

Exercise for improving outcomes after osteoporotic vertebral fracture

Lora M Giangregorio¹, Norma J MacIntyre², Lehana Thabane³, Carly J Skidmore¹, and Alexandra Papaioannou⁴

¹Department of Kinesiology, University of Waterloo, Waterloo, Canada

²School of Rehabilitation Science, McMaster University, Hamilton, Canada

³Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada

⁴Department of Medicine, McMaster University, Hamilton, Canada

Abstract

Background—Vertebral fractures are associated with increased morbidity (e.g., pain, reduced quality of life), and mortality. Therapeutic exercise is a non-pharmacologic conservative treatment that is often recommended for patients with vertebral fractures to reduce pain and restore functional movement.

Objectives—Our objectives were to evaluate the benefits and harms of exercise interventions of four weeks or greater (alone or as part of a physical therapy intervention) versus non-exercise/non-active physical therapy intervention, no intervention or place boon the incidence of future fractures and adverse events among adults with a history of osteoporotic vertebral fracture(s). We were also examined the effects of exercise on the following secondary outcomes: falls, pain, posture, physical function, balance, mobility, muscle function, quality of life and bone mineral density of the lumbar spine or hip measured using dual-energy X-ray absorptiometry (DXA). We also reported exercise adherence.

Contact address: Lora M Giangregorio, Department of Kinesiology, University of Waterloo, 200 University Ave W, Waterloo, Ontario, N2L 3G1, Canada. lmgiangr@uwaterloo.ca.

CONTRIBUTIONS OF AUTHORS

LG, AP, and LT conceived the idea for the review. LG, NM, AP, and LT contributed to the design of the protocol, in addition to Ann Cranney and Richard Cook, who were not involved in the preparation of the full report. CS assisted with the literature review and data abstraction, and with preparing figures. LG and NM reviewed full texts for eligibility and independently extracted data from included trials. LG was responsible for writing the protocol and incorporating all authors' feedback.

DECLARATIONS OF INTEREST

L. Giangregorio: None relevant to this paper. She has received a research grant from Merck Frosst that was awarded in conjunction with an Early Researcher Award from the Ontario Ministry of Research and Innovation.

N. MacIntyre: None relevant to this paper.

L. Thabane: is a member of an Advisory Board for Network Meta-analysis for GSK.

C. Skidmore: None

A. Papaioannou: None relevant to this paper. She is or has been a consultant, or on a speaker's bureau, or received unrestricted grants for the following: Amgen, Eli Lilly, Merck Frosst Canada, Novartis, Pfizer, Warner Chilcott.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We reorganized the list of outcomes and removed fear of falling from the list because none of the studies assessed fear of falling. In most cases we chose not to perform any meta-analyses because of the limited number of studies, the availability of data for pooling or the variability across studies with respect to outcomes chosen.

Search methods—We searched the following databases: *The Cochrane Library* (Issue 11 of 12, November 2011), MEDLINE (2005 to 2011), EMBASE (1988 to November 23, 2011), CINAHL (Cumulative Index to Nursing and Allied Health Literature, 1982 to November 23, 2011), AMED (1985 to November 2011), and PEDro (Physiotherapy Evidence Database, www.pedro.fhs.usyd.edu.au/index.html, 1929 to November 23, 2011). Ongoing and recently completed trials were identified by searching the World Health Organization International Clinical Trials Registry Platform (to December 2009). Conference proceedings were searched via ISI and SCOPUS, and targeted searches of proceedings of the American Congress of Rehabilitation Medicine and American Society for Bone and Mineral Research. Search terms or MeSH headings included terms such as vertebral fracture AND exercise OR physical therapy.

Selection criteria—We considered all randomized controlled trials and quasi-randomized trials comparing exercise or active physical therapy interventions with placebo/non-exercise/non-active physical therapy interventions or no intervention implemented in individuals with a history of vertebral fracture and evaluating the outcomes of interest.

Data collection and analysis—Two review authors independently selected trials and extracted data using a pre-tested data abstraction form. Disagreements were resolved by consensus, or third party adjudication. The Cochrane Collaboration's tool for assessing risk of bias was used to evaluate each study. Studies were grouped according to duration of follow-up (i.e., a) four to 12 weeks; b) 16 to 24 weeks; and c) 52 weeks); a study could be represented in more than one group depending on the number of follow-up assessments. For continuous data, we report mean differences (MDs) of the change or percentage change from baseline. Data from two studies were pooled for one outcome using a fixed-effect model.

Main results—Seven trials (488 participants, four male participants) were included. Substantial variability across the seven trials prevented any meaningful pooling of data for most outcomes. No trials assessed the effect of exercise on incident fractures, adverse events or incident falls. Individual trials reported that exercise could improve pain, performance on the Timed Up and Go test, walking speed, back extensor strength, trunk muscle endurance, and quality of life. However, the findings should be interpreted with caution given that there were also reports of no significant difference between exercise and control groups for pain, Timed Up and Go test performance, trunk extensor muscle strength and quality of life. Pooled analyses from two studies revealed a significant between-group difference in favour of exercise for Timed Up and Go performance (MD -1.13 seconds, 95% confidence interval (CI) -1.85 to -0.42 , $P = 0.002$). Individual studies also reported no significant between-group differences for posture or bone mineral density. Adherence to exercise varied across studies. The risk of bias across all studies was variable; low risk across most domains in four studies, and unclear or high risk in most domains for three studies.

Authors' conclusions—No definitive conclusions can be made regarding the benefits of exercise for individuals with vertebral fracture. Although individual trials did report benefits for some pain, physical function and quality of life outcomes, the findings should be interpreted with caution given that findings were inconsistent and the quality of evidence was very low. The small number of trials and variability across trials limited our ability to pool outcomes or make conclusions. Evidence regarding the effects of exercise after vertebral fracture, particularly for men, is scarce. A high-quality randomized trial is needed to inform exercise prescription for individuals with vertebral fractures.

INDEX TERMS

*Exercise Therapy; Back Pain [therapy]; Muscle Strength; Osteoporotic Fractures [*therapy]; Quality of Life; Randomized Controlled Trials as Topic; Spinal Fractures [*therapy]; Walking; Female; Humans; Male

PLAIN LANGUAGE SUMMARY

Exercise for improving outcomes after osteoporotic spine fracture

Researchers in the Cochrane Collaboration conducted a review of the effect of exercise in individuals with osteoporotic spine fractures. After searching for all relevant studies, they found seven studies with a total of 488 people.

What are osteoporotic spine fractures and exercise?

Bone is a living part of your body. Throughout your life, old bone is removed and replaced with new, stronger bone. In someone with osteoporosis, old bone is removed faster than the new bone can replace it, making bones weaker and more likely to break. Exercise is often recommended for people with osteoporosis. Exercise programs may need to be modified for individuals at high risk of fracture, such as individuals with spine fractures due to osteoporosis. It is possible that exercise, if not done correctly, could increase the risk of fracture.

What happens to people with osteoporotic spine fracture who exercise?

In people with a spine fracture due to osteoporosis it is uncertain whether exercise has any effect on pain, speed of moving between sitting and walking tasks, walking speed or quality of life.

No studies were found that looked at whether people had fractures or falls after starting an exercise program.

We often do not have precise information about side effects and complications. This is particularly true for rare but serious side effects. It is uncertain whether exercise will cause any harm.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON [Explanation]

Exercise for improving outcomes after osteoporotic vertebral fracture						
Patient or population: individuals with osteoporotic vertebral fracture						
Settings: outpatient						
Intervention: exercise						
Outcomes	Illustrative comparative risks*		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Exercise				
Fractures ¹ - not measured	See comment	See comment	Not estimable ¹	-	See comment	No studies examined fracture as an outcome.
Adverse events ²	See comment	See comment	Not estimable ²	369 (5)	See comment	Adverse events were reported in the results of 5 studies, but there was no clear description of how they were recorded or monitored in the methods
Falls ³ - not measured	See comment	See comment	Not estimable ³	-	See comment	No studies examined falls as an outcome.
Pain Follow-up: 4 to 24 weeks	See comment	See comment	Not estimable	289 (4 studies)	⊕○○○ very low ^{4,5,6,7,8}	Data could not be pooled. The trials were too diverse with respect to the outcome measures chosen, the duration of follow-up, and the interventions implemented. There was an error in reporting of data for one trial
Physical Function: Mobility Timed Up and Go (TUG) Follow-up: 4 to 52 weeks	The mean mobility in the control group for the largest study ¹¹ was 7.9 (2.6) seconds	The TUG score in the intervention groups was 1.13 (1.85 to 0.42) seconds lower ⁹		119 (3 studies)	⊕○○○ very low ^{4,5,7,8,9}	Two studies were pooled. Two studies were excluded from pooling because of : 1) required data were not available in the report or from authors; 2) multi-component intervention. Notably,

Exercise for improving outcomes after osteoporotic vertebral fracture						
Patient or population: individuals with osteoporotic vertebral fracture						
Settings: outpatient						
Intervention: exercise						
Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Exercise				
						these trials did not observe a significant effect of exercise on Tim ed Up and Go Performance
Physical Function: Mobility Walking speed Follow-up: 12 weeks	The mean mobility in the control group was 0.6 (-0.3 to 1.4) seconds	The mean mobility in the intervention group was 1.3 (0.6 to 2) seconds lower		89 (1 study)	⊕○○○ very low ^{4,8,10}	Only one trial measured this outcome.
Quality of life QUALEFFO-41, OQLQ Follow-up: 12 to 52 weeks	See comment	See comment	Not estimable	183 (3 studies)	⊕○○○ very low ^{4,5,8,9}	Data could not be pooled. The trials were too diverse with respect to the outcome measures chosen, the duration of follow-up, and the interventions implemented

* The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval;

GRADE Working Group grades of evidence

⊕⊕⊕⊕**High quality:** Further research is very unlikely to change our confidence in the estimate of effect.

⊕⊕⊕○**Moderate quality:** Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

⊕⊕○○**Low quality:** Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

⊕○○○**Very low quality:** We are very uncertain about the estimate.

¹ No studies examined fracture as an outcome.

² Adverse events were reported in the results of 4 studies, but there was no clear description of how they were recorded or monitored in the methods.

³ No studies examined falls as an outcome.

⁴ Blinding was not possible due to the nature of the intervention.

⁵ Sequence generation, allocation concealment or how missing data were handled were not adequately described in one or more studies.

⁶ Outcome assessors were not blind to group allocation in one study.

⁷ Data could not be pooled because the trials were too diverse with respect to the outcome measures chosen, the duration of follow-up, and the interventions implemented

⁸ Sample size is small resulting in an imprecise estimate.

⁹ The available data for one study were not sufficient to determine group means and SDs to permit it to be pooled.

¹⁰ There is only one study so the estimate has not been replicated.

¹¹ The baseline TUG score for the control group from Bergland et al was used.

BACKGROUND

Description of the condition

Osteoporosis is a disease of the skeleton characterized by low bone mineral density and deterioration of bone tissue, resulting in an increased risk of fragility fracture (WHO 2003). A fragility fracture is a fracture that occurs with minimal trauma, such as a fall from a standing height. The wrist, hip and vertebra are the most common sites for osteoporotic fragility fractures (Kanis 2001). Fragility fractures, particularly those of the hip and vertebra, are associated with increased mortality and significant morbidity including impaired mobility, pain, deformity, sleep disturbance, depression, fear of future fracture and falling and reduced quality of life (Adachi 2001; Cauley 2000; Papaioannou 2002; Petrella 2000; Wiktorowicz 2001).

Estimating the prevalence and impact of vertebral fractures is difficult; only about 30% of vertebral fractures come to clinical attention as they depend on a report of pain or height loss that triggers the clinician to order a radiograph (Papaioannou 2002). Further, many fractures are not reported even when present on X-ray (Papaioannou 2003b). Estimates of the prevalence of vertebral fractures among men and women have been reported to be similar; prevalent vertebral deformities were found in 23.5% of women and 21.5% of men aged 50 years and older in a Canadian population-based study (Jackson 2000). Similarly, in the United States a prevalence rate of 25.3% in women 50 years and over has been reported (Cooper 1993). Among individuals presenting with a fragility fracture in the United Kingdom, 25% had evidence of vertebral deformity on X-ray (Gallacher 2007). The absolute risk of a subsequent vertebral fracture among women with a prevalent vertebral fracture and osteoporosis based on bone mineral density has been reported to be 50%, compared to 9% among women with no fracture and normal bone mineral density (Cauley 2007). Increasing age and a history of fragility fracture predict subsequent fractures independent of bone mineral density (Kanis 2004). Therefore, individuals who experience vertebral fracture should be targeted for fracture prevention strategies.

Description of the intervention

The management of osteoporosis is multi-modal, and includes pharmacological and non-pharmacological interventions (Avenell 2009; Wells 2008a; Wells 2008b; Wells 2008c). Exercise alone or as part of physical therapy management is often recommended as a beneficial non-pharmacological treatment to slow the rate of bone loss; National and international osteoporosis organizations emphasize the importance of physical activity for preserving bone health. Exercise interventions designed specifically for individuals with vertebral fracture may include postural correction and modified trunk and lower extremity

muscle strengthening exercises with the goal of regaining or maintaining normal spine curvatures, increasing spine stability and improving functional movement and mobility while avoiding postures and physical activities that may increase the risk of fracture (Bennell 2000).

How the intervention might work

Meta-analyses exploring the impact of exercise on bone mineral density in postmenopausal women suggest that weight-bearing exercise or resistance training may decrease the rate of bone loss in women, but may not result in large increases in bone mass measured using dual-energy X-ray absorptiometry (DXA) (Kelley 1998; Martyn-St James 2009; Moayeri 2008;). Meta-analyses of the effect of exercise on bone mineral density in adulthood report point estimates varying from -0.7% to 3.22%, which may be of questionable benefit in the prevention of fractures (Howe 2011; Kelley 1998). Further, studies to date include primarily healthy women without a history of vertebral fracture or in some cases with normal bone mineral density, which may not be generalizable to individuals with vertebral fractures. Exercise may have important effects on bone strength that are not reflected in DXA-based assessments of bone mineral density (Polidoulis 2012). There is also evidence that exercise improves muscle strength and balance and prevent falls (Sherrington 2011), which may indirectly prevent fractures. Moreover, exercise can improve physical function in frail older adults (Chou 2012). However, as with studies of effects of exercise on bone mineral density, restrictive study sampling limits generalizability to individuals with vertebral fractures. Further, individuals with vertebral fractures may have hyperkyphosis, which increases spinal loading (Briggs 2007; Bruno 2012). Exercises that are aimed at reducing kyphotic posture may reduce the risk of vertebral fracture.

Why it is important to do this review

Evidence based clinical practice guidelines for exercise prescription specific to individuals with vertebral fractures are lacking. Exercise guidelines developed for healthy older adults may not be appropriate for individuals with vertebral fractures; the type(s) of exercises and their intensity may need to be modified. Systematic reviews on the topics of interventions for improving mobility after hip fracture, and rehabilitation after distal radius fracture have been developed by The Cochrane Collaboration (Handoll 2006; Handoll 2007). A systematic review of exercise interventions in individuals with vertebral fracture, published in 2010 highlighted the paucity of work in the area (Dusdal 2010). Several of the trials in the latter review were not exclusively individuals with vertebral fracture or were not randomized trials. Further, newer trials may have been published since their search was performed. It is also possible that the risk of fracture might be increased with exercise in individuals with vertebral fractures, but the risk has not been quantified. Therefore, to identify knowledge gaps and inform evidence-based guidelines for exercise prescription specifically in individuals with vertebral fractures, we chose to perform an updated synthesis of trials on the effects of exercise interventions for improving outcomes for individuals with vertebral fracture.

