

Summary of judgements

Domain	All adults	Older people
Benefits	Small; trivial	Uncertain
Harms	Trivial; uncertain	Uncertain
Balance benefits to harms	Probably favours the intervention	Probably favours the intervention
Overall certainty	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; varies	Moderate costs; varies
Equity and human rights	Probably increased	Probably increased
Acceptability	Yes	Yes
Feasibility	Yes; probably yes	Yes; probably yes

GRADE evidence profile tables by comparator

GRADE Table 1: ***What are the benefits and harms of education/advice 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	Education or advice	Sham	Relative (95% CI)	Absolute (95% CI)		
ALL ADULTS												
Pain (high-income country, unclassified presence of leg pain) (follow-up: closest to 3 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
1[1]	randomized trials	very serious[a]	not serious[b]	serious[c]	very serious[d]	none	40	40	-	MD **0.22 higher** (0.05 higher to 0.39 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on pain stratified by gender, race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Function (high-income country, unclassified presence of leg pain) (follow-up: closest to 3 months; assessed with: ODI; benefit indicated by lower values; scale: 0 to 50)												
1[1]	randomized trials	very serious[a]	not serious[b]	serious[c]	very serious[d]	none	40	40	-	MD **0.2 higher** (5.7 lower to 6.1 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on function stratified by gender, race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Fear avoidance (high-income country, unclassified presence of leg pain) (follow-up: closest to 3 months; assessed with: FABQ-PA; benefit indicated by lower values; scale: 0 to 24)												
1[1]	randomized trials	very serious[a]	not serious[b]	serious[c]	very serious[d]	none	40	40	-	MD **5.41 higher** (0.28 higher to 10.54 higher)	⨁◯◯◯ Very low	CRITICAL
Fear avoidance (high-income country, unclassified presence of leg pain) (follow-up: closest to 3 months; assessed with: FABQ-W; benefit indicated by lower values; scale: 0 to 42)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	Sham	Relative (95% CI)	Absolute (95% CI)		
1[1]	randomized trials	very serious[a]	not serious[b]	serious[c]	very serious[d]	none	40	40	-	MD **2.64 higher** (0.54 lower to 5.82 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on fear avoidance stratified by gender, race/ethnicity or in adults in low- or lower middle-income countries not identified												
Trials on health-related quality of life, depression, catastrophizing, anxiety or self-efficacy not identified												
0												
Trials on social participation, change in use of medications, adverse events/harms or health literacy not identified												
0												
OLDER ADULTS (aged 60 years or more)												
Trials on pain, function, health-related quality of life, psychological functioning, change in use of medications, falls or adverse events/harms not identified												
0												

CI: confidence interval; **FABQ-PA:** Fear Avoidance Beliefs Questionnaire-Physical Activity outcomes**; FABQ-W:** Fear Avoidance Beliefs Questionnaire-Work outcomes**; MD:** mean difference; **NRS:** numerical rating scale; **ODI:** Oswestry Disability Index; **OIS:** Optimal Information Size

The following was used to guide the ratings.
Risk of bias: *Not serious:* all or most of the weight (>50%) comes from overall low risk of bias trial(s). *Serious:* some of the weight (<50%) comes from overall low risk of bias trial(s). *Very serious:* all or most of the weight (>50%) comes from overall high or unclear risk of bias trial(s).
Inconsistency: *Not serious:* high extent of similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 0% and 40%, which might not be important. *Serious:* some extent of similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 30% and 60%, which could not be explained due to small subgroups and may represent moderate heterogeneity. *Very serious:* little or no similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 50% and 90% or 75% and 100%, which could not be explained due to small subgroups and may represent substantial or considerable heterogeneity, respectively.
Indirectness: *Not serious:* trial(s) were conducted in different countries or settings. *Serious:* trial(s) were conducted from a single country/setting. *Very serious:* evidence is not directly related to PICO question.
Imprecision: *Not serious:* Optimal Information Size (OIS) was reached (i.e., sample sizes with at least 200 participants per group may provide prognostic balance); and the entire confidence interval lies on one side of the threshold that may be considered clinically important (≥10% scale range or SMD ≥0.2 for continuous variables, ≥10% for binary variables), such that the clinical course of action would not differ if the upper versus the lower boundary of the confidence interval represented the truth. *Serious:* OIS would not have been reached (sample sizes with less than 200 participants per group); if the OIS was reached, the clinical course of action might differ if the upper versus the lower boundary of the confidence interval represented the truth. *Very serious:* similar to 'serious' but to a greater extent (e.g., very small sample sizes and confidence intervals crossing appreciable benefit and harm).
Other considerations: *Not serious:* Publication bias is undetected. *Serious/very serious:* Publication bias is strongly suspected.

Explanations
a. We downgraded twice due to two risk of bias domains with high risk and greater than two domains with unclear risk.
b. Inconsistency: We did not downgrade; however, there are no additional studies with which to compare these findings.

c. Indirectness: We downgraded once. This is a single trial from a single country (high-income).
d. Imprecision: We downgraded twice due to small sample size (OIS would have not been reached).

References

1.Jassi FJ, Del Antonio TT,Azevedo BO,Moraes R,George SZ,Chaves TC. Star-Shape Kinesio Taping Is Not Better Than a Minimal Intervention or Sham Kinesio Taping for Pain Intensity and Postural Control in Chronic Low Back Pain: A Randomized Controlled Trial. Arch Phys Med Rehabil; 2021.

