Should exercises be painful in the management of chronic musculoskeletal pain? A systematic review and meta-analysis

Benjamin E Smith,1,2 Paul Hendrick,3 Toby O Smith,4 Marcus Bateman,1 Fiona Moffatt,3 Michael S Rathleff,5,6 James Selfe,7 Pip Logan2

ABSTRACT
Background Chronic musculoskeletal disorders are a prevalent and costly global health issue. A new form of exercise therapy focused on loading and resistance programmes that temporarily aggravates a patient’s pain has been proposed. The object of this review was to compare the effect of exercises where pain is allowed/encouraged compared with non-painful exercises on pain, function or disability in patients with chronic musculoskeletal pain within randomised controlled trials.

Methods Two authors independently selected studies and appraised risk of bias. Methodological quality was evaluated using the Cochrane risk of bias tool, and the Grading of Recommendations Assessment system was used to evaluate the quality of evidence.

Results The literature search identified 9081 potentially eligible studies. Nine papers (from seven trials) with 385 participants met the inclusion criteria. There was short-term significant difference in pain, with moderate quality evidence for a small effect size of −0.27 (−0.54 to −0.05) in favour of painful exercises. For pain in the medium and long term, and function and disability in the short, medium and long term, there was no significant difference.

Conclusion Protocols using painful exercises offer a small but significant benefit over pain-free exercises in the short term, with moderate quality of evidence. In the medium and long term there is no clear superiority of one treatment over another. Pain during therapeutic exercise for chronic musculoskeletal pain need not be a barrier to successful outcomes. Further research is warranted to fully evaluate the effectiveness of loading and resistance programmes into pain for chronic musculoskeletal disorders.

PROSPERO registration CRD42016038882.

BACKGROUND
Musculoskeletal disorders are one of the most prevalent and costly disorders globally.1 2 Low back pain is considered the leading cause of years lived with disability worldwide, ahead of conditions such as depression, diabetes, cardiovascular disease and cancer, with a global point prevalence of 9.4%.3 4 Neck pain and other musculoskeletal pain ranks fourth and sixth in terms of years lived with disability, with a global point prevalence of 5% and 8%, respectively.5 6 In the UK, an estimated one in four people suffer from chronic musculoskeletal disorders,7 with an estimated economic consequence of 8.8 million working days lost.8

Previous systematic reviews have assessed the effectiveness of various interventions for musculoskeletal disorders, including pharmaceutical therapies,9–12 psychological-based therapies13–16 and physical-based therapies, including manual therapy17–19 and exercise.20–24 These have all presented poor to moderate results in terms of effectiveness at improving pain and function, and have identified limitations in the quality of included trials when drawing conclusions.

There is a high level of uncertainty and lack of sufficient level 1 evidence on which to base treatment for people with musculoskeletal disorders. A systematic review of self-management interventions for chronic musculoskeletal pain concluded that strong evidence existed that changes in the psychological factors, self-efficacy and depression were predictors of outcomes, irrespective of the intervention delivered, and strong evidence existed that positive changes in patients’ pain catastrophising and physical activity were mediating factors.25 Experimental studies have also demonstrated that stimulus context and the emotional response to pain affect the experience of pain,26–28 and have led to the development of desensitisation interventions for chronic musculoskeletal disorders.29–31

It has been proposed that modern treatment therapies for chronic musculoskeletal pain and disorders should be designed around loading and resistance programmes targeting movements and activities that can temporarily reproduce and aggravate patients’ pain and symptoms.31–33 Pain does not correlate with tissue damage,34 and psychological factors such as catastrophising and fear avoidance behaviours play an important role in the shaping of the physiologial responses to pain, and therefore the development and maintenance of chronic pain.35 It is thought that such an exercise programme could facilitate the reconceptualisation of pain by addressing fear avoidance and catastrophising beliefs within a framework of ‘hurt not equalling harm’.36 37 Through this, proponents support the prescription of exercises into pain for chronic musculoskeletal pain and disorders.31 37 38 We define ‘exercise into pain’ as a therapeutic exercise where pain is encouraged or allowed.

No previous systematic reviews have evaluated the effectiveness of exercises into pain for chronic musculoskeletal pain. Therefore the object of this review was to compare the effect of exercises into pain compared with non-painful exercises on pain, function or disability in patients with chronic musculoskeletal pain within randomised controlled

CrossMark

trials (RCTs), specifically exercises that were prescribed with instructions for patients to experience pain, or where patients were told it was acceptable and safe to experience pain, and to compare any difference in contextual factors and prescription parameters of the prescribed exercise intervention.

METHODS
This systematic review followed the recommendations of the PRISMA statement, and was registered with the International Prospective Register of Systematic Reviews (PROSPERO; http://www.crd.york.ac.uk/prospero/, reference CRD42016038882).

Search strategy
An electronic database search was conducted on titles and abstract from inception to October 2016 on the following databases: the Allied and Complimentary Medicine Database, the Cumulative Index to Nursing and Allied Health Literature, the Cochrane Library, Embase, Medline, SPORTDiscus and Web of Science. For the keywords and keywords search strategy used, please see table 1. The database searches were accompanied by hand searches of the reference list of included articles, and the grey literature and ongoing trials were searched using the following databases: OpenGrey, WHO International Clinical Trials Registry Platform, ClinicalTrials.gov and the bjsports-2016-097383 portfolio.