OBJECTIVES

The primary objective of the review was to evaluate the benefits and harms of exercise interventions of four weeks or greater (alone or as part of a physical therapy intervention) versus non-exercise/non-active physical therapy intervention, no intervention or placebo on the incidence of future fractures and adverse events among adults with a history of osteoporotic vertebral fracture(s).

The secondary objectives of the review were to evaluate the effect of exercise interventions of four weeks or greater (alone or as part of a physical therapy intervention) versus non-exercise/non-active physical therapy intervention, no intervention or placebo on the following health-related outcomes among adults with a history of osteoporotic vertebral fracture(s): incident falls; pain; posture; physical function; balance; muscle strength or endurance; health-related quality of life; and bone mineral density of the lumbar spine or hip measured using DXA. We also describe adherence to the intervention(s).

METHODS

Criteria for considering studies for this review

Types of studies—We considered for inclusion all randomized controlled trials (including those in which the treatment allocation was inadequately concealed) or quasi-randomized trials comparing an exercise intervention (alone or as part of a physical therapy intervention) to a non-exercise/non-active physical therapy intervention, no intervention or placebo implemented in individuals with a history of vertebral fracture.

Types of participants—We included studies of men and women over the age of 40 years with a history of non-traumatic or minimal trauma osteoporotic fracture of one or more vertebrae. A non-traumatic fracture was defined as a fracture that occurs spontaneously. A minimal trauma fracture was defined as a fracture that occurs following a:

- fall from standing height;
- fall from sitting position;
- fall from supine position (bed or reclining deck chair < 1 metre high);
- fall after having missed one to three steps in a staircase;
- movement outside of the typical plane of motion or coughing (Bessette 2008).

Types of interventions—Treatment: Trials that involved exercise of any kind, such as: muscle strengthening or resistance training exercises, aerobic exercise, balance training, Tai Chi, or individualized exercise prescribed by a physical therapist were included. Trials examining modalities or devices that did not include an active physical activity component were excluded. Trials that included co-interventions were not excluded. For example, multimodal physical therapy interventions were included if one group received exercise as part of the multimodal intervention and the comparison group received a non-exercise intervention or no intervention. Trials had to include an intervention of at least four weeks duration with subsequent outcome assessment; studies with interventions of less than four weeks duration

or with all outcome assessments prior to four weeks of intervention were excluded. All variations of frequency, intensity and duration of intervention during each session were considered. Comparators: a) non-exercise/non-active physical therapy intervention (e.g. educational intervention); b) no intervention; or c) placebo.

Types of outcome measures

Major outcomes

1. Incident fragility fractures of the hip, vertebra or other sites:
 - fractures confirmed on X-ray; or
 - reduction in vertebral height of greater than 20% (Schousboe 2008) as measured using the Vertebral Fracture Assessment protocol with DXA.
2. Adverse events (other than fragility fractures or falls):
 - serious adverse events that may or may not be related to the study intervention, defined as “any untoward medical occurrence that results in death, is life-threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability/incapacity” (FDA 1995);
 - adverse events, defined as “any unfavorable and unintended sign, symptom, or disease” (FDA 1995). The adverse event should be temporally associated with study participation.

Minor outcomes

1. Incident falls:
 - self-reported falls;
 - falls documented in medical records.
2. Pain:
 - self-report questionnaires specifically developed and validated to assess pain.
3. Posture: measures of postural alignment or spine curvature, e.g., measurements made using clinical devices such as an inclinometer, or from radiographs.
4. Physical function:
 - self-reported questionnaires specifically developed and validated to assess physical function; subscales from validated self-report questionnaires related to physical function, e.g., SF-36;
 - performance-based measures of physical function e.g., documented use of walking aids, six-minute walk test or Timed Up and Go test;
 - other indices of physical function as described in each study.
5. Muscle Function

- Muscle strength or endurance of key muscle groups, e.g., back extensors or lower limb muscles, measured quantitatively.
6. Balance
 - Balance, including self-report questionnaires and performance-based measures e.g., assessments of balance performance using a force plate.
 7. Health-related quality of life: generic and disease-specific self-report questionnaires.
 8. Bone mineral density of the lumbar spine or hip measured using DXA.
 9. Adherence to or acceptability of the intervention(s).

Search methods for identification of studies

Electronic searches—We searched the following databases: *The Cochrane Library* (Issue 11 of 12, November 2011), MEDLINE (2005 to 2011), EM-BASE (1988 to November 23, 2011), CINAHL (Cumulative Index to Nursing and Allied Health Literature, 1982 to November 23, 2011), AMED (1985 to November 2011), and PEDro (Physiotherapy Evidence Database, www.pedro.fhs.usyd.edu.au/index.html, 1929 to November 23, 2011). Ongoing and recently completed trials were identified by searching the World Health Organization International Clinical Trials Registry Platform (to December 2009). We did not apply any language restrictions. MEDLINE searches were undertaken using MeSH headings and text words for vertebral fracture, exercise and physical therapy. The MEDLINE strategy (Appendix 1) was modified for use in *The Cochrane Library*, EMBASE, CINAHL, PEDro and the World Health Organization International Clinical Trials Registry Platform (Appendix 2; Appendix 3; Appendix 4; Appendix 5; Appendix 6; Appendix 7).

Searching other resources—We searched the reference lists of included articles for additional references. We searched conference proceedings using ISI and SCOPUS. We also searched the conference proceedings available online for the American Congress of Rehabilitation Medicine and American Society for Bone and Mineral Research.

Data collection and analysis

Selection of studies—Two review authors (CS, LG) reviewed the title, abstract and descriptors of identified studies for possible inclusion. From the full text, two review authors (LG, NM) independently assessed potentially eligible trials for inclusion. The percentage agreement between review authors during level two screening was 90%. Any disagreement was resolved through discussion. We contacted authors of articles when additional information was needed.

Data extraction and management—Each trial was independently reviewed by two review authors (LG, NM), and data were extracted using a pre-tested data abstraction form. The abstraction form was tested on two articles by the two review authors. Any disagreement was resolved by consensus or third party adjudication. Review authors did not review their own trials.

Assessment of risk of bias in included studies—Risk of bias was assessed independently by two review authors (LG, NM). An assessment tool (Table 1) was developed based on the recommendations in chapter 8 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). The following criteria were used to evaluate risk of bias: random sequence generation; allocation concealment; blinding of participants and personnel; blinding of outcome assessors; addressing incomplete outcome data; and selective outcome reporting. We reviewed each study for the presence or absence of each criterion, and coded it as low risk of bias, unclear risk of bias (uncertain risk of bias) or high risk of bias. Any disagreement regarding risk of bias was resolved by consensus.

Measures of treatment effect—We calculated mean differences (95% confidence intervals, (CI)) for continuous outcomes. We had planned to calculate risk ratios (RR) and corresponding 95% CIs for binary outcomes but there were none to pool. We planned to use the standardized mean difference, as described in chapter 9.4.6 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011), to pool an outcome measured with different scales but the data available for pooling did not permit these calculations. We did not pool data for any outcomes except one because of the limited number of trials. To convert CI to standard deviations, we used the methods described in chapter 7.7.3.2 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).

Unit of analysis issues—No trials were included in pooling that required adjustment for clustering, or correction for design effect in clustered trials that did not adjust for clustering in the analyses and therefore, we did not need to impute any Intra-Cluster Correlation (ICC) to calculate the Variance Inflation Factor (VIF) as described in our protocol.

Dealing with missing data—We reported levels of attrition for included trials. We contacted the authors of trials to provide missing data when data were not reported for some outcomes or groups. We planned to impute missing data and test the effects of imputation on the conclusions that we made (Little 1987).

We analyzed participant data in the group to which they were allocated, independent of whether they received the allocated intervention or not. We planned to re-analyze data for trials where participants were not analyzed in the group to which they were allocated, if there was sufficient information in the trial report or if the data could be retrieved from the study authors. One study (Malmros 1999) did not perform an intention-to-treat analysis, so the raw data were obtained and missing data were imputed using the last observation carried forward method, with the exception of one subscale of the balance test for one individual, where we imputed the mean of all participants in that group because no data were available for any time point. We performed an analysis of variance on the change from baseline for all available time points at which outcomes were assessed (one or more outcomes were assessed at five weeks, 10 weeks and 22 weeks).

Assessment of heterogeneity—We used the Chi² test and I² statistic to quantify any unexplained heterogeneity, where an I² of less than 25% was considered low heterogeneity, an I² of 25 to 50% was considered moderate heterogeneity and an I² of over 50% was considered high heterogeneity (Higgins 2003).

Assessment of reporting biases—We planned to use funnel plots to evaluate publication bias, but there were not enough trials to do so. We had planned to use the capture-mark-recapture technique (i.e., the Horizon Estimate) to estimate the total number of articles in the domain of clinical trials of exercise for individuals with vertebral fracture and the proportion of these articles we were able to capture with our search strategy (Kastner 2007; Kastner 2009). We could not perform the Horizon Estimate analysis due to the small number of studies.

Data synthesis—We planned to pool results of exercise with comparable outcomes using a random-effects approach (95% CI) in Review Manager (RevMan 5). In almost all cases, we did not pool data because of the heterogeneity across trials and the number of trials, or the lack of available data in publications and after queries to authors. In the instance of only two available trials with comparable outcomes, we used a fixed-effect approach (95% CI) in RevMan 5.

The comparative risks, relative effects, number of participants studied and grade of evidence related to the following outcomes are presented in a 'Summary of findings' table as described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011): fractures, adverse events, falls, pain, physical function and quality of life. The GRADE approach was used to rate the quality of the body of evidence for each outcome (Higgins 2011).

Subgroup analysis and investigation of heterogeneity—Heterogeneity between trials may be related to a number of participant-related or intervention-related variables. We had planned to test a number of hypotheses regarding sources of heterogeneity, if applicable, in subgroup analyses, where subgroups would be defined according to the following: level of supervision (continual, intermittent, none), site (home-based versus centre-based), type of exercise, target goal of exercise or positioning during exercise (e.g., strength training, aerobic training, postural retraining, combination, open versus closed chain, dynamic versus static), dose (frequency and duration), intensity, compliance (greater than 75% of sessions versus less than or equal to 75% of sessions), randomized versus quasi-randomized and individual versus cluster-randomized designs. We did not perform these *a priori* sub-group analyses because of the limited number of trials and inability to pool data, but retain the description of our plan to inform future work. We planned to use meta-regression to explain the heterogeneity attributable to these variables. We hypothesized that participants who were continually supervised (i.e., supervision during every session) would achieve greater benefit versus intermittent supervision (i.e., supervision during some, but not all sessions), which would be better than no supervision (i.e., preliminary instructions provided but no consistent monitoring, or less than 5% of sessions monitored) because a greater degree of supervision would provide more opportunity for individualized prescription, feedback regarding proper form, and less fear of adverse events.

Sensitivity analysis—We had planned to use sensitivity analyses to explore the effect of study quality (adequate versus inadequate allocation concealment and blinding) and number of participants with multiple spine fractures at baseline by including them as potential

predictors in the meta-regression when testing for heterogeneity. We were not able to do this because of the limited number of trials.

RESULTS

Description of studies

See: Characteristics of included studies; Characteristics of excluded studies.

Results of the search—We identified 594 references (Figure 1). After reviewing the titles and abstracts, 20 full-text articles and three abstracts were retrieved in the case of journal publications. When abstracts were identified from conference proceedings, they were cross-referenced with the results of the search to ensure that data were not included twice. At least three attempts were made via email to contact the authors of the abstracts to provide additional information when necessary.

Included studies—Seven full journal articles (Bennell 2010; Bergland 2010; Bergstrom 2011, Gold 2004; Malmros 1998; Papaioannou 2003a; Yang 2007) with a total of 488 participants were eligible for inclusion. One study (Gold 2004) was a cluster-randomized trial, where each cluster was a retirement residence. The variance inflation factor (used to adjust the standard deviation in the treatment and control groups which accounts for average cluster size and the ICC) was not reported for this trial, but the site was controlled for in the analysis.

Bennell 2010 was the only trial to include men; four of the 11 participants in the intervention group and none of the nine in the control group were men. Six articles were published in English, and one article was published in Chinese and translated to English (Yang 2007). One published abstract was identified (Miyakoshi 2010), however, attempts to contact the authors to confirm unreported details or data were not successful and the abstract was excluded. The interventions are described in detail in Characteristics of included studies. Characteristics that may introduce heterogeneity in the findings are described below, and include participant characteristics, exercise frequency, intensity and duration, the setting and level of supervision, adherence to the intervention, the outcomes chosen, as well as co-interventions or comparator group activities.

Participant characteristics—All trials had history of vertebral fracture as an inclusion criterion, but the way vertebral fracture was defined varied between studies. Studies defined fracture based on the presence of symptoms (Malmros 1998) or on morphometric changes observed on X-ray, and among those, the definition varied e.g., no definition provided (Bergland 2010; Bergstrom 2011), height reduction of 15% (Papaioannou 2003a) or 20% (Bennell 2010; Gold 2004; Yang 2007), using DXA or radiography. Two studies (Bennell 2010; Malmros 1998) had pain as an inclusion criterion; future studies that aim to evaluate effects on pain associated with vertebral fracture should consider studying effects on individuals with pain at baseline. Variability within a study population or between populations in the severity of vertebral fracture or in the presence of symptoms may result in variable effects of exercise or in generalizability.

Frequency, intensity and duration of the exercise intervention, and duration of follow-up

—There was considerable diversity in the frequency, intensity, and duration of interventions, as well as follow-up periods. Bergland 2010, Bergstrom 2011 and Malmros 1998 reported a recommended exercise frequency of twice per week, with the remaining four studies (Bennell 2010; Gold 2004; Papaioannou 2003a; Yang 2007) reporting a frequency of three times per week. Bennell 2010 reported that muscle strengthening exercises were to be performed three times a week, and that posture training and range of motion exercises should be performed daily. None of the studies reported a specific intensity; when intensity was reported it was often adjusted according to clinical presentation. Four of the studies evaluated exercise interventions that were performed for one to three months (Bennell 2010, Bergland 2010, Malmros 1998, Yang 2007). Bergstrom 2011 implemented a four-month intervention where the stated goal was back muscle strengthening, however, 30 repetitions were prescribed for each exercise. Gold 2004 implemented a six-month exercise intervention (a thrice-weekly group exercise class), and Papaioannou 2003a asked participants to exercise at home thrice weekly for 12 months. In all studies, outcome assessment occurred before and immediately after the intervention period. Two studies had a follow-up outcome assessment when the exercise ceased and also some time later: Bergland 2010 implemented a twice-weekly exercise class for 12 weeks, and outcome assessment was performed at 12 weeks and 12 months, and Malmros 1998 implemented a twice-weekly exercise class for 10 weeks, and outcomes assessment was performed at 10 weeks and 22 weeks. Papaioannou 2003a performed an interim analysis at the half way point, six months after randomization. Gold 2004 also collected interim data at the half way point, three months after randomization, however, the data were not reported, and are not available for inclusion in this review. Gold 2004 had a Phase two where the control group received the intervention for six months, and the intervention group practiced self-maintenance. We did not include the outcomes of Phase two as part of this review as it was no longer a randomized controlled trial.

Setting and supervision—One study included home-based exercise (Papaioannou 2003b), two studies combined a clinic-based physical therapist-led intervention that included a home exercise program (Bennell 2010; Malmros 1999), and the remaining four studies implemented a centre-based intervention (Bergland 2010; Bergstrom 2011; Gold 2004; Yang 2007). Three studies reported continual supervision (Bergland 2010; Bergstrom 2011; Gold 2004), three studies reported intermittent supervision (Bennell 2010; Malmros 1998; Papaioannou 2003a) and for one study, the level of supervision was unclear (Yang 2007). Two interventions were conducted for six months or longer using the following arrangements: 1) intermittent supervision was provided for the first six months followed by telephone follow-up for the final six months (Papaioannou 2003a); 2) continual supervision was provided throughout the six-month intervention in a group exercise class (Gold 2004).