GRADE Table 2: ***What are the benefits and harms of education/advice 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 to no intervention or interventions where the effect of education/ advice could be isolated?***

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
ALL ADULTS												
Pain (follow-up: closest to 3 months; assessed with: NRS, VAS, Chronic Pain Questionnaire; benefit indicated by lower values; scale: 0 to 10)												
10[1,2,3,4,5,6,7,8,9,10]	randomized trials	very serious[a]	serious[b]	not serious[c]	serious[d]	none	430	428	-	MD **1.1 lower** (1.63 lower to 0.56 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in males (follow-up: closest to 3 months; assessed with: VAS, Chronic Pain Questionnaire; benefit indicated by lower values; scale: 0 to 10)												
2[1,4]	randomized trials	very serious[a]	not serious[e]	not serious[f]	serious	none	225	225	-	MD **1.12 lower** (1.5 lower to 0.74 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in females and males (follow-up: closest to 3 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
7[2,3,6,7,8,9,10]	randomized trials	very serious[a]	serious[g]	not serious[c]	serious[h]	none	187	186	-	MD **1.16 lower** (2.08 lower to 0.23 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in females (follow-up: closest to 3 months; assessed with: NRS, VAS; benefit indicated by lower values; scale: 0 to 10)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
1[5]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	18	17	-	MD **0.69 lower** (1.56 lower to 0.18 higher)	⨁◯◯◯ Very low	CRITICAL
Pain in people with unclassified presence of leg pain (follow-up: closest to 3 months; assessed with: NRS, VAS, Chronic Pain Questionnaire; benefit indicated by lower values; scale: 0 to 10)												
6[1,3,4,6,8,10]	randomized trials	very serious[a]	serious[l]	not serious[c]	serious[d]	none	349	351	-	MD **1.01 lower** (1.85 lower to 0.17 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in people without leg pain (follow-up: closest to 3 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
2[2,9]	randomized trials	very serious[a]	serious[m]	serious[n]	very serious[k]	none	34	34	-	MD **1.33 lower** (12.08 lower to 9.42 higher)	⨁◯◯◯ Very low	CRITICAL
Pain in people either with or without non-radicular leg pain (follow-up: closest to 3 months; assessed with: NRS, VAS; benefit indicated by lower values; scale: 0 to 10)												
2[5,7]	randomized trials	very serious[a]	serious[b]	not serious[o]	very serious[k]	none	49	43	-	MD **1.15 lower** (7.99 lower to 5.69 higher)	⨁◯◯◯ Very low	CRITICAL
Pain in trials undertaken in low- or lower middle-income countries (follow-up: closest to 3 months; assessed with: VAS, Chronic Pain Questionnaire; benefit indicated by lower values; scale: 0 to 10)												
2[1,4]	randomized trials	very serious[a]	not serious[e]	not serious[f]	serious[d]	none	225	225	-	MD **1.12 lower** (1.5 lower to 0.74 lower)	⨁◯◯◯ Very low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
Pain in trials undertaken in high to upper-middle income countries (follow-up: closest to 3 months; assessed with: NRS, VAS, benefit indicated by lower values; scale: 0 to 10)												
8[2,3,5,6,7,8,9,10]	randomized trials	very serious[a]	serious[p]	not serious[c]	serious[d]	none	205	203	-	MD **1.09 lower** (1.86 lower to 0.31 lower)	⨁◯◯◯ Very low	CRITICAL
Pain stratified by race/ethnicity												
0												
Pain (education intervention: mixed content) (follow-up: closest to 3 months; assessed with: NRS, VAS, Chronic Pain Questionnaire; benefit indicated by lower values; scale: 0 to 10)												
5[1,3,4,6,10]	randomized trials	very serious[a]	not serious[q]	not serious[c]	serious[r]	none	329	332	-	MD **0.8 lower** (1.41 lower to 0.19 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (education intervention: pain neuroscience) (follow-up: closest to 3 months; assessed with: NRS, VAS; benefit indicated by lower values; scale: 0 to 10)												
5[2,5,7,8,9]	randomized trials	very serious[a]	serious[p]	not serious[o]	serious[h]	none	101	96	-	MD **1.47 lower** (2.57 lower to 0.37 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (education intervention delivery mode: combined verbal and written and/or electronic) (follow-up: closest to 3 months; assessed with: NRS, VAS, Chronic Pain Questionnaire; benefit indicated by lower values; scale: 0 to 10)												
7[1,2,4,5,8,9,10]	randomized trials	very serious[a]	serious[s]	not serious[c]	serious[t]	none	322	319	-	MD **1.21 lower** (1.84 lower to 0.57 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (education intervention delivery mode: verbal) (follow-up: closest to 3 months; assessed with: NRS, VAS; benefit indicated by lower values; scale: 0 to 10)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
3[3,6,7]	randomized trials	very serious[a]	serious[u]	not serious[c]	very serious[v]	none	108	109	-	MD **0.68 lower** (3.19 lower to 1.83 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (after removing high risk of bias studies) (follow-up: closest to 3 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
2[2,6]	randomized trials	very serious[a]	very serious[w]	not serious[c]	very serious[x]	none	102	102	-	MD **1.1 lower** (13.41 lower to 11.22 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (follow-up: closest to 6 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
1[6]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	74	74	-	MD **0.55 lower** (1.49 lower to 0.39 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (follow-up: closest to 12 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)												
1[6]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	74	74	-	MD **1.35 lower** (2.34 lower to 0.36 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (follow-up: 2 years; assessed with: VAS; benefit indicated by lower values; scale: 0 to 100)												
1[11]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	40	50	-	MD **8 lower** (18.14 lower to 2.14 higher)	⨁◯◯◯ Very low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
Function (follow-up: closest to 3 months; assessed with: RMDQ, ODI, Chronic Pain Questionnaire, Quebec Back Pain Disability Scale; benefit indicated by lower values)												
10[1,2,3,4,5,6,7,8,9,10]	randomized trials	very serious[a]	serious[p]	not serious[c]	serious[d]	none	430	428	-	SMD **0.51 lower** (0.89 lower to 0.12 lower)	⨁◯◯◯ Very low	CRITICAL
Function in males (follow-up: closest to 3 months; assessed with: RMDQ, Chronic Pain Questionnaire; benefit indicated by lower values)												
2[1,4]	randomized trials	very serious[a]	not serious[e]	not serious[f]	serious[y]	none	225	225	-	SMD **0.4 lower** (0.79 lower to 0)	⨁◯◯◯ Very low	CRITICAL
Function in females and males (follow-up: closest to 3 months; assessed with: RMDQ, ODI, Chronic Pain Questionnaire, Quebec Back Pain Disability Scale; benefit indicated by lower values)												