For inclusion, the studies had to meet the following criteria: adults recruited from the general population with any musculoskeletal pain or disorder greater than 3 months; participants with pain suggestive of non-musculoskeletal pain, for example, headache, migraine, bowel/stomach pain, cancer, fibromyalgia, chest pain, and breathing difficulties were excluded. Studies had to have a primary treatment arm of therapeutic exercises that was advised to be purposely painful, or where pain was allowed or tolerated. The comparison group had to use therapeutic exercises that were pain-free. Included studies were required to report pain, disability or function. Studies had to be full RCT published in English. Studies that were not randomised or quasi-random were excluded.

Study selection
One reviewer (BES) undertook the searches. Titles and abstracts were screened by one reviewer (BES), with potential eligible papers retrieved and independently screened by two reviewers (BES and PH). Initial inclusion agreement was 81%, and using Cohen’s statistic method the kappa agreement was $k=0.47$, which is considered ‘fair to moderate’ agreement. All initial disagreements were due to intervention criteria, specifically the levels of pain during the therapeutic exercises in each intervention arm, and were resolved through consensus. Three trials needed further information with regard to their control exercise to ascertain if they met the inclusion criteria, and all three were contacted. All three responded with further information, and after discussion there was consensus to include two of the three trials.

Data extraction
The following data were extracted from the included articles: trial design, participant information, intervention and control exercise, setting, follow-up periods and outcome data. The data were independently extracted and transcribed to a standard table by one reviewer (BES), and then 25% of the data were independently extracted and transcribed to a standard table by one reviewer (BES), with potential eligible papers retrieved and independently screened by two reviewers (BES and PH). Initial inclusion agreement was 81%, and using Cohen’s statistic method the kappa agreement was $k=0.47$, which is considered ‘fair to moderate’ agreement. All initial disagreements were due to intervention criteria, specifically the levels of pain during the therapeutic exercises in each intervention arm, and were resolved through consensus. Three trials needed further information with regard to their control exercise to ascertain if they met the inclusion criteria, and all three were contacted. All three responded with further information, and after discussion there was consensus to include two of the three trials.

Quality assessment
Each included study was appraised independently by two reviewers (BES and PH) for methodological quality using the Cochrane risk of bias tool for randomised clinical trials. The tool was originally developed in 2008, and updated in 2011, and is based on seven key bias domains: sequence generation and allocation concealment (both within the domain of selection bias or allocation bias), blinding of participants and personnel (performance bias), blinding of outcome assessors (detection bias), incomplete outcome data (attrition bias) and selective reporting (reporting bias). For each domain the reviewers judged the risk of bias as ‘high’, ‘low’ or ‘unclear’. Percentage agreement between the two reviewers for the individual risk of bias domains for the Cochrane risk of bias tool was 86%, with a kappa of $k=0.76$, which is considered ‘substantial or good’, and disagreements were resolved through consensus.

We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system to rate the overall

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Search strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Randomised controlled trials as</td>
</tr>
<tr>
<td>2</td>
<td>Topic/</td>
</tr>
<tr>
<td>3</td>
<td>randomised controlled trial.pt</td>
</tr>
<tr>
<td>4</td>
<td>controlled clinical trial.pt</td>
</tr>
<tr>
<td>5</td>
<td>or/1-3</td>
</tr>
<tr>
<td>6</td>
<td>Exp Pain</td>
</tr>
<tr>
<td>7</td>
<td>Exp Musculoskeletal Disease</td>
</tr>
<tr>
<td>8</td>
<td>Exp Musculoskeletal Pain</td>
</tr>
<tr>
<td>9</td>
<td>Or/5-7</td>
</tr>
<tr>
<td>10</td>
<td>Rehabilitation</td>
</tr>
<tr>
<td>11</td>
<td>Bone</td>
</tr>
<tr>
<td>12</td>
<td>Joint</td>
</tr>
<tr>
<td>13</td>
<td>Muscle</td>
</tr>
<tr>
<td>14</td>
<td>Exp Exercise therapy</td>
</tr>
<tr>
<td>15</td>
<td>Physiotherapy</td>
</tr>
<tr>
<td>16</td>
<td>Physical Therapy</td>
</tr>
<tr>
<td>17</td>
<td>Physical-therapy</td>
</tr>
<tr>
<td>18</td>
<td>Exp Exercise Or/9-17</td>
</tr>
<tr>
<td>19</td>
<td>(exercise adj7 pain$).af</td>
</tr>
<tr>
<td>20</td>
<td>High load</td>
</tr>
<tr>
<td>21</td>
<td>Loaded$</td>
</tr>
<tr>
<td>22</td>
<td>Resistance$</td>
</tr>
<tr>
<td>23</td>
<td>Eccentric$</td>
</tr>
<tr>
<td>24</td>
<td>Concentric$</td>
</tr>
<tr>
<td>25</td>
<td>Weight loaded</td>
</tr>
<tr>
<td>26</td>
<td>Weight-loaded</td>
</tr>
<tr>
<td>27</td>
<td>Weight resistance</td>
</tr>
<tr>
<td>28</td>
<td>Weight-resistance</td>
</tr>
<tr>
<td>29</td>
<td>High-load</td>
</tr>
<tr>
<td>30</td>
<td>Heavy load</td>
</tr>
<tr>
<td>31</td>
<td>Heavy-load</td>
</tr>
<tr>
<td>32</td>
<td>Direction$ preference</td>
</tr>
<tr>
<td>33</td>
<td>Directional-preference</td>
</tr>
<tr>
<td>34</td>
<td>Or/19-33</td>
</tr>
<tr>
<td>35</td>
<td>4 and 8 and 18 and 34 (limited to English)</td>
</tr>
</tbody>
</table>
quality of the body of evidence in each pooled analysis.58 We did not evaluate the publication bias domain in this review as it is not recommended to assess funnel plot asymmetry with a meta-analysis of fewer than 10 trials.59 A GRADE profile was completed for each pooled estimate. Where only single trials were available, evidence from studies with <400 participants was downgraded for inconsistency and imprecision and rated as low-quality evidence. Three reviewers assessed these factors for each outcome and agreed by consensus (BES, PH and TOS).