Adherence—The four studies of exercise interventions of four to 12 weeks (Bennell 2010; Bergland 2010; Malmros 1998; Yang 2007) can provide insight into short-term adherence to exercise in women with vertebral fracture. The way adherence was reported varied: Malmros 1998 reported an average adherence of 100% (90% to 100%); Bennell 2010 reported that eight of 11 (73%) of participants achieved 100% adherence to therapy sessions and a median

adherence to home exercise sessions of 95%, with an overall minimum and maximum adherence of 34% and 100%; Bergland 2010 reported that the mean number of completed sessions was 19.5 out of a possible 24 (SD 4.4), and that 24% of participants completed < 19 sessions; and Yang 2007 reported that all participants completed 100% of treatment. Regarding longer-term adherence, Gold 2004 reported a mean attendance to exercise classes of 58% over six months. In Papaioannou 2003a, 62% of participants reported performing their exercises at home three times a week at six months, which declined to 46% at 12 months. An average adherence was not reported in Bergstrom 2011, but it was noted that two individuals could not complete the intervention, two individuals in the intervention group were lost to follow-up and 28 individuals achieved 90% attendance to exercise sessions. For five studies (Bergland 2010; Bergstrom 2011; Gold 2004; Malmros 1998; Papaioannou 2003a), it was not clear whether drop-outs were considered in the estimates of adherence.

Outcomes—Below we report the outcomes and associated measures that were evaluated in the included studies. Webber 2003 included a comparison of vertebral height as a continuous measure obtained in a subset of participants from Papaioannou 2003a, but this study was excluded for several reasons (see Characteristics of Excluded studies).

Fractures, adverse events and falls: No studies evaluated fractures or falls as an outcome. Fractures and falls were reported as adverse events in some studies (Table 2). No studies evaluated between-group differences in adverse events.

Pain: Pain was measured in four studies (Bennell 2010; Gold 2004; Malmros 1998; Yang 2007): Bennell 2010 used an 11-point scale to assess pain intensity on movement in the previous week and pain intensity at rest in the previous week; Gold 2004 measured pain with activities using the pain subscale of the Functional Status Index; Malmros 1999 used an 11-point scale to assess pain intensity in the previous week; and Yang 2007 used a 10 cm visual analogue scale to assess pain intensity (0 = no pain, 10 cm = worst pain, no reference timeframe provided), however, the means provided in the results are all higher than 10, suggesting that they converted the data to a 100 mm scale. Malmros 1999 also used a five-point categorical scale for participants to rate analgesic use, but we did not analyze these data.

Posture: Posture was measured in Bennell 2010 with an inclinometer. Bergstrom 2011 indicated that posture in maximum extension was measured using a kyphometer.

Physical function: Several studies examined between-group differences in self-reported physical function using subscales of a quality of life tool; the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO-41) (Bennell 2010; Bergland 2010) or the Osteoporosis Quality of Life Questionnaire (OQLQ) (Papaioannou 2003a). Physical function was measured using a modified version of the self-reported Oswestry Disability Index (Malmros 1998). A 10-point scale to assess restriction of everyday activities in the previous week was used in Bennell 2010

Several studies examined between-group differences in performance-based measures of physical function; walking speed (Bergland 2010), the Timed Up and Go test (Bennell 2010; Bergland 2010; Yang 2007; Papaioannou 2003a) and the functional reach test (Bergland 2010). Yang 2007 measured time to get up from a supine position.

Balance: Balance was measured with centre of pressure variability recorded using a force plate (Malmros 1998; Papaioannou 2003a).

Muscle function: Trunk muscle endurance was measured by Bennell 2010 using the Timed Loaded Standing test. Bergstrom 2011 measured back extensor muscle strength in standing using an isometric dynamometer. Back extensor muscle strength was measured by Malmros 1998 with a strain gauge. Gold 2004 measured peak isometric torque of the back extensor muscles using the B-200 Isostation.

Health-related quality of life: Quality of life was measured using generic (Bennell 2010; Bergland 2010; Papaioannou 2003a) and disease-specific (Bennell 2010; Bergland 2010; Papaioannou 2003a) questionnaires; questionnaires included the Assessment of Quality of Life (AQoL), the General Health Questionnaire, the Sickness Impact Profile, the QUALEFFO-41 and the OQLQ. Malmros 1999 used a non-validated questionnaire that asked participants to score their quality of life compared to their perceived quality of life at baseline; we did not analyze these data. Although there were several studies that measured QOL, we chose not to pool these data because of the diversity in follow-up times and lack of required data or composite scores for pooling.

Bone mineral density: Papaioannou 2003a was the only study to measure bone mineral density.

Co-Interventions and Comparison Groups—Bennell 2010 implemented a physical therapy intervention that included manual therapy, massage and taping in addition to a home exercise program. The control group did not receive any intervention. An exercise intervention implemented by Gold 2004 was combined with a twice-weekly coping class for 45 minutes designed to reduce psychological concerns common to individuals with vertebral fracture. The control group attended a once-weekly 45-minute class where general health concerns were discussed. In all other studies, the control group received no intervention. No studies excluded individuals on medications for osteoporosis.

Excluded studies—Thirteen full journal articles and three abstracts identified during the search were not included in the review. Two articles (Bada 2009; Sinaki 1995) were reviews or book chapters, five articles (Bautmans 2010; Borgo 2010; Hongo 2007; Lord 1996; Smith 1998) and one abstract did not exclusively study individuals with vertebral fracture, one did not randomize participants to groups (Sinaki 1984), one was a repeat publication of Malmros 1998 but in Danish (Malmros 1999), two were descriptive studies (Rittchen 1991; Schwinning 1992), one included an outcome that was not part of our inclusion criteria (vertebral height as a continuous measure, Webber 2003) and was a subgroup analysis from an included study (Papaioannou 2003a), and one compared standard physical therapy to casting or bracing in individuals with an acute fracture, where individuals less than 40 years

of age were included (Stadhouders 2009). One abstract (Miyakoshi 2010) did not include enough information to verify that it was a randomized trial or that the control group received no intervention, and no estimates of effect were provided. Two abstracts (Shipp 2004; Shipp 2007) were identified that included secondary analyses from the study by Gold 2004. Because the data came from a study that is included and for which we could obtain the relevant information, we report on the findings that exist in the abstract but did not include them as separate studies, clearly identifying that these data came from published abstracts only.

Risk of bias in included studies

Two review authors assessed risk of bias using the scheme presented in Table 1, where YES answers referred to a low risk of bias, NO answers referred to a possibility of bias, and UNCLEAR meant it was difficult to determine whether there was potential for bias with respect to the domain in question. Discrepancies between review authors were resolved via consensus. The results are provided in the Characteristics of included studies table, Figure 2 and Figure 3.

Allocation—In Bennell 2010 and Gold 2004 the allocation sequence was adequately generated and concealed. Malmros 1998 used the drawing of sealed envelopes, so allocation concealment was not ideal. Bergstrom 2011 had participants pick their allocation out of a hat; it was not clear if this was done with replacement nor if the study nurse who performed randomization was the same one that performed the outcome assessment (Bergstrom 2011). In some cases, sequence generation or allocation concealment was not described with sufficient detail (Bergland 2010; Papaioannou 2003a; Yang 2007).

Blinding—Blinding of outcome assessors was confirmed in all studies except for Yang 2007 and Bergstrom 2011. Because of the nature of the intervention, neither participants nor personnel administering the intervention were blind to group allocation for any study.

Incomplete outcome data—In Gold 2004, data were excluded from 63 of a total of 185 participants for the muscle strength outcome because of equipment problems; analyses for this outcome used a sample size of 122. Malmros 1998 performed an as-treated analysis, so we chose to reanalyze the data as intention-to-treat. Papaioannou 2003a reports an intention-to-treat analysis, but it was not clear how missing data were handled. Bergstrom 2011 performed an intention-to-treat analysis for the primary outcome (back extensor muscle strength) but it was not clear if an intention-to-treat or per-protocol analysis was performed for the posture outcome.

Selective reporting—Five trials (Bergland 2010; Bergstrom 2011; Malmros 1998; Papaioannou 2003a; Yang 2007) did not report a clinical trial registration number. Webber 2003 published data on vertebral height obtained in a subset of participants from the trial by Papaioannou 2003a that were not included in the main paper, however these data were evaluated for hypothesis-generating purposes and were not powered to test efficacy of the exercise intervention. Bergstrom 2011 reported that there was no significant difference in C7

to wall distance, but the outcome is not described in the methods, and it is not clear if an intention-to-treat or per-protocol analysis was performed.

Gold 2004 reported data for three primary outcomes. However, the data belong to a much larger database incorporating a number of studies. Not all of the outcomes collected were reported in the paper. Among these, trunk muscle endurance (measured using the Timed Loaded Standing test) and walking endurance (measured using the Six-Minute Walk test) were reported in published abstracts (Shipp 2004; Shipp 2007), but with insufficient detail to enable analysis in the current review. We have been in contact with the authors, who agreed to provide access to the data, but these were not available in time for this version of the review. The authors confirmed that the three outcomes were chosen *a priori* as the only primary outcomes.

The clinical trial registration number for Bennell 2010 was reported, and upon reviewing the registered trial, two outcomes were listed that were not reported in the published article: standing balance on a force platform, and Human Activity Profile self-report questionnaire. Communication with the authors revealed that they were not able to process the standing balance data, so the choice to omit the data was due to logistical difficulties. The Human Activity Profile data were collected for descriptive purposes.

Other potential sources of bias—We chose to use a last observation carried forward method of imputation when reanalyzing the data from Malmros 1998, which may introduce bias in the estimate of treatment effects, especially when compared to methods such as multiple imputation. However, because we were not the authors of the data, we could not verify the source of missingness to confirm whether the data were missing at random. The study reports that data were missing from five individuals (three treatment, two control) because of adverse events, and two other instances were due to participants not completing assessments, but in some cases data points were absent with no known reason.

Effects of interventions

See: **Summary of findings for the main comparison** Exercise for improving outcomes after osteoporotic vertebral fracture

The small number of trials and the variability in the outcomes assessed across studies, the measurement tools chosen for a given outcome and the duration of follow-ups prevented meaningful pooling in meta-analyses for all outcomes except for the Timed Up and Go test (Summary of findings for the main comparison). Studies were grouped according to primary and secondary objectives, and duration of follow-up, where a study could be represented in more than one group depending on the number of follow-up assessments: a) four to 12 weeks (Bennell 2010; Bergland 2010; Malmros 1998; Yang 2007); b) 16 to 24 weeks (Bergstrom 2011; Gold 2004; Malmros 1998; Papaioannou 2003a); and c) 52 weeks (Bergland 2010; Papaioannou 2003a).

There were a few instances where two or more studies included comparable outcomes. However, there were a number of factors that limited our ability to pool data. One study compared a multi-component intervention, including exercise, taping, massage and manual

therapy, compared with no intervention and there was no comparison group that received the same intervention with the exception of exercise (Bennell 2010). Therefore, the effect of exercise could not be estimated and we were unable to pool these data with the other trials. We could not pool data from Papaioannou 2003a because the required means or standard deviations were not provided in the report and the raw data were not available. We chose not to pool data from Gold 2004 because the required means and standard deviations were not included in the report, and because it was a cluster-trial and site was included as a covariate. Because of the limited number of studies and the variability across studies with respect to interventions, outcome measures and analysis, we have chosen to present a narrative synthesis of the findings, grouped according to outcome and follow-up time. The exception was the Timed Up and Go outcome measure, where data from two trials (Bergland 2010; Yang 2007) were pooled (Analysis 1.1). Whether group mean values or mean changes are compared is indicated when known. In Bennell 2010, group means adjusted for baseline value were compared using ANCOVA for each variable. Bennell 2010; Bergland 2010; Bergstrom 2011; Papaioannou 2003a; Yang 2007 did not correct for multiple comparisons.

Fractures—No studies measured fractures as an outcome. Fractures were reported as adverse events (Table 2).

Adverse events—No studies specifically indicated that adverse events were included as an outcome, or described a method for assessing and recording adverse events throughout the trial. Adverse events were reported in the results section of four trials (Bennell 2010; Bergstrom 2011; Gold 2004; Papaioannou 2003a), and one trial (Malmros 1999) indicated that there were five adverse events unrelated to study participation, but did not describe them. Table 2 summarizes the reported adverse events. Four events, including three fractures, were directly attributable to exercise.

Falls—There were no studies that evaluated the effects of exercise on the incidence of falls.

Pain

After four to 12 weeks: No significant effect of exercise was observed for pain after four weeks of exercise in Yang 2007, or after five weeks in our intention-to-treat analysis of data from Malmros 1999. After a 10-week multi-modal physical therapy intervention that included exercise, Bennell 2010 reported a significant between-group difference in favour of the intervention for pain on movement (mean change score -1.8 points [95% CI -3.5 to -0.1], $P < 0.05$) and pain at rest (mean change score -2.0 points [95% CI -3.8 to -0.2], $P < 0.05$). In our intention-to-treat analysis, Malmros 1999 demonstrated a significant between-group difference in change scores in favour of the intervention group for pain (mean -1.03 points [95% CI -1.37 to -0.69], $P = 0.013$) after a 10-week exercise program implemented by physical therapists. However, if a Bonferroni correction was made to account for multiple comparisons (at least six outcomes at one or more time points), the effect is no longer significant. It should be noted that in the original paper, Malmros 1999 did not report whether there were significant between-group differences for pain at any time point, but stated across all time points (five weeks, 10 weeks, 22 weeks) that “a significant difference was found between the course of values from the two study groups ($P = 0.02$).”

After 16 to 24 weeks: For pain with activities, no significant between-group difference was observed in Gold 2004. The significant difference in change scores we reported from our intention-to-treat analysis of data from Malmros 1999 after 10 weeks of exercises was no longer significant (mean -0.73 points [95% CI -1.0 to 0.36], $P = 0.09$) at 22 weeks follow-up when contact with the physical therapist has ceased.

After 52 weeks: No studies measured pain at this time point.

Posture

After four to 12 weeks: Bennell 2010 reported no significant between-group difference in degrees of thoracic kyphosis (mean between-group difference in change scores -2.9 [95% CI -7.9 to 2.1] degrees).

After 16 to 24 weeks: Bergstrom 2011 reported no significant difference in posture after the intervention.

After 52 weeks: No studies measured posture at this time point.

Physical function

After four to 12 weeks: Bennell 2010 reported no between-group difference in activity restriction, but did report a significant improvement in the physical function subscale of the QUALEFFO-41 (mean between-group difference in change scores -4.8 points [95% CI -9.2 to -0.5], $P < 0.05$) on a zero to 100 scale, where lower scores indicate better physical function. Bergland 2010 reported no significant improvement in the physical function subscale of the QUALEFFO-41 (mean change score -2.1 points [95% CI -4.9 to 0.8] for the intervention group versus -0.6 points [95% CI -2.6 to 1.5] for the control group, $P = 0.40$). No significant effect of exercise on self-reported physical function using the Oswestry questionnaire was observed in our intention-to-treat analysis after 10 weeks by Malmros 1999, which is similar to what was reported in the original paper.