7[2,3,6,7,8,9,10]	randomized trials	very serious[a]	serious[z]	not serious[o]	serious[aa]	none	187	186	-	SMD **0.55 lower** (1.22 lower to 0.13 higher)	⨁◯◯◯ Very low	CRITICAL
Function in females (follow-up: closest to 3 months; assessed with: RMDQ; benefit indicated by lower values)												
1[5]	randomized trials	very serious[a]	not serious[i]	serious[i]	very serious[k]	none	18	17	-	SMD **0.58 lower** (1.26 lower to 0.1 higher)	⨁◯◯◯ Very low	CRITICAL
Function in people with unclassified presence of leg pain (follow-up: closest to 3 months; assessed with: RMDQ, ODI, Chronic Pain Questionnaire; benefit indicated by lower values)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
6[1,3,4,6,8,10]	randomized trials	very serious[a]	not serious[ab]	not serious[c]	serious[d]	none	349	351	-	SMD **0.35 lower** (0.62 lower to 0.07 lower)	⨁◯◯◯ Very low	CRITICAL
Function in people without leg pain (follow-up: closest to 3 months; assessed with: RMDQ; benefit indicated by lower values)												
2[2,9]	randomized trials	very serious[a]	not serious[e]	serious[n]	very serious[k]	none	34	34	-	SMD **1.46 lower** (3.33 lower to 0.41 higher)	⨁◯◯◯ Very low	CRITICAL
Function in people either with or without non-radicular leg pain (follow-up: closest to 3 months; assessed with: RMDQ, Chronic Pain Questionnaire; benefit indicated by lower values)												
2[5,7]	randomized trials	very serious[a]	not serious[e]	not serious[o]	very serious[k]	none	47	43	-	SMD **0.49 lower** (1.41 lower to 0.43 higher)	⨁◯◯◯ Very low	CRITICAL
Function in trials undertaken in low- or lower middle-income countries (follow-up: closest to 3 months; assessed with: RMDQ, Chronic Pain Questionnaire; benefit indicated by lower values)												
2[1,4]	randomized trials	very serious[a]	not serious[q]	not serious[f]	serious[y]	none	225	225	-	SMD **0.4 lower** (0.79 lower to 0)	⨁◯◯◯ Very low	CRITICAL
Function in trials undertaken in high to upper-middle income countries (follow-up: closest to 3 months; assessed with: RMDQ, ODI, Quebec Back Pain Disability Scale; benefit indicated by lower values)												
8[2,3,5,6,7,8,9,10]	randomized trials	very serious[a]	serious[ac]	not serious[o]	serious[ad]	none	205	203	-	SMD **0.55 lower** (1.1 lower to 0)	⨁◯◯◯ Very low	CRITICAL
Function stratified by race/ethnicity												
0												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
Function (education intervention: mixed content) (follow-up: closest to 3 months; assessed with: RMDQ, ODI, Chronic Pain Questionnaire; benefit indicated by lower values)												
5[1,3,4,6,10]	randomized trials	very serious[a]	not serious[ae]	not serious[c]	serious[y]	none	329	332	-	SMD **0.28 lower** (0.68 lower to 0.11 higher)	⨁◯◯◯ Very low	CRITICAL
Function (education intervention: pain neuroscience) (follow-up: closest to 3 months; assessed with: RMDQ, ODI, Quebec Back Pain Disability Scale; benefit indicated by lower values)												
5[2,5,7,8,9]	randomized trials	very serious[a]	not serious[af]	not serious[o]	serious[ag]	none	101	96	-	SMD **0.87 lower** (1.46 lower to 0.28 lower)	⨁◯◯◯ Very low	CRITICAL
Function (education intervention delivery mode: combined verbal, written, and/or electronic) (follow-up: closest to 3 months; assessed with: RMDQ, ODI, Chronic Pain Questionnaire; benefit indicated by lower values)												
7[1,2,4,5,8,9,10]	randomized trials	very serious[a]	serious[ah]	not serious[c]	serious[ai]	none	322	319	-	SMD **0.68 lower** (1.08 lower to 0.28 lower)	⨁◯◯◯ Very low	CRITICAL
Function (education intervention delivery mode: verbal) (follow-up: closest to 3 months; assessed with: RMDQ, ODI, Quebec Back Pain Disability Scale; benefit indicated by lower values)												
3[3,6,7]	randomized trials	very serious[a]	serious[aj]	not serious[o]	very serious[v]	none	108	109	-	SMD **0.08 lower** (1.52 lower to 1.36 higher)	⨁◯◯◯ Very low	CRITICAL
Function (after removing high risk of bias studies) (follow-up: closest to 3 months; assessed with: RMDQ, ODI; benefit indicated by lower values)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
2[2,6]	randomized trials	very serious[a]	very serious[w]	not serious[o]	very serious[x]	none	102	102	-	SMD **0.74 lower** (9.46 lower to 7.98 higher)	⨁◯◯◯ Very low	CRITICAL
Function (follow-up: closest to 6 months; assessed with: ODI; benefit indicated by lower values; scale: 0 to 100)												
1[6]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	74	74	-	MD **2.86 lower** (7.51 lower to 1.79 higher)	⨁◯◯◯ Very low	CRITICAL
Function (follow-up: closest to 12 months; assessed with: ODI; benefit indicated by lower values; scale: 0 to 100)												
1[6]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	74	74	-	MD **4.66 lower** (9.68 lower to 0.36 higher)	⨁◯◯◯ Very low	CRITICAL
Function (follow-up: 2 years; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												
1[11]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	40	50	-	MD **1.5 lower** (3.42 lower to 0.42 higher)	⨁◯◯◯ Very low	CRITICAL
Health-related quality of life (unclassified presence of leg pain) (follow-up: closest to 3 months; assessed with: SF-36 (PCS); benefit indicated by higher values; scale: 0 to 100)												
2[4,10]	randomized trials	very serious[a]	not serious[e]	not serious[c]	serious[ak]	none	150	149	-	MD **24.27 higher** (12.93 higher to 35.61 higher)	⨁◯◯◯ Very low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
Health-related quality of life (unclassified presence of leg pain) (follow-up: closest to 3 months; assessed with: SF-36 (MCS); benefit indicated by higher values; scale: 0 to 100)												
2[4,10]	randomized trials	very serious[a]	very serious[al]	not serious[c]	very serious[x]	none	125	125	-	MD **13.99 higher** (62.04 lower to 90.03 higher)	⨁◯◯◯ Very low	CRITICAL
Health-related quality of life (follow-up: closest to 3 months; assessed with: WHOQOL-BREF; benefit indicated by higher values; scale: 26 to 130)												
1[3]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	8	9	-	MD **9.4 lower** (17 lower to 1.8 lower)	⨁◯◯◯ Very low	CRITICAL
Fear avoidance (high-income country) (follow-up: closest to 3 months; assessed with: TSK, TSK-11; benefit indicated by lower values)												
5[2,5,7,8,9]	randomized trials	very serious[a]	serious[am]	not serious[o]	serious[ag]	none	72	70	-	SMD **1.4 lower** (2.51 lower to 0.29 lower)	⨁◯◯◯ Very low	CRITICAL
Fear avoidance in females and males (follow-up: closest to 3 months; assessed with: TSK, TSK-11; benefit indicated by lower values)												
4[2,7,8,9]	randomized trials	very serious[a]	serious[an]	not serious[o]	serious[aa]	none	83	79	-	SMD **1.57 lower** (3.21 lower to 0.07 higher)	⨁◯◯◯ Very low	CRITICAL
Fear avoidance in females (follow-up: closest to 3 months; assessed with: TSK-11; benefit indicated by lower values; scale: 11 to 44)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