The quality of evidence was defined as the following: (1) high quality—further research is unlikely to change our confidence in the estimate of effect; the Cochrane risk of bias tool identified no risks of bias and all domains in the GRADE classification were fulfilled; (2) moderate quality—further research is likely to have an important impact on our confidence in the estimate of effect, and one of the domains in the GRADE classification was not fulfilled; (3) low quality—further research is likely to have an important impact on our confidence and is likely to change the estimate; two of the domains were not fulfilled in the GRADE classification; and (4) very low quality—we are uncertain about the estimate; three of the domains in the GRADE classification were not fulfilled.60 61

Statistical analysis
Clinical heterogeneity was assessed through visual examination of the data extraction table on details related to participant characteristics, intervention, study design and process in the included studies. Based on this assessment, the reviewers judged there to be low clinical heterogeneity and accordingly it was appropriate to perform a meta-analysis where feasible. The primary outcome was a measure of pain, disability or function. As pain scores were reported on different scales, we used the standardised mean difference (SMD).62 We a priori defined effect size interpretation as 0.2 for a ‘small’ effect size, 0.5 for a ‘medium’ effect size and 0.8 for a ‘large’ effect size, as suggested by Cohen (1988).63 If data were not available, the associated corresponding author was contacted. Failing this, the mean and SD were estimated, assuming normal distribution, from medians and IQRs.64 Statis-
tical between-study heterogeneity was assessed with the I² statistic. We considered 0%–25% as low, 26%–74% moderate and 75% and over as high statistical heterogeneity.65 When outcomes presented with low statistical heterogeneity, data were pooled using a fixed-effects model.66 When analyses presented with moderate or high statistical heterogeneity, a DerSimonian and Laird random-effects model was adopted.67

All data analyses were performed using the OpenMetaAnalyst software.68

Sensitivity analysis
A sensitivity analysis was performed for the primary and secondary analyses using only trials that presented with a low risk of bias.56 In addition we carried out a sensitivity analysis to assess the impact of studies where mean and SD were estimated from medians and IQRs, and outcome measures of pain were pooled scores set within pain domains from patient-reported outcome measures, for example, the Shoulder Pain and Disability Index (SPADI).69

RESULTS
Study identification
The search results are presented in figure 1. The database search produced 9081 results, with no additional findings from reference list searches or unpublished searches. After duplicates were removed, 37 papers were appropriate for full-text review. After full-text review, 28 articles were excluded, 5 were due to participants not meeting the criteria, 26 because the intervention did not meet the criteria, 3 because of study design not meeting criteria, and 1 due to inappropriate outcome measures. Some articles were excluded for multiple reasons. Therefore nine articles were included in the final review. Of the included articles, there were two occurrences of the same trial reporting different time points over two publications.43 70–72

Characteristics of included trials
A summary of the characteristics and main findings of the included trials can be found in table 2. The two occurrences of the same trial reporting different time points over two articles were analysed as single trials to prevent multiplicity in analyses.43 70–72 All trials investigated home-based exercises, had a roughly even composition of women and men (46% women), with similar mean ages of participants (mean age 47, range 19–83). One trial included low back pain,43 72 three included shoulder pain,47 52 70 71 51 two included Achilles pain53 72 and one included plantar heel pain.51

Three trials used a Visual Analogue Scale to measure pain,43 70–72 74 two trials used the SPADI,47 52 one used the Knee Injury and Osteoarthritis Outcome Score (KOOS),73 and one used the Foot Function Index (FFI) including pain at worse and pain on first step on a numerical rating scale (0–10).51

Where pain outcomes were included within patient-reported outcome measures, these data were extracted.47 52 73 Two trials that used the SPADI had insufficient data in the publication to complete a meta-analysis for pain.47 52 and both were contacted and asked to supply pain domain data. Littlewood et al47 replied and provided all the available data; however, Maenhout et al66 did not respond. One trial reported outcomes in medians and IQRs,74 and was contacted and asked for further data. They
70 patients recruited from occupational healthcare services in Sweden (mean age 42, 56% female); inclusion criteria included: (a) adults with low back pain >3 months duration and (b) with or without leg pain.