Maximum walking speed over 20 metres in Bergland 2010 was significantly improved after three months of exercise (mean change score -1.3 seconds [95% CI -2.0 to -0.6] for the intervention group versus 0.6 seconds [95% CI -0.3 to 1.4] for the control group, effect size 0.5 , $P < 0.001$). Functional reach was also significantly improved (mean change score 1.7 cm [95% CI 0.1 to 3.1] for the intervention group versus -2.2 cm [95% CI -3.8 to -0.7] for the control group, effect size = 0.6 , $P < 0.001$). Performance on the Timed Up and Go test was significantly better in two studies after exercising for four weeks (mean Timed Up and Go score 11.87 seconds [SD = 2.13] for the intervention group versus 14.4 seconds [SD = 3.08] for the control group, $P < 0.05$, Yang 2007) and after exercising for 12 weeks (mean change score -0.5 seconds [95% CI -0.9 to 0.1] for the intervention group versus 0.4 seconds [95% CI -0.2 to 1.1] for the control group, effect size 0.2 , $P < 0.026$, Bergland 2010). In contrast, one study showed no significant between-group difference in performance on the Timed Up and Go test after 10 weeks of follow-up (mean between-group difference in change scores 0.5 seconds [95% CI -1.6 to 0.6], Bennell 2010). When data from two of the studies (Bergland 2010; Yang 2007) were pooled, the effect of exercise

on performance of the Timed Up and Go test was -1.13 seconds [95% CI -1.85 to -0.42], $P = 0.002$ (Analysis 1.1). Time to get up from a supine position was significantly improved relative to controls after four weeks of exercise (mean time 4.87 seconds [SD = 1.17] for the intervention group versus 7.73 seconds [SD = 1.65] for the control group, $P < 0.001$, Yang 2007).

After 16 to 24 weeks: No significant between-group differences were observed in Papaioannou 2003a for Timed Up and Go test performance or physical function or activities of daily living subscales of the OQLQ.

After 52 weeks: Although Bergland 2010 reported no significant between-group difference in change scores for the physical function subscale of the QUALEFFO-41 after the 12-week intervention, it was significant at one-year follow-up (mean change score -2.5 points [95% CI -5.0 to -0.03] for the intervention group versus -1.0 points [95% CI -1.5 to 3.4] for the control group, effect size 0.3, $P < 0.047$). Between-group differences in the activities of daily living subscale of the OQLQ were significant after one year of exercise (mean score 0.34 points [95% CI -0.11 to 0.79], $P < 0.04$) in Papaioannou 2003a, but not in the physical function subscale.

Maximum walking speed over 20 metres in Bergland 2010 was still significantly improved at 12-month follow-up of a three-month exercise program (mean change score -0.9 seconds [95% CI -1.4 to 0.3] for the intervention group versus 0.6 seconds [95% CI -0.6 to 1.8] for the control group, effect size 0.4 seconds, $P < 0.019$). No significant between-group differences were observed in Papaioannou 2003a or Bergland 2010, for Timed Up and Go test performance. Functional reach was no longer significant at 12-month follow-up in Bergland 2010 (mean change score 1.1 cm [95% CI -0.7 to 2.7] for the intervention group versus -0.3 cm [95% CI -1.0 to -1.6] for the control group).

Balance

After 12 weeks or less: No studies measured balance at this time point.

After 16 to 24 week: Tests of postural sway revealed a significant between-group difference in favour of exercise for the range of displacement during the eyes closed condition (mean change score -0.80 cm [95% CI -1.45 to -0.15], $P = 0.01$), but not for any of the other measured postural sway displacement of velocity variables in the eyes closed condition, or for any postural sway variables in the eyes open condition in Papaioannou 2003a. No significant effect of exercise on postural sway was observed after 22 weeks in our analysis of data from Malmros 1999; in the original paper they reported a trend towards an improvement in balance with exercise.

After 52 weeks: Papaioannou 2003a reported that displacement in lateral and anteroposterior directions and velocity of movement (measured with a force plate) were significantly better in the intervention group compared with control ($P < 0.01$), but no data were provided, and it is not clear if this is for the eyes open or eyes closed condition.

Muscle function

After 4 to 12 weeks: A significant effect of a multi-modal physical therapy intervention including exercise was observed for trunk muscle endurance in Bennell 2010, measured via Timed Loaded Standing (mean change score 46.7 seconds [95% CI 16.1 to 77.3], $P < 0.05$). No statistically significant effect of exercise on back extensor muscle strength was observed after five or 10 weeks by Malmros 1999; the P value reported in the original paper for the 10 week comparison was $P = 0.09$.

After 16 to 24 weeks: For trunk extension muscle strength, a significant between-group difference in favour of exercise was observed in Gold 2004 (between-group difference in change score 10.68 foot pounds [95% CI 6.98 to 14.39], $P < 0.001$, $n = 122$, subgroup of total sample $n = 185$). From an abstract (Shipp 2007) using the study sample from Gold 2004, there was no statistically significant between-group difference in change scores for trunk and arm muscle endurance. Bergstrom 2011 reported no significant difference in back extensor muscle strength when control and intervention groups were compared in intention-to-treat analyses (254 ± 85 N versus 302 ± 108 N, $P = 0.74$). They reported a significant between-group difference in back extensor muscle strength ($P = 0.029$) in a per-protocol analysis, where eight individuals who dropped out or were not compliant with exercise or control activities were excluded, and adjustment was made for baseline differences.

Health-related quality of life: Disease-specific scales

After 4 to 12 weeks: Bennell 2010 reported no significant differences in QUALEFFO-41 total score (mean between-group difference -7.1 points [95% CI -14.9 to 0.8]) or in any subscales other than physical function (see Physical Function section above), after 10 weeks of a multi-modal physical therapy intervention that included exercise. Bergland 2010 reported no significant improvement in the total score (mean change score -2.1 points [95% CI -4.2 to -0.02] for the intervention group versus 0.2 points [95% CI -2.2 to 2.5] for the control group) or any of the subscales of the QUALEFFO-41, with the exception of the mental function subscale (mean change score -5.1 points [95% CI -8.1 to -1.6] for the intervention group versus 2.6 [95% CI -1.8 to 6.9] points for the control group, effect size 0.6 , $P < 0.006$).

After 16 to 24 weeks: Significant between-group differences favouring exercise were observed in several OQLQ subscales in Papaioannou 2003a: symptom (mean change score 0.44 points [95% CI 0.16 to 0.73], $P < 0.003$); emotion (mean change score 0.34 points [95% CI 0.02 to 0.66], $P < 0.01$); and leisure/social (mean change score 0.39 points [95% CI -0.02 to 0.81], $P = 0.03$).

After 52 weeks: Although no significant improvement in QUALEFFO-41 total score was observed after a three-month exercise program, Bergland 2010 reported a significant between-group difference at 12-month follow-up (mean change score -3.3 points [95% CI -5.2 to -1.3] for the intervention group versus -0.4 points [95% CI -2.0 to -2.7] for the control group, effect size 0.3 , $P < 0.019$). The between-group difference in the mental function subscale of the QUAL-EFFO-41 observed at three months was maintained at 12-month follow-up (mean change score -2.6 points [95% CI -6.2 to -0.9] for the intervention

group versus 2.7 points [95% CI -1.0 to 6.5] for the control group, effect size 0.4, $P < 0.04$), and between-group difference in the pain subscale (mean change score -13.6 points [95% CI -19.3 to -7.8] for the intervention group versus -1.8 points [95% CI -7.7 to 4.1] for the control group, effect size 0.5, $P < 0.005$) was also significant at this time point. A significant between-group difference favouring exercise was observed in Papaioannou 2003a in the OQLQ symptom subscale (total OQLQ data not reported): symptom (mean change score 0.38 points [95% CI -0.05 to 0.81], $P < 0.02$). Papaioannou 2003a reported that individuals who completed at least three days of exercise a week had significantly greater improvements in OQLQ symptoms score ($P = 0.017$) and total OQLQ score ($P = 0.048$).

Health-related quality of life: Generic scales

After 4 to 12 weeks: Bennell 2010 reported no significant between-group difference in AQL scores (mean change score 0.10 [95% CI -0.04 to 0.24] points). Bergland 2010 reported a significant between-group difference for GHQ total score (mean change score -3.7 points for the intervention group [95% CI -5.5 to -1.9] versus -0.2 points for the control group [95% CI -2.1 to 1.7], effect size 0.4, $P < 0.009$).

After 16 to 24 weeks: No significant between-group differences were observed in Papaioannou 2003a for the total score or subscales of the Sickness Impact Profile.

After 52 weeks: Although total GHQ scores were improved relative to controls after three months of exercise in Bergland 2010, the between-group difference was no longer significant at 12-month follow-up (mean change score -2.8 points [95% CI -4.6 to -1.0] for the intervention group versus -1.1 points for the control group [95% CI -2.8 to 0.7]).

Bone health outcomes

After 52 weeks: Papaioannou 2003a reported no significant effect of thrice-weekly home exercise for one year on lumbar spine or femoral neck bone mineral density (no data reported).

DISCUSSION

Summary of main results

We have summarized the findings from seven published randomized controlled trials (RCTs). The limited number of studies and the diversity in outcomes reported, measurement tools used for a given construct, and duration of follow-up prevented any meaningful pooling of data. Among studies that had similar outcomes and follow-up durations, there were contradictory findings. Our ability to draw conclusions is limited by the very low quality of the body of evidence.

No trials measured fractures, adverse events or falls as outcomes. Notably, some adverse events occurred that were directly associated with the interventions, including three fractures. Most studies did not report if and how adverse event data were collected in the methodology section, and adverse events were reported inconsistently. It would be important

for future studies to systematically collect and report adverse events as a study outcome, and identify those related to the intervention.

Individual studies demonstrated significant between-group differences in favour of exercise for pain, maximum walking speed, Timed Up and Go test performance, time to get up from a supine position, trunk muscle endurance, back extensor muscle strength, range of displacement during a postural sway test with eyes closed, and quality of life (either as a total score or as one of the subscales of a disease-specific or generic quality of life tool). However, there were also individual studies reporting no significant effects of exercise on some of these, and on other outcome measures representing the same construct; this was true for pain, posture, Timed Up and Go, most elements of a postural sway test, back extensor muscle strength, and quality of life (disease-specific or generic). The significant effects of exercise on mobility outcomes are promising, but the magnitude of the effects were small (between-group differences ~1 second for Timed Up and Go and ~2 seconds for walking speed). Pooled analyses of two trials revealed a small but significant between-group difference in favour of exercise for Timed Up and Go test performance, but these findings should be viewed with caution given that two other trials that could not be included in the pooled analysis did not find significant between-group differences. Further, there was substantial variability in quality across all studies.

Only two studies implemented an intervention or follow-up of 52 weeks, and significant between-group differences in favour of exercise were reported for maximum walking speed, and displacement and velocity of movement on a postural sway test in one of the two studies, and significant between-group differences for disease-specific health-related quality of life (total score or select subscales) were reported in both. However, there were also reports of no significant effect of exercise on Timed Up and Go test performance, some elements of a postural sway test, generic health-related quality of life and bone mineral density (lumbar spine and proximal femur).

Adherence to exercise varied across studies; adherence appeared higher among studies that included supervised, patient-specific assessment and prescription, even if it was intermittent. In the two studies with long-term follow-up, adherence decreased when the level of supervision decreased.

Overall completeness and applicability of evidence

The quality of reporting has improved since the advent of the CONSORT guidelines, but there is room for further improvement. The group means and standard deviations, estimates of effect, 95% confidence intervals and P values were not consistently reported for all outcomes, which would be helpful if future versions of this review are able to pool data. In some cases between-group differences are reported as actual differences and in other cases they are reported as change scores. Ideally one is chosen as a primary outcome but both are reported, or the raw data are archived so they can be shared with researchers performing knowledge syntheses. All of the exercise programs included muscle strengthening, however, the specific exercise regimens were not always adequately described (e.g., which muscles were targeted or magnitude and progression of resistance or number of repetitions). Regarding reporting of the exercise intervention, Bennell 2010 included a table listing

details of individual exercises, which is ideal for readers who will attempt to apply the research to practice; for future studies, this could be included in the body of the paper or as an appendix. It should be mandatory to report the protocol for progressing the exercises. There have been no studies to date examining the effects of exercise in men with vertebral fracture, either as a single group or a subgroup of a larger trial including both genders.

Quality of the evidence

The small number of studies, variability in study design, and high or unclear risk of bias due to some methodological components limits the conclusions that can be made to date regarding the effectiveness of exercise for women with vertebral fracture. All studies were rated as high risk of bias for “blinding of participants and personnel” because the nature of exercise limits the ability to blind participants or those implementing the intervention. All studies were rated as unclear or high risk for at least one other criterion in the assessment of risk of bias (Characteristics of included studies and Figure 3), and the quality of studies was quite variable. In many cases the protocol was not publicly available, as this did not become a requirement until more recently. The imprecision in the estimates of effect for many outcomes (because of inconsistent findings across studies) also contributes to our conclusion that the body of evidence around the benefits of exercise for women with vertebral fracture is of very low quality.

Potential biases in the review process

We restricted the review to randomized controlled trials. An inclusion criterion for all trials was a history of vertebral fracture, but the way vertebral fracture was defined varied between studies (e.g., clinical versus morphometric, variable definitions of morphometric fracture), which may create dissimilar study populations and variable effects, and may affect generalizability. For example, Malmros 1998 and Bennell 2010 included only women with a painful vertebral fracture at baseline, whereas Gold 2004 defined fractures based on morphometry; those with symptomatic fractures may have a different response to exercise for outcomes such as pain or quality of life. When we were able to communicate with authors or review a clinical trial registration, we were able to uncover instances of incomplete outcome reporting (Bennell 2010; Gold 2004). The reasons provided by the authors for the incomplete outcome reporting (e.g., difficulty with equipment, reporting the *a priori* defined outcomes) did not appear to be attempts to hide information. However, it supports the need for clinical trial registration, and for thoroughness in evaluations of risk of bias related to incomplete outcome reporting. There were a few instances of incomplete outcome data (e.g., as-treated analysis, subgroup analysis on participants for whom outcomes could be assessed, no information about how missing data were treated). It is very possible that the lack of effect or contradictory findings observed were related to studies being underpowered, as well as variability in participants, exercise interventions, co-interventions, comparators, follow-up times and outcomes chosen. For example, the multi-modal physical therapy intervention used by Bennell 2010 including manual therapy, massage, taping and exercises may have had interactive effects, and it may not be appropriate to pool this trial with others that intervened with exercise only. Three trials were excluded because of the heterogeneity of study participants (Bautmans 2010; Hongo 2007; Smith 1998). The participants in Bautmans 2010 and Hongo 2007 were women with

postmenopausal osteoporosis, and individuals with vertebral fractures at baseline were not excluded (unless they were recent or symptomatic), so it is possible that a subgroup of participants in each of the studies were representative of the participants who were of interest in the current review. We chose not to include these studies to avoid heterogeneity of effects, as those without vertebral fracture may be different with respect to response to exercise than those with vertebral fracture. Bautmans 2010 reported an improvement in thoracic kyphosis (measured with Spinal-Mouse) in the intervention group compared with controls, but no significant between-group difference in quality of life (QUALEFFO-41) or back pain (Visual Analog Scale); the intervention included postural taping, spinal mobilization, and postural exercises over 18 sessions with a physical therapist. Hongo 2007 included women with and without vertebral fractures, and has the same authors as those in Miyakoshi 2010. However, although the outcomes and authors are similar, the interventions described in each trial are slightly different with respect to exercises chosen, follow-up time, and exercise frequency so it is likely that they are independent trials. No between-group difference in spinal range of motion was reported in Hongo 2007; the exercise intervention was one set of 10 repetitions of back extension exercises performed in prone lying five days a week for four months. Hongo 2007 and Miyakoshi 2010 reported a significant between-group difference in favour of exercise for back extensor muscle strength and quality of life (Japanese Osteoporosis Quality of Life questionnaire).