1[5]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	18	17	-	MD **7.59 lower** (12.63 lower to 2.55 lower)	⨁◯◯◯ Very low	CRITICAL
Fear avoidance in people without leg pain (follow-up: closest to 3 months; assessed with: TSK, TSK-11; benefit indicated by lower values)												
2[2,9]	randomized trials	very serious[a]	not serious[ao]	not serious[o]	very serious[k]	none	34	34	-	SMD **2.12 lower** (7.61 lower to 3.37 higher)	⨁◯◯◯ Very low	CRITICAL
Fear avoidance in people either with or without non-radicular leg pain (follow-up: closest to 3 months; assessed with: TSK, TSK-11; benefit indicated by lower values)												
2[5,7]	randomized trials	very serious[a]	not serious[ap]	not serious[o]	very serious[k]	none	47	43	-	SMD **0.67 lower** (3.89 lower to 2.55 higher)	⨁◯◯◯ Very low	CRITICAL
Fear avoidance in people with unclassified presence of leg pain (follow-up: closest to 3 months; assessed with: TSK; benefit indicated by lower values)												
1[8]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	20	19	-	SMD **1.52 lower** (2.24 lower to 0.8 lower)	⨁◯◯◯ Very low	CRITICAL
Fear avoidance (after removing high risk of bias studies) (follow-up: closest to 3 months; assessed with: TSK, TSK-11; benefit indicated by lower values)												
1[2]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	28	28	-	SMD **1.95 lower** (2.59 lower to 1.31 lower)	⨁◯◯◯ Very low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
Trials on fear avoidance stratified by race/ethnicity or low- or lower middle-income countries not identified												
0												
Fear avoidance (follow-up: 2 years; assessed with: FABQ; benefit indicated by lower values; scale: 13 to 78)												
1[11]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	40	50	-	MD **1 lower** (7.13 lower to 5.13 higher)	⨁◯◯◯ Very low	CRITICAL
Catastrophizing (follow-up: closest to 3 months; assessed with: Pain Catastrophizing Scale; benefit indicated by lower values; scale: 0 to 52)												
2[2,5]	randomized trials	very serious[a]	serious[aq]	not serious[o]	very serious[k]	none	46	45	-	MD **10.19 lower** (55.46 lower to 35.07 higher)	⨁◯◯◯ Very low	CRITICAL
Catastrophizing (females and males, no leg pain) (follow-up: closest to 3 months; assessed with: Pain Catastrophizing Scale; benefit indicated by lower values; scale: 0 to 52)												
1[2]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	28	28	-	MD **13.9 lower** (17.16 lower to 10.64 lower)	⨁◯◯◯ Very low	CRITICAL
Catastrophizing (females, either with or without non-radicular leg pain) (follow-up: closest to 3 months; assessed with: Pain Catastrophizing Scale; benefit indicated by lower values; scale: 0 to 52)												
1[5]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	18	17	-	MD **6.77 lower** (8.48 lower to 5.06 lower)	⨁◯◯◯ Very low	CRITICAL
Catastrophizing in trials undertaken in high to upper-middle income countries (follow-up: closest to 3 months; assessed with: Pain Catastrophizing Scale; benefit indicated by lower values; scale: 0 to 52)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
2[2,5]	randomized trials	very serious[a]	serious[aq]	not serious[o]	very serious[k]	none	46	45	-	MD **10.19 lower** (55.46 lower to 35.07 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on catastrophizing stratified by race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Catastrophizing (after removing high risk of bias studies) (follow-up: closest to 3 months; assessed with: Pain Catastrophizing Scale; benefit indicated by lower values; scale: 0 to 52)												
1[2]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	28	28	-	MD **13.9 lower** (17.16 lower to 10.64 lower)	⨁◯◯◯ Very low	CRITICAL
Depression (females and males, low-income country, either with or without leg pain unclassified) (follow-up: closest to 2 weeks; assessed with: Multidisciplinary Work-related LBP Predictor Questionnaire, Emotional Coping subscale; benefit indicated by higher values; scale: 4 to 20)												
1[12]	randomized trials	very serious[a]	not serious[i]	serious[j]	serious[ag]	none	63	62	-	MD **2.1 higher** (1.05 higher to 3.15 higher)	⨁◯◯◯ Very low	CRITICAL
Depression (females and males, low-income country, either with or without leg pain unclassified) (follow-up: closest to 6 months; assessed with: Multidisciplinary Work-related LBP Predictor Questionnaire, Emotional Coping subscale; benefit indicated by higher values; scale: 4 to 20)												
1[12]	randomized trials	very serious[a]	not serious[i]	serious[j]	serious[ag]	none	63	62	-	MD **1.5 higher** (0.5 higher to 2.5 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on anxiety, depression stratified by gender, race/ethnicity or in high to upper middle-income countries not identified												
0												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
Self-efficacy (females and males, low-income country, either with or without leg pain unclassified) (follow-up: closest to 2 weeks; assessed with: Multidisciplinary Work-related LBP Predictor Questionnaire, Self-efficacy subscale; benefit indicated by higher values; scale: 7 to 35)												
1[12]	randomized trials	very serious[a]	not serious[i]	serious[j]	serious[ag]	none	63	62	-	MD **4.4 higher** (2.77 higher to 6.03 higher)	⨁◯◯◯ Very low	CRITICAL
Self-efficacy (females and males, low-income country, either with or without leg pain unclassified) (follow-up: closest to 6 months; assessed with: Multidisciplinary Work-related LBP Predictor Questionnaire, Self-efficacy subscale; benefit indicated by higher values; scale: 7 to 35)												
1[12,ar]	randomized trials	very serious[a]	not serious[i]	serious[j]	serious[ag]	none	63	62	-	MD **1.6 higher** (0.04 higher to 3.16 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on elf-efficacy stratified by gender, race/ethnicity or in high to upper middle-income countries not identified												
0												
Social participation (paid work) (females and males, high-income country, unclassified presence of leg pain) (follow-up: 2 years; assessed with: number of sickness absence days; benefit indicated by lower values)												
1[11]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	40	50	-	MD **11 lower** (44 lower to 22 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on social participation stratified by gender, race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Trials on change in use of medications or health literacy not identified												
0												
Adverse events/harms (people with uncertain presence of leg pain, high-income country) (follow-up: 2 years)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
1[11]	randomized trials	very serious[a]	not serious[i]	serious[j]	serious[ag]	none	The trial author reported that no adverse events were reported by participants (n=90) during the interventions.				⨁◯◯◯ Very low	CRITICAL