Physiotherapy clinic, sports centre and home setting

1. n=35; group exercises based at a sports centre (5 participants in each group), with pain up to 50 mm Visual Analogue Scale acceptable, such that the pain subsided after each set of exercises; 12 treatment sessions over an 8-week period (weeks 1–4, 2 sessions per week; weeks 5–8, 1 session per week); 60 min in duration; no home exercises

2. n=35; pain-free individual exercises at a physiotherapy centre; 12 treatment sessions over an 8-week period (weeks 1–4, 2 sessions per week; weeks 5–8, 1 session per week); 20–30 min in duration; exercises involved improving control around joint neutral positions; in supine, four-point kneeling sitting, and/or standing positions; Plus home exercises, 10 repetitions 2–3x a day

Main outcome assessed at baseline, 2-month and 12-month follow-up was average pain on a Visual Analogue Scale (0–100mm) and Roland-Morris Disability Questionnaire.

Group 1: mean pain at baseline 43 (SD 24), 2 months 22 (SD 21), 12 months 24 (SD 27) and 24 months 27 (SD 27)

Group 2: mean pain at baseline 47 (SD 28), 2 months 30 (SD 26), 12 months 25 (SD 22) and 24 months 30 (SD 29)

Main outcome of Constant-Murley score (C-M) (0–100), along with shoulder assessment scores and pain scores taken at baseline, 3, 6 and 12 months, including pain at rest measured on Visual Analogue Scale (0–100 mm). Significantly more patients in the control group decided to have surgery (63%) than those in the specific exercise group (24%; p<0.0001).

Group 1: mean C-M at baseline 48 (SD 15), 3 months 72 (SD 19) and 12 months 83 (SD 14)

Group 2: mean C-M at baseline 43 (SD 15), 3 months 52 (SD 23) and 12 months 76 (SD 18)

Group 1: mean at rest at baseline 15 (SD 19), 3 months 10 (SD 14) and 12 months 2 (SD 6)

Group 2: mean at rest at baseline 20 (SD 21), 3 months 20 (SD 25) and 12 months 4 (SD 13)

Both groups had significant improvements in all outcomes at 3 months and 1 year follow-up. Significantly more patients in the control group decided to have surgery (63%) than those in the specific exercise group (24%; p<0.0001).

Main outcome of the Shoulder Pain and Disability Index (SPADI) (0–100) from baseline to all three follow-up points. There were no statistically significant differences between the groups across all the outcomes at 3, 6 or 12 months, (p=0.75, 0.19 and 0.32, respectively).
### Table 2  Continued