Agreements and disagreements with other studies or reviews

The observed benefit of exercise for walking speed is consistent with the findings of a recent meta-analysis demonstrating that exercise interventions in frail older adults can improve walking speed (Chou 2012). A systematic review entitled “Effects of therapeutic exercise for persons with osteoporotic vertebral fractures: a systematic review” was published in 2010 (Dusdal 2010). We are in agreement with their position that no definitive conclusions can be drawn about the effects of exercise in this population, and that future research is needed. Several differences between the reviews should be acknowledged. We limited our review to randomized controlled trials of exercise where all of the participants had to have at least one osteoporotic vertebral fracture. Dusdal 2010 included non-randomized trials, and specified that to be included 50% of the study participants had to have had at least one vertebral fracture. Their criteria resulted in the inclusion of an additional three studies (Bautmans 2010; Hongo 2007; Sinaki 1984) and one abstract (Smith 1998) that we had officially excluded (see Excluded studies). We excluded an additional article that may have met their eligibility criteria (Webber 2003). We included two full-length journal publications that they did not; Bergland 2010 was published at the same time as their review and Bergstrom 2011 was published after it. We identified information from secondary data analyses (Shipp 2004; Shipp 2007) under the trial by Gold 2004, albeit we did exclude the abstracts as stand-alone contributors to the review. We chose to re-analyze the data from Malmros 1998 because an intention-to-treat analysis was not published and our re-analysis resulted in a different interpretation of the findings of that study and provided data regarding the effects of the intervention at five, 10 and 22 weeks of follow-up. We did not examine the effects of exercise on range of motion, but it was reported in Dusdal 2010. Finally, we emphasized between-group comparisons only, whereas Dusdal 2010 discussed both within- and between-group comparisons. This is relevant because in some cases the within-group

comparisons were significant but the between-group comparisons were not, and this may affect the interpretation of the efficacy of exercise.

AUTHORS' CONCLUSIONS

Implications for practice

No definitive conclusions can be made regarding the benefits of exercise for women with vertebral fracture, although individual trials did report benefits for some pain, physical function, balance, back extensor muscle strength, trunk muscle endurance and quality of life. These findings should be interpreted with caution, especially given the very low quality of the body of evidence and the contradictory findings reported. The effects on mobility (e.g., improvements in walking speed of 1.3 seconds over 20 metres, or a one second improvement in Timed Up and Go score) were small, and may not be clinically significant. Because all of the exercise programs included muscle strengthening, it is reasonable to suggest that if a reader would like to achieve the potential benefits suggested by any of the trials, muscle strengthening must be included in the exercise prescription. Back extensor muscle strength was often emphasized, but it cannot be confirmed if potential benefits are attributable to the inclusion of progressive resisted back extensor muscle exercises. It does point to some consensus among researchers in the field that this is an important impairment to target. However, Bennell 2010 and Shipp 2007 measured trunk muscle endurance. Supervision may affect adherence. Many of the exercise programs were supervised by a physical therapist, so it is unknown whether an intervention would be more or less effective in a different setting.

Implications for research

The experience of vertebral fracture confers a unique effect on quality of life, fall risk, posture, and pain (Adachi 2001; Cauley 2000; Kado 2007). It may not be appropriate to generalize results of exercise studies carried out in postmenopausal women to individuals with vertebral fracture because the appropriate exercise prescription may not be the same, and the safety and efficacy of exercise may differ. The scarcity and very low quality of available data on the effects of exercise on men and women with vertebral fracture points to the need for a large randomized trial to inform exercise prescription in this population at high-risk of future fracture. All of the trials but one cluster-randomized trial Gold 2004 had less than 100 participants, and only three trials had more than 60 participants. Future research should account for adherence and attrition in sample size calculations: the adherence to trials of thrice-weekly exercise for six months or greater is often less than 60%, and the drop-out rate can be 20% or greater, suggesting that there is a need to develop strategies for enhancing adherence or facilitating behaviour change. Only one study (Bergland 2010) reported a sample size calculation. Outcomes important to people with osteoporotic spine fracture and the health systems, such as falls, fractures, disability and health services costs should be considered for inclusion in future trials. Adverse events should be assessed in both groups (intervention and comparator) as a secondary outcome to ensure that the risks do not outweigh any benefits of an exercise intervention. Future work should also consider whether the number or severity of fractures at baseline, vitamin D levels or bisphosphonate use modify the effect. From a theoretical perspective, many of the interventions target back extensor muscle endurance or back extensor muscle strength to

reduce hyperkyphosis and resist forward flexion of the spine. Future research should specifically state whether the intervention targets muscle strength, endurance or both, and specifically assess change in both strength and endurance outcome measures. Furthermore, it may be prudent to investigate whether change in strength or endurance account for change in other outcome measures such as posture and quantify these effects on vertebral fractures.

Acknowledgments

The project was supported in part by the CIHR grant for the first author under the CIHR Randomized Controlled Trials Mentoring Program. We would like to thank Elizabeth Ghogomu and the Cochrane Musculoskeletal Group for their support with the preparation of this manuscript. Many thanks to Louise Falzon and Tamara Rader for contributing their expertise to the design and conduct of the literature search. Thank you to Drs. Malmros, Bergland, Bennell, Summers and Mortensen for their assistance with our queries. Thank you to Ann Cranney and Richard Cook for your contributions to the protocol.

L. Giangregorio holds a Canadian Institutes for Health Research (CIHR) New Investigator Award and an Early Researcher Award from the Ontario Ministry of Research and Innovation which contributed to the available time and personnel to complete the review. N. MacIntyre holds an award from the CIHR Randomized Controlled Trials Mentoring Program, and L. Giangregorio is a past recipient of this award also. L. Thabane and A. Papaioannou are CIHR clinical trials mentors. A. Papaioannou is the CIHR-Eli Lilly Chair in Osteoporosis and Fracture Prevention.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- Canadian Institutes of Health Research, Canada.

Lora Giangregorio is the recipient of the Randomized Controlled Trials Mentoring Program Award (2008–2010) and is a CIHR New Investigator (2011–2016).

Norma MacIntyre is the recipient of the Randomized Controlled Trials Mentoring Program Award (2010–2012).

- Canadian Institutes of Health Research/Eli Lilly, Canada.

A. Papaioannou holds the Canadian Institutes for Health Research/Eli Lilly Chair in Osteoporosis, Falls and Fracture Prevention.

References to studies included in this review

* Indicates the major publication for the study

Bennell 2010 {published data only}. Bennell KL, Matthews B, Greig A, Briggs A, Kelly A, Sherburn M, et al. Effects of an exercise and manual therapy program on physical impairments, function and quality-of-life in people with osteoporotic vertebral fracture: a randomised, single-blind controlled pilot trial. *BMC Musculoskelet Disord.* 2010; 11:36. [PubMed: 20163739]

Bergland 2010 {published data only}. Bergland A, Thorsen H, Karesen R. Effect of exercise on mobility, balance, and health-related quality of life in osteoporotic women with a history of vertebral fracture: a randomized, controlled trial. *Osteoporosis International.* 2011; 22(6):1863–71. [PubMed: 21060992]

Bergstrom 2011 {published data only}. Bergstrom I, Bergstrom K, Kronhed A, Karlsson S, Brinck J. Back extensor training increases muscle strength in postmenopausal women with osteoporosis, kyphosis and vertebral fractures. *Advances in Physiotherapy.* 2011; 13(3):110–7.

Gold 2004 {published data only}. Gold DT, Shipp KM, Pieper CF, Duncan PW, Martinez S, Lyles KW. Group treatment improves trunk strength and psychological status in older women with vertebral fractures: results of a randomized, clinical trial. *Journal of the American Geriatrics Society.* 2004; 52(9):1471–8. [PubMed: 15341548]

- Malmros 1998 {published data only}. Malmros B, Mortensen L, Jensen MB, Charles P. Positive effects of physiotherapy on chronic pain and performance in osteoporosis. *Osteoporos Int.* 1998; 8(3):215–21. [PubMed: 9797905]
- Papaioannou 2003a {published data only}. Papaioannou A, Adachi JD, Winegard K, Ferko N, Parkinson W, Cook RJ, et al. Efficacy of home-based exercise for improving quality of life among elderly women with symptomatic osteoporosis-related vertebral fractures. *Osteoporosis International.* 2003; 14(8):677–82. [PubMed: 12879220]
- Yang 2007 {published data only}. Yang L, He CQ, Xie W, Lan Q. Effect of pain-free exercises on female osteoporosis patients with spinal compressive fracture. *Journal of Clinical Rehabilitative Tissue Engineering Research.* 2007; 11(45):9108–111.

References to studies excluded from this review

- Bada 2009 {published data only}. Bada AM, Rodriguez JE. What are the best treatments for reducing pain due to compression fractures? *Evidence Based Practice.* 2009; 12(1):4–5.
- Bautmans 2010 {published data only}. Bautmans I, Van Arken J, Van Mackelenberg M, Mets T. Rehabilitation using manual mobilization for thoracic kyphosis in elderly postmenopausal patients with osteoporosis. *Journal of Rehabilitation Medicine.* 2010; 42:129–35. [PubMed: 20140408]
- Borgo 2010 {published data only}. Borgo MJ, Sinaki M. Back progressive resistive exercise program to reduce risk of vertebral fractures. *Journal für Mineralstoffwechsel.* 2010; 17(2):66–71.
- Hongo 2007 {published data only}. Hongo M, Itoi E, Sinaki M, Miyakoshi N, Shimada Y, Maekawa S, et al. Effect of low-intensity back exercise on quality of life and back extensor strength in patients with osteoporosis: a randomized controlled trial. *Osteoporosis International.* 2007; 18(10):1389–95. [PubMed: 17572835]
- Lord 1996 {published data only}. Lord SR, Ward JA, Williams P, Zivanovic E. The effects of a community exercise program on fracture risk factors in older women. *Osteoporosis International.* 1996; 6(5):361–7. [PubMed: 8931030]
- Malmros 1999 {published data only}. Malmros B, Mortensen L, Jensen M, Charles P. Effect of specific physiotherapy on chronic pain, functional level and quality of life in osteoporosis: A prospective randomized single-blind placebo controlled study. *Ugeskrift Laeger.* 1999; 161(33):4636–41.
- Miyakoshi 2010 {published data only}. Miyakoshi N, Horikawa A, Hongo M, Kasukawa Y, Shimada Y. Effects of a simple, low-intensity back extension exercise on quality of life and back extensor strength in patients with osteoporosis and vertebral fractures: a 6-month randomized controlled trial. *Journal of Bone and Mineral Research.* 2010 Vol. Abstract.
- Rittchen 1991 {published data only}. Rittchen A, Petravic B. Frühfunktionelle Behandlung der stabilen Kompressionsbrüche der Brust- und Lendenwirbelsäule. *Krankengymnastik.* 1991; 43(4):336–9.
- Schwinning 1992 {published data only}. Schwinning M. Krankengymnastik nach Wirbelfrakturen der BWS-LWS ohne neurologische symptomatik. *Krankengymnastik.* 1992; 44(1):34–8.
- Shipp 2004 {published data only}. Shipp KM, Pieper CF, Gold DT, Lyles KW. Change in walking endurance among women with prevalent vertebral fractures: associations with concurrent change in impairment, functional status, and disability. *Journal of Bone and Mineral Research.* 2004 Vol. Abstract M454.
- Shipp 2007 {published data only}. Shipp KM, Gold DT, Pieper CF, Lyles KW. Improving trunk strength and endurance in older women with vertebral fractures. *Journal of Bone and Mineral Research.* 2007 Vol. Abstract W398.
- Sinaki 1984 {published data only}. Sinaki M, Mikkelsen BA. Postmenopausal spinal osteoporosis: Flexion versus extension exercises. *Archives of Physical Medicine and Rehabilitation.* 1984; 65:593–6. [PubMed: 6487063]
- Sinaki 1995 {published data only}. Sinaki M. Rehabilitation of osteoporotic fractures of the spine. *Physical Medicine and Rehabilitation.* 1995; 9(1):105–23.

- Smith 1998 {published data only}. Smith, AA., Summers, GD., Baxendale, A., Butterly, RJ. Current research in osteoporosis and bone mineral measurement. London: British Institute of Radiology; 1998. A randomized controlled trial of the effects of weight-bearing exercise and resistance training on risk factors for fracture in women with osteoporosis; p. 128
- Stadhouders 2009 {published data only}. Stadhouders A, Buskens E, Vergroesen DA, Fidler MW, de Nies F, Oner FC. Nonoperative treatment of thoracic and lumbar spine fractures: a prospective randomized study of different treatment options. *Journal of Orthopaedic Trauma*. 2009; 23(8): 588–94. [PubMed: 19704275]
- Webber 2003 {published data only}. Webber CE, Papaioannou A, Winegard KJ, Adachi JD, Parkinson W, Ferko NC, et al. A 6-mo home-based exercise program may slow vertebral height loss. *Journal of Clinical Densitometry*. 2003; 6(4):391–400. [PubMed: 14716053]

Additional references

- Adachi 2001. Adachi JD, Loannidis G, Berger C, Joseph L, Papaioannou A, Pickard L, et al. The influence of osteoporotic fractures on health-related quality of life in community-dwelling men and women across Canada. *Osteoporosis International*. 2001; 12(11):903–8. [PubMed: 11804016]
- Avenell 2009. Avenell A, Gillespie WJ, Gillespie LD, O'Connell D. Vitamin D and vitamin D analogues for preventing fractures associated with involutional and post-menopausal osteoporosis. *Cochrane Database of Systematic Reviews*. 2009; (2)doi: 10.1002/14651858.CD000227.pub3
- Bennell 2000. Bennell K, Khan K, McKay H. The role of physiotherapy in the prevention and treatment of osteoporosis. *Manual Therapy*. 2000; 5(4):198–213. [PubMed: 11052899]
- Bessette 2008. Bessette L, Ste-Marie LG, Jean S, Davison KS, Beaulieu M, Baranci M, et al. Recognizing osteoporosis and its consequences in Quebec (ROCQ); background, rationale and methods of an anti-fracture patient health-management programme. *Contemporary Clinical Trials*. 2008; 29(2):194–210. [PubMed: 17766187]
- Briggs 2007. Briggs AM, van Dieën JH, Wrigley TV, Greig AM, Phillips B, Lo SK, et al. Thoracic kyphosis affects spinal loads and trunk muscle force. *Physical Therapy*. 2007; 87(5):595–607. [PubMed: 17472956]
- Bruno 2012. Bruno AG, Anderson DE, D'Agostino J, Bouxsein ML. The effect of thoracic kyphosis and sagittal plane alignment on vertebral compressive loading. *Journal of Bone and Mineral Research*. 2012; 27(10):2144–51. [PubMed: 22589006]
- Cauley 2000. Cauley JA, Thompson DE, Ensrud KC, Scott JC, Black D. Risk of mortality following clinical fractures. *Osteoporosis International*. 2000; 11(7):556–61. [PubMed: 11069188]
- Cauley 2007. Cauley JA, Hochberg MC, Lui LY, Palermo L, Ensrud KE, Hillier TA, et al. Long-term risk of incident vertebral fractures. *JAMA*. 2007; 298(23):2761–7. [PubMed: 18165669]
- Chou 2012. Chou CH, Hwang CL, Wu YT. Effect of exercise on physical function, daily living activities, and quality of life in the frail older adults: a meta-analysis. *Archives of Physical Medicine and Rehabilitation*. 2012; 93(2):237–44. [PubMed: 22289232]
- Cooper 1993. Cooper C, O'Neill T, Silman A. The epidemiology of vertebral fractures. *European Vertebral Osteoporosis Study Group*. *Bone*. 1993; 14(Suppl1):S89–97. [PubMed: 8110529]
- Dusdal 2010. Dusdal K, Grundmanis J, Luttin K, Ritchie P, Rompre C, Sidhu R, et al. Effects of therapeutic exercise for persons with osteoporotic vertebral fractures: a systematic review. *Osteoporosis International*. 2010; 22:755–69. [PubMed: 21161506]
- FDA 1995. Clinical Safety Data Management: Definitions and Standards for Expedited Reporting. FDA Center for Drug and Evaluation Research; 1995. Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.
- Gallacher 2007. Gallacher SJ, Gallagher AP, McQuillan C, Mitchell PJ, Dixon T. The prevalence of vertebral fracture amongst patients presenting with non-vertebral fractures. *Osteoporosis International*. 2007; 18(2):185–92. [PubMed: 17109062]