OLDER ADULTS (aged 60 years or more)

Pain (high-income country) (follow-up: closest to 3 months; assessed with: NRS, VAS; benefit indicated by lower values; scale: 0 to 10)

2[3,5]	randomized trials	very serious[a]	not serious[e]	not serious[o]	very serious[k]	none	23	26	-	MD **0.5 lower** (5.42 lower to 4.41 higher)	⨁◯◯◯ Very low	CRITICAL

Pain (females, either with or without non-radicular leg pain) (follow-up: closest to 3 months; assessed with: NRS; benefit indicated by lower values; scale: 0 to 10)

1[5]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	18	17	-	MD **0.69 lower** (1.56 lower to 0.18 higher)	⨁◯◯◯ Very low	CRITICAL

Pain (females and males, unclassified presence of leg pain) (follow-up: closest to 3 months; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)

1[3]	randomized trials	very serious[a]	not serious[i]	serious[j]	very serious[k]	none	5	9	-	**0.3 higher** (2.38 lower to 2.98 higher)	⨁◯◯◯ Very low	CRITICAL

Trials on pain stratified by race/ethnicity or in adults in low- or lower middle-income countries not identified

0												

Function (high-income country) (follow-up: closest to 3 months; assessed with: RMDQ; benefit indicated by lower values)

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
2[3,5]	randomized trials	very serious[a]	very serious[as]	not serious[c]	very serious[k]	none	23	26	-	SMD **0.02 lower** (9.79 lower to 9.76 higher)	⨁◯◯◯ Very low	CRITICAL
Function (females, either with or without non-radicular leg pain) (follow-up: closest to 3 months; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												
1[5]	randomized trials	very serious[a]	not serious[i]	serious[i]	very serious[k]	none	18	17	-	MD **1.12 lower** (2.37 lower to 0.13 higher)	⨁◯◯◯ Very low	CRITICAL
Function (females and males, unclassified presence of leg pain) (follow-up: closest to 3 months; assessed with: RMDQ; benefit indicated by lower values; scale: 0 to 24)												
1[3]	randomized trials	very serious[a]	not serious[i]	serious[i]	very serious[k]	none	5	9	-	MD **4.52 higher** (0.46 higher to 8.58 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on function stratified by race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Fear avoidance (females, high-income country, either with or without non-radicular leg pain) (follow-up: closest to 3 months; assessed with: TSK-11; benefit indicated by lower values)												
1[5]	randomized trials	very serious[a]	not serious[i]	serious[i]	very serious[k]	none	18	17	-	SMD **0.97 lower** (1.68 lower to 0.27 lower)	⨁◯◯◯ Very low	CRITICAL
Trials on fear avoidance in males, stratified by race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Trials on health-related quality of life, depression, catastrophizing, anxiety, self-efficacy, change in use of medications, falls or adverse events/harms not identified												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	No treatment	Relative (95% CI)	Absolute (95% CI)		
0												

CI: confidence interval; **FABQ:** Fear Avoidance Beliefs Questionnaire; **LBP**: low back pain; **MCS:** mental component summary; **MD:** mean difference; **n/a:** non-applicable; **NRS:** numerical rating scale; **ODI:** Oswestry Disability Index; **OIS:** Optimal Information Size; **PCS:** Physical Component Summary; **RMDQ:** Rolland Morris Disability Questionnaire; **SF-36:** short form health survey; **SMD:** standardized mean difference; **TSK:**Tampa Scale of Kinesiophopia; **VAS:** Visual Analogue Scale; **WHOQOL-BREF:** World Health Organization Quality of Life Questionnaire – Brief version