<table>
<thead>
<tr>
<th>Study characteristics</th>
<th>Participant characteristics</th>
<th>Intervention and setting</th>
<th>Outcome data/results</th>
</tr>
</thead>
</table>
| Maenhout et al (2013)22 | 61 patients recruited from a shoulder surgeon’s clinic in Belgium (mean age 39.8, 41% female); inclusion criteria included (a) adults with >3 months of shoulder pain, (b) painful arc, (c) 2 out of 3 impingement tests, (d) pain on palpation of rotator cuff tendons | Physiotherapy and home setting  
1. n=31; the same exercises as group 2, plus a heavy loaded eccentric exercise of abduction within the scapular plane; 3 sets of 15 repetitions, such that the patient experiences pain on the last set; up to 5/10 Visual Analogue Scale, such that the pain subsided by the following morning.  
2. n=30; pain-free, traditional rotator cuff exercises of internal and external rotation with a resisted rubber band; performed once a day, with 3 sets of 10 repetitions; both groups had exercise prescription and monitoring through 9 physiotherapy appointments over 12 weeks | Main outcome of the SPADI (0–100) at baseline, 6 weeks and 12 weeks  
Group 1: mean at baseline 44.3 (SD 11.5), 6 weeks 17.7 (SD 12.0) and 12 weeks 14.5 (SD 11.4).  
Group 2: mean at baseline 42.0 (SD 11.0), 6 weeks 25.4 (SD 11.9) and 12 weeks 17.0 (SD 11.4).  
In both groups pain and function, measured with the SPADI score, improved significantly over time (p<0.001). When comparing between groups, improvement of the SPADI score was not significantly different. |
| Nørregaard et al (2007)47 | 45 patients recruited from a clinic of sports medicine in Denmark (mean age 42, 49% female); inclusion criteria included (a) adults with Achilles pain >3 months of shoulder pain, (b) local thickening >2 mm on ultrasound, (c) diffuse posterior ankle pain | Sports medicine clinic and home setting  
1. n=21; information leaflet with home exercise programme on; to be performed twice a day, for 12 weeks; 1 follow-up appointment at 3 months; 3 sets of 15 repetitions of eccentric calf exercises, with knee straight and semi-flexed; patients told to expect pain during the exercises, but to avoid increasing daily pain or morning stiffness  
2. n=24; information leaflet with home exercise programme on; to be performed twice a day, for 12 weeks; 1 follow-up appointment at 3 months; pain-free standing stretches for gastrocnemius and soleus; 5 repetitions of 30 s each | Outcome measures were tenderness on palpation, ultrasound and pain, as measured by the Knee Injury and Osteoarthritis Outcome Score (KOOS) (0–4) and patient’s global assessment; follow-up was at baseline, 3, 6, 9, 12 weeks and 1 year  
Group 1: mean pain domain from KOOS at baseline 1.6 (SD 0.6), 3 weeks 0.1 (SD 0.1), 6 weeks 0.3 (SD 0.1), 9 weeks 0.4 (SD 0.2), 12 weeks 0.4 (SD 0.2) and 1 year score was 1.0 (SD 0.2)  
Group 2: mean pain domain from KOOS at baseline 1.6 (SD 0.6), 3 weeks 0.2 (SD 0.1), 6 weeks 0.3 (SD 0.1), 9 weeks 0.3 (SD 0.2), 12 weeks 0.4 (SD 0.2) and 1 year score was 0.7 (SD 0.2)  
There were significant improvements in all dimensions of the KOOS compared with baseline, with no differences between group differences. |
| Rathleff et al (2015)51 | 48 patients recruited from a university hospital, regional hospital and private clinic in Denmark (mean age 46, 66% female); inclusion criteria included (a) adults with plantar fasciitis >3 months, (b) pain on palpation, (c) diffuse posterior ankle pain | Home base exercises  
1. n=24; information leaflet, heel inserts and a prescription of a high-load strengthening programme; consisting of single calf raises with a towel rolled up under the toes for maximum toe extension, activating the windlass mechanism; each calf raises was 3 s up, 2 s pause, 3 s down; weight was added in rucksacks, starting at 1 repetition maximum for three sets, and slowly progressed over 3 months; patients were advised to perform the exercise every other day; exercises were allowed to be painful, with no postincrease in pain.  
2. n=24; information leaflet, heel inserts and a prescription of pain-free* plantar-specific stretches; patients were asked to stretch the plantar fascia in a cross-legged position by extending their toes, hold for 10 s, 10 times, 3× a day for 3 months | Primary outcome was Foot Function Index at 1, 3, 6 and 12 months, including pain at worse and pain on first step on a numerical rating scale (0–10).  
Mean scores for group 1 pain at worse at baseline was 7.9 (SD 1.7), 1 month 6.1 (95% CI 5.1 to 7.2), 3 months 3.5 (95% CI 2.3 to 4.7), 6 months 2.5 (95% CI 1.4 to 3.6) and 12 months 2.9 (95% CI 1.7 to 4.0). Mean scores for group 2 pain at worse at baseline was 7.5 (SD 1.6), 1 month 6.1 (95% CI 5.2 to 7.1), 3 months 6.1 (95% CI 4.4 to 7.7), 6 months 3.4 (95% CI 2.0 to 4.7) and 12 months 1.8 (95% CI 0.7 to 3.0).  
At 3 months group 1 had significantly lower pain scores than group 2 (p<0.05). At months 1, 6 and 12, there was no significant difference between groups. |
| Silbernagel et al (2001)74 | 40 patients recruited from mailings to hospitals, clinics and sports clubs in Sweden (mean age 45, 23% female); inclusion criteria included (a) adults with Achilles pain >3 months | Clinic and home setting  
1. n=22; progressive eccentric exercise programme to be performed 2× a day, plus three sets of six different stretching exercises, 20× each, as well as balance, toe/heel walking exercises; weekly physiotherapy contact for 12 weeks; pain was allowed during the exercises up to 5/10 Visual Analogue Scale, such that the pain subsided by the following morning with no morning stiffness  
2. n=18; 3× a day of regular concentric and eccentric calf strengthening plus two sets of the stretching exercises from group 1. Physiotherapy 3–5 x during the 12 weeks. Exercises must be pain-free. | Outcomes of pain on palpation (Visual Analogue Scale) (0–100 mm) taken at baseline, 6 weeks, 3 and 6 months. Other outcomes included pain on walking and pain on stairs (yes/no), various objective measures, plus a non-validated functional questionnaire  
Median ±IQR scores for pain on palpation for group 1 at baseline was 49±26.2, 6 weeks 40±27.5, 3 months 35±24.8 and 6 months 21±20.  
Median ±IQR scores for group 2 at baseline was 27±21.5, 6 weeks 20±20, 3 months 31±26 and 6 months 9±17.5.  
There was a significant decrease in pain on palpation in both treatment groups; no significant differences between groups were seen. |

*Information not within publication, authors contacted for clarification.  
ROM, range of motion.
Narrative synthesis of disability and function outcomes

Of the seven trials, six reported some form of patient-reported outcome measure of disability or function. One reported the Roland-Morris Disability Questionnaire, one reported Constant-Murley and the Disabilities of the Arm Shoulder and Hand score, two reported the SPADI, one reported the KOOS, and one reported the FFI. With the exception of Rathleff et al, there was clinically significant improvements in all outcomes, with no clear superiority. At 3-month follow-up for Rathleff et al, the intervention group had a statistically significant lower FFI than the control group (p=0.016). At 1, 6 and 12 months, there were no differences between groups (p>0.34).

Contextual factors

With regard to the parameters of pain in the exercise intervention, the participants were advised to adhere to, each trial gave different instructions, the key differences being if pain was allowed, or recommended. In addition other differences were if an acceptable level of pain measured on a pain scale was advised, and a time frame for the pain to subside by, for instance, if the pain had to subside immediately, by the next session, or by the next day. Clinically significant improvements in patient-reported outcome measures were reported across all interventions and control exercises, and all time points. It is not clear from the data if one approach was superior to the others.