- Handoll 2006. Handoll HH, Madhok R, Howe TE. Rehabilitation for distal radial fractures in adults. *Cochrane Database of Systematic Reviews*. 2006; (3)doi: 10.1002/14651858.CD003324.pub2
- Handoll 2007. Handoll HH, Sherrington C. Mobilisation strategies after hip fracture surgery in adults. *Cochrane Database of Systematic Reviews*. 2007; (1)doi: 10.1002/14651858.CD001704.pub3
- Higgins 2003. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003; 327(7414):557–60. [PubMed: 12958120]
- Higgins 2011. Higgins, JPT., Green, S., editors. *The Cochrane Collaboration. The Cochrane Collaboration*; 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1* [updated March 2011]. Available from www.cochrane-handbook.org
- Howe 2011. Howe TE, Shea B, Dawson LJ, Downie F, Murray A, Ross C, et al. Exercise for preventing and treating osteoporosis in postmenopausal women. *Cochrane Database of Systematic Reviews*. 2011; (7)doi: 10.1002/14651858.CD000333.pub2
- Jackson 2000. Jackson SA, Tenenhouse A, Robertson L. Vertebral fracture definition from population-based data: preliminary results from the Canadian Multicenter Osteoporosis Study (CaMos). *Osteoporosis International*. 2000; 11(8):680–7. [PubMed: 11095171]
- Kado 2007. Kado DM, Huang MH, Nguyen CB, Barrett-Connor E, Greendale GA. Hyperkyphotic posture and risk of injurious falls in older persons: the Rancho Bernardo Study. *Journals of Gerontology Series A, Biological Sciences and Medical Sciences*. 2007; 62(6):652–7.
- Kanis 2001. Kanis JA, Oden A, Johnell O, Jonsson B, De Laet C, Dawson A. The burden of osteoporotic fractures: a method for setting intervention thresholds. *Osteoporosis International*. 2001; 12(5):417–27. [PubMed: 11444092]
- Kanis 2004. Kanis JA, Johnell O, De Laet C, Johansson H, Oden A, Delmas P, et al. A meta-analysis of previous fracture and subsequent fracture risk. *Bone*. 2004; 35(2):375–82. [PubMed: 15268886]
- Kastner 2007. Kastner M, Straus S, Goldsmith CH. Estimating the Horizon of articles to decide when to stop searching in systematic reviews: an example using a systematic review of RCTs evaluating osteoporosis clinical decision support tools. *AMIA Annual Symposium Proceedings*. 2007:389–93. [PubMed: 18693864]
- Kastner 2009. Kastner M, Straus SE, McKibbin KA, Goldsmith CH. The capture-mark-recapture technique can be used as a stopping rule when searching in systematic reviews. *Journal of Clinical Epidemiology*. 2009; 62(2):149–57. [PubMed: 18722088]
- Kelley 1998. Kelley GA. Exercise and regional bone mineral density in postmenopausal women: a meta-analytic review of randomized trials. *American Journal of Physical Medicine and Rehabilitation*. 1998; 77(1):76–87. [PubMed: 9482383]
- Little 1987. Little, RJA., Rubin, DB. *Statistical Analysis with Missing Data*. 2. New York: John Wiley and Sons; 1987.
- Martyn-St James 2009. Martyn-St James M, Carroll S. A meta-analysis of impact exercise on postmenopausal bone loss: the case for mixed loading exercise programmes. *British Journal of Sports Medicine*. 2009; 43(12):898–908. [PubMed: 18981037]
- Moayyeri 2008. Moayyeri A. The association between physical activity and osteoporotic fractures: a review of the evidence and implications for future research. *Annals of Epidemiology*. 2008; 18(11):827–35. [PubMed: 18809340]
- Papaioannou 2002. Papaioannou A, Watts NB, Kendler DL, Yuen CK, Adachi JD, Ferko N. Diagnosis and management of vertebral fractures in elderly adults. *American Journal of Medicine*. 2002; 113(3):220–8. [PubMed: 12208381]
- Papaioannou 2003b. Papaioannou A, Parkinson W, Ferko N, Probyn L, Ioannidis G, Jurriaans E, et al. Prevalence of vertebral fractures among patients with chronic obstructive pulmonary disease in Canada. *Osteoporosis International*. 2003; 14(11):913–7. [PubMed: 14551675]
- Petrella 2000. Petrella RJ, Payne M, Myers A, Overend T, Chesworth B. Physical function and fear of falling after hip fracture rehabilitation in the elderly. *American Journal of Physical Medicine and Rehabilitation*. 2000; 79(2):154–60. [PubMed: 10744190]
- Polidoulis 2012. Polidoulis I, Beyene J, Cheung AM. The effect of exercise on pQCT parameters of bone structure and strength in postmenopausal women--a systematic review and meta-analysis of randomized controlled trials. *Osteoporosis International*. Jan; 23(1):39–51.

- Schousboe 2008. Schousboe JT, Vokes T, Broy SB, Ferrar L, McKiernan F, Roux C, et al. Vertebral Fracture Assessment: the 2007 ISCD Official Positions. *Journal of Clinical Densitometry*. 2008; 11(1):92–108. [PubMed: 18442755]
- Sherrington 2011. Sherrington C, Tiedemann A, Fairhall N, Close J, Lord SR. Exercise for preventing falls in older adults: an updated meta-analysis and best practice recommendations. *NSW Public Health Bulletin*. 2011; 22(3–4):78–83. [PubMed: 21632004]
- Wells 2008a. Wells GA, Cranney A, Peterson J, Boucher M, Shea B, Welch V, et al. Alendronate for the primary and secondary prevention of osteoporotic fractures in postmenopausal women. *Cochrane Database of Systematic Reviews*. 2008; (1)doi: 10.1002/14651858.CD001155.pub2
- Wells 2008b. Wells GA, Cranney A, Peterson J, Boucher M, Shea B, Welch V, et al. Etidronate for the primary and secondary prevention of osteoporotic fractures in postmenopausal women. *Cochrane Database of Systematic Reviews*. 2008; (1)doi: 10.1002/14651858.CD003376.pub3
- Wells 2008c. Wells GA, Cranney A, Peterson J, Boucher M, Shea B, Welch V, et al. Risedronate for the primary and secondary prevention of osteoporotic fractures in postmenopausal women. *Cochrane Database of Systematic Reviews*. 2008; (1)doi: 10.1002/14651858.CD004523.pub3
- WHO 2003. WHO Scientific Group on the Prevention and Management of Osteoporosis. Prevention and management of osteoporosis: report of a WHO scientific group. *WHO Technical Report Series*. 2003; 921
- Wiktorowicz 2001. Wiktorowicz ME, Goeree R, Papaioannou A, Adachi JD, Papadimitropoulos E. Economic implications of hip fracture: health service use, institutional care and cost in Canada. *Osteoporosis International*. 2001; 12(4):271–8. [PubMed: 11420776]

APPENDICES

Appendix 1. MEDLINE search strategy

1. Spinal Fractures/
2. exp Spine/
3. (spine\$ or spinal or vertebra\$).tw.
4. exp Fractures, Bone/
5. fractur\$.tw.
6. 2 or 3
7. 4 or 5
8. 6 and 7
9. Vertebroplasty/
10. vertebroplast\$.tw.
11. kyphoplast\$.tw.
12. or/1,8–11
13. exp EXERCISE/
14. Physical Exertion/
15. exp Sports/
16. exp PLIABILITY/

17. exertion\$.tw.
18. exercis\$.tw.
19. sport\$.tw.
20. exp Physical Therapy Modalities/
21. ((physical or motion) adj5 (fitness or therap\$)).tw.
22. (physical\$ adj2 endur\$).tw.
23. manipulat\$.tw.
24. Muscle Strength/
25. ((muscle or strength\$ or extensor or weight) adj (training or resistance)).tw.
26. resistance training.tw.
27. exp rehabilitation/
28. rh.fs.
29. rehab\$.tw.
30. kinesiotherap\$.tw.
31. or/13–30
32. 12 and 31
33. randomized controlled trial.pt.
34. controlled clinical trial.pt.
35. randomized.ab.
36. placebo.ab.
37. drug therapy.fs.
38. randomly.ab.
39. trial.ab.
40. groups.ab.
41. or/33–40
42. (animals not (humans and animals)).sh.
43. 41 not 42
44. 32 and 43

Appendix 2. EMBASE search strategy

1. exp spine fracture/
2. exp spine/

3. (spine\$ or spinal or vertebra\$).tw.
4. exp Fracture/
5. fractur\$.tw.
6. 2 or 3
7. 4 or 5
8. 6 and 7
9. kyphoplasty/ or percutaneous vertebroplasty/
10. vertebroplast\$.tw.
11. kyphoplast\$.tw.
12. or/1,8–11
13. exp EXERCISE/
14. pliability/
15. exertion\$.tw.
16. exercis\$.tw.
17. exp physiotherapy/
18. ((physical or motion) adj5 (fitness or therap\$)).tw.
19. (physical\$ adj2 endur\$).tw.
20. manipulat\$.tw.
21. muscle strength/
22. ((muscle or strength\$ or extensor or weight) adj (training or resistance)).tw.
23. resistance training.tw.
24. exp Rehabilitation/
25. rh.fs.
26. rehab\$.tw.
27. kinesiotherap\$.tw.
28. or/13–27
29. 12 and 28
30. random\$.ti,ab.
31. factorial\$.ti,ab.
32. (crossover\$ or cross over\$ or cross-over\$).ti,ab.
33. placebo\$.ti,ab.
34. (doubl\$ adj blind\$).ti,ab.

35. (singl\$ adj blind\$.ti,ab.
36. assign\$.ti,ab.
37. allocat\$.ti,ab.
38. volunteer\$.ti,ab.
39. crossover procedure.sh.
40. double blind procedure.sh.
41. randomized controlled trial.sh.
42. single blind procedure.sh.
43. or/30–42
44. exp animal/ or nonhuman/ or exp animal experiment/
45. exp human/
46. 44 and 45
47. 44 not 46
48. 43 not 47
49. 29 and 48

Appendix 3. The Cochrane Library search strategy

- #1 MeSH descriptor Spinal Fractures, this term only
- #2 MeSH descriptor Spine explode all trees
- #3 (spine* or spinal or vertebra*):ti,ab
- #4 MeSH descriptor Fractures, Bone explode all trees
- #5 fractur*:ti,ab
- #6 (#2 OR #3)
- #7 (#4 OR #5)
- #8 (#6 AND #7)
- #9 MeSH descriptor Vertebroplasty explode all trees
- #10 vertebroplast*:ti,ab
- #11 kyphoplast*:ti,ab
- #12 (#1 OR #8 OR #9 OR #10 OR #11)
- #13 MeSH descriptor Exercise explode all trees
- #14 MeSH descriptor Physical Exertion, this term only
- #15 MeSH descriptor Sports explode all trees

- #16 MeSH descriptor Pliability, this term only
- #17 exertion*:ti,ab
- #18 exercis*:ti,ab
- #19 sport*:ti,ab
- #20 MeSH descriptor Physical Therapy Modalities explode all trees
- #21 ((physical or motion) near/5 (fitness or therap*)):ti,ab
- #22 (physical* near/2 endur*):ti,ab
- #23 manipul*:ti,ab
- #24 MeSH descriptor Muscle Strength, this term only
- #25 ((muscle or strength* or extensor or weight) next (training or resistance)):ti,ab
- #26 resistance next training:ti,ab
- #27 MeSH descriptor Rehabilitation explode all trees
- #28 Any MeSH descriptor with qualifier: RH
- #29 rehab*:ti,ab
- #30 kinesiotherap*:ti,ab
- #31 (#13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30)
- #32 (#12 AND #31)

Appendix 4. CINAHL search strategy

S55 S41 and S54

S54 S42 or S43 or S44 or S45 or S46 or S47 or S48 or S49 or S50 or S51 or S52 or S53

S53 TI Allocat* random* or AB Allocat* random*

S52 (MH "Quantitative Studies")

S51 (MH "Placebos")

S50 TI Placebo* or AB Placebo*

S49 TI Random* allocat* or AB Random* allocat*

S48 (MH "Random Assignment")

S47 TI Randomi?ed control* trial* or AB Randomi?ed control* trial*

S46 AB singl* blind* or AB singl* mask* or AB doub* blind* or AB doubl* mask* or AB trebl* blind* or AB trebl* mask* or AB tripl* blind* or AB tripl* mask*

S45 TI singl* blind* or TI singl* mask* or TI doub* blind* or TI doubl* mask* or TI trebl* blind* or TI trebl* mask* or TI tripl* blind* or TI tripl* mask*

S44 TI clinical* trial* or AB clinical* trial*

S43 PT clinical trial

S42 (MH “Clinical Trials+”)

S41 S12 and S40

S40 S13 or S14 or S15 or S16 or S17 or S18 or S19 or S20 or S21 or S22 or S23 or S24 or S25 or S26 or S27 or S28 or S29 or S30 or S31 or S32 or S33 or S34 or S35 or S36 or S37 or S38 or S39

S39 ti kinesiotherap* or ab kinesiotherap*

S38 ti rehab* or ab rehab*

S37 (MH “Rehabilitation+”)

S36 ti resistance training or ab resistance training

S35 ti weight resistance or ab weight resistance

S34 ti weight training or ab weight training

S33 ti extensor resistance or ab extensor resistance

S32 ti extensor training or ab extensor training

S31 ti strength resistance or ab strength resistance

S30 ti strength training or ab strength training

S29 ti muscle resistance or ab muscle resistance

S28 ti muscle training or ab muscle training

S27 (MH “Muscle Strength”)

S26 ti manipul* or ab manipul*

S25 ti physical* N2 endur* or ab physical* N2 endur*

S24 ti motion N5 therap* or ab motion N5 therap*

S23 ti motion N5 fitness or ab motion N5 fitness

S22 ti physical N5 therap* or ab physical N5 therap*

S21 ti physical N5 fitness or ab physical N5 fitness

S20 (MH “Physical Therapy+”)

S19 ti sport* or ab sport*

S18 ti exercis* or ab exercis*

S17 ti exertion* or ab exertion*

S16 (MH “Pliability”)

S15 (MH “Sports+”)
 S14 (MH “Exertion+”)
 S13 (MH “Exercise+”)
 S12 S1 or S9 or S10 or S11
 S11 ti kyphoplast* or ab kyphoplast*
 S10 ti vertebroplast* or ab vertebroplast*
 S9 S7 and S8
 S8 S5 or S6
 S7 S2 or S3 or S4
 S6 ti fractur* or ab fractur*
 S5 (MH “Fractures+”)
 S4 ab spine* or spinal or vertebra*
 S3 ti spine* or spinal or vertebra*
 S2 (MH “Spine+”)
 S1 (MH “Spinal Fractures+”)

Appendix 5. AMED search strategy

1. spinal fractures/
2. exp spine/
3. (spine\$ or spinal or vertebra\$).tw.
4. exp fractures bone/
5. fractur\$.tw.
6. 2 or 3
7. 4 or 5
8. 6 and 7
9. vertebroplast\$.tw.
10. kyphoplast\$.tw.
11. or/1,8–10
12. exp exercise/
13. exp exertion/
14. exp sports/
15. pliability/

16. exertion\$.tw.
17. exercis\$.tw.
18. sport\$.tw.
19. Physiotherapy/
20. exp physical therapy modalities/
21. ((physical or motion) adj5 (fitness or therap\$)).tw.
22. (physical\$ adj2 endur\$).tw.
23. manipulat\$.tw.
24. muscle strength/
25. ((muscle or strength\$ or extensor or weight) adj (training or resistance)).tw.
26. resistance training.tw.
27. rehabilitation/
28. rehab\$.tw.
29. kinesiotherap\$.tw.
30. or/12–29
31. 11 and 30

Appendix 6. PEDro search strategy

Search 1

Spinal fracture* in Abstract & Title AND

Fitness training in Therapy AND

Clinical trial in Method

Search 2

Spinal fracture* in Abstract & Title AND

Strength training in Therapy AND

Clinical trial in Method

Search 3

Vertebroplast* in Abstract & Title AND

Fitness training in Therapy AND

Clinical trial in Method

Search 4

Vertebroplast* in Abstract & Title AND

Strength training in Therapy AND
 Clinical trial in Method
 Search 5
 kyphoplast* in Abstract & Title AND
 Fitness training in Therapy AND
 Clinical trial in Method
 Search 6
 kyphoplast* in Abstract & Title AND
 Strength training in Therapy AND
 Clinical trial in Method

Appendix 7. World Health Organization International Clinical Trials Registry Platform search strategy

Search 1
 Spinal fracture* in Title
 Search 2
 Vertebroplast* in Title
 Search 3
 kyphoplast* in Title

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Bennell 2010	
Methods	Randomized controlled trial
Participants	# Randomized: 20 (11 intervention, 9 control) Lost to follow-up: 0 Age: 53–90 (66.2 intervention, 66.3 control) Country: Australia Inclusion criteria: at least 1 painful vertebral crush or wedge fracture (20% reduction in anterior height relative to posterior) sustained 3–24 months previously, osteoporosis (lumbar spine or proximal femur BMD T-score \leq -2.5), 5 years postmenopause, stable medications for 6 months, English speaking, community dwelling Exclusion criteria: contraindications to exercise (heart or pulmonary disease, inflammatory or joint disease or severe osteoarthritis, psychiatric condition, neuromuscular condition), secondary causes of osteoporosis e.g., osteomalacia, radicular signs or symptoms, back pain radiating into lower limb, physiotherapy in the last 6 months or previous participation in pain management program for back pain, allergy to adhesive tape or skin condition preventing use of tape
Interventions	Experimental: <i>Frequency:</i> Strengthening 3 times/week, physical therapy once per week, posture training and range of motion exercises daily <i>Intensity:</i> varies according to clinical presentation <i>Time:</i> not specified <i>Type:</i> Multi-modal physical therapy intervention including manual therapy, massage, taping, posture training and exercise prescription

	<p><i>Duration of intervention:</i> 10 weeks <i>Supervisor:</i> Physical therapist <i>Supervision/Setting:</i> Intermittent, clinic and home <i>Progression of intensity:</i> unknown Control: No intervention Duration of follow-up after baseline assessment: 11 weeks</p>	
Outcomes	<p>Pain: 11-point rating scale for pain at rest and pain with activities Pain/physical function: 11-point scale for restriction of daily activities Quality of life: Assessment of Quality of Life, QUALEFFO-41 Mobility: Timed Up and Go Posture: Dualer Electric Inclinometer Trunk and arm endurance: Timed Loaded Standing</p>	
Notes	<p>Adherence: 82% Adverse Events: <i>Due to intervention:</i></p> <ul style="list-style-type: none"> • shoulder pain (2, 10%) • flare up of wrist injury • sore wrist • sore knee • tape irritation <p><i>Unclear if due to intervention:</i></p> <ul style="list-style-type: none"> • None <p><i>Not due to intervention:</i></p> <ul style="list-style-type: none"> • None <p>Between-group comparison as actual difference or change scores: actual difference adjusted for baseline values (ANCOVA)</p>	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The randomisation sequence was generated a priori using the random number function in Excel (Microsoft Corporation, USA) by an independent investigator not directly involved in assessment of participants. These were concealed in opaque envelopes and stored in a locked central location." Comment: probably done.
Allocation concealment (selection bias)	Low risk	"The envelopes were opened sequentially at the time of intervention assignment by an administrator who informed the treating therapist by facsimilie or email." Comment: probably done.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Comment: Participants and personnel delivering the intervention were not blind to group allocation due to nature of intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	"Participants were assessed at baseline and at 11 weeks by an assessor blinded to group allocation." Comment: probably done.
Incomplete outcome data (attrition bias) All outcomes	Low risk	"All analyses were conducted on an intention-to-treat principle using all randomized participants using the Statistical Package for the Social Sciences (Norusis/ SPSS Inc., Chicago IL, USA)." Comment: Probably done. All participants completed the trial
Selective reporting (reporting bias)	Unclear risk	Comment: The clinical trial registration number was reported, and upon reviewing the registered trial, two outcomes were listed that were not reported in the published article: standing balance on a force platform, and Human Activity Profile. Communication with the authors revealed that they were not able to process the standing balance data, so the choice to omit the data was due to logistical difficulties. The Human Activity Profile Data was collected for descriptive purposes

Bergland 2010		
Methods	Randomized controlled trial	
Participants	# Randomized: 89 (47 intervention, 42 control) Lost to follow-up: 11 (6 intervention, 5 control) at 3 months, 6 additional (1 intervention, 5 control) at 12 months Age: 71.4 (70.8 intervention, 72 control) Country: Norway Inclusion criteria: living at home, ambulatory, 60 years and older, and at least one vertebral fracture verified by x-ray, osteoporosis (lumbar spine or proximal femur BMD T-score ≤ -2.5) Exclusion criteria: recent vertebral fractures, unable to complete questionnaires, Mini-Mental State Examination ≤ 23	
Interventions	Experimental: <i>Frequency:</i> 2 times a week <i>Intensity:</i> adjust to ability <i>Time:</i> 60 minutes <i>Type:</i> balance training, agility training, posture exercises <i>Duration of intervention:</i> 12 weeks <i>Supervisor:</i> Physiotherapist <i>Supervision/Setting:</i> continual supervision in a group exercise class <i>Progression of intensity:</i> yes Control: Usual activities Duration of follow-up after baseline assessment: 12 weeks and 52 weeks	
Outcomes	Maximum walking speed over 20 metres in seconds Balance: Functional Reach test Mobility: Timed Up and Go Quality of life: QUALEFFO-41, General Health Questionnaire	
Notes	Adherence: 81% Adverse Events: None reported Between-group comparison as actual difference or change scores: change scores	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Participants were randomly assigned by a computer-generated list to two groups, intervention (IT) and control (CT)."
Allocation concealment (selection bias)	Low risk	"Researchers not involved in the study performed the randomization by using lots in sealed opaque envelopes." Comment: probably done
Blinding of participants and personnel (performance bias) All outcomes	High risk	Comment: Participants and personnel delivering the intervention were not blind to group allocation due to nature of intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	"All the subjects included in the study were assessed three times by a physiotherapist who did not know which of the two groups the participants belonged to." Comment: probably done.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Analysis was on an intention-to-treat basis, in which missing data at 3 or 12 months for subjects who did not complete the program were replaced by baseline test values." Comment: probably done.
Selective reporting (reporting bias)	Unclear risk	Comment: The protocol is not publicly available for review or the clinical trial registration number was not provided
Bergstrom 2011		
Methods	Randomized controlled trial	
Participants	# Randomized: 36 (20 intervention, 16 control) Lost to follow-up: 3 (2 intervention, 1 control) at 4 months Age: 73.2 intervention, 74.1 control Country: Sweden	

	Inclusion criteria: at least one vertebral fracture (at least 6 months prior), osteoporosis (states diagnosis by DXA but no criteria), bisphosphonate therapy for 4 months, kyphosis (no criteria stated), postmenopausal (no time criteria stated) Exclusion criteria: metabolic disease or secondary osteoporosis, cardiovascular or musculoskeletal complications that would prevent full participation, already participating in exercise at same level	
Interventions	Experimental: <i>Frequency:</i> 2 times a week <i>Intensity:</i> individualized to ability, 30 repetitions of each exercise <i>Time:</i> 60 minutes <i>Type:</i> warm up, exercises for back extensors, shoulder flexors and shoulder abductors, scapular retraction and push-ups, bird dog (thoracic and lumbar extension) <i>Duration of intervention:</i> 16 weeks <i>Supervisor:</i> Physiotherapist <i>Supervision/Setting:</i> continual supervision in a group exercise class <i>Progression of intensity:</i> yes Control: Not described Duration of follow-up after baseline assessment: 16 weeks	
Outcomes	Back Extension Strength: In standing with an isometric dynamometer (in Newtons) Height Posture - kyphometer and C7 to wall distance	
Notes	Adherence: not specifically stated. Per-protocol analyses included those with at least 90% attendance to sessions, two were excluded because of poor (< 50%) attendance, and two were lost to follow-up Although goal was strengthening, 30 repetitions were performed, suggesting that endurance would be the goal	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Comment: Participants picked allocation out of hat.
Allocation concealment (selection bias)	High risk	Comment: Study nurse prepared materials for participant randomization
Blinding of participants and personnel (performance bias) All outcomes	High risk	Comment: Participants and personnel delivering the intervention were not blind to group allocation due to nature of intervention
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	"All measurements were ...blinded as to the participant. The study nurse examined and evaluated all women." Comment: It is not clear if the study nurse that performed randomization also performed outcome assessment
Incomplete outcome data (attrition bias) All outcomes	High risk	Comment: Per-protocol analysis performed only for all secondary outcomes. Although intention-to-treat analysis was performed for primary outcome, per-protocol analysis was also performed in two ways (with and without adjustment for baseline levels of outcome) and the per-protocol analysis is emphasized in interpretation
Selective reporting (reporting bias)	Unclear risk	Comment: No clinical trial registration number available.
Gold 2004		
Methods	Cluster-randomized controlled trial (unit of randomization was retirement facility)	
Participants	# Randomized: 185 (94 intervention, 91 control) Lost to follow-up: 9 intervention, 5 control Age: 81 Country: USA Inclusion criteria: Women, at least one vertebral fracture (reduction in vertebral height compared to adjacent vertebrae of 20% or greater) Exclusion criteria: Vertebral fracture in the last 6 months, hip fracture in the last 12 months, more than 2 errors on the short portable mental status questionnaire, corrected vision worse than 20/40, active cardiac, pulmonary, neurological disease or rheumatoid arthritis, injurious fall in the last 6 months, not an independent ambulator	
Interventions	Experimental:	

	<p><i>Frequency:</i> 3 times/week for exercise, 2 times/week for coping class <i>Intensity:</i> not specified <i>Time:</i> 45 min each for exercise class and coping class <i>Type:</i> stretching, strengthening with weights and resistance bands, instruction on optimal alignment and body mechanics, coping class <i>Duration of intervention:</i> 6 months <i>Supervisor:</i> physical therapist <i>Supervision/Setting:</i> Group class in retirement facility <i>Progression of intensity:</i> Yes Control: weekly class on general health information Duration of follow-up after baseline assessment: 6 months</p>	
Outcomes	<p>Trunk extension strength: Peak isometric strength in foot pounds-mean of 3 maximal trials using B-200 isostation Pain - subscale of Functional Status Index Psychological symptoms - Global Severity Index of the Hopkins Symptom Checklist 90 - Revised</p>	
Notes	<p>Adherence: 58% Adverse Events: <i>Due to intervention:</i></p> <ul style="list-style-type: none"> • fracture of costal cartilage while rolling from supine to prone • pain <p><i>Unclear if due to intervention:</i></p> <ul style="list-style-type: none"> • hip fracture when lifting leg into measurement device (6 month assessment) • metatarsal fracture when weight dropped on foot <p><i>Not due to intervention:</i></p> <ul style="list-style-type: none"> • none <p>Between-group comparison as actual difference or change scores: change scores</p>	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomization was performed using a random number generator with equal allocation to the two arms." Comment: probably done.
Allocation concealment (selection bias)	Low risk	"All researchers except the biostatistician were masked to allocation status until a site was enrolled." Comment: probably done.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	"Participants were unaware of the content of the intervention that they did not receive initially." Comment: probably done.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	"All personnel involved with subject contacts, data collection, and intervention administration were masked to the intervention status of the sites and to the study hypotheses throughout the trial." Comment: probably done.
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Comment: Data were excluded from 63 of a total of 185 participants for the muscle strength outcome because of equipment problems; analyses for this outcome used a sample size of 122
Selective reporting (reporting bias)	Unclear risk	Comment: Gold 2004 reported on data from three primary outcomes. However, the data belongs to a much larger database that includes data from a number of studies, and a number of other outcomes were collected but were not reported in the paper. Data on walking endurance (six minute walk distance) was presented at the American Society for Bone and Mineral Research Annual Meeting in 2004, and data on trunk endurance (Timed Loaded Standing) was presented at the American Society for Bone and Mineral Research Annual Meeting in 2007. We have been in contact with the authors to get access to these and other

		unreported outcomes, but it was not available in time for this review. We were informed that the three primary outcomes were chosen a priori
Malmros 1998		
Methods	Randomized controlled trial	
Participants	# Randomized: 53 Lost to follow-up: 3 intervention, 2 control, no ITT analysis Age: 65 intervention, 68 control Country: Denmark Inclusion criteria: Women, at least 1 vertebral fracture, 55–75 years of age, back pain in the last 5 years (not from a new fracture) Exclusion criteria: no acute back pain from a new fracture, severe heart or pulmonary disease, psychiatric disease, medicine or alcohol abuse, rheumatoid disease	
Interventions	Experimental: <i>Frequency:</i> 2 times/ week <i>Intensity:</i> not specified <i>Time:</i> 1 hour <i>Type:</i> balance training, muscle strengthening, including flexors and extensors of trunk, stretching <i>Duration of intervention:</i> 10 weeks <i>Supervisor:</i> Physical therapist <i>Supervision/Setting:</i> Continual 1:1, centre-based <i>Progression of intensity:</i> Yes Control: Usual activities Duration of follow-up after baseline assessment: Weekly for 22 weeks	
Outcomes	Pain: 11-point scale Physical function: modified Oswestry questionnaire, Back extension strength in kg measured with strain gauge Quality of life: self-designed Balance: postural sway in x and y with eyes open and closed, tandem and with platform tilted, sum of four scores	
Notes	Adherence: 100% (except for those excluded because of adverse events) Adverse Events: 5 events of unknown etiology described as not due to intervention: 3 in intervention group and 2 in control group Between-group comparison as actual difference or change scores: both	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The participants were randomized consecutively and in masked fashion by the drawing of sealed envelopes containing the name of the group." Comment: probably not random sequence generation.
Allocation concealment (selection bias)	Low risk	Comment: see above. It is unclear who prepared or had access to the envelopes or performed randomization
Blinding of participants and personnel (performance bias) All outcomes	High risk	Comment: Participants and personnel delivering the intervention were not blind to group allocation due to nature of intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	"The tester was the same at all visits and was masked with regard to the group to which the participant was allocated." Comment: probably done
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: The original study would be assessed with a high risk of bias because an as-treated analysis was performed. However, our analysis presented in this review represents an intention-to-treat analysis, so the risk of bias in the findings we report is low
Selective reporting (reporting bias)	Unclear risk	Comment: The protocol is not publicly available for review, but the authors did provide the raw data, and there does not appear to be selective outcome reporting
Papaioannou 2003a		
Methods	Randomized controlled trial	

Participants	# Randomized: 74 (37 intervention, 37 control) Lost to follow-up: 14 at 6 months, 17 in total at 12 months Age: 71.6 intervention, 72.2 control Country: Canada Inclusion criteria: Women, at least one vertebral fracture (anterior height reduction of 15%), osteoporosis (lumbar spine T-score of -2.5) Exclusion criteria: resting heart rate > 100 bpm, uncontrolled hypertension, secondary causes of bone loss, another diagnosis for back pain, unable to stand independently for 3 minutes	
Interventions	Experimental: <i>Frequency:</i> 3 times/week <i>Intensity:</i> not specified <i>Time:</i> 60 minutes <i>Type:</i> stretching, strengthening with elastic band, aerobic activity <i>Duration of intervention:</i> 52 weeks <i>Supervisor:</i> Kinesiologist with MSc. <i>Supervision/Setting:</i> Intermittent home visits and phone calls, home exercise <i>Progression of intensity:</i> Yes Control: Usual activities Duration of follow-up after baseline assessment: 52 weeks	
Outcomes	Quality of life: Sickness Impact Profile, Osteoporosis Quality of Life Questionnaire Balance: postural sway Mobility: Timed Up and Go Bone density: Dual-energy x-ray absorptiometry at lumbar spine and femoral neck *Webber et al report changes in vertebral height in a sub-group from this study - not included in the review	
Notes	Adherence: 62% at 6 months, 46% at 12 months (those that were adherent to at least 3 times/week) Adverse Events: <i>Due to intervention:</i> <ul style="list-style-type: none"> • none reported <i>Unclear if due to intervention:</i> <ul style="list-style-type: none"> • fear of falling (n = 4) • pain or illness (n = 10) <i>Not due to intervention:</i> <ul style="list-style-type: none"> • 1 death Between-group comparison as actual difference or change scores: change scores	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Comment: The randomization process was not described.
Allocation concealment (selection bias)	Unclear risk	Comment: The allocation concealment was not described.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Comment: Participants and personnel delivering the intervention were not blind to group allocation due to nature of intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	"The research assistant who collected all data was blinded to intervention status." Comment: probably done.
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Comment: It is reported that an intention-to-treat analysis was performed, but it is not clear what was done with missing data

Selective reporting (reporting bias)	Unclear risk	Comment: Webber 2003 published data on vertebral height changes from a subset of participants. The protocol is not publicly available for review
Yang 2007		
Methods	Randomized controlled trial	
Participants	# Randomized: 30 (15 intervention, 15 control) Lost to follow-up: none Age: 66.4 (67.4 intervention, 65.5 control) Country: China Inclusion criteria: Women, more than one vertebral fracture (vertebrae compressed 20%), Chinese standard for primary osteoporosis diagnosis, willing to accept treatment Exclusion criteria: vertebral fractures due to trauma or malignancy, contraindications to exercise, cannot complete evaluation or treatment	
Interventions	Experimental: <i>Frequency:</i> twice per week <i>Intensity:</i> not specified <i>Time:</i> 10–30 minutes <i>Type:</i> strengthening and flexibility exercises <i>Duration of intervention:</i> 4 weeks <i>Supervisor:</i> Not reported <i>Supervision/Setting:</i> Unclear <i>Progression of intensity:</i> No Control: No description Duration of follow-up after baseline assessment: 4 weeks	
Outcomes	Pain on a 100 mm visual analogue scale Physical function: Time to sit up from a supine position Mobility: Timed Up and Go	
Notes	Adherence: 100% Adverse Events: None reported Between-group comparison as actual difference or change scores: actual difference	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Comment: There is not enough information provided in the methods to evaluate risk
Allocation concealment (selection bias)	Unclear risk	Comment: There is not enough information provided in the methods to evaluate risk
Blinding of participants and personnel (performance bias) All outcomes	High risk	Comment: Participants and personnel delivering the intervention were not blind to group allocation due to nature of intervention
Blinding of outcome assessment (detection bias) All outcomes	High risk	Comment: Outcome assessors were not blind to group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Comment: All participants returned for follow-up.
Selective reporting (reporting bias)	Unclear risk	Comment: There is not enough information provided in the methods to evaluate risk. The protocol is not publicly available

BMD: bone mineral density

QUALEFFO-41: Quality of Life Questionnaire of the European Foundation for Osteoporosis

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Bada 2009	A brief descriptive review of treatments for vertebral fracture
Bautmans 2010	Participants are a heterogeneous group of patients with and without vertebral fracture
Borgo 2010	Participants who had a vertebral fracture at baseline were excluded from study participation
Hongo 2007	Participants are a heterogeneous group of patients with and without vertebral fracture. Although the outcomes and authors are similar to that reported in Miyakoshi 2010, the interventions described in each trial are slightly different with respect to exercises chosen, follow-up time, and exercise frequency so it is likely that they are independent trials. Three attempts to contact the authors via email were made with no response
Lord 1996	Participants did not have vertebral fracture.
Malmros 1999	Same study as Malmros 1998 but in Danish.
Miyakoshi 2010	No information about control group activities to confirm that they did not receive an intervention. No information about randomization to confirm it was a randomized trial. No estimates of effect provided (P values only)
Rittchen 1991	Not a randomized controlled trial, but a descriptive paper about the management of patients with acute fractures
Schwinning 1992	Not a randomized controlled trial, but a descriptive paper about the management of patients with acute fractures
Shipp 2004	An abstract reporting data from Gold 2004. There is no formal between-group comparison in this abstract, but there is one sentence implying that a between-group difference in six-minute walk distance was not significant, which we have included in the review. The focus of the abstract was to investigate whether changes in six minute walk distance was related to changes in other measures of disability. We were in contact with the authors of Gold 2004 to get access to all unpublished outcomes from the larger data set, but were not able to obtain them in time for this version of the review
Shipp 2007	An abstract reporting previously published data from Gold 2004, as well as unpublished data on one outcome. We were in contact with the authors of Gold 2004 to get access to all unpublished outcomes from the larger data set, but were not able to obtain them in time for this version of the review. We did report the findings for the unpublished outcome in our review, but there was not enough information in the abstract to warrant its inclusion independently
Sinaki 1984	Not a randomized controlled trial.
Sinaki 1995	Not a randomized controlled trial.
Smith 1998	Not all participants had a vertebral fracture. Participants had several types of fracture e.g., vertebral, hip, radius, clavicle No information on randomization process provided in published report, or in methodology section of unpublished report that we obtained from one of the authors. Therefore it is not possible to confirm that it was truly a randomized trial
Stadhouders 2009	Includes patients with acute traumatic vertebral fracture only Includes patients who are less than 40 years of age. Comparison group is using a brace.
Webber 2003	A secondary analysis of data from a subgroup of patients in Papaioannou 2003a, using an outcome that was not specified in the protocol, namely vertebral height as a continuous measure

DATA AND ANALYSES

Comparison 1

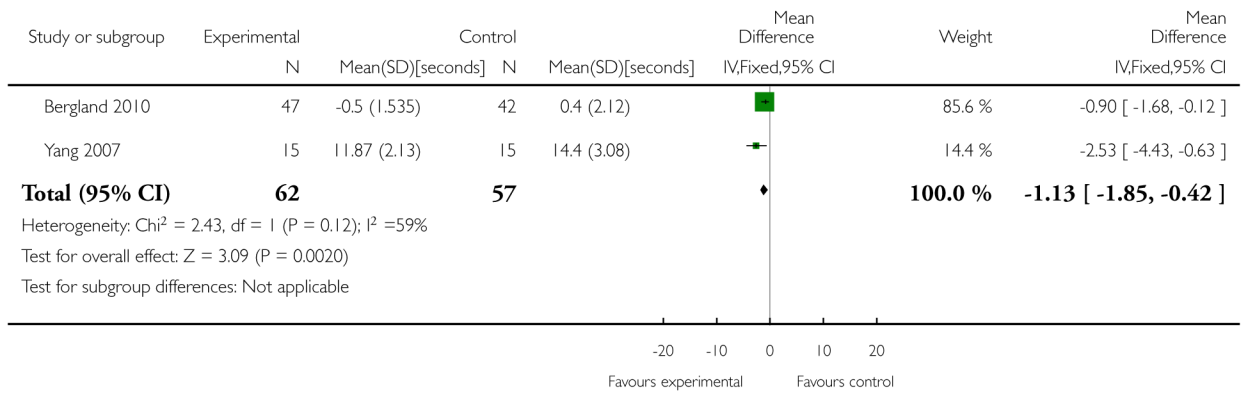
Exercise versus Control

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Timed Up and Go	2	119	Mean Difference (IV, Fixed, 95% CI)	-1.13 [-1.85, -0.42]

Review: Exercise for improving outcomes after osteoporotic vertebral fracture

Comparison: 1 Exercise versus Control

Outcome: 1 Timed Up and Go



Analysis 1.1.

Comparison 1 Exercise versus Control, Outcome 1 Timed Up and Go.

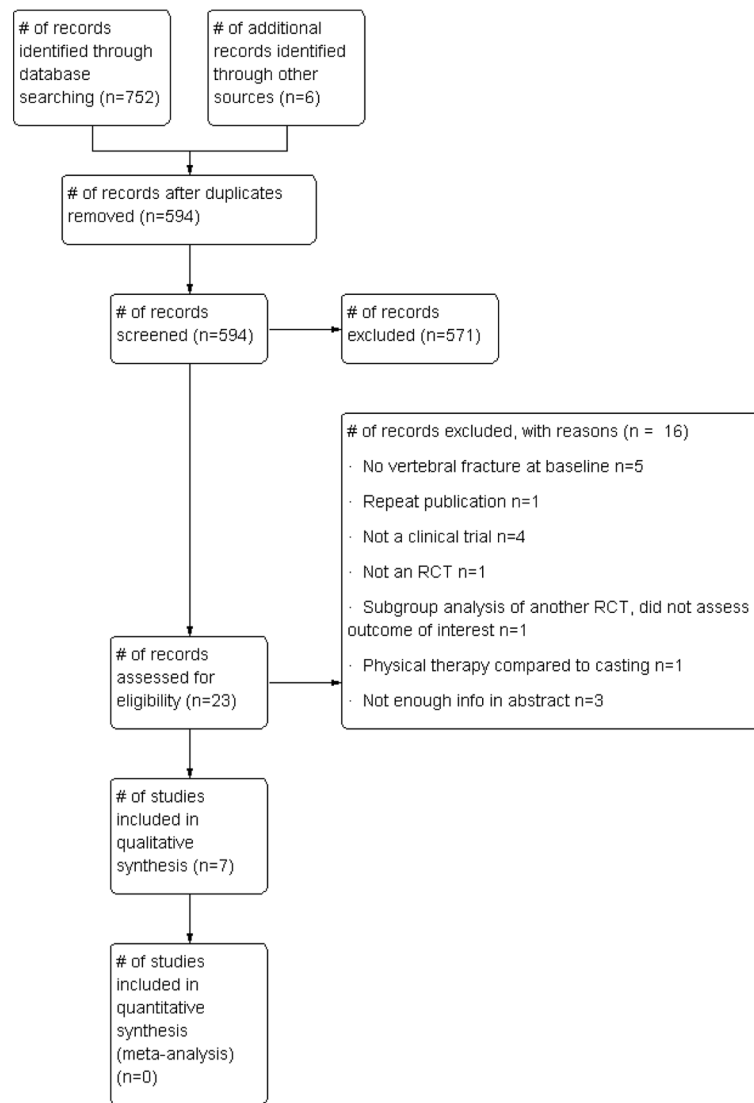


Figure 1.
PRISMA study flow diagram.

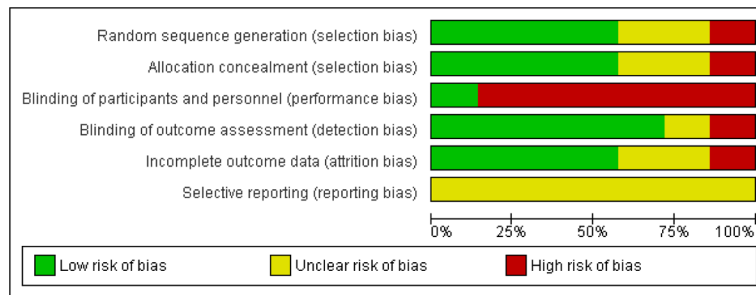


Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)
Bennell 2010	+	+	-	+	+	?
Bergland 2010	+	+	-	+	+	?
Bergstrom 2011	-	-	-	?	-	?
Gold 2004	+	+	+	+	?	?
Malmros 1998	+	+	-	+	+	?
Papaioannou 2003a	?	?	-	+	?	?
Yang 2007	?	?	-	-	+	?

Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

Table 1

Methodological quality assessment scheme (adapted from The Cochrane Collaboration's tool for assessing risk of bias)

Domain	Score	Domain Description	Comments
Was the allocation sequence adequately generated?	YES UNCLEAR NO	There is a random component in sequence generation Method of randomization not stated or unclear Quasi-randomized, nonrandom component in sequence generation	
Was allocation adequately concealed prior to or during randomization?	YES UNCLEAR NO	Participants/investigators could not foresee assignments Method of allocation concealment not stated or unclear Participants/investigators could possibly foresee assignments, quasirandomized	
Were outcome assessors blinded to treatment status?	YES UNCLEAR NO	Blinding of outcome assessment, or outcomes unlikely to be affected by lack of blinding Insufficient information to determine if blinding did or did not occur No blinding, incomplete blinding, chance blinding could be broken, AND lack of blinding is likely to introduce bias	
Were incomplete outcome data adequately addressed?	YES UNCLEAR NO	No missing data, or missing data are: balanced across groups, unlikely to affect outcome, imputed, ITT analysis Insufficient information about attrition/exclusions Missing data likely to affect outcome or be related to outcome, as-treated analysis, inappropriate imputation	
Are reports of the study free of selective outcome reporting?	YES UNCLEAR NO	Protocol is available and measurement methods for pre-specified outcomes defined and reported as defined, or key expected outcomes have been defined and reported Insufficient information to judge whether or not selective outcome reporting has occurred Incomplete or absent reporting, or key outcomes not reported that would be expected, measurement methods not specified	

ITT: intention-to-treat

Table 2

Adverse Events Reported in Exercise Trials in Individuals with Vertebral Fracture

Adverse Event	Number of incidences per study	Due to Intervention	Cause	Resulted in Study Withdrawal	Study
Death	1	No	Unknown	Yes	Papaioannou 2003a
Fracture of costal cartilage	1	Yes	Prone Exercise	Unknown	Gold 2004
Rib Fracture	1	Yes	Rolling from supine to prone	Unknown	Gold 2004
Hip Fracture	1	No	Study physical examination	Unknown	Gold 2004
Metatarsal Fracture	1	Yes	2lb weight fell on foot	Unknown	Gold 2004
Myeloma diagnosis	1	No		No	Bergstrom 2011
Pain	4	Unknown	Soft tissue origin	Unclear - resulted in missed classes	Gold 2004; Bergstrom 2011
Pain or illness	10	Unknown	Unknown	Yes	Papaioannou 2003a
Pain or injury	5	Unknown	Unknown	No	Bennell 2010
Irritation to tape	1	Yes	Reaction to tape material	No	Bennell 2010
Fear of falling or fall	4	Unknown	Unknown	Yes	Papaioannou 2003a
Undescribed adverse events that caused study withdrawal	5	Author indicated they were unrelated	Unknown	Yes	Malmros 1999

The adverse events here are reported in the results of each study, but not all studies mentioned adverse events. There was no clear indication in any of the studies that adverse events were systematically monitored.