The following was used to guide the ratings.
Risk of bias: *Not serious:* all or most of the weight (>50%) comes from overall low risk of bias trial(s). *Serious:* some of the weight (<50%) comes from overall low risk of bias trial(s). *Very serious:* all or most of the weight (>50%) comes from overall high or unclear risk of bias trial(s).
Inconsistency: *Not serious:* high extent of similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 0% and 40%, which might not be important. *Serious:* some extent of similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 30% and 60%, which could not be explained due to small subgroups and may represent moderate heterogeneity. *Very serious:* little or no similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 50% and 90% or 75% and 100%, which could not be explained due to small subgroups and may represent substantial or considerable heterogeneity, respectively.
Indirectness: *Not serious:* trial(s) were conducted in different countries or settings. *Serious:* trial(s) were conducted from a single country/setting. *Very serious:* evidence is not directly related to PICO question.
Imprecision: *Not serious:* Optimal Information Size (OIS) was reached (i.e., sample sizes with at least 200 participants per group may provide prognostic balance); and the entire confidence interval lies on one side of the threshold that may be considered clinically important (≥10% scale range or SMD ≥0.2 for continuous variables, ≥10% for binary variables), such that the clinical course of action would not differ if the upper versus the lower boundary of the confidence interval represented the truth. *Serious:* OIS would not have been reached (sample sizes with less than 200 participants per group); if the OIS was reached, the clinical course of action might differ if the upper versus the lower boundary of the confidence interval represented the truth. *Very serious:* similar to 'serious' but to a greater extent (e.g., very small sample sizes and confidence intervals crossing appreciable benefit and harm).
Other considerations: *Not serious:* Publication bias is undetected. *Serious/very serious:* Publication bias is strongly suspected.

Explanations
a. Risk of bias: We downgraded twice. All of the trials were rated as overall high or unclear risk of bias.
b. Inconsistency: We downgraded once. There is similarity in the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 54%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
c. Indirectness: We did not downgrade because the trials were conducted in different countries (high and low- or lower middle-income).
d. Imprecision: We downgraded once (studies have small sample sizes ranging from 5 to 125 participants per group). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval does not cross null; however, one of the boundaries crosses the pre-specified threshold (≥ 10% scale range or SMD ≥ 0.2).
e. 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%).
f. Indirectness: We did not downgrade because the trials were conducted in different countries (low- or lower middle-income).
g. Inconsistency: We downgraded once. The point estimates are mostly in the same direction with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 68%). This could not be explained due to small subgroups and may represent substantial heterogeneity.
h. Imprecision: We downgraded once due to small sample size (the OIS would not have been reached). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval does not cross the null; however, one of the boundaries crosses the pre-specified threshold (≥ 10% scale range or SMD ≥ 0.2).
i. Inconsistency: We did not downgrade; however, there are no additional studies with which to compare these findings.
j. Indirectness: We downgraded once. This is a single trial from a single centre (high or upper-middle income).
k. Imprecision: We downgraded twice due to small sample size (the OIS would not have been reached).
l. Inconsistency: We downgraded once. There is similarity in the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 58%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
m. Inconsistency: We downgraded once. There is some overlap in the confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 79%). This could not be explained due to small subgroups and may represent substantial heterogeneity.
n. Indirectness: We downgraded once because the trials were conducted in the same country (high-income).
o. Indirectness: We did not downgrade because the trials were conducted in different countries (high or upper-middle income).

p. Inconsistency: We downgraded once. The point estimates are or are mostly in the same direction with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 64%). This could not be explained due to small subgroups and may represent substantial heterogeneity.
q. Inconsistency: We did not downgrade. The point estimates are mostly similar with overlapping confidence intervals; statistical heterogeneity is between 0% and 40%, which might not be important (i.e., I2 = 30%).
r. Imprecision: We downgraded once (studies have small sample sizes ranging from 5 to 125 participants per group). The point estimate did not reach the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval does not cross null; however, one of the boundaries crosses the pre-specified threshold (≥ 10% scale range or SMD ≥ 0.2).
s. Inconsistency: We downgraded once. There is similarity in most or all of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 52%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
t. Imprecision: We downgraded once (studies have small sample sizes ranging from 6 to 125 participants per group). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval does not cross null; however, one of the boundaries crosses the pre-specified threshold (≥ 10% scale range or SMD ≥ 0.2).
u. Inconsistency: We downgraded once. There are overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 57%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
v. Imprecision: We downgraded twice. The sample size is small (OIS would not have been achieved). The point estimate did not reach the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval crossed the null with the boundaries crossing the thresholds for what may be considered appreciable benefit and harm (MD ≥ 10% scale range or SMD ≥ 0.2).
w. Inconsistency: We downgraded twice. The point estimates differ without overlapping confidence intervals. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 94%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
x. Imprecision: We downgraded twice. The sample size is small (OIS would not have been achieved). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval crossed the null with the boundaries crossing the thresholds for what may be considered appreciable benefit and harm (MD ≥ 10% scale range or SMD ≥ 0.2).
y. Imprecision: We downgraded once (studies have sample sizes ranging from 100 to 125 participants per group). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval crosses the null.
z. Inconsistency: We downgraded once. There similarity is some of the point estimates with some overlap in the confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 76%). This could not be explained due to small subgroups and may represent substantial heterogeneity.
aa. Imprecision: We downgraded once . The sample size is small (OIS would not have been achieved). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval crosses the null.
ab. Inconsistency: We downgraded once. There is similarity in most or all of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 49%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
ac. Inconsistency: We downgraded once. The point estimates are mostly in the same direction with overlapping confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 72%). This could not be explained due to small subgroups and may represent substantial heterogeneity.
ad. Imprecision: We downgraded once (studies have sample sizes ranging from 5 to 74 participants per group). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval crosses the null.
ae. Inconsistency: We did not downgrade. There is similarity in most of 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.
af. 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.
ag. Imprecision: We downgraded once due to small sample size (the OIS would not have been reached).
ah. Inconsistency: We downgraded once. There is similarity in the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 60%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
ai. Imprecision: We downgraded once (studies have small sample sizes ranging from 6 to 125 participants per group).
aj. Inconsistency: We downgraded once. There is similarity in most or all of the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 59%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
ak. Imprecision: We downgraded once (studies have sample sizes ranging from 24 to 125 participants per group). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval does not cross the null.
al. Inconsistency: We downgraded twice. The point estimates are in the same direction with no overlap of confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 89%); this could not be explained due to small subgroups and may represent substantial heterogeneity.
am. Inconsistency: We downgraded once. There is similarity in most of the point estimates and overlap in the confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 78%). This could not be explained due to small subgroups and may represent substantial heterogeneity.
an. Inconsistency: We downgraded once. There is similarity in some of the point estimates and some overlap in the confidence intervals. Statistical heterogeneity is between 50% and 90% (i.e., I2 = 83%). This could not be explained due to small subgroups and may represent substantial heterogeneity.