Meta-analysis of pain

Short-term results

Six trials with 385 participants reported post-treatment effect on pain. Combining the results of these trials demonstrated significant benefit (SMD) of exercises into pain compared with pain-free exercises for musculoskeletal pain in the short term, with a small effect size of −0.28 (95% CI −0.49 to −0.08; figure 4). Statistical heterogeneity was negligible, I²=0%. The quality of evidence (GRADE) was rated as ‘low quality’ due to trial design, low participant numbers (table 3).

For sensitivity analysis in the short term, we repeated the meta-analysis, removing two trials that used a patient-reported outcome measures index and had high dropout rates, and the Silbernagel et al trial where the mean and SD were estimated from medians and IQRs. The results of the data synthesis produced very similar results, with a small effect size of −0.27 (95% CI −0.54 to −0.05), with low statistical heterogeneity of I²=22%. The quality of evidence (GRADE) was rated as ‘moderate quality’ due to low participant numbers (table 3).

Medium-term results

In the medium-term follow-up, meta-analysis demonstrated significant benefit (SMD) for exercises into pain compared with pain-free exercises for musculoskeletal pain, with a medium effect size of −0.59 (95% CI −1.03 to −0.15) (see figure 5). The statistical heterogeneity was moderate, I²=50%. The quality of evidence (GRADE) was rated as ‘low quality’ due to trial design and low participant numbers (table 3).

Sensitivity analysis was not possible for medium-term results as two trials were excluded, one for using a patient-reported outcome measures index, and one due to means and SD being estimated from medians and IQRs. The one remaining trial showed no significant difference in the medium term. The quality of evidence (GRADE) was rated as ‘low quality’ due to it being only from a single trial (table 3).

Trial quality and bias

The two papers reporting long-term outcomes for the trials that reported different time points made reference to the short-term outcome papers with regard to design parameters; therefore, trial quality and bias were assessed accordingly.

No trial had greater than three ‘high risk’ of bias scores for a domain (figure 2).

The greatest risk of bias was with the blinding of participants and personnel (100%) (figure 3). The greatest amount of uncertainty was with regard to selective reporting bias, as many of the trials failed to include trials register details, or protocol details (44%). Other common areas of bias with the included trials were with attrition bias, one trial failed to adequately describe attrition, and two trials had large dropout rates; however, Littlewood et al received a ‘low risk’ score as their participant attrition was balanced across the intervention and control groups, and an intention-to-treat analysis was performed. The risk of bias assessment tool highlights common trial write-up errors, with a number of papers failing to give an appropriate level of detail to adequately assess selection bias risk (33%).

Meta-analysis of pain

Short-term results

Six trials with 385 participants reported post-treatment effect on pain. Combining the results of these trials demonstrated significant benefit (SMD) of exercises into pain compared with pain-free exercises for musculoskeletal pain in the short term, with a small effect size of −0.28 (95% CI −0.49 to −0.08; figure 4). Statistical heterogeneity was negligible, I²=0%. The quality of evidence (GRADE) was rated as ‘low quality’ due to trial design and low participant numbers (table 3).

For sensitivity analysis in the short term, we repeated the meta-analysis, removing two trials that used a patient-reported outcome measures index and had high dropout rates, and the Silbernagel et al trial where the mean and SD were estimated from medians and IQRs. The results of the data synthesis produced very similar results, with a small effect size of −0.27 (95% CI −0.54 to −0.05), with low statistical heterogeneity of I²=22%. The quality of evidence (GRADE) was rated as ‘moderate quality’ due to low participant numbers (table 3).

Medium-term results

In the medium-term follow-up, meta-analysis demonstrated significant benefit (SMD) for exercises into pain compared with pain-free exercises for musculoskeletal pain, with a medium effect size of −0.59 (95% CI −1.03 to −0.15) (see figure 5). The statistical heterogeneity was moderate, I²=50%. The quality of evidence (GRADE) was rated as ‘low quality’ due to trial design and low participant numbers (table 3).

Sensitivity analysis was not possible for medium-term results as two trials were excluded, one for using a patient-reported outcome measures index, and one due to means and SD being estimated from medians and IQRs. The one remaining trial showed no significant difference in the medium term. The quality of evidence (GRADE) was rated as ‘low quality’ due to it being only from a single trial (table 3).
Long-term results
In the long term follow-up, meta-analysis demonstrated no statistical difference between exercises into pain and pain-free exercises, with an effect size of 0.01 (95% CI −0.14 to 0.40) (figure 6). The statistical heterogeneity was high, I²=70%. The quality of evidence (GRADE) was rated as ‘very low quality’ due to trial design, heterogeneity and low participant numbers (table 3).

For sensitivity analysis in the long term, we repeated the meta-analysis, removing the two trials that used a patient-reported outcome measures index.52 73 The results of the data synthesis found no statistical difference between exercises into pain and pain-free exercises, with an effect size of 0.13 (95% CI −0.14 to 0.40). The statistical heterogeneity was negligible, I²=0%. The quality of evidence (GRADE) was rated as ‘moderate quality’ due to low participant numbers (table 3).

DISCUSSION
Summary of main findings
There was a significant short-term benefit for exercises into pain over pain-free exercises for patient-reported outcomes of pain, with a small effect size and moderate quality of evidence. There appears to be no difference at medium-term or long term follow-up, with the quality of the evidence rated as moderate to low.

