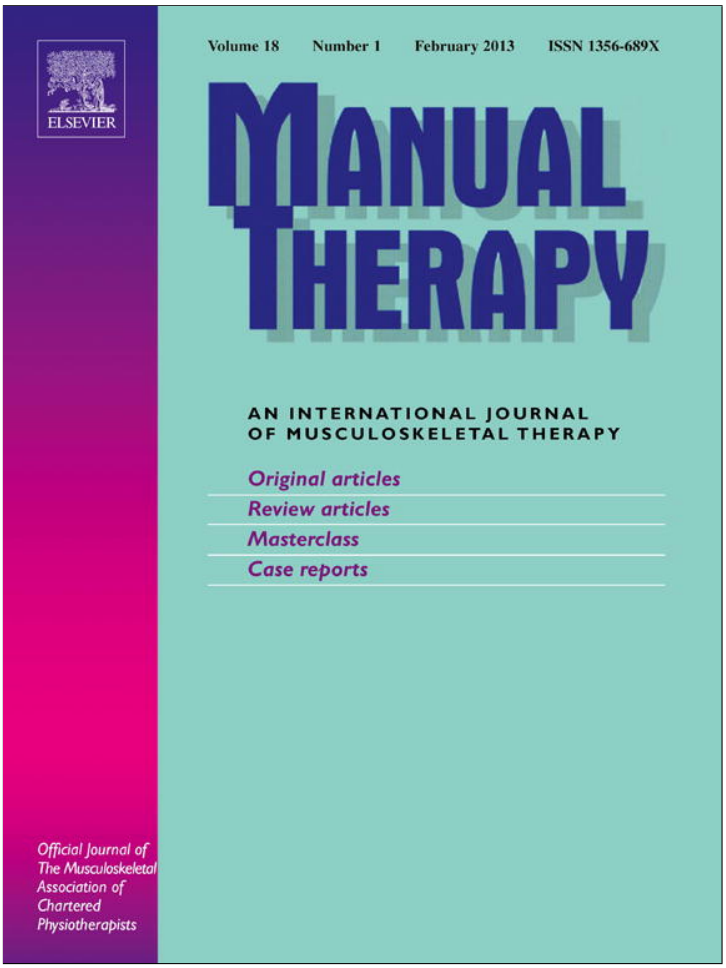
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Manual Therapy 18 (2013) 4–25

Contents lists available at SciVerse ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Systematic review

The effectiveness of physiotherapy functional restoration for post-acute low back pain: A systematic review

Matthew C. Richards [a,*], Jon J. Ford [a], Sarah L. Slater [a], Andrew J. Hahne [a], Luke D. Surkitt [a], Megan Davidson [b], Joan M. McMeeken [c]

[a] Low Back Research Team, Musculoskeletal Research Centre, Department of Physiotherapy, School of Allied Health, Faculty of Health Sciences, La Trobe University, 3086, Australia[1]
[b] Department of Physiotherapy, School of Allied Health, Faculty of Health Sciences, La Trobe University, 3086, Australia
[c] School of Physiotherapy, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, 3010, Australia

ARTICLE INFO

Article history:
Received 4 March 2012
Received in revised form
6 June 2012
Accepted 11 June 2012

Keywords:
Systematic review
Back pain
Physiotherapy
Functional restoration

ABSTRACT

Background: The effectiveness of multidisciplinary treatment for post-acute (>6 weeks) low back pain (LBP) has been established. Physiotherapists have sufficient training to conduct less intensive functional restoration. The effectiveness of physiotherapy functional restoration (PFR) has not been evaluated using current systematic review methodology.
Objectives: To determine the effects of PFR for post-acute LBP.
Data sources: Electronic databases searched include: MEDLINE, EMBASE, CINAHL, PsycINFO, PEDro and Cochrane CENTRAL.
Trial eligibility criteria: Randomised controlled trials of physiotherapy treatment for post-acute LBP combining exercise and cognitive-behavioural intervention compared with other intervention, no intervention or placebo.
Trial appraisal and synthesis methods: Two authors independently extracted data. Risk of bias was assessed using the PEDro scale and overall quality of the body of evidence was assessed using GRADE (Grading of Recommendations, Assessment, Development and Evaluation). Treatment effect sizes and 95% confidence intervals were calculated for pain, function and sick leave.
Results: Sixteen trials were included. Heterogeneity prevented meta-analysis for most comparisons. Meta-analyses showed moderate to high quality evidence of significant but small effects favouring PFR compared with advice for intermediate term function and intermediate and long term pain. There was however low to moderate quality evidence that PFR was no more effective than a range of other treatment types. Heterogeneous trials frequently contributed to very low quality evidence.
Conclusions: Moderate to high quality evidence was found of small effects favouring PFR compared with advice. Preliminary evidence suggested PFR is not different to other treatment types. Further high quality research is required replicating existing trial protocols.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

The high prevalence (Walker et al., 2004) and burden (Dagenais et al., 2008) of low back pain (LBP) is well established. The condition is typically characterised by recurrent episodes of pain (Stanton et al., 2009), with most sufferers experiencing persistent problems at 12 months (Hestbaek et al., 2003). Most of the societal costs, estimated to be at least $US100 billion annually (Katz, 2006), are due to post-acute LBP (Maetzel and Li, 2002; Dagenais et al., 2008) which can be defined as pain of at least six weeks duration (Hartigan et al., 1996).

Psychosocial distress negatively impacts the course of LBP (Hayden et al., 2009) and the comorbidity of such distress and LBP ranges from 28% to 36% (von Korff et al., 2005; Leijon and Mulder, 2009; Australian Institute of Health and Welfare, 2010). Functional restoration addresses the physical, psychological and social dimensions of LBP (Poiraudeau et al., 2007) via "a multimodal pain management program that employs a comprehensive cognitive-behavioural treatment orientation to help patients better cope with, and manage their pain…while undergoing the sports medicine physical approach to correct functional deficits" (Mayer et al., 1985). Multidisciplinary functional restoration has demonstrated

* Corresponding author. Tel.: +61 3 9479 5801; fax: +61 3 9479 5768.
E-mail address: matt.c.richards@gmail.com (M.C. Richards).
[1] The work is attributed to this department and institution.

1356-689X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.math.2012.06.005

Author's personal copy

moderate effect sizes for the outcomes of pain, function and work status in post-acute LBP (Chou and Huffman, 2007; Poiraudeau et al., 2007; van Geen et al., 2007; Norlund et al., 2009) and is recommended for this population in clinical guidelines (Koes et al., 2010). However, multidisciplinary programs are perceived to be more expensive and less accessible compared with those provided by a single discipline (Karjalainen et al., 2001; van Geen et al., 2007; Gatchel and Mayer, 2008).

Physiotherapists are trained in the assessment and management of post-acute LBP using exercise and cognitive-behavioural strategies (Bekkering et al., 2003; van der Windt et al., 2008). There has been no systematic review published using current best practice methodology (Furlan et al., 2009) specifically evaluating the effectiveness of functional restoration provided by physiotherapists. Existing reviews have included trials of both physiotherapy and multidisciplinary interventions without separate evaluation (George, 2008; Macedo et al., 2010; Schaafsma et al., 2010). Another review (Bunzli et al., 2011) only included trials evaluating operant conditioning (a specific type of cognitive-behavioural approach) as provided by physiotherapists and did not use current systematic review methodology including the presentation of effect sizes. Therefore, the aim of this systematic review was to evaluate the effectiveness of physiotherapy functional restoration (PFR) for post-acute LBP using the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) approach.

2. Methods

2.1. Data sources and searches

One reviewer (MR) performed a computerised search (Appendix A) for relevant trials. Searches were conducted to 31/12/2011 in MEDLINE (Ovid 1950-), EMBASE (Ovid 1980-), PsycINFO (Ovid 1806-), CINAHL (Ebsco 1982-), Cochrane Central Register of Controlled Trials (CENTRAL) and Physiotherapy Evidence Database (PEDro). Search terms for randomised controlled trials (RCT) and LBP were used as recommended by the Cochrane Back Review Group (2008), and experiential studies (Wong et al., 2006; Zhang et al., 2006). Cognitive-behavioural and exercise search terms were determined by the authors with guidance from a previous review (Schonstein et al., 2003). Bibliographies of related reviews and trials were searched for relevant studies. Grey literature was not searched.

2.2. Trial selection

Two reviewers (MR, SS) independently screened titles and abstracts. Full texts of all trials included by at least one reviewer were obtained and both reviewers (MR, SS) independently applied the exclusion criteria. A third reviewer (JF) was available to resolve any disagreements regarding eligibility and provided translation of German text. No other language translation was required. Full selection criteria are provided in Appendix D.

2.2.1. Participants

Trials with participants aged $\geq$18 years with LBP of >6 weeks duration were included. If a trial had a mixed sample, it was required to have $\geq$70% of participants experiencing LBP >6 weeks duration to be included. Trials were excluded where participants had diagnosed serious or non-mechanical pathologies.

2.2.2. Interventions

Only physiotherapy programs with both exercise and cognitive-behavioural components without invasive techniques or significant levels of passive intervention were included. Included trials either described a clear cognitive-behavioural approach (Henschke et al., 2010) or used the following terms: psychological, cognitive, behavioural, relaxation, operant, social, coping, respondent or counselling. Functional restoration requires at least moderate amounts of practitioner contact time (Poiraudeau et al., 2007), therefore trials were only included if they utilised at least 3 hours of total intervention time or a minimum of ten sessions.

2.2.3. Outcomes

Outcomes of interest included pain, function and sick leave (Deyo et al., 1998; Bombardier, 2000; Kent and Keating, 2008). Where a trial used multiple measures of pain, function or sick leave, the primary outcome measure was used (Macedo et al., 2010).

2.3. Data extraction and risk of bias assessment

Two reviewers (MR, SS) independently extracted and recorded data using a previously developed standardised computer spreadsheet (Hahne et al., 2010; Slater et al., 2012; Surkitt et al., 2012). Data extracted included trial setting, sample characteristics, interventions, comparisons, outcomes and adverse events. Missing data were either requested from the authors or calculated using the methods described in the Cochrane Handbook (Higgins et al., 2011).

Follow-up periods were categorised as short term (less than 3 months after randomisation), intermediate term (3 months up to 12 months), and long term (12 months or more) (van Tulder et al., 2003). Where a trial presented the same outcome more than once within a follow-up period, the earliest outcome was presented (Hayden et al., 2005), except for varying results in which case all outcomes were presented.

The reviewers independently assessed risk of bias using the PEDro Scale (Maher et al., 2003) (Table 1), shown to have sufficient validity (de Morton, 2009) and reliability (Maher et al., 2003). Trials that fulfilled $\geq$6 of 10 criteria were judged to have high methodological quality (Maher, 2000). Recommended criteria (Higgins et al., 2011) were used to evaluate clinical relevance including assessment of minimal clinically important difference (Table 1).

2.4. Data synthesis and analysis

Effect sizes were reported in line with suggested recommendations for systematic reviews (Higgins et al., 2011). Hedges adjusted-g standardised mean difference (SMD) (Hedges and Olkin, 1985) was used to calculate the treatment effect and 95% confidence interval (CI) for continuous outcomes. The SMD is the difference in mean outcome between groups divided by the pooled standard deviation (SD) of the outcome among participants (Higgins et al., 2011). Positive treatment effects for PFR were assigned positive SMD values, with 0.2, 0.5 and 0.8 representing small, moderate and large effect sizes respectively (Cohen, 1988). Relative risk (RR) and 95% CI were calculated for each dichotomous variable (Herbert, 2000) and standardised such that RR >1 indicated an increased risk of the event occurring in the PFR group relative to the comparison group. When unavailable, data were calculated from median values, mean change, graphical data, standard error (Hozo et al., 2005), baseline SD (Higgins et al., 2011) or from other trials within the review utilising the same outcome measure (Furlan et al., 2009).

Pooling of data in a meta-analysis using computer software Revman 5.1 (2011) was planned if $\geq$2 trials were evaluated as clinically homogenous (similar participant, intervention, outcome and comparison characteristics). When clinically homogenous trials were identified they were assessed for statistical heterogeneity (Higgins et al., 2011), which was considered likely if p-values of <0.1 were obtained on the χ^2 test or if the I^2 statistic was >25%

Author's personal copy

Table 1
Methodological quality criteria and clinical relevance criteria.

Item	Methodological quality criteria
1	Were the eligibility criteria specified?
2	Were participants randomly allocated to groups?
3	Was allocation concealed?
4	Were groups similar at baseline for the most important prognostic indicators?
5	Were all participants blinded?
6	Were all therapists who administered therapy blinded?
7	Were all assessors who measured at least one key outcome blinded?
8	Were measures of at least one key outcome obtained from >85% of the participants initially allocated to groups?
9	Did all participants (for whom outcome measures were available) receive the treatment or control condition as allocated or, where this was not the case, was data for a least one key outcome analysed by intention to treat?
10	Were the results of between-group statistical comparisons reported for at least one key outcome?
11	Did the study provide both point measures and measures of variability for at least one key outcome?
Item	**Clinical relevance criteria**
1	Are the patients described in detail so that you can decide whether they are comparable to those that you see in your practice?
2	Are the interventions and treatment settings described well enough so that you can provide the same for your patients?
3	Were all clinically relevant outcomes measured and reported?
4	Is the size of the effect clinically important? Clinical importance was measured by comparing the between group mean differences with published measures of minimal clinically important difference (Bombardier et al., 2001; Furlan et al., 2009), defined as two points on the Numerical Pain Rating Scale (NPRS) (Ostelo & de Vet, 2005), 1.5 on the Visual Analogue Scale (VAS) (Ostelo & de Vet, 2005), 10% on the Oswestry Disability Questionnaire (ODQ) (Ostelo & de Vet, 2005) and five points on the Roland Morris Disability Questionnaire (RMDQ) (Ostelo & de Vet, 2005).
5	Are the likely treatment benefits worth the potential harms?

Note: criteria 2–11 of methodological quality criteria contribute to the methodological quality score.

(Hahne et al., 2010). A fixed effects model was chosen based on the high threshold for statistical and clinical homogeneity that needed to be reached in order for trials to qualify for meta-analysis, hence random variation between these studies was assumed to be minimal. There was no planned sensitivity or subgroup analysis.

Quality of the body of evidence was determined using the GRADE approach which analyses the following domains: trial design limitations due to risk of bias (utilising the PEDro score), inconsistency of results, indirectness, imprecision of results and publication bias (Atkins et al., 2004; Furlan et al., 2009) (Appendix C).

3. Results

3.1. Trial selection

Fig. 1 outlines the reasons for trial exclusion. Authors (Anema et al., 2007; Pengel et al., 2007) provided additional data and clarification of standard error data (Heymans et al., 2006). Long term sick leave data in Hlobil (2005) was excluded due to data irregularity. Sixteen trials and 20 publications were included. Follow-up data were presented in secondary publications (Frost et al., 1998; Friedrich et al., 2005; Hlobil et al., 2005; Anema et al., 2007).

3.1.1. Trial characteristics

The characteristics of the 16 included trials are listed in Table 2. One trial (Pengel et al., 2007) investigated two different PFR interventions, the first described as exercise and the second as exercise and advice. Both satisfied the review's inclusion criteria for PFR and were therefore analysed separately.

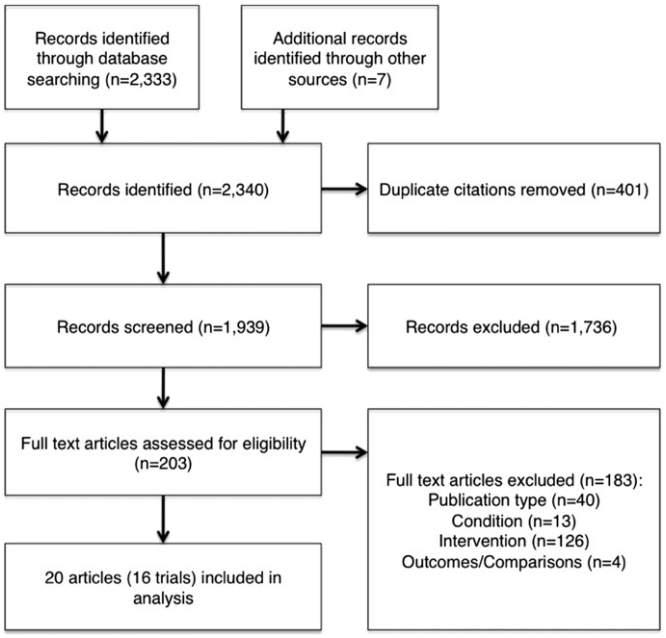


Fig. 1. Flow chart of article progression through the selection process.

Pengel (2007) reported mild adverse effects from 21 participants with no withdrawals, and Heymans et al. (2006) reported four diagnoses of herniated disc and three cases of increased LBP during the trial with no withdrawals. One trial (Staal et al., 2004) decided to not monitor mild adverse events to avoid focus on "disabilities" and there was no mention of adverse events in a follow-up publication (Hlobil et al., 2005). Five trials reported no adverse effects (UK BEAM Trial Team, 2004; Anema et al., 2007; Critchley et al., 2007; van der Roer et al., 2008; Paoloni et al., 2011). The remaining trials did not mention adverse events.

The mean differences, treatment effect sizes and associated 95% CI for the individual trials are grouped according to comparison treatment followed by outcome (Appendix B). Clinical heterogeneity prevented pooling of results apart from two pairs of trials: Staal et al. (2004) + Steenstra et al. (2006) and UK BEAM Trial Team (2004) + Johnson et al. (2007). Qualitative evaluation of the evidence was made for each comparison outcome based on the GRADE results. The reasons for downgrading quality of the body of evidence are presented in parenthesis for each separate outcome, apart from very low quality evidence where reasons for downgrading are provided across outcomes. Details of GRADE domain criteria are presented in Appendix C.

3.2. Risk of bias assessment

The results of the risk of bias assessment for each of the 20 publications are listed in Table 3. The publications had an average score of 6.9 out of 10 (range 5–8) with three publications (Lindström et al., 1992; Kankaanpaa et al., 1999; Woods and Asmundson, 2008) having low methodological quality. Due to well known feasibility issues in physiotherapy research (Herbert et al., 2005) no publications blinded treating therapist or participants.

3.3. PFR versus placebo

Two trials (Kankaanpaa et al., 1999; Pengel et al., 2007) compared PFR with placebo, one of which (Pengel et al., 2007) compared two variations of PFR treatment (one with further advice) with placebo (Appendix B: GRADE Table 1). There was only

Table 2
Characteristics of included trials.

Trial	Participant characteristics	Intervention(s)	Comparison(s)	Outcome measures	Follow-up
Carr et al. (2005) Risk of bias = 8 Clinical relevance = 2	• Patients with LBP >6 weeks duration were referred by their GP or hospital consultants • Average age = 42 • *N* = 237	• Back to fitness program in hospital or community health centre settings included low impact aerobics, strengthening and stretching exercises and relaxation, where a cognitive-behavioural approach guided messages with a reference to a published treatment manual (Klaber Moffett and Frost, 2000) – 8 h long classes conducted over a 4 week period	• Individual physiotherapy treatment at the discretion of the physiotherapist	• Function (RMDQ)	• 3 months • 12 months
Critchley et al. (2007) Risk of bias = 7 Clinical relevance = 1	• Participants with LBP >12 weeks duration were referred by specialist or primary care practitioner. • Average age = 44 • *N* = 212	• Pain management program in the hospital setting included structured back pain education, group general strengthening, stretching, and exercises progressed according to pacing principles. A cognitive-behavioural approach was used to reduce fear of movement and re-injury and encourage self-management. Education promoted hurt not equating to harm, a graded return to usual activities with goal-setting and positive coping strategies – maximum of 8 sessions of 90 min	A. Individual physiotherapy: joint mobilisation/manipulation and massage, exercise prescription and back care advice – up to 12 sessions of 30 min (comparison = other therapy) B. Spinal stabilisation: transversus abdominis and lumbar multifidus muscle training followed by group exercise – maximum of 8 sessions of 90 min (comparison = other exercise therapy)	• Pain (VAS 100-point) • Function (RMDQ) • Sick Leave (days off in last 6 months)	• 6 months • 12 months • 18 months
Friedrich et al. (1998) Risk of bias = 7 Clinical relevance = 4 Friedrich et al. (2005) Risk of bias = 6 Clinical relevance = 4	• Patients referred to hospital outpatient department with 4 months LBP, or at least 3 episodes in past 6 months, and current episode >2 months • Average age = 44 • *N* = 93	• Exercise which is sub-maximal, gradually increased, personalised and progressive, with a focus on stretching, strengthening, endurance, co-ordination, aiming to restore normal function, promotion of a home program to overcome fears about pain (anticipated or actual) that might contribute to avoidance of physical activity, encouragement to be physically active, postural education and ongoing maintenance for prevention of recurrent pain (Lindgren et al., 1993). "Motivation" including extensive counselling and information regarding regular exercise with explanation of dependence on behaviour to enhance locus of control, tailoring of the program to the patient, reinforcement techniques, positive feedback, rewards for exercise compliance, oral agreements and written treatment contracts individually negotiated (including rewards or punishments for gradual exercise increase compliance), and completion of an exercise diary – 10 sessions of 25 min, 2–3 per week + further 5 sessions of motivation	• Exercise sessions (10) without motivation	• Pain (VAS 100-point) • Function (LBOS) • Sick Leave (working – y/n)	• 3.5 weeks • 4 months • 12 months • 5 years

(*continued on next page*)

Author's personal copy

Author's personal copy

Table 2 (*continued*)

Trial	Participant characteristics	Intervention(s)	Comparison(s)	Outcome measures	Follow-up
Frost et al. (1995) Risk of bias = 8 Clinical relevance = 2 Frost et al. (1998) Risk of bias = 8 Clinical relevance = 2	• Patients with LBP >6 months in duration were referred by orthopaedic consultants • Average age = 36 • *N* = 81	• Fitness program involved psychological principles, a warm-up, stretching, progressive circuit of 15 exercises, further stretching and light aerobic exercise, with a fitness rather than disabled approach and hurt versus harm message with the aim to better their own previous personal achievements in the gym (in addition to the back school control intervention) – 8 h long sessions over 4 weeks	• Back school and home program (two sessions) included discussion of the main problem, functional anatomy, applied body mechanics, advice regarding functional activity and exercise, relaxation, prevention, ergonomics and an injury prevention video	• Pain (NPRS 101-point) • Function (ODI)	• Post-treatment • 6 months • 2 years
Heymans et al. (2006) Risk of bias = 8 Clinical relevance = 4	• Workers within occupational health services with non-specific LBP, sick-listed for 3–6 weeks • Average age = 40 • *N* = 299	• High intensity back school with cognitive-behavioural therapy, a time contingent increase in activity, individualised work-specific exercise, gradually increased work-simulating and strength exercise, home-based exercise (Vlaeyen et al., 1995) (in addition to the usual care control) – 16 h long sessions over 8 weeks	A. Usual care + low intensity back school, advice to exercise despite pain and a general exercise program – four 2-h sessions (comparison = other cognitive-behavioural therapy) B. Usual care according to Dutch OP guidelines (comparison = advice)	• Pain (VAS) • Function (RMDQ) • Sick Leave (working – y/n)	• 3 months • 6 months
Johnson et al. (2007) Risk of bias = 7 Clinical relevance = 2	• LBP patients, either identified by their GP or identified on GP records, completed a questionnaire at 3 months, with those reaching persistent pain and function thresholds assessed • Average age = 48 • *N* = 234	• Active intervention program with 4–10 patients per group, promoted resumption of activity, enabled patient control of pain through physical exercise and psychological self-help techniques, with pacing, problem solving, regulation of activity, challenging distorted cognitions about activity and harm, assisted return to normal activity/work, prepared for independent management of future episodes, individualised behavioural and activity homework (with follow-up and problem solving) and encouraged self-management with a cognitive-behavioural approach (Main and Spanswick, 2000) (in addition to advice pack control) – eight 2-h long sessions over 6 weeks	• Advice pack: booklet and audio cassette contained advice on self-management suitable for patients with persistent LBP; leaflets on pain and activity, pacing, goal setting, stress, posture and body mechanics, sleep hygiene, beds and sleeping, flare-up plans and when to see the GP	• Pain (VAS 100-point) • Function (RMDQ)	• 3 months • 9 months • 15 months
Kankaanpaa et al. (1999) Risk of bias = 5 Clinical relevance = 2	• Chronic non-specific LBP patients recruited through an occupational health centre, with moderate functional disability and occasional work absences • Average age = 39 • *N* = 54	• Active rehabilitation program, 4–5 patients per group, with physical exercise with specific equipment, stretching, relaxation, ergonomic advice, gradually increasing loads from baseline, behavioural support including "good prognosis" for LBP and encouragement to continue home exercise to one-year follow-up – 90 min sessions twice/week over 12 weeks	• Passive treatment (non-exercise control – dosages aimed for low efficacy) including thermal therapy and massage individually once/week over 1 month (4 sessions)	• Pain over past 6 weeks (VAS 100-point) • Function (PDI)	• 12 weeks • 6 months • 12 months

Klaber Moffett (1999) Risk of bias = 7 Clinical relevance = 1	• Mechanical LBP 4 weeks to 6 months in duration, referred by GP, medically fit for exercise • Average age = 42 • *N* = 187	• Exercise with a cognitive-behavioural approach, 10 patients per class, using simple equipment, included stretching, low impact aerobic exercise, strengthening, encouragement of normal movement, discouragement from viewing themselves as invalids, positive feedback, self-reliance promotion – 8 sessions (each 60 min) over 4 weeks	• Usual care by patient's doctor (no attempt to regulate treatment)	• Pain (ABPS) • Function (RMDQ)	• 6 weeks • 6 months • 12 months
Lindström et al. (1992) Risk of bias = 5 Clinical relevance = 4	• Industrial workers with new episode LBP, were included when on sick leave for over 6 weeks (no sick-leave due to LBP over the previous 12 weeks), and randomised after 8 weeks sick leave • Average age = 41 • *N* = 98	• Graded activity program to restore occupational function to facilitate return to work involved a workplace visit, back school education and an individual, sub-maximal, gradually increased exercise program, with an operant conditioning behavioural approach (work specific), including education that it is safe to move while regaining function – return to work was the program end-point	• Physician care involving traditional care, general rest, analgesics and unspecific physical treatment modalities	• Sick Leave (working – y/n @ 6 & 12 weeks) • Sick Leave (average total @12 months)	• 6 weeks • 12 weeks • 12 months
Paoloni et al. (2011) Risk of bias = 8 Clinical relevance = 1	• Chronic LBP patients aged between 30 and 80, with symptoms >12 weeks in duration, recruited through an outpatient facility • Average age = 63 • *N* = 26	• Therapeutic exercise program including relaxation, stretching, active exercise for the abdominal, psoas, ischiotibial, pelvic, lumbar and thoracic extensor muscles. Exercise intensity was adapted to participant ability. Groups of up to 5 participants were guided by a senior physiotherapist with over 10 years experience for 30 min sessions, 3 times per week for 4 weeks	• "Kinesio taping" involving application of 3 (20 × 5 cm) strips of tape between T12 and L5 spinal levels, replaced every 3 days for 4 weeks	• Pain (VAS 10-point) • Function (RMDQ)	• 4 weeks
Pengel et al. (2007) Risk of bias = 8 Clinical relevance = 2	• Persons with non-specific LBP lasting for between 6 and 12 weeks (sub-acute) were recruited by direct referral by a health care professional, through a hospital PT wait list or newspaper advertising • Average age = 50 • *N* = 259	A. Exercise (Lindström et al., 1992): 12 sessions of an individualised, progressive, sub-maximal, and functional program with aerobic exercise, stretching, functional activities, and trunk and limb strengthening, incorporated cognitive-behavioural therapy, progressive goal setting, self-monitoring, self-reinforcement and ongoing self-management and 3 sessions of sham advice (patients given the opportunity to talk about their LBP with therapist empathy) for 6 weeks B. Exercise (as above) and 3 sessions of "advice" (Indahl et al., 1995): encouraged a graded return to normal activities, described the benign nature of LBP, addressed unhelpful beliefs (not to be overly careful or avoid light activity) for 6 weeks	A. Advice (3 sessions) and placebo exercise (sham electrotherapy for 12 sessions) for 6 weeks (comparison = advice) B. Placebo advice (empathic discussion for 3 sessions) and placebo exercise (12 sessions) for 6 weeks (comparison = placebo)	• Pain over past week (VAS 10-point) • Function (PSFS)	• 6 weeks • 3 months • 12 months

(*continued on next page*)

Author's personal copy

Author's personal copy

Table 2 (*continued*)

Trial	Participant characteristics	Intervention(s)	Comparison(s)	Outcome measures	Follow-up
Staal et al. (2004) Risk of bias = 8 Clinical relevance = 2 Hlobil et al. (2005) Risk of bias = 8 Clinical relevance = 2	• Airline workers listed as full or partially absent from work with non-specific LBP of at least 4 weeks duration • Average age = 38 • *N* = 134	• Graded activity intervention with cognitive-behavioural principles included reassurance that nothing is seriously wrong in patients' backs, a tailored program of aerobic exercise, strengthening, work-specific exercise from an established baseline level, goal-orientated exercise with feedback, praise for quota achievement, and education that it is safe to move despite pain (in addition to usual care control) – twice weekly hour long sessions for 3 months or until a complete return to work is achieved	• Usual care according to Dutch OP guidelines (active approach)	• Pain (NPRS 11-point) • Function (RMDQ)	• 3 months • 6 months • 12 months
Steenstra et al. (2006) Risk of bias = 7 Clinical relevance = 4 Anema et al. (2007) Risk of bias = 7 Clinical relevance = 4	• Workers on LBP-related sick-leave of more than 8 weeks • Average age = 42 • *N* = 112	• Graded activity (Staal et al., 2004) aimed to restore occupational function through an individual, sub-maximal, gradually increased exercise program, with an operant conditioning approach, tailored to the patient – 60 min sessions, twice per week (26 maximum) or until a lasting return to work is achieved	• Usual care according to Dutch OP guidelines (Staal et al., 2003) (active approach)	• Pain (VAS 10-point) • Function (RMDQ) • Sick Leave (days off work over the first 6 months)	• 12 weeks • 6 months • 12 months
UK BEAM Trial Team (2004) Risk of bias = 6 Clinical relevance = 2	• LBP patients with duration of symptoms every day for 4 weeks prior to randomisation, or LBP for 21 of the 28 days prior to randomisation and for 21 of the 28 days prior to that, requiring at least a score of 4 on the RMDQ, identified by research nurses or directly from GPs or their staff and from computer records • Average age = 43 • *N* = 1001	• Exercise program (Frost et al., 1995; Klaber Moffett et al., 1999) with group classes (up to 10 patients) incorporating a cognitive-behavioural approach – up to 8 h sessions over 6–8 weeks and one refresher class at 12 weeks	A. Inter-discipline agreed manipulation, up to 8 sessions (20 min) over 12 weeks (comparison = other therapy) B. Best care in general practice, advice to continue normal activities and avoid rest, provision of a self-management publication (Roland et al., 1996) (comparison = advice)	• Pain (Von Korff) • Function (RMDQ)	• 3 months • 12 months
van der Roer et al. (2008) Risk of bias = 7 Clinical relevance = 2	• New episode LBP patients with duration >12 weeks and health insurance with a particular insurance company allowing funding of treatment, recruited from PT practices • Average age = 42 • *N* = 114	• Intensive group training (van der Roer et al., 2004) combining exercise, back school and behavioural principles, assessment, education, goal setting, signing a treatment contract and evaluating treatment goals – 10 individual sessions; and training according to operant conditioning principles based on a baseline level of functional capacity – 20 group sessions	• Dutch PT guidelines (Bekkering et al., 2003) to provide adequate information, active approach, with behavioural principles – number of sessions at the discretion of PT (average = 13)	• Pain (NPRS 11-point) • Function (RMDQ)	• 6 weeks • 13 weeks • 6 months • 12 months
Woods and Asmundson (2008) Risk of bias = 5 Clinical relevance = 0	• Chronic LBP, recruited via newspaper, email, hospital posters • Average age = 46 • *N* = 83	• Graded activity involving individualised quota-based exercise, based on operant conditioning principles with positive reinforcement – eight 45 min sessions provided twice/week for 4 weeks	A. Psychologist-run program of graded exposure to activities in fear hierarchy with cognitive-behavioural education (comparison = other cognitive-behavioural therapy) B. Wait list (comparison = minimal intervention)	• Pain (MPQ-SF) • Function (PDI)	• 4 weeks • 8 weeks

Author's personal copy

Table 3
Trial methodological quality and clinical relevance.

	Methodological quality												Clinical relevance					
	1	2	3	4	5	6	7	8	9	10	11	Score	1	2	3	4	5	Total
Anema et al. (2007)	✓	✓	✓	✓	–	–	–	✓	✓	✓	✓	7	✓	✓	–	✓	✓	4
Carr et al. (2005)	✓	✓	✓	✓	–	–	✓	✓	✓	✓	✓	8	✓	✓	–	–	–	2
Critchley et al. (2007)	✓	✓	✓	✓	–	–	✓	–	✓	✓	✓	7	–	✓	–	–	–	1
Friedrich et al. (1998)	✓	✓	–	✓	–	–	✓	✓	✓	✓	✓	7	✓	✓	–	✓	✓	4
Friedrich et al. (2005)	✓	✓	–	✓	–	–	–	✓	✓	✓	✓	6	✓	✓	–	✓	✓	4
Frost et al. (1995)	✓	✓	✓	✓	–	–	✓	✓	✓	✓	✓	8	✓	✓	–	–	–	2
Frost et al. (1998)	✓	✓	✓	✓	–	–	✓	✓	✓	✓	✓	8	✓	✓	–	–	–	2
Heymans et al. (2006)	✓	✓	✓	✓	–	–	✓	✓	✓	✓	✓	8	✓	✓	–	✓	✓	4
Hlobil et al. (2005)	✓	✓	✓	✓	–	–	✓	✓	✓	✓	✓	8	✓	✓	–	–	–	2
Johnson et al. (2007)	✓	✓	✓	✓	–	–	–	✓	✓	✓	✓	7	✓	✓	–	–	–	2
Kankaanpaa et al. (1999)	✓	✓	✓	✓	–	–	–	–	–	✓	✓	5	–	–	–	✓	✓	2
Klaber Moffett et al. (1999)	✓	✓	✓	✓	–	–	–	✓	✓	✓	✓	7	–	✓	–	–	–	1
Lindström et al. (1992)	✓	✓	–	✓	–	–	–	✓	–	✓	✓	5	✓	✓	–	✓	✓	4
Paoloni et al. (2011)	–	✓	✓	✓	–	–	✓	✓	✓	✓	✓	8	✓	–	–	–	–	1
Pengel et al. (2007)	✓	✓	✓	✓	–	–	✓	✓	✓	✓	✓	8	✓	✓	–	–	–	2
Staal et al. (2004)	✓	✓	✓	✓	–	–	✓	✓	✓	✓	✓	8	✓	✓	–	–	–	2
Steenstra et al. (2006)	✓	✓	✓	✓	–	–	–	✓	✓	✓	✓	7	✓	✓	–	✓	✓	4
UK BEAM Trial Team (2004)	✓	✓	✓	✓	–	–	–	–	✓	✓	✓	6	✓	✓	–	–	–	2
van der Roer et al. (2008)	✓	✓	✓	✓	–	–	–	✓	✓	✓	✓	7	✓	✓	–	–	–	2
Woods and Asmundson (2008)	✓	✓	–	✓	–	–	–	–	✓	✓	✓	5	–	–	–	–	–	0
Total	19	20	16	20	0	0	10	16	18	20	20		16	17	0	7	7	

Note: criteria 2–11 of methodological quality criteria contribute to the methodological quality score.

very low evidence (limitations, inconsistency and indirectness) for all outcomes.

3.4. PFR versus Dutch occupational physician guideline advice for sub-acute sick-listed workers

A pair of clinically homogenous trials (Staal et al., 2004; Steenstra et al., 2006) compared PFR with advice for sub-acute sick-listed workers (Appendix B: GRADE Table 2A). Dutch occupational physician guideline advice provided sick leave guidance and explained good prognosis and activity as not harmful. Both trials specifically recruited sick-listed workers affecting generalisability of findings and therefore the GRADE domain for indirectness was downgraded for all outcomes. There was moderate quality evidence (indirectness) from the meta-analysis showing a small effect on intermediate term function favouring PFR.

In this pair of trials, meta-analysis was not possible on pain and long term functional outcomes due to statistical heterogeneity. Qualitative analysis revealed low quality evidence (inconsistency and indirectness) and conflicting results for intermediate and long term pain. There was also low quality evidence (inconsistency and indirectness) showing no effect on long term function.

Short term follow up and sick leave were only measured in one of the trials (Steenstra et al., 2006). There was low quality evidence (indirectness and imprecision) for small to moderate effects favouring PFR for short term pain and function. There was also low quality evidence (indirectness and imprecision) of a small effect on long term sick leave favouring advice.

3.5. PFR versus advice

A pair of clinically homogenous trials (UK BEAM Trial Team, 2004; Johnson et al., 2007) compared PFR in a post-acute general practice population with non-discipline specific advice regarding self-management strategies without a focus on sick leave guidance as above (Appendix B: GRADE Table 2B). These trials were clinically heterogeneous (population, intervention and comparison, detailed in Table 2) with those in the meta-analysis above and were therefore analysed separately. There was high quality evidence from the meta-analyses of small effects favouring PFR for intermediate and long term pain and intermediate term function, although there was no effect on long term function.

3.6. PFR (other) versus advice (other)

Two trials (Heymans et al., 2006; Pengel et al., 2007) compared PFR to advice using treatment protocols significantly different to those in the two pairs of trials described above (Appendix B: GRADE Table 2C). One trial (Pengel et al., 2007) compared two variations of PFR treatment (one with further advice). Clinical heterogeneity (population and intervention) prevented meta-analysis being conducted on these two trials, or pooling of data from either trial with the four other trials above that utilised advice as a comparison group. Qualitative evaluation showed low quality evidence (indirectness and imprecision) from one trial (Pengel et al., 2007) showing no difference for short term pain, and short and long term function. There was also low quality evidence (indirectness and imprecision) from one trial (Heymans et al., 2006) showing no difference for intermediate term sick leave. There was only very low evidence (inconsistency, indirectness and imprecision) for the remaining outcomes.

3.7. PFR versus other cognitive-behavioural therapy

Four trials (Frost et al., 1995; Heymans et al., 2006; van der Roer et al., 2008; Woods and Asmundson, 2008) compared PFR with other cognitive-behavioural therapy (Appendix B: GRADE Table 3). There was moderate quality evidence (imprecision) from one trial (van der Roer et al., 2008) showing no difference for long term pain. There was low quality evidence (indirectness) from two trials (Heymans et al., 2006; van der Roer et al., 2008) showing no difference for intermediate term pain. There was low quality evidence (indirectness and imprecision) from a single trial (Heymans et al., 2006) showing that other cognitive-behavioural therapy was more effective than PFR for intermediate term sick leave. There was only very low evidence (limitations, inconsistency, indirectness and imprecision) for the remaining outcomes.

3.8. PFR versus other exercise therapy

Two trials compared PFR with other exercise therapy (Friedrich et al., 1998; Critchley et al., 2007) (Appendix B: GRADE Table 4).

Author's personal copy

There was moderate quality evidence (imprecision) from a single trial (Friedrich et al., 1998) showing no difference for short term pain, function and sick leave. There was moderate quality evidence (indirectness) from both trials showing no difference for intermediate term pain, function and sick leave. There was low quality evidence (inconsistency and indirectness) from both trials showing varying results for long term pain, function and sick leave.

3.9. PFR versus other therapy

Four trials (UK BEAM Trial Team, 2004; Carr et al., 2005; Critchley et al., 2007; Paoloni et al., 2011) compared PFR with other therapy (Appendix B: GRADE Table 5). There was moderate quality evidence (indirectness) from two trials (UK BEAM Trial Team, 2004; Critchley et al., 2007) showing no difference for intermediate and long term pain. There was moderate quality evidence (indirectness) from three trials showing no difference for long term function. There was moderate quality evidence (imprecision) from one trial (Critchley et al., 2007) showing no difference for intermediate and long term sick leave. There was low quality evidence (indirectness and imprecision) from one trial (Paoloni et al., 2011) showing no difference for short term pain and function, and low quality evidence (inconsistency and indirectness) from three trials (UK BEAM Trial Team, 2004; Carr et al., 2005; Critchley et al., 2007) showing varying results for intermediate term function.

3.10. PFR versus minimal intervention

There is significant literature supporting the effectiveness of advice compared to other minimal interventions (Engers et al., 2008). Therefore, for the purposes of this review, minimal intervention was defined as usual care or waiting list (Macedo et al., 2010) and reported separately from advice. Three trials (Lindström et al., 1992; Klaber Moffett et al., 1999; Woods and Asmundson, 2008) evaluated PFR compared to minimal intervention (Appendix B: GRADE Table 6). There was moderate quality evidence (imprecision) from a single trial (Klaber Moffett et al., 1999) showing no difference for intermediate and long term pain, and intermediate and long term function. There was low quality evidence (limitations and indirectness) from two trials (Klaber Moffett et al., 1999; Woods and Asmundson, 2008) showing no difference for short term pain and function. There was only very low evidence (limitations, indirectness and imprecision) for the remaining outcomes.

4. Discussion

Physiotherapy functional restoration has the potential to be a valuable treatment for post-acute LBP. Due to the clinical and statistical heterogeneity between the various included trials, a meta-analysis was not appropriate in all but four trials. When conducted, meta-analyses showed effect sizes of questionable clinical importance.

Qualitative evaluation of comparisons using the GRADE evaluation revealed moderate evidence that PFR was no more effective than other types of treatment (16 out of 56) and otherwise either very low quality or varying evidence (19 out of 56). Overall the results of this review suggest that PFR when applied to people with post-acute LBP may be more effective in improving pain and function than evidence-based advice and no more effective than other types of treatment. None of the trials in this review were designed to assess equivalence according to the revised CONSORT guidelines (Piaggio et al., 2006). Of the trials indicating no difference, methodology did not satisfy the requirements of an "equivalence trial" in accordance with current guidelines (Piaggio et al., 2006). Consequently, it was not possible to conclusively prove there was genuinely "no difference" or to be able to establish "non-inferiority". There were however trials with significant population samples, the largest of which (UK BEAM Trial Team, 2004) may have been sufficient to demonstrate equivalence between interventions had the trial been designed with this goal in mind.

This is the first systematic review to specifically evaluate the effectiveness of PFR independently of trials on multidisciplinary functional restoration. This distinction is important for a number of reasons. Firstly, physiotherapists are taught skills that give them the potential to apply cognitive-behavioural principles (Bekkering et al., 2003) necessary for PFR, and additional training for physiotherapists in this area is common (Klaber Moffett and Frost, 2000; Staal et al., 2004; Hay et al., 2005; Johnson et al., 2007). Secondly, it has been argued that the high intensity of a multidisciplinary approach is a necessary requirement for effective functional restoration (Guzman et al., 2001; Poiraudeau et al., 2007; Gatchel and Mayer, 2008), however all 16 trials in this review utilised an intensity of program of less than 100 hours in total or 30 hours per week, which are the commonly recommended thresholds for an "intensive" program (Guzman et al., 2001; van Geen et al., 2007). Thirdly, physiotherapy services are more easily accessible and have the potential to be more cost effective than multidisciplinary programs (Karjalainen et al., 2001; van Geen et al., 2007; Gatchel and Mayer, 2008). The literature suggests positive effects for both PFR and multi-disciplinary programs. Further research could compare these two treatment approaches utilising a non-inferiority head to head study incorporating a secondary cost-effectiveness analysis.

A recent systematic review by Bunzli et al. (2011) reported statistically significant differences in favour of PFR using an operant conditioning approach compared to placebo and other cognitive-behavioural interventions. However, operant conditioning is only a subgroup of cognitive-behavioural interventions commonly used by physiotherapists (Ostelo et al., 2005). The review by Bunzli et al. did not include Friedrich et al. (1998), Kankaanpaa et al. (1999), Johnson et al. (2007) and Paoloni et al. (2011), and otherwise included three trials of primarily acute LBP not included in this review, two of which attributed a lack of intervention effect to recruiting an acute population (Hay et al., 2005; George et al., 2008). In addition, Bunzli et al. did not present effect sizes (Higgins et al., 2011) which enable evaluation of evidence consistency (Furlan et al., 2009).

There is general agreement that multidisciplinary FR is effective in reducing pain and improving sick leave (Guzman et al., 2001; Poiraudeau et al., 2007; Norlund et al., 2009). Other reviews that have included, but not exclusively targeted, trials of PFR show similar results to this review (Henschke et al., 2010; Macedo et al., 2010; Schaafsma et al., 2010).

The modest results on the effectiveness of PFR demonstrated in both this review and other recent reviews may be due to between trial heterogeneity. In the 16 included trials there were nine different types of PFR, 11 different comparison treatments, five measures of pain, five measures of function, and follow up at 14 different time points. In several trials there were inadequate descriptions of the treatment protocol (Kankaanpaa et al., 1999; Woods and Asmundson, 2008; Paoloni et al., 2011) or the sample recruited (Kankaanpaa et al., 1999; Klaber Moffett et al., 1999; Critchley et al., 2007; Woods and Asmundson, 2008). The principles and component parts of FR are complex requiring robust methods to ensure treatment integrity (van der Windt et al., 2008). Only six out of 16 trials (Friedrich et al., 1998; Staal et al., 2004; Steenstra et al., 2006; Johnson et al., 2007; Pengel et al., 2007; van der Roer et al., 2008) incorporated such methods to train and monitor the treatment quality of the treating physiotherapists. Even trials determined in this review as being clinically homogenous had potential for heterogeneity that may have affected the results. For example, the trials of Staal et al. (2004) and Steenstra et al. (2006) were based on the same treatment protocol, participant population and compensation environment. However, participants in

Author's personal copy

the Steenstra et al. trial had not returned to work after specific workplace intervention prior to the PFR, a fact which the authors acknowledge may have impacted on the results (Steenstra et al., 2006).

The potential impact of heterogeneity between trials on synthesising evidence has been well described (Ford et al., 2007; Fritz et al., 2007). Future trials evaluating the effectiveness of PFR should address this issue in a number of areas. Replication of well described high quality trials that demonstrate substantial effect sizes is an important research principle that has the potential of providing strong evidence when subsequent meta-analysis is performed (Kamper et al., 2010). A number of trials in our review demonstrated substantial effect sizes (0.5 and above) favouring PFR (Frost et al., 1995; Friedrich et al., 1998; Kankaanpaa et al., 1999; Steenstra et al., 2006; Pengel et al., 2007) with some of these describing detailed treatment protocols (Frost et al., 1995; Friedrich et al., 1998; Steenstra et al., 2006; Pengel et al., 2007). Future research should consider replication of these studies. Eligibility criteria for future trials should also be made more stringent to selectively recruit participants more likely to be responsive to PFR. Although this may have the effect of reducing generalisability, it is a necessary measure to ensure specificity of treatment is attained (Koes et al., 2006; Hay et al., 2008; Sowden et al., 2011). In this review, several studies applied eligibility criteria in this manner, with kinesiophobia (Woods and Asmundson, 2008) and disability (Kankaanpaa et al., 1999; UK BEAM Trial Team, 2004; Johnson et al., 2007) thresholds.

4.1. Limitations

Systematic reviews are most valuable when meta-analysis is conducted (Higgins et al., 2011), however between trial heterogeneity limited the use of this methodology in most comparisons. A number of trials analysed the specific population of sick-listed workers (Lindström et al., 1992; Staal et al., 2004; Heymans et al., 2006; Steenstra et al., 2006) limiting generalisability of findings (Furlan et al., 2009; BMJ Clinical Evidence, 2011). One trial (Steenstra et al., 2006) included a workplace intervention prior to a graded activity program and the authors acknowledged a potential impact on sick leave outcomes. It is possible that this review may have been effected by publication bias (Higgins et al., 2011).

5. Conclusion

Functional restoration is commonly recommended in evidence-based guidelines for post-acute LBP, however the effectiveness of PFR had not been evaluated in a high quality systematic review. Moderate to high quality evidence was limited to meta-analysis findings of small effects favouring PFR compared with advice for intermediate and long term follow up. Low to moderate quality evidence was found suggesting PFR is not different to other types of treatment. Although several trials concluded that there was no difference between groups, none were designed to conclusively demonstrate equivalence. Further high quality research on PFR is required replicating existing trial protocols.

Financial support

There are no declarations of financial support.

Conflicts of interest

There are no declarations of conflicts of interest.

Appendix A. Search strategy

Medline

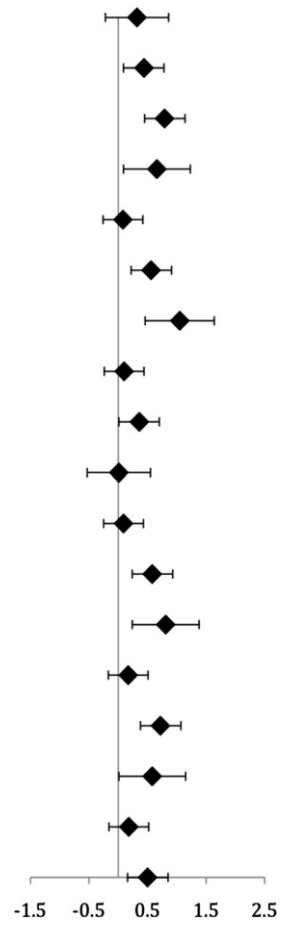
1. exp Back Pain/
2. (back adj pain).ab,ti.
3. (lumbar adj pain).ab,ti.
4. backache.ab,ti.
5. lumba$.ab,ti.
6. spondylosis.ab,ti.
7. coccyx.ab,ti.
8. coccydynia.ab,ti.
9. dorsalgia.ab,ti.
10. sciatica.ab,ti.
11. sciatica/
12. exp spine/
13. exp Low Back Pain/
14. LBP.ab,ti.
15. or/1–14
16. randomi?ed controlled trial.pt.
17. RCT.ti,ab.
18. controlled clinical trial.pt.
19. Exp Clinical Trial/
20. trial.ab,ti.
21. groups.ab,ti.
22. comparative stud$.ab,ti.
23. evaluation stud$.ab,ti.
24. follow?up stud$.ab,ti.
25. (control$ or prospectiv$ or volunteer$).ab,ti.
26. double-blind method/
27. single-blind method/
28. Random Allocation/
29. placebo$.ab,ti.
30. dt.fs.
31. Research Design/
32. ((singl$ or doubl$ or tripl$ or trebl$) adj25 (blind$ or mask$)).ab,ti.
33. random$.ab,ti.
34. (clin$ adj25 trial$).ab,ti.
35. (versus or vs).ab,ti.
36. (latin adj square).ab,ti.
37. Cross-over Studies/
38. or/16–37
39. (animals not (humans and animals)).sh.
40. 38 not 39
41. conditioning.ab,ti.
42. hardening.ab,ti.
43. functional restoration.ab,ti.
44. exercis$.ab,ti.
45. rehabil$.ab,ti.
46. graded.ab,ti.
47. stabili$.ab,ti
48. strength$.ab,ti.
49. or/41–48
50. exp Behaviour Therapy/
51. Conditioning, Operant/
52. exp “Reinforcement (Psychology)”/
53. operant.ab,ti.
54. respondent.ab,ti.
55. behav$.ab,ti.
56. Cognitive Therapy/
57. cognit$.ab,ti.
58. relaxation.ab,ti.
59. exp Relaxation/
60. or/50–59
61. 15 AND 40 AND 49 AND 60

Author's personal copy

Appendix B. Quality assessment (GRADE) and summary of findings for PFR versus comparison groups

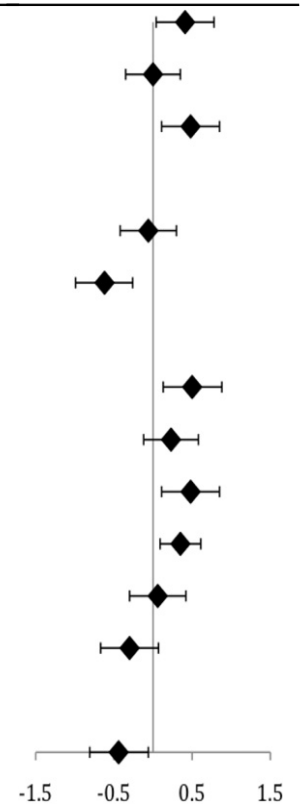
GRADE Table 1
Summary of findings: PFR versus placebo.

Outcome/ timeframe	Limitations	Inconsistency	Indirectness	Imprecision	Trial	PFR (*n*)	Placebo (*n*)	Effect size SMD/RR (95% CI)	GRADE quality	Forest plot of SMD (95% CI), positive values in favour of PFR
Pain/short	−1 limitations†	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Kankaanpaa	30	24	0.32 (−0.22, 0.86)	Very low	
					Pengel PFR1	65	68	**0.44 (0.09, 0.78)**		
					Pengel PFR2	63	68	**0.79 (0.45, 1.14)**		
Pain/intermediate	−1 limitations†	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Kankaanpaa	28	22	**0.66 (0.09, 1.23)**	Very low	
					Pengel PFR1	65	68	0.08 (−0.26, 0.42)		
					Pengel PFR2	63	68	**0.56 (0.22, 0.91)**		
Pain/long	−1 limitations†	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Kankaanpaa	27	22	**1.05 (0.46, 1.64)**	Very low	
					Pengel PFR1	65	68	0.10 (−0.24, 0.44)		
					Pengel PFR2	63	68	**0.36 (0.01, 0.70)**		
Function/short	−1 limitations†	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Kankaanpaa	30	24	0.01 (−0.53, 0.55)	Very low	
					Pengel PFR1	65	68	0.09 (−0.25, 0.43)		
					Pengel PFR2	63	68	**0.58 (0.24, 0.93)**		
Function/intermediate	−1 limitations†	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Kankaanpaa	28	22	**0.81 (0.24, 1.38)**	Very low	
					Pengel PFR1	65	68	0.17 (−0.17, 0.51)		
					Pengel PFR2	63	68	**0.72 (0.38, 1.07)**		
Function/long	−1 limitations†	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Kankaanpaa	27	22	**0.58 (0.01, 1.15)**	Very low	
					Pengel PFR1	65	68	0.18 (−0.16, 0.52)		
					Pengel PFR2	63	68	**0.50 (0.16, 0.85)**		



GRADE Table 2A
Summary of findings: PFR (Lindstrom et al., 1992) versus Dutch occupational physician guideline advice for sub-acute sick-listed workers.

Outcome/timeframe	Limitations	Inconsistency	Indirectness	Imprecision	Trial	PFR (n)	Advice (n)	Effect size SMD/RR (95% CI)	GRADE quality	Forest plot of SMD (95% CI), positive values in favour of PFR
Pain/short	No serious limitations	N/A	−1 indirectness§	−1 imprecision*	Steenstra	55	57	**0.41 (0.04, 0.78)**	Low	
Pain/intermediate	No serious limitations	−1 inconsistency‖	−1 indirectness§	No serious imprecision	Staal	61	63	0.00 (−0.35, 0.35)	Low	
					Steenstra	55	57	**0.48 (0.11, 0.85)**		
				Staal and Steenstra: $p = 0.07$, $I^2 = 70\%$ − statistical heterogeneity						
Pain/long	No serious limitations	−1 inconsistency‖	−1 indirectness§	No serious imprecision	Staal	60	59	−0.06 (−0.42, 0.30)	Low	
					Steenstra	55	57	**−0.62 (−0.99, −0.24)**		
				Staal and Steenstra: $p = 0.04$, $I^2 = 77\%$ − statistical heterogeneity						
Function/short	No serious limitations	N/A	−1 indirectness§	−1 imprecision*	Steenstra	55	57	**0.50 (0.13, 0.88)**	Low	
Function/intermediate	No serious limitations	No serious inconsistency	−1 indirectness§	No serious imprecision	Staal	62	64	0.23 (−0.12, 0.58)	Moderate	
					Steenstra	55	57	**0.48 (0.11, 0.85)**		
				Staal and Steenstra: $p = 0.34$, $I^2 = 0\%$ − pooled analysis		116	117	**0.35 (0.09, 0.61)**		
Function/long	No serious limitations	−1 inconsistency‖	−1 indirectness§	No serious imprecision	Staal	60	60	0.06 (−0.30, 0.42)	Low	
					Steenstra	55	57	−0.30 (−0.67, 0.07)		
				Staal and Steenstra: $p = 0.17$, $I^2 = 48\%$ − statistical heterogeneity						
Sick leave/long	No serious limitations	N/A	−1 indirectness§	−1 imprecision*	Steenstra	55	57	**−0.44 (−0.81, −0.06)**	Low	

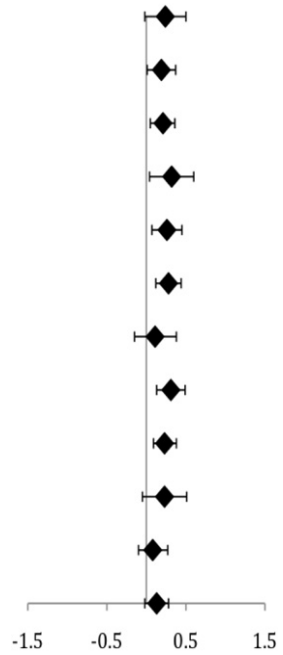


Author's personal copy

Author's personal copy

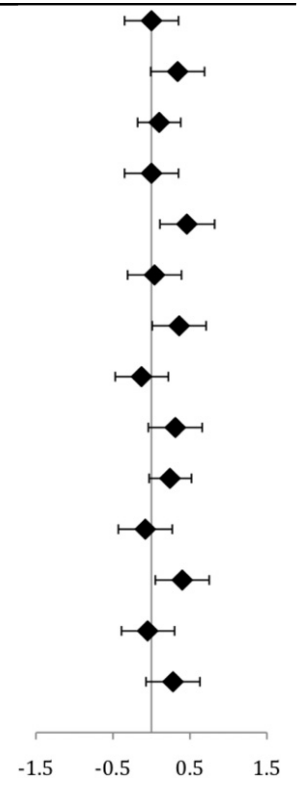
GRADE Table 2B
Summary of findings: PFR (Frost et al., 1995; Klaber Moffett et al., 1999) versus advice.

Outcome/timeframe	Limitations	Inconsistency	Indirectness	Imprecision	Trial	PFR (*n*)	Advice (*n*)	Effect size SMD/RR (95% CI)	GRADE quality	Forest plot of SMD (95% CI), positive values in favour of PFR
Pain/intermediate	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Johnson	110	113	0.24 (−0.02, 0.50)	High	
					UK BEAM	204	239	**0.19 (0.001^, 0.37)**		
				Johnson and UK BEAM: $p = 0.75$, $I^2 = 0\%$ − pooled analysis		314	352	**0.21 (0.05, 0.36)**		
Pain/long	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Johnson	102	94	**0.32 (0.04, 0.60)**	High	
					UK BEAM	200	235	**0.26 (0.07, 0.45)**		
				Johnson and UK BEAM: $p = 0.76$, $I^2 = 0\%$ − pooled analysis		302	329	**0.28 (0.12, 0.44)**		
Function/intermediate	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Johnson	110	113	0.11 (−0.15, 0.38)	High	
					UK BEAM	225	256	**0.31 (0.13, 0.49)**		
				Johnson and UK BEAM: $p = 0.27$, $I^2 = 17\%$ − pooled analysis		335	369	**0.23 (0.09, 0.38)**		
Function/long	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	Johnson	101	94	0.23 (−0.05, 0.51)	High	
					UK BEAM	216	248	0.08 (−0.10, 0.27)		
				Johnson and UK BEAM: $p = 0.39$, $I^2 = 0\%$ − pooled analysis		317	342	0.13 (−0.02, 0.28)		



GRADE Table 2C
Summary of findings: PFR (other) versus advice (other).

Outcome/timeframe	Limitations	Inconsistency	Indirectness	Imprecision	Trial	PFR (*n*)	Advice (*n*)	Effect size SMD/RR (95% CI)	GRADE quality	Forest plot of SMD (95% CI), positive values in favour of PFR
Pain/short	No serious limitations	No serious inconsistency	−1 indirectness‡	−1 imprecision*	Pengel PFR1	65	63	0.00 (−0.35, 0.35)	Low	
					Pengel PFR2	63	63	0.34 (−0.01, 0.69)		
Pain/intermediate	No serious limitations	−1 inconsistency#	−2 indirectness‡§	No serious imprecision	Heymans	98	103	0.10 (−0.18, 0.38)	Very low	
					Pengel PFR1	65	63	0.00 (−0.35, 0.35)		
					Pengel PFR2	63	63	**0.46 (0.11, 0.82)**		
Pain/long	No serious limitations	−1 inconsistency#	−1 indirectness‡	−1 imprecision*	Pengel PFR1	65	63	0.04 (−0.31, 0.39)	Very low	
					Pengel PFR2	63	63	**0.36 (0.01, 0.71)**		
Function/short	No serious limitations	No serious inconsistency	−1 indirectness‡	−1 imprecision*	Pengel PFR1	65	63	−0.13 (−0.47, 0.22)	Low	
					Pengel PFR2	63	63	0.31 (−0.04, 0.66)		
Function/intermediate	No serious limitations	−1 inconsistency#	−2 indirectness‡§	No serious imprecision	Heymans	98	103	0.24 (−0.03, 0.52)	Very low	
					Pengel PFR1	65	63	−0.08 (−0.43, 0.27)		
					Pengel PFR2	63	63	**0.40 (0.05, 0.75)**		
Function/long	No serious limitations	No serious inconsistency	−1 indirectness‡	−1 imprecision*	Pengel PFR1	65	63	−0.05 (−0.39, 0.30)	Low	
					Pengel PFR2	63	63	0.28 (−0.07, 0.63)		
Sick leave/intermediate	No serious limitations	N/A	−1 indirectness§	−1 imprecision*	Heymans	98	103	0.90 (0.70, 1.15) (dichotomous)	Low	

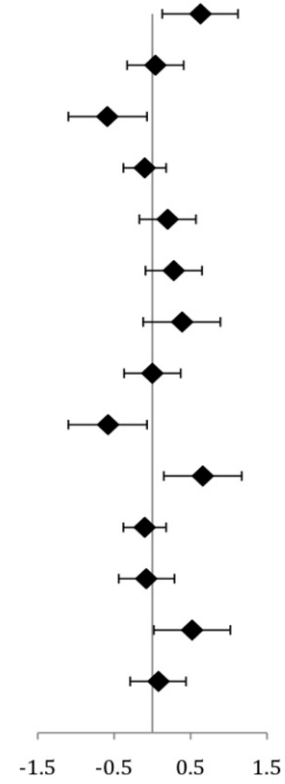


Author's personal copy

Author's personal copy

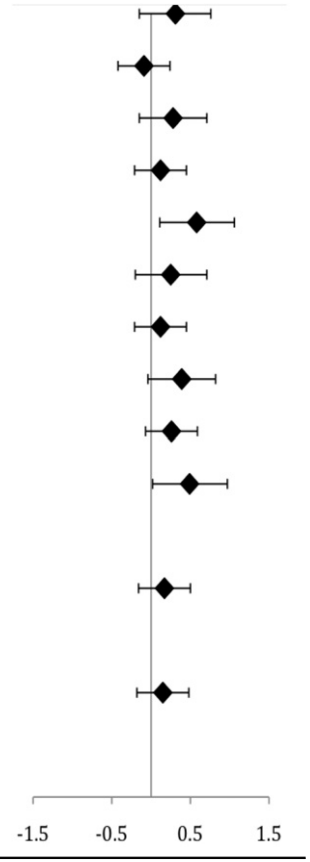
GRADE Table 3
Summary of findings: PFR versus other cognitive-behavioural therapy.

Outcome/timeframe	Limitations	Inconsistency	Indirectness	Imprecision	Trial	PFR (*n*)	Other behavioural (*n*)	Effect size SMD/RR (95% CI)	GRADE quality	Forest plot of SMD (95% CI), positive values in favour of PFR
Pain/short	−1 limitations†	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Frost	33	32	**0.63 (0.13, 1.12)**	Very low	
					van der Roer	60	54	0.04 (−0.33, 0.41)		
					Woods	25	36	**−0.59 (−1.10, −0.07)**		
Pain/intermediate	No serious limitations	No serious inconsistency	−2 indirectness‡§	No serious imprecision	Heymans	98	98	−0.10 (−0.38, 0.18)	Low	
					van der Roer	60	54	0.20 (−0.17, 0.57)		
Pain/long	No serious limitations	N/A	No serious indirectness	−1 imprecision*	van der Roer	60	54	0.28 (−0.09, 0.65)	Moderate	
Function/short	−1 limitations†	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Frost	29	32	0.39 (−0.12, 0.89)	Very low	
					van der Roer	60	54	0.00 (−0.37, 0.37)		
					Woods	25	36	**−0.58 (−1.10, −0.07)**		
Function/intermediate	No serious limitations	−1 inconsistency#	−2 indirectness‡§	No serious imprecision	Frost	29	32	**0.66 (0.15, 1.17)**	Very low	
					Heymans	98	98	−0.10 (−0.38, 0.18)		
					van der Roer	60	54	−0.08 (−0.44, 0.29)		
Function/long	No serious limitations	−1 inconsistency#	−1 indirectness‡	−1 imprecision*	Frost	31	31	**0.52 (0.02, 1.02)**	Very low	
					van der Roer	60	54	0.08 (−0.29, 0.44)		
Sick leave/intermediate	No serious limitations	N/A	−1 indirectness§	−1 imprecision*	Heymans	98	98	**0.80 (0.62, 0.98)** (dichotomous)	Low	



GRADE Table 4
Summary of findings: PFR versus other exercise therapy.

Outcome/timeframe	Limitations	Inconsistency	Indirectness	Imprecision	Trial	PFR (*n*)	Exercise Alone (*n*)	Effect size SMD/RR (95% CI)	GRADE quality	Forest plot of SMD (95% CI), positive values in favour of PFR
Pain/short	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Friedrich	39	36	0.31 (−0.15, 0.76)	Moderate	
Pain/intermediate	No serious limitations	No serious inconsistency	−1 indirectness‡	No serious imprecision	Critchley	69	72	−0.09 (−0.42, 0.24)	Moderate	
					Friedrich	43	41	0.28 (−0.15, 0.71)		
Pain/long	No serious limitations	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Critchley	69	72	0.12 (−0.21, 0.45)	Low	
					Friedrich	34	35	**0.58 (0.11, 1.06)**		
Function/short	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Friedrich	38	36	0.25 (−0.20, 0.71)	Moderate	
Function/intermediate	No serious limitations	No serious inconsistency	−1 indirectness‡	No serious imprecision	Critchley	69	72	0.12 (−0.21, 0.45)	Moderate	
					Friedrich	43	41	0.39 (−0.04, 0.82)		
Function/long	No serious limitations	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Critchley	69	72	0.26 (−0.07, 0.59)	Low	
					Friedrich	34	35	**0.49 (0.02, 0.97)**		
Sick Leave/short	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Friedrich	38	36	1.12 (0.70, 1.76) (dichotomous)	Moderate	
Sick leave/intermediate	No serious limitations	No serious inconsistency	−1 indirectness‡	No serious imprecision	Critchley	69	72	0.17 (−0.16, 0.50)	Moderate	
					Friedrich	43	41	1.36 (0.95, 1.95) (dichotomous)		
Sick leave/long	No serious limitations	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Critchley	69	72	(−0.18, 0.48)	Low	
					Friedrich (12 months)	34	35	1.29 (0.91, 1.83) (dichotomous)		
					(5-years)	26	30	**1.49 (1.05, 2.13)** (dichotomous)		

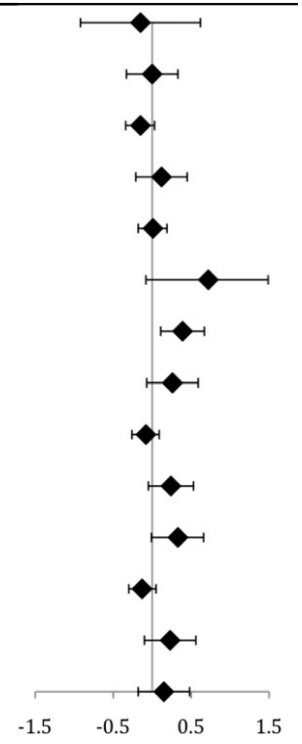


Author's personal copy

Author's personal copy

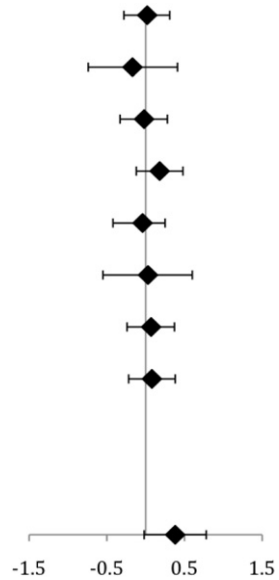
GRADE Table 5
Summary of findings: PFR versus other therapy.

Outcome/timeframe	Limitations	Inconsistency	Indirectness	Imprecision	Trial	PFR (*n*)	Other Therapy (*n*)	Effect Size SMD/RR (95% CI)	GRADE Quality	Forest plot of SMD (95% CI), positive values in favour of PFR
Pain/short	No serious limitations	N/A	−1 indirectness§	−1 imprecision*	Paoloni	13	13	−0.15 (−0.92, 0.62)	Low	
Pain/intermediate	No serious limitations	No serious inconsistency	−1 indirectness‡	No serious imprecision	Critchley	69	71	0.00 (−0.33, 0.33)	Moderate	
					UK BEAM	204	275	−0.15 (−0.34, 0.03)		
Pain/long	No serious limitations	No serious inconsistency	−1 indirectness‡	No serious imprecision	Critchley	69	71	0.12 (−0.21, 0.45)	Moderate	
					UK BEAM	200	264	0.01 (−0.18, 0.19)		
Function/short	No serious limitations	N/A	−1 indirectness§	−1 imprecision*	Paoloni	13	13	0.72 (−0.08, 1.51)	Low	
Function/intermediate	No serious limitations	−1 inconsistency#	−1 indirectness‡	No serious imprecision	Carr	94	109	**0.39 (0.11, 0.67)**	Low	
					Critchley	69	71	0.26 (−0.07, 0.59)		
					UK BEAM	225	287	−0.08 (−0.26, 0.09)		
Function/long	No serious limitations	No serious inconsistency	−1 indirectness‡	No serious imprecision	Carr	92	88	0.24 (−0.05, 0.53)	Moderate	
					Critchley	69	71	0.33 (−0.003^, 0.66)		
					UK BEAM	216	273	−0.13 (−0.30, 0.05)		
Sick leave/intermediate	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Critchley	69	71	0.23 (−0.10, 0.56)	Moderate	
Sick leave/long	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Critchley	69	71	0.15 (−0.18, 0.48)	Moderate	



GRADE Table 6
Summary of findings: PFR versus minimal intervention.

Outcome/timeframe	Limitations	Inconsistency	Indirectness	Imprecision	Trial	PFR (n)	Minimal Intervention (n)	Effect Size SMD/RR (95% CI)	GRADE Quality	Forest plot of SMD (95% CI), positive values in favour of PFR
Pain/short	−1 limitations†	No serious inconsistency	−1 indirectness‡	No serious imprecision	Klaber Moffett	85	94	0.02 (−0.28, 0.31)	Low	
					Woods	25	22	−0.17 (−0.74, 0.41)		
Pain/intermediate	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Klaber Moffett	77	86	−0.02 (−0.33, 0.28)	Moderate	
Pain/long	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Klaber Moffett	83	88	0.18 (−0.12, 0.48)	Moderate	
Function/short	−1 limitations†	No serious inconsistency	−1 indirectness‡	No serious imprecision	Klaber Moffett	85	94	−0.04 (−0.34, 0.25)	Low	
					Woods	25	22	0.03 (−0.55, 0.60)		
Function/intermediate	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Klaber Moffett	77	86	0.07 (−0.24, 0.37)	Moderate	
Function/long	No serious limitations	N/A	No serious indirectness	−1 imprecision*	Klaber Moffett	83	88	0.08 (−0.22, 0.38)	Moderate	
Sick leave/short	−1 limitations†	N/A	−1 indirectness§	−1 imprecision*	Lindström (6 weeks)	49	49	1.45 (0.96, 2.18) (dichotomous)	Very Low	
					Lindström (12 weeks)	49	49	**1.39 (1.05, 1.84)** (dichotomous)		
Sick leave/long	−1 limitations†	N/A	−1 indirectness§	−1 imprecision*	Lindström	49	49	0.38 (−0.02, 0.78)	Very Low	



GRADE table legend:
† downgrading due to >25% of trials at a high risk of bias (score <6/10) (Higgins et al., 2011)
# downgrading due to <75% of comparisons showing consistent effects and/or clinical significance (Atkins et al., 2004)
|| downgrading due to statistical heterogeneity (Gross et al., 2010)
‡ downgrading due to clinical heterogeneity (BMJ Clinical Evidence, 2011)
§ downgrading due to trials with restricted sample population (Furlan et al., 2009; BMJ Clinical Evidence, 2011)
* downgrading due to sparse data (single trial or $n < 200$ for multiple trials)
^ CI calculated to 3 decimal places to determine statistical significance
Effect size in bold denotes statistical significance.
Refer to list of short forms (Appendix E) for other abbreviations.

Author's personal copy

Author's personal copy

Appendix C. GRADE quality of the body of evidence domain definitions

Methodological quality limitation refers to the risk of bias of included trials. The methodological quality of trials was rated using the PEDro scale which has demonstrable reliability (Maher et al., 2003; Bhogal et al., 2005). Trials with a score of six or more were considered at low risk of bias. If 75% or more of the included trials scored six or more on the PEDro scale then this domain was adjudged as having "no limitations" (Atkins et al., 2004; Furlan et al., 2009; Schaafsma et al., 2010). A quality point was deducted when less than 75% of trials scored six or more.

Consistency refers to the similarity of treatment effect estimates for each outcome across the trials. Trial results were considered consistent when directions, effect size and statistical significance were considered similar enough to draw the same conclusion (Atkins et al., 2004). Prior to meta-analysis, clinically homogenous trials were first assessed for statistical heterogeneity and in this situation consistency was defined as absence of statistical heterogeneity (Gross et al., 2010). A quality point was deducted for statistically heterogeneity (BMJ Clinical Evidence, 2011). In the case of clinically heterogeneous trials a quality point was deducted for inconsistent or conflicting (or varying) results (BMJ Clinical Evidence, 2011). "Consistency in direction was defined as 75% or more of the included trials showing either benefit or no benefit, and consistency of effect when 75% or more of the trials showing a clinically important or unimportant treatment effect" (Higgins et al., 2011) based on the minimum clinically important difference for the outcome measures relevant to the review. This domain was not applicable when there was only one trial intervention per outcome.

Directness refers to the extent to which the people, interventions and outcome measures are similar to those of interest (Atkins et al., 2004; Gross et al., 2010). A quality point was deducted for limited generalisability of findings due to restricted sample population where results were more applicable to a specific population (Furlan et al., 2009; BMJ Clinical Evidence, 2011). Directness was also influenced by intervention differences (Brozek et al., 2009), where a quality point was deducted for clinical heterogeneity between trials (BMJ Clinical Evidence, 2011). Downgrading was conducted for some (−1) or major (−2) uncertainty about directness (Atkins et al., 2004) where a score of −2 was given where there was a problem with 2 or more elements (BMJ Clinical Evidence, 2011). Consistent with other domains, a threshold for directness was applied where 75% or more of included trials satisfied the above criteria.

Precision refers to the number of trials, population and the events for each outcome (Furlan et al., 2009). A quality point was deducted for imprecision when:

- Only one trial reported an outcome for the chosen comparison (Furlan et al., 2009)
- There was sparse data with less than 200 participants per comparison (BMJ Clinical Evidence, 2011)

Publication (reporting) bias as described within the Cochrane Handbook (Higgins et al., 2011) was only considered present if actual evidence was found (Gross et al., 2010).

Appendix D. Systematic review article selection protocol

Stage 1 – Article exclusion (based on title and abstract assessment)

Title and abstract are to be analysed by 2 independent reviewers. During this process, the following criteria must be met:

1. *Trial type:* Exclude if the trial is clearly other than human adult randomised or controlled or clinical trial. (ie: it is labelled a case report, case series, comment, conference paper, retrospective/compliance/predictor/uncontrolled). NB: Trials are not to be excluded at this stage if the randomisation is questionable. (eg: controlled trials with specific allocation or blinding issues). NB: Articles are not to be excluded on the basis of language.
2. *Condition:* Exclude if low back pain (LBP) is not a condition studied. Exclude non-mechanical or specific LBP conditions including: surgery, infection, neoplasm, metastasis, osteoporosis, fractures, inflammatory conditions, spinal cord lesions, pregnancy.
3. *Intervention:* Exclude if it is clear that exercise (or similar) is not a treatment component. Exclude if intervention is multidisciplinary (where members of more than one profession participate in the treatment program). Exclude if intervention is conducted by a profession other than "physiotherapy" or "physical therapy". NB: combined exercise and non-exercise treatments may be described by specific terms not indicating profession, such as: functional restoration, back school, multimodal, operant, conditioning, hardening, rehabilitation, graded activity – include for full text analysis if uncertain.

Stage 2 – Article exclusion (based on full text

Articles must satisfy ALL of the following criteria to be included in the systematic review:

1. *Publication type:* Only full reports of journal articles are to be included.
2. *Trial type:* Only randomised controlled trials (RCT) with human subjects are to be included. Assume it is an RCT if it uses the word "random" when referring to allocation to groups. Non-acceptable randomisation procedures include any "predictable" or "systematic" allocation method, such as alternate allocation, or allocation via date of birth. Cross-over trials can only be included if they present group outcome data before the cross-over occurs.
3. *Participants:* Exclude if participants are either pregnant or under the age of 18 years, or if >30% of participants have symptoms <6 weeks in duration
4. Condition: Include non-specific LBP or mechanical LBP not attributed to a recognisable pathology (as per title and abstract criteria). LBP is pain, muscle tension, or stiffness localised below the costal margin and above the inferior gluteal folds, with or without leg pain (sciatica – pain that radiates from the back into the buttock or leg and is most commonly caused by prolapse of an intervertebral disk; the term may also be used to describe pain anywhere along the course of the sciatic nerve) (van Tulder, 2003). For trials examining a mixed population, >70% of participants are required to have LBP as their primary condition
5. *Intervention:* The PFR intervention must include:

   A. Both an exercise component and a behavioural component. The "behavioural" component requires either the use of at least one of the following descriptors: "psychological, cognitive, behavioural, relaxation, operant, social, coping, respondent, hypnosis, counseling", or the description of a clear behavioural approach (an example of an insufficient description is "participants were told that exercising with pain is alright" as this does not sufficiently describe a behavioural approach)

   B. at least 3 hours of total intervention time, or a minimum of 10 sessions
6. *Comparison groups:* Must be able to assess the effectiveness of PFR. Exclude trials with interventions which both involve PFR

Author's personal copy

where one has an additional unrelated component (eg: PFR v PFR + socialisation: this is studying the effect of socialisation and not PFR)

7. *Treaters:* The treatment must only be run by one or more physiotherapist(s) or physical therapist(s). Multi-disciplinary programs are to be excluded.
8. *Treatment must be non-invasive:* Exclude treatments including an invasive treatment component (surgery, injections/deep skin penetration)
9. *Treatment must be predominantly active:* Exclude trials where passive treatment modalities (eg: manual therapy, acupuncture, electrotherapy) were utilised in at least 50% of sessions. Combining PFR and ongoing passive treatment contradicts behavioural messages such as self-management promotion.
10. *Outcome measures:* Must have between-group comparison for at least one outcome measure or present data to allow comparison (eg: mean score with standard deviations for each group for continuous variables, or the number of subjects in each group who achieved a certain outcome is provided for categorical/dichotomous data). One measure of pain, function or sick leave must be included.

Appendix E. Log of short forms

ABPS	Aberdeen back pain scale (function)
CBT	cognitive behavioural therapy
CI	confidence interval
GP	general practitioner
GRADE	grading of recommendations, assessment, development and evaluation
LBOS	low back outcome score (function)
LBP	low back pain
MPQ-SF	McGill pain questionnaire-short form (pain)
N	sample
N/A	not applicable
NPRS	numerical pain rating scale (pain)
OP	occupational physician
PDI	pain disability index (function)
PFR	physiotherapy functional restoration
PSFS	patient specific functional scale (function)
PT	physiotherapy
RCT	randomised controlled trial
RMDQ	Roland-Morris disability questionnaire (function)
RR	relative risk
SD	standard deviation
SMD	standardised mean difference
VAS	visual analogue scale (pain)

References

Anema J, Steenstra I, Bongers P, de Vet H, Knol D, Loisel P, et al. Multidisciplinary rehabilitation for subacute low back pain: graded activity or workplace intervention or both? Spine 2007;32(3):291–8.

Atkins D, Best D, Briss P, Eccles M, Falck-Ytter Y, Flottorp S. Grading quality of evidence and strength of recommendations. The GRADE Working Group. British Medical Journal (Clinical Research Ed.) 2004;328(7454):1490.

Australian Institute of Health and Welfare. When musculoskeletal conditions and mental disorders occur together. Cat. no. AUS 129. Canberra: AIHW; 2010.

Bekkering G, Hendriks H, Koes B, Oostendorp R, Ostelo R, Thomassen J, et al. Dutch physiotherapy guidelines for low back pain. Physiotherapy 2003;89(2):82–96.

Bhogal S, Teasell R, Foley N, Speechley M. The PEDro scale provides a more comprehensive measure of methodological quality than the Jada scale in stroke rehabilitation literature. Journal of Clinical Epidemiology 2005;58(7):668–73.

BMJ Clinical Evidence. What is GRADE? [BMJ Clinical Evidence website]. Available from: http://clinicalevidence.bmj.com/ceweb/about/about-grade.jsp; 2011 [accessed 24.02.11].

Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine 2000;25(24):3100–3.

Bombardier C, Hayden J, Beaton D. Minimal clinically important difference. Low back pain: outcome measures. Journal of Rheumatology 2001;28(2):431–8.

Brozek J, Akl E, Alonso-Coello P, Lang D, Jaeschke R, Williams J, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. Allergy 2009;64(5):669–77.

Bunzli S, Gillham D, Esterman A. Physiotherapy-provided operant conditioning in the management of low back pain disability: a systematic review. Physiotherapy Research International 2011;16(1):4–19.

Carr J, Klaber Moffett J, Howarth E, Richmond S, Torgerson D, Jackson D, et al. A randomized trial comparing a group exercise programme for back pain patients with individual physiotherapy in a severely deprived area. Disability and Rehabilitation 2005;27(16):929–37.

Chou R, Huffman L. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. Annals of Internal Medicine 2007;147(7):492–504.

Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

Critchley D, Ratcliffe J, Noonan S, Jones R, Hurley M. Effectiveness and cost-effectiveness of three types of physiotherapy used to reduce chronic low back pain disability. Spine 2007;32(14):1474–81.

Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine Journal 2008;8(1):8–20.

de Morton N. The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. Australian Journal of Physiotherapy 2009;55(2):129–33.

Deyo R, Battie M, Beurskens A, Bombardier C, Croft P, Koes B, et al. Outcome measure for low back pain research: a proposal for standardized use. Spine 1998;23(18):2003–13.

Engers A, Jellema P, Wensing M, van der Windt D, Grol R, van Tulder M. Individual patient education for low back pain. Cochrane Database of Systematic Reviews 2008;1:CD004057.

Ford J, Story I, O'Sullivan P, McMeeken J. Classification systems for low back pain: a review of the methodology for development and validation. Physical Therapy Reviews 2007;12(1):33–42.

Friedrich M, Gittler G, Halberstadt Y, Cermak T, Heiller I. Combined exercise and motivation program: effect on the compliance and level of disability of patients with chronic low back pain: a randomised controlled trial. Archives of Physical Medicine and Rehabilitation 1998;79(5):475–87.

Friedrich M, Gittler G, Arendasy M, Friedrich K. Long-term effect of a combined exercise and motivational program on the level of disability of patients with chronic low back pain. Spine 2005;30(9):995–1000.

Fritz J, Cleland J, Childs J. Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. Journal of Orthopaedic and Sports Physical Therapy 2007;37(6):290–302.

Frost H, Klaber Moffett J, Moser J, Fairbank J. Randomised controlled trial for evaluation of fitness programme for patients with chronic low back pain. British Medical Journal (Clinical Research Ed.) 1995;310(6973):151–4.

Frost H, Lamb S, Klaber-Moffett J, Fairbank J, Moser J. A fitness programme for patients with chronic low back pain: 2-year follow-up of a randomised controlled trial. Pain 1998;75:273–9.

Furlan A, Pennick V, Bombardier C, van Tulder M. 2009 Updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine 2009;34(18):1929–41.

Gatchel R, Mayer T. Evidence-informed management of chronic low back pain with functional restoration. The Spine Journal 2008;8(1):65–9.

George S. What is the effectiveness of a biopsychosocial approach to individual physiotherapy care for chronic low back pain? The Internet Journal of Allied Health Sciences and Practice 2008;6(1):1–10.

George S, Zeppieri Jr G, Cere A, Cere M, Borut M, Hodges M, et al. A randomized trial of behavioral physical therapy interventions for acute and sub-acute low back pain. Pain 2008;140(1):145–57.

Gross A, Miller J, D'Sylva J, Burnie S, Goldsmith C, Graham N, et al. Manipulation or mobilisation for neck pain: a Cochrane review. Manual Therapy 2010;15:315–33.

Guzman J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: systematic review. British Medical Journal (Clinical Research Ed.) 2001;322(7301):1511–6.

Hahne A, Ford J, McMeeken J. Conservative management of lumbar disc herniation with associated radiculopathy: a systematic review. Spine 2010;35(11):E488–504.

Hartigan C, Miller L, Liewehr S. Rehabilitation of acute and subacute low back and neck pain in the work-injured patient. Orthopedic Clinics of North America 1996;27(4):841–60.

Hay E, Dunn K, Hill J, Lewis M, Mason E, Konstantinou K, et al. A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT back trial study protocol. BMC Musculoskeletal Disorders 2008;9:58.

Hay E, Mullis R, Lewis M, Vohora K, Main C, Watson P, et al. Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. Lancet 2005;365(9476):2024–30.

Hayden J, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. Journal of Clinical Epidemiology 2009;62(8):781–96.

Author's personal copy

Hayden J, van Tulder M, Malmivaara A, Koes B. Exercise therapy for treatment of non-specific low back pain. Cochrane Database of Systematic Reviews 2005;3: CD000335.
Hedges L, Olkin I. Statistical methods for meta-analysis. 1st ed. Orlando, FL: Academic Press; 1985.
Henschke N, Ostelo R, van Tulder M, Vlaeyen J, Morley S, Assendelft W, et al. Behavioural treatment for chronic low-back pain. Cochrane Database of Systematic Reviews 2010;7:CD002014.
Herbert R. How to estimate treatment effects from reports of clinical trials. II: dichotomous outcomes. Australian Journal of Physiotherapy 2000;46(4):309–13.
Herbert R, Jamtvedt G, Mead J, Hagen K. Practical evidence-based physiotherapy. London: Elsevier Butterworth-Heinemann; 2005.
Hestbaek L, Leboeuf-Yde C, Manniche C. Low back pain: what is the long-term course? A review of studies of general patient populations. European Spine Journal 2003;12(2):149–65.
Heymans M, de Vet H, Bongers P, Knol D, Koes B, van Mechelen W. The effectiveness of high-intensity versus low-intensity back schools in an occupational setting. Spine 2006;31(10):1075–82.
Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available from: www.cochrane-handbook.org; 2011 [accessed 25.05.12].
Hlobil H, Staal J, Twisk J, Koke A, Ariens G, Smid T, et al. The effects of a graded activity intervention for low back pain in occupational health on sick leave, functional status and pain: 12-month results of a randomized controlled trial. Journal of Occupational Rehabilitation 2005;15(4):569–80.
Hozo S, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range and the size of the sample. BMC Medical Research Methodology 2005;5:13.
Indahl A, Velund L, Reikeraas O. Good prognosis for low back pain when left untampered. Spine 1995;20(4):473–7.
Johnson R, Jones G, Wiles N, Chaddock C, Potter R, Roberts C, et al. Active exercise, education, and cognitive behavioral therapy for persistent disabling low back pain. Spine 2007;32(15):1578–85.
Kamper S, Maher C, Hancock M, Koes B, Croft P, Hay E. Treatment-based subgroups of low back pain: a guide to appraisal of research studies and a summary of current evidence. Best Practice and Research: Clinical Rheumatology 2010; 24(2):181–91.
Kankaanpaa M, Taimela S, Airaksinen O, Hanninen O. The efficacy of active rehabilitation in chronic low back pain. Spine 1999;24(10):1034–42.
Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, et al. Multidisciplinary biopsychosocial rehabilitation for subacute low back pain in working-age adults. Spine 2001;26(3):262–9.
Katz J. Lumbar disc disorders and low-back pain: socioecomonic factors and consequences. Journal of Bone and Joint Surgery. American Volume 2006;88-(Suppl. 2):21–4.
Kent P, Keating J. Can we predict poor recovery from recent-onset nonspecific low back pain? A systematic review. Manual Therapy 2008;13(1):12–28.
Klaber Moffett J, Frost H. Back to fitness programme: the manual for physiotherapists to set up the classes. Physiotherapy 2000;86(6):295–305.
Klaber Moffett J, Torgerson D, Bell-Syer S, Jackson D, Llewelyn Phillips H, Farrin A, et al. A randomised trial of exercise for primary care back pain patients: clinical outcomes, costs and preferences. British Medical Journal (Clinical Research Ed.) 1999;319:279–83.
Koes B, van Tulder M, Lin C, Macedo L, McAuley J, Maher C. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. European Spine Journal 2010;19(12):2075–94.
Koes B, van Tulder M, Thomas S. Diagnosis and treatment of low back pain. British Medical Journal (Clinical Research Ed.) 2006;332(7555):1430–4.
Leijon O, Mulder M. Prevalence of low back pain and concurrent psychological distress over a 16-year period. Occupational and Environmental Medicine 2009; 66(2):137–9.
Lindgren K, Sihvonen T, Leino E, Pitkanen M, Manninen H. Exercise therapy effects on functional radiographic findings and segmental electromyographic activity in lumbar spine instability. Archives of Physical Medicine and Rehabilitation 1993;74(9):933–9.
Lindström I, Öhlund C, Wallin L, Peterson L, Nachemson A. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Physical Therapy 1992;72(4):279–93.
Macedo L, Smeets R, Maher C, Latimer J, McAuley J. Graded activity and graded exposure for persistent nonspecific low back pain: a systematic review. Physical Therapy 2010;90(6):860–79.
Maetzel A, Li L. The economic burdon of low back pain: a review of studies published between 1996 and 2001. Best Practice and Research Clinical Rheumatology 2002;16(1):23–30.
Maher C, Sherrington C, Herbert R, Moseley A, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Physical Therapy 2003; 83(8):713–21.
Maher CG. A systematic review of workplace interventions to prevent low back pain. Australian Journal of Physiotherapy 2000;46(4):259–69.
Main C, Spanswick C. Pain management: an interdisciplinary approach. Edinburgh: Churchill-Livingstone; 2000.
Mayer T, Gatchel R, Kishino N, Keeley J, Capra P, Mayer H, et al. Objective assessment of spine function following industrial injury; a prospective study with comparison group and one-year follow up. Spine 1985;10(6):482–93.
Norlund A, Ropponen A, Alexanderson K. Multidisciplinary interventions: review of studies of return to work after rehabilitation for low back pain. Journal of Rehabilitation Medicine 2009;41(3):115–21.
Ostelo R, de Vet H. Clinically important outcomes in low back pain. Best Practice and Research in Clinical Rheumatology 2005;19(4):593–607.
Ostelo R, van Tulder M, Vlaeyen J, Linton S, Morley S, Assendelft W. Behavioural treatment for chronic low-back pain. Cochrane Database of Systematic Reviews 2005;3:CD002014.
Paoloni M, Bernetti A, Fratocchi G, Mangone M, Parrinello L, Del Pilar Cooper M, et al. Kinesio Taping applied to lumbar muscles influences clinical and electromyographic characteristics in chronic low back pain patients. European Journal of Physical and Rehabilitation Medicine 2011; 47(2):237–44.
Pengel L, Refshauge K, Maher C, Nicholas M, Herbert R, McNair P. Physiotherapy directed exercise, advice, or both for subacute low back pain. Annals of Internal Medicine 2007;146(11):787–96.
Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of non-inferiority and equivalence randomized trials: an extension of the CONSORT statement. Journal of the American Medical Association 2006;295(10): 1152–60.
Poiraudeau S, Rannou F, Revel M. Functional restoration programs for low back pain: a systematic review. Annales de Readaptation et de Medecine Physique 2007;50(6):425–9.
Version 5.1Review Manager (Revman) [computer program]. Copenhagen: The Nordic Cochrane Centre; 2011.
Roland M, Waddell G, Klaber Moffett J, Burton A, Main C, Cantrell T. The back book. Norwich: Stationary Office; 1996.
Schaafsma F, Schonstein E, Whelan K, Ulvestad E, Kenny D, Verbeek J. Physical conditioning programs for improving work outcomes in workers with back pain. Cochrane Database of Systematic Reviews 2010;1:CD001822.
Schonstein E, Kenny D, Keating J, Koes B. Work conditioning, work hardening and functional restoration for workers with back and neck pain. Cochrane Database of Systematic Reviews 2003;3:CD001822.
Slater SL, Ford JJ, Richards MC, Taylor NF, Surkitt LD, Hahne AJ. The effectiveness of sub-group specific manual therapy for low back pain: a systematic review. Manual Therapy 2012;17(3):201–12.
Sowden G, Hill J, Konstantinou K, Khanna M, Main C, Salmon P, et al. Subgrouping for targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study. Family Practice [serial online]. Available from: http://fampra.oxfordjournals.org; 2011 [accessed 05.07.11].
Staal J, Hlobil H, Twisk J, Smid T, Köke A, van Mechelen W. Graded activity for low back pain in occupational health care. Annals of Internal Medicine 2004;140(2): 77–84.
Staal J, Hlobil H, van Tulder M. Occupational health guidelines for the management of low back pain: an international comparison. Occupational and Environmental Medicine 2003;60(9):618–26.
Stanton TR, Latimer J, Maher CG, Hancock M. Definitions of recurrence of an episode of low back pain: a systematic review. Spine (Phila Pa 1976) 2009;34(9): E316–22.
Steenstra I, Anema J, Bongers P, de Vet H, Knol D, van Mechelen W. The effectiveness of graded activity for low back pain in occupational healthcare. Occupational and Environmental Medicine 2006;63(11):718–25.
Surkitt LD, Ford JJ, Hahne AJ, Pizzari T, McMeeken JM. Efficacy of directional preference management for low back pain: a systematic review. Physical Therapy 2012;92(5):652–65.
UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. British Medical Journal (Clinical Research Ed.) 2004;329(7479): 1377.
van der Roer N, van Tulder M, Barendse J, Knol D, van Mechelen W, de Vet H. Intensive group training protocol versus guideline physiotherapy for patients with chronic low back pain: a randomised controlled trial. European Spine Journal 2008;17(9):1193–200.
van der Roer N, van Tulder M, Barendse J, van Mechelen W, Franken W, Ooms A, et al. Cost-effectiveness of an intensive group training protocol compared to physiotherapy guideline care for sub-acute and chronic low back pain: design of a randomised controlled trial with an economic evaluation. BMC 2004;5:45.
van der Windt D, Hay E, Jellema P, Main C. Psychosocial interventions for low back pain in primary care: lessons learned from recent trials. Spine 2008;33(1): 81–9.
van Geen J, Edelaar M, Janssen M, van Eijk J. The long term effect of multidisciplinary back training: a systematic review. Spine 2007;32(2):249–55.
van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003;28(12):1290–9.
Vlaeyen J, Haazen I, Schuerman J, Kole-Snijders A, van Eek H. Behavioral rehabilitation of chronic low back pain: comparison of an operant treatment, an operant-cognitive treatment and an operant-respondent treatment. British Journal of Clinical Psychology 1995;34:95–118.
von Korff M, Crane P, Lane M, Miglioretti D, Simon G, Saunders K, et al. Chronic spinal pain and physical-mental comorbidity in the United States: results from the national comorbidity survey replication. Pain 2005;113(3): 331–9.

Author's personal copy

Walker B, Muller R, Grant W. Low back pain in Australian adults. Prevalence and associated disability. Journal of Manipulative and Physiological Therapeutics 2004;27(4):238–44.
Wong S, Wilczynski N, Haynes R. Optimal CINAHL search strategies for identifying therapy studies and review articles. Journal of Nursing Scholarship 2006;38(2):194–9.
Woods M, Asmundson G. Evaluating the efficacy of graded in vivo exposure for the treatment of fear in patients with chronic back pain: a randomized controlled clinical trial. Pain 2008;136(3):271–80.
Zhang L, Ajiferuke I, Sampson M. Optimizing search strategies to identify randomized controlled trials in MEDLINE. BMC Medical Research Methodology 2006;6:23.