ao. 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 = 34%).
ap. 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 = 34%).
aq. Inconsistency: We downgraded once. The point estimates differ without overlapping confidence intervals, but are in the same direction. Statistical heterogeneity is between 75% and 100% (i.e., I2 = 93%); this could not be explained due to small subgroups and may represent considerable heterogeneity.
ar. An additional report of the same trial (Shojaei 2017, Ref. ID 22030) also assessed self-efficacy at 6 months with another scale (The Behaviour Questionnaire). We reported the estimate obtained with the Multidisciplinary Work-related LBP Predictor Questionnaire (self-efficacy subscale), since it was also used to assess self-efficacy in the immediate term (closest to 2 weeks) (Shojaei 2017, Ref. ID 25009).
as. We downgraded twice because there was high statistical heterogeneity (I2 = 81%) which could not be explained due to small subgroups. Education was favoured in Kim 2022 (SMD = -0.59; 95% CI -1.26 to 0.10); no treatment was favoured in da Silva 2014 (SMD =1.03; 95% CI -0.15 to 2.21).

References

1.Ayanniyi O, Ige GO. Back care education on peasant farmers suffering from chronic mechanical low back pain. Journal of Experimental and Integrative Medicine; 2015.
2.Bodes Pardo G, Lluch Girbes E,Roussel NA,Gallego Izquierdo T,Jimenez Penick V,Pecos Martin D. Pain Neurophysiology Education and Therapeutic Exercise for Patients With Chronic Low Back Pain: A Single-Blind Randomized Controlled Trial. Arch Phys Med Rehabil; 2018.
3.da Silva TMJC, da Silva NN,de Souza Rocha SH,et al. Back school program for back pain: education or physical exercise?. ConScientiae Saúde; 2014.
4.Ibrahimi Ghavamabadi L, Mohammadi A,Behzadi A,Dehaghi BF. Effectiveness of the training program on the low back pain and functional disability in industrial workers. International Journal of Health Promotion and Education; 2022.
5.Kim KS, An J,Kim JO,Lee MY,Lee BH. Effects of Pain Neuroscience Education Combined with Lumbar Stabilization Exercise on Strength and Pain in Patients with Chronic Low Back Pain: Randomized Controlled Trial. J Pers Med; 2022.
6.Miyamoto GC, Fagundes FRC,de Melo do Espirito Santo C,et al. Education With Therapeutic Alliance Did Not Improve Symptoms in Patients With Chronic Low Back Pain and Low Risk of Poor Prognosis Compared to Education Without Therapeutic Alliance: A Randomized Controlled Trial. 2021.
7.Pires D, Cruz EB,Caeiro C. Aquatic exercise and pain neurophysiology education versus aquatic exercise alone for patients with chronic low back pain: a randomized controlled trial. Clin Rehabil; 2015.
8.Saracoglu I, Arik MI,Afsar E,Gokpinar HH. The effectiveness of pain neuroscience education combined with manual therapy and home exercise for chronic low back pain: A single-blind randomized controlled trial. Physiother Theory Pract; 2020.
9.Tellez-Garcia M, de-la-Llave-Rincon AI,Salom-Moreno J,Palacios-Cena M,Ortega-Santiago R,Fernandez-de-Las-Penas C. Neuroscience education in addition to trigger point dry needling for the management of patients with mechanical chronic low back pain: A preliminary clinical trial. J Bodyw Mov Ther; 2015.
10.Zhang Y, Wan L,Wang X. The effect of health education in patients with chronic low back pain. J Int Med Res; 2014.
11.Rantonen J, Karppinen J,Vehtari A,et al. Effectiveness of three interventions for secondary prevention of low back pain in the occupational health setting - a randomized controlled trial with a natural course control. 2018.
12.Shojaei S, Sadat Tavafian S,Reza Jamshidi A,Wagner J,Reza Sepahvandi M. Social Cognitive Theory-Based Intervention and Low Back Pain among Health Care Workers in Qom Hospitals of Iran. 2017.