Clinical and research implications
Traditionally, healthcare practitioners have been reluctant to encourage patients to continue with exercise into pain when they are treating chronic musculoskeletal pain,76 with some research suggesting clinicians’ fear being the primary deterrent.77 The results of our systematic review show that there does not appear to be a scientific basis for this fear in relation to outcome measures of pain, and also potentially function and disability. This is an important point when considering what advice is given on any short-term exacerbations of musculoskeletal pain during physical activity or exercise by healthcare practitioners, particularly when physical inactivity is one of the 10 leading risk factors for death worldwide,78 and when an estimated €1.9 billion a year in healthcare and €9.4 billion a year in economic costs in the UK are attributable to physical inactivity.79

A theoretical rationale for a positive response to exercises into pain is the positive impact on the central nervous system.31 37 Specifically, the exercise addresses psychological factors such as fear avoidance, kinesiophobia and catastrophising, and is set within a framework of ‘hurt not equalling harm’, thus, in time, reducing the overall sensitivity on the central nervous system, with a modified pain output.31 37 The exercise-induced endogenous analgesia effect

Table 3 GRADE summary of findings table

<table>
<thead>
<tr>
<th>Follow-up</th>
<th>Number of participants(trials)</th>
<th>SMD (95% CI)</th>
<th>Design</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short term</td>
<td>385 (6 trials)</td>
<td>−0.28 (−0.49 to −0.08)</td>
<td>Limitations*</td>
<td>No inconsistency</td>
<td>No indirectness</td>
<td>Imprecision†</td>
<td>Low</td>
</tr>
<tr>
<td>Medium term</td>
<td>173 (3 trials)</td>
<td>−0.59 (−1.03 to −0.15)</td>
<td>Limitations*</td>
<td>No inconsistency</td>
<td>No indirectness</td>
<td>Imprecision†</td>
<td>Low</td>
</tr>
<tr>
<td>Long term</td>
<td>345 (5 trials)</td>
<td>0.01 (−0.39 to 0.41)</td>
<td>Limitations*</td>
<td>Inconsistency†</td>
<td>No indirectness</td>
<td>Imprecision†</td>
<td>Very low</td>
</tr>
</tbody>
</table>

Sensitivity analysis

<table>
<thead>
<tr>
<th>Follow-up</th>
<th>Number of participants(trials)</th>
<th>SMD (95% CI)</th>
<th>Design</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short term</td>
<td>215 (3 trials)</td>
<td>−0.27 (−0.54 to −0.05)</td>
<td>No limitations</td>
<td>No inconsistency</td>
<td>No indirectness</td>
<td>Imprecision†</td>
<td>Moderate</td>
</tr>
<tr>
<td>Medium term</td>
<td>40 (1 trials)</td>
<td>−0.32 (−0.95 to 0.31)</td>
<td>No limitations</td>
<td>Inconsistency†</td>
<td>No indirectness</td>
<td>Imprecision†</td>
<td>Low</td>
</tr>
<tr>
<td>Long term</td>
<td>215 (3 trials)</td>
<td>0.13 (−0.14 to 0.40)</td>
<td>No limitations</td>
<td>No inconsistency</td>
<td>No indirectness</td>
<td>Imprecision†</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

* Lack of blinding of participants and personnel, attrition bias, unable to adequately assess selection bias risk.
† Large statistical heterogeneity; I²=70%.
‡ Only single trial available, ≤400 participants therefore downgraded for inconsistency and imprecision.
§ Short term, ≤3 months; medium term, >3 and <12 months; long term, ≥12 months.

High quality: further research is unlikely to change our confidence in the estimate of effect.
Moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect.
Low quality: further research is very likely to have an important impact on our confidence in the estimate of effect.
Very low quality: we are uncertain about the estimate.
GRADE, Grading of Recommendations Assessment, Development and Evaluation; SMD, standardised mean difference.
is thought to occur due to a release of endogenous opioids and activation of spinal inhibitory mechanisms.80–84 However, a recent systematic review has established that no firm conclusions could be reached about pain modulation during exercise therapy for chronic musculoskeletal pain.85 Indeed one experimental study has shown a dysfunction of endogenous analgesia in patients with musculoskeletal pain,86 and therefore exercising non-painful body parts with patients with chronic musculoskeletal pain has been recommended.87 However, it is worth noting that empirical data within this field are greatly lacking, and this systematic review shows that painful exercises may even improve the clinical outcomes. Additionally, exercise prescription in the included trials was primarily based on strength and conditioning principles, with the exception of Littlewood et al.52 suggesting a tissue-focused approach, and therefore could still have been giving a ‘hurt is harm’ message to the majority of participants.

Significant improvements in patient-reported pain can be achieved with a range of contextual factors, such as varying degrees of pain experiences and postrecovery time for therapeutic exercise. In addition to the aspect of pain, an important difference between the intervention arm and the control arm is the higher loads, or levels of resistance, employed with the exercises into pain, and it is unknown if the difference in responses can be attributable to these two elements of the different exercise programmes. Research has shown a ‘dose response’ to exercise for musculoskeletal pain—the more incremental exercise (with appropriate recovery period) a person does the greater his/her improvements in pain88–90; the short-term benefits of exercises into pain over pain-free exercises could be explained by this dose effect, or response to load/resistance. However to our knowledge the optimal ‘dose’ of therapeutic exercise for musculoskeletal pain has not been established. Furthermore, little is known if it is possible or appropriate to identify individuals most suitable to exercise interventions.

Our review only investigated patient-reported outcome measures of pain and function/disability. It has been hypothesised that exercise therapy, where it has been advised that the experience of pain is safe and allowed, may address other patient-reported outcome measures—fear avoidance, self-efficacy and catastrophising beliefs37 38—and therefore may lead to improvements in functional outcome measures—fear avoidance, self-efficacy and catastrophising beliefs37 38—and therefore may lead to improvements in function, quality of life and disability, despite pain levels. Unfortunately none of the trials included in this review recorded the level of pain patients actually experienced during their exercise programme, preventing any detailed attempt to fully explain any mechanisms of effect. This aspect of exercise prescription clearly warrants further investigation in relation to chronic musculoskeletal pain. Any future trials should consider the role of pain with exercises and clearly define the parameters employed to ensure translation of findings into practice and further evaluation of optimal ‘dosage’.

Strengths and limitations of included trials

We chose not to perform subgroup analyses by anatomical region and/or tissue structures. The labelling of musculoskeletal structures as sources of pain has been debated for many years, with polarising opinions.91 92 However, the diagnostic labelling of patients into tissue-specific pathology characteristically suffers from poor reliability and validity.93–98 A strength of this review is that despite the trials including subjects suffering from musculoskeletal pain at different body locations, there exists low statistical heterogeneity at short-term follow-up and for the sensitivity analyses carried out.

The overall quality of the included papers can be considered relativity high, with only three domains in the Cochrane risk of bias tool (disregarding blinding of participants) demonstrating clear risk of bias across all domains for all trials. However taking into account other factors assessed with the GRADE analysis, the quality of the evidence was rated as moderate to low. Therefore our results can be considered to have moderate to low internal validity, with future research likely to alter our conclusions.

The main source of bias within the included trials were blinding; no trial blinded the participants. Knowledge of group assignment may affect participants’ behaviour, for example with patient-reported outcome measures such as pain scales or compliance with therapy interventions.99 However, it is accepted that blinding in physiotherapy and physical intervention trials is difficult to achieve.24 Another limitation of the included trials is the high level of attrition suffered by some of the trials in both treatment arms. For example Littlewood et al.52 suffered from 51% dropout at 12-month follow-up. A high level of attrition can overestimate the treatment effect size and could bias the results of our meta-analysis. However, we minimised the risk of bias on our results by conducting a sensitivity analysis on trials with a large dropout, identified using the Cochrane risk of bias tool and assessed level of evidence using the GRADE classification.

Limitations of this review

For pragmatic reasons one reviewer screened titles and abstracts. An extensive literature search was carried out, with two reviewers independently screening full texts for inclusion, and a sample of the data extraction independently verified. Additionally an attempt was made to retrieve unpublished trials; however, it may be that not all trials were retrieved, particularly considering we did not search for papers published in languages other than English and US spelling was used in the search terms. This review excluded trials where participants had a diagnosis of more widespread pain disorders like fibromyalgia.

Figure 5  Forest plot of exercises into pain versus pain-free exercises—medium term. Negative values favour painful intervention, whereas positive favour pain-free.

Figure 6  Forest plot of exercises into pain versus pain-free exercises—long term. Negative values favour painful intervention, whereas positive favour pain-free. AMED, Allied and Complementary Medicine Database; CINAHL, Cumulative Index to Nursing and Allied Health Literature.
CONCLUSION
The results of this systematic review indicates that protocols using exercises into pain offer a small but significant benefit over pain-free exercises in the short term, with moderate quality of the evidence for outcomes of pain in chronic musculoskeletal pain in adults. There appears to be no difference at medium-term or long-term follow-up, with moderate to low quality of evidence, demonstrating pain need not be ruled out or avoided in adults with chronic musculoskeletal pain.

What are the findings?

► Protocols using exercises into pain for chronic musculoskeletal pain offer a small but significant benefit over pain-free exercises in the short term.

► Adults with musculoskeletal pain can achieve significant improvements in patient-reported outcomes with varying degrees of pain experiences and postrecovery time with therapeutic exercise.

► Pain during therapeutic exercise for chronic musculoskeletal pain need not be a barrier to successful outcomes.

► Protocols using exercises into pain may have higher loads and dose of exercise.

Correction notice This paper has been amended since it was published Online First. The authors have noticed that figure 4 was a duplication of figure 6. The correct figure 4 has now been uploaded.

Contributors BE5 was responsible for conception and design, publication screening, acquisition of data, analysis and interpretation, and drafting and revising the manuscript. PH was responsible for conception and design, publication screening, acquisition of data, data interpretation, and reviewing and revising the manuscript. TDS was responsible for conception and design, data interpretation, and reviewing and revising the manuscript. All authors were involved in interpretation, reviewing revisions to the manuscript and final approval of the version to be published. All have read and approved the final version.

Funding This report is an independent research arising from a Clinical Doctoral Research fellowship, Benjamin E Smith, ICA-CRF-2015-01-002, supported by the National Institute for Health Research (NIHR) and Health Education England (HEE).

Disclaimer The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR, HEE or the Department of Health.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

42 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.


59 ISCA/Cebro. The economic cost of physical inactivity in Europe. *ISCA / Cebro Rep Published Online First*: 2015.