GRADE Table 3: ***What are the benefits and harms of education/advice 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	Education or advice	Usual care	Relative (95% CI)	Absolute (95% CI)		
ALL ADULTS												
Pain (high or upper-middle income country) (follow-up: closest to 3 months; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
2[1,2]	randomized trials	very serious[a]	serious[b]	not serious[c]	very serious[d]	none	83	77	-	MD **2.49 lower** (10.73 lower to 5.75 higher)	⨁◯◯◯ Very low	CRITICAL
Pain in people with and without radicular leg pain (follow-up: closest to 3 months; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
1[1]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	42	48	-	MD **1.8 lower** (3.03 lower to 0.57 lower)	⨁◯◯◯ Very low	CRITICAL
Pain in people with and without non-radicular leg pain (follow-up: closest to 3 months; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												
1[2]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	41	29	-	MD **3.1 lower** (4.14 lower to 2.06 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (high-income country, either with or without non-radicular leg pain) (follow-up: closest to 6 months; assessed with: VAS; benefit indicated by lower values; scale: 0 to 10)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	Usual care	Relative (95% CI)	Absolute (95% CI)		
1[2]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	41	29	-	MD **2.1 lower** (3.13 lower to 1.07 lower)	⨁◯◯◯ Very low	CRITICAL
Trials on pain stratified by gender, race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Function (high-income country, either with or without non-radicular leg pain) (follow-up: closest to 3 months; assessed with: ODI; benefit indicated by lower values; scale: 0 to 50)												
1[2]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	41	29	-	MD **7.8 lower** (14.28 lower to 1.32 lower)	⨁◯◯◯ Very low	CRITICAL
Function (high-income country, either with or without non-radicular leg pain) (follow-up: closest to 6 months; assessed with: ODI; benefit indicated by lower values; scale: 0 to 50)												
1[2]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	41	29	-	MD **9.2 lower** (16.5 lower to 1.9 lower)	⨁◯◯◯ Very low	CRITICAL
Trials on function stratified by gender, race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Health-related quality of life (high-income country, either with or without non-radicular leg pain) (follow-up: closest to 3 months; assessed with: SF-36 (PCS); benefit indicated by higher values; scale: 0 to 100)												
1[2]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	41	29	-	MD **2.5 higher** (1.41 lower to 6.41 higher)	⨁◯◯◯ Very low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	Usual care	Relative (95% CI)	Absolute (95% CI)		
Health-related quality of life (high-income country, either with or without non-radicular leg pain) (follow-up: closest to 3 months; assessed with: SF-36 (MCS); benefit indicated by higher values; scale: 0 to 100)												
1[2]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	41	29	-	MD **9.4 higher** (2.7 higher to 16.1 higher)	⨁◯◯◯ Very low	CRITICAL
Health-related quality of life (high-income country, either with or without non-radicular leg pain) (follow-up: closest to 6 months; assessed with: SF-36 (PCS); benefit indicated by higher values; scale: 0 to 100)												
1[2]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	41	29	-	MD **2.4 higher** (1.56 lower to 6.36 higher)	⨁◯◯◯ Very low	CRITICAL
Health-related quality of life (high-income country, either with or without non-radicular leg pain) (follow-up: closest to 6 months; assessed with: SF-36 (MCS); benefit indicated by higher values; scale: 0 to 100)												
1[2]	randomized trials	very serious[a]	not serious[e]	serious[f]	very serious[g]	none	41	29	-	MD **7.2 higher** (0.53 higher to 13.87 higher)	⨁◯◯◯ Very low	CRITICAL
Trials on health-related quality of life stratified by gender, race/ethnicity or in adults in low- or lower middle-income countries not identified												
0												
Trials on psychological functioning, social participation, change in use of medications, health literacy or adverse events/harms not identified												
0												

OLDER ADULTS (aged 60 years or more)

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Education or advice	Usual care	Relative (95% CI)	Absolute (95% CI)		
Trials on pain, function, health-related quality of life, psychological functioning, change in use of medications, falls or adverse events/harms not identified												
0												

CI: confidence interval; **MD:** mean difference; **MCS:** mental component summary**; ODI:** Oswestry Disability Index; **OIS:** Optimal Information Size; **PCS:** Physical Component Summary**; SF-36:** short form health survey; **VAS:** Visual Analogue Scale

The following was used to guide the ratings.
Risk of bias: *Not serious:* all or most of the weight (>50%) comes from overall low risk of bias trial(s). *Serious:* some of the weight (<50%) comes from overall low risk of bias trial(s). *Very serious:* all or most of the weight (>50%) comes from overall high or unclear risk of bias trial(s).
Inconsistency: *Not serious:* high extent of similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 0% and 40%, which might not be important. *Serious:* some extent of similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 30% and 60%, which could not be explained due to small subgroups and may represent moderate heterogeneity. *Very serious:* little or no similarity of point estimates and overlap of confidence intervals; statistical heterogeneity (I^2) is between 50% and 90% or 75% and 100%, which could not be explained due to small subgroups and may represent substantial or considerable heterogeneity, respectively.
Indirectness: *Not serious:* trial(s) were conducted in different countries or settings. *Serious:* trial(s) were conducted from a single country/setting. *Very serious:* evidence is not directly related to PICO question.
Imprecision: *Not serious:* Optimal Information Size (OIS) was reached (i.e., sample sizes with at least 200 participants per group may provide prognostic balance); and the entire confidence interval lies on one side of the threshold that may be considered clinically important (≥10% scale range or SMD ≥0.2 for continuous variables, ≥10% for binary variables), such that the clinical course of action would not differ if the upper versus the lower boundary of the confidence interval represented the truth. *Serious:* OIS would not have been reached (sample sizes with less than 200 participants per group); if the OIS was reached, the clinical course of action might differ if the upper versus the lower boundary of the confidence interval represented the truth. *Very serious:* similar to 'serious' but to a greater extent (e.g., very small sample sizes and confidence intervals crossing appreciable benefit and harm).
Other considerations: *Not serious:* Publication bias is undetected. *Serious/very serious:* Publication bias is strongly suspected.

Explanations
a. Risk of bias: We downgraded twice. Trials were rated as overall high or unclear risk of bias.
b. Inconsistency: We downgraded once. There is similarity in the point estimates with overlapping confidence intervals. Statistical heterogeneity is between 30% and 60% (i.e., I2 = 60%); this could not be explained due to small subgroups and may represent moderate heterogeneity.
c. Indirectness: We did not downgrade because the trials were conducted in different countries (high or upper-middle income).
d. Imprecision: We downgraded twice due to small sample size (OIS would not have been reached). The point estimate reached the pre-specified threshold for what may be considered clinically important (MD ≥ 10% scale range or SMD ≥ 0.2). The confidence interval crosses the null.
e. Inconsistency: We did not downgrade; however, there are no additional studies with which to compare these findings.
f. Indirectness: We downgraded once. This is a single trial from a single centre (high-income country).
g. Imprecision: We downgraded twice due to small sample size (the OIS would not have been reached).

References
1.Akca NK, Aydin G, Gumus K. Effect of Body Mechanics Brief Education in the Clinical Setting on Pain Patients with Lumbar Disc Hernia: A Randomized Controlled Trial. International Journal of Caring Sciences; 2017.
2.Morone G, Paolucci T, Alcuri MR et al. Quality of life improved by multidisciplinary back school program in patients with chronic non-specific low back pain: a single blind randomized controlled trial. 2